Computers, Environment and Urban Systems 39 (2013) 107–120

Contents lists available at SciVerse ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Visualizing gridded time series data with self organizing maps: An application to multi-year snow dynamics in the Northern Hemisphere

Ninghua Wang [a,*], Trent W. Biggs [b], André Skupin [b]

[a] Joint Doctoral Program in Geography, San Diego State University and University of California, Santa Barbara, United States
[b] Department of Geography, San Diego State University, United States

ARTICLE INFO

Article history:
Received 9 March 2012
Received in revised form 16 September 2012
Accepted 22 October 2012
Available online 24 November 2012

Keywords:
Snow water equivalent
Self organizing map
Visualization
Spatialization
Spatio-temporal analysis
Snow dynamics
Gridded time series

ABSTRACT

Gridded time-series data are increasingly available for climatology research. Microwave imagery of snow water equivalent (SWE) has been accumulated at daily basis for over two decades, but complex spatial-temporal patterns in SWE dataset pose great challenges for exploration and interpretation. This paper introduces the use of several perspectives from a tri-space conceptualization of a time series of SWE grids combined with dimensionality reduction via the self-organizing map (SOM) method. While SOM has been predominantly viewed as a clustering mechanism within climatology research, we expand the visual-analytic potential of SOM for climate research with a series of conceptual, computational, and visual transformations. Specifically, we apply a medium-resolution SOM to an SWE dataset covering the Northern Hemisphere over a 20-year period, with high temporal resolution. Through clustering and visualization a number of distinct SWE patterns are identified, including mountainous, coastal, and continental regions. A subset of cells from six areas are selected for transition analysis, including mountainous (Sierra Nevada, Western Himalaya, Eastern Himalaya) and continental (central Siberia, western Russia and Midwest United States) regions. By combining with trajectory analysis, this SOM documents notable transitions in seasonal SWE accumulation and melt patterns in mountain ranges, suggesting that SWE in some geographic locations alternates between different discrete annual patterns. In the Sierra Nevada, transitions to classes with high SWE are shown to be related to the Southern Oscillation Index, demonstrating that the annual patterns and transitions in SWE regime identified by the SOM correspond to synoptic climate conditions.

1. Introduction

Environmental monitoring data captured through multi-temporal satellite imagery is increasingly becoming a vital resource for a range of applications, including hydrologic models, climate change research, and land cover analysis. Where the coverage of large geographic areas combines with long time spans and high temporal resolution, one quickly faces the question of how to make sense of the resulting large data sets. Finding temporal and spatial structures, patterns, and trends is the task given to a range of methods, some focusing on computational approaches, like most clustering techniques, and others leveraging abilities of the human cognitive system through visualization, including animation. Self organizing maps (SOMs) have the potential to combine both computational and visualization approaches to understanding time series of gridded data. Although the clustering capacity of SOM is well recognized (Fassnacht & Derry, 2010; Hewitson & Crane, 2002; Hsu, Gupta, Gao, Sorooshian, & Imam, 2002; Kalteh, Hjorth, & Berndtsson, 2008; Sheridan & Lee, 2011), the potential of SOM to support multi-perspective analysis of detailed attribute patterns from time series data has received much less attention. Analytical aspects missing from the literature include the use of thorough multi-perspective conceptualization of time series data, such as the tri-space approach (Kolovos, Skupin, Jerrett, & Christakos, 2010; Skupin, 2010), and the use of much larger numbers of neurons than in traditional cluster applications of SOM (Skupin & Esperbé, 2008, 2011).

Here we demonstrate the visual-analytic potential of SOM as applied to a gridded time series of snow water equivalent (SWE). SWE was chosen both due to its importance in earth system processes, and in consideration of newly available, gridded datasets quantifying the geographic distribution of SWE for time series up to 20 years in length. Snow impacts a wide range of Earth surface processes, including albedo, atmospheric circulation, precipitation patterns (Bamzai & Shukla, 1999), runoff processes (Dunne & Black, 1971), and regional stream flow (Yang, Robinson, Zhao, Estilow, & Ye, 2003; Yang, Zhao, Armstrong, & Robinson, 2009). The pattern of accumulation and melt is particularly important for mountain

* Corresponding author. Address: 5500 Campanile Dr., San Diego, CA 92182, United States. Tel./fax: +1 619 594 8038.

E-mail address: wangn@rohan.sdsu.edu (N. Wang).

0198-9715/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2012.10.005

chains, which are "walls of water" that provide water resources for human activities in adjacent valleys and foreland basins (Viviroli, Dürr, Messerli, Meybeck, & Weingartner, 2007). Seasonal and interannual patterns of snowmelt have important consequences for anticipating the effects of climate change on surface climate and water resources (Gillan, Harper, & Moore, 2010; Miller, Bashford, & Strem, 2003; Singh & Kumar, 1997; Stewart, Cayan, & Dettinger, 2004). Optical remote sensing can be used to map snow covered area, while microwave imagery can be used to estimate snow water equivalent (Albert, 1993; Farmer, Nelson, Wulder, & Derksen, 2009; Grody, 2008; Saraf, Foster, Singh, & Tarafdar, 1999). A recently-completed time-series of SWE has been made available for a 20-year time period (1988–2007), but its decadal and annual structures and the evolution of regional difference over time have not yet been adequately explored. Existing maps of SWE and its climatology have been based on mean monthly or annual snow depth (Armstrong, Brodzik, Knowles, & Savoie, 2007). Spatial patterns in the annual pattern of accumulation and melt and their interannual variability are not well documented. Basic questions about the accumulation and melt of snow require different computational tools to explore the high dimensionality of gridded time series. For example, how do rates of accumulation and melt differ between mountains and continents? Do snow dynamics differ among mountain ranges – for example, are the rates of snow accumulation and melt in parts of the Sierra Nevada similar to the Himalaya or the Alps? How interannually variable is SWE, and how can that variability, including aspects of melt and accumulation not captured by annual totals, be visualized? These fundamental questions about the behavior of snow at the hemispheric scale cannot be easily explored without data analytical tools that can reduce the dimensionality of complex data for identification of patterns in snow dynamics.

One major challenge in spatio-temporal data analysis, including application of SOM techniques, is how to determine the granularity of a chosen representation such that meaningful groupings of observations emerge, while also being able to distinguish finer-grained variations among individual observations. In the context of exploratory analysis of multi-attribute data set, this can translate into a tension between clustering and dimensionality reduction (Agarwal & Skupin, 2008). Previous studies have tended to focus on the clustering ability of SOM, which distinguishes itself by explicitly representing topological relationships among clusters, typically in two dimensions. Instead of such low-resolution SOMs (i.e., small number of neurons), our study deploys medium-resolution SOMs (i.e., several hundred neurons) that allow for making more detailed distinctions among the SWE observations of geographic cells. Meanwhile, *k*-means clustering is then applied to neurons and the respective cluster memberships are propagated to geographic cells for subsequent visualization.

The initial SWE data set is interpreted within a tri-space framework aimed at accommodating multiple loci (here geographic cells) and multiple time periods (Kolovos et al., 2010). This tri-space approach (Skupin, 2010) consists of a systematic set of six conceptual perspectives on time series of multi-variate data. It subsumes existing solutions that address particular aspects of time series analysis (Andrienko et al, 2010; Guo, Chen, MacEachren, & Liao, 2006; Guo, Gahegan, MacEachren, & Zhou, 2005; Skupin & Hagelman, 2003). Though the SWE data set used in this study contains only a single attribute, it accommodates a nuanced representation of both long-term snow regimes and interannual transitions. Data thus conceptualized and preprocessed undergo a process of spatialization approach (Skupin & Fabrikant, 2003, 2008) aimed at making high-dimensional structures accessible to human cognition. Two fundamentally different conceptualizations of global SWE dynamics are adopted in this study, resulting in a single, long-term (i.e., "synoptic") map and a dynamic map of annual transitions (i.e., "annualized map"). The latter represents – within the climatology domain – a notable new approach for representing inter-annual SWE transitions as trajectories across a high-dimensional space. The correlation between the year-to-year movement of SWE trajectories and several synoptic climatic indicators was tested in several geographic regions.

The central goals of the research were to (1) demonstrate the use of SOMs to visually analyze long-term and interannual patterns in gridded SWE over the Northern Hemisphere (2) identify dominant modes of SWE patterns and their interannual transitions in select locations and (3) identify relationships between the interannual transitions as identified by the SOM with synoptic climatic indices.

The rest of the article is organized as follows: Section 2 introduces considerations and procedures of the SOM visual analytics process. Results are presented in Section 3 and findings and future directions are discussed in Section 4. Section 5 concludes this article.

2. Methods

In the reduction of multi-dimensional gridded data into a simpler, lower-dimensional representation, both computational and visualization approaches can be used. The SOM method can accomplish both: (1) it computationally captures major structures in multi-dimensional data and (2) the two-dimensional arrangement of those neurons exposes such structures to the power of visualization. It takes a set of *n*-dimensional vectors as input for making modifications to a set of *n*-dimensional neuron vectors (Kohonen, 2001, Agarwal and Skupin, 2008). The corresponding neurons are typically arranged as a two-dimensional lattice. During SOM training, individual input vectors are presented to the neuron lattice and the best-matching neuron vector (also known as best-matching unit or BMU) is determined, based on some measure of *n*-dimensional similarity. The BMU, as well as the vectors of neighboring neurons (as determined by lattice topology) are then adjusted to provide a better match to the input vector. A region of neurons in the lattice is thus conditioned to potentially provide BMUs for forthcoming input vectors that are similar to the prior one, as well as to repulse input vectors that are dissimilar. Over the course of thousands or millions of training runs, the SOM thus begins to replicate major structures existing in the n-dimensional input space. Following training, a SOM can be utilized in a number of ways. These include examination of the distribution of vector weights across the neuron lattice, as well the projection of *n*-dimensional vectors onto the SOM via computation of their BMUs (Agarwal and Skupin, 2008).

With a gridded data set, geographic cells – rather than distinct objects – act as loci to which one or more attributes can be attached over one or more time periods. Such data can be reorganized in different ways, for example to accommodate a chosen focus on loci or time slices (Andrienko et al., 2010; Guo et al., 2005, 2006; Skupin & Hagelman, 2003), with each focus manifesting a particular tri-space perspective (Skupin, 2010).

Two different conceptualizations of multitemporal SWE observations for geographic cells drive this study:

(a) Synoptic conceptualization: a view of geographical cells as existing in an *n*-dimensional space, with *n* being the number of time slices for which SWE was determined. In the case of our data set, this conceptualization expresses long-term snow cover characteristics in the Northern Hemisphere, based on a 20-year time series for each cell, with a temporal resolution of eight days. Each cell in the dataset belongs to a single neuron in the SOM representation of the data.