Anomaly Detection and Characterization in Spatial Time Series Data: A Cluster-Centric Approach

Hesam Izakian, *Student Member, IEEE*, and Witold Pedrycz, *Fellow, IEEE*

***Abstract*—Anomaly detection in spatial time series (spatiotemporal data) is a challenging problem with numerous potential applications. A comprehensive anomaly detection approach not only should be able to detect and identify the emerging anomalies but has to characterize the essence of these anomalies by visualizing the structures revealed within data in a way that is understandable to the end-user as well. In this paper, we consider fuzzy c-means (FCM) as a conceptual and algorithmic setting to deal with the problem of anomaly detection. Using a sliding window, the time series are divided into a number of subsequences, and the available spatiotemporal structure within each time window is discovered using the FCM method. In the sequel, an anomaly score is assigned to each cluster, and using a fuzzy relation formed between revealed structures, a propagation of anomalies occurring in consecutive time intervals is visualized. To illustrate the proposed method, several datasets (synthetic data, a simulated disease outbreak scenario, and Alberta temperature data) have been investigated.**

***Index Terms*—Anomaly detection, anomaly propagation, fuzzy c-means (FCM), fuzzy relation, reconstruction criterion, spatial time series data.**

I. Introduction

SPATIAL time series data are commonly encountered in numerous application areas such as aerology, agriculture, and medical science. In these types of data, each datum is composed of two components, namely, a spatial part comprising location information (e.g., x–y coordinates or latitude–longitude pairs) and temporal part including one or more time series describing some temporal phenomena reported in successive time steps. Anomaly detection in these types of data refers to detecting any *unexpected* changes in a subsequence of a set of spatially adjacent time series. This problem occurs in numerous application areas. For example, aerologists are interested to detect anomalies in climate patterns to predict future consequences. Experts in health surveillance systems are interested in detecting anomalies of disease incidence to control possible future outbreaks. A general framework for sequence-based anomaly detection involves using a sliding window to generate a set of subsequences and then determining those subsequences which exhibit the highest differences in comparison with the others as anomalies [10]. Three main scenarios can be distinguished here.

1) In the simplest case, there is only a single time series and the objective is to find a subsequence in this time series showing the highest difference from all other subsequences of this time series.
2) In the second case, there are a number of time series, and the objective is to find a subsequence of a set of time series exhibiting high dissimilarity from other parts of data. In fact, in this problem, we consider a set of time series at the same time to quantify dissimilarity.
3) Finally, the last case is concerned with spatial time series. Here, we have to tackle another constraint present in the problem, which deals with the spatial position of time series when defining dissimilarity between subsequences. The objective is to find subsequences of a set of spatially neighboring time series, showing a large difference from the other parts of data.

Anomaly detection in spatial time series is a challenging problem since defining a spatial neighborhood of a set of time series encountering some unexpected changes is not straightforward. Moreover, the definition of unexpected changes may not be precise since we do not know what type of changes is expected and what type is not. Because of the problems identified above, using a brute force method to consider all possible states (spatially adjacent time series with subsequences encountering unexpected changes) is not efficient, and for large size of data, such an approach is not feasible at all.

In this paper, we propose a clustering-based technique for anomaly detection and characterization in spatial time series data. Clustering is an efficient instrument to visualize and understand the structure present within data. Fuzzy c-means (FCM) proposed by Dunn [1] and Bezdek [2] is one of the most commonly used clustering techniques in fuzzy set community in which, instead of assigning data to individual cluster, the Boolean-like nature of assignment is relaxed by admitting membership grades assuming values in the unit interval. This technique has been employed successfully in clustering spatial, temporal, and spatiotemporal data [3], [9], [16], [28], [30], [31].

The main objectives of the proposed study are as follows:

1) revealing and visualizing the available structure within spatiotemporal data in successive time steps (intervals);
2) assigning an anomaly score to each revealed spatiotemporal cluster and quantifying the level of occurred unexpected changes in the structure of data;

Manuscript received July 28, 2013; revised November 6, 2013; accepted December 25, 2013. Date of publication January 24, 2014; date of current version November 25, 2014. This work was supported in part by Alberta Innovates—Technology Futures and Alberta Advanced Education and Technology, Natural Sciences and Engineering Research Council of Canada, and the Canada Research Chair Program.

H. Izakian is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2V4 (e-mail: izakian@ualberta.ca).

W. Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2V4, with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia, and also with the System Research Institute, Polish Academy of Sciences, Warsaw 00-716, Poland (e-mail: wpedrycz@ualberta.ca).

Digital Object Identifier 10.1109/TFUZZ.2014.2302456

3) constructing relations between clusters present in successive time windows to visualize and quantify anomaly propagation over time.

In more detail, a sliding window to generate a set of subsequences over the spatial time series data has been considered and FCM is employed to visualize the structure available within the resulting spatiotemporal subsequences. For each single temporal subsequence, an anomaly score is determined by taking into account its historical behavior, and the estimated anomaly scores are aggregated within each cluster. Furthermore, using a gradient-based optimized fuzzy relation, the revealed clusters in consecutive time steps are mapped to each other. Although one may compare the revealed clusters in different time windows using some techniques reported in the literature [25], [32], [35], [36], the proposed fuzzy relation technique shows some advantages as a vehicle to visualize the propagation of anomalies over time and space.

Detecting anomalies in spatial time series data using spatiotemporal clustering is a novel idea proposed in this study. The method introduced here visualizes structures present in different time windows to make them understandable to the end-user. Moreover, using the fuzzy relation-based model of relationships, the revealed clusters in spatiotemporal subsequences can be tracked from the structure identified in the past, leading to a thorough temporal analysis of propagation of anomalies.

This paper is structured into seven sections. In Section II, we review the related works available in the literature. In Section III, the problem is formulated and the proposed idea is described. In Section IV, a spatiotemporal clustering approach is discussed. Section V introduces anomaly detection in spatiotemporal data and describes a characterization technique of anomalies. Section VI reports the experimental studies, while conclusions are covered in Section VII.

II. Literature Review

Most studies reported in the literature deal with anomaly detection in time series; however, they do not consider spatiotemporal data. These methods can be divided into a set of categories comprising similarity-based, clustering-based, classification-based, and modeling-based techniques.

Similarity-based techniques: One straightforward method for anomaly detection in time series data is to assign an anomaly score to each time series according to its similarity to the other time series existing in dataset. A suitable distance function or resemblance measure can be considered to be a similarity/dissimilarity measure. In [14], an anomaly detection technique has been proposed for light curves in catalogues of periodic variable stars. By considering N time series $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ present in the dataset, the anomaly score of a certain time series $\boldsymbol{x}_i$ was expressed as

$$\Re^2_{\boldsymbol{x}_i} = \frac{1}{N-1} \sum_{\substack{j=1,2,\ldots,N \\ j \neq i}} r^2(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{1}$$

where $r(\boldsymbol{x}_i, \boldsymbol{x}_j)$ was the cross correlation present between time series $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. A lower value of the score of the cross correlation corresponds to a higher level of anomaly of the time series. Keogh *et al.* [10] proposed detecting the maximal different subsequence within a longer time series (called discords) using a 1-nearest neighbor technique. Formally, by considering a time series of length $m, t_1, t_2, \ldots, t_m$, a discord with length n $(n < m)$ was a subsequence $t_p, t_{p+1}, \ldots, t_{p+n-1}$ for $1 \leq p \leq m - n + 1$ with the highest distance to its none overlapping nearest neighbor. By representing time series using a symbolic aggregate approximation, the authors proposed an algorithm that was faster than the brute force method. In [24], a compression-based dissimilarity measure was proposed for anomaly detection in time series data. For two sequences $\boldsymbol{x}$ and $\boldsymbol{y}$, the dissimilarity measure was expressed as

$$\mathrm{CDM}(\boldsymbol{x}, \boldsymbol{y}) = \frac{C(\boldsymbol{xy})}{C(\boldsymbol{x}) + C(\boldsymbol{y})} \tag{2}$$

where $C(\boldsymbol{x})$ was size (given in bytes) of compressed file containing time series $\boldsymbol{x}$, and $C(\boldsymbol{xy})$ was size of compressed file containing concatenated sequences $\boldsymbol{x}$ and $\boldsymbol{y}$. Considering $\boldsymbol{x}$ to be a subsequence of a longer time series $\boldsymbol{y}$, CDM $(\boldsymbol{x}, \boldsymbol{y})$ was considered as anomaly score of $\boldsymbol{x}$.

In [21], a method is proposed to detect outliers in spatial, temporal, and spatiotemporal earth science data. Dimensionality reduction was performed during the preprocessing step using a L_2-norm, and a global outlier was estimated for each separate location and separate timestamp. A distance-based outlier detection approach has been considered for temporal part of data, and a neighborhood-based outlier detection method was exploited for spatial part of data. Moreover, spatially anomalous units encountering a high deviation from the historical trends are considered as spatiotemporal outliers.

Clustering-based techniques: Clustering is another method used for anomaly detection in time series data. In this method, time series are clustered using an appropriate clustering technique and the revealed cluster centers are exploited to assign an anomaly score to each time series. In [15], an FCM clustering was used to cluster a set of time series data, and a reconstruction criterion [4] was employed to reconstruct time series with the aid of the revealed cluster centers. Finally, a reconstruction error was used to assign an anomaly score to each time series. In [17], a set of training sequences was clustered using a k-medoids clustering, and for each test sequence, its inverse similarity to its closest medoid was considered as the anomaly score. Formally, an anomaly score $A(\boldsymbol{s}_q)$ was assigned to each test sequence $\boldsymbol{s}_q$ as

$$A(\boldsymbol{s}_q) = \left(\min_{\forall i} (S(\boldsymbol{s}_q, \boldsymbol{v}_i)) \right)^{-1} \tag{3}$$

where $\boldsymbol{v}_i$ was the ith medoid, and $S(.)$ denotes a similarity measure computed for two sequences.

Classification-based techniques: Classification techniques also are in interest for anomaly detection in time series data. A common method in this category is to train a classifier using a set of training normal time series and then use a classifier to assign an anomaly score to each test time series. In [34], the time series data are projected onto a phase space, and then, novel events in time series are interpreted as outliers of normal

distribution of vectors in the phase space. A single-class support vector machine was employed as the outlier detector. Dasgupta and Forrest [33] proposed an anomaly detection inspired by the negative selection mechanism of the immune system. Normal data were considered as "self" and anomalies were considered as "nonself" patterns. Moreover, Gao *et al.* [29] proposed a neural network for event extraction in time series data. They showed that neural network could characterize the properties of homeostatic dynamics and model the dynamic relation between endogenous and exogenous variables in financial time series.

Modeling-based techniques: Time series modeling techniques form another group of anomaly detection approaches reported in the literature. An autoregressive (AR) model [23] is one of the commonly used techniques for this purpose. The AR model assumes that for a value of the time series in time t, x_t can be approximated using the values of its p values present in the previous time instants. Formally, we have

$$x_t = \sum_{i=1}^{p} q_i x_{t-i} + \varepsilon_t \tag{4}$$

where p is the order of the model, $q_i, i = 1, 2, \ldots, p$ are its parameters, and ε_t is white noise. Takeuchi and Yamanishi [7] proposed a two-stage time series learning model to detect change points in time series data. Considering $\{x_t | t = 1, 2, \ldots\}$ as the input time series, at the first step of the algorithm, a sequence of probability density functions $\{p_t | t = 1, 2, \ldots\}$ was constructed using an AR model, and for each point x_t in time series, a logarithmic loss score was calculated as $\text{Score}(x_t) = -\log p_{t-1}(x_t | x^{t-1})$. A higher score for x_t indicates that this point is an outlier with a higher likelihood. In the next step, using a sliding window, a new time series was constructed as an average of the calculated scores obtained in the previous step. The new time series was fitted again using an AR model and new loss scores were calculated for the new time series. Higher values of the scores for the points in the new time series indicate that they are change points with a higher probability. In [8], a self-organizing map was employed to characterize the time evolution in AR processes. The regions of the map that the AR process was expected to move are identified and the anomalous changes of the AR process has been detected. The method was applied to a real-world industrial process. In [5], multivariate time series are modeled using a weighted graph representation, where each node of the graph corresponds to a data point or a subsequence in a time series and each edge was weighted through a similarity measure between nodes. Considering that p is the number of variables in multivariate time series, the similarity between timestamps i and j in time series was calculated with the aid of the RBF function as follows:

$$K(i, j) = \exp\left[-\frac{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}{\sigma^2}\right]. \tag{5}$$

To calculate the connectivity of each node, the constructed graph was considered as a Markov chain with a transition matrix S, where the element in ith row and jth column denotes the transition probability from node i to j. The transition matrix was normalized and the connectivity value of each node was calculated and nodes with a low value of connectivity were considered as anomalies. Khatkhate *et al.* [11] proposed modeling time series data through a hidden Markov model (called the D-Markov machine model) from a symbolic representation. For each time epoch t_k, an anomaly measure was defined as

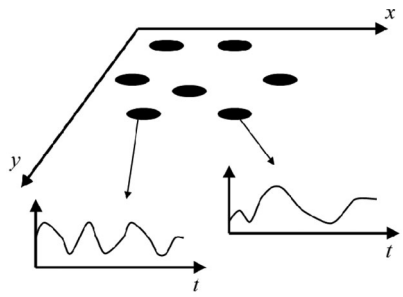

Fig. 1. Essence of spatial time series data.

$$\hat{M}(t_k) = \sum_{l=1}^{k} d(\boldsymbol{p}^l, \boldsymbol{p}^{l-1}) \tag{6}$$

where $d(.,.)$ was a distance function, while probability vectors $\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots$ are obtained at epochs $t_1, t_2, \ldots$ based on the respective time series data. In [13], a dynamic Bayesian network was employed to develop two automated anomaly detection techniques. These methods can be applied to single sensor data streams (called uncoupled detection), as well as several data streams at once (called coupled anomaly detection). The efficiency of the proposed methods was investigated for two wind speed data streams to perform a data quality assurance and control. Dereszynski and Dietterich [18] proposed a real-time data quality control in sensor networks. This method models the spatial relationships among sensors using a Bayesian network. To exploit the temporal correlations, the model was extended to a dynamic Bayesian network. It was able to detect failure observations and predict their true values. In [27], expectation-based scan statistics [26] were proposed in order to monitor a set of spatially located time series for detecting emerging spatial patterns. For this purpose, an expected number of events was calculated, and a set of spatial regions containing a significantly high number of events was detected. A survey of time series anomaly detection approaches can be found in [12].

III. Problem Formulation and the Proposed Idea

Let us consider N data $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ each comprising a spatial part, and a time series part. For the kth data $\boldsymbol{x}_k$, the concatenation $\boldsymbol{x}_k = [\boldsymbol{x}_k(s) \,|\boldsymbol{x}_k(t)\,]$ is realized, where $\boldsymbol{x}_k(s)$ represents its spatial part and $\boldsymbol{x}_k(t)$ refers to its time series part (or its representation). By considering r features (usually, $r = 2$) for the spatial part and q features for the time series part, altogether, we have the following representation for the kth data with dimensionality $n = r + q$:

$$\boldsymbol{x}_k = [x_{k1}(s), \ldots, x_{kr}(s) | x_{k1}(t), \ldots, x_{kq}(t)]^{\mathrm{T}}. \tag{7}$$

Fig. 1 visualizes the nature of the spatial time series data. There are a number of spatial points in x–y coordinates, and for each spatial point, there is one (or more) time series representing the measurements of a phenomenon in different time steps.

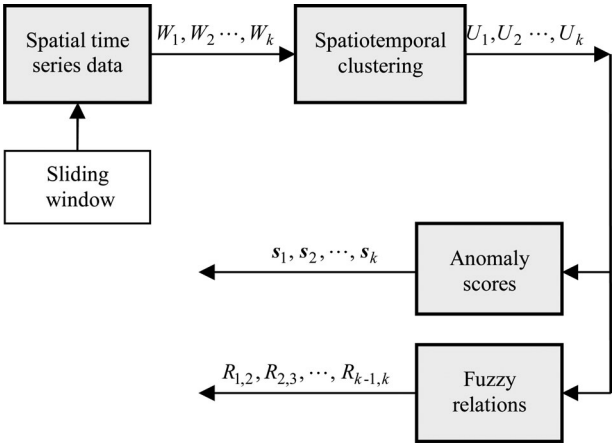

Fig. 2. Overall scheme of the proposed method for anomaly detection and characterization in spatial time series data.

These types of data are also referred to as georeferenced time series [20].

Our objective is to detect and characterize any unexpected changes in a subsequence of a set of spatially neighboring time series. Fig. 2 shows the overall scheme of the proposed method by presenting a bird's eye view at the introduced approach.

Let us briefly highlight the main processing realized here. At the first step, a sliding window moves across the time coordinate of data. Since there are N spatial time series data, the time window at each step includes N subsequences. By considering the spatial information and the generated subsequences, we form a set of spatiotemporal subsequences $W_1, W_2, \ldots, W_k$. In fact, the sliding window allows us to look at the data at different time intervals. At the second step, the available structure in each set of spatiotemporal subsequences $W_i, i = 1, 2, \ldots k$ is revealed using a spatiotemporal clustering approach proposed in [3]. The result of this step is a collection of partition matrices $U_1, U_2, \ldots, U_k$, each describing a set of clusters existing within the spatiotemporal subsequences. In the next step, the revealed partition matrices are exploited in two different ways. The first one is to assign an anomaly score to the revealed clusters in each time window, and the second way is to construct a fuzzy relation between clusters present in successive time steps to visualize the behavior of data in time.

IV. Clustering Spatiotemporal Data Using a Reconstruction Criterion

For the sake of completeness, we briefly recall the method in [3] for clustering spatiotemporal data. Let us consider N spatiotemporal data $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$. FCM aims to describe these N data by the use of c information granules (cluster centers) $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c$ and a partition matrix $U = [u_{ik}]$, $i = 1, 2, \ldots, c$, and $k = 1, 2, \ldots, N$. This structure arises as a result of the minimization of the following objective function:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m d^2(\boldsymbol{v}_i, \boldsymbol{x}_k) \tag{8}$$

where $m > 1$ is a fuzzification coefficient, and $d(.,.)$ is a distance function. In clustering spatiotemporal data, the challenging problem is to control the effect of each part of data (spatial and temporal components) in the clustering process. To deal with this problem, the following composite (aggregate) distance function was proposed:

$$d_\lambda^2(\boldsymbol{v}_i, \boldsymbol{x}_k) = \|\boldsymbol{v}_i(s) - \boldsymbol{x}_k(s)\|^2 + \lambda \|\boldsymbol{v}_i(t) - \boldsymbol{x}_k(t)\|^2, \lambda \geq 0 \tag{9}$$

where $\|.\|$ denotes Euclidean distance, $\boldsymbol{x}_k(s)$ is the spatial part of kth data, and $\boldsymbol{x}_k(t)$ stands for the time series part of kth data (or its representation). The parameter λ standing in the above distance function helps us strike a sound balance between the impact of the spatial and temporal components of the data in the clustering process. By assigning $\lambda = 0$, we remove the effect of temporal part of data and consider only a spatial part when forming clusters. Assigning higher value to λ increases the impact of the temporal part in the clustering process. By considering the distance (9), one calculates the cluster centers and the partition matrix in an iterative fashion as follows:

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^m \boldsymbol{x}_k}{\sum_{k=1}^{N} u_{ik}^m} \tag{10}$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_\lambda^2(\boldsymbol{v}_i, \boldsymbol{x}_k)}{d_\lambda^2(\boldsymbol{v}_j, \boldsymbol{x}_k)} \right)^{1/(m-1)}} \tag{11}$$

where $i = 1, 2, \ldots, c$, and $k = 1, 2, \ldots, N$. The coefficient λ plays an important role in the clustering process, and a suitable value of this coefficient has to be determined. In [3], the optimization of the value of this coefficient is guided by the two-optimization criteria, namely, a reconstruction criterion [4] and a prediction criterion [22]. Here, we consider the reconstruction criterion to determine the optimal value of λ. The essence of this criterion is to reconstruct the original spatiotemporal data using the estimated prototypes and partition matrix. We evaluate the quality of clusters in the process of information granulation and degranulation. The kth spatiotemporal object is reconstructed as

$$\hat{\boldsymbol{x}}_k = \frac{\sum_{i=1}^{c} u_{ik}^m \boldsymbol{v}_i}{\sum_{i=1}^{c} u_{ik}^m} \tag{12}$$

and a reconstruction error (that is a squared normalized Euclidean distance between the original and reconstructed spatiotemporal objects) is considered to evaluate the quality of clusters comes as follows:

$$E(\lambda) = \sum_{k=1}^{N} \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\|^2. \tag{13}$$

The lower the value of $E(\lambda)$, the better the reconstruction capabilities of the spatiotemporal clusters. To find an optimal value of λ, one may consider different values of this parameter in a range and select a value that minimizes (13). In other words, $\lambda_{\text{opt}} = \arg\min_\lambda E(\lambda)$. For more details, see [3].

V. Anomaly Detection and Characterization

Clustering spatiotemporal subsequences within different time windows leads to revealing a set of structures within data present at different time steps. Here, for each single subsequence inside a time window, an anomaly score is estimated based on its historical behavior, and then, the estimated anomaly scores are aggregated to determine an anomaly score for each cluster inside each time window.

Let us consider the jth time window W_j. Since there are N spatial time series in dataset, W_j contains N subsequences. To assign an anomaly score for each subsequence, there are a number of methods proposed in the literature and some of them were reviewed in Section II. The strategy to estimate an anomaly score for a subsequence depends on the nature of data and the application purpose. In this paper, the anomaly score of a subsequence is considered as the average squared Euclidean distance to its previous subsequences. Formally, considering $\boldsymbol{x}_{kj}$ to be a subsequence of spatial time series $\boldsymbol{x}_k$ falling within the window W_j, its anomaly score is expressed as follows:

$$f_{kj} = \frac{1}{j-1}\sum_{i=1}^{j-1} \|\boldsymbol{x}_{kj} - \boldsymbol{x}_{ki}\|^2. \tag{14}$$

The intuition behind this measure is that usually normal subsequences are very similar to the subsequences present in the previous time intervals. As the result, (14) generates a high score for anomalous subsequences, while normal subsequences exhibit a lower score.

After computing an anomaly score for each single subsequence inside W_j, the anomaly scores are aggregated to estimate an anomaly score for each cluster inside W_j. Assuming that U is the partition matrix resulting from clustering of spatiotemporal data corresponding to the time window W_j, the anomaly scores for the clusters located in W_j, $\boldsymbol{s}_j = \{s_i, i = 1, 2, \ldots, c_j\}$ can be estimated using

$$s_i = \frac{\sum_{k=1}^{N} u_{ik} f_{kj}}{\sum_{k=1}^{N} u_{ik}}, \quad i = 1, 2, \ldots, c_j \tag{15}$$

where c_j is number of clusters in W_j, and f_{kj} is the anomaly score corresponding to kth spatiotemporal data $\boldsymbol{x}_k$ inside time window W_j. Higher value of s_i indicates that the subsequences belonging to ith cluster of W_j are more anomalous. On the other hand, a lower value of s_i indicates that the subsequences corresponding to this cluster are similar to the subsequences in their previous time intervals, and as a result, the level of unexpected changes (anomalies) is lower.

Although detecting anomalous part of data is critical, in many applications analyzing and characterizing the detected anomalies (e.g., identifying the source of anomalies and visualizing anomaly propagation during time evolution) are equally important. We develop an approach to establish dependences among the revealed clusters within different time windows and quantify them in the form of a certain fuzzy relation.

The detected structures that are positioned in different time windows come in the form of a family of partition matrices; see Fig. 2. We are interested in analyzing the relationships available between these structures. As the FCM algorithm is initialized randomly, the order of clusters encountered in these matrices could be different for different runs and different time windows. Furthermore, there are different numbers of clusters obtained for various time windows. As the result, we have to form a fuzzy relation using which we map clusters present in some partition matrix U_1 to the clusters in U_2. The entries of this fuzzy relation describe degrees to which clusters in U_1 are related to clusters in U_2.

Let us assume the two partition matrices U_1 and U_2 of dimensionality $c_1 \times N$ and $c_2 \times N$, respectively, where N is number of data, c_1 is number of clusters in U_1, and c_2 is number of clusters in U_2. Each data object $\boldsymbol{x}_k$ is described in U_1 as a collection of membership degrees $\boldsymbol{ux}_k = [ux_{1,k}, ux_{2,k}, \ldots, ux_{c_1,k}]^{\mathrm{T}}$, $k = 1, 2, \ldots, N$, and similarly, each data object $\boldsymbol{y}_k$ is represented in U_2 through $\boldsymbol{uy}_k = [uy_{1,k}, uy_{2,k}, \ldots, uy_{c_2,k}]^{\mathrm{T}}, k = 1, 2, \ldots, N$. Our intent is to form a relational dependency that maps clusters present in U_1 onto the clusters occurring in U_2. We construct a fuzzy relation R of dimensionality $c_1 \times c_2$ whose entries are optimized in a way that the following performance index becomes minimized:

$$\begin{aligned} Q &= \sum_{k=1}^{N} \|\boldsymbol{ux}_k - R \circ \boldsymbol{uy}_k\|^2 \\ &= \sum_{k=1}^{N}\sum_{i=1}^{c_1} \left(ux_{i,k} - \max_{j=1,2,\ldots,c_2} (r_{i,j}\, t\, uy_{j,k}) \right)^2 \end{aligned} \tag{16}$$

where $R = [r_{i,j}], i = 1, 2, \ldots, c_1, j = 1, 2, \ldots, c_2$ is a fuzzy relation to be determined, $\circ$ denotes a sup-t composition with "t" being a t-norm, and $ux_{i,k}$ is the ith element of $\boldsymbol{ux}_k$. By considering a gradient descent optimization approach to minimize (16), we update the entries of R in an iterative fashion for $i = 1, 2, \ldots, c_1$, and $j = 1, 2, \ldots, c_2$. For the sth element of R, say $r_{s,t}$, we have

$$r_{s,t}(\text{iter}+1) = \left\langle r_{s,t}(\text{iter}) - \alpha \frac{\partial Q}{\partial r_{s,t}(\text{iter})} \right\rangle \tag{17}$$

where $\langle\ \rangle$ indicates that the resulting values of $r_{s,t}(\text{iter}+1)$ are confined to the [0,1] interval, α is a positive learning rate controlling intensity of learning, and iter stands for iteration index. By choosing the max–min composition operator (although the solution can be derived for different types of max-t and min-s compositions involving various t-norms and t-conorms), we have

$$\frac{\partial Q}{\partial r_{s,t}} = 2\left(\max_{j=1,2,\ldots,c_2} (\min(r_{i,j}, uy_{j,k})) - ux_{i,k} \right) \times \varphi_{s,t,i,k} \tag{18}$$

where

$$\varphi_{s,t,i,k} = \frac{\partial}{\partial r_{s,t}} \max_{j=1,2,\ldots,c_2} (\min(r_{i,j}, uy_{j,k})). \tag{19}$$

By virtue of the nature of the minimum and maximum operations, we have

$$\varphi_{s,t,i,k} = \begin{cases} 1, & \text{if } i = s \text{ and } r_{s,t} < uy_{t,k} \text{ and} \\ & \quad r_{s,t} > \max\limits_{\substack{j=1,2,\ldots,c_2,\\ j\neq t}} (\min(r_{s,j}, uy_{j,k})) \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Using the above optimization, the fuzzy relation R can be estimated. Each row in R corresponds with a given cluster in U_1 and each column in R stands for a cluster in U_2. As an example, let us assume that the following fuzzy relation is given:

$$R = \begin{bmatrix} 0.2 & 0.9 & 0.3 \\ 0.4 & 0.2 & 0.8 \\ 0.1 & 0.05 & 0.2 \end{bmatrix}. \quad (21)$$

This fuzzy relation indicates that the first cluster in U_1 is related with the first, second, and third clusters in U_2 with degrees of 0.2, 0.9, and 0.3, respectively. The second cluster in U_1 is related to the first, second, and third clusters in U_2 with degrees 0.4, 0.2, and 0.8, respectively. When a chain of structures in successive time steps (resulting from clustering data in different time windows) is considered, the above fuzzy relations characterize the behavior of clusters (linkages among them) and specify the origin of an arbitrary cluster.

VI. Experimental Studies

In this section, to illustrate the proposed approach, a synthetic dataset, as well as a simulated outbreak scenario and a simulated anomaly in the Alberta temperature dataset, has been studied.

A. Synthetic Dataset

A synthetic dataset with ten spatial time series labeled as "a" to "j" has been generated. Fig. 3(a) shows the spatial coordinates, and Fig. 3(b) shows the time series part of data.

The length of each time series is 90, and there are some visible changes in the temporal part of data points d, e, i, and j at 61st time moment. In fact, these points are located spatially in the topmost of y coordinate. Two time windows–one covering time steps from 1 to 50 (called W_1) and another one covering time steps from 41 to 90 (called W_2)–have been considered in this experiment. By concatenating the spatial part of data with the specified time series parts, two spatiotemporal subsequences are formed. Note that selecting the length and position of time windows is application dependent and the end-user may assign values for these parameters in an interactive manner when analyzing some initial results. To cluster the generated spatiotemporal data corresponding to time windows W_1 and W_2, the number of clusters is varied from 2 to 4, and the fuzzification coefficient m was set to 2.0. Different values of λ in range [0,100] have been considered leading to the optimal value of this parameter. For this value, the corresponding clusters have been selected.

Fig. 4 displays the values of the reconstruction error for $c = 3$ and different values of λ. For the segments W_1 and W_2, the optimal values of λ are 0.00 and 0.05, respectively.

Fig. 5 shows the revealed spatiotemporal clusters using the reconstruction criterion in the form of contour plot of member-

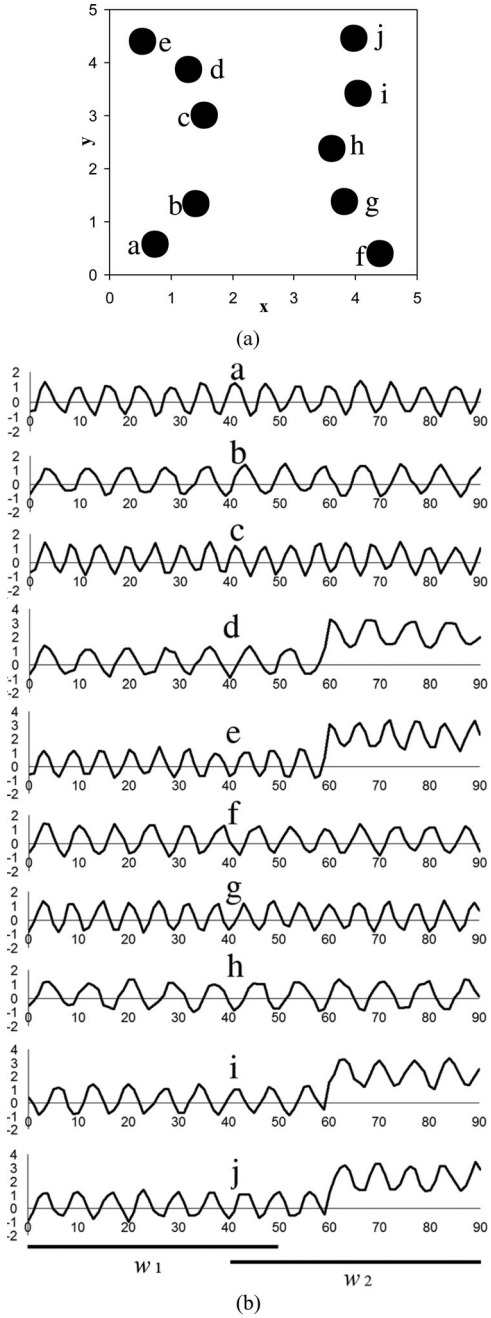

Fig. 3. Synthetic dataset. (a) Spatial part and (b) the associated time series.

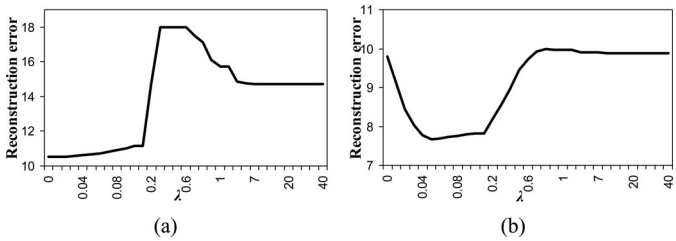

Fig. 4. Reconstruction error versus different values of λ for clustering spatiotemporal data for (a) W_1 and (b) W_2. The number of clusters was $c = 3$.

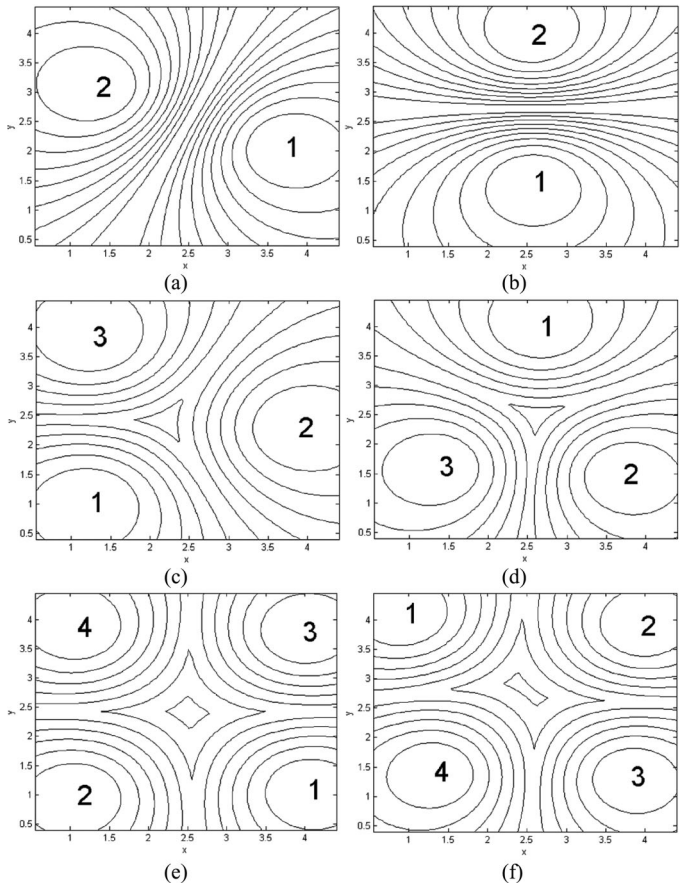

Fig. 5. Spatiotemporal clusters of the generated synthetic dataset for $c = 2$, 3, and 4. (a), (c), and (e) Time window W_1 and (b), (d), and (f) W_2.

TABLE I
ANOMALY SCORES OF SPATIOTEMPORAL CLUSTERS INSIDE TIME WINDOW W_2 FOR THE NUMBER OF CLUSTERS VARYING FROM 2 TO 4

Case study	Clusters			
	1	2	3	4
c=2	53.73	**125.75**	–	–
c=3	**130.33**	56.39	66.64	–
c=4	**111.47**	**139.88**	45.19	56.26

ship degrees for W_1 and W_2 and number of clusters $c = 2$, 3, and 4. For each time window, its corresponding subsequences in time series part along with the spatial part of data have been considered for clustering purposes. The number inside each cluster represents the order of this cluster in its corresponding partition matrix. There are some visible differences between clusters within time window W_1 and clusters inside W_2 because of some existing changes in time series part of data in steps 61–90.

Table I shows the values of the anomaly score [calculated using (14) and (15)] corresponding to each cluster inside time window W_2 for the number of clusters $c = 2$, 3, and 4. Since W_1 is the first time window and there are no historical data for it, the clusters inside this time window were considered as normal, and their anomaly score is set to 0. As shown in this table, the second cluster in Fig. 5(b), the first cluster in Fig. 5(d), and the first and second clusters in Fig. 5(f) exhibit a high anomaly score, indicating a high level of anomaly in the time series part of data in these clusters.

TABLE II
ESTIMATED FUZZY RELATIONS DESCRIBING TRANSITIONS FROM PARTITION MATRICES CORRESPONDING TO W_1 AND W_2; $c = 2$, 3, AND 4

clusters	W_1 to W_2
c=2	$\begin{pmatrix} \mathbf{0.58} & \mathbf{0.40} \\ \mathbf{0.42} & \mathbf{0.60} \end{pmatrix}$
c=3	$\begin{pmatrix} 0.13 & 0.16 & \mathbf{0.87} \\ \mathbf{0.35} & \mathbf{0.97} & 0.14 \\ \mathbf{0.59} & 0.16 & 0.21 \end{pmatrix}$
c=4	$\begin{pmatrix} 0.04 & 0.04 & \mathbf{0.79} & 0.14 \\ 0.03 & 0.03 & 0.12 & \mathbf{0.89} \\ 0.07 & \mathbf{0.89} & 0.06 & 0.01 \\ \mathbf{0.85} & 0.05 & 0.15 & 0.19 \end{pmatrix}$

In the gradient-based method, considering a high value for learning rate α may lead to some oscillations in the produced values of the performance index and may eventually lead to a danger of falling into local optima. In contrast, by selecting a very small value of the learning rate, we end up with a very slow learning. Different values of this parameter have been examined, and finally, $\alpha = 0.01$ was selected. The learning was terminated once there was no more visible reduction in the values of the performance index i.e., $Q(\text{iter}) - Q(\text{iter} + 1) < 0.001$. Assuming that U_1 is the partition matrix resulted from clustering data within the window W_1, and U_2 is the partition matrix associated with W_2, Table II shows the fuzzy relations from U_1 to U_2 for the number of clusters $c = 2$, 3, and 4.

Let us consider the fuzzy relation for $c = 4$ in Table II. It indicates that cluster 1 of W_1 [see Fig. 5(e)] is mainly (with 0.79 degree) related to cluster 3 of W_2 [see Fig. 5(f)], and cluster 2 of W_1 is mainly related to cluster 4 of W_2. Furthermore, clusters 3 and 4 of W_1 are in a strong relation with clusters 2 and 1 of W_2, respectively. Using the estimated fuzzy relations, one may visualize the propagation of anomalies during time evolution. For example, let us consider the anomalous cluster 2 in Fig. 5(f). From the estimated fuzzy relations, one may conclude that this cluster is coming from cluster 3 in Fig. 5(e), and so on. Fig. 6(a)–(c) shows the values of the defined performance index (16), after 50 iterations of the proposed gradient-based method. Usually, the objective function is minimized in the first few iterations.

B. Simulated Outbreak Scenario

The North American Animal Disease Spread Model (NAADSM) [6] is used in order to simulate a livestock disease outbreak across the province of Alberta, Canada. NAADSM is a unit (herd)-based stochastic state-transition spread simulation model for contagious diseases among animals. In this method, each infected susceptible unit may have four disease states comprising latent period, subclinically infectious period, clinically infectious period, and naturally immune period. To simulate a disease spread among a group of units in a map, a set of spread parameters should be determined. Some of these parameters are as follows: duration of each disease state in the form of a probability density function, rate of animal shipment,

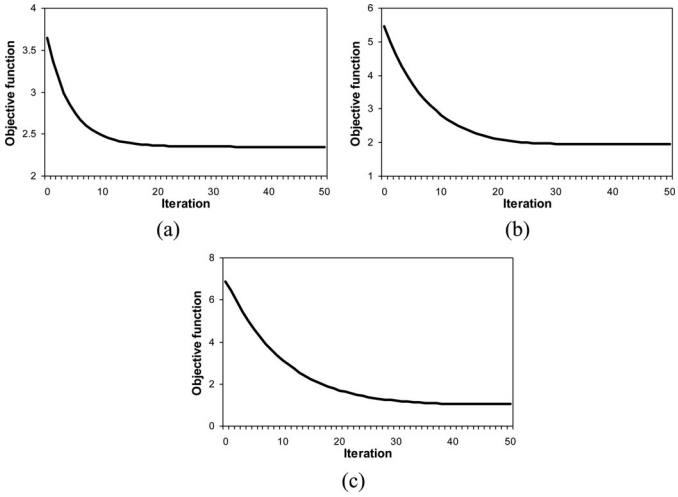

Fig. 6. Objective function of the gradient-based method. (a) $c = 2$, (b) $c = 3$, and (c) $c = 4$.

movement distance, shipping delay, wind direction, maximum distance of spread, airborne transport delay, etc. In addition, some spread control policies including disease detection, quarantine, destruction, and vaccination can be determined.

We considered 246 stations in Alberta agriculture and rural development system (www.agric.gov.ab.ca), and for each station, a number of cattle herds have been generated randomly. Moreover, the population of each herd is considered as a random number located in a certain range. Using NAADSM and considering some values for the above-mentioned parameters, an outbreak with a period of 100 days has been simulated. In the resulting outbreak dataset, for each station, the spatial coordinates are provided in the form of latitude–longitude and there is a time series with length 100 measuring the rate of infected herds within each station for each day. Fig. 7(a) shows the spatial part of the simulated data. As shown in this figure, an outbreak (triangles) occurred in the south part of the province. The highlighted station in this figure (named Del Bonita) is the start point of the outbreak. Fig. 7(b) shows the rate of infected herds corresponding to this station during the simulation, and Fig. 7(c) shows the rate of infected herds in the entire province for 100 days.

In the first step of the experiments, the latitude–longitude pairs are mapped to Cartesian coordinates to be utilized in the calculations of the Euclidean distance. In selecting time windows to generate subsequences, two parameters, namely the length of time windows and the length of overlap between two successive time windows, are considered. As mentioned earlier, these parameters can be selected by the end-user in an interactive manner when running the system and analyzing initial solutions. Moreover, this selection can be based on the application and the nature of data. For example, in periodic time series, the length of time windows can be equal to the length of the period of time series. Another option is using time windows of variable length.

In this experiment, we simply used a sliding window of length 20, and in each step, the window moves in 10 days. As the result, the subsequences inside the following time windows are considered: W_1: days 1–20, W_2: days 11–30, W_3: days 21–40,

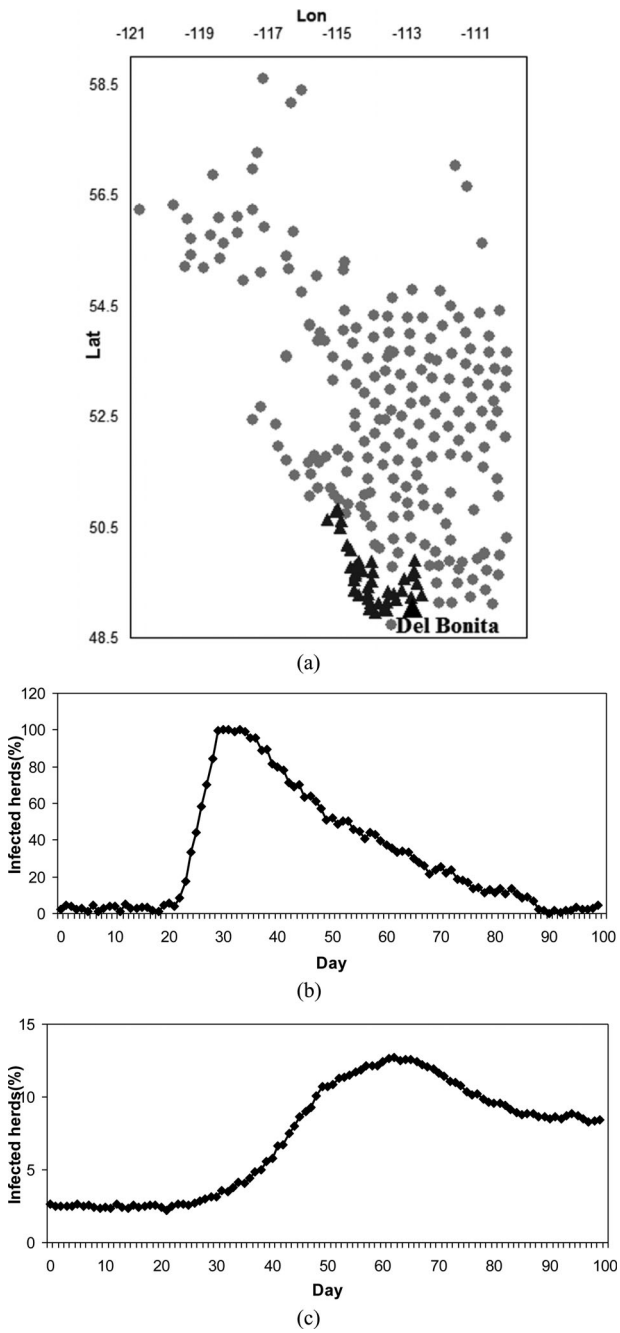

Fig. 7. (a) Spatial part of simulated outbreak, (b) time series corresponding to the station Del Bonita, and (c) rate of infected herds across the province in 100 days.

W_4: days 31–50, W_5: days 41–60, W_6: days 51–70, W_7: days 61–80, W_8: days 71–90, and W_9: days 81–100.

In the next step, the spatial part of data is concatenated to the above-generated temporal subsequences and the resulting spatiotemporal subsequences are clustered using the reconstruction criterion described in Section IV. Different values of parameter λ in range [0,100] are considered for this purpose. A challenging problem here is to find an appropriate number of clusters for each time window (cluster validity index). Although, for this problem, numerous approaches have been reported in the literature [19], most of them are not suitable for data having different parts with different natures (e.g., spatial time series

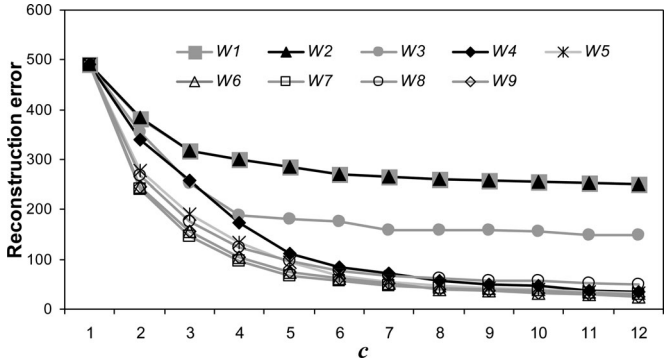

Fig. 8. (a) Reconstruction error versus different number of clusters for time windows W_1–W_9.

TABLE III
SELECTED NUMBER OF CLUSTERS FOR DIFFERENT TIME WINDOWS

Time window	c	Time window	c
W_1	3	W_6	5
W_2	3	W_7	5
W_3	4	W_8	5
W_4	5	W_9	5
W_5	5		

data here). In this paper, we employed the reconstruction criterion in order to find an appropriate number of clusters for each set of subsequences inside time windows. For this purpose, the temporal part of data inside each time window along with the spatial part is clustered using the reconstruction criterion for different number of clusters, and the number of clusters that can reduce the reconstruction error effectively has been chosen. In fact, lower amount of reconstruction error specifies a higher quality of clusters in terms of granulation and degranulation [3].

Fig. 8 shows the reconstruction error versus number of clusters $c = 1$–12 for the time windows W_1–W_9.

For $c = 1$, we simply considered the average value of data objects (both spatial part and temporal part) as the cluster center and the membership degree of each data object to that cluster center is set to 1. In addition, in all experiments, the fuzzification coefficient m was set to 2. As shown in Fig. 8, for all the defined time windows, usually increasing the number of clusters decreases the reconstruction error, and after some steps, this reduction in reconstruction error has been flattened. Consequently, we select the number of clusters at the point where the values of the reconstruction error start exhibiting a saturation effect (no further substantial changes of error are reported when increasing the values of "c").

Table III shows the selected number of clusters for different time windows. Note that in most cases, determining the number of clusters is application dependent and the user can choose the number of clusters based of the nature of the problem under consideration.

Fig. 9(a)–(i) shows the revealed spatiotemporal clusters for different time windows. The stars represent spatial cluster centers, and the number positioned next to each cluster center represents the order of that cluster in its corresponding partition matrix. As shown in these figures, the outbreak has been started in time window W_3 in the south part of the province around the Del Bonita station [see Fig. 7(a)]. In time window W_4, the outbreak moves in two ways: the northern part and the western (left-hand) part of the map. As can be seen from the clusters coming from next time windows, it continues to propagate to the western part of the province. In fact, one of the advantages of the proposed technique is that it can visualize the dynamic change (migration) of anomaly during time evolution.

TABLE IV
ANOMALY SCORES REPORTED FOR DIFFERENT CLUSTERS REVEALED IN TIME WINDOWS W_2–W_9

Time widow	Clusters				
	1	2	3	4	5
W_2	0.01	0.04	0.01	–	–
W_3	0.07	**6.47**	0.05	0.04	–
W_4	0.10	0.16	**4.89**	**7.54**	0.13
W_5	**7.70**	**7.97**	0.23	0.14	0.18
W_6	0.56	0.37	0.17	0.16	**7.73**
W_7	0.27	0.50	**6.29**	0.13	0.12
W_8	0.63	0.38	**5.45**	0.29	0.24
W_9	0.21	**7.07**	0.53	**2.98**	0.30

Table IV reports the estimated anomaly scores calculated using (14) and (15) for different clusters revealed for time windows W_2–W_9. Since W_1 is the first generated time window and there are no historical data for that, the subsequences inside this time window considered as normal. As shown in Table IV, the first anomalous cluster has been detected in time window W_3 and cluster 2 present in this time window is an anomalous one with the anomaly score of 6.47. In the next time window, W_4, both clusters 3 and 4 exhibit a high anomaly score, and in W_5, clusters 1 and 2 are anomalous clusters. In the time windows W_6–W_8, there is one anomalous cluster, and finally, in the window W_9, two anomalous clusters have been found. Now, let us analyze the movement of clusters over time using the proposed fuzzy relational model.

Table V shows the estimated fuzzy relations obtained for successive structures corresponding to the generated time windows. For the gradient-based method, which is used to construct fuzzy relations, the learning rate α was set to 0.01, and the learning has been terminated once there was no significant reduction observed in the performance index (we adhere to the same effect as exercised in the previous experiment).

The fuzzy relation obtained for the transition from W_1 to W_2 indicates that clusters 1, 2, and 3 in W_1, in the next time step, move to clusters 3, 1, and 2 in W_2, respectively. Fig. 9(a) and (b) visualizes these transitions. Considering fuzzy relation from W_2 to W_3, one may conclude that cluster 2 from W_2 moves to clusters 2 and 3 of W_3. Since cluster 2 of W_3 exhibits a high anomaly score, this indicates that some parts of cluster 2 of W_2 in the next step encountered some anomalies in the temporal part of data.

Using the fuzzy relations presented in Table V, one may visualize the evolution of clusters in different time windows

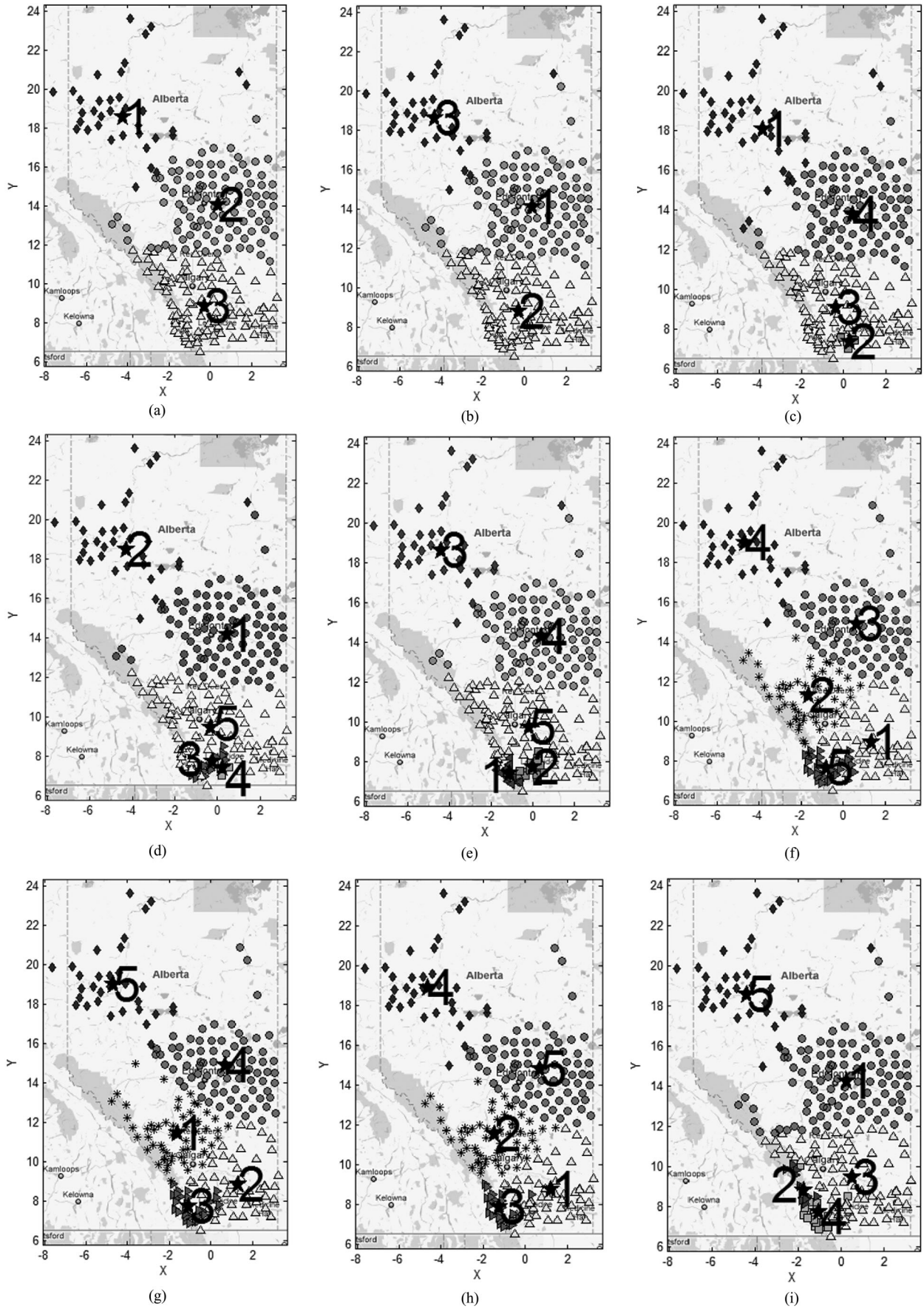

Fig. 9. Revealed spatiotemporal clusters for time windows (a) W_1, (b) W_2, (c) W_3, (d) W_4, (e) W_5, (f) W_6, (g) W_7, (h) W_8, and (i) W_9.

TABLE V
RELATION BETWEEN ANY TWO CONSECUTIVE REVEALED STRUCTURES

Time window	Fuzzy relation (R)
W_1 to W_2	$\begin{pmatrix} 0.02 & 0 & 0.95 \\ 0.93 & 0.05 & 0.04 \\ 0.04 & 0.93 & 0.01 \end{pmatrix}$
W_2 to W_3	$\begin{pmatrix} 0.05 & 0.05 & 0 & 0.90 \\ 0.02 & 0.93 & 1.00 & 0.02 \\ 0.92 & 0.02 & 0 & 0 \end{pmatrix}$
W_3 to W_4	$\begin{pmatrix} 0.07 & 0.88 & 0.08 & 0.01 & 0 \\ 0 & 0 & 0 & 0.96 & 0 \\ 0.10 & 0.03 & 0.68 & 0.02 & 0.80 \\ 0.77 & 0.09 & 0.15 & 0.02 & 0.15 \end{pmatrix}$
W_4 to W_5	$\begin{pmatrix} 0.09 & 0.08 & 0.02 & 0.97 & 0.04 \\ 0.08 & 0.07 & 0.98 & 0 & 0 \\ 0.13 & 0.54 & 0 & 0 & 0 \\ 0.05 & 0.31 & 0 & 0 & 0 \\ 0.44 & 0.08 & 0.01 & 0.01 & 0.97 \end{pmatrix}$
W_5 to W_6	$\begin{pmatrix} 0.01 & 0.01 & 0.01 & 0.01 & 0.46 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.38 \\ 0.03 & 0.05 & 0.03 & 0.98 & 0.03 \\ 0.14 & 0.19 & 0.94 & 0.02 & 0.03 \\ 0.82 & 0.66 & 0.03 & 0.01 & 0.04 \end{pmatrix}$
W_6 to W_7	$\begin{pmatrix} 0.01 & 0.95 & 0.06 & 0.01 & 0.01 \\ 0.96 & 0.02 & 0.06 & 0.02 & 0.01 \\ 0.03 & 0.01 & 0.02 & 0.97 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.97 \\ 0.01 & 0.01 & 0.83 & 0.01 & 0.01 \end{pmatrix}$
W_7 to W_8	$\begin{pmatrix} 0.03 & 0.99 & 0.03 & 0.01 & 0.02 \\ 0.95 & 0.01 & 0.04 & 0.01 & 0.01 \\ 0.04 & 0.01 & 0.91 & 0.01 & 0.01 \\ 0.02 & 0.02 & 0.02 & 0.01 & 0.96 \\ 0.01 & 0.01 & 0.01 & 0.98 & 0.01 \end{pmatrix}$
W_8 to W_9	$\begin{pmatrix} 0.01 & 0.09 & 0.65 & 0.07 & 0.01 \\ 0.17 & 0.12 & 0.37 & 0.06 & 0.02 \\ 0.01 & 0.56 & 0.01 & 0.76 & 0.01 \\ 0.01 & 0.06 & 0.01 & 0.02 & 0.96 \\ 0.75 & 0.08 & 0.04 & 0.05 & 0.02 \end{pmatrix}$

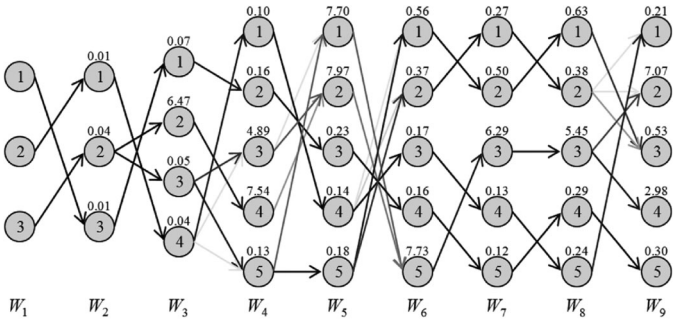

Fig. 10. Graph representation of anomaly scores and fuzzy relations reported in Tables IV and V.

using a graph-oriented representation. In Fig. 10, nodes represent clusters, edges stand for relations (associations) between clusters, and each layer of nodes represents single-step windows. The numbers displayed over the nodes of the graph represent anomaly scores reported in Table IV, and the level of shading of the edges corresponds to the membership value present in the corresponding entry of the fuzzy relation. In other words, for the membership grades close to 1, the edges are black, and for values of membership close to 0, the links are almost invisible.

The constructed graph in Fig. 10 shows the evolution of clusters in time windows W_1–W_9. Using this structure, one may track normal and anomalous clusters. Let us consider cluster 3

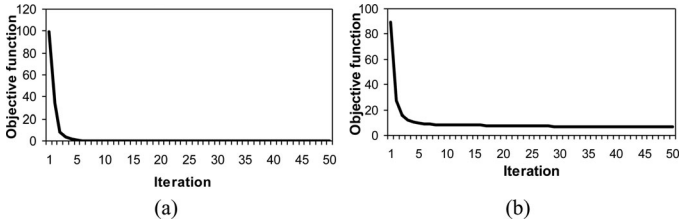

Fig. 11. Values of the minimized objective function reported in 50 iterations of the learning scheme: optimization of the relationships from (a) W_1 to W_2, and (b) W_2 to W_3.

from time window W_1. As shown in Fig. 10, it moves to cluster 2 in the next time window, W_2. In W_3, this cluster split into clusters 2 and 3. Cluster 2 is anomalous and is related to anomalous cluster 4 in time window W_4. On the other hand, some data of cluster 3 in W_3 encounter with some anomalies in the next time window resulting the appearance of cluster 3 in W_4. Both anomalous clusters in W_4 merge into cluster 2 in time window W_5. A new anomalous cluster (cluster 1) has emerged in this time window from cluster 5 of W_4. Anomalous clusters 1 and 2 of W_5 are merged into cluster 5 in time window W_6. It moves to cluster 3 in W_7 and then moves to cluster 3 in W_8. Finally, this cluster splits into clusters 2 and 4 in W_9.

Fig. 11(a) and (b) shows the amount of the performance index defined in (16) to calculate the fuzzy relations from W_1 to W_2, and from W_2 to W_3 respectively. The final value of the objective function in Fig. 11(a) is very close to 0, while in Fig. 11(b), the amount of the final objective function is higher. The reason is that the revealed structures in time windows W_1 and W_2 are very similar so that the available structure in W_1 can be estimated using the structure in W_2 and the estimated fuzzy relation. On the other hand, the available structures in time windows W_2 and W_3 are different and the resulting objective function has a higher extent.

C. Alberta Temperature Data

In this section, we look at the Alberta daily temperature data for the first 70 days of years 2009 and 2010 (www.agric.gov.ab.ca) for the stations shown in Fig. 7(a). The spatial time series corresponding to year 2010 are used to realize anomaly detection, while the data corresponding to 2009 are assumed to be normal and are used in calculating anomaly scores. A sliding window with length 30 moving 20 time steps in each movement is considered in this experiment. The following time windows are realized: W_1 for days 1–30, W_2 for days 21–50, and W_3 for days 41–70 all from year 2010. Stations located in the western part of Alberta are selected and their daily temperature is increased 20 °C in days 25–45 to produce some anomalies. By concatenating the spatial part of data to the temporal subsequences W_1, W_2, and W_3, three spatiotemporal subsequences are realized, and using the described method in the previous sub-section, the number of clusters for them are estimated as 3, 4, and 3, respectively. Instead of using the expressed method in (14), an anomaly score for each subsequence is calculated as its squared Euclidean distance to the corresponding time window in the previous year. For example, to estimate an

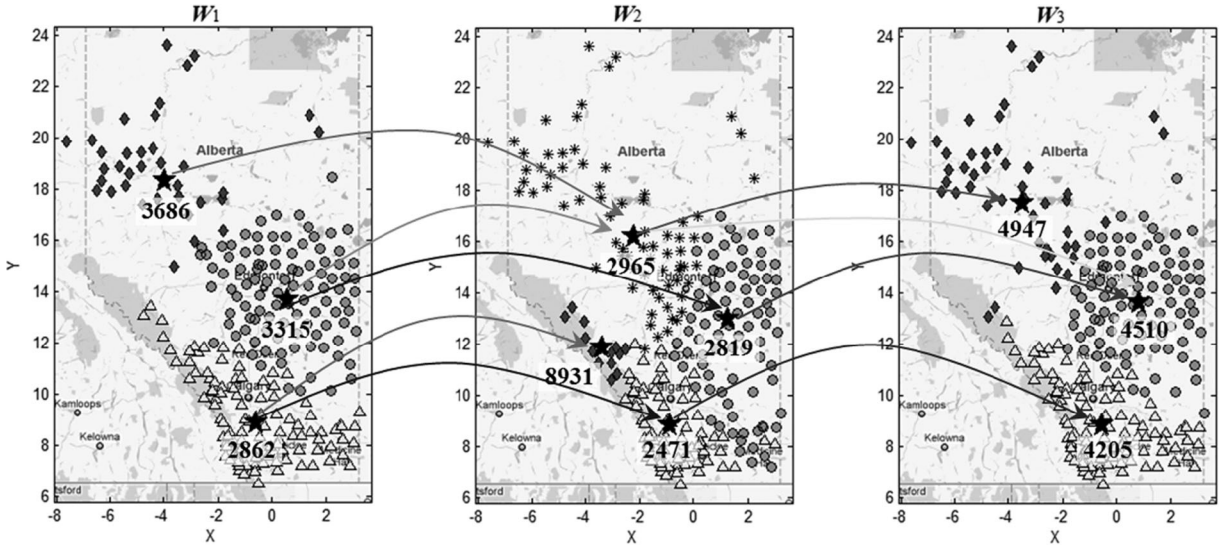

Fig. 12. Revealed clusters for spatiotemporal data corresponding to W_1 to W_3. Numbers inside clusters show estimated anomaly scores, and arcs between clusters represent fuzzy relations.

anomaly score for a subsequence inside W_1, we compare it with its corresponding subsequence for days 1–30 in year 2009. The intuition behind this technique is that temperature data in the same period of different years usually follow a similar pattern, and as the result, subsequences that are not similar to the previous year period are regarded as anomalous subsequences. After estimating anomaly score for each single subsequence, (15) has been exploited to aggregate the estimated anomaly scores inside each cluster. Fig. 12 shows the revealed clusters using the reconstruction criterion for the generated spatiotemporal data W_1, W_2, and W_3. Stars inside clusters show the spatial prototypes, and the numbers under this symbol are anomaly scores. Arcs between clusters in consecutive time windows represent the membership value present in the corresponding entry of the fuzzy relation. For the membership grades close to 1, the edges are black, and for values of membership close to 0, the links become almost invisible. As shown in this figure, the cluster located in the western side of the map in time window W_2 has a high anomaly score in comparison with the other clusters. From the input arc coming to this cluster, one may conclude that it has been emerged from the cluster located in the southern part of the map in time window W_1. There is no a strong relation between this cluster and the clusters in time window W_3, meaning that it has been disappeared in next time window.

The described technique in this paper is a general framework for anomaly detection and characterization in spatial time series data. As shown from the three above experiments, it can visualize the structure of data in an understandable form for the end-user and provide an anomaly score for each cluster measuring the level of unexpected changes of data for each time interval. The proposed gradient-based fuzzy relation method visualizes the propagation of anomalies during time evolution so that one may track an anomalous cluster from the revealed structure in the past. In comparison with traditional methods (e.g., SatScan [26]), the proposed approach is more general and can detect and visualize different types of anomalies and their propagation during time evolution in data. It is transparent and does not require complicated preprocessing steps for implementation. Furthermore, it provides a flexible and interactive way for the end-user to select a suitable technique to assign anomaly score to each single subsequence.

VII. Conclusion

A novel technique for anomaly detection and characterization for spatial time series data has been proposed in this paper. A sliding window with a fixed length has been considered to generate a set of spatiotemporal subsequences in consecutive time steps. Then, a spatiotemporal clustering has been employed to reveal existing structure within each time window. Considering the historical behavior of each single subsequence inside each time window, an anomaly score has been estimated, and then, the estimated anomaly scores are aggregated to assign an anomaly score to each cluster inside each time window. In the next step, to analyze the behavior of clusters and propagation of anomalies during time evolution, a gradient-based method has been proposed to produce a fuzzy relation between any two consecutive revealed structures. The technique is illustrated using an illustrative synthetic datasets, an outbreak scenario simulated using NAADSM over a set of agriculture stations in Alberta province, and the Alberta temperature dataset. We have showed that the proposed method helps detect and characterize the incident anomalies in the form that is understandable and user-friendly by strongly supporting the visualization and comprehension of the revealed structures.

References

[1] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1974.

[2] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Functions*. New York, NY, USA: Plenum, 1981.

[3] H. Izakian, W. Pedrycz, and I. Jamal, "Clustering spatio-temporal data: An augmented fuzzy c-means," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 855–868, Oct. 2013.
[4] W. Pedrycz and J. V. De Oliveira, "A development of fuzzy encoding and decoding through fuzzy clustering," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 4, pp. 829–837, Apr. 2008.
[5] H. Cheng, P. Tan, C. Potter, and S. Klooster, "A robust graph-based algorithm for detection and characterization of anomalies in noisy multivariate time series," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Pisa, Italy, 2008, pp. 349–358.
[6] N. Harvey, A. Reeves, M. Schoenbaum, F. Zagmutt-Vergara, C. Dubé, A. Hill, B. Corso, B. McNab, C. Cartwright, and M. Salman, "The North American animal disease spread model: A simulation model to assist decision making in evaluating animal disease incursions," *Preventive Veterinary Med.*, vol. 82, no. 3/4, pp. 176–197, Dec. 2007.
[7] J. Takeuchi and K. Yamanishi, "A unifying framework for detecting outliers and change points from time series," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 4, pp. 482–489, Apr. 2006.
[8] C. Brighenti and M. A. Sanz-Bobi, "Auto-regressive processes explained by self-organized maps: Application to the detection of abnormal behavior in industrial processes," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2078–2090, Dec. 2011.
[9] W. Ma, L. Jiao, M. Gong, and C. Li, "Image change detection based on an improved rough fuzzy c-means clustering algorithm," *Int. J. Mach. Learn. Cybern.*, 2013. DOI: 10.1007/s13042-013-0174-4.
[10] E. Keogh, J. Lin, and A. Fu, "Hot SAX: Efficiently finding the most unusual time series subsequence," in *Proc. Fifth IEEE Int. Conf. Data Mining*, 2005, pp. 226–233.
[11] A. Khatkhate, A. Ray, E. Keller, S. Gupta, and S. C. Chin, "Symbolic time series analysis for anomaly detection in mechanical systems," *IEEE Trans. Mechatronics*, vol. 11, no. 4, pp. 439–447, Aug. 2006.
[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, pp. 832–839, May 2012.
[13] D. J. Hill, B. S. Minsker, and E. Amir, "Real-time Bayesian anomaly detection for environmental sensor data," in *Proc. 32nd Congr. Int. Assoc. Hydraulic Eng. Res.*, 2007.
[14] P. Protopapas, J. M. Giammarco, L. Faccioli, M. F. Struble, R. Dave, and C. Alcock, "Finding outlier light curves in catalogues of periodic variable stars," *Monthly Notices Royal Astron. Soc*, vol. 369, no. 2, pp. 677–696, 2006.
[15] H. Izakian and W. Pedrycz, "Anomaly detection in time series data using a fuzzy c-means clustering," in *Proc. Joint IFSA World Congr. NAFIPS Annu. Meeting*, 2013, pp. 1513–1518.
[16] H. Izakian and W. Pedrycz, "Agreement-based fuzzy c-means for clustering data with blocks of features," *Neurocomputing*, vol. 127, pp. 266–280, Mar. 2014.
[17] V. Chandola, V. Mithal, and V. Kumar, "Comparative evaluation of anomaly detection techniques for sequence data," in *Proc. 8th IEEE Int. Conf. Data Mining*, Pisa, Italy, 2008, pp. 743–748.
[18] E. W. Dereszynski and T. G. Dietterich, "Spatio-temporal models for data-anomaly detection in dynamic environmental monitoring campaigns," *ACM Trans. Sensor Netw.*, vol. 8, no. 1, Aug. 2011.
[19] W. Wang and Y. Zhanga, "On fuzzy cluster validity indices," *Fuzzy Sets Syst*, vol. 158, no. 19, pp. 2095–2117, Oct. 2007.
[20] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal clustering," in *Data Mining and Knowledge Discovery Handbook*. New York, NY, USA: Springer, part 6, 2010, pp. 855–874.
[21] M. Das and S. Parthasarathy, "Anomaly detection and spatio-temporal analysis of global climate system," in *Proc. Third Int. Workshop Knowl. Discovery Sensor Data*, 2009, pp. 142–150.
[22] W. Pedrycz and A. Bargiela, "Fuzzy clustering with semantically distinct families of variables: Descriptive and predictive aspects," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 1952–1958, Oct. 2010.
[23] G. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1994.
[24] E. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Towards parameter-free data mining," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 206–215.
[25] D. T. Anderson, A. Zare, and S. Price, "Comparing fuzzy, probabilistic, and possibilistic partitions using the Earth Mover's distance," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 4, pp. 766–775, Aug. 2013.
[26] M. Kulldorff, R. Heffernan, J. Hartman, R. Assunção, and F. Mostashari, "A space-time permutation scan statistic for disease outbreak detection," *PLoS Med.*, vol. 2, no. 3, pp. 216–224, Mar. 2005.
[27] D. B. Neill, "Expectation-based scan statistics for monitoring spatial time series data," *Int. J. Forecasting*, vol. 25, no. 3, pp. 498–517, Sep. 2009.
[28] W. Pedrycz, "Proximity-based clustering: A search for structural consistency in data with semantic blocks of features," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 978–982, Oct. 2013.
[29] D. Gao, Y. Kinouchi, K. Ito, and X. Zhao, "Neural networks for event extraction from time series: A back propagation algorithm approach," *Future Gener. Comput. Syst.*, vol. 21, no. 7, pp. 1096–1105, Jul. 2005.
[30] X. X. Zhang, H. X. Li, and C. K. Qi, "Spatially constrained fuzzy-clustering-based sensor placement for spatio-temporal fuzzy-control system," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 946–957, Oct. 2010.
[31] P. D'Urso, "Fuzzy clustering for data time arrays with inlier and outlier time trajectories," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 5, pp. 583–604, Oct. 2005.
[32] R. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, no. 7, pp. 833–841, May 2007.
[33] D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology," in *Proc. 5th Int. Conf. Intell. Syst.*, 1996, pp. 82–87.
[34] J. Ma and S. Perkins, "Time series novelty detection using one-class support vector machines," in *Proc. Int. Joint Conf. Neural Netw.*, 2003, pp. 1741–1745.
[35] D. Anderson, J. Bezdek, M. Popescu, and J. Keller, "Comparing fuzzy, probabilistic and possibilistic partitions," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 906–918, Oct. 2010.
[36] E. Hullermeier, M. Rifqi, S. Henzgen, and R. Senge, "Comparing fuzzy partitions: A generalization of the Rand index and related measures," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 3, pp. 546–556, Jun. 2012.

Hesam Izakian (S'12) received the M.S. degree in computer engineering from the University of Isfahan, Isfahan, Iran, and is currently working toward the Ph.D. degree under the supervision of Prof. W. Pedrycz with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada.

His research interests include computational intelligence, knowledge discovery and data mining, pattern recognition, and software engineering.

Witold Pedrycz (M'88–SM'94–F'99) received the M.Sc., Ph.D., and D.Sci. degrees from the Silesian University of Technology, Gliwice, Poland.

He is currently a Professor and Canada Research Chair (CRC computational intelligence) with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia. He is the author of 14 research monographs covering various aspects of computational intelligence and software engineering. His main research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in these areas.

Prof. Pedrycz was elected as a Fellow of the Royal Society of Canada in 2012. He has been a member of numerous program committees of IEEE conferences in the area of fuzzy sets and neurocomputing. In 2009, he was elected as a foreign member of the Polish Academy of Sciences, Warsaw, Poland. He is intensively involved in editorial activities. He is an Editor-in-Chief of *Information Sciences* and Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART A: SYSTEMS AND HUMANS. He currently serves as an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a member of a number of editorial boards of other international journals. In 2007, he received the prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Council. He received the IEEE Canada Computer Engineering Medal in 2008. In 2009, he received a Cajastur Prize for soft computing from the European Centre for Soft Computing for "*pioneering and multifaceted contributions to granular computing*."