

# Investigating long-term trends of climate change and their spatial variations caused by regional and local environments through data mining

XIE Yichun<sup>1,2</sup>, ZHANG Yang<sup>3</sup>, LAN Hai<sup>4</sup>, MAO Lishen<sup>1</sup>, ZENG Shi<sup>5</sup>, CHEN Yulu<sup>1</sup>

1. Institute for Geospatial Research and Education, Eastern Michigan University, Ypsilanti, Michigan 48197, USA;

2. Guangzhou Institute of Geography, Guangzhou 510070, China;

3. Department of Computer Science, Indiana University, Bloomington, Indiana 47405, USA;

4. Department of Computer Science, New York University, NY 10012, USA;

5. Center for Advanced Spatial Analysis, University College London, London WC1E 6BT, UK

**Abstract:** Climate change is a global phenomenon but is modified by regional and local environmental conditions. Moreover, climate change exhibits remarkable cyclical oscillations and disturbances, which often mask and distort the long-term trends of climate change we would like to identify. Inspired by recent advancements in data mining, we experimented with empirical mode decomposition (EMD) technique to extract long-term change trends from climate data. We applied GIS elevation model to construct 3D EMD trend surface to visualize spatial variations of climate change over regions and biomes. We then computed various time-series similarity measures and plot them to examine spatial patterns across meteorological stations. We conducted a case study in Inner Mongolia based on daily records of precipitation and temperature at 45 meteorological stations from 1959 to 2010. The EMD curves effectively illustrated the long-term trends of climate change. The EMD 3D surfaces revealed regional variations of climate change, while the EMD similarity plots disclosed cross-station deviations. In brief, the change trends of temperature were significantly different from those of precipitation. Noticeable regional patterns and local disturbances of the changes in both temperature and precipitation were identified. The trends of change were modified by regional and local topographies and land covers.

**Keywords:** climate change; empirical mode decomposition; Inner Mongolia; similarity plot; trend surface

## 1 Introduction

Global climate change has been widely recognized as a new natural threat to biodiversity (Dawson *et al.*, 2011) and human welfare (de Sherbinin, 2011) in the 21st century. Many modern techniques have been developed by scientists to collect evidences concerning cli-

**Received:** 2017-08-23 **Accepted:** 2017-12-10

**Foundation:** Guangdong Innovative and Entrepreneurial Research Team Program, No.2016ZT06D336; GDAS Special Project of Science and Technology Development, No.2017GDASCX-0101

**Author:** Xie Yichun (1956–), PhD and Professor, specialized in urban modelling, ecological modelling and environmental modelling. E-mail: [yxie@emich.edu](mailto:yxie@emich.edu)

mate change contained in tree rings, ice cores, greenhouse gas levels, shoreline changes, glacier and permafrost melt, radiocarbon dating, air and sea temperature, phenology, etc. Among them, weather conditions – daily temperature and rainfall – have long been recorded and analysed to provide a direct evidence of climate change. For instance, climate change has been manifested in increased global temperatures, but also in increasing frequency of extreme weather events such as floods and droughts, severe winds, and increased temperature extremes of both hot and cold (National Climate Assessment, 2014; Rahmani and Zarghami, 2015; Xia *et al.*, 2015).

Global climate change (e.g., global warming, increased/decreased precipitation, and extreme weather event) is regarded as one of the primary factors that are impacting grassland ecosystems (Kysely *et al.*, 2012; Piras *et al.*, 2015; Ribalaygua *et al.*, 2013). For instance, the Mongolia Plateau is getting warmer and drier. The average temperature of Mongolia increased by 1.5°C to 2.5°C in the 1990s and 2000s (Lu *et al.*, 2009), while the annual precipitation decreased by about 7.0% during the same period (Wang *et al.*, 2013). Global climate change is assumed to affect growth condition as well as spatial distribution of plant communities (Li S *et al.*, 2013). Drying and warming climate usually increases fluctuation of vegetation productivity (Bai, 2008; Gong, 2015) and leads to plant community degradation (Li and Xie, 2013; Xie *et al.*, 2017).

Moreover, ecosystem researchers have recognized that different plant communities show varied responses to climate change (Brown *et al.*, 2013). Although precipitation and temperature are two main climate factors affecting vegetation response (Han, 2016; Wang *et al.*, 2012), plant communities respond differently to short-term or long-term changes of precipitation and temperature, respectively (Chuai, 2007; Yuan, 2015). Second, vegetation responses to climate change display significant local and regional variations (Bai, 2008; Lu *et al.*, 2009; Xie *et al.*, 2017). Third, climate change displays apparent regional variations. For instance, mountainous regions are usually more sensitive to climate change than flat and low elevation regions (Damsø *et al.*, 2015). Unique locational arrangements of land masses, oceans or water bodies, dominant air flows, topographies and elevations can have significant impacts on local meteorological conditions and thus lead to distinct regional patterns of climate change (Cheng *et al.*, 2015). Even within the same river basin, the yields of rice and wheat could display significant regional variations because of the influence of agro-climatic factors, such as variation in temperature, length of maturity period and leaf area index (Mishra *et al.*, 2013; Swain and Thomas, 2010). Climate change in combination with land use change could make water quality and land productivity deteriorate swiftly and lead to noticeable spatial variations (Jordan *et al.*, 2014).

Furthermore, studies of vegetation response to climate change are facing many challenges. Climate changes contain remarkable cyclical oscillations and disturbances, which often mask and distort the long-term trends we would like to identify (Kennedy *et al.*, 2014). Annual cycles are critical phenomena of temperature and precipitation. When analysing climate change, we should bear in mind that the seasonal changes are interacting with a lot of other dynamics, such as long-term warming/cooling or drying/wetting trends, abrupt weather events (flooding, drought, hot wave, and cold front), pollutant emissions, solar activity cycles, etc. Traditional analytical methods based on the comparisons between minimum, average and maximum values of temperature and precipitation are not capable of separating

long-term trends from cyclical fluctuations and abrupt changes or capturing temporal dynamics or regional patterns of climate change (Chamaille *et al.*, 2007). As a result, it is almost impossible to study long-term interactions between meteorological conditions and the underlying landscape, vegetation and topography by simply analyzing the records of temperature and precipitation. Therefore it is desirable to apply an effective data analysis method to break down climate variations into individual processes, i.e., cyclical, long-term and abrupt components. Only with this type of data mining and pre-processing is it feasible to investigate spatial patterns and interactions between climate change and regional environmental factors.

In this paper, we will examine diverse responses of plant communities to climate change and their spatial variations, visualize temporal trajectories and spatial patterns of climate change at regional scale, and compare spatial variations of climate change across meteorological stations. Since current GIS tools are not adequately supporting analysis and visualization of temporal trajectories and spatial patterns of climate change, we integrate recent computational data mining approaches with GIS. In particular, we will synthesize advanced signal processing and denoising techniques to extract long-term trends of climate change. We will also adopt the similarity analysis and surface visualization methods often seen in big data analytics to visualize spatial variations of the identified change trends and to analyse their relationships with landscape, vegetation and topography at region, biome and weather-station scales.

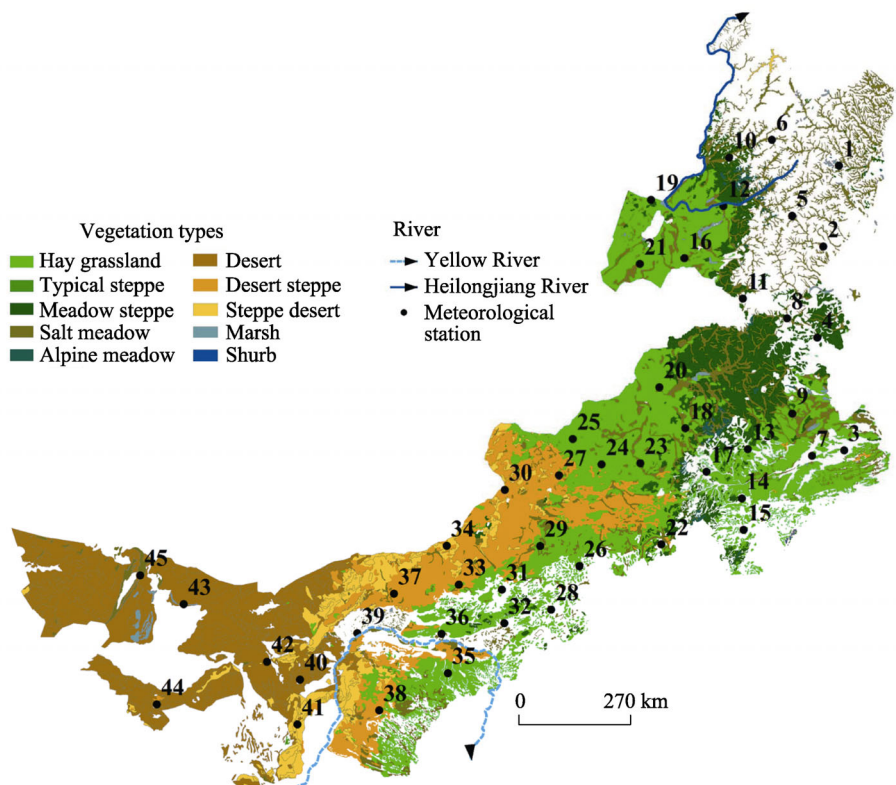
## 2 The study area, data and method

### 2.1 The study area and data

Inner Mongolia Autonomous Region (IMAR, 37°24'–53°23'N, 97°12'–126°04'E) is located in China's northern border region with a total area of about 1,180,000 km<sup>2</sup> (Figure 1). IMAR is also located in the southern portion of the Mongolian Plateau with an average altitude of 1000–1200 m. The climate of the steppe area is a typical temperate continental climate, with an annual precipitation of 50–450 mm and an annual average temperature of 1–10°C. The climate in the study area experiences a gradual transition from humid and semi-humid regions to semi-arid and arid regions from east to west. Precipitation shows a gradual decrease from northeast to southwest, while temperatures gradually increase from the northeast toward the southwest (Li J *et al.*, 2013).

Grassland is the dominant land cover in IMAR, which is concentrated in the central part of IMAR, while most of the forest is located in the northeastern section dominated by broad-leaf and needle-leaf forests and cropland in the southern and eastern regions (Li J *et al.*, 2013). Traditionally, grazing has long been the primary economic activity in IMAR. However, since the late 20th century, due to many coupled natural and human factors, such as climate change, economic development, population growth, and overgrazing, grassland ecosystems in IMAR suffered severe degradation and even desertification in some areas. Grasslands occupy almost 40% of the earth's land surface, support nearly one third of global population (Gibson, 2009), and boast many important ecological functions, including soil and water conservation, carbon sequestration, wildlife habitat, etc. (Carlier *et al.*, 2009). In addition, the Mongolian Plateau is the largest stretch of grasslands remaining on the earth, in

which IMAR is an important part (Xie *et al.*, 2017). Therefore, the selection of IMAR grassland as the case study has important implications in both academic research and policy management.



**Figure 1** The map of the study area (Meteorological stations are numbered from east to west)

The vegetation map of Inner Mongolian Autonomous Region in 2010 was provided by Inner Mongolian Institute of Grassland Surveying and Planning (Li S *et al.*, 2013). Meteorological data was extracted from the China Meteorological Data Service Center (CMDC, 2013), consisting of 50 surface meteorological stations distributed in Inner Mongolia (Figure 1) over 50 years, 1959–2010. The data includes the longitude and latitude information of each meteorological station, daily precipitation and temperature of each meteorological station.

**2.2 The research methods**

The paper synthesizes three groups of methods: (1) trends analysis derived from computational data mining (empirical mode decomposition – EMD) to extract long-term trends of change from cyclical climate data; (2) similarity analysis stemmed from data mining to examine spatial variations of climate change across meteorological stations; and (3) GIS 3D and 2D visualization techniques to reveal regional and cross-station patterns of plan community responses to climate change.

**(1) Trends analysis of cyclic climate change**

Data mining is a newly acknowledged-field that has received vast attention from computer science and information science researchers. Data mining refers to an analytic process designed to search for consistent patterns and/or systematic relationships between variables

from large volumes of data (also known as “big data”) (Jain and Srivastava, 2013). One vital goal of data mining is to construct models and then to apply these models to new data to generate predictions (Huang *et al.*, 1998).

Signal processing and intelligent recognition is one technique within an increasingly growing data mining toolbox. Data mining in many fields involves constantly monitoring real-time conditions on the basis of signals collected by sensors. The datasets of these signals are usually recorded or saved in the form of time series. Therefore, suitable signal processing techniques are needed in order to extract information from such signals and to disclose underlying dynamics embedded in these time-series data (Gao and Yan, 2011). These techniques serve two purposes to represent the reality sensed by various sensors: first to determine the parameters to create an abstract model of the reality, and second to confirm the model that can represent the reality to certain degree (Huang *et al.*, 1998). The challenge is to represent the real world as close as possible while eliminating as much noises and abrupt interference on data/signal as possible. In the past, many methods were developed to achieve this goal according to the special field in which the data analysis is applied. In general, these signal processing techniques could be grouped into three approaches: Fourier transform, wavelet transform, and empirical mode decomposition.

Fourier methods traditionally are used to approximate any general function as a sum of trigonometric functions (Grafakos and Teschl, 2013). Wavelet transforms, also called “mathematical microscopes (Bovik, 2009, p. 463)”, include a suite of signal processing techniques that are developed to filter signals by using a different centre frequency in the band-pass filter, in which small scales of the noise frequency can be removed to get good-quality and useful signals (Portilla *et al.*, 2003). Wavelet-based denoising at various scales aims to achieve high resolution in both the time and frequency domains (Dai *et al.*, 2006; Chen and Xu, 2005). The principle of EMD is to decompose the signal into a group of similar sinusoidal signals, which was defined by the signal itself, named the intrinsic mode functions (IMFS), and a residue (Huang, 1998; Gloersen and Huang, 2003; Rao and Hsu, 2008). The IMFS reveal the status of the signal in various scales, and the residue told us the trend of the signal, which is the statistic we are interested in this paper.

In recent years, due to the increasing popularity of computational data mining, the EMD method has been applied in many research fields, such as, removing noise of time series data (Huang *et al.*, 2001; Peng *et al.*, 2005); analysing the properties of time series data in finance (Huang *et al.*, 2003); and applications in hydrology and environment (Rao and Hsu, 2008), goaf surface deformation (Zhang, 2011), temperature trend extraction (Xian *et al.*, 2008), image-based land cover classification (Demir and Ertürk, 2010) and vegetation analysis (Chen and Xu, 2005; Ghasemi *et al.*, 2013; Cheng *et al.*, 2014). Mathematically, EMD iteratively applies an intrinsic mode function (IMF), which decomposes complex signals into a number of distinct, simple, and non-sinusoidal signals along with a trend curve. This EMD trend curve reflects the trend of change of time-series signals and is used to identify long-term trend of changes hidden in the cyclical datasets of climate change and ecological evolution.

The mathematical algorithm of EMD is not the same as Fourier method and the Discrete Wavelet Transform although they belong to the same family of time domain signal analysis techniques (Zhang *et al.*, 2015). EMD can be applied to decompose non-linear and non-stationary datasets in comparison with Fourier. Contrasting to wavelet denoising, EMD

method has similar distortion magnitude in the process of denoising signals (Luan *et al.*, 2004). However, the application of EMD denoising does not require to set up a priori classification function as wavelet algorithm does. This enables EMD method becomes more robust and stable.

(2) Similarity measurement of spatial variations of climate change across meteorological stations

We adopt a time series analysis of climate change to characterize climate change dynamics and reveal their spatial variations across meteorological stations. Various methods have been developed based on temporal trajectories to characterize changes in ecosystem dynamics in recent years (Lhermitte *et al.*, 2011). These methods are also suitable to the studies of climate change due to the similar nature of temporal dynamics. The key techniques in these time series analyses are methods of identifying similarities or dissimilarities between two sequences of measurements (Goshtasby, 2012; Zastrow, 2015).

(a) Euclidean distance

$$d_{st}^2 = (x_s - x_t)(x_s - x_t)' \quad (1)$$

where  $x_s$  is a data vector with the size of  $(1 \times s)$ , and  $x_t$  is another data vector with the size of  $(1 \times t)$ . In our case,  $s = t$ . The symbol,  $'$ , represents the transposition operation of a vector (e.g.  $(x_s - x_t)'$  represents the transposed vector of the original vector  $(x_s - x_t)$ ).

(b) Standardized Euclidean distance

$$d_{st}^2 = (x_s - x_t)V^{-1}(x_s - x_t)' \quad (2)$$

where  $V$  is the  $n$ -by- $n$  diagonal matrix whose  $j$ th diagonal element is  $S(j)^2$ , where  $S$  is the vector of standard deviations. Each coordinate difference between rows in  $X$  is scaled by dividing by the corresponding element of the standard deviation.

(c) City block metric

Manhattan distance assumes that in going from one pixel to the other it is only possible to travel directly along pixel grid lines and diagonal moves are not allowed. Therefore, the distance between centroid is given by:

$$d_{st} = \sum_{j=1}^n |x_{sj} - x_{tj}| \quad (3)$$

(d) Chebyshev distance

Chebyshev distance or the  $L_\infty$  metric (Luan *et al.*, 2004), is defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension.

$$d_{st} = \max_j \{|x_{sj} - x_{tj}|\} \quad (4)$$

(e) Cosine distance

Cosine distance measures the similarity between two vectors of an inner product space in terms of the cosine of the angle between them. This technique has been used to measure cohesion within clusters in the field of data mining.

$$d_{st} = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}} \quad (5)$$

(f) Correlation coefficient

Pearson correlation coefficient is suitable for determining the similarity between images

with intensities that are known to be linearly related (Goshtasby, 2012). The correlation coefficient between sequences  $X = \{x_i: i = 1, \dots, n\}$  and  $Y = \{y_i: i = 1, \dots, n\}$  is defined by

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \right\}^{\frac{1}{2}} \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 \right\}^{\frac{1}{2}}} \quad (6)$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

(g) Spearman distance

A similarity measure relating to the Pearson correlation coefficient is Spearman's rank correlation or Spearman's Rho (Goshtasby, 2012). If image intensities do not contain ties when they are ordered from the smallest to the largest, then by replacing the intensities with their ranks and calculating the Pearson correlation coefficient between the ranks in two images, which will give Spearman rank correlation. This is equivalent to calculating (Goshtasby, 2012):

$$d_{st} = 1 - \frac{(r_s - \bar{r}_s)(r_t - \bar{r}_t)'}{\sqrt{(r_s - \bar{r}_s)(r_s - \bar{r}_s)'} \sqrt{(r_t - \bar{r}_t)(r_t - \bar{r}_t)'}} \quad (7)$$

where

$r_{sj}$  is the rank of  $x_{sj}$  taken over  $x_{1j}, x_{2j}, \dots, x_{mj}$

$r_s$  and  $r_t$  are the coordinate-wise rank vectors of  $x_s$  and  $x_t$ , i.e.,  $r_s = (r_{s1}, r_{s2}, \dots, r_{sn})$

$$\bar{r}_s = \frac{1}{n} \sum_j r_{sj} = \frac{(n+1)}{2}$$

$$\bar{r}_t = \frac{1}{n} \sum_j r_{tj} = \frac{(n+1)}{2}$$

Spearman rank correlation is less sensitive to outliers, impulse noise and occlusion. It is also less sensitive to nonlinear intensity difference between images than Pearson correlation coefficient (Goshtasby, 2012). Because of these features, Spearman rank correlation has been used to measure trends in data as a function of time or distance.

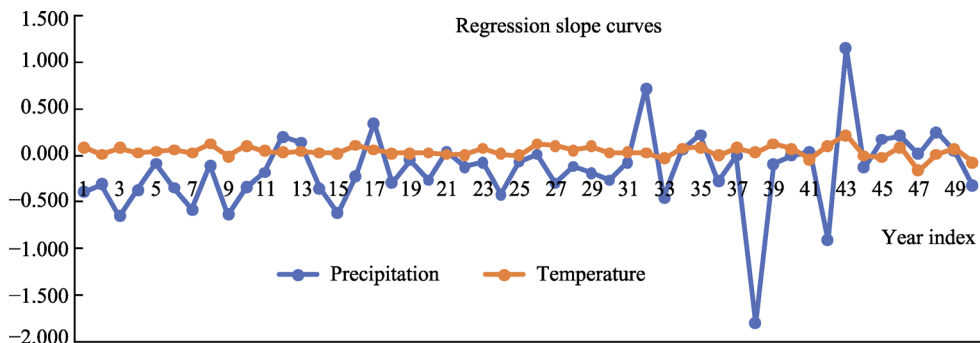
(3) Visual data mining methods for identifying regional and cross-station patterns

In geographical information science, 3D surface models are important tools in GIS for conducting spatial analysis and visualizing the outcomes of the spatial analysis (Li *et al.*, 2005). A typical example is the digital elevation model (DEM) that represents the earth's elevation surface. In this paper, we are using the same technique to visualize the EMD trend line as an elevation surface over the entire study area. These EMD 3D trend surfaces clearly visualize the regional patterns of climate change during the study period (1959 – 2010) over the study area and can be examined visually. In addition, 2D contour lines derived from the 3D surfaces are also plotted to provide different views of the regional patterns of climate change. The 3D and 2D surfaces clearly depict regional patterns of plant community responses to climate change and help reveal topographical impacts on varied responses of plant communities to climate change.

The similarity measures are also plotted as 2D maps to visualize spatial variations of precipitation and temperature across the meteorological stations. The 2D similarity plots visibly reveal geographical differences of precipitation and temperature in the study area.

### 3 The case study of climate change in Inner Mongolia

As a starting point, we calculated annual precipitation by adding daily precipitation and annual average temperature by averaging daily temperature at each station. As a result, we got 51 records of annual precipitation and annual average temperature for all 50 meteorological stations. In order to identify long-term trends of climate change at 50 stations, we ran regression analysis of precipitation against the year and temperature against the year, respectively. We then plotted the regression slopes of precipitation and temperature as curves over the stations from east to west (Figure 2). Apparently, 38 out of 50 stations observed a decline in precipitation because they had negative slopes (under the 0.000 horizontal line). Moreover, the precipitation slope curve showed obvious ups and downs, revealing significant differences of change among the stations. On the contrary, the temperature slope curve was much smoother. 44 out of 50 stations witnessed an increase in temperature.



**Figure 2** Regression slopes of precipitation and temperature, 1959–2010, at meteorological stations

Next, we examined variations of climate change over three scales, regional, vegetation type, and meteorological station, by using the methods we introduced before. The daily precipitation and temperature data from 1959 to 2010 over 45 meteorological stations were analysed. Five stations were located in urban areas and were excluded from the analysis due to the paper’s focus of exploring different responses of vegetation communities to climate change. Through EMD method, we transformed the daily climate data at each station incrementally into a long-term trend curve. The trend curves of temperature and precipitation were visualized as 3D surfaces and 2D plots by all stations from east to west. Furthermore, we applied the above-mentioned eight different similarity measurements to the EMD curves for all 45 stations. For each similarity measurement, we created a similarity matrix of 45×45 stations and then examined its sensitivity of spatial variations by stations. For sensitive similarity measurements, we created 2D plots to illustrate station-wide variations.

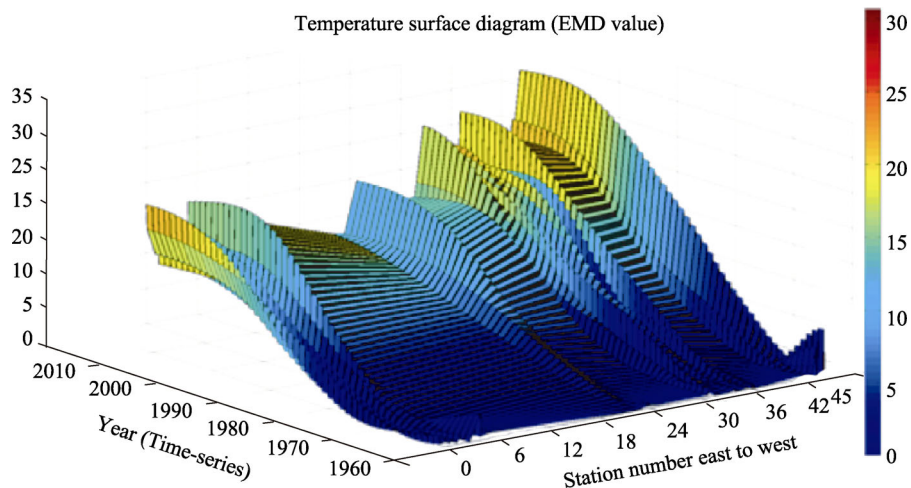
In the first visualization, we produced a vegetation map with the meteorological stations numbered from the east to the west and with ten types of vegetation mapped in different colours. The vegetation map helped explain the east-to-west distribution patterns of different biomes. From the vegetation map we could see the “wettest” grasslands (the meadow types) matched with the Da Hinggan Mountains stretching north to south in the eastern region of the study area and the Heilongjiang River (Figure 1). The typical steppe and hay grassland are located in the northeastern-central section and the southwestern-central section. The desert-type grasslands occupy the northwestern-central section and the western section. The dryness increased toward the west. The spatial pattern of biomes deeply affected the trends



of precipitation and temperature changes from 1959 to 2010.

### 3.1 The temporal trends of temperature changes

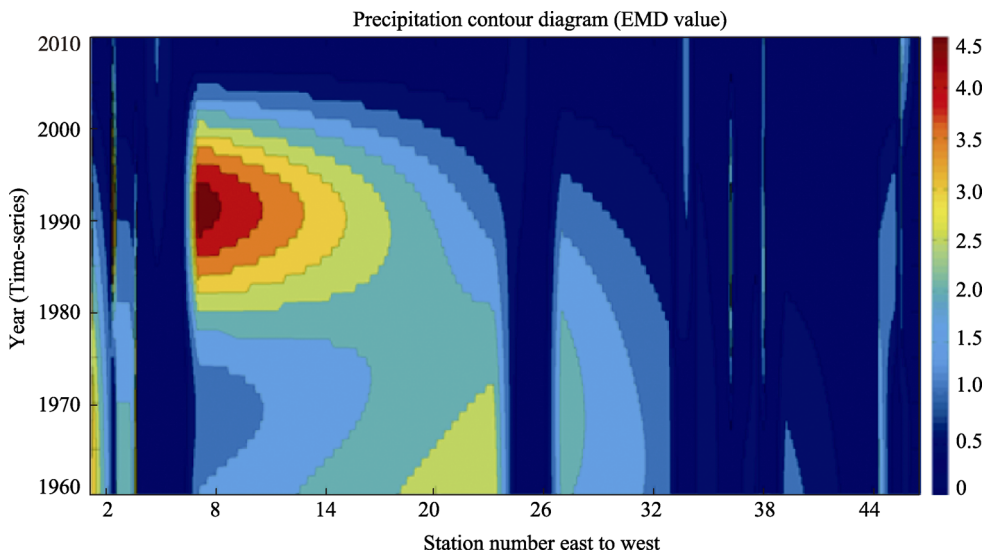
Several trends of temperature changes were identified on Figures 3 and 4: (1) temperature witnessed an overall and consistent increase in the past 50 years; (2) temperature changes were moderate in the first 35 years but dramatic in the recent 15 years; (3) temperature changes showed a regional pattern, becoming warmer from east to west (because the widths of cold temperature bands were decreasing from east to west but the widths of warm temperature bands were increasing); (4) the regional pattern of temperature changes revealed noticeable spatial variations; and (5) regional variations of temperature changes were impacted by biomes. For example, around station 12 in meadow area close to the Heilongjiang River (Figures 1, 3 and 4), the relative EMD value increased from 0 to 10, station 5 in mountain area from 0 to 20 and station 40 in desert area from 0 to 30. In other words, the change trends differed locally and regionally. The desert and mountain areas responded quicker to the temperature increase than the wet biomes.



**Figure 3** Temperature EMD surface diagram

It was clear that the EMD values constantly increased over most stations in the past 50 years (Figure 3). The range of the EMD value changes was 10 to 30, depending on the types of biome and topography where the stations were located. In general, the EMD values increased by 10–12 points before 1995 but jumped additional 18–20 points from 1995–2010 (Figure 4). Moreover, the trend curves of EMD were raised from east to west except for several unique regional patterns (Figures 3 and 4). Firstly, most of the stations of 1–12, are located in the east edges of the Hulun Buir Steppe and the valleys of Da Xinggan Mountains with relatively less productive biomes (primarily salt meadow and some typical steppe). Therefore, the EMD changes were over 16–18 points. However, there were a couple of stations (2 and 6) around the Hulun Lake, which showed that the EMD changes were less than 15 points. Secondly, from stations 16 to 32, except for stations 25 and 26, we saw a wide stretch of blue belt and the EMD increase was a little bit over 10 points (Figure 4). Most of these stations were located in Xilinhot Typical Steppe, which was the heartland of Eurasian temperate grasslands (Xie *et al.*, 2009), and, hence, the EMD increases were moderate. Finally, from station 34 and further west, we witnessed an extensive brown and red stretch of

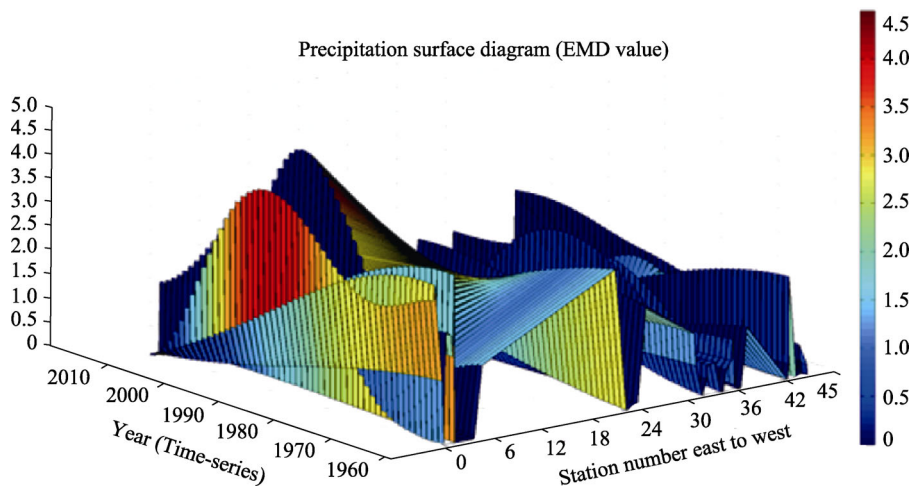
higher EMD values, which signified apparent warming impacts. Most of these stations were located in either deserts or desert steppe with much drier environments. Therefore, the EMD curves were raised much higher than the remaining areas (Figure 3). Clearly different environments showed diverse susceptibility to climate change.



**Figure 4** Temperature 2D contour diagram

**3.2 The temporal trends of precipitation change**

The temporal trends and spatial patterns of precipitation changes in the past 50 years were different from those of temperature changes. First, the precipitation generally decreased from east to west but showed complicated temporal trajectories and spatial patterns. Secondly, there was a singular precipitation increase event (reflected by higher EMU values) in the past 50 years. This precipitation increase was centered on the Hulun Lake area (Xin Barag Left Banner) and extended to the east section of the Hulun Buir Steppe and the northern section of the Da Xinggan Mountains (Figures 5 and 6). This increase event started from



**Figure 5** Precipitation surface diagram

1975, reached the maximum around 1990–1994, and returned to normal in 2004, lasting almost 30 years. Thirdly, the precipitation variation in the west section of the Hulun Buir Steppe and the Xilin Gol Steppe gradually decreased. Moreover, the decrease was more apparent in the wetter areas. For instance, the precipitation EMD curves over the western Hulun Buir and the eastern Xilinhot (stations 14–24) declined more noticeably and quickly than those over the western Xilinhot (stations 27–32). Fourthly, over the deserts and desert steppes (stations 4–6, 25–26 and 34–44), the precipitation EMD curves did not show a clear increasing trend (Figure 6). Finally, the EMD curves over the desert areas displayed moderate temporal changes (Figure 5) although no clear linear trends were identified.

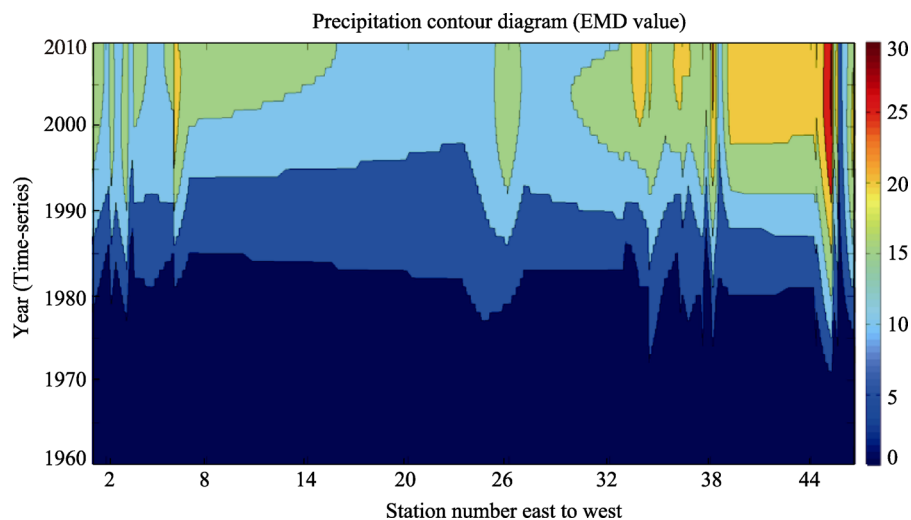


Figure 6 Precipitation contour diagram

3.3 Similarity measurement results

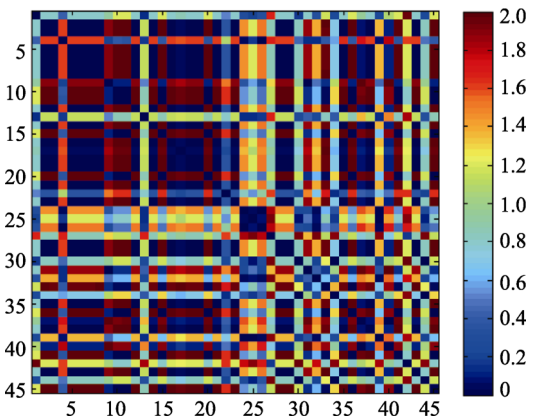
We applied the above-mentioned seven similarity measurements to the EMD dataset of 45 meteorological stations. We computed the similarity measurements of the EMD values across 45 meteorological stations. In other words, we compared similarity degrees among the meteorological stations. We also calculated the root mean square error (RMSE) to examine the sensitivity levels of these similarity measures (Table 1). The values of RMSE for temperature were usually lower than 0.30 and showed slight changes between different similarity measures. Thus, the spatial variation of temperature across meteorological stations was

Table 1 Seven similarity measures: Root Mean Square Errors (RMSE)

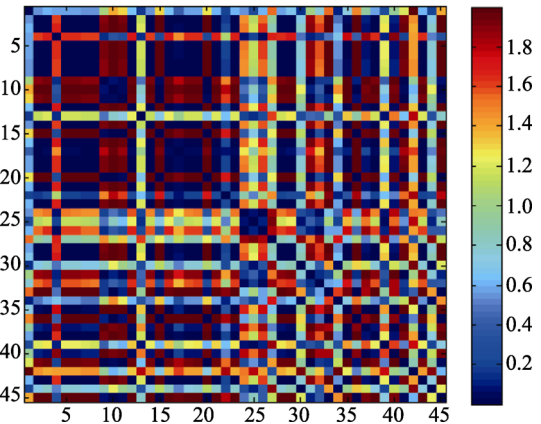
Precipitation	RMSE	Temperature	RMSE
Chebyshev	0.4249	Chebyshev	0.3008
City block	0.3870	City block	0.2600
Correlation	0.5956	Correlation	0.2896
Cosine	0.3898	Cosine	0.2264
Euclidean distance (E.D.)	0.4170	Euclidean distance (E.D.)	0.2654
Spearman	0.5912	Spearman	0.2805
Standardized E.D.	0.4413	Standardized E.D.	0.3042

gradual and not dramatic. However, the RMSE values for precipitation were usually above 0.38. Especially, the RMSE values of correlation distance (C.D.) and Spearman distance (S.D.) measurements were above 0.50 and almost close to 0.60. The higher RMSE values indicated that the similarity measures were poor for precipitation (Veerasamy *et al.*, 2011), which reflected significant variations in precipitation among the meteorological stations. For the purpose of revealing spatial variations across the stations, we were reporting C.D. and S.D. measures in the following graphics.

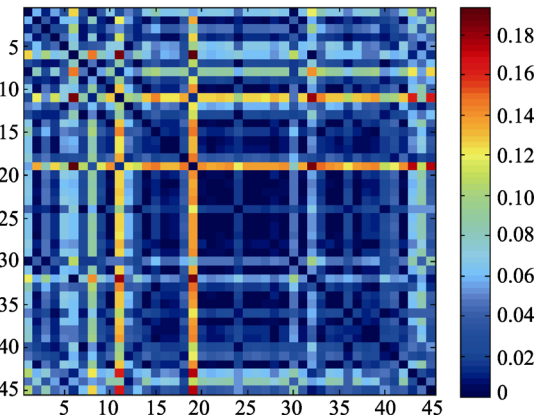
Figures 7 and 8 were the S.D. and C.D. cross-difference maps of 45 stations in precipitation, respectively. Both displayed clear and similar patterns. These maps could be clearly read in four quads. The upper left (U-L) quad revealed the similarities or differences of the stations numbered 1 to 23. In comparison, the similarities between them were the smallest. Moreover, significant different similarities were identified around stations 3, 9–11, 13, 15, and 20 in this quad. Seen from Figure 1, we found out that these stations were either far away in distance from the stations numbered around them, or these stations were located in different biomes. The lower right (L-R) quad was another extreme, displaying dense small cells and thus indicating apparent differences between these stations numbered above 23. The different similarity measures in these two quads were consistent with the biomes observed on ground. The stations in U-L were located in meadow and typical grasslands, while the stations in L-R were largely located in deserts, desert steppes and steppe deserts. The remaining two quads depicted similarity comparisons between the stations located in quads 1 and 4. In other words, the similarity differences were identical in these quads. The variations were intermediate between U-L and



**Figure 7** Cross-station plot of Spearman distance for precipitation



**Figure 8** Cross-station plot of correlation distance for precipitation

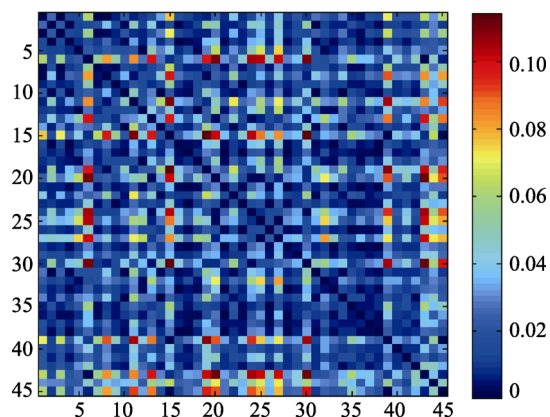


**Figure 9** Cross-station plot of Spearman distance for temperature

L-R quads.

Figures 9 and 10 were the S.D. and C.D. cross-difference maps of 45 stations by temperature, respectively. The temperature similarity maps showed different characteristics from the precipitation maps. First, there were no apparent spatial patterns. Second, the similarity differences of the meteorological stations in terms of temperature were generally smaller in comparison with precipitation. Third, the S.D. temperature map was revealing more information than the C.D. map. For instance, the S.D. map indicated

that there existed some temperature irregularities among the stations numbered 1 to 11. On the other hand, the C.D. map displayed more descrambled similarities.



**Figure 10** Cross-station plot of correlation distance for temperature

#### 4 Conclusions and discussion

We developed a statistical-cum-visual method on the basis of data-mining techniques that became available in recent years in order to investigate long-term trends and examine spatial patterns of ecological, environmental and geographical processes that are signified with cyclical or seasonable dynamics. We applied the EMD technique to extract long-term climate change trends and used 3D surface maps and 2.0D contour maps to visualize differences of change trends in three scales of region, biome and station. Moreover, we experimented with another data-mining technique, the similarity measurement, and compared seven types of commonly used similarity measures. We also visualized these similarity measures by using the cross-station plots. We tested these methods through a case study of investigating climate change in Inner Mongolia based on the daily observations of precipitation and temperature from 1959 to 2010 at 45 meteorological stations.

The case study confirmed that the selected data-mining methods and geo-visualization techniques innovatively and effectively revealed long-term climate change trends and visualized spatial patterns of climate change in three scales, region, biome and station. Moreover, two data-mining methods, EMD and the similarity measurement, complemented each other, disclosing different characteristics of spatial patterns of climate change.

Temperature, during the study period (1959–2010), increased across the study area based on the EMD trend-curves. However, the increases of temperature revealed significant temporal variations and spatial patterns. Temperature increased gently before 1995 and dramatically after 1995. Temperature increased slowly in the biomes of meadow and typical grasslands but quickly in the desert-type grasslands. When the environment was drier, the temperature increased more quickly. Furthermore, the similarity measurement illustrated that temperature trend (EMD) curves showed slight variation over meteorological stations. The cross-station similarity plots of temperature by the stations did not reveal noticeable spatial patterns.

On the other hand, the change trends of precipitation in the past 50 years on the basis of

EMD curves displayed complicated temporal trajectories and spatial patterns. A significant increase centre was accompanied by an overall decrease in other areas. From the perspectives of regional and biome scales, no clear linear trends were identified. Furthermore, from the station point of view, based on the similarity measures and cross-station plots, similar change trends of precipitation were found for the stations located in the meadow and typical grasslands. However, this finding didn't hold for the stations located in desert-type grasslands.

The above findings provided very convincing evidences to support the IPCC predictions that the climate change varied significantly by location and through time. The influences of climate change showed different temporal trends and spatial disparities at varied scales. The reactions to climate change displayed different trajectories over different regions, biomes and locations. The integrated data-mining-cum-visual method was very effective in revealing change trends and their spatial patterns of climate changes. The methods developed in this study are also suitable for investigating long-term trends and spatial patterns of other ecological processes that are signified with cyclical or seasonable fluctuations.

Finally, there is a noticeable limitation of current method. Although the primary gradients of precipitation and temperature changes are from east toward west, there is a significant distance in the south-north direction. Moreover, the meteorological stations are not located in the same latitude. Using the westward locations of the meteorological stations to describe the spatial pattern of east-toward-west vegetation responses to climate change neglected spatial variations of south-toward-north changes.

## References

- Bai Y, Wu J, Xing Q *et al.*, 2008. Primary production and rain use efficiency across a precipitation gradient on the Mongolia Plateau. *Ecology*, 89(8): 2140–2153.
- Bovik A C, 2009. The Essential Guide to Image Processing. Burlington, MA: Academic Press (Elsevier), 853.
- Brown D G, Agrawal A, Sass D A *et al.*, 2013. Responses to climate and economic risks and opportunities across national and ecological boundaries: Changing household strategies on the Mongolian Plateau. *Environmental Research Letters*, 8(045011): 9. doi: 10.1088/1748-9326/8/4/045011.
- Cade B S, Noon B R, 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8): 412–420.
- Carlier L, Rotar I, Vlahova M *et al.*, 2009. Importance and functions of grasslands. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 37(1): 25–30.
- Chamaille S, Jammes B, Fritz H *et al.*, 2007. Short communication detecting climate changes of concern in highly variable environment. *Journal of Arid Environments*, 71(3): 321–326. doi: 10.1016/j.jaridenv.2007.05.005.
- Chen J, Xu Y, 2005. Application of EMD to signal trend extraction. *Journal of Vibration, Measurement and Diagnosis*, 25(2): 101–104.
- Cheng F Y, Jian S P, Yang Z M *et al.*, 2015. Influence of regional climate change on meteorological characteristics and their subsequent effect on ozone dispersion in Taiwan. *Atmospheric Environment* 103: 66–81.
- Cheng T, Riaño D, Ustin S L, 2014. Detecting diurnal and seasonal variation in canopy water content of nut tree orchards from airborne imaging spectroscopy data using continuous wavelet analysis. *Remote Sensing of Environment*, 143: 39–53.
- Chuai X W, Huang X J, Wang W J *et al.*, 2013. NDVI, temperature and precipitation changes and their relationships with different vegetation types during 1998–2007 in Inner Mongolia, China. *International Journal of Climatology*, 33(7): 1696–1706.
- Cianfrani C, Satizábal H F, Randin C, 2015. A spatial modelling framework for assessing climate change impacts

- on freshwater ecosystems: Response of brown trout (*Salmo trutta* L.) biomass to warming water temperature. *Ecological Modelling*, 313: 1–12.
- Cleve B M, 2004. Fourier Analysis. Society for Industrial and Applied Mathematics 235–253. doi: <http://dx.doi.org/10.1137/1.9780898717952>.
- CMDC (China Meteorological Data Service Center), 2013. Inner Mongolian Meteorological Stations and Data. <https://data.cma.cn/en> (the data was acquired in 2013).
- Connors R W, Harlow C A, 1980. A theoretical comparison of texture algorithms. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2(3): 204–222.
- Coselmon M M, Balter J M, McShan D L *et al.*, 2004. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Medical Physics*, 31(11): 2942–2948.
- Dai G, Liu F, 2007. Instantaneous parameters extraction based on wavelet denoising and EMD. *Acta Metrologica Sinica*, 28(2): 158–162.
- Dai W, Ding X, Zhu J, 2006. EMD filter method and its application in GPS multipath. *Acta Geodaetica et Cartographica Sinica*, 35(11): 321–327.
- Damsø T, Kjær T, Christensen T B, 2016. Local climate action plans in climate change mitigation: Examining the case of Denmark. *Energy Policy* 89: 74–83. doi: <http://dx.doi.org/10.1016/j.enpol.2015.11.013>.
- Dawson I K, Vinceti B, Weber J C, 2011. Climate change and tree genetic resource management: maintaining and enhancing the productivity and value of smallholder tropical agroforestry landscapes: A review. *Agroforestry Systems*, 81(1): 67–78.
- Demir B, Ertürk S, 2010. Empirical mode decomposition of hyperspectral images for support vector machine classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11): 4071–4084.
- Gao R X, Yan X, 2011. Wavelets: Theory and Applications for Manufacturing. Springer Science and Business Media LLC. doi: 10.1007/978-1-4419-1545-0.
- Gauthier T, 2001. Detecting trends using Spearman's rank correlation, coefficient. *Environ Forensics*, 2: 359–362.
- Ghasemi N, Sahebi M R, Mohammadzadeh A, 2013. Biomass estimation of a temperate deciduous forest using wavelet analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2): 765–776.
- Gloersen P, Huang N E, 2003. Comparison of inter annual intrinsic modes in hemispheric sea ice covers and other geophysical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 41(5): 1062–1074.
- Gong Z, Kawamura K, Ishikawa N *et al.*, 2015. MODIS NDVI and vegetation phenology dynamics in the Inner Mongolia grassland. *Solid Earth Discussions*, 7(3): 2381–2411.
- Goshtasby A, 2012. Advances in Computer Vision and Pattern Recognition. Springer-Verlag London Limited. doi: [http://10.1007/978-1-4471-2458-0\\_2](http://10.1007/978-1-4471-2458-0_2).
- Grafakos L, Teschl G, 2013. On Fourier transforms of radial functions and distributions. *Journal of Fourier Analysis Applications*, 19: 167–179.
- Han F, Kang S, Buyantuev A *et al.*, 2016. Effects of climate change on primary production in the Inner Mongolia Plateau, China. *International Journal of Remote Sensing*, 37(23): 5551–5564. doi: 10.1080/01431161.2016.1230286.
- Han M, Liu Y H, Xi J H, 2007. Noise smoothing for nonlinear time series using wavelet soft threshold. *IEEE Signal Processing Letters*, 14(1): 62–65.
- He Y, Wen H, Yuan H, 2007. Multi-wavelet-based deformation monitoring signal processing. *Hydropower Automation and Dam Monitoring*, 31(1): 61–71.
- Huang D, Ding X, Chen Y, 2001. Wavelet filters based separation of GPS multipath effects and engineering structural vibrations. *Acta Geodaetica et Cartographica Sinica*, 30(1): 36–41.
- Huang N E, Wu M C, Long S R *et al.*, 2003. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *The Royal Society*, 459(2037): 2317–2345. doi: 10.1098/rspa.2003.1123.
- Huang N E, Wu M L, Qu W *et al.*, 2003. Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 19: 245–268. doi: 10.1002/asmb.501.
- Huang S, Li P, Yang B, 2005. Study on the characteristics of multipath effects in GPS dynamic deformation



- monitoring. *Geomatics and Information Science of Wuhan University*, 30(10): 877–879.
- Jain N, Srivastava V, 2013. Data mining techniques: A survey paper. *International Journal of Research in Engineering and Technology*, 2(11): 2319–1163.
- Jordan Y C, Ghulam A, Chu M L, 2014. Assessing the impacts of future urban development patterns and climate changes on total suspended sediment loading in surface waters using geoinformatics. *Journal of Environmental Informatics*, 24(2): 65–79.
- Kennedy M, Basu B, 2014. An analysis of the climate change architecture. *Renewable and Sustainable Energy Reviews*, 34: 185–193.
- Koenker R, 2005. Quantile Regression. Cambridge University Press.
- Kotlyar M, Fuhrman S, Ableson A *et al.*, 2002. Spearman correlation identifies statistically significant gene expression clusters in spinal cord development and injury. *Neurochemical Research*, 27(10): 1133–1140.
- Kysely J, Beguería S, Beranová R *et al.*, 2012. Different patterns of climate change scenarios for short-term and multi-day precipitation extremes in the Mediterranean. *Global and Planetary Change*, 98: 63–72.
- Lhermitte S, Verbesselt J, Verstraeten W *et al.*, 2011. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sensing of Environment*, 115: 3129–3152.
- Li J, Cui Y, Liu J *et al.*, 2013. Estimation and analysis of net primary productivity by integrating MODIS remote sensing data with a light use efficiency model. *Ecological Modelling*, 252: 3–10. doi: 10.1016/j.ecolmodel.2012.11.026.
- Li S, Xie Y, 2013. Investigating coupled impacts of climate change and socioeconomic transformation on desertification by using multi-temporal landsat images: A case study in Central Xilingol, China. *IEEE Geoscience and Remote Sensing Letters*, 10(5): 1244–1248. doi: 10.1109/LGRS.2013.2257158.
- Li S, Xie Y, Brown D *et al.*, 2013. Spatial variability of the adaptation of grassland vegetation to climatic change in Inner Mongolia of China. *Applied Geography*, 43: 1–12. doi: 10.1016/j.apgeog.2013.05.008.
- Li Z, Zhu Q, Gold C, 2005. Digital Terrain Modeling: Principles and Methodology. CRC Press.
- Liu L, Liu C, Jiang C, 2007. Novel EMD algorithm and its application. *Journal of System Simulation*, 19(2): 446–447.
- Lu N, Wilske B, Ni J *et al.*, 2009. Climate change in Inner Mongolia from 1955 to 2005: Trends at regional, biome and local scales. *Environmental Research Letters*, 4: 045006. doi: 10.1088/1748-9326/4/4/045006.
- Luan Y, Fan Y, Xue L, 2004. Under ground space study on prediction model of trend. *Term for Ground Surface Movement*, 24(1): 14–18.
- Mishra A, Singh R, Raghuwanshi N S *et al.*, 2013. Spatial variability of climate change impacts on yield of rice and wheat in the Indian Ganga Basin. *Science of the Total Environment*, 468: S132–S138.
- Nunes J C, Gu Y S, Delechelle E, 2005. Texture analysis based on local analysis of the bidimensional empirical mode decomposition. *Machine Vision and Applications*, 16(3): 177–188.
- Pearson, K. 1895. Contributions to the mathematical theory of evolution (III): Regression, heredity, and panmixia. *Proceeding of the Royal Society of London*, 59(353–358): 67–71.
- Peng Z K, Tse P W, Chu F L, 2005. A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Original Research Article Mechanical Systems and Signal Processing*, 19(5): 974–988.
- Piras M, Mascaro G, Deidda R *et al.*, 2016. Impacts of climate change on precipitation and discharge extremes through the use of statistical downscaling approaches in a Mediterranean basin. *Science of the Total Environment*, 543: 952–964.
- Portilla J, Strela V, Wainwright M J *et al.*, 2003. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transaction on Image Processing*, 12(11): 1338–1351.
- Rahmani M A, Zarghami M, 2015. The use of statistical weather generator, hybrid data driven and system dynamics models for water resources management under climate change. *Journal of Environmental Informatics*, 25(1): 23–35.
- Rao K S, Hsu C, 2008. Antenna system supporting multiple frequency bands and multiple beams. *Antennas and Propagation*, 56(10): 3327–3329.



- Ribalaygua J, Pino M, Pórtolés J *et al.*, 2013. Climate change scenarios for temperature and precipitation in Aragón (Spain). *Science of the Total Environment*, 463: 1015–1030.
- Sherbinin de A, Castro M, Gemenne F, 2011. Preparing for resettlement associated with climate change. *Science*, 334(6055): 456–457.
- Spearman C, 1904. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1): 72–101.
- Swain D K, Thomas D, 2010. Climate change impact assessment and evaluation of agro-adaptation measures for rice production in eastern India. *Journal of Environmental Informatics*, 16(2): 94–101.
- Veerasamy R, Rajak H, Jain A *et al.*, 2011. Validation of QSAR models: Strategies and importance. *International Journal of Drug Design and Discovery*, 2(3): 511–519.
- Wang J, Brown D, Agrawal A, 2013. Climate adaptation, local institutions, and rural livelihoods: A comparative study of herder communities in Mongolia and Inner Mongolia, China. *Global Environmental Change*, 23(6): 1673–1683.
- Wang Z, Fang J, Tang Z *et al.*, 2012. Relative role of contemporary environment versus history in shaping diversity patterns of China's woody plants. *Ecography*, 35(12): 1124–1133. doi: 10.1111/j.1600-0587.2011.06781.x.
- Xia X H, Wu Q, Mou X L *et al.*, 2015. Potential impacts of climate change on the water quality of different water bodies. *Journal of Environmental Informatics*, 25(2): 85–98.
- Xian X, Lin Z S, Cheng X X *et al.*, 2008. Regional features of the temperature trend in China based on empirical mode decomposition. *Journal of Geographical Sciences*, 18(2): 166–176.
- Xie Y, Crary D, Bai Y *et al.*, 2017. Modelling grassland ecosystem responses to coupled climate and socioeconomic influences from multi-spatial-and-temporal scales. *Journal of Environmental Informatics*, 1684–8799. doi: 10.3808/jei.201600337.
- Xie Y, Sha Z, Yu M *et al.*, 2009. A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China. *Ecological Modelling*, 220: 1810–1818.
- Yuan X, Li L, Chen X *et al.*, 2015. Effects of precipitation intensity and temperature on NDVI-based grass change over northern China during the period from 1982 to 2011. *Remote Sensing*, 7(8): 10164–10183. doi: 10.3390/rs70810164.
- Yue S, Pilon P, Cavadias G, 2002. Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, 259(1): 254–271.
- Zastrow M, 2015. Data visualization: Science on the map: Easy-to-use mapping tools give researchers the power to create beautiful visualizations of geographic data. *Nature*, 519: 119–120.
- Zhang A B, Chen T Y, Liu X X *et al.*, 2015. Monitoring data filter and deformation information extraction based on wavelet filter and empirical mode decomposition. *Applied Mechanics and Materials*, 742: 261–271.
- Zhang A B, Gao J X, Zhang Z J, 2011. Deformation analysis and prediction of building above old mine goaf based on multiscale method. *Rock and Soil Mechanics*, 32(8): 2423–2428.
- Zheng T, Yang L, 2007. Discussion and improvement on empirical mode decomposition algorithm. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 46(1): 1–6.
- Zheng Z, 2001. Wavelet Transformation and the Application of Its MATLAB Tools. Beijing: Earthquake Press.
- Zhong P, Ding X, Zheng D, 2006. Separation of structural vibrations and GPS multipath signals using Vondrak filter. *Journal of Central South University of Technology*, 37(6): 1189–1195.