Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

Comparison of Time and Spatial Scales in Global Temperature Data

Ming Luo[1], Igor Zurbenko[2]
[1,2] Department of Epidemiology and Biostatistics, State University of New York at Albany, One University Place, Room 131, Rensselaer, NY 12144-3456

Abstract
To capture the important spatio-temporal patterns in a global temperature dataset, we applied correlation analysis on recovered signals based on a nonparametric smoothing algorithm, the Kolmogorov-Zurbenko Spline (KZS). In our study, KZS was applied on monthly average temperature observations from thousand of weather stations over the global to generate smoothed monthly temperature deviations from normal levels over the time period of 1893 to 2008. These signals, which were separated out on different time and spatial scales, can be viewed as global long term climate trend and El Nino-like movement. Then the correlation relationships of the reconstructed signals were studied with bootstrapping. We found this method was decisive for understanding the key features of the data, and it revealed an interesting corresponding relation between time and geographical distance. A movie for global maps of long term temperature trend has been developed as well as El Nino scales over the global, and both provide striking revelations in global temperature anomalies for specific regions.

Key Words: non-parametric method, smoothing, temporal-spatial data, correlation analysis, El Nino-like movement, global long-term trend

1. Introduction

The general feature of the global climate variability is an important topic for climate researchers. In this paper, we are interested in revealing and displaying the basic change pattern of climate movement at the global level. From the statistical prospective, this task is a typical space-time modeling problem, and the major challenge lies in the complex spatio-temporal dependencies over multiple time and spatial scales [1, 2].

First, the presence of various scales of motion in time complicates the analysis and interpretation of the data. In time series studies, the basic strategy for this kind of problem is to decompose temporal data into components that we are interested in [3]. However, in our study, the weather observation series are spatially correlated and geographically distributed over the globe. Before decomposing time series data, we need deal with the spatial variations and separate out the spatial components first. Our paper will describe how to utilize linear regression equation to remove major spatial variations in the global temperature data.

On the other hand, the general profile of global climate movement should not be simply treated as panel data analysis or multiple time series problem. It involved assimilating, gridding, smoothing and interpolating surface climate records to form a unified temperature field covering the land area of the Earth [4]. Considering the number of weather stations used in the study, the statistical method used here should be simple, and

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

therefore, highly efficient. Otherwise, the computation could be very time-consuming. Our paper will introduced Kolmogorov-Zurbenko Filter (KZ) [5] as the smoothing and interpolating model for both time and space dimensions. The algorithm of KZ Filter is based on averaging data in a moving window that can be easily extended to multi-dimensions. This means that it is a local nonparametric smoothing method and therefore is more appropriate for meteorological data compared to other surface-fitting procedures like thin-plate spline [6]. The simplicity, flexibility and extensibility of KZ filter make it an efficient and robust tool for spatial smoothing of global temperature data.

To decompose time series, spectral analysis will be utilized to understand the time series structure [7]. In our case, there are over thousands of weather observation series involved. Therefore, the issue is how to find the common frequency components in these signals, and then decompose the time series data based on their common structure. Here, to simplify the problem, we proposal to exam the average spectrum distribution of global temperature series to identify various scales of motion for temperature varieties. The bootstrapping scheme of generating average periodogram will be illustrated in the paper. The advantage and limitation of this approach will also be carefully discussed.

Based on the result of spectral analysis, we can identify the long-term component and El Niño-like movement of surface climate over global land areas for the period of 1880 to 2008. We will extract those components utilizing KZ Filters' powerful capability to reconstruct high resolution signals buried in high background noises [5]. In the later part of the paper, we will describe how to select the KZ parameters to meet this purpose.

KZ Filters' outputs enable us to visualize spatial temperature change over time with movies of thermal maps. We will show how the long-term trend movie facilitates capturing the spatial energy input-consuming patterns over the globe, and how the El Niño-like movement movie helps to understand the short term temperature fluctuation and its spatial feature. The correlation analysis over these global climate components revealed striking spatial/temporal correlation patterns and will also be discussed.

The technologies and methods discussed in our paper is the basic part of the framework for a new approach to investigate the climate changes on the Earth. We believe that this approach is more spatial-temporal statistics oriented compared with previous "reanalysis" works [4, 6, 8-9]. This will provide us a solid base and can serve for a number of applications in the future.

2. Method

2.1 Data Source and Preparation

Our data source is the monthly mean temperature data from the Global Historical Climate Network Dataset (GHCN) version 3 [10]. Considering the data coverage, the study period is from 1880 to 2011.

There are more than seven thousand stations in GHCN monthly mean temperature data set. However, these stations are not evenly distributed (Figure 1). To avoid inadequate spatial sampling, we have to aggregate these station records to the grid of 1°×1° latitude-longitude resolution by averaging all the temperature records in each grid cell. Otherwise, it may enlarge the civilization bias [10] if we use the station data directly. This gridded dataset covers most land surface except Antarctic (Figure 2).

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

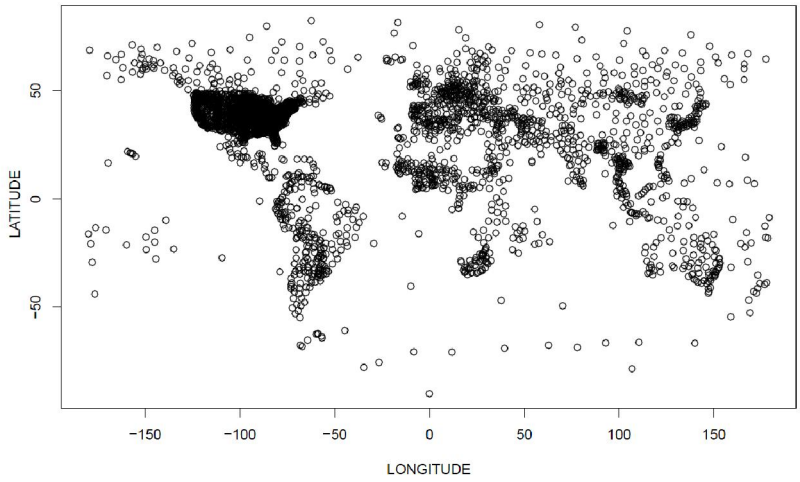


Figure 1: The Distribution of the GHCN Weather Stations

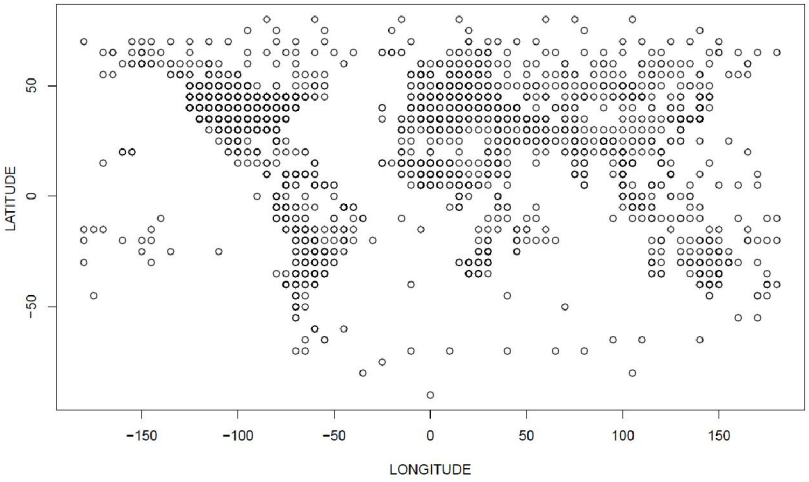


Figure 2: The Distribution of the Gridded Climate Data Cells

2.2 The Regression Model for Spatial Variation

It has been noticed that long-term monthly temperature is majorly affected by latitude and altitude [12]. These effects are commonly known as latitude pattern and mountain effect. As a matter of fact, they are the major spatial components of global temperature data. However, we are not interested in these normal patterns in space. We are more care of spatial anomaly.

The major spatial variations of the global temperature can be count for by connecting monthly temperature records with geographic factors through regression equation. The information of station location (latitude, longitude) and altitude are available in GHCN data set; the altitude for each grid point can be taken as the average altitude for stations in the cell (See Figure 3). Equation (1) is used to address these major spatial variations of global temperature associated with latitude, altitude and their interactions (Figure 4).

$$T(y) = a_0 + a_1 cos^2(y) + a_2 l + a_3 sin(y) + a_4 S + a_5 l{:}sin(y) + a_6 l{:}S + a_7 l{:}E{:}cos^2(y) \quad (1)$$

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

where T(y) is the sea-level mean temperature on latitude y, l is the variable for altitude. S and E is the dummy for Antarctic (y < 70°) and equator (-10° < y < 10°), respectively. Symbol ":" represents interaction between variables.

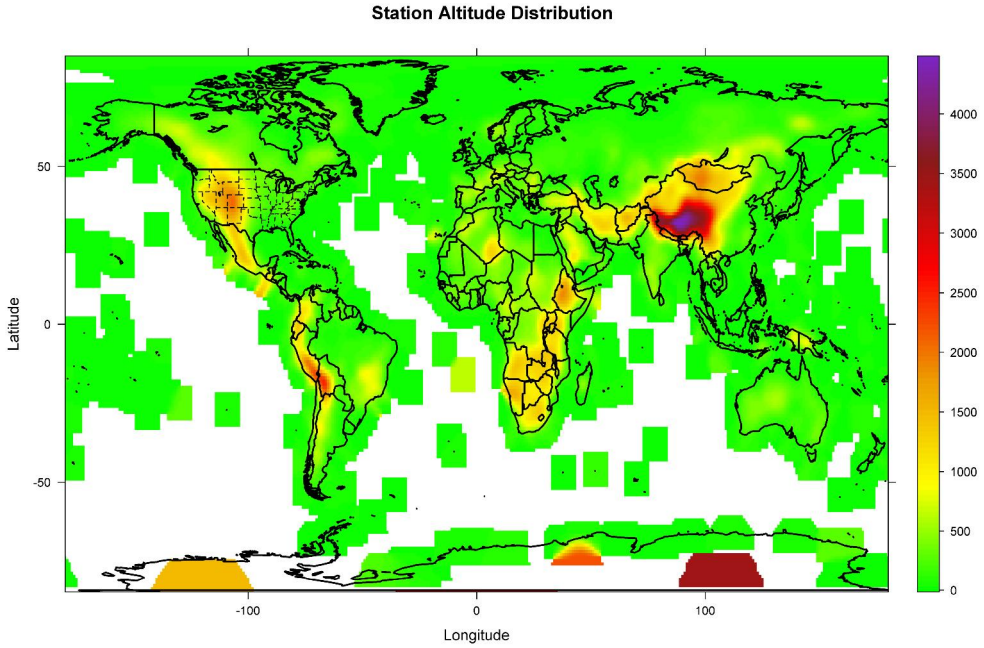


Figure 3: The Spatial Smoothed Distribution of Station Altitude

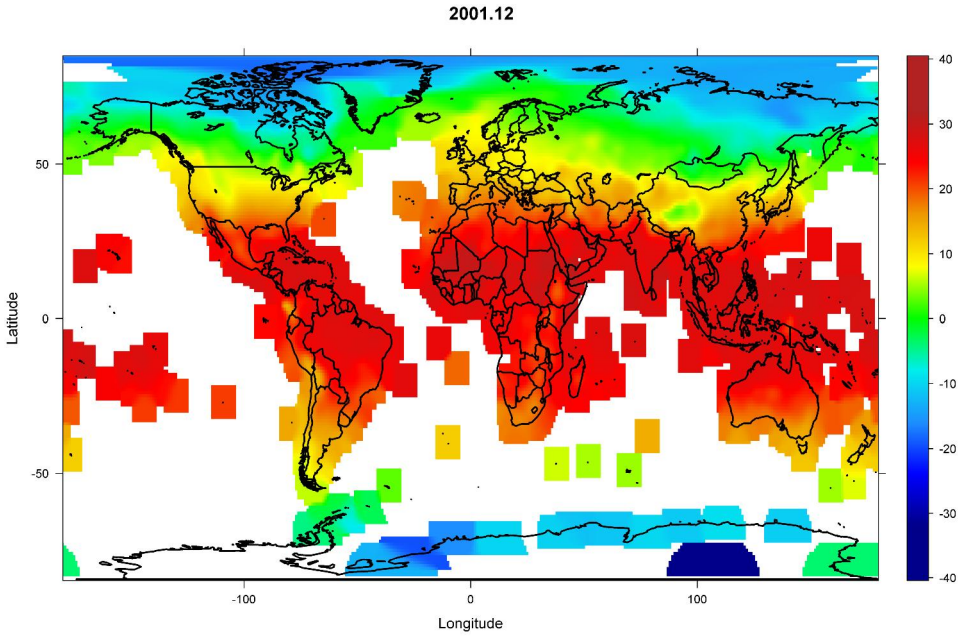


Figure 4: Latitude Pattern and Altitude Effect of Monthly Temperature

Every item in equation (1) can be justified physically: coefficient a_1 is for the cosine-square latitude pattern of temperature [12]; $-a_2$ is the average lapse rate[1] of mountain

[1] It is the rate at which temperature declines with altitude near the ground surface.

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

effect; a_3 is used to address the temperature difference for same latitude area in north and south hemisphere[2]; a_4 is the extra adjustment because temperatures in Antarctic are lower than the prediction of the cosine-square law [13]; while $-a_5$, $-a_6$ and $-a_7$ is the lapse rate adjustments for south hemisphere, Antarctic and equator region, respectively.

It is worth to point out that the regression is performed on latitudes-altitudes grid (1° × 50 m), instead of stations, to avoid underestimation of the lapse rate [14] caused by the extremely uneven distribution of the station observations on different altitude levels. This latitudes-altitudes grid data is generated by aggregating the latitudes-longitude gridded data in pervious section.

The estimated coefficients (a_1 to a_7) will be used to adjust the station observations to produce the spatial anomaly that we are looking for. Technically, the residuals of the regression equation represent the spatial anomaly that may be caused by some other factors not in the equation - for example, the ocean upward wind, the density of water vapour in the air, or the cloudiness in some area [13].

2.3 Spectral Analysis

The average spectral feature of time series for stations over the global can be identified with bootstrapping on KZ periodogram (KZP) [5]:

1) Randomly select 3000 stations over the globe, and for all stations with 50+ year records, calculate 5% DZ smoothed KZ periodogram over (0, 0.06);
2) Evenly divide (0, 0.06) as 133 bins, calculate the mean values of smoothed KZ periodogram on the 133 bins as the global raw periodogram;
3) Smooth the global raw periodogram with 5% DZ smoothing parameter.

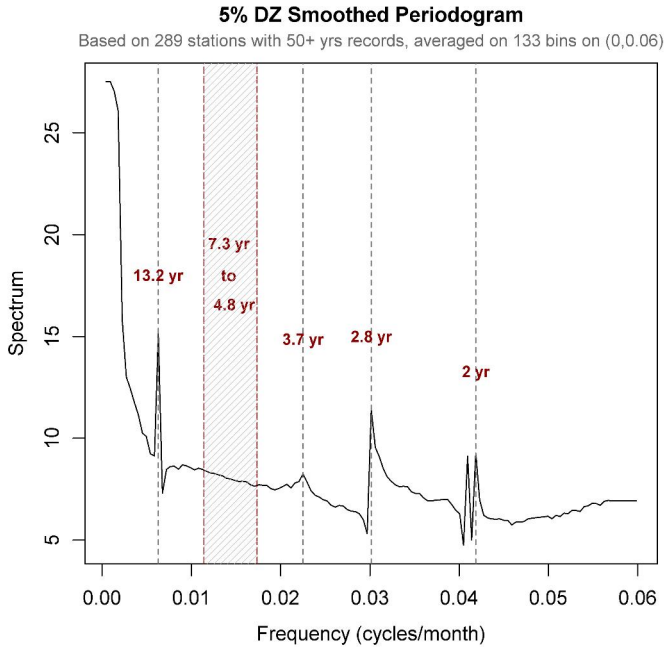


Figure 5: Average Spectrum Distribution for Stations over the Globe

[2] We used *sin*(*y*) instead of dummy variable to avoid sudden temperature change when crossing the equator. Same for the lapse rate difference of north and south hemisphere.

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

The final periodogram is the average spectral distribution for most stations on the Earth (Figure 5). It is consistent with the spectral analysis of the 350-year long historical Central England temperature (CET) time series [12]. Based on this result, we can identify the components that we are interested: global long-term trend (longer than 6 to 7 years) and El Niño-like temperature variation (1.5 to 5 years).

2.4 Temporal Decomposition of the Data

In the following sections, we use the notion **KZ**(m, k) refer KZ filters that iteratively applying moving average of m points k times. Then the transfer function of **KZ**(m, k) is:

$$|\psi_{m,k}(\lambda)| = \left\{\frac{1}{m}\frac{\sin(m\lambda/2)}{\sin(\lambda/2)}\right\}^k$$

The cut-off frequency of **KZ**(m, k) is:

$$\lambda_0 \approx 2\sqrt{6}\sqrt{\frac{1-\alpha^{1/2k}}{m^2-\alpha^{1/2k}}}, \quad \alpha \in (0,1)$$

In our case, m is the time steps in month, and α is usually 0.5. When design the KZ filter system to separate components with different scales, the area with shadow in Figure 5 can be used as the cut-off separation range. The purpose is to use it as a "buffer zone" to prevent the mixing of long-term and El Niño-like movement.

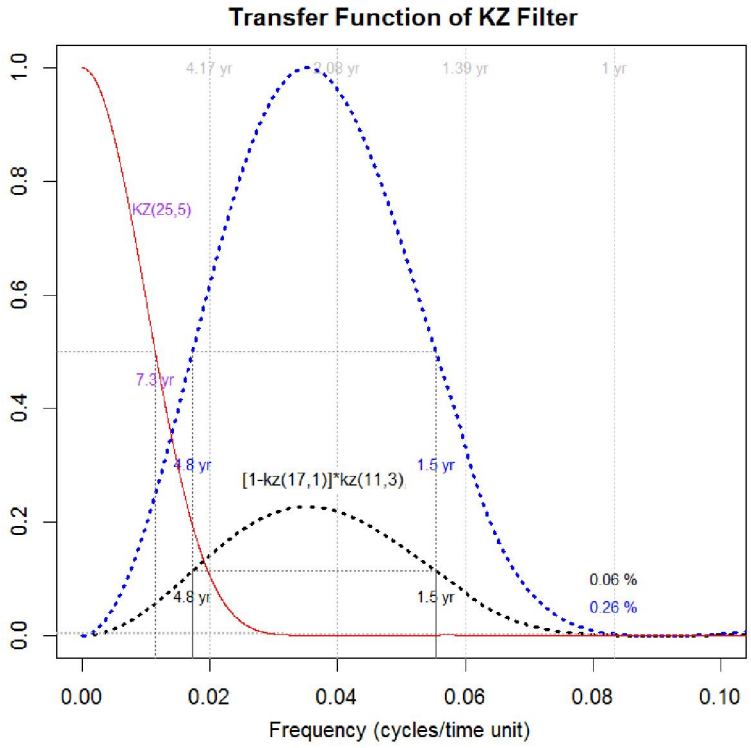


Figure 6: Using Transfer Function to Select KZ System Parameters

As we can see in Figure 6, the transfer function of **KZ**(25, 5) will pass signals with 7.3+ years cycle. This low-pass feature makes it perfect for separating the long-term trend. The notion of **KZ**(25, 5) means that this KZ is the moving average of 25 months of the temperature records and iterate for 5 times. Its result will be our long-term component.

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

Also in Figure 6, the band-pass filter [**1-KZ**(17,1)]***KZ**(11,3) has cut-off at 4.8 and 1.5 years, and it works well for the El Nino-like component. To implement this KZ filter, one need to apply moving average on the window of 17 months for one time, then use the original monthly data minus the result; after this, average on moving window of 11 months and iterate for 3 times. The result is needed to be enlarged by some scale coefficient (about 4) to counterbalance the amplitude attenuation.

After filtering out long-term component and El Niño-like component, we smoothed them with KZ filter **KZ**((3°, 3°), 5). This will interpolate the series data of cell points and extend them into the space. It enables us to generate the movies for their thermal plots evaluated over time. The movie was implemented with slides generated by R lattice package on global map, aimed to visualize climate change over time and space, and facilitate capturing important events of global climate change.

2.5 Correlation Analysis

We applied correlation analysis on the generated long-term component and El Niño-like component with a re-sampling scheme:

(1) For each pair of randomly selected grid points, the correlation coefficient for their time series data over common time period was calculated;
(2) Draw scatter-plot for distance-correlation relationship based on more than 10,000 samples; same for angle-correlation relationship;
(3) Applied KZ filter on distance-correlation data with parameter m = 300 and k = 3; plotted the relationship;
(4) Applied two-dimensional KZ filter on distance, angle-correlation data with parameter m = (300, 5) and k = 3; plotted their relationship with thermal plot.

Since we had removed most of the important spatial variations for both components, we expected the correlation should be essentially the same on each direction. This procedure was designed to check the isotropic assumption as well as the assumptions that used to choose the spatial smoothing parameters in the previous section.

3. Results

3.1 Correlation Patterns

We checked the spectral structure of both components as a verification of our design. Figure 7 shows the typical spectral distribution for both components. Obviously, the KZ filters separated the two components as desired, and there is no energy leaking around annual frequency (0.08333).

The Figure 8 exhibits the spatial dependency of El Niño-like component. The average correlation coefficients are greater than 0.5 when the distances of two location pairs are less than 1180 km. When the distance increases to 1850 km, the average correlation is still above 0.25. The result is consistent with Hansen's work about temperature anomaly [4], and is much larger than the support range of KZ filters used for spatial smoothing. It suggests that the dependency comes from the component itself instead of KZ filters.

Figure 9 gives the distribution of correlations on both the distances and directions for the sampled pair of grid points. This thermal plot demonstrates that the spatial correlation of El Niño-like movement is not related to the direction, especially in some distance ranges

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

(e.g., less than 2,500 km). Therefore, it is consistent with the marginal result shown in Figure 8, and it supports our model assumption.

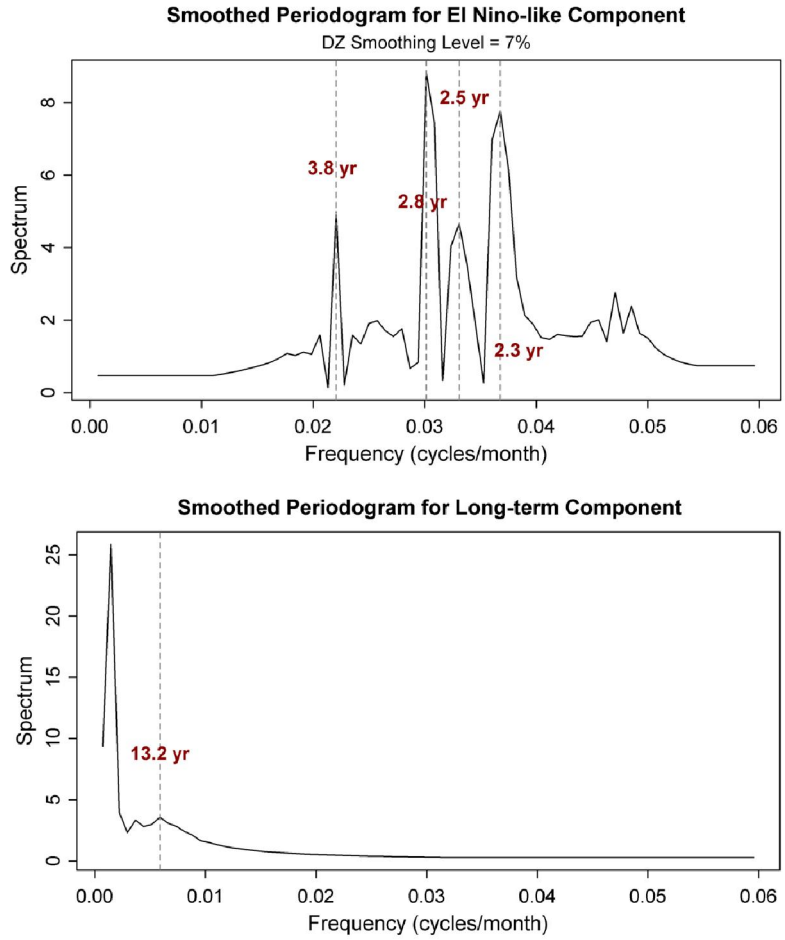
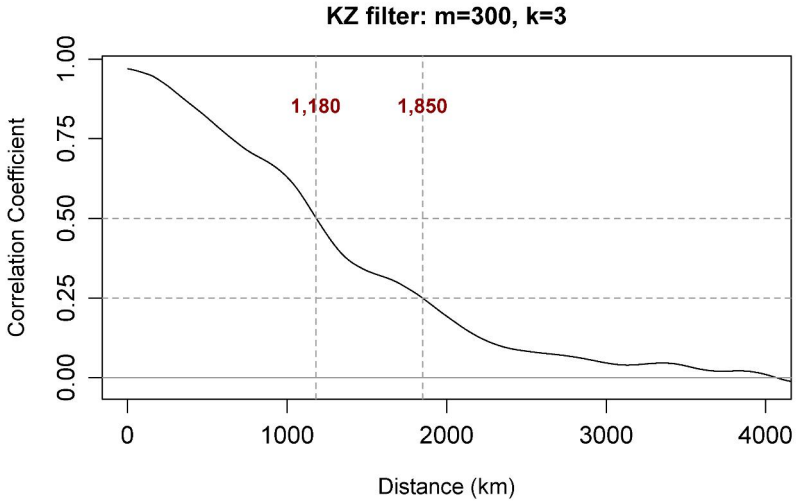


Figure 7: The Typical Spectral Distribution for Reconstructed Components



Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

Figure 8: Average Spatial Correlation for El Niño Component

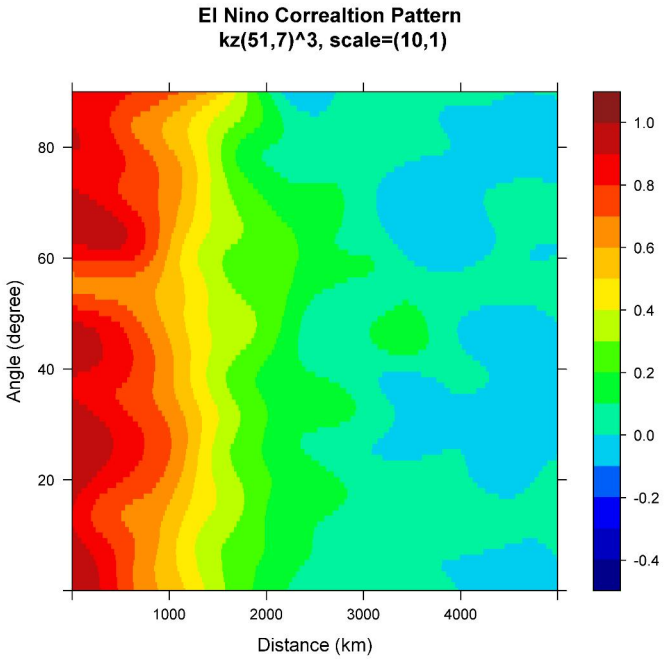


Figure 9: Spatial Correlation for El Niño-like Component in Two Directions

3.2. Global Long Term Component and El Nino-like Component

Figure 10 exhibits the global long-term component after the latitude pattern and altitude effects are adjusted. Please note the regions on the North Atlantic Ocean near the Northern European and Iceland - its long-term mean temperature is 5°C to 10°C higher than other areas' on the same latitude. The second notable area is along 23.5° latitude with red color, from north-western India to Saudi Arabia and Sarhra Desert — for most places in this region, their average temperatures are more than 5°C higher compared with other areas' on the same latitude.

The global long-term component also changed with time. For most areas of the world, climate in early 1970s is cooler than climate of 2000s. We can see this in the long-term component movie.

El Nino-like oscillations has smaller magnitude than the global long-term trend, but it changes much faster. The spatial patterns of El Niño-like signal are also more complex and keep changing with time. As an example, Figure 11 shows that the serious drought in 2001 to 2002. In the Eastern USA, the serious drought could be connected to the unusual high temperatures in the winter of 2001 to 2002; while in the Western USA, the drought could be explained by the increased long-term temperature in this region (Figure 10).

4. Discussion

We had described the methodology to separate the global long term temperature trend and El Niño-like oscillations from a seasonal changed raw temperature dataset. Specially designed KZ filters help us reconstruct the desired signal from background with high

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

noise. The setting of KZ filter parameters has been confirmed by the results of spectral analysis and correlation analysis. We also believe that our method to separate spatial components is reasonable and had been supported by the data.

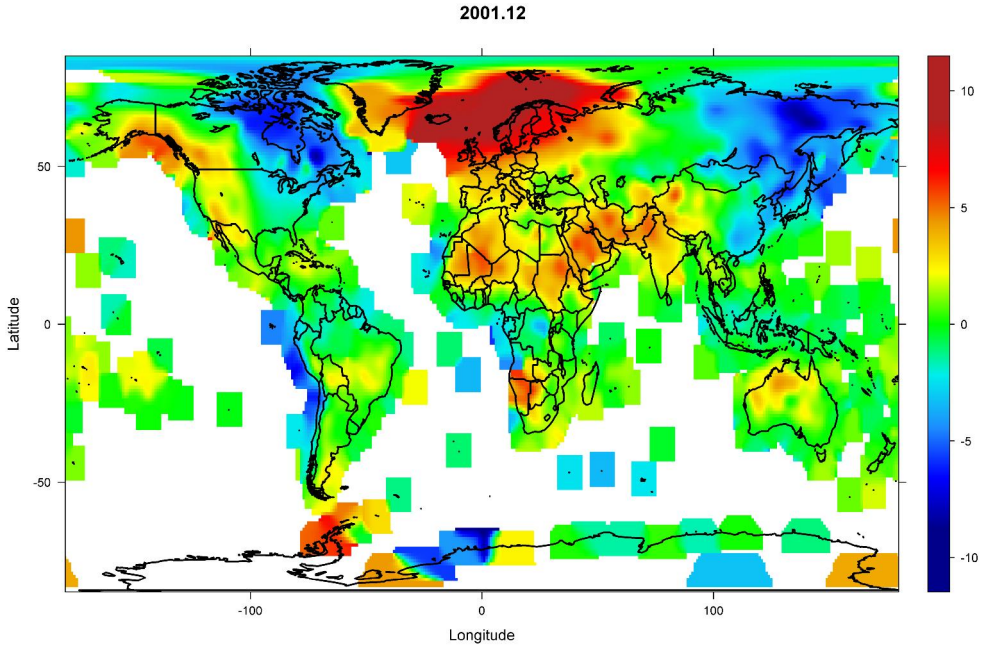


Figure 10: Global Long-term Component on Dec., 2001.

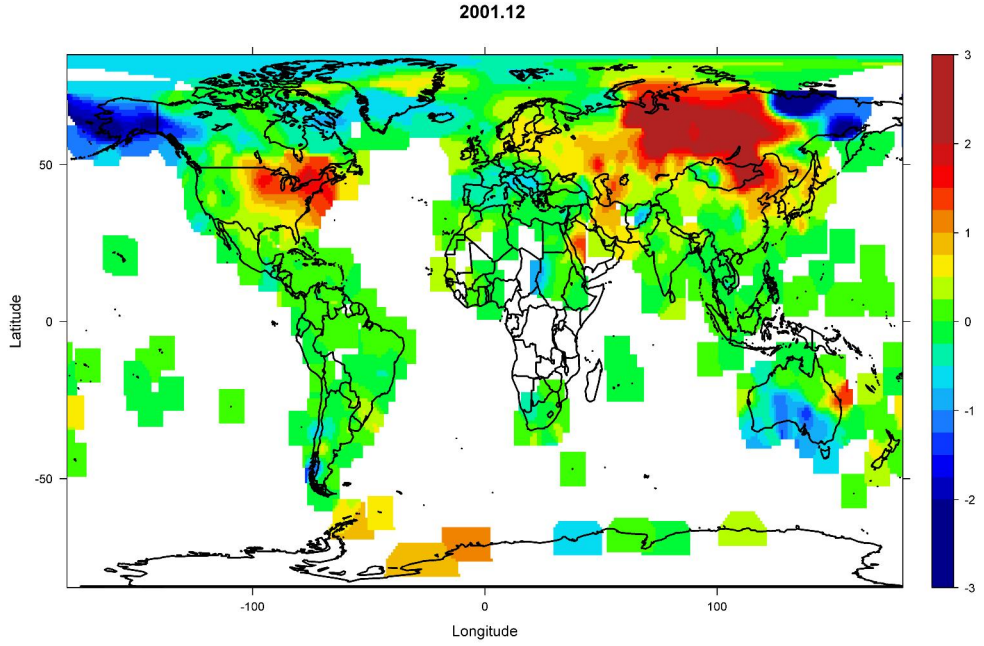


Figure 11: El Niño-like Component on Dec., 2001.

Generally speaking, the global long term component evaluates very slowly along the time line. What interested us is the spatial pattern of the long term mean temperature, and it can be explained from the energy balance perspective. Usually grid cells with red color in long term movie means more energy input in related regions. For example, Northern

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

European area is warm due to the heat that the North Atlantic Current absorbed from southern ocean and carried to this area; so do the pacific coast of Canada and Alaska. Mean while, the surplus energy for Sahara and Arizona area come from the extra solar irradiation associated with extreme low humidity and precipitation, low cloud coverage, or high elevations and other local conditions. The long-term trend maps could be a good tool to predict the climate changes for related regions.

Spatial construction of filters provides outcomes from different data support on distance more then 500km (smoothing scale of the filter). Our spatial images are very well organized on much larger scales corresponding to the support regions for El Niño-like and long term scales we are investigating. Visual El Niño-like scales are of order more than 1,000 km to 2,000 km. We had observed temperature correlations on scale larger than our smoothing scale. Evidence from spectral analysis, correlation analysis and images displayed in our paper clearly support of El Nino-like fluctuations in atmospheric temperature over entire globe.

The average spectrum distribution provides us important information about the common frequency range for most station series over the globe. It simplified the distributed parameter problem by taking average. This provides us a base to separate the global temperature movement as long-term and El Nino-like component. But we must realize that it is a trade-off between efficiency and accuracy. The result show here is for the most common situations. However, the advantage of this method is lies in the movies of the changes. Visualization of climate components may provide us new inspiration to look into climate change study.

References

[1] Finkenstadt B, Leonhard H, Valerie I, Statistical Methods for Spatio-Temporal Systems, Chapman & Hall/CRC, Boca Raton, 2007.

[2] Porcu E, Montero JM, Schlather M, Advances and Challenges in Space-time Modelling of Natural Events, Springer-Verlag Berlin Heidelberg, 2012.

[3] Rao ST, Zurbenko IG, Neagu R, Porter PS, Ku JY, Henry RF, Space and Time Scales in Ambient Ozone Data, Bulletin of the American Meteorological Society, Vol. 78, No. 10, 1997, pp. 2153-2166.

[4] Hansen J, Ruedy R, Sato M, and Lo K, 2010: Global Surface Temperature Change, Reviews of Geophysics, Vol. 48, No. RG4004, 2010, doi:10.1029/2010RG000345.

[5] Yang W and Zurbenko IG, Kolmogorov–Zurbenko Filters, Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 2, No. 3, 2010, pp. 340–351.

[6] New MG, Hulme M, Jones PD, Representing Twentieth-Century Space-Time Climate Variability, Part II: Development of 1901–96 Monthly Grids of Terrestrial Surface Climate. J Clim 13:2217-2238.

[7] Tsakiri KG and Zurbenko IG, Prediction of Ozone Concentrations using Atmospheric Variables, Air Quality, Atmosphere & Health, Vol. 4, No. 2, 2011, pp. 111-120.

Generated by Foxit PDF Creator @ Foxit Software
http://www.foxitsoftware.com For evaluation only.

[8] Caesar JL and Vose R, Large Scale Changes in Observed Daily Maximum and Minimum Temperatures: Creation and Analysis of a New Gridded Data Set, Journal of Geophysical Research, Vol. 111, No. D05101, 2006, doi:10.1029/2005JD006280.

[9] Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvi A, Very High Resolution Interpolated Climate Surfaces for Global Land Areas, International Journal of Climatology, Vol. 25, No. 15, 2005, pp. 1965–1978.

[10] Lawrimore JH, Menne MJ, Gleason BE, Wil-liams CN, Wuertz DB, Vose RS and Rennie J, An Overview of the Global Historical Climatology Network Monthly Mean Temperature Data Set, Version 3, Journal of Geophysical Research, Vol. 116, No. D19121, 2011, 18 p. doi:10.1029/2011JD016187.

[11] Karl TR, Diaz HF, and Kukla G, Urbanization: Its Detection and Effect in the United States Climate Record, Journal of Climate, Vol. 1, No. 11, 1988, pp. 1099–1123.

[12] Zurbenko IG, Cyr DD, Climate Fluctuations in Time and Space, Climate Research, Vol. 46, No. 2, 2011, pp. 67-76.

[13] Linacre ET, Climate Data and Resources: A Reference and Guide, Routledge, London, 1992.

[14] Dodson R, Marks D, Daily Air Temperature Interpolated At High Spatial Resolution over a Large Mountainous Region, Climate Research, Vol. 8, No.1, 1997, pp. 1-20.