

Understanding Climate Change Patterns with Multivariate Geovisualization

Hai Jin and Diansheng Guo

Department of Geography
University of South Carolina
Columbia, SC 29208

E-mail: jinhai.cn@gmail.com; guod@sc.edu

Abstract— Climate change has been a challenging and urgent research problem for many related research fields. Climate change trends and patterns are complex, which may involve many factors and vary across space and time. However, most existing visualization and mapping approaches for climate data analysis are limited to one variable or one perspective at a time. For example, it is common to map the surface temperature anomaly at different locations or plot trends of time series. Although such approaches are useful in presenting information and knowledge, they have limited capability to support discovery and understanding of unknown complex patterns from data that span across multiple dimensions. This paper introduces the application of a multivariate geovisualization approach to explore and understand complex climate change patterns across multiple perspectives, including the geographic space, time, and multiple variables.

Keywords—geovisualization, climate change, multivariate mapping

I. INTRODUCTION

Climate changes, particularly regional temperature increases, are affecting physical, biological, and human systems on all continents [17]. Understanding climate change patterns and their trends across time (such as decadal, annual, and monthly changes) and variations across geographic space is of great significance. There are enormous on-going research efforts on the collection of climate data, the analysis of climate changes, and the modeling of climate processes. This paper focuses on the exploratory analysis of climate data and the discovery of change patterns.

Most existing climate mapping and visualization approaches focus on one variable or one type of patterns at a time. These approaches are useful to *communicate* discovered trends and knowledge but they have limited capability to support the process of *data exploration* and *knowledge discovery*. Moreover, since they focus on one perspective at a time, they are not able to help understand complex patterns that span across multiple perspectives.

For example, a map of surface temperature anomaly across the world [18] or a temporal plot of seasonal trends of anomalies [19] can only present patterns from a specific perspective, with the map showing spatial patterns and the plot showing temporal patterns. However, it is natural and important for an analyst to wonder: Do these temporal patterns in the plot vary across space? Do the spatial patterns

in the map change from year to year? The answer to these questions will lead to more complex patterns and subsequently improve our understanding of climate changes, which inherently are complex and involve multiple dimensions and factors.

This paper presents the application of a multivariate geovisualization approach [4, 5] to analyze climate data across space and time and help the analyst explore answers to the above questions. The approach leverages both visual and computational approaches to synthesize data, accentuate patterns, integrate multiple views, and support user interactions.

II. BACKGROUND

A. Climate Change Analysis

Climate change is a subject of great interest and importance. Many research efforts have been devoted to the study of climate change patterns over time and their variations across space [1, 15, 16]. For example, Deser and Blackmon [2] described the low-frequency variability of the surface climate over the North Atlantic during winter from 1900 to 1989. Kushnir [12] studied the interdecadal patterns and variations of the North Atlantic sea surface temperature and its associated ocean-atmosphere relationship. Hurrell [7] examined the relationship between large decadal variations over the North Atlantic and the North Atlantic Oscillation. At the global level, Jones et al. [9] analyzed the surface air temperature and found that, for the last century, two 20-year periods (i.e., 1925-1944 and 1978-1997) saw the greatest increase in the surface temperature. Hansen et al. [6] analyzed global surface temperature changes and anomalies in the tropical Pacific Ocean.

Above are a few samples out of a large body of research on climate change analysis. Although each study is unique, several common elements can be found in these analyses. *First*, the changes are multi-dimensional, across time (seasonal, annual, decadal, etc.), space (stations, regions, continents, oceans, and global), and climate variables (surface temperature, precipitation, etc., although surface temperature has been the primary focus). *Second*, the multidimensional changes are connected and interact with each other. For example, temporal trends vary from place to place and regional patterns change over time. *Third*, to analyze and present climate change patterns, mapping and visual graphics are indispensable and commonly used.

B. Mapping and Visualization of Climate Data

Most climate analysis maps or visual graphics are univariate, i.e., showing only one variable at a time, such as a map of the annual temperature anomalies [18] (which obviously ignores the seasonal variations) or a trend plot of the seasonal patterns of temperature anomalies [19] (which ignores the geographic variation of these patterns). The primary limitation of these univariate mapping and visual approaches is that they cannot present or discover more complex patterns that span across multiple dimensions.

Multivariate mapping has long been an interesting research topic. Three primary approaches have been used: (1) multivariate representation that depicts each dimension (variable) independently through some attribute of the display and then integrates all variable depictions into one map using composite glyphs, attributes of color, or other methods [3, 21, 22]; (2) dimension reduction that projects multivariate information to two (or three) dimensions and then map the result (e.g., [5]), and (3) multiple linked views that show one (or more) variables per view [13, 14]. If we treat time series as a special case of multivariate data, then it is possible to map the spatial variation of temporal patterns.

There is a great potential for multivariate mapping to bring in new capabilities for climate data analysis as it can present more than one variables in a single map. Moreover, many of latest visualization and mapping techniques also support interactive user exploration and multiple linked views. These capabilities can significantly facilitate the discovery and understanding of more complex patterns.

This paper adopts a multivariate geovisualization

approach [4, 5] to analyze climate data across space and time. Section III will briefly introduce the methodology and Section IV presents the climate change analysis, visualization, and pattern interpretation.

III. MULTIVARIATE GEOVISUALIZATION

The multivariate clustering and geovisualization method [4, 5] adopted in this research in an integrated approach that couples a self-organizing map (SOM), a multidimensional visualization component (i.e., a parallel coordinate plot (PCP)), and a multivariate mapping component.

A SOM is a multivariate clustering method that seeks clusters in multivariate data and orders the clusters in a two-dimensional layout so that nearby clusters are similar to each other [11]. It projects multivariate data to a 2D space. The SOM component used in this paper is shown in Figure 1 (top left). Each cluster (node) in the SOM is associated with a multivariate vector that represents the centroid of the cluster in the multivariate space. The size of the circle on top of the node is linearly scaled according to the number of data items contained in that cluster.

A two-dimensional color scheme is used to assign a unique color to each SOM node so that nearby and thus similar clusters have similar colors. Each data item inside the same cluster will have the same color of their containing cluster. Thus, colors can be used to locate and identify the same cluster or similar data items. These colors are passed on to every visualization components. For example, in the four views in Figure 1, the red color represents the same group of data items. The layer of hexagons behind the nodes

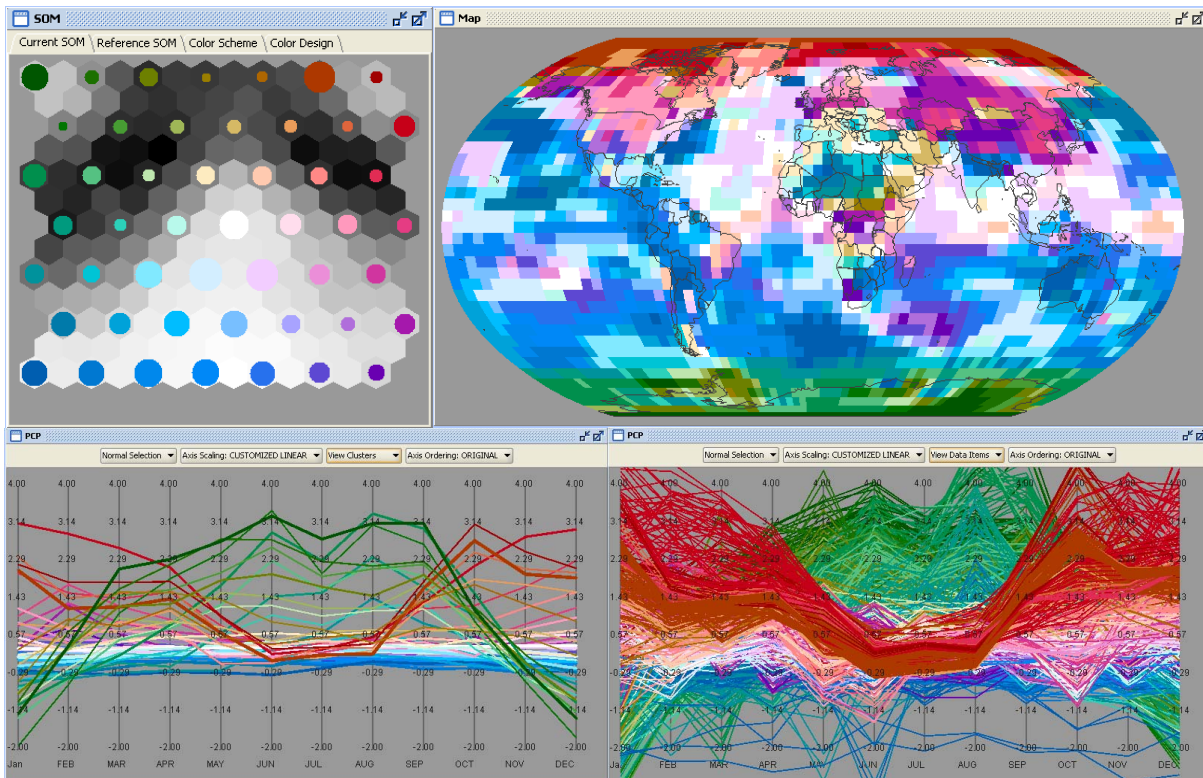


Figure 1. Seasonal and spatial patterns of surface temperature anomaly (1998-2007).

is shaded to show the multivariate dissimilarity between neighboring nodes. Clusters in a darker area are less similar to each other than those in the brighter area. For example, the green clusters (top left corner in the SOM view) and the red clusters (top right corner) are two distinct groups of clusters while the blue and purple clusters are more similar to each other. This can be confirmed by the patterns shown in the PCP (Figure 1, bottom left).

A parallel coordinate plot (PCP) is a multidimensional visualization technique, which draws axes as vertical parallel lines, equally spaced [8]. A point in a d -dimensional space is represented as a polyline with vertices on the parallel axes. In Figure 1, the PCP shows 12 “variables”, which are surface temperature anomaly for each month.

The PCP can visualize the data at two different levels: the cluster level and the data item level. A PCP at the cluster level is shown in Figure 1 (bottom left), where each string represents a cluster with its mean vector, and has the same color as the cluster does in the SOM. The thickness of each string is proportional to the cluster size. At the data item level (Figure 1, bottom right), each string in the PCP represents an individual data item, and has the same color as that of its cluster. By comparing the PCP at the two different levels, it is obvious that the aggregation of data items into clusters effectively helps discover major patterns in the data. The colors significantly improve the visual understanding of data and patterns.

With the colors derived by the SOM, a multivariate mapping component is readily constructed (Figure 1, top right). The map shows the spatial variation of the seasonal patterns (which are encoded with colors). To interpret the meaning of a color, one needs to look up the PCP, which serves as the legend for the multivariate map. The following section will explain in the detail of data and the patterns shown in Figures 1, 2, and 3.

IV. CLIMATE CHANGE ANALYSIS

A. Data

The climate data used in this research is a spatio-temporal data set of monthly mean surface air temperature for 60 years (Jan. 1948—Dec. 2007), which is part of the NCEP/NCAR reanalysis data archive [10]. It has a global coverage ranging from 90N to 90S and from 0E to 357.5E based on a matrix of 2.5° latitude by 2.5° longitude grids.

The analysis focuses on the surface air temperature. An anomaly value is calculated for each month and each grid cell in order to analyze temperature changes. First, a 40-year average temperature (1948-1987) is calculated for each grid cell and each month, and this will be used as the “normal” value. Then, the 10-year average temperature for each 10-year period (1948-1957, 1958-1967, 1968-1977, 1978-1987, 1988-1997, 1998-2007) is calculated for each grid cell and each month. Finally, the anomaly value is the difference between the 40-year average and the 10-year average for each period. In other words, each combination of a grid cell, a 10-year period, and a month has an anomaly value.

A positive anomaly value represents a temperature increase for a specific month and grid cell during a specific

10-year period. To smooth the data and remove some noise, the grid cells are aggregated at a $5^\circ \times 5^\circ$ resolution by merging each 2×2 block of the original grid cells. Thus, the original data is transformed to a $2664 \times 12 \times 6$ cube, which has 2664 spatial objects (grid cells) and 12 variables (monthly anomaly) for 6 decades.

B. Seasonal and Spatial Patterns

Let us start with a simple case, focusing only on one decade, 1998—2007. We consider the 12 monthly anomaly values for this decade as a “multivariate” vector. Thus each grid cell has vector and there are 2664 such vectors. These vectors are input to the self-organizing map (SOM), which groups them into clusters based on the Euclidean distance (dissimilarity) among the multivariate vectors. As briefly introduced in Section III, the SOM component (1) orders the discovered clusters in a 2D layout so that nearby clusters are similar to each other; and (2) assigns a color to each cluster according to a 2D color scheme so that similar clusters have similar colors. Figure 1 (top-left) shows the SOM clusters of the decadal monthly temperature anomalies for 1998-2007.

The SOM derived 49 clusters. The PCP and the map can help understand these clusters. The shaded hexagons suggest that there are three high-level groups of clusters: the green/light-green nodes at the top left corner, the red/reddish nodes at the top right corner, and the rest (blue/purple/white) nodes at the bottom and the center.

From the map and the PCP, we understand that the red/reddish clusters, primarily located in the Arctic area, have high positive values for winter months (November, December, January, and February) and around zero (i.e., no change) for summer months (June, July, and August). This means that, for the past decade, the Arctic area were much warmer than before in winter but relatively stable during summer. In contrast, the green/light-green clusters, mainly located in the Antarctic area, have high positive values for its winter months (April – September) and negative values for its summer months (January, February, November, and December). This means that, for the past decade, the Antarctic area were much warmer in winter but actually got cooler during summer.

The purple clusters, as seen in the PCP, also represent a warming trend (mostly for Dec., Jan., Feb., and March), but it is much less dramatic than those in the Polar area. These purple clusters are mainly over the continental regions in the northern Hemisphere.

C. Seasonal, Spatial, and Temporal Patterns

Now let us include other 5 decades as well. There are $2664 \times 6 = 15984$ monthly temperature anomaly vectors, which are grouped by the SOM into 49 clusters. (Note: the number of clusters can be configured in the SOM component.) Figure 2 shows SOM clusters, the PCP, and map matrix (one map for each decade and thus six maps). A cluster now may contain vectors from different decades.

It is worth emphasizing that, the colors are not associated with the actual meaning of clusters. For example, the red color is not always assigned to the cluster of a warming trend. Moreover, if the input data changed, such as adding

more data as in this case, the color for the same data item or cluster may change. However, from the analyst's point of view, it is desirable to keep the meaning of colors stable. The SOM component has several options to stabilize the color

meaning such as manually changing the color assignment or automatically matching clusters and colors in different runs.

The meaning of each color in Figure 2 is very similar to (although not exactly the same as) those in Figure 1. For

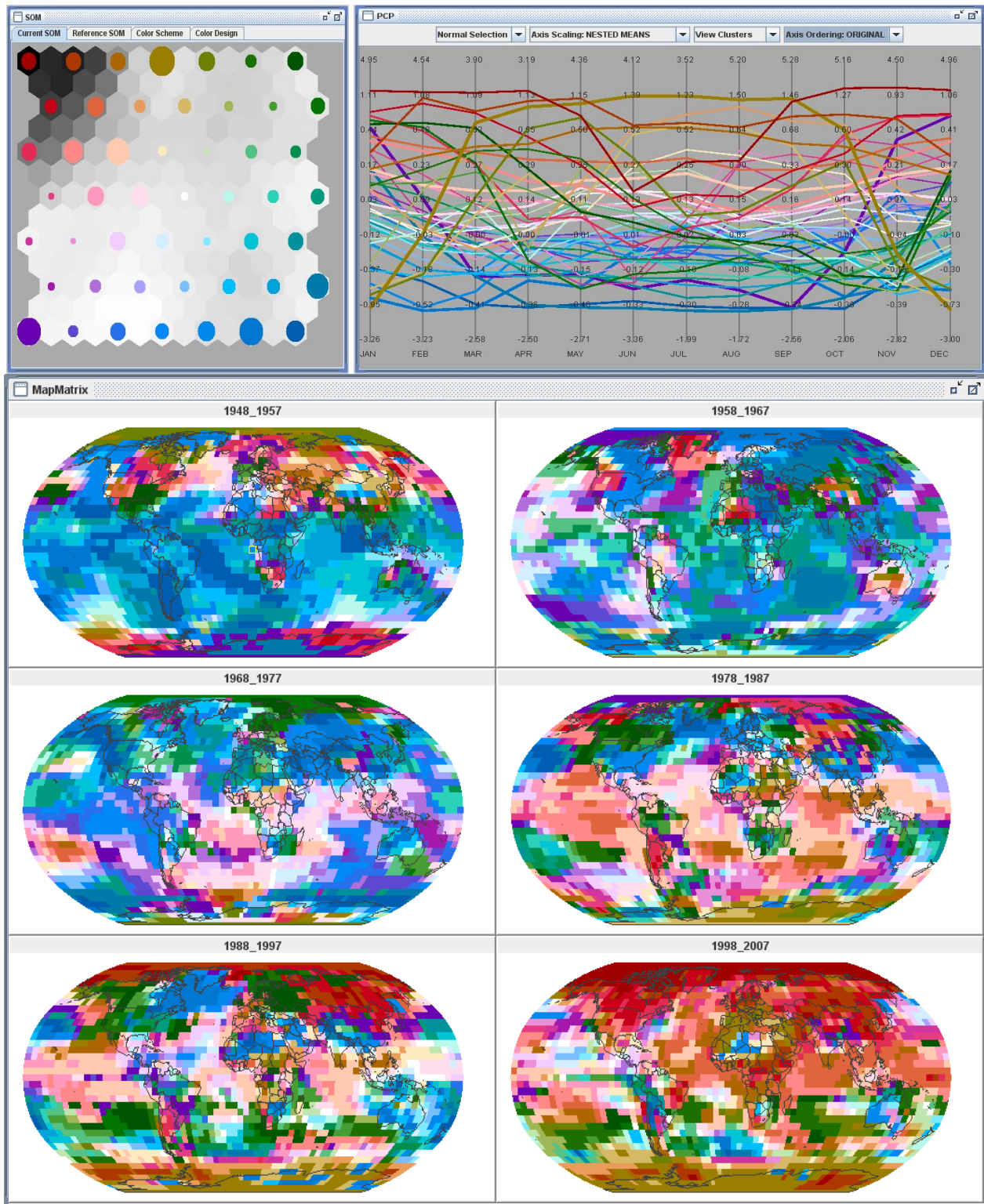


Figure 2. An overview of multivariate (seasonal), spatial and temporal patterns of surface temperature change (anomaly) for 6 decades.

example, the red clusters still represent very strong warming trends, especially for Nov., Dec., Jan., and Feb. These clusters again mainly locate in the Arctic area, not only for 1998-2007 but also for 1988-1997. The blue clusters represent a stable or even cooler trends (relative to the 40-year average for 1948-1987), which were mainly present for the first three decades but diminish quickly for the recent decades. Other clusters, such as green, purple, and those between red and green, represent unique seasonal change patterns, with different warming months and magnitudes.

The integrated views, including the SOM, the PCP, and the map matrix, allow us understand both the spatial variations and temporal trends of the seasonal temperature change patterns. Moreover, these views support interactive user selections and highlighting that can significantly improve our understanding of each cluster and their spatial-temporal variations.

D. User Interaction and Exploration

User interactions can help better understand patterns in two aspects. First, from a cognitive perspective, the human mind may struggle to precisely perceive and differentiate too many colors and match them across different views. With user selection and highlighting, users can understand a specific pattern and its connection to other views in greater accuracy. Figure 3 shows a selection of eight clusters of warming trends, which are highlighted in all maps.

Second, patterns at the cluster level may lose important details. The user can select one or more clusters and view the data items in each cluster to understand the interval variance. For example, Figure 1 shows one PCP at the cluster level and one PCP at the data item level. Since the map always shows data at the item level, one can also select an area in a map and examine its corresponding patterns in other views.

The integrated system supports a variety of user interactions to explore multivariate patterns from different perspectives and at different detail levels. A selection made in one component will be highlighted in all other components simultaneously. A selection can be progressively refined via adding or subtracting new selections. The user can select at either the cluster level or the data item level. If one selects some data items at the data item level the SOM will not only highlight the clusters that contain the selected data items but also change the circles to wedges to show the percentages of the selected items in the corresponding clusters.

V. DISCUSSION AND CONCLUSION

This paper presents a preliminary application of an integrated approach to multivariate clustering and geovisualization to explore climate change patterns. The approach is built on the synergy of both computational and visual methods, to synthesize different perspective information and present an overview of complex patterns across multiple dimensions. Through the static linking (colors) and user interactions, complex relationships can be easily perceived and understood. The approach can effectively help explore and understand complex patterns in multivariate and spatiotemporal climate data sets.

Data preprocessing, transformation, and integration requires further attention. Given the complexity of potential patterns, the exploratory nature of such analysis, it is desirable to have an integrated data processing component that can allow the analyst interactive process, transform, and integrate various data sources on the fly and examine the corresponding changes in patterns. For example, the analyst may want to calculate the anomaly in a different way such as separating natural variability from forced changes [20].

The analysis and visualization of climate change patterns presented in the paper focus on fixed spatial (grid cells) and temporal resolutions (monthly and decadal aggregations). It is necessary to extend the approach to allow the examination and comparison of patterns across spatial scales (e.g., grid cells, regions, continents, and hemispheres) and temporal scales (e.g., month, year, decade, and multiple decades).

Exploratory analysis, such as the mapping and visualization presented in this paper, is usually the beginning of a multi-step and iterative knowledge discovery process. Once patterns and relationships are discovered, further steps are needed to validate the findings and understand its implications. The presented approach is implemented with a flexible component-based structure. Thus it is possible to add more or different components to carry out further analysis.

The software for the presented approach is available at <http://www.SpatialDataMining.org>. Color versions of the figures in this paper are also available at the website.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. 0748813.

REFERENCES

- [1] T.L. Delworth, and M.E. Mann, "Observed and Simulated Multidecadal Variability in the Northern Hemisphere," *Climate Dynamics*, vol. 16, no. 9, pp. 661-676, Sep, 2000.
- [2] C. Deser, and M.L. Blackmon, "Surface Climate Variations over the North-Atlantic Ocean During Winter - 1900-1989," *Journal of Climate*, vol. 6, no. 9, pp. 1743-1753, Sep, 1993.
- [3] M. Gahegan, "Scatterplots and Scenes: Visualization Techniques for Exploratory Spatial Analysis," *Computers, Environment and Urban Systems*, vol. 22, no. 1, pp. 43-56, 1998.
- [4] D. Guo, J. Chen, A.M. MacEachren, and K. Liao, "A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP)," *Ieee Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1461-1474, 2006.
- [5] D. Guo, M. Gahegan, A.M. MacEachren, and B. Zhou, "Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach," *Cartography and Geographic Information Science*, vol. 32, no. 2, pp. 113-132, 2005.
- [6] J. Hansen, M. Sato, R. Ruedy, K. Lo, D. Lea, and M. Medina-Elizade, "Global Temperature Change," *Proceedings of the National Academy of Sciences*, vol. 103, no. 39, pp. 14288, 2006.
- [7] J.W. Hurrell, "Decadal Trends in the North-Atlantic Oscillation - Regional Temperature and Precipitation," *Science*, vol. 269, no. 5224, pp. 676-679, Aug, 1995.
- [8] A. Inselberg, "The Plane with Parallel Coordinates," *The Visual Computer*, vol. 1, pp. 69-97, 1985.
- [9] P.D. Jones, M. New, D.E. Parker, S. Martin, and I.G. Rigor, "Surface Air Temperature and Its Changes over the Past 150 Years," *Reviews of Geophysics*, vol. 37, no. 2, pp. 173-199, May, 1999.

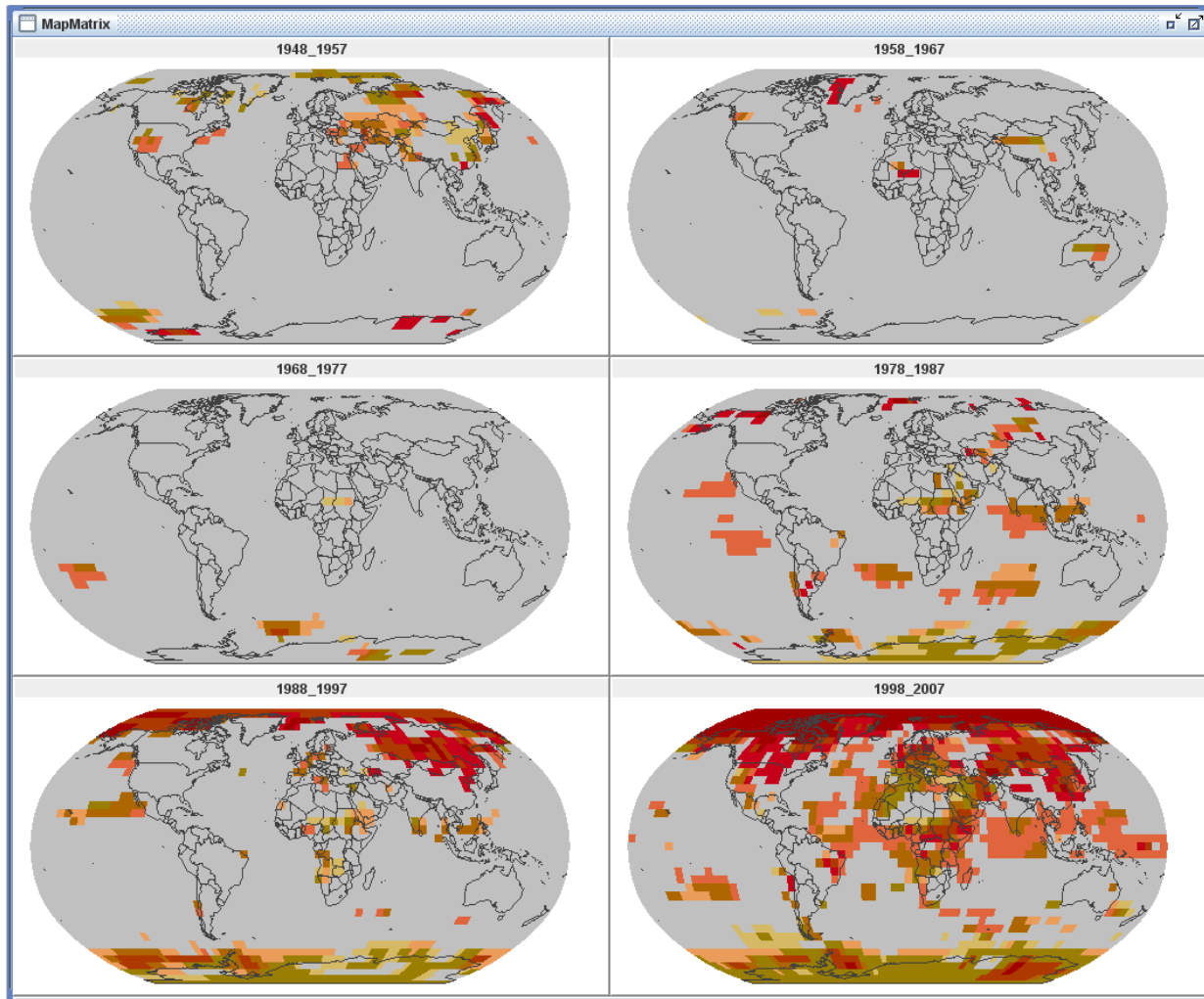


Figure 3. The analyst selected several clusters in the SOM that represent warming trends (see the PCP in Figure 2). Their spatial and temporal distributions are clearly shown in the map matrix.

- [10] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, and J. Woollen, "The Ncep/Ncar 40-Year Reanalysis Project," *Bulletin of the American Meteorological Society*, vol. 77, no. 3, pp. 437-471, 1996.
- [11] T. Knhonen, *Self-Organizing Maps, 2nd Ed.*, Berlin: Springer, 2001.
- [12] Y. Kushnir, "Interdecadal Variations in North-Atlantic Sea-Surface Temperature and Associated Atmospheric Conditions," *Journal of Climate*, vol. 7, no. 1, pp. 141-157, Jan, 1994.
- [13] A.M. MacEachren, M. Wachowicz, R. Edsall, D. Haug, and R. Masters, "Constructing Knowledge from Multivariate Spatiotemporal Data: Integrating Geographical Visualization with Knowledge Discovery in Database Methods," *International Journal of Geographical Information Science*, vol. 13, no. 4, pp. 311-334, 1999.
- [14] M. Monmonier, "Geographic Brushing: Enhancing Exploratory Analysis of the Scatterplot Matrix," *Geographical Analysis*, vol. 21, no. 1, pp. 81-84, 1989.
- [15] M. New, M. Hulme, and P. Jones, "Representing Twentieth-Century Space-Time Climate Variability. Part I: Development of a 1961-90 Mean Monthly Terrestrial Climatology," *Journal of Climate*, vol. 12, no. 3, pp. 829-856, Mar, 1999.
- [16] D.E. Parker, P.D. Jones, C.K. Folland, and A. Bevan, "Interdecadal Changes of Surface-Temperature since the Late-19th-Century," *Journal of Geophysical Research-Atmospheres*, vol. 99, no. D7, pp. 14373-14399, Jul, 1994.
- [17] M. Parry, O. Canziani, J. Palutikof, P.v.d. Linden, and C. Hanson eds., "Climate Change 2007: Impacts, Adaptation and Vulnerability," Cambridge University Press, 2007.
- [18] T.C. Peterson, and M.O. Baringer, "State of the Climate in 2008," *Bull. Amer. Meteor. Soc.*, vol. 90, no. 8, pp. S1-S196, 2009.
- [19] A.R. Stine, P. Huybers, and I.Y. Fung, "Changes in the Phase of the Annual Cycle of Surface Temperature," *Nature*, vol. 457, no. 7228, pp. 435-440, 2009.
- [20] K.L. Swanson, G. Sugihara, and A.A. Tsonis, "Long-Term Natural Variability and 20th Century Climate Change," *Proceedings of the National Academy of Sciences*, vol. 106, no. 38, pp. 16120-16123, September 22, 2009, 2009.
- [21] C.M. Wittenbrink, E. Saxon, J.J. Furman, A. Pang, and S. Lodha, "Glyphs for Visualizing Uncertainty in Environmental Vector Fields," *Ieee Transactions on Visualization and Computer Graphics*, vol. 2, no. 3, pp. 266-279, 1995.
- [22] X. Zhang, and M. Pazner, "The Icon Imagemap Technique for Multivariate Geospatial Data Visualization: Approach and Software System," *Cartography and Geographic Information Science*, vol. 31, no. 1, pp. 29-41, 2004.