Article

Tri-Clustering Based Exploration of Temporal Resolution Impacts on Spatio-Temporal Clusters in Geo-Referenced Time Series

Xiaojing Wu ^{1,*} and Donghai Zheng ²

¹ Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

² Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100864, China; zhengd@itpcas.ac.cn

* Correspondence: xiaojingwu@bnu.edu.cn

Received: 11 March 2020; Accepted: 27 March 2020; Published: 30 March 2020

Abstract: Unprecedented amounts of spatio-temporal data instigates an urgent need for patterns exploration in it. Clustering analysis is useful in extracting patterns from big data by grouping similar data elements into clusters. Compared with one-way clustering and co-clustering methods, tri-clustering methods are more capable of exploring complex patterns. However, the explored patterns or clusters could be different due to varying temporal resolutions of input data. This study presents a tri-clustering based method to explore the impacts of different temporal resolutions on spatio-temporal clusters identified in geo-referenced time series (GTS), one type of spatio-temporal data. Dutch daily temperature data at 28 stations over 20 years was used to illustrate this study. The temperature data at daily, monthly, and yearly resolutions were subjected to the Bregman cube average tri-clustering algorithm with I-divergence (BCAT_I) to detect spatio-temporal clusters, which were then compared in terms of patterns exhibited, compositions, and changed elements. Results confirm the temporal resolution impacts on the spatio-temporal clusters identified in the Dutch temperature data: most compositions of clusters are varying when changing the temporal resolutions of input data in the GTS. Nevertheless, there is almost no change of elements in certain clusters (12 stations in the northeast of the country; years 1996, 2010) at all temporal resolutions, suggesting them as the “true” clusters in the case study dataset.

Keywords: tri-clustering; spatio-temporal clusters; geo-referenced time series; Modifiable Temporal Unit Problem (MTUP); Dutch temperature

1. Introduction

The advancement in data acquisition techniques (e.g., remote sensing, GPS, mobile phone etc.) along with data sharing services has significantly promoted the accumulation of spatio-temporal data [1,2]. Such unprecedented amounts of data at multiple spatial and temporal resolutions instigates an urgent need for patterns exploration to obtain useful information in it [3,4]. One popular type of spatio-temporal data is geo-referenced time series (GTS), which are time series of one or more attributes' values observed at stationary locations and time intervals [5,6]. A common example of GTS is daily temperatures recorded at meteorological stations.

As an important data mining task, clustering is useful for exploring patterns in GTS by assigning similar data elements into the same cluster and dissimilar elements into different ones [7,8]. As a result, it provides both an overview of data at cluster levels and investigation of details on single clusters [9,10]. According to the involved dimensions in the clustering analysis, clustering methods for GTS are categorized as one-way clustering, co-clustering, and tri-clustering methods [11,12].

One-way clustering methods analyze 2D GTS that is typically organized into a data table with locations as rows and timestamps as columns [6]. The analysis can be performed either from the spatial or temporal aspect. For example, in the analysis from the spatial aspect, locations are grouped into location-clusters with similar data elements along all timestamps (Figure 1b). There have been extensive studies that employed one-way clustering methods to analyze GTS from the spatial or temporal aspect [13–16]. Varying from one-way clustering, co-clustering methods analyze 2D GTS in the data table from both spatial and temporal aspects simultaneously [6,7]. By concurrently grouping locations and timestamps into location-clusters and timestamp-clusters, the co-clustering results are their intersections, namely co-clusters, with similar data elements along both locations and timestamps (Figure 1c). Many studies have been conducted on co-clustering analysis of GTS for the exploration of concurrent spatio-temporal patterns [17–19].

Figure 1. One-way clustering (a,b), co-clustering (a,c) and tri-clustering (d,e) methods.

Tri-clustering methods analyze 3D GTS that is organized into a data cube with its three dimensions as locations, timestamps, and any third, e.g., another timestamps or locations [20,21]. Take 3D GTS with one spatial (locations) and two nested temporal (timestamps) dimensions for example. By concurrently grouping locations and two nested timestamps (e.g., years and months) into location-clusters—timestamp1-clusters and timestamp2-clusters—the tri-clustering results are their intersections, namely tri-clusters, with similar data elements along all three dimensions. Compared with one-way clustering and co-clustering methods, tri-clustering methods have advantages of exploring more patterns in the GTS with more detailed data and, thereby, extract more useful information [12].

Another important issue in the clustering analysis of the GTS is the temporal resolution. Changes of temporal resolutions in the input data could lead to different clustering results because of the temporal aggregation effects [22,23]. While the spatial aggregation effects on patterns explored in spatio-temporal data have been well studied as the Modifiable Area Unit Problem (MAUP), the issues related to the temporal dimensions become to attract attention in recent years but require more studies [24–26]. In 2011, Coltekin et al. [27] first proposed Modifiable Temporal Unit Problem (MTUP) and defined it in an analogy to MAUP with temporal resolution as one essential aspect. Afterward, a few studies have analyzed the effects of temporal resolutions on the explored patterns [22,23,28–30]. Among them, Cheng and Adepeju [22] examined the temporal resolution effects on the detected spatio-temporal clusters in point data. However, to our knowledge, no study has been conducted to explore the effects of temporal resolutions on the tri-clustering results in GTS.

Thus, this study uses a tri-clustering based method to explore the impacts of different temporal resolutions on the identified spatio-temporal clusters in GTS. Dutch daily temperatures at 28 stations from 1992 to 2011 are used as the case study dataset. Briefly, the tri-clustering method is used to explore the temperature dataset at daily, monthly, and yearly resolutions and then the detected spatio-temporal clusters at different temporal resolutions are compared to examine their impacts. The contributions of this study are: (1) this study designs an experiment based on a tri-clustering method to extract patterns in the GTS. In comparison with other clustering methods, tri-clustering methods have more capability of exploring complex patterns and revealing useful information in the data; (2) this study examines the effects of changing temporal resolutions on the spatio-temporal clusters identified in the GTS. With the case study dataset, this study compares detected clusters at different temporal resolutions and reveals the temporal resolution impacts on the tri-clustering results.

The structure of this paper is organized in the following manner: Section 2 introduces the Dutch daily temperature data, the tri-clustering method, and the experiment on tri-clustering analysis of the dataset at multiple temporal resolutions. Thereafter, the identified spatio-temporal clusters at different resolutions are briefly described and then compared in detail in Section 3. Finally, the results are discussed in Section 4 and conclusions are drawn in Section 5.

2. Materials and Methods

In this section, the study area and case study dataset are first introduced, then the procedure of the tri-clustering analysis is explained, and finally, the workflow of our experiment is described.

2.1. Study Area and Dataset

In this study, the Netherlands is chosen as the study area because of its location in Europe (Figure 2). Bordering with the North Sea in the north and west, weather in this area of the Netherlands is determined more by the moderate maritime climate. In contrast, the weather in the east and south of the Netherlands is more influenced by the continental climate of the neighbouring Germany and Belgium.

Figure 2. The Thiessen polygons map of Dutch weather stations.

As mentioned above, Dutch daily temperature data collected at 28 stations over 20 years (1992–2011) is used as the case study dataset, which is available from the Royal Netherlands Meteorological Institute (KNMI). To generate monthly and yearly data as input data of different temporal resolutions for the

tri-clustering analysis, the daily data was aggregated using the averaged value, which is the most widely used method in downscaling weather data [31]. With stations' coordinates also available from KNMI, the Thiessen polygons map was generated to indicate the region influenced by each station and used to visualize the tri-clustering results.

2.2. Tri-Clustering Analysis

Since the tri-clustering algorithm was first proposed in 2005, tri-clustering analysis has been employed for the exploration of patterns in many applications [11,20,21,32,33]. As the first tri-clustering algorithm, TRICLUSTER identifies tri-clusters by using multigraphs of ranges and constrained maximal cliques. Zhao and Zaki [20] applied this tri-clustering algorithm to gene expression data for the exploration of coherent patterns over time. Sim et al. [32] developed a tri-clustering algorithm named mining-correlated 3D subspace cluster (MIC), which optimizes correlation information to identify highly correlated tri-clusters. They used MIC to analyze financial stock data. Amar et al. [33] proposed the three-way module inference via Gibbs sampling (TWIGS), which uses normal-gamma assumption and Gibbs sampler to mine tri-clusters in 3D true-valued biological datasets. Recently, Wu et al. [21] developed the Bregman cube average tri-clustering algorithm with I-divergence (BCAT_I). It calculates the amounts of shared information among three variables using the mutual information in the field of information theory and then searches the optimal tri-clusters by optimizing the divergence of mutual information among the original data cube and the tri-clustered one. They applied BCAT_I to analyze time series of temperature data, which is a GTS dataset. Wu et al. [12] applied the same algorithm to identify tri-clusters in time series of air pollution data. As it has been proven effective for analyzing GTS, BCAT_I is also used in this study.

To illustrate the optimization process of BCAT_I, Dutch monthly temperature data is used as an example. The temperature data can be organized into a 3D data cube where rows are stations, columns are years, depths are 12 months, and elements are monthly temperatures. The data cube can be seen as a 3D co-occurrence matrix, O_{sym} , among three variables, which include a spatial variable taking values at 28 stations and two nested temporal variables taking values over 20 years and 12 months separately. The 3D data matrix, the number of station-clusters, year-clusters, and month-clusters are input parameters for BCAT_I while the output is the optimized tri-clusters. The pseudocode of BCAT_I in Figure 3 summarized the optimization process of the algorithm in three steps.

The first step is the random initialization, in which 28 stations, 20 years, and 12 months are randomly mapped to station-clusters, year-clusters, and month-clusters, respectively. The average value of each tri-cluster is calculated, which is used to replace elements within each tri-cluster and generate the tri-clustered matrix $\hat{O}_{sym}$. In the next step, the objective function of BCAT_I is built using the information divergence between the original and tri-clustered 3D matrix, denoted by $D_I(\cdot||\cdot)$. The function measures the similarity between O_{sym} and $\hat{O}_{sym}$. With more similar elements within each tri-cluster and different ones between the tri-clusters, the two matrices are more similar and the objective function has smaller values. The final step is to iteratively update the membership of station-clusters, year-clusters, and month-clusters to optimize the objective function. To this end, all stations are assigned to corresponding station-clusters under the situation that the minimum value of the objective function is obtained (Step 3.1). The same assignments are conducted for all years and months respectively (Step 3.2 and 3.3). After each iteration of assignments, the objective function monotonically decreases until the convergence is reached, i.e., the difference between the values of the objective function in two continuous iterations is smaller than a predefined threshold [34]. Then the optimized tri-clustering results are yielded.

Figure 3. The optimization process of BCAT_I exemplified using Dutch monthly temperature data.

2.3. Experiment: Tri-Clustering Dutch Temperature Data at Multiple Temporal Resolutions

To compare spatio-temporal clusters detected at different temporal resolutions, an experiment was designed in which the BCAT_I algorithm in Section 2.2 was used to analyze the Dutch temperature data at daily, monthly, and yearly resolutions. The workflow of the experiment is shown in Figure 4. First, to identify spatio-temporal clusters at the daily resolution, Dutch daily temperature data was organized into a 3D data cube with 28 stations, 20 years, and 365 days (29th February in leap years was removed) as its three dimensions (Figure 4a). Elements of the data cube are daily temperatures. Such a data cube can be seen as a 3D data matrix with a size of 28 (stations) $\times$ 20 (years) $\times$ 365 (days), which was subjected to BCAT_I to identify station-clusters, year-clusters, and day-clusters. Second, to detect spatio-temporal clusters at the monthly resolution, daily temperature data was averaged to generate monthly temperature data, which was organized into a 3D data cube with stations, years, and months as its three dimensions (Figure 4b). The elements of the data cube are monthly temperatures. Such a data cube can be regarded as a 3D data matrix with a size of 28 (stations) $\times$ 20 (years) $\times$ 12 (months), which was then analyzed using BCAT_I to identify station-clusters, year-clusters, and month-clusters. Thereafter, to identify spatio-temporal clusters at the yearly resolution, yearly temperature data was generated using the averaged value within each year for each station of the case study dataset. Even though there are only two dimensions, i.e., stations and years in the yearly temperature data, it can still be organized into a 3D data cube with stations, years, and 1 as its three dimensions (Figure 4c) and elements of the data cube are the yearly temperatures. BCAT_I was used to identify the station-clusters and year-clusters in the yearly temperature data. Finally, spatio-temporal clusters identified at daily, monthly, and yearly resolutions are compared in terms of patterns exhibited by clusters, compositions of clusters, and changed elements of clusters. Since the tri-clustering results at these three temporal resolutions all include station-clusters and year-clusters, they were compared to examine the impacts of different temporal resolutions on the tri-clustering results.

Figure 4. Workflow of tri-clustering analysis of Dutch temperature data at daily (a) monthly (b) and yearly (c) resolutions.

Tri-clustering analysis in our experiment requires several predefined parameters, e.g., the numbers of station-clusters, year-clusters, day-clusters, and month-clusters. Regarding parameters at daily resolutions, the numbers of station-clusters, year-clusters, and day-clusters were empirically set as four, four, and eight according to previous studies on the same dataset [21]. For parameters at the monthly resolution, the numbers of station-clusters and year-clusters were both set as four for comparisons among different temporal resolutions. The number of month-clusters was chosen as four with the expectation of 12 months falling into four seasons. For parameters at the yearly resolution, the numbers of station-cluster and year-clusters were set as four for comparison and the number of clusters in the third dimension was set as 1. Besides, the predefined threshold for reaching convergence and the number of iterations for BCAT_I analysis at all temporal resolutions were set to 10^{-6} and 2000 to guarantee the optimal tri-clustering results.

3. Results

In this section, spatio-temporal clusters identified by BCAT_I at daily, monthly, and yearly resolutions are first briefly described. Then clusters at different resolutions are compared in terms of patterns exhibited, compositions, and changed elements.

3.1. Spatio-Temporal Clusters at Daily Resolution

The daily temperature data cube with the size as $28 \times 20 \times 365$ was subjected to BCAT_I to identify spatio-temporal clusters. After the tri-clustering analysis, 28 stations, 20 years, and 365 days were grouped into four station-clusters, four year-clusters, and eight day-clusters. The small multiples in Figure 5a display the elements within each station-cluster and their spatial distribution in the Netherlands with colors. The deeper blue the color is, the lower the temperature of that station-cluster is. The timeline in Figure 5b shows the elements within each year-cluster and their temporal distribution from 1992 to 2011, with increasing temperatures from year-cluster1 to year-cluster4.

Figure 5. Spatio-temporal clusters ((a) station-clusters; (b) year-clusters) identified in Dutch temperature data at the daily resolution.

The spatial distribution of station-clusters in Figure 5a shows that four regions are partitioned from northeast to southwest of the Netherlands and there is increasing temperature patterns in this direction. Station-clusters in the northeast that border with Germany have lower temperatures while those in the southwest that neighbor with the North Sea have higher temperatures. Such results confirm the aforementioned fact that the weather in the northeast of the country is determined more by the continental climate while that in the southwest is more influenced by the moderate maritime climate. These results are also supported by previous studies on Dutch temperature data [21,35]. The spatial distribution also shows that most elements within each station-clusters are spatially adjacent. The temporal distribution of year-clusters in Figure 5b shows that there is general increasing temperature patterns from 1992 to 2011, especially in recent years after 1999. It is noticeable that half of all years (10/20) belong to the year-cluster4 with the highest temperature while only two years (1996 and 2010) belong to the year-cluster1 with the lowest temperature.

3.2. Spatio-Temporal Clusters at Monthly Resolution

The monthly temperature data cube with the size as $28 \times 20 \times 12$ was analyzed by BCAT_I to detect spatio-temporal clusters at the monthly resolution. After the analysis, 28 stations, 20 years, and 12 months were mapped to four station-clusters, four year-clusters, and four month-clusters. The small multiples in Figure 6a show the elements of station-clusters and their spatial distribution in the Netherlands with colors. The deeper blue color means a lower temperature of the station-cluster. The linear timeline in Figure 6b displays the elements of four year-clusters and their temporal distribution from 1992 to 2011. The temperature increases from year-cluster1 to year-cluster4.

Figure 6. Spatio-temporal clusters ((a) station-clusters; (b) year-clusters)) identified in Dutch temperature data at the monthly resolution.

The spatial distribution of station-clusters in Figure 6a also shows that the whole country is partitioned into four regions with increasing temperature patterns from northeast to southwest. However, station-clusters in the south that border with northern Belgium and the North Sea have high temperatures at the monthly resolution. Figure 6a also shows that stations of each station-clusters are spatially adjacent. The temporal distribution of year-clusters in Figure 6b shows that the temperature experienced a continuous variability from 1992 to 2011 at this temporal resolution. Nevertheless, more than half of all years (13/20) belong to the year-clusters3 and 4 with high temperatures, whereas two years (1996 and 2010) belong to the year-cluster1 with the lowest temperature.

3.3. Spatio-Temporal Clusters at Yearly Resolution

The yearly temperature data cube with the size as $28 \times 20 \times 1$ was subjected to BCAT_I to identify spatio-temporal clusters at the yearly resolution. After the tri-clustering analysis, 28 stations and 20 years were mapped to four station-clusters and four year-clusters. The small multiples in Figure 7a show the spatial distribution of elements within each station-cluster in the Netherlands with colors: the deeper blue color means lower temperature. The timeline in Figure 7b shows the temporal distribution of elements of each year-cluster from 1992 to 2011, with increasing temperatures from year-cluster1 to year-cluster4.

Figure 7. Spatio-temporal clusters ((a) station-clusters; (b) year-clusters) identified in Dutch temperature data at the yearly resolution.

Figure 7a shows that the spatial distribution of station-clusters at the yearly resolution is the same as that at the daily resolution. The linear timeline in Figure 7b shows general increasing temperature patterns of Dutch yearly temperatures from 1992 to 2011. Most years (17/20) belong to the year-clusters3 and 4 with high temperatures while three years (1993, 1996, and 2010) belong to the year-clusters1 and 2 with low temperatures.

3.4. Comparisons of Spatio-Temporal Clusters at Different Temporal Resolutions

The station-clusters and year-clusters detected in Dutch temperature data at daily, monthly and yearly resolutions were compared in terms of patterns exhibited by clusters, compositions, and the changed elements of the clusters. To facilitate the comparison, the spatial distributions of station-clusters at these three temporal resolutions are displayed side-by-side in the small multiples in Figure 8a. Each map in the small multiples shows the spatial coverage of station-clusters at one temporal resolution using deeper blue colors to indicate station-clusters with lower temperatures. The rectangle view in Figure 8b is used to display the temporal distribution of year-clusters over 20 years at these three resolutions. It provides straightforward comparisons of year-clusters at different resolutions using a darker red color to indicate year-clusters with higher temperatures [36]. Besides, the numbers of elements that changed in each of four station-clusters and four year-clusters from the daily to monthly resolution, from the daily to yearly resolution, and from the monthly to yearly resolution are listed in Table 1. The positive number indicates an increase of elements, whereas a negative number means a decrease.

The small multiples in Figure 8a show that even though station-clusters at all temporal resolutions exhibit the patterns of increasing temperatures from the northeast to southwest of the Netherlands, the composition of station-clusters at the monthly resolution is different from that at the daily and yearly resolutions. As shown in Table 1, several stations in station-clusters3 and 4, i.e., the southwest of the country, experienced changes of station-clusters from daily to monthly and monthly to yearly resolutions. Take station-cluster4 with the highest temperature for example. From the daily to monthly resolution, four more stations (344 Rotterdam, 210 Valkenburg, 240 Schiphol, 235 De Kooy) that border with the North Sea were divided into this station-cluster. From the monthly to yearly resolution, these four stations lost their membership of station-cluster4. Such a difference is because the stations in the southwest with similar daily and yearly temperatures exhibit different variability among monthly temperatures. In contrast, there are almost no changes of elements in station-cluster1 and 2 at these three

temporal resolutions, which indicates that most stations in the northeast have the same membership of station-clusters at all resolutions.

Figure 8. Comparisons of station-clusters (a) and year-clusters (b) in Dutch temperature data at daily, monthly and yearly resolutions.

Table 1. The number of elements changed in station-clusters and year-clusters at different temporal resolutions.

Temporal Resolutions		num. of Changed Elements							
		Station-Cluster1	Station-Cluster2	Station-Cluster3	Station-Cluster4	Year-Cluster1	Year-Cluster2	Year-Cluster3	Year-Cluster4
daily	→ monthly	0	−1	−3	4	−7	4	3	0
daily	→ yearly	0	0	0	0	−4	5	−1	0
monthly	→ yearly	0	1	3	−4	3	1	−4	0

The rectangle view in Figure 8b shows that although year-clusters at the daily and yearly resolution exhibit the patterns of increasing temperatures over the study period, compositions of year-clusters are different at all temporal resolutions. As shown in Table 1, year-clusters1 and 2 have the largest number of changed elements at different temporal resolutions, especially the former. For year-cluster1, 10 recent years (1999, 2000, 2001, 2002, 2004, 2005, 2006, 2007, 2008, 2011) belong to this year-cluster at the daily resolution while three years (1995, 2001, 2006) and six years (1999, 2000, 2002, 2006, 2007, 2011) belong to it at the monthly and yearly resolutions, respectively. It is possible because the year-cluster at the daily resolution only indicates the similarity of daily temperatures for all years and so does the year-clusters at the monthly and yearly resolutions [6]. Years with similar yearly temperatures might have different variability among daily and monthly temperatures.

4. Discussion

As displayed in Figure 8 and Table 1, even though elements changed in several station-clusters and year-clusters at different temporal resolutions, there is no or little changes in stations of station-clusters1 and 2 and years of year-cluster1 at all temporal resolutions. These stations include five stations in station-cluster1 (270 Leeuwarden, 280 Eelde, 286 Nieuw Beerta, 279 Hoogeveen, 278 Heino) and seven stations in station-cluster2 (267 Stavoren, 269 Lelystad, 273 Markness, 275 Deelen, 277 Lauwersoog, 283 Hupsel, 290 Twenthe) that are located in the northeast of the country. The years of year-cluster1 with the lowest temperature are 1996 and 2010. These stable clusters suggest that different temporal resolutions of input data in the tri-clustering analysis have no or little impact on them, which can

be seen as “true” clusters in the case study dataset [22]. On the contrary, the station-clusters and year-clusters with large numbers of changed elements at all temporal resolutions, e.g., year-clusters1 and 2, suggest that there are strong effects of different temporal resolutions on them. These unstable clusters need further analysis at other temporal resolutions.

Figure 8 and Table 1 also show that there is no change of compositions of all station-clusters from the daily to yearly resolution, which might suggest the two temporal resolutions as the suitable temporal resolution for the case study dataset. Even though stations of all station-clusters changed at the monthly resolution, they stayed the same at the daily and yearly resolutions, which might imply that these two resolutions are suitable ones for analyzing station-clusters in Dutch temperature data. Under this situation, the analysis related to station-clusters in the data could use the coarse resolution, i.e., the yearly resolution, which would significantly reduce the computational time, given that the tri-clustering method is quite time-consuming [12].

5. Conclusions

This study presented a tri-clustering based method to explore the impacts of changing temporal resolutions of input data on the identified spatio-temporal clusters in GTS. To illustrate the study, Dutch daily temperature data collected at 28 stations from 1992 to 2011 was used. More specifically, the Bregman cube average tri-clustering algorithm with I-divergence (BCAT_I) was employed to identify spatio-temporal clusters in the data at daily, monthly and yearly resolutions. Then, spatio-temporal clusters detected at these three resolutions were compared in terms of patterns exhibited, compositions, and changed elements to examine the temporal resolution effects.

Results show that temporal resolutions indeed have impacts on the spatio-temporal clusters identified in a GTS. Compositions of station-clusters at daily and yearly resolutions are different from that at the monthly resolution and those of year-clusters are varying at all resolutions. However, there is almost no change at the stations in station-clusters1 and 2 (12 stations in the northeast of the country) and in the years in year-cluster4 (1996, 2010) at all three resolutions, which suggests them as the “true” clusters in the case study dataset. Besides, compositions of station-clusters are the same at daily and yearly resolutions, which might imply them as suitable temporal resolutions for the spatial analysis of the dataset.

In summary, the tri-clustering based method proposed in this study effectively explores the temporal resolution impacts on the spatio-temporal clusters identified in GTS. However, one limitation of the method is that the tri-clustering algorithm (BCAT_I) used in this study requires heavy computational effort because of its high computational complexity. Therefore, future work will focus on the optimization of this tri-clustering algorithm to reduce the running time of the experiment. Besides, other tri-clustering algorithms could also be used in the future to save the running time.

Author Contributions: Conceptualization, Xiaojing Wu and Donghai Zheng; Methodology, Xiaojing Wu; Writing-Original Draft Preparation, Xiaojing Wu; Writing-Review & Editing, Donghai Zheng. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Postdoctoral Science Foundation Grant, grant number 2018M641246 and the National Natural Science Foundation of China, grant number 41901317.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Z.; Yang, C.; Liu, K.; Hu, F.; Jin, B. Automatic scaling hadoop in the cloud for efficient process of big geospatial data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 173. [\[CrossRef\]](#)
2. Sagl, G.; Loidl, M.; Beinat, E. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. *ISPRS Int. J. Geo-Inf.* **2012**, *1*, 256–271. [\[CrossRef\]](#)
3. Shekhar, S.; Jiang, Z.; Ali, R.Y.; Eftelioglu, E.; Tang, X.; Gunturi, V.; Zhou, X. Spatiotemporal data mining: A computational perspective. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2306–2338. [\[CrossRef\]](#)

4. Miller, H.J.; Han, J. Geographic Data Mining and Knowledge Discovery: An Overview. In *Geographic Data Mining and Knowledge Discovery*, 2nd ed.; Miller, H.J., Han, J., Eds.; Taylor & Francis Group: London, UK, 2009; pp. 1–26.
5. Kisilevich, S.; Mansmann, F.; Nanni, M.; Rinzivillo, S. Spatio-Temporal Clustering. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: New York, NY, USA, 2010; pp. 855–874.
6. Wu, X.J.; Zurita-Milla, R.; Kraak, M.J. Co-clustering geo-referenced time series: Exploring spatio-temporal patterns in Dutch temperature data. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 624–642. [[CrossRef](#)]
7. Han, J.; Kamber, M.; Pei, J. *Data Mining Concepts and Techniques*; Morgan Kaufman MIT Press: Burlington, MA, USA, 2012.
8. Mueller, E.; Sandoval, J.; Mudigonda, S.; Elliott, M. A cluster-based machine learning ensemble approach for geospatial data: Estimation of health insurance status in Missouri. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 13. [[CrossRef](#)]
9. Andrienko, G.; Andrienko, N.; Rinzivillo, S.; Nanni, M.; Pedreschi, D.; Giannotti, F. Interactive Visual Clustering of Large Collections of Trajectories. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST), Atlantic City, NJ, USA, 12–13 October 2009.
10. Wang, H.; Du, Y.; Sun, Y.; Liang, F.; Yi, J.; Wang, N. Clustering Complex Trajectories Based on Topologic Similarity and Spatial Proximity: A Case Study of the Mesoscale Ocean Eddies in the South China Sea. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 574. [[CrossRef](#)]
11. Henriques, R.; Madeira, S.C. Triclustering algorithms for three-dimensional data analysis: A comprehensive survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 95. [[CrossRef](#)]
12. Wu, X.; Cheng, C.; Zurita-Milla, R.; Song, C. An overview of clustering methods for geo-referenced time series: From one-way clustering to co- and tri-clustering. *Int. J. Geogr. Inf. Sci.* **2020**, 1–27. [[CrossRef](#)]
13. Mills, R.T.; Hoffman, F.M.; Kumar, J.; Hargrove, W.W. Cluster analysis-based approaches for geospatiotemporal data mining of massive data sets for identification of forest threats. *Proc. Comput. Sci.* **2011**, *4*, 1612–1621. [[CrossRef](#)]
14. Andrienko, G.; Andrienko, N.; Bremm, S.; Schreck, T.; Von Landesberger, T.; Bak, P.; Keim, D. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Comput. Gr. Forum* **2010**, *29*, 913–922. [[CrossRef](#)]
15. Hagenauer, J.; Helbich, M. Hierarchical self-organizing maps for clustering spatiotemporal data. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2026–2042. [[CrossRef](#)]
16. White, M.A.; Hoffman, F.; Hargrove, W.W.; Nemani, R.R. A global framework for monitoring phenological responses to climate change. *Geophys. Res. Lett.* **2005**, *32*, L04705. [[CrossRef](#)]
17. Wu, X.; Zurita-Milla, R.; Kraak, M.-J. A novel analysis of spring phenological patterns over Europe based on co-clustering. *J. Geophys. Res. Biogeosci.* **2016**, *121*, 1434–1448. [[CrossRef](#)]
18. Andreo, V.; Izquierdo-Verdiguier, E.; Zurita-Milla, R.; Rosà, R.; Rizzoli, A.; Papa, A. Identifying Favorable Spatio-Temporal Conditions for West Nile Virus Outbreaks by Co-Clustering of Modis LST Indices Time Series. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018.
19. Ullah, S.; Daud, H.; Dass, S.C.; Khan, H.N.; Khalil, A. Detecting space-time disease clusters with arbitrary shapes and sizes using a co-clustering approach. *Geospatial Health* **2017**, *12*, 567.
20. Zhao, L.; Zaki, M.J. Triclust: An Effective Algorithm for Mining Coherent Clusters in 3D Microarray Data. In Proceedings of the 2005 Acm Sigmod International Conference on Management of Data, Baltimore, MD, USA, 14–16 June 2005.
21. Wu, X.; Zurita-Milla, R.; Izquierdo Verdiguier, E.; Kraak, M.-J. Triclustering Georeferenced Time Series for Analyzing Patterns of Intra-Annual Variability in Temperature. *Ann. Am. Assoc. Geogr.* **2018**, *108*, 71–87. [[CrossRef](#)]
22. Cheng, T.; Adepeju, M. Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PLoS ONE* **2014**, *9*. [[CrossRef](#)]
23. Liu, X.; Huang, Q.; Li, Z.; Wu, M. The Impact of MTUP to Explore Online Trajectories for Human Mobility Studies. In Proceedings of the 1st Acm Sigspatial Workshop on Prediction of Human Mobility, Redondo Beach, CA, USA, 7–10 November 2017.
24. Openshaw, S. *The Modifiable Unit Problem*. Geo Books; Headley Brothers Ltd. Kent: Norwick, UK, 1983.
25. Jiang, B.; Brandt, S.A. A fractal perspective on scale in geography. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 95. [[CrossRef](#)]

26. Josselin, D.; Louvet, R. Impact of the Scale on Several Metrics Used in Geographical Object-Based Image Analysis: Does GEOBIA Mitigate the Modifiable Areal Unit Problem (MAUP)? *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 156. [[CrossRef](#)]
27. Coltekin, A.; Sabbata, S.C.; Willi, D.; Vontobel, I.; Pfister, S.; Kuhn, M.; Lacayo, M. Modifiable Temporal Unit Problem. In Proceedings of the ISPRS/ICA workshop Persistent problems in geographic visualization (ICC2011), Paris, France, 2–7 July 2011.
28. de Jong, R.; de Bruin, S. Linear trends in seasonal vegetation time series and the modifiable temporal unit problem. *Biogeosciences* **2012**, *9*, 71–77. [[CrossRef](#)]
29. Wu, X.; Zurita-Milla, R.; Kraak, M.-J. Visual discovery of synchronization in weather data at multiple temporal resolutions. *Cartogr. J.* **2013**, *50*, 247–256. [[CrossRef](#)]
30. Zhao, Z.; Shaw, S.-L.; Yin, L.; Fang, Z.; Yang, X.; Zhang, F.; Wu, S. The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1471–1495. [[CrossRef](#)]
31. Estrella, N.; Sparks, T.; Menzel, A. Trends and temperature response in the phenology of crops in Germany. *Glob. Chang. Biol.* **2007**, *13*, 1737–1747. [[CrossRef](#)]
32. Sim, K.; Aung, Z.; Gopalkrishnan, V. Discovering Correlated Subspace Clusters In 3D Continuous-Valued Data. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010.
33. Amar, D.; Yekutieli, D.; Maron-Katz, A.; Hendler, T.; Shamir, R. A hierarchical Bayesian model for flexible module discovery in three-way time-series data. *Bioinformatics* **2015**, *31*, i17–i26. [[CrossRef](#)] [[PubMed](#)]
34. Banerjee, A.; Dhillon, I.; Ghosh, J.; Merugu, S.; Modha, D.S. A generalized maximum entropy approach to bregman co-clustering and matrix approximation. *J. Mach. Learn. Res.* **2007**, *8*, 1919–1986.
35. Lenderink, G.; Mok, H.; Lee, T.; Van Oldenborgh, G. Scaling and trends of hourly precipitation extremes in two different climate zones—Hong Kong and the Netherlands. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3033–3041. [[CrossRef](#)]
36. Nocke, T.; Schumann, H.; Böhm, U. Methods for the visualization of clustered climate data. *Comput. Stat.* **2004**, *19*, 75–94. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).