Overview

Spatiotemporal change footprint pattern discovery: an inter-disciplinary survey

Xun Zhou,* Shashi Shekhar and Reem Y. Ali

Given a definition of change and a dataset about spatiotemporal (ST) phenomena, ST change footprint discovery is the process of identifying the location and/or time of such changes from the dataset. Change footprint discovery is fundamentally important for the study of climate change, the tracking of disease, and many other applications. Methods for detecting change footprints have emerged from a diverse set of research areas, ranging from time series analysis and remote sensing to spatial statistics. Researchers have much to learn from one another, but are stymied by inconsistent use of terminology and varied definitions of change across disciplines. Existing reviews focus on discovery methods for only one or a few types of change footprints (e.g., point change in a time series). To facilitate sharing of insights across disciplines, we conducted a multi-disciplinary review of ST change patterns and their respective discovery methods. We developed a taxonomy of possible ST change footprints and classified our review findings accordingly. This exercise allowed us to identify gaps in the research that we consider ripe for exploration, most notably change pattern discovery in vector ST datasets. In addition, we illustrate how such pattern discovery might proceed using two case studies from historical GIS.

How to cite this article:
WIREs Data Mining Knowl Discov 2014, 4:1–23. doi: 10.1002/widm.1113

INTRODUCTION

Given a definition of change and a dataset about a spatiotemporal (ST) phenomenon, ST change footprint pattern discovery is the process of identifying the location and/or time of such changes in the data. Discovering footprint patterns of change from large datasets is an increasingly important activity in application domains ranging from climate science to public health. Data science (e.g., data mining, machine learning, statistics) researchers have developed numerous techniques to facilitate the discovery of such patterns. Addressing domain-specific challenges, they have often worked in distinct research settings, most notably time series analysis, image analysis, and spatial statistics. An interdisciplinary review classifying and summarizing different change footprint patterns and techniques may provide domain users valuable guidance for tool selection to solve their problems. More importantly, a survey analyzing the current research accomplishments also helps data scientists identify future research needs for data science.

Change footprint pattern discovery is an important task in a number of applications. We list a few major applications and their respective domain questions that ST change pattern discovery may help answer.

Statistical quality control: Change pattern discovery techniques have long been applied in industrial process monitoring. The main goal is to detect system faults.[1] A typical question is: at what time did the signal change?

Remote Sensing: Remote sensing techniques provide images of an area at different times (e.g., before and after a flood). By comparing two or more snapshots of the study area, one can answer a question like ‘which

*Correspondence to: xun@cs.umn.edu

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.

road or bridge has been damaged?' or 'which area has been flooded'. This task is referred to as 'change detection.'[2–4] Decision makers can quickly assess the loss in a disaster and respond correspondingly.

Public Health: An important mission of public health is to monitor epidemic disease outbreak. Specifically, it aims to find regions where the risk of infection increases.[5] This can also be viewed as a change pattern discovery problem.

Public Safety: Similar to public health, public safety applications are interested in finding the change pattern of crime risk. A region with an increase in number of crime reports may indicate the need of more policing attention.[6]

Ecology and Environmental Science: Ecologists and environmental scientists are interested in finding spatial change patterns of ecosystems, such as the boundaries between habitats of different species or eco-zones,[7] or the shrinking or expansion of certain land types, such as the desertification process.

Climate Science: Climate science is obviously interested in a number of change-related questions. The biggest one is to verify and quantify global climate change. Specifically, they may ask questions such as when did the precipitation in the Sahel region in Africa change from normal to very limited? Is the intensity, frequency, or duration of extreme climate events (e.g., flood) changing?[8]

We reviewed survey papers in various disciplines and found they mostly focused on a limited types of change footprint pattern. For example, the literature on time series change detection[2,9–12] focuses on change(s) that occur at single time points. Remote sensing and image change detection survey papers[13–15] mainly focus on finding regions (collections of pixels) of change between two imagery snapshots. A tutorial by Wong and Neill[16] discusses the problem of 'event detection' which aims to discover abnormal behavior of data. The authors covered change points in time series, spatial clusters (polygon footprint), and spatiotemporal clusters (ST volume footprint), but not changes with other interesting footprints such as temporal intervals and spatial boundaries.[17,18] Comprehensive reviews that compare techniques across disciplines are lacking.

Classifying ST change footprint patterns and techniques across disciplines is challenging. First of all, there is no unified definition of ST change. This makes it hard to compare and contrast change discovery techniques across disciplines. Second, different research disciplines employ similar terminology to describe different phenomena. For example, the term 'abrupt change' in time series analysis refers to a change in the statistical distribution of data at a certain time point.[9] However, in spatial Wombling, it describes a significant difference of value across boundaries separating different spatial areas.[19] This is referred to as the synonym problem. Compounding the confusion, patterns exhibiting the same ST footprint may be named differently. For example, the output of spatial Wombling, which is a set of boundaries highlighting significant changes of value, has been called (zones of) abrupt change,[19] rapid change,[20] etc, by different researchers. This is referred to as the homonym problem. Finally, ST change patterns defy easy classification due to their inherently complex nature. It is also hard to select the set of key features that best classifies a particular pattern.

In this article, we define an interdisciplinary framework for ST change pattern discovery and make the following specific contributions.

- We propose a taxonomy that classifies ST changes based on their ST footprints.
- We review representative techniques in use today by data scientists for the discovery of each change pattern.
- We analyze gaps where research is lacking in an ST change footprint pattern family, and we suggest problems that merit more exploration by the data science community.

This paper examines only two aspects of ST changes: where and when a change occurs. Understanding the physical mechanism of changes (why) and modeling of various change patterns from applications (how) are beyond the scope of this paper. Also, this paper focuses on the discovery of change footprint patterns from available data. Prediction methods for future changes are not addressed here. Finally, this paper only reviews representative techniques only in the context of change footprint pattern discovery. It is not meant to provide a comprehensive survey of the large body of techniques in time series analysis (e.g., spectral decomposition), image processing, and spatial.

Figure 1 shows the main components of a change footprint pattern discovery process: ST data from an application is the input of the problem. A definition of a change pattern is given based on the underlying application. Finally, a method (e.g., statistical, computational) that discovers the pattern from the data will produce the ST footprints. The rest of the chapter is organized according to this framework: The 'Data Input for Spatiotemporal Change Pattern Discovery' section surveys different ST data types and statistical data models for change pattern discovery. The 'Definitions of Spatiotemporal

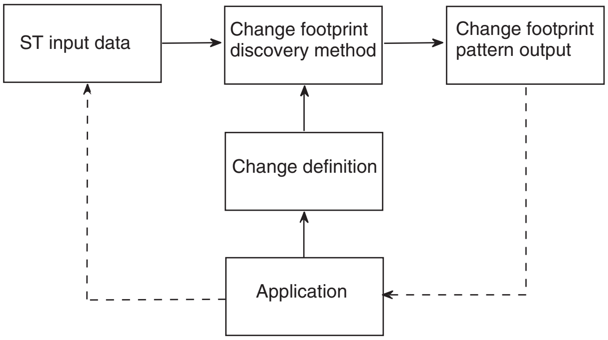

FIGURE 1 | A flow chart showing ST change footprint pattern discovery process.

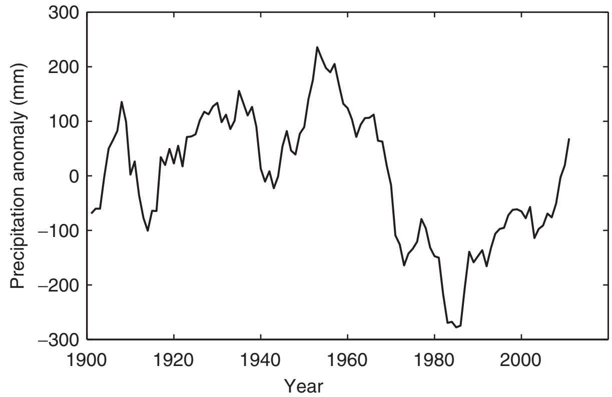

FIGURE 2 | An example of time series from climate science.

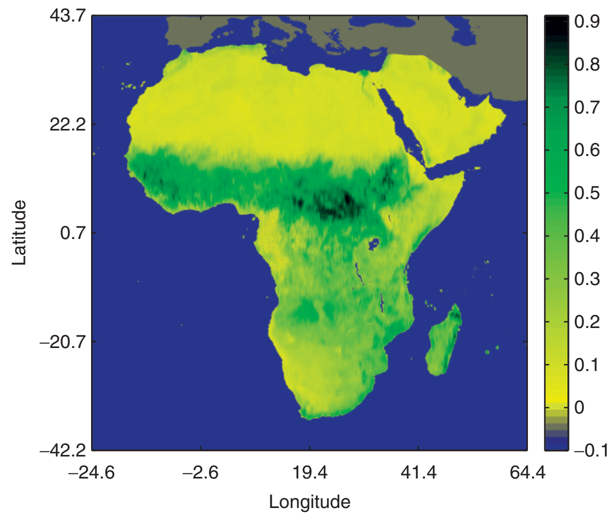

FIGURE 3 | A spatial raster dataset showing vegetation cover (in NDVI value) of Africa (best viewed in color).

Change' section presents four common definitions of change pattern. We propose a taxonomy of ST change footprint patterns and classify existing change footprint pattern discovery techniques in the 'A Taxonomy of Spatiotemporal Change Footprint Patterns' section. In the 'Future directions and research needs' section we list a few patterns that we believe merit attention by researchers in future research. Finally, we conclude the paper with future plans.

DATA INPUT FOR SPATIOTEMPORAL CHANGE PATTERN DISCOVERY

Data generated by various applications that are used for change footprint discovery include temporal data, spatial data, and ST data. Temporal data here refers to time series. Spatial data can be categorized into two types, namely, spatial raster data and spatial vector data.[21] ST data include raster series, and ST vector data.

A *time series* is a sequence of data points, measured typically at successive time instants spaced at uniform time intervals (e.g., second, day, year, etc). Many applications generate time series. For example, the readings of a sensor that monitors the quality of products form a time series. The yearly rainfall amounts in a spatial region is also a time series. Figure 2 shows an example of a time series generated by climate science data. It represents the annual precipitation anomaly (mean value removed from real value) from 1900 to 2010 in the Sahel region of Africa.[22]

Spatial raster data represents the space as a finite grid structure (i.e., spatial framework) where a number of functions representing application specific nonspatial features are assigned to each grid. For example, the longitude–latitude system is a spatial framework; the data representing the precipitation amount over the surface of the world is a typical spatial raster dataset. A satellite image is also a typical spatial raster dataset. Figure 3 is an example of raster data. It represents the vegetation cover in Africa measured in normalized difference vegetation index (NDVI) value where each pixel equals approximately 8 km by 8 km on the ground.[23–25]

Relationships and interactions between different spatial raster fields are specified by field operations.[26] Depending on their scale, operations may be local, focal, or zonal.[27] Local operations determine the output at a single location depending on attributes at this location. For example, 'find locations with a precipitation greater than 500 mm' is a local operation. Focal operations determine their output based on attributes in an assumed small neighborhood of an input location. For example, computing the gradient of precipitation value over the world's land surface is a focal operation. Zonal operations usually

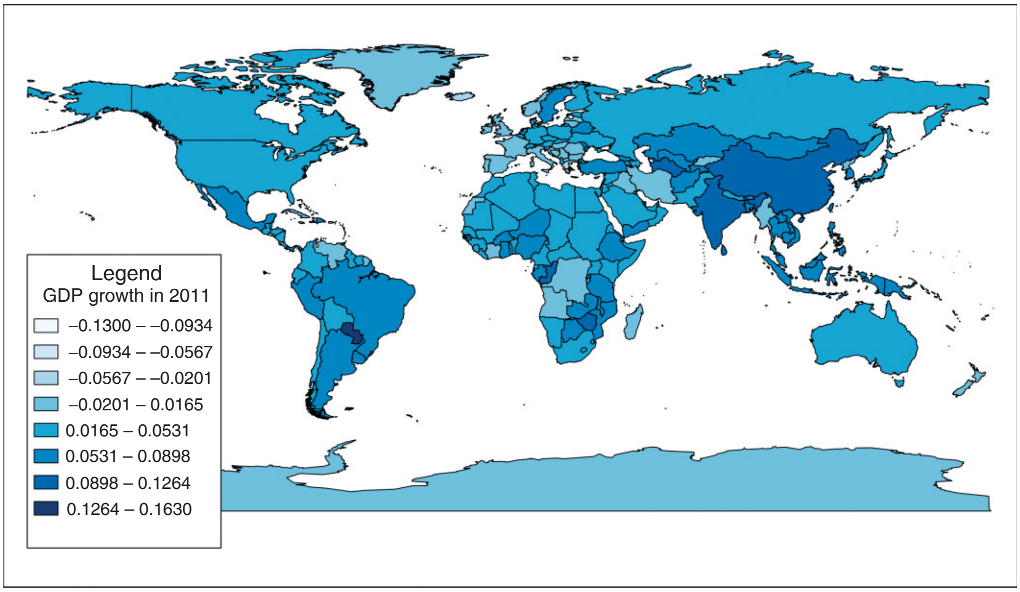

FIGURE 4 | A spatial vector dataset showing GDP growth in countries around the world in 2011 (best viewed in color).

employ aggregate operators of locations in a region. For example, determining the average precipitation in Africa is a zonal operation.

Spatial vector data uses geometric shapes to represent spatial objects. These shapes include points, line segments, polygons, as well as combinations of these shapes. For example, cities in a map can be represented as points. Road segments are usually represented as line segments. Counties, States, and countries are often modeled as polygons. Figure 4 shows an example of spatial vector data where each country in the world is represented by a polygon object and the GDP growth in 2011 of each country is an associated attribute.[28]

ST data are associated with temporal information in addition to the spatial data types. An ST raster series represents a spatial field at a number of successive snapshots. For example, a sequence of satellite images showing the vegetation cover in Africa taken every other week forms a spatial raster series dataset. From a temporal perspective, the data at each location/pixel of the raster field also forms a time series of the same length.[29] Thus this type of dataset is also referred to as the spatial time series.[30]

ST vector data are objects with both spatial and temporal information. For example, the disease reports in the state of Minnesota in the past 5 years form a point-collection sequence where the locations of the points may vary. The historical coverage of a bird's habitat also forms a vector series, where the habitat in each year can be represented as a polygon.

Statistical Models of ST Data

Time series are traditionally modeled as independent, identically distributed (i.i.d.) samples drawn from an underlying distribution. For example, it is quite common to assume that the readings of a sensor follow a Gaussian process[9] where the deviation between each reading and a fixed mean value follows a normal distribution. However, it is often the case that time-referenced attributes are auto-correlated, meaning that the value at time t is dependent on the value at time t-1. More complicated models of time series have been applied to model temporal data, such as the Markov chain, where the probability distribution of a value at time t depends only on the value at t-1.[31]

Spatial data were viewed by traditional statistics as i.i.d. data samples from a distribution. In contrast, spatial statistics, a branch of statistics that studies the modeling and analysis of spatial data, propose different models to honor the unique nature of spatial data, i.e., spatial autocorrelation and spatial heterogeneity. First of all, spatial data is highly self-correlated. This has been recognized as a fundamental observation in geography and named as the first law of geography: 'Everything is related to everything, but nearby things are more related than distant things.'[32] Spatial heterogeneity refers to the fact that an underlying process varies from place to place.

As a result, three spatial statistical models have been developed to model spatial data, namely, the geostatistical model, the lattice model, and the point process model.[33] The geostatistical model deals with a continuous spatial surface with discrete sampled locations (e.g., ground observational stations to record rainfall data). Tools such as Kriging are used to interpolate un-sampled values. The lattice model (a.k.a. the areal model) deals with continuous space partitioned into regular grids or irregular polygons (e.g., the counties in Minnesota). Interactions between partitions are characterized by the spatial neighborhood relationship (e.g., topological connectivity). Data models such as the spatial autoregressive model (SAR) and Markov random field (MRF) can be applied on such datasets. Finally, point process is used to model the distribution of spatial points in a spatial framework. For example, crime locations in a city can be modeled as a point process. The distribution of points may be completely random, clustered, or de-clustered.

DEFINITIONS OF SPATIOTEMPORAL CHANGE

Although the same term 'change' is used to name patterns in various applications, the underlying phenomena may differ significantly. This section briefly summarizes four main ways for defining a change in data. Since the modeling of change is not the focus of this paper, our review of the definitions should not be considered comprehensive.

Specifically, change in data can be defined mainly in the following four different ways.

Change in Statistical Parameter

Data in some applications are assumed to be random samples drawn from an underlying process. A change is thus defined as a shift in the statistical distribution of the data. For example, in statistical quality control, sensor readings are expected to follow a certain statistical distribution (e.g., Gaussian).[9] If a fault occurs, the mean or variance of data will change. This type of definition may make different assumptions on the underlying data distribution. Definitions with parametric models assume that the underlying distributions are of a certain kind (e.g., Gaussian distribution) while definitions with nonparametric models do not. Figure 5(a) shows an example of a 'change of mean' in a time series. As can be seen, the mean of the data before and after the highlighted time points are significantly different.

Change in Actual Value

The second type of definition is based on the actual values of the data. The definition of change is initially modeled (mathematically) in calculus, where a difference between a data value and its neighborhoods in location or time is viewed as a change. In a one-dimensional continuous function, the magnitude of change is usually characterized by the derivative function, while on two-dimensional surfaces, a change is usually characterized by the magnitude of the gradient. For a discrete function (e.g., time series), the change between two data points can be characterized by the slope of the line connecting the two. For example, in spatial statistics, boundary analysis (a.k.a. spatial Wombling) is done by finding the significant changes between neighboring locations. Figure 5(b) a 'change of value' in a time series. The time period highlighted has a steep slope, indicating it is a change.

Change in Models Fitted to Data

A third type of definition focuses on the change in the trend/behavior of the data. A number of function models are fitted to the data where a change in one or more of the models is defined as an instance of change.[34] For example, climate scientists studying the trend of global precipitation want to be able to detect a turning point where the rainfall changes from increasing to decreasing. In such scenarios, a time series can be fitted using a certain number of straight lines by minimizing the error (e.g., least square error). Hence, a change in the data is defined as a discontinuity between two consecutive linear functions. The models can also be nonlinear (e.g., polynomial).[35] Figure 5(c) illustrates a 'change in linear models' in a time series. The highlighted time points represent 'change' since they are the break points between different linear segments fitted to the data.

Change in Derived Attributes

Some applications define change patterns indirectly. First, they establish a classification or prediction model of the data. Then they run the model and derive new attributes, such as predicted value or a categorical class label of the data. For example, in time series change detection, a predictive model can be learnt based on a training dataset. The future values are then predicted and compared with actual values. A difference between the prediction and actual value is considered a change.[36,37]

A TAXONOMY OF SPATIOTEMPORAL CHANGE FOOTPRINT PATTERNS

We are now ready to propose a taxonomy of the change footprint patterns from space-time perspective.

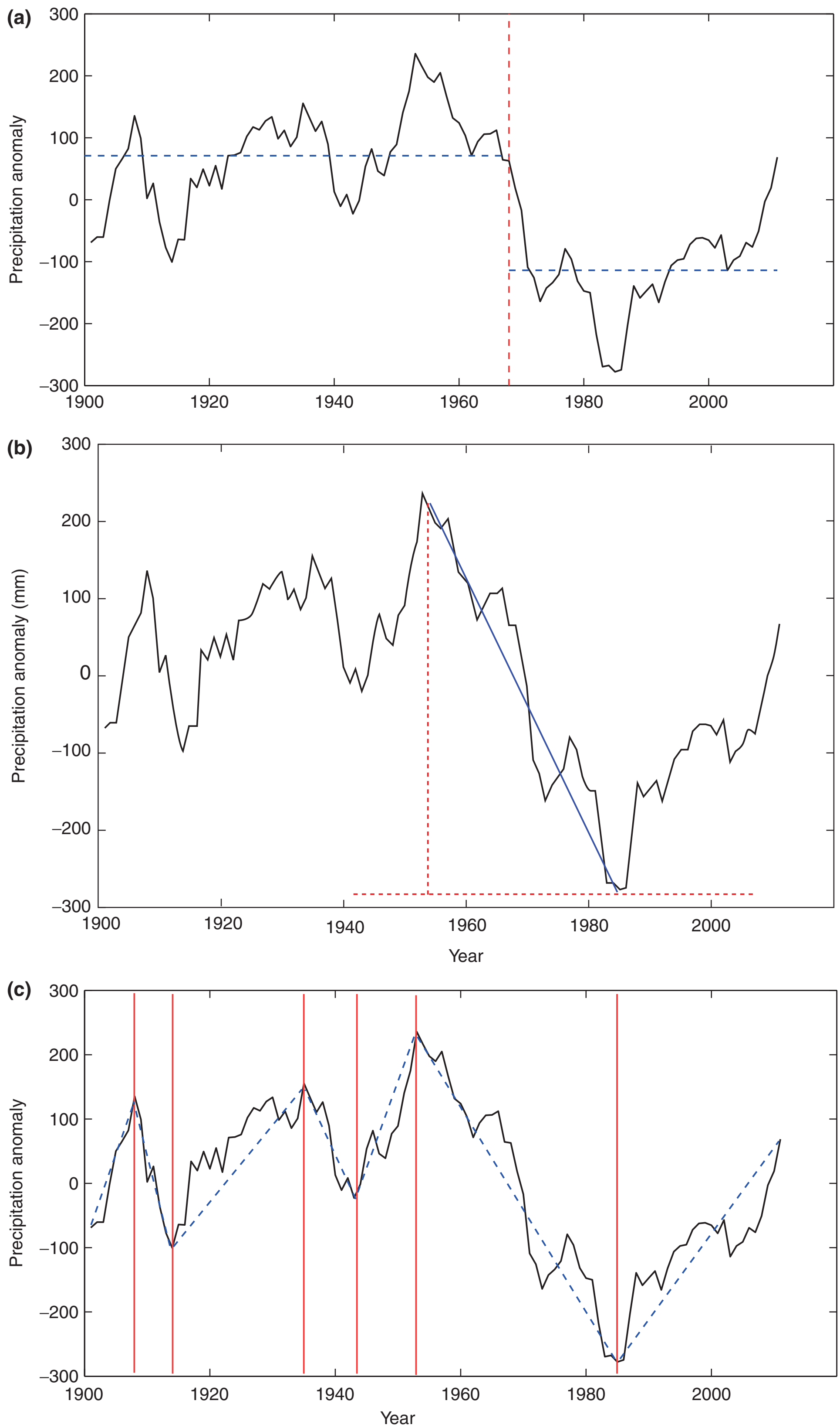

FIGURE 5 | Three sets of results for one time series dataset using three different definitions of change. (a) Statistical parameter change in a time series. (b) Value change in a time series. (c) Change in model fitted on a time series.

TABLE 1 | Classification of ST Raster Change Footprints with Examples of Typical Questions. Empty Cells Indicate Patterns that Yet to be Studied in Depth

		Temporal Footprint			
		Single Snapshot (T1)	Between Few Snapshots (T2)	Points in Time Series (T3)	Intervals in Time Series (T4)
Spatial footprint	Local R1	*R1T1 (N/A)*		Change point detection (e.g., CUSUM) *R1T3*	Change interval discovery on time series *R1T4*
	Focal R2	Edge Detection (Lattice Wombling) *R2T1*	Remote sensing image change detection (R1T2, R2T2, R3T2)		
	Zonal R3	Interesting Sub-path discovery (e.g., ecotones) *R3T1*			

We also describe representative techniques used to discover these patterns. Since the goal of this paper is to identify broad gaps in the research, readers interested in a more comprehensive discussion are invited to consult more specialized literature on specific change patterns. Also, we restrict the discussion to the context of change footprint pattern discovery. The larger realm of general techniques in time series analysis and image processing (e.g., trend analysis, spectral decomposition) is not included here.

We begin by classifying change footprints along two dimensions: temporal and spatial. Temporal footprints are of four types, namely, single snapshot (T1), set of snapshots (T2), point in a long series (T3), and interval in a long series (T4). Single snapshot (T1) means that the change is purely spatial without any temporal context. 'Set of snapshots' (T2) indicates that the change is between two or more versions of the same spatial field, e.g., different satellite images of the same region. T3 refers to a single time instance in a long series of data. T4 is a long time period in a long series of data.

Spatial footprints in our taxonomy are classified as raster-based or vector-based. Specifically, raster footprints are further classified based on the scale of the pattern, namely, local, focal, and zonal patterns. Vector-based patterns are further classified into four types, including point(s), line(s), Polygon(s), and network footprint patterns. In each of the two parts, we examine all the combinations of the four temporal footprints and the corresponding spatial footprints. In addition, purely temporal patterns (e.g., change points in a time series) are considered as having local spatial footprint and are discussed under the raster-based change footprints only.

Taxonomy of Change Patterns with Raster-based Spatial Footprint

For raster spatial data, we classify the change footprints based on whether their scale is local, focal, or zonal (Table 1). A local change footprint (R1) involves only the attribute at individual locations. A focal change footprint (R2) is between a location and a spatial neighborhood of it. Change patterns with a zonal footprint (R3) refer to the change that occurs in a spatial region (collection of locations as a whole pattern).

As can be observed in Table 1, our taxonomy yields 11 types of change patterns. Among these, we know that three groups have been well studied by the research community, namely, purely spatial change patterns (R2T1, R3T1), time series change patterns (R1T3, R1T4), and image (snapshot) change patterns (R1T2, R2T2, R3T2). In contrast, remaining patterns have received little attention in the literature of change pattern discovery. We first illustrate the known patterns and major discovery techniques and then discuss the remaining ones in the future research needs sections.

Patterns with Purely Spatial Footprint

Focal Spatial Change (R2T1)

A focal spatial change pattern describes a change that occurs between a location and its spatial neighborhood. For example, given a contiguous spatial field, a focal spatial change pattern may be defined as a collection of locations with high gradient. This pattern can be used to characterize phenomena such as the boundaries between different ecological zones where environmental attributes (e.g., gene frequency) change sharply,[7] or as disconnections of value between different soil types.[38] Previous

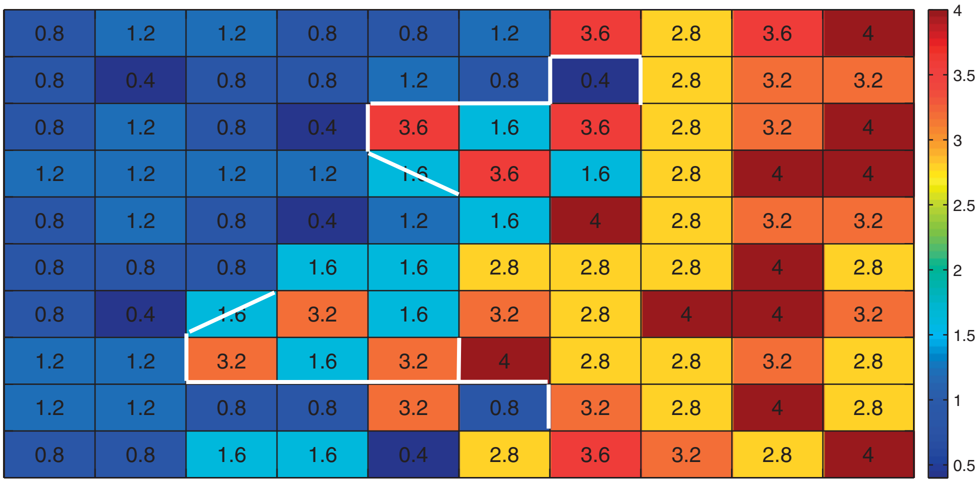

FIGURE 6 | An example of lattice Wombling results on a raster field (best viewed in color).

investigation of such patterns has given various names to it, including 'spatial boundary', 'spatial barriers',[39,40] 'spatial (zones) of abrupt change',[41] 'spatial rapid change',[20] etc. A common goal of such work is to find locations with 'significant difference' or 'abrupt change' against its neighboring locations. By finding a collection of such locations, one may draw a boundary or curve between different homogeneous regions. First addressed by Womble,[17] the spatial boundary analysis problem is also known as the spatial Wombling problem. In ecology study, it is also referred to as 'edge detection'.[42]

Spatial Wombling techniques for the discovery of focal spatial change footprint have been developed. A simple approach is called lattice Wombling.[7,42] Given a raster spatial data field, the lattice Wombling algorithm evaluates all the grid intersections by computing the change magnitude based on the values in the four surrounding cells. In a regular gridded field, the magnitude can be computed as the mean gradient along the four directions. Connecting grid intersections with the top k% will generate boundary linear change footprint pattern. Figure 6 shows an example of a spatial raster field and two collections of focal changes discovered by lattice Wombling.

Zonal Spatial Change (R3T1)

A zonal spatial change pattern describes a spatial zone in a raster field where a transition of data attributes occurs. This pattern may characterize phenomena such as rapid environmental change across different ecological zones. For example, the Sahel region in Africa is a transitional zone between the Sahara Desert and the tropical savanna and grassland. Vegetation cover increases rapidly from north to south. These areas are also referred to as ecotones.[43] Compared to the boundary footprint formed by a collection of focal changes, a spatial zonal change footprint pattern has two-dimensional footprint and exhibits more information about the spatial process within it. This also distinguishes the zonal change footprint pattern from the boundary/line patterns discovered by edge detection or cartography line generation.[44]

Zhou et al.[45] proposed an indirect way to approach this problem by discovering change sub-paths along orthogonal (e.g., longitudinal and latitudinal) directions. The problem is converted to finding 'interesting sub-paths' in each spatial path. Given a spatial path, an algebraic (nonmonotonic) interest measure, and a Boolean test, the technique employs a sub-path enumeration and pruning (SEP) approach to efficiently find all the dominant interesting sub-paths. It enumerates and evaluates all the sub-paths in a top-down manner with proper pruning. This computational framework allows the user to specify interest measures to define the pattern. For example, a change sub-path can be modeled by a slope or high linear regression coefficient. In the same work, the authors also proposed an algebraic interest measure called a 'sameness degree' based on an aggregate function of piecewise difference. Figure 7 shows the African vegetation cover dataset (a) and the longitudinal change (north to south) sub-paths outline the footprint of the Sahel region (marked in red), as shown in (b).

Patterns with Purely Temporal Footprint

Time Point Change (R1T3)

A time point change refers to a change occurring at a single time point or in one unit time intervals in a time series. Time point change patterns have been explored

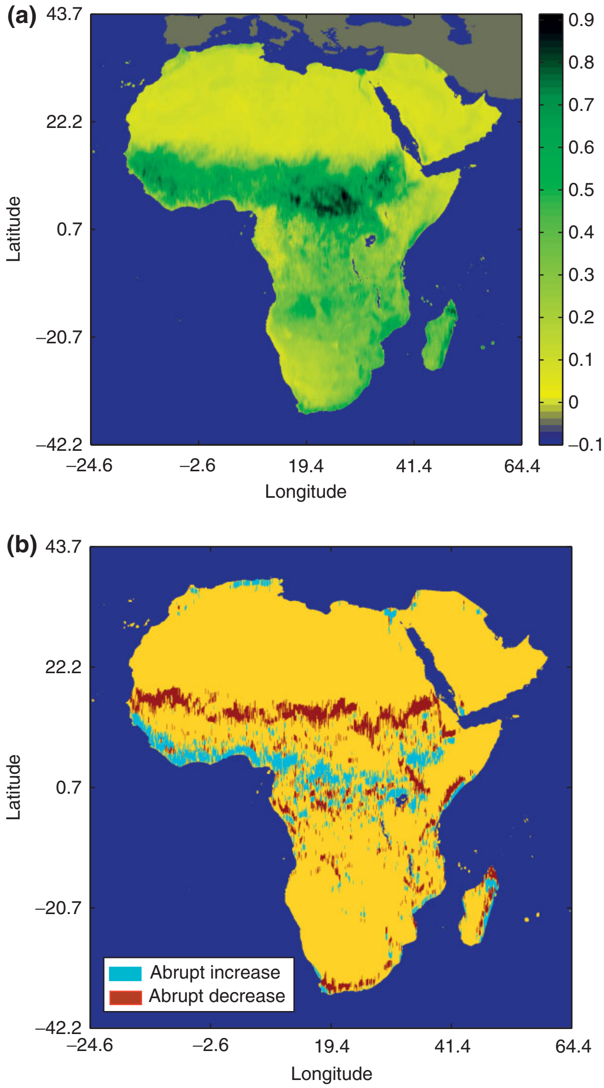

FIGURE 7 | An example of spatial zonal change footprint (best viewed in color). (a) Vegetation cover (in NDVI) in Africa, August, 1981. (b) Footprints of spatial zonal change patterns with longitudinal changes in vegetation cover of Africa.

in various applications to find phenomenon such as the time instance of system fault using sensor signal series, year of abrupt climate change in rainfall and temperature time series, and time of land cover change in remote sensing satellite data series at a location. As discussed previously, such changes can be defined as a shift in statistical parameters (e.g., mean), a change in the model fitted to the data, or a change in derived attributes (e.g., prediction results). In the literature of time series analysis, this problem is also referred to as 'change point detection', 'abrupt change detection', or simply 'time series change detection'.

There is a large body of literature on techniques that identify time point change patterns in time series.[2,46–51] Several survey papers have classified and reviewed existing techniques on change point detection.[9–11,52–60] We introduce one of the major techniques, namely, the CUmulative SUM (CUSUM) Chart.[61] CUSUM is a sequential technique that assumes that observations are recorded at regular intervals and estimates the point at which the change took place by detecting changes in a parameter, θ, of the data distribution. Given a time series x_1, x_2, … x_n at time t_1, t_2, … t_n, the approach keeps a cumulative sum of a score $S_i = \sum_1^i s_i$ at each time point. In an online detection scheme, once the sum of the score S_i exceeds a threshold h, a change can be identified. In the scenario of offline detection, the change point can be identified when the difference between current S_i and the historical minimum S_j is maximized. The score s_i can be defined by the mean, likelihood ratio, or standard deviation of the data. For example, the mean-based score can be written as $s_i = x(i) - \mu_0$, where μ_0 is the normal mean of the data. Figure 8 shows the output of the CUSUM with this setting on the Sahel rainfall index time series data, using an offline manner. The normal mean was estimated using the mean of the entire dataset. As can be seen, the score (represented by the green curve) reaches the maximum at year 1967, marking 1967 as the change point identified by CUSUM. A number of other techniques based on the CUSUM framework appear in the literature as well.[62,63]

Another major technique for change point detection is time series segmentation.[64–67] For simplicity, we introduce linear segmentation methods. Linear segmentation algorithms take a time series as an input and return a piecewise linear representation of the time series by approximating it with a number of straight lines. The problem may be approached in several ways. The sliding window approach starts from the left-most point, attempting to approximate points to the right with a longer segment. If the fitting error exceeds a user specified threshold when adding a new data point to the current segment, the current segment is finalized, and a new segment starts from the new point. This is done repeatedly until the all the data points are examined. Figure 9 shows the output of a sliding window approach, which finds 9 change points in a time series (represented by vertical dotted lines). The threshold for adding a new point to the current approximate segment is set to 'at most 5° change in slope of the segment'. Top-down and bottom-up are two other approaches for time series segmentation. The top-down approach starts with the entire time series and iteratively finds the best change point for the current segment.[68,69] Bottom-up

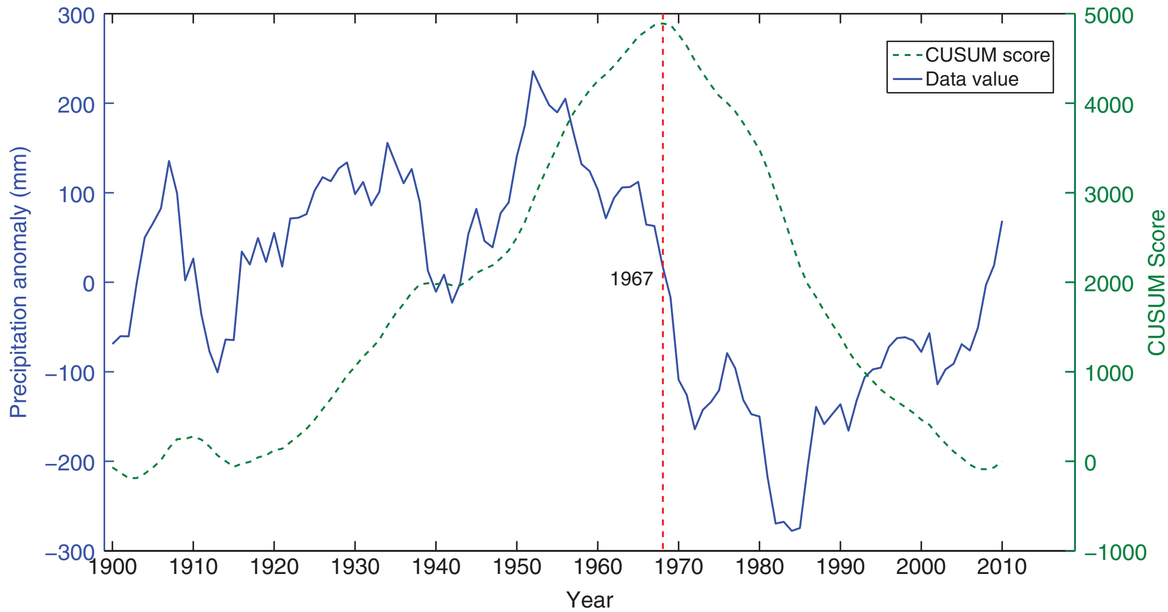

FIGURE 8 | A time point change footprint pattern discovered by CUSUM.

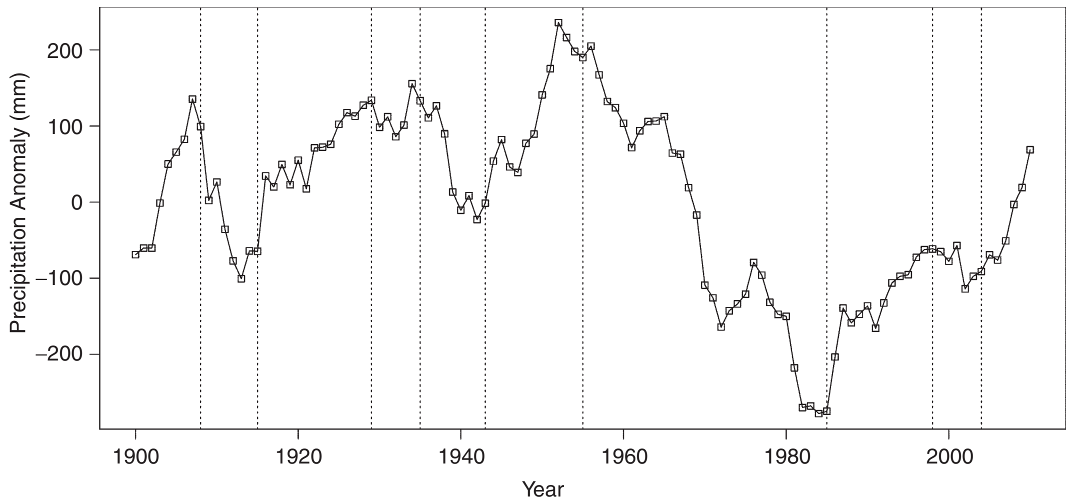

FIGURE 9 | Example of time point change identified by time series segmentation.

approaches start with the unit segments and iteratively merge two segments with minimum cost.[70]

Time Interval Change (R1T4)

A time interval change pattern is a change in a time series that may last for a number of consecutive time points. In contrast to the abrupt change characterized by the point change pattern, the change interval pattern outlines the duration of rapid or gradual change processes. The time interval change pattern is mostly addressed in a specific domain context, such as climate change and land cover change. For example, the Sahel region in Africa experienced a steady yet fast decrease in rainfall in the late 1960s.[71,72] Given a time series of annual precipitation in the Sahel region, this decrease over time can be characterized as a change interval starting from 1968 to 1973 where the precipitation dropped 12% in 5 years.[21] This pattern has also been called 'abrupt change'[21] or an 'interval/temporal sub-path of abrupt change'.[45]

The interesting sub-path discovery technique introduced previously[45] can also be applied on temporal data to find temporal change interval footprints. For example, given a smoothed Sahel rainfall index time series, a 'sameness degree' interest measure can be applied to a few major intervals with precipitation increase or decrease. Figure 10 shows the result of this approach on the smoothed Sahel rainfall

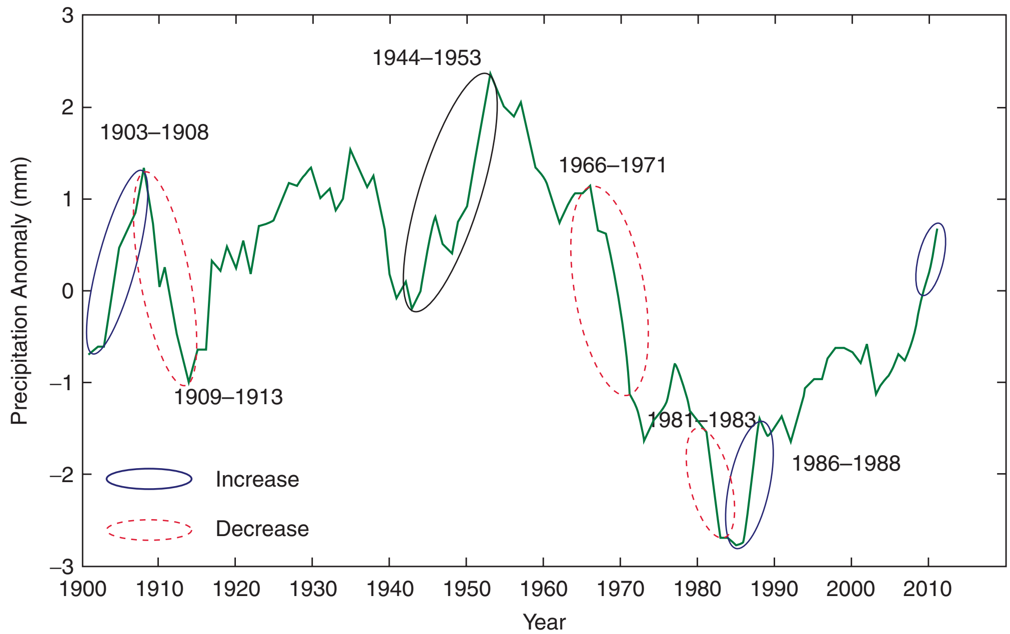

FIGURE 10 | Example of interval change on a time series found by the interesting sub-path discovery method.

index dataset, where a number of time intervals with persistently rapid change are identified.

Patterns with Spatial and Temporal Footprint

Spatial changes between snapshots (R1T2, R2T2, R3T2)

Application domains such as remote sensing and medical image processing are particularly interested in discovering changes that occur between two or more snapshots (e.g., satellite images) of the same spatial framework. For example, in remote sensing, changes between satellite images help identify land cover change due to human activity, natural disasters, or climate change.[73–75] This problem is widely studied as the 'change detection' problem.[76–82]

Given two geographically aligned raster images, the change detection problem aims to find a collection of pixels that have significant changes between the two images. Formally it can be written as:

$$B(x) = \begin{cases} 1, & \text{if there is a signifcant} \\ & \text{change at pixel x} \\ 0, & \text{otherwise} \end{cases}$$

where B is the desired binary image of decisions.[83] In this definition, a change at a pixel is assumed to be independent of changes at other pixels. We thus classify this pattern as a local change between snapshots (R1T2). In addition, alternative definitions have assumed that a change at a pixel is also associated with its neighborhoods.[84] For example, the pixel values in each block may be assumed to follow a Gaussian distribution.[85] We refer to this type of change footprint pattern as a focal spatial change between snapshots (R2T3).

Researchers in remote sensing and image processing have also tried to apply image change detection to objects instead of pixels.[86–88] The assumption is that images are composed of homogeneous segments (i.e., objects). A change pattern between two images is defined as a significant difference of an object in the two snapshots. Since the change footprint is a group of pixels (object), we classify these patterns as a zonal spatial change between snapshots (R3T2).

A well-known technique for detecting a local change footprint is simple differencing. The technique starts by calculating the differences between the corresponding pixels' intensities in the two images. A change at a pixel is flagged if the difference at the pixel exceeds a certain threshold. Many methods for selecting this threshold have been discussed in the literature. A variation of this technique known as Change Vector Analysis (CVA) has been used for multi-spectral images.[89] In this case, each pixel is represented by a feature vector with features representing different spectral channels. The difference between the feature vectors of corresponding pixels is used to detect a change at the pixel. Figure 11 shows an example output of simple differencing. Dots in Figure 11(c) show the locations with changes exceeding 60% of the maximum magnitude between the two images of land surface impervious data[90] shown in Figure 11(a) and (b).

Alternative approaches have also been proposed to discover focal change footprints between images. For example, the block-based density ratio test detects change based on a group of pixels, known as a block,[91,92] rather than a pixel-by-pixel approach. This technique employs hypothesis testing where the null hypothesis corresponds to no change at a given pixel and the alternative hypothesis corresponds to a change. The likelihood ratio is calculated at the pixel and then compared to a defined threshold. If the likelihood ratio exceeds the threshold, the alternative hypothesis is selected; otherwise, the null hypothesis is selected.

Object-based approaches in remote sensing[76,88,93] employ image segmentation techniques to partition the image into homogeneous 'objects'.[44] Classification methods (e.g., decision tree) are then used to classify object pairs in the two images into no-change classes (e.g., water–water) or change classes (e.g., bare land to built-up).

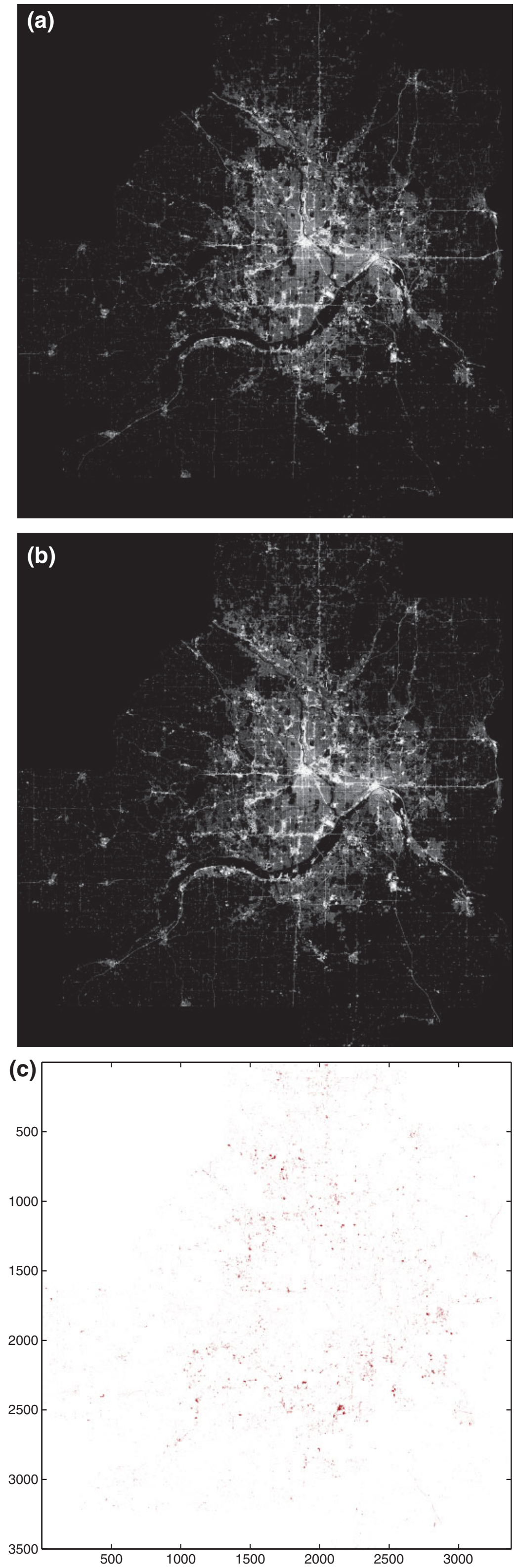

FIGURE 11 | An example of pixel-wise change detection outputs. (a) Impervious surface image of an area in 1986. (b) Impervious surface image of the same area in 1991. (c) Locations with difference exceeding 60% of the maximum magnitude between (a) and (b).

Taxonomy of Change Patterns with Vector-based Spatial Footprint

Spatial footprints on vector data are classified as points (V1), line segments (V2), polygons (V3), and spatial networks (V4). Table 2 shows our classification of vector-related change footprint patterns.

As can be observed, among the 20 possible combinations, we identified three change footprint patterns that have been explored in the literature: line changes (V2T1), polygon changes (V3T1), and ST volumes (polygon with time interval) changes (V3T4).

Patterns with Purely Spatial Footprint

Spatial Lines Change (V2T1)

Given a set of connected polygons, boundary line segments can be identified such that the attributes in the two adjacent polygons are significantly different. As noted previously, the problem of finding spatial boundaries is commonly known as the spatial Wombling problem. In this particular setting, the problem is known as areal Wombling or polygon Wombling.[19] The spatial line change footprint pattern is of interest to applications such as public health where disease information may be collected in an aggregated format (e.g., total cases in each county) due to confidentiality requirement.[19] Areal Wombling helps find boundaries separating high-risk and low-risk counties.

Areal Wombling can be done in a similar way as lattice Wombling. The change magnitude at each

TABLE 2 | Classification of ST Vector Change Footprints

		Temporal			
		Single Snapshot (T1)	Set of Snapshots (T2)	Point in Time series (T3)	Interval in time series (T4)
Spatial footprint	Point(s) (V1)				
	Line segment (V2)	Find boundary lines separating areas with high and low risk of disease (V2T1)			
	Polygon (V3)	Find the regions where the risk of disease is higher inside than others (V3T1)			Find a region and a time interval where the risk of disease is increasing during this interval (V3T4)
	Spatial network (V4)				

Patterns that have received research attention are displayed with examples.

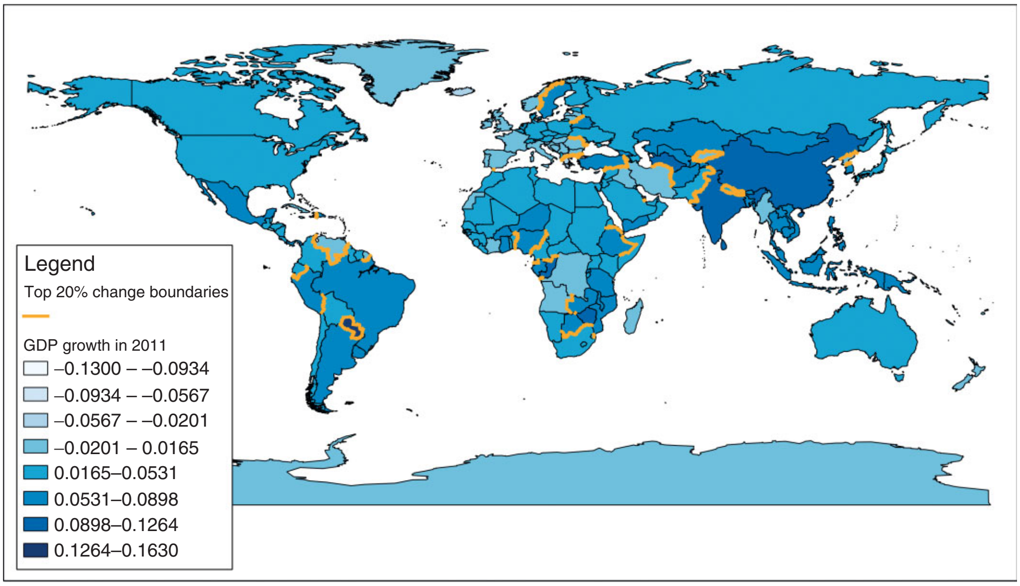

FIGURE 12 | Change boundary (line) footprints on the world GDP growth data.

boundary line can be computed as the difference of attribute values between adjacent polygons. The changes exceeding a threshold or in the top k% form the output footprint. Figure 12 shows an example of boundaries highlighting significant difference of national GDP growth in 2011.[28] It is done by selecting the boundaries with the top 20% highest difference in GDP growth between the two neighboring countries.

Statistical approaches have been designed to evaluate the significance of these footprints. For example, in a hierarchical Bayesian approach,[19] a boundary likelihood value (BLV) is estimated for each boundary using the Markov Chain Monte Carlo (MCMC) method. Similar approaches have been developed on geostatistical[18] and point process[94] data models.

Spatial polygon change (V3T1)

A change with a polygon footprint delineates a region where some attributes have significantly changed. For example, given a spatial point process dataset representing disease cases, a spatial polygon change pattern draws a polygon in which the density of points (disease) inside the polygon is significantly higher than outside. Such patterns can also be identified on a set of polygons (representing geopolitical regions) where the total disease count of each polygon is specified. This problem is referred to as spatial cluster detection.[95] In public health/epidemiology, finding spatial clusters with a higher density of disease is of great interest to understand the distribution and spread of disease.

Kulldorff et al.[95] proposed a spatial *scan statistics* framework for disease outbreak detection. Given a number of fixed locations (e.g., hospitals) and the number of disease cases at each location, the early version of this work focuses on finding the most likely spatial region where the relative risk of disease inside is higher than outside. The spatial scan statistics employs a likelihood ratio test where the null hypothesis H_0 is that the probability of disease inside a region is the same as outside the region, and the alternative hypothesis H_1 is that there is a higher probability of disease inside than outside. Assuming that the total disease count in a region follows a Poisson model, the likelihood ratio can be formally written as: $\frac{\left(\frac{n_z}{\mu_z}\right)^{n_z}\left(\frac{n_G-n_z}{\mu_G-\mu_z}\right)^{n_G-n_z}}{\left(\frac{n_G}{\mu_G}\right)^{n_G}} I\left(\frac{n_z}{\mu_z} > \frac{n_G-n_z}{\mu_G-\mu_z}\right)$, where n_G and n_z are the number of observations in the entire area and in candidate region Z respectively, and μ_G and μ_z are the expected number of disease reports in the entire area and the region Z. The expected numbers may be derived from a baseline population at risk or estimated using nonparametric models according to the size of the region. All the spatial regions represented by a circle or ellipsoids in the spatial framework are enumerated and the one that maximizes the likelihood ratio score is identified as the candidate. Finally, the statistical significance of the candidate cluster is tested by a Monte Carlo simulation. Figure 13(a) shows an example of spatial point dataset where each point represents a fixed location (numbered 1–20). The number of disease cases and population in each location are labeled (i.e., cases/population). The most likely cluster outlined by the red circle is shown in Figure 13(b), assuming the data follows a discrete Poisson model. As can be observed, locations in the circle have a higher rate of disease compared to the outside (about 1/100).

Later extensions on the same ideas have explored normal,[96] exponential,[97] ordinal,[98] and nonparametric models.[99] The scan statistic has also been generalized to handle ST point (event report) datasets[100,101] and irregular-shaped clusters.[102] Computational efficiency of the scan statistic method was further optimized by employing a top-down pruning of the search space.[103] A Bayesian version of scan statistics uses an inference instead of the frequentist hypothesis testing.[104]

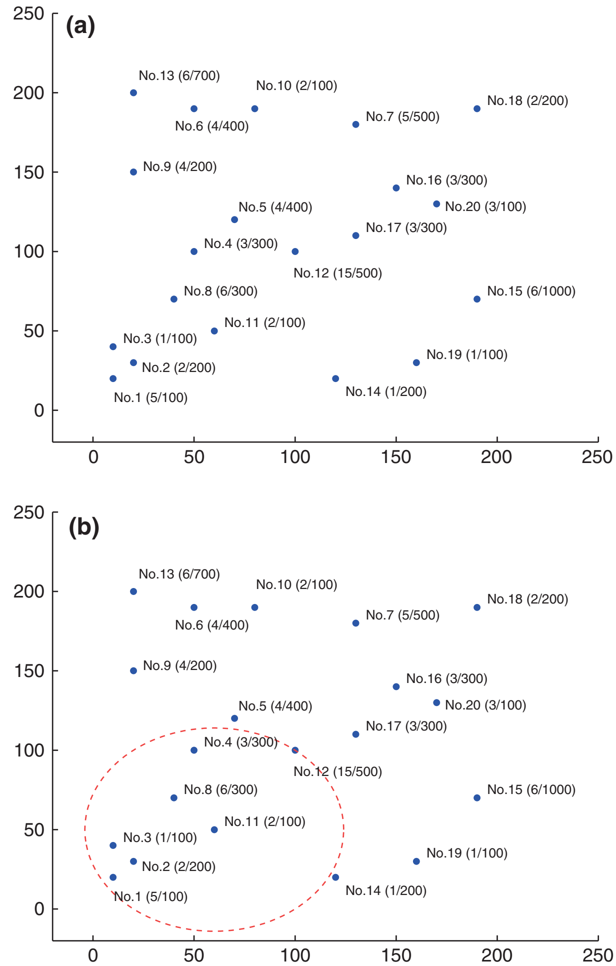

FIGURE 13 | Input and output example of spatial scan statistics. (a) A point process dataset. (b) The mostly likely cluster discovered

SaTScan[106] is a software tool developed for discovering spatial and ST clusters which integrates the above models.

Patterns with Spatial and Temporal Footprint

ST Volume Change Footprint (V3T4)

The ST volume change footprint represents a change process occurring in a spatial region (characterize by a polygon) during a time interval. It quantifies both the spatial coverage and the temporal duration of a nonstationary ST process. For example, an outbreak event of a disease can be defined as an increase in disease reports in a certain region during a certain time window up to the current time. Change patterns known to have an ST volume footprint include the ST scan statistics (introduced above) and emerging ST clusters defined by Neill et al.[107]

Given an ST point process dataset, and a baseline risk probability P, an emerging cluster is defined as a spatial region S and a time window

TABLE 3 | A Full List of Raster ST Change Footprint Patterns

		Temporal Footprint			
		Single Snapshot (T1)	Between Few Snapshots (T2)	Points in Time Series (T3)	Intervals in Time Series (T4)
Spatial footprint	Local R1	*R1T1 (N/A)*		Change point detection (e.g., CUSUM) *R1T3*	Change interval discovery on time series *R1T4*
	Focal R2	Edge Detection (Lattice Wombling) *R2T1*	Remote sensing image change detection (T2)	Find a location where the time series changes differently than its spatial neighbors at a time point. *R2T3*	Find a location where the time series changes differently than its spatial neighbors during a time interval. *R2T4*
	Zonal R3	Interesting Sub-path discovery (e.g., ecotones) *R3T1*		Find a region where the aggregate/summary time series has a change at a time point. *R3T3*	Find a region where the aggregate time series has a persistent change during a time interval. *R3T4*

Patterns yet to be further addressed are also illustrated by typical questions.

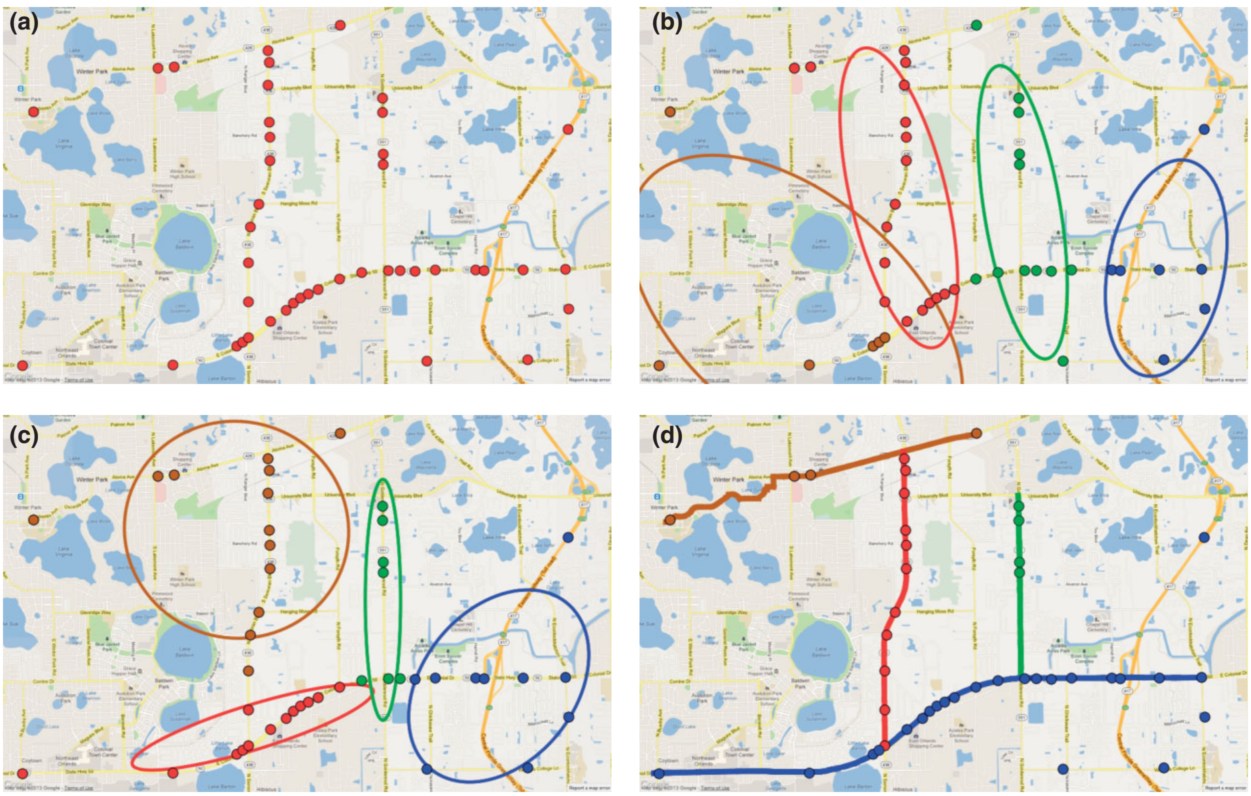

FIGURE 14 | An example of network footprint to summarize high activity patterns. (a) A spatial network with activities. (b) Polygon footprint to summarize activities with Euclidian distance (by CrimeStat)[105] (c) Polygon footprint to summarize activities with network distance (by CrimeStat). (d) Network footprint to summarize activities.

starting at $t_{\min}$ such that the risk of each instance in S in days $t_{\min}$, $t_{\min+1}$, ..., T is monotonically increasing: $q_i \times P$, where $1 \leq q_{\min} \leq q_{\min+1} \leq \ldots \leq q_T$. The null and alternative hypotheses are defined as follows: H_0: the probability that one instance is at risk in the cluster for each day are the same and equals the expected value P, and H_1: The above probability in days $t_{\min}$, $t_{\min+1}$, ..., T is $q \times P$ where $1 \leq q_1 \leq q_2 \leq \ldots \leq q_T$. The following likelihood ratio is tested over all the spatial regions S and start time $t_{\min}$

TABLE 4 | A Full List of Vector ST Change Footprint Patterns

		Temporal			
		Single Snapshot (T1)	Set of Snapshots (T2)	Point in Time series (T3)	Interval in Time Series (T4)
Spatial footprint	Point(s) (V1)	N/A	Find significant shift of locations of a point process at two different time (V1T2)	Which county seats abruptly change their location? When? (V1T3)	Find county seats that change their locations frequently during some time period in Chinese history. (V1T4)
	Line segment (V2)	Find the significant boundaries highlighting the high-risk and low-risk of disease (V2T1)	What is the difference in direction of a road/river after earthquake? (V2T2)	When and where did a person change his/her route to work? (V2T3)	When & where did the Mississippi river significantly change its route in the last century? (V2T4)
	Polygon (V3)	Find the area where the risk of disease is higher than others (*V3T1*)	Which county in China has changed its area significantly between 800 AD and 900 AD? (V3T2)	Which bird habitats significantly shank their area due to urban sprawl? When? (V3T3)	Find a region and a time interval where the risk of disease is increasing during this interval (V4T3)
	Spatial network (V4)	Find the sub-network of the road networks where the risk of crime becomes higher than other parts (V4T1)	Find the sub-network that summarizes the change between the two snapshots of the network (V4T2)	Find a time/places when/where the railroad network significantly grows (V4T3)	Find the expanding direction of the network during a time period (V4T4)

Patterns that may be further studied are also illustrated by typical questions.

pairs: $\left(\mathrm{Max}_{1\le q_{t_{\min}}\cdots\le q_T}\prod q_t^{C_t}\cdot e^{-q_t B_t}/e^{-B}\right)$, where C_t and B_t are the total observations and total expected number (baseline) of disease reports, respectively, in day t in S, and B is the total number of expected diseases (baseline) for the entire window in S. The most likely pair $(S, t_{\min})$ is the output pattern.

FUTURE DIRECTIONS AND RESEARCH NEEDS

An important goal of our change footprint taxonomy was to uncover gaps in the research. In this section, we discuss the change footprint patterns that have yet to receive much, if any, attention by researchers. Some may be interesting to data scientists and could be explored in the future. To highlight interesting vector change footprint patterns, we show two case studies from historical GIS applications at the end of this section. We believe that they represent exciting examples of avenues for future research.

Raster-based Patterns

Table 1 shows that most of the raster footprint patterns have been explored. However, we did identify four footprint patterns that may be further investigated, i.e., spatial focal and zonal changes that occur at a time point or a time interval in time as shown in Table 3. Hence we suggest some change phenomena that appear to exhibit such footprints.

Spatial Focal Change at a Time Point

A spatial focal change at a time point can be modeled in a spatial time series dataset. One example of a spatial focal change at a time point may be described as follows: given a spatial time series dataset, find a location s and a time point t where the time series at s starts to behave very differently from time series in its spatial neighborhood at t. This pattern may be of interest to research in climate science and remote sensing. For example, given a spatial time series dataset representing land cover indices, one may be interested in finding locations and the corresponding

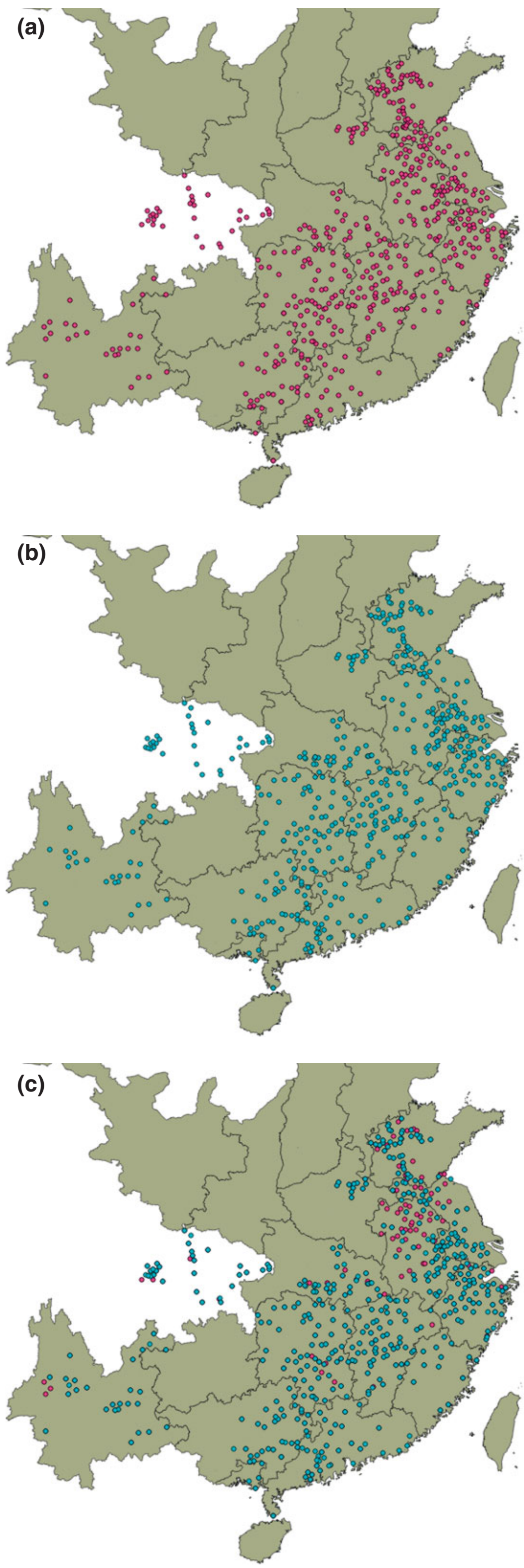

FIGURE 15 | A comparison of county seats locations at 300 and 400 AD in China (best viewed in color). (a) County seats of China in 300 AD. (b) County seats of China in 400 AD. (c) Change of county seat locations between 300 AD and 400 AD.

time point of an abrupt land cover change due to forest fire or other human introduced events. Due to spatial autocorrelation, we know that neighboring time series tend to behave similarly in normal scenarios. A focal change in space can distinguish a time series from its neighbors in the case of anomalous events, even if the change on the time series itself is not very abrupt.

Similarly, we can define a spatial focal change in a time interval as an event that a time series exhibits behavior significantly different from its spatial neighbors during a contiguous time period.

Spatial Zonal Change at a Time Point/in an Interval

This pattern may be characterized as an abrupt change (point) or rapid/gradual change (interval) that occurs on the aggregated time series of a spatial region. This pattern may be explored to address problems such as finding regional climate change patterns, and automatic finding of large scale of land cover change, etc.

Vector-based Patterns

Compared to raster ST change footprint patterns, vector ST change patterns have been much less intensively explored. The existing patterns and techniques focus mostly on finding changes in non-ST attributes (e.g., count of disease) with vector spatial footprints. We believe that there is also great value in understanding the change patterns of ST attributes of vector spatial objects. For example, attention may be given to spatial change patterns with network footprints (V4T1). Figure 14 shows primary results from exploration on activity summarization with a network footprint[108] where the activities on spatial networks in (a) are better summarized using the paths in (d) rather than using the ellipsoids or circles in (b) and (c). Future extension of this work may focus on finding statistically significant high crime risk in some path/sub-networks. Such work may help public safety authorities better understand crime patterns and respond accordingly. In Table 4 we list numerous examples of questions users may pose in a variety of applications in addition to those listed in Table 2.

Change Footprint Patterns in Nonscientific Domains

Change footprint patterns are valuable to study even in domains far removed from scientific domains. One field where it is critical to understand how spatial relationships change over time is the study of history.[109] Such changes can be modeled as ST change with vector spatial footprints. Specifically,

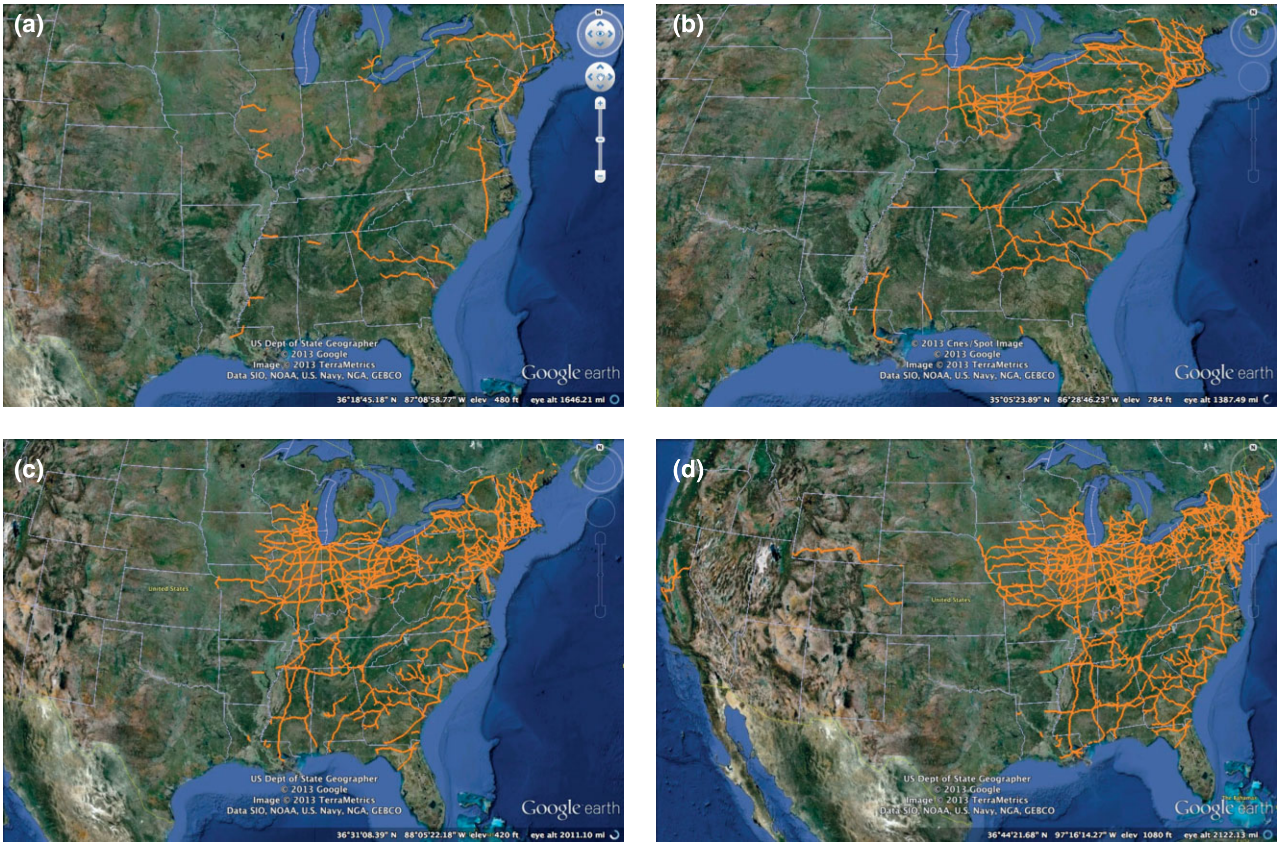

FIGURE 16 | U.S. railroad network expansion in the 19th century (best viewed in color). (a) U.S. railroad network in 1840. (b) U.S. railroad network in 1850. (c) U.S. railroad network in 1861. (d) U.S. railroad network in 1870.

we illustrate two application examples: the China historical jurisdiction change footprint patterns and the U.S. railroad change footprint patterns.

Case study 1

Historians studying Chinese history to trace changes in administrative hierarchies down to the 'Xian'(county) level from 221 BC to 1911 AD. The county seats (capital) in Chinese history are usually used to represent the real location of the counties. Vector datasets representing the county seats at different times are available from the China Historical GIS project.[110,111] In this dataset, each historical county seat is represented as a ST point, where the location, valid time period, and non-ST attributes (e.g., name) are available. For example, Figure 15(a)–(c) show different snapshots of the county seat locations in 300 and 400 AD, and the difference between the 2 years in Chinese history.

Given such an ST vector dataset, one may define a few change patterns with vector ST footprints. For individual points, one may identify the time period with frequent change of location. A regional pattern (e.g., polygon) summarizing a large number of location shifts during a given time period (e.g., 300–400 AD) can be identified to indicate the dynamic of the geo-political relationships in this area. For example, a significant change in point locations can be found in the east of the map as shown in Figure 14(c).

Case Study 2

Railroad growth in the 19th century is another subject of interest to historians. The expansion of the U.S. railroad network contributed significantly to the immigration and employment in the country. The complexity and scale of railroad operations and their uneven extension across the landscape created both intensive and extensive changes in American communities.[112] Understanding the change pattern of the railroad network helps understand U.S. social and economic development in the 19th century.[113]

A historical railroad network dataset is available for this study.[114] Figure 16 shows a few snapshots of this datasets in 1840, 1850, 1861, and 1870.

Given such an ST vector dataset, one can define change patterns with line/network spatial footprints. One possible pattern may be a time interval during which the spatial network grew significantly (e.g., total length, coverage) in a particular area (e.g., Mid-west). For example, in the snapshots a significant expansion of the network can be identified between 1840 and 1851. Another pattern that could be described is the directions in which the spatial network expanded during a certain time period (e.g., 1840–1851). For example, in the given dataset, the network expands toward the west (e.g., Midwest) and southwest between 1840 and 1870.

CONCLUSION AND FUTURE PLAN

This paper proposes a taxonomy of ST change footprints that may be of use to researchers across multiple research domains. We built the taxonomy after conducting a multi-disciplinary review of research in ST change pattern discovery. Our taxonomy achieves two valuable goals. First, it classifies a wide variety of ST change footprints that have already received attention in different domains. Previously, much of this research was hidden from view, so to speak, due to the lack of common terminology across disciplines for discussing similar phenomena. Second, our taxonomy reveals gaps in the research, that is, change footprint patterns that have yet to be studied despite their potential applicability to many real-world problems. We especially note the need for research on ST change footprints on vector data.

In the future, we plan to incorporate other aspects of change patterns in our classification. Currently, the taxonomy mainly focuses on only univariate (single non-ST attribute) techniques. The next step will be to include multivariate definitions and approaches. Also, we want to address issues such as computational structure and statistical modelling of change patterns.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant No. 1029711, IIS-1320580, 0940818 and IIS-1218168 as well as USDOD under Grant No. HM1582-08-1-0017, and HM0210-13-1-0005. We would like to thank the members of the Spatial Database and Spatial Data Mining Group at the University of Minnesota for their helpful comments. We also thank Prof. Joseph Knight for helpful advices and Kim Koffolt for helping improve the readability of this paper.

REFERENCES

1. Box GE, Luceno A, del Carmen Paniagua-Quiñones M. *Statistical Control by Monitoring and Adjustment*, vol. 898. Hoboken, NJ: Wiley; 2011.
2. Lunetta RS, Elvidge CD. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London: Taylor & Francis Ltd; 1999.
3. Singh A. Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 1989, 10:989–1003.
4. Eastman JR, McKendry J. *Change and Time Series Analysis in GIS*. Geneva: UNITAR; 1991.
5. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995, 14:799–810.
6. Chainey S, Tompson L, Uhlig S. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* 2008, 21:4–28.
7. Fagan WF, Fortin MJ, Soykan C. Integrating edge detection and dynamic modeling in quantitative analyses of ecological boundaries. *BioScience* 2003, 53:730–738.
8. Hansen J, Sato M, Ruedy R. Perception of climate change. *Proceedings of the National Academy of Sciences* 2012, 109:E2415–E2423.
9. Basseville M, Nikiforov IV. *Detection of Abrupt Changes: Theory and Application*, vol. 15. Englewood Cliffs, NJ: Prentice Hall; 1993.
10. Shaban S. Change point problem and two-phase regression: an annotated bibliography. *International Statistical Review* 1980, 48:83–93.
11. Zacks S. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. *Recent advances in statistics: papers in honor of Herman Chernoff on his sixtieth birthday* 1983:245.
12. Chen J, Gupta AK. On change point detection and estimation. *Communications in Statistics-Simulation and Computation* 2001, 30:665–697.
13. Coppin P, Jonckheere I, Nackaerts K, Muys B, Lambin E. Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing* 2004, 25:1565–1596.

14. Barkat M. *Signal Detection and Estimation*. Boston, MA: Artech house; 2005.
15. Lu D, Mausel P, Brondizio E, Moran E. Change detection techniques. *International Journal of Remote Sensing* 2004, 25:2365–2401.
16. Wong WK, Neill DB. Tutorial on event detection. Presentation in ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2009, Paris, France; Available at: http://www.pptsearch.net/download.php?fid=109746 (Accessed March 1, 2013).
17. Womble WH. Differential systematics. *Science* 1951, 114:315.
18. Banerjee S, Gelfand AE. Bayesian Wombling: curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association* 2006, 101:1487–1501.
19. Lu H, Carlin BP. Bayesian areal wombling for geographical boundary analysis. *Geographical Analysis* 2005, 37:265–285.
20. Oden NL, Sokal RR, Fortin MJ, Goebl H. Categorical wombling: detecting regions of significant change in spatially located categorical variables. *Geographical Analysis* 1993, 25:315–336.
21. Shekhar S, Evans MR, Kang JM, Mohan P. Identifying patterns in spatial information: a survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2011, 1:193–214.
22. Joint Institute for the Study of the Atmosphere and Ocean(JISAO). Sahel Rainfall Index. Available at: http://jisao.washington.edu/data/sahel/ (Accessed October 15, 2012).
23. Tucker CJ, Pinzon JE, Brown ME. Global Inventory Modeling and Mapping Studies, Global Land Cover Facility: University of Maryland, Maryland; 2004, 1–20.
24. Pinzon J, Brown ME, Tucker CJ. Satellite time series correction of orbital drift artifacts using empirical mode decomposition In: Huang N, ed. *Hilbert-Huang Transform: Introduction and Applications*; 2005, 167–186.
25. Tucker CJ, Pinzon JE, Brown ME, Slayback D, Pak EW, Mahoney R, Vermote E, El Saleous N. An Extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing* 2005, 26: 4485–5598.
26. Shekhar S, Chawla S. *Spatial Databases: A Tour*. Prentice Hall; Englewood Cliffs, NJ, 2003.
27. Worboys M, Duckham M. *GIS: A Computing Perspective*. Boca Raton, FL: CRC; 2004. ISBN: 0415283752.
28. CIA World Factbook. Available at: http://www.cia.gov/library/publications/the-world-factbook/ (Accessed January 13, 2013).
29. Boriah S, Kumar V, Steinbach M, Potter C, Klooster S. Land cover change detection: a case study. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2008, 857–865.
30. Zhang P, Huang Y, Shekhar S, Kumar V. Correlation analysis of spatial time series datasets: a filter-and-refine approach. In: *Advances in Knowledge Discovery and Data Mining*. New York: Springer; 2003, 532–544.
31. Gilks WR, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*, vol. 2. Boca Raton, FL: Chapman & Hall/CRC; 1995.
32. Tobler WR. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 1970, 46:234–240.
33. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC; 2003.
34. Chandola V, Hui D, Gu L, Bhaduri B, Vatsavai RR. Using Time Series Segmentation for Deriving Vegetation Phenology Indices from MODIS NDVI Data. Data Mining Workshops (ICDMW), 2010 IEEE International Conference on 2010, 202–208.
35. Guralnik V, Srivastava J. Event detection from time series data. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, CA: ACM, 1999.
36. Fang Y, Ganguly AR, Singh N, Vijayaraj V, Feierabend N, Potere DT. Online change detection: Monitoring land cover from remotely sensed data. In: Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on: IEEE: Hong Kong, 2006.
37. Zhu Z, Woodcock CE, Olofsson P. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment* 2012, 122:75–91.
38. Gabriel E, Allard D. Evaluating the sampling pattern when detecting zones of abrupt change. *Environmental and Ecological Statistics* 2008, 15:469–489.
39. Sokal RR, Thomson BA. Spatial genetic structure of human populations in Japan. *Human Biology* 1998, 70:1.
40. Bocquet-Appel J-P, Jakobi L. Barriers to the spatial diffusion for the demographic transition in Western Europe. *Spatial Analysis of Biodemographic Data* 1996, 16:117–129.
41. Barbujani G, Oden NL, Sokal RR. Detecting regions of abrupt change in maps of biological variables. *Systematic Biology* 1989, 38:376–389.
42. Fortin M-J. Edge detection algorithms for two-dimensional ecological data. *Ecology* 1994:956–965.
43. Strayer DL, Power ME, Fagan WF, Pickett ST, Belnap J. A classification of ecological boundaries. *BioScience* 2003, 53:723–729.

44. Douglas DH, Peucker TK. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization* 1973, 10:112–122.
45. Zhou X, Shekhar S, Mohan P, Liess S, Snyder PK. Discovering interesting sub-paths in spatiotemporal datasets: a summary of results. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2011, 44–53.
46. Kawahara Y, Sugiyama M. Change-point detection in time-series data by direct density-ratio estimation. In: Proceedings of 2009 SIAM International Conference on Data Mining (SDM2009), 2009, 389–400.
47. Jandhyala VK, Fotopoulos SB, Hawkins DM. Detection and estimation of abrupt changes in the variability of a process. *Computational Statistics & Data Analysis* 2002, 40:1–19.
48. Adams RP, MacKay DJ. Bayesian online changepoint detection. Technical Report. University of Cambridge, Cambridge, *arXiv preprint arXiv:0710.3742*, 2007.
49. Wolfe DA, Schechtman E. Nonparametric statistical procedures for the changepoint problem. *Journal of Statistical Planning and Inference* 1984, 9:389–396.
50. Chernoff H, Zacks S. Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics* 1964, 35:999–1018.
51. Hawkins DM. Fitting multiple change-point models to data. *Computational Statistics & Data Analysis* 2001, 37:323–341.
52. BarryHawkins D, Hartigan JA. A Bayesian analysis for change point problems. *Journal of the American Statistical Association* 1993:309–319.
53. Siegmund D. Boundary crossing probabilities and statistical applications. *The Annals of Statistics* 1986, 14:361–404.
54. Krishnaiah P, Miao B. 19 Review about estimation of change points. *Handbook of Statistics* 1988, 7:375–402.
55. Csörgő M, Horvath L. 20 Nonparametric methods for changepoint problems. *Handbook of Statistics* 1988, 7:403–425.
56. Lai TL. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995:613–658.
57. Basseville M. Detecting changes in signals and systems—a survey. *Automatica* 1988, 24:309–326.
58. Bhattacharyya GK. 5 Tests of randomness against trend or serial correlations. *Handbook of Statistics* 1984, 4:89–111.
59. Gustafsson F, Gustafsson F. *Adaptive Filtering and Change Detection*, vol. 1. Boca Raton, FL: Wiley Londres; 2000.
60. Brodsky BE, Darkhovsky BS. *Nonparametric Methods in Change Point Problems*, vol. 243. New York: Springer; 1993.
61. Page ES. Continuous inspection schemes. *Biometrika JSTOR* 1954, 41:100–115.
62. Kucera J, Barbosa P, Strobl P. Cumulative sum charts-a novel technique for processing daily time series of modis data for burnt area mapping in Portugal. In: Analysis of Multi-temporal Remote Sensing Images, 2007. MultiTemp 2007. International Workshop on the: IEEE, Provinciehuis Leuven, Belgium ; 2007.
63. Nopiah Z, Baharin M, Abdullah S, Khairir M, Nidzwan C. The Detection of Abrupt Changes in Fatigue Data by using Cumulative Sum (CUSUM) Method. In: Recent Advances in Applied and Theoretical Mechanics: Proceedings of the 4th International Conference on Applied and Theoretical Mechanics (Mechanics' 08), Cairo, Egypt; 2008.
64. Bingham E, Gionis A, Haiminen N, Hiisilä H, Mannila H, Terzi E. Segmentation and dimensionality reduction. In: 2006 SIAM Conference on Data Mining, Bethesda, MD; 2006.
65. Himberg J, Korpiaho K, Mannila H, Tikanmaki J, Toivonen HT. Time series segmentation for context recognition in mobile devices. In: Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on: IEEE, San Jose, CA; 2001.
66. Keogh E, Chu S, Hart D, Pazzani M. An online algorithm for segmenting time series. In: Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on: IEEE, San Jose, CA; 2001.
67. Keogh E, Chu S, Hart D, Pazzani M. Segmenting time series: a survey and novel approach. *Data Mining in Time Series Databases* 2004, 57:1–22.
68. Shatkay H, Zdonik SB. Approximate queries and representations for large data sequences. In: Data Engineering, 1996. Proceedings of the Twelfth International Conference on: IEEE, New Orleans, Louisiana; 1996.
69. Li C-S, Yu PS, Castelli V. MALM: a framework for mining sequence database at multiple abstraction levels. In: Proceedings of the seventh international conference on Information and knowledge management: ACM, Seattle, WH, 1998.
70. Keogh E, Pazzani MJ. *An enhanced Representation of Time Series which Allows Fast and Accurate Classification, Clustering and Relevance Feedback*. KDD; 1998.
71. Dai A, Lamb PJ, Trenberth KE, Hulme M, Jones PD, Xie P. The recent Sahel drought is real. *International Journal of Climatology* 2004, 24:1323–1331.
72. Narisma GT, Foley JA, Licker R, Ramankutty N. Abrupt changes in rainfall during the twentieth century. *Geophysical Research Letters* 2007, 34:L06710.
73. Bujor F, Trouvé E, Valet L, Nicolas J-M, Rudant J-P. Application of log-cumulants to the detection of

spatiotemporal discontinuities in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 2004, 42:2073–2084.

74. Kosugi Y, Sakamoto M, Fukunishi M, Lu W, Doihara T, Kakumoto S. Urban change detection related to earthquakes using an adaptive nonlinear mapping of high-resolution images. *IEEE Transactions on Geoscience and Remote Sensing Letters* 2004, 1:152–156.
75. Di Martino G, Iodice A, Riccio D, Ruello G. A novel approach for disaster monitoring: fractal models and tools. *IEEE Transactions on Geoscience and Remote Sensing* 2007, 45:1559–1570.
76. Im J, Jensen JR. A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote Sensing of Environment* 2005, 99:326–340.
77. Bazi Y, Bruzzone L, Melgani F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 2005, 43:874–887.
78. Inglada J, Mercier G. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing* 2007, 45:1432–1445.
79. Bovolo F, Bruzzone L. A split-based approach to unsupervised change detection in large-size multitemporal images: application to tsunami-damage assessment. *IEEE Transactions on Geoscience and Remote Sensing* 2007, 45:1658–1670.
80. Celik T. Change detection in satellite images using a genetic algorithm approach. *IEEE Transactions on Geoscience and Remote Sensing* 2010, 7:386–390.
81. Dianat R, Kasaei S. Change detection in optical remote sensing images using difference-based methods and spatial information. *IEEE Transactions on Geoscience and Remote Sensing Letters* 2010, 7:215–219.
82. Moser G, Angiati E, Serpico SB. Multiscale unsupervised change detection on optical images by Markov random fields and wavelets. *IEEE Transactions on Geoscience and Remote Sensing Letters* 2011, 8:725–729.
83. Radke RJ, Andra S, Al-Kofahi O, Roysam B. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing* 2005, 14:294–307.
84. Thoma R, Bierling M. Motion compensating interpolation considering covered and uncovered background. *Signal Processing: Image Communication* 1989, 1:191–212.
85. Aach T, Kaup A. Bayesian algorithms for adaptive change detection in image sequences using Markov random fields. *Signal Processing: Image Communication* 1995, 7:147–160.
86. Chen G, Hay GJ, Carvalho LM, Wulder MA. Object-based change detection. *International Journal of Remote Sensing* 2012, 33:4434–4457.
87. Desclée B, Bogaert P, Defourny P. Forest change detection by statistical object-based method. *Remote Sensing of Environment* 2006, 102:1–11.
88. Im J, Jensen J, Tullis J. Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing* 2008, 29:399–423.
89. Bruzzone L, Prieto DF. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Transactions on Image Processing* 2002, 11:452–466.
90. Remote sensing and geospatial analysis Laboratory, University of Minnesota, Twin Cities Metro Area Impervious Surface Data 1998 and 2002, St Paul, MN. Available at: http://land.umn.edu (Accessed January 20, 2013).
91. Aach T, Kaup A, Mester R. Statistical model-based change detection in moving video. *Signal Processing* 1993, 31:165–180.
92. Rignot EJ, van Zyl JJ. Change detection techniques for ERS-1 SAR data. *IEEE Transactions on Geoscience and Remote Sensing* 1993, 31:896–906.
93. Yakimovsky Y. Boundary and object detection in real world images. *Journal of the ACM (JACM)* 1976, 23:599–618.
94. Liang S, Banerjee S, Carlin BP. Bayesian wombling for spatial point processes. *Biometrics* 2009, 65:1243–1253.
95. Kulldorff M. A spatial scan statistic. *Communications in Statistics-Theory and Methods* 1997, 26:1481–1496.
96. Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics* 2009, 8:58.
97. Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. *Biometrics* 2007, 63:109–118.
98. Jung I, Kulldorff M, Klassen AC. A spatial scan statistic for ordinal data. *Statistics in Medicine* 2007, 26:1594–1607.
99. Kulldorff M, Heffernan R, Hartman J, Assuncão R, Mostashari F. A space–time permutation scan statistic for disease outbreak detection. *PLoS Medicine* 2005, 2:e59.
100. Kulldorff M, Athas W, Feurer E, Miller B, Key C. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health* 1998, 88:1377–1380.
101. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of*

the Royal Statistical Society: Series A (Statistics in Society) 2001, 164:61–72.

102. Costa MA, Kulldorff M, Assuncao RM. A space time permutation scan statistic with irregular shape for disease outbreak detection. *Advances in Disease Surveillance* 2007, 4:243.
103. Neill DB, Moore AW. Rapid detection of significant spatial clusters. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 2004, 256–265.
104. Neill D, Moore A, Cooper G. A Bayesian spatial scan statistic. *Advances in Neural Information Processing Systems* 2006, 18:1003.
105. Ned Levine. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3)*. Houston, TX/Washington, DC: Ned Levine & Associates/National Institute of Justice; 2010.
106. Kulldorff M. SaTScan user guide for version 7.0. 2008. (*Accessed January* 18, 2006).
107. Neill DB, Moore AW, Sabhnani M, Daniel K. Detection of emerging space-time clusters. In: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining: ACM; New York: 2005.
108. Oliver D, Shekhar S, Kang JM, Laubscher Renee, Carlan V, Bannur A. A K-Main Routes Approach to Spatial Network Activity Summarization. IEEE Transactions on Knowledge and Data Engineering, IEEE computer Society Digital Library. Washington DC: IEEE Computer Society, 2013.
109. Hillier A, Knowles AK. *Placing History: How Maps, Spatial Data, and GIS are Changing Historical Scholarship*. Redlands, CA: ESRI press; 2008.
110. Berman ML, GIS CH. A data model for historical GIS: the CHGIS time series. Cambridge, MA, Harvard Yenching Institute Technical Report 2003.
111. Harvard University and Fudan University. The China Historical GIS Project. Available at: http://www.fas.harvard.edu/~chgis/index.html (Accessed January 20, 2013).
112. University of Nebraska-Lincoln. Railroads and the Making of Modern America. Available at: http://railroads.unl.edu/ (Accessed January 20, 2013).
113. Thomas WG. *The Iron Way: Railroads, the Civil War, and the Making of Modern America*. New Haven, CT: Yale University Press; 2011.
114. University of Nebraska-Lincoln. Historical GIS: The 1840-1845-1850-1861-1870 Railroad System in America, State and National shape files. 2011. Available at: http://railroads.unl.edu/resources/ (Accessed January 20, 2013).

