Spatio-Temporal Analysis Frequency Separation and Block Bootstraps of Periodically Correlated Time Series

Edward Valachovic[1]
[1]Department of Epidemiology and Biostatistics, School of Public Health, University at Albany, State University of New York, One University Place, Rensselaer, NY 12144

Abstract
This study investigates the benefits of spatio-temporal analysis frequency separation (STAFS) strategies prior to bootstrapping periodically correlated time series. Analysis of time series data that contains periodically correlated (PC) principal components, such as seasonal, daily, or other cyclostationary processes, benefits from separating these components from other interfering frequencies or PC components. Bootstrapping allows estimation of a statistic's sampling distribution using random sampling with replacement, while block bootstrapping is a model-free resampling strategy for time series data and extensions of the block bootstrap help preserve the correlation structures of PC processes. Frequency separated periodic block bootstrapping (FSPBB) is introduced to separate different periodic components by frequency and bootstrap the PC components to improve estimation of the periodic characteristics of those component processes as well as the full time series. Data simulation studies are used to compare block bootstrapping results of periodically correlated time series with and without prior frequency separation.

Key Words: Spatio – Temporal Analysis Frequency Separation, Frequency Separated Periodic Block Bootstrap, Periodically Correlated Time Series, Cyclostationary Process Frequency Interference, Seasonal Mean, Kolmogorov-Zurbenko Iterated Moving Average

1. Introduction

1.1 Background

To better understand the characteristics of a set of data and the population that the set of data represents, bootstrapping, or resampling with replacement from a given set of data, is useful to estimate the sampling distribution of statistics such as means and variances. Random sampling with replacement results in independent draws that form a resample of the same length as the original data set first detailed by Efron (1979). In time series data, successive data points, or observations, are ordered in time and more on time series including definitions, notation, and examples can be found in Wei (1990). The ordering can be in space, space-time, or another combination of dimensions, generally described as spatio-temporal data, but without loss of generality the data can still be referred to as time series and the metric that data is ordered in can be time. It is often the case that the ordered data points may be correlated with prior observations in the time series. Consequently, independently sampling data points from the dataset to form a new ordered resample will not replicate correlations between a given and prior data points. The correlation between successive time series data points is destroyed. Many bootstrapping schemes destroy correlation structures between in time series data.

Block bootstrapping methods attempt to replicate and preserve correlations in spatio-temporal data. Block bootstrapping often involves a general strategy of splitting the time series into blocks, and then randomly sampling the blocks to form the resamples. An example of this is the Moving Block Bootstrap first introduced by Kunsch (1989). This can help replicate the correlation structure between given data points and, at least up to a limited number, some prior observations. Clearly, correlations between data points that are separated by more than the length of the bootstrapped blocks will have correlation structures destroyed.

Many fields that relate to natural or physical processes involve spatio-temporal data that is driven by periodic factors or components. Periodic components of a given period, p, exhibit a unique type of correlation structure between observations in spatio-temporal data. Such a component of a time series is periodically correlated (PC). A PC time series with a period p has strong correlation between data points that are p time points removed. In general, block bootstrapping has difficulty reproducing the correlation structure of PC time series. First, arbitrarily bootstrapping with block lengths less than the given period, p, will sever these correlations. Additionally, bootstrapping with block lengths greater than the given period, p, will preserve the correlation within each block, however there is no reason that randomly sampled adjacent blocks will retain the periodic correlation structure when forming the resample. With this bootstrap design, there is nothing to force the first data point of a sampled block to be the next step in each cycle of period p that follows the last data point of the prior sampled block. This is illustrated in the following figure. For this reason, many block bootstrapping strategies are not well suited for PC time series.

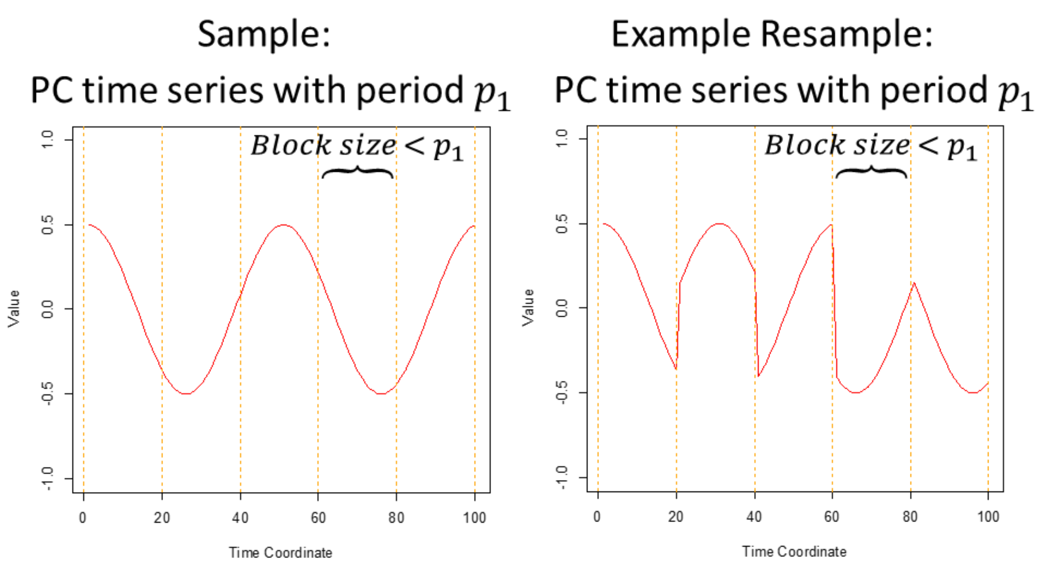

Figure 1: Illustration of resampling a PC time series component of period p_1 with a block boostrap of block length less than p_1 and destroying the PC structure.

1.2 Existing Methods and Challenge

A block bootstrapping strategy that is better suited for PC time series with a PC component of period, p, is one that considers bootstrapping blocks whose length in some way accounts for the period p. A seasonal block bootstrap was proposed by Politis (2001). There, the block length is restricted to be a multiple of the period p. Regardless of whichever step in the cycle a block begins with, all other blocks in the resample will likewise start and end with a common step in the cycle of period p. This strategy will preserve correlation structures between data points that are p points removed from each other in time series data. This can be seen in the following figure. Some other block bootstrapping methods for PC time series include that of Chan et al. (2004) as well as that presented by Dudek et al. (2014).

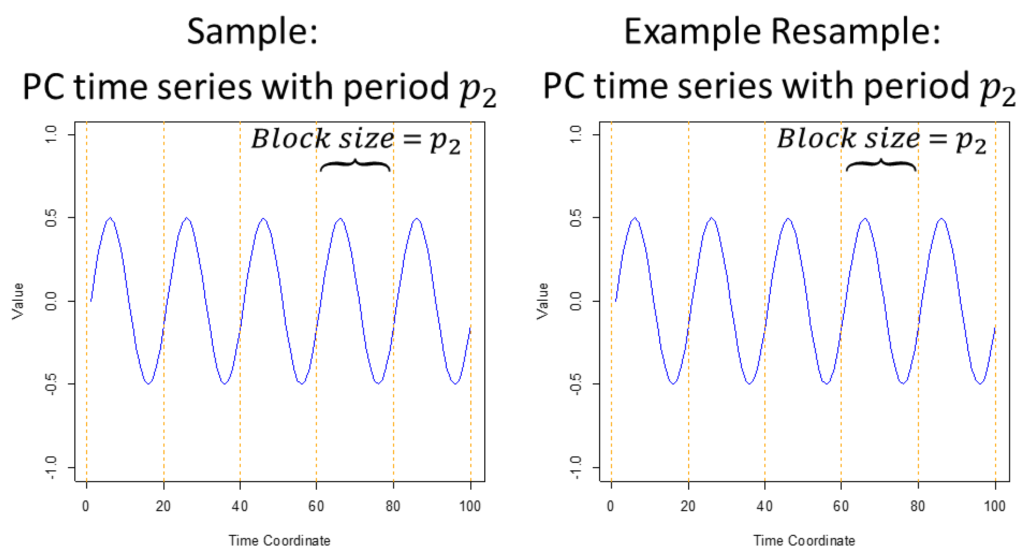

Figure 2: Illustration of resampling a PC time series component of period, p_2, with a block bootstrap strategy that incorporates the period, p_2, in the design of block length, replicating the PC structure.

In many real-world problems, time series will be a function of the accumulated influence of multiple periodically correlated (MPC) components. These problems could arise in meteorology, environmental sciences, physics, economics, and other fields where factors that operate on different time scales with different periodicities play significant roles in determining the time series. While the block bootstrapping methods described above attempt to preserve one PC correlation structure, the same reasoning that explains their success underscores the challenge presented by MPC components contained within a time series. If there are two or more PC components in a MPC time series, only in rare circumstances will the period of one PC component be an integer multiple of another. In this more trivial case, block bootstrapping can be performed on the longer period and still bootstrap the shorter without eliminating the correlation structure of either. More likely, in MPC time series with multiple PC components, each with their own correlation structure, one periodicity will not be an inter multiple of another. With MPC, these block bootstrapping methods can preserve one but not multiple periodic components. The choice of block length suitable for preserving one correlation structure will fail to preserve that of a different PC component. This is illustrated in the following figure. Any choice of block length for resampling from correlated time series with two or more component periodicities will essentially destroy at least one inherent periodic component's correlation structure.

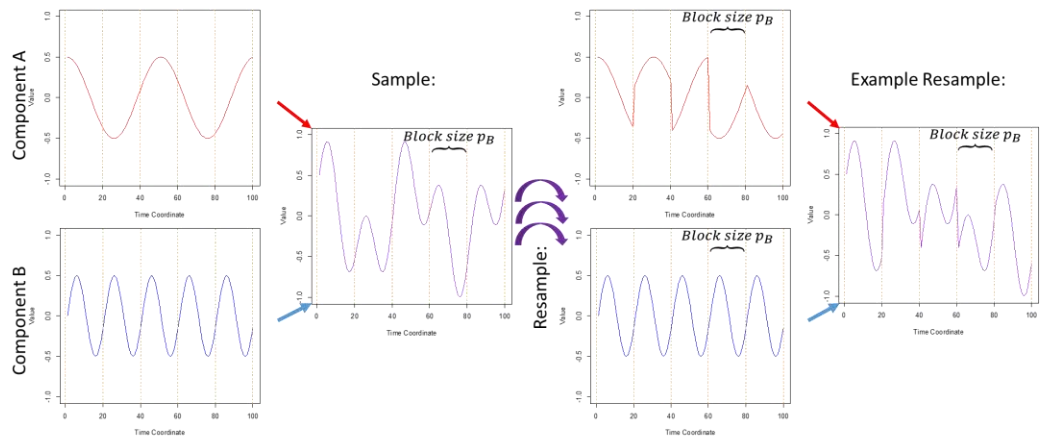

Figure 3: Illustration of resampling a MPC time series with PC component A of period, p_A, and a different PC component B of period, p_B, with a block bootstrapping strategy that incorporates the period, p_B, in the design of block length, replicating the PC structure of component B but failing to replicate the PC structure of component A.

1.3 Solution

The objective is to bootstrap MPC time series with two or more PC components while preserving the correlation structures of each periodic component present. One potential solution would be to choose a block length that is an integer multiple of the period of both components. In effect, the block length would be a common multiple and at best a least common multiple (LCM) of the two periods. This however is problematic because unless the LCM is relatively small, and as a result the periods themselves are very small, the block length will be relatively large compared to the data set length, or more likely, many times longer than the available data. This problem is compounded, with even higher demand for data, if a third or additional components are present, and the LCM is for 3 or more components.

Another solution is inspired by the periodogram, which is a spectral density or frequency domain representation of a time series and is described in Wei (1989). Two or more different periodic components operate at different frequencies, or the reciprocal of period, and signals of two different frequencies may sum in the time domain but have zero correlation between them. Different frequencies operate independently. Separating and filtering a MPC time series according to its spectral density by the individual PC component frequencies, would create a set of component PC time series each with one PC structure. This can be seen in the following figure. The set of component PC time series each with one PC structure could then individually be block bootstrapped with the appropriate block size to preserve that correlation structure. Frequency separated periodic block bootstrapping (FSPBB) should preserve all correlation structures within a MPC time series.

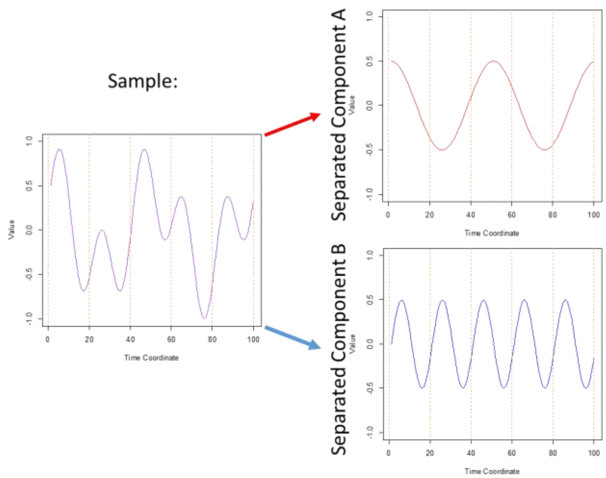

Figure 4: Illustration of separating a MPC time series into PC components

There are several potential advantages to using FSPBB on an MPC time series. The block bootstrapping techniques previously mentioned can be used on the individual PC time series of each component that result from separating the MPC time series. This strategy may be useful if the primary purpose is to investigate one PC component individually, rather than collectively as the MPC time series. This may be useful to gain insight into other components that may not be the primary intended focus of a study. This could be used if the primary purpose is investigating an individual PC component free from the interference, perhaps from a nuisance PC component present in the time domain of a MPC time series. Another potential use of this strategy beyond reproducing a single PC component, is recombining block bootstrapped components to investigate a subset of PC processes of interest. Recombining bootstrapped components can replicate the correlation structure of any combination of components up to the full MPC time series. This was not previously possible the by block bootstrapping the MPC time series directly. This strategy is illustrated in the figure below.

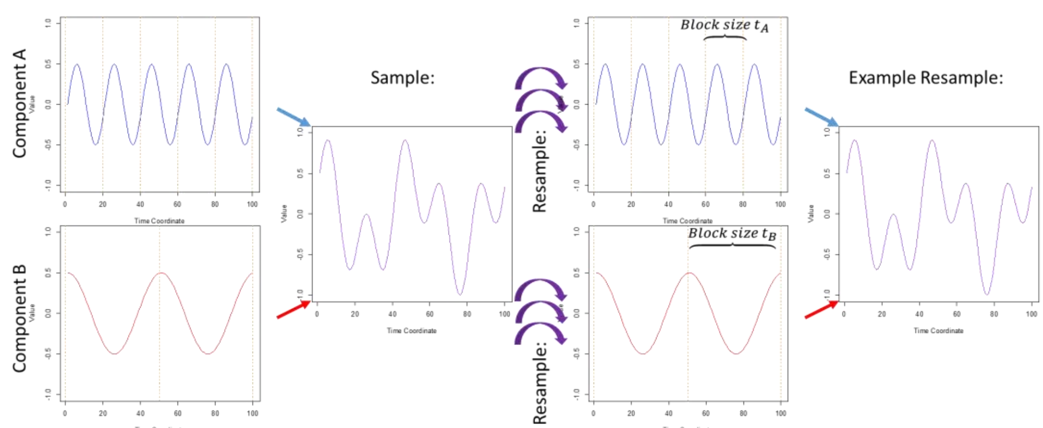

Figure 5: Illustration of resampling a MPC time series with PC component A of period, p_A, and a different PC component B of period, p_B. After PC component separation, the strategy incorporates the period, p_A and p_B, in the design of block length for block bootstrapping respectively, replicating the PC structure of component A and component B.

This FSPBB strategy can be accomplished by applying filters to the original MPC time series. It is necessary to separate and filter the MPC time series into portions of the spectral density, each containing one PC component frequency. With only two PC components this can be accomplished by applying low and high-pass filters designed so that the filter cut-offs, or the boundary between passed and attenuated bandwidth, sits between the PC component frequencies. If there are more than two PC components, bandpass filters can be used in addition to low and high pass filters with the same design feature. It may also be beneficial to FSPBB even in the case of a single PC time series, as it will separate the PC component from other potential components, even those that are not PC such as noise. This may improve investigation of the properties of the component itself.

There is a choice of filters of these types, and one such filter is the Kolmogorov-Zurbenko (KZ) filter, and its extensions. These filters are described in Zurbenko (1986). KZ filters are a class of low-pass filters, but their extensions include bandpass filters, and in combination with difference filters, are a flexible way to produce low, high, and bandpass filters with fine control over filter cut-off frequencies. KZ filters are iterations of a simple moving average. These filters are well suited for computing processes that will include bootstrapping algorithms. Also, the parameters of the filter provide a clear and direct explanation relating to the desired problem.

Kolmogorov-Zurbenko (KZ) filters and their extensions can separate portions of the frequency domain to exclude interfering frequencies as detailed in Yang and Zurbenko (2010). The frequency separated spatio-temporal components can then independently be used to reveal important details about patterns and processes often hidden within the original data, as well as associations with possible factors operating at similar spatio-temporal scales. This is the idea behind spatio-temporal analysis by frequency separation (STAFS). These filters are used to isolate frequencies in a variety of fields such as the environmental sciences, meteorology, and climatology. Examples include investigating ozone in Rao et al. (1997) and Tsakiri and Zurbenko (2010), air quality in Kang et al. (2013), global temperature in Ming and Zurbenko (2012), atmospherics in Zurbenko and Potrzeba (2010) and climate in Zurbenko and Cyr (2011). They are used to model disease such as skin cancer in Valachovic and Zurbenko (2013, 2014), and diabetes by Arndorfer and Zurbenko (2017). Valachovic and Zurbenko uses frequency separation to identify hidden PC components in skin cancer time series and perform multivariate analysis on these component factors (2017). Many of these examples highlight the use of Kolmogorov-Zurbenko filters to smooth data, reduce random variation, interpolate missing observations, and specifically separate and filter portions of the frequency domain prior to analysis. These features make KZ filters and their extensions ideal for use in FSPBB.

2. Methods

2.1 Statistical Analysis Tools

The Kolmogorov-Zurbenko (KZ) filter is the iteration of a moving average of length m, a positive odd integer defined in Zurbenko (1986). It is a filter with two parameters. The parameter m is the filter window length and the parameter k is the number of iterations. KZ filters are low pass filters that strongly attenuate signals of frequency $1/m$ and higher while passing lower frequencies. Applied to a random process $\{X(t): t \in \mathbb{Z}\}$ a KZ filter with m time points, and k iterations is defined as:

Equation 1: Kolmogorov-Zurbenko Filter

$$KZ_{m,k}\big(X(t)\big) = \sum_{u=-k(m-1)/2}^{k(m-1)/2} \frac{a_u^{m,k}}{m^k} X(t+u)$$

The coefficients $a_u^{m,k}$ are the polynomial coefficients from:

$$\sum_{r=0}^{k(m-1)} z^r a_{r-k(m-1)/2}^{m,k} = (1 + z + \cdots + z^{m-1})^k$$

One advantage of the KZ filter is the computational ease with which statistical software can apply it in an iterated form. As an application of k iterations of a moving average filter of *m* time points, the Kolmogorov-Zurbenko filter can be produced according to the following algorithm:

Equation 2: Kolmogorov-Zurbenko filter as an iterated algorithm

$$KZ_{m,1}\big(X(t)\big) = \sum_{u=-(m-1)/2}^{(m-1)/2} \frac{a_u^{m,1}}{m^1} X(t+u) = \frac{1}{m} \sum_{u=-(m-1)/2}^{(m-1)/2} X(t+u)$$

$$KZ_{m,2}\big(X(t)\big) = \frac{1}{m} \sum_{u=-k(m-1)/2}^{k(m-1)/2} KZ_{m,1}\big(X(t+u)\big)$$

$$\vdots$$

$$KZ_{m,k}\big(X(t)\big) = \frac{1}{m} \sum_{u=-k(m-1)/2}^{k(m-1)/2} KZ_{m,k-1}\big(X(t+u)\big)$$

The transfer function is the linear mapping that describes how input frequencies are transferred to outputs. The energy transfer function is the square of the transfer function and as such is symmetric about zero. The energy transfer function of the KZ filter at frequency λ is seen in the following equation. This shows how with only a few iterations, a KZ filter, strongly attenuates signals of frequency $1/m$ and higher while passing lower frequencies:

Equation 3: Kolmogorov-Zurbenko energy transfer function

$$|B(\lambda)|^2 = \left(\frac{\sin(\pi m \lambda)}{m \sin(\pi \lambda)}\right)^{2k}$$

The cut-off frequency is a limit or boundary at which the energy transferred through a filter is suppressed or diminished rather than allowed to pass through. The point where output power is $\alpha \in (0,1)$ times that of the input can be used as the boundary, and it is common to use $\alpha = 1/2$ or the half power point, a power ratio in $10 * log_{10}$ of -3 decibels units. The cut-off frequency, where the transfer function for a KZ filter is:

Equation 4: Kolmogorov-Zurbenko cut-off frequency

$$\lambda_0 \approx \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{\frac{1}{2k}}}{m^2-(1/2)^{\frac{1}{2k}}}}$$

Where the KZ filter is a low pass filter, strongly filtering signals of a frequency at or above the frequency equivalent to $1/m$, the related Kolmogorov-Zurbenko Fourier Transform (KZFT) filter is a band pass filter. KZFT is a filter applied to a random process $\{X(t): t \in T\}$ that has parameters m time points, and k iterations but is shifted to center at a frequency ν and is defined:

Equation 5: Kolmogorov-Zurbenko Fourier Transform

$$KZFT_{m,k,\nu}\big(X(t)\big) = \sum_{u=-k(m-1)/2}^{k(m-1)/2} \frac{a_u^{m,k}}{m^k} e^{-i2m\nu u} X(t+u)$$

The coefficients $a_u^{m,k}$ are the polynomial coefficients from:

$$\sum_{r=0}^{k(m-1)} z^r a_{r-k(m-1)/2}^{m,k} = (1+z+\dots+z^{m-1})^k$$

Where the KZ filter is symmetric around zero, the KZFT is a symmetric band pass filter around frequency ν. Practical use of the KZFT filter is similar to the KZ filter since it can be produced in statistical software. The energy transfer function of the KZFT filter at a frequency λ with parameters m, k, and ν is given below.

Equation 6: Kolmogorov-Zurbenko Fourier Transform energy transfer function

$$|B(\lambda-\nu)|^2 = \left(\frac{\sin(\pi m(\lambda-\nu))}{m\sin(\pi(\lambda-\nu))}\right)^{2k}$$

It follows that the cut off frequency is:

Equation 7: Kolmogorov-Zurbenko Fourier Transform cut-off frequency

$$|\lambda_0-\nu| \approx \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{\frac{1}{2k}}}{m^2-(1/2\ )^{\frac{1}{2k}}}}$$

For these filters, the cut-off frequency boundaries then become useful to determine the region of the spectra that is passed and that which is suppressed or filtered.

2.2 Statistical Theory

Kolmogorov-Zurbenko filters can PC components in MPC time series in most circumstances. Yet, it is important to understand the limitations and conditions under which the filter is applicable, as well as have detailed guidance for the implementation of the filters to accomplish their intended purpose. While any that two different frequencies can theoretically be separated by Kolmogorov-Zurbenko (KZ) filters with appropriate chosen filter parameters with a sufficiently large spatio-temporal series length or number of observations, n, so that each frequency is outside of the filter cut-off from the other frequency, this does not mean that two frequencies are separable for any set of data. While possible to separate any two different frequencies with KZ filters, regardless of how close they may be, doing so in practice requires increasingly lengthy time series and increasingly high parameters values, m and k. With finite datasets the ability to increase KZ parameters is limited.

However, the cut-off frequency can be used to derive a set of conditions necessary so that appropriate KZ filters can be assured of separating frequencies, while minimizing interference between filtered spectral components subject to the limitations of the data. The research in Valachovic (2020) answers what is the closest that any two frequencies can be and still be separated by KZ filters for a fixed time series length. That work also answers what is the minimum number of observations necessary to separate them by KZ filters, for two different given frequencies? This work uses these criteria to guide filtration for signal separation. The first proposition provides the minimal required length of time series data for separation of two given frequencies. The second proposition provides a means to calculate how close two frequencies may be for a fixed time series length, *n*, and still be separated with given KZ filters.

Proposition 1: Given $\lambda_i = 1/d_i$ and $\lambda_j = 1/d_j$ different spatio-temporal frequencies where $\lambda_i \neq \lambda_j$, and given k_i, k_j parameters of KZFT filters, where $m_i = m_j = m_{i,j} \equiv$

$$max\left(ceiling\left(\sqrt{(1/2)^{1/2k_i} + \frac{1-(1/2)^{1/2k_i}}{\frac{\pi^2}{6}\left(\frac{|\lambda_i-\lambda_j|}{2}\right)^2}}\right), ceiling\left(\sqrt{(1/2)^{1/2k_j} + \frac{1-(1/2)^{1/2k_j}}{\frac{\pi^2}{6}\left(\frac{|\lambda_j-\lambda_j|}{2}\right)^2}}\right)\right)$$

if $\left|\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{1/2k_i}}{m_{i,j}^2-(1/2)^{1/2k_i}}}\right| \leq \frac{|\lambda_i-\lambda_j|}{2}$ and $\left|\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{1/2k_j}}{m_{i,j}^2-(1/2)^{1/2k_j}}}\right| \leq \frac{|\lambda_i-\lambda_j|}{2}$ then

$$n_{i,j} \geq max\left(m_{i,j}, ceiling\left(\frac{1}{\lambda_i}\right), ceiling\left(\frac{1}{\lambda_j}\right)\right).$$

Proposition 1 may be best suited for research design, prior to observation when the time series length may be adjusted. The related question is, with a fixed spatio-temporal sequence length, *n*, what is the closest that two frequencies may be and still be separated with given KZ filters. This will determine if frequency separation is a viable strategy for block bootstrapping of MPC time series. This proposition provides the minimum separable frequency difference between any number of frequencies and is most useful for existing time series data of fixed length.

Proposition 2: If *n* is the given number of observations, and $\lambda_i = 1/d_i$ and $\lambda_j = 1/d_j$ are two frequencies so that $\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{1/2k_i}}{{m_i}^2-(1/2)^{1/2k_i}}} \leq \frac{|\lambda_1-\lambda_2|}{2}$, $i = 1, 2$ where m_1, k_1, λ_1 and m_2, k_2, λ_2 are parameters of KZFT filters, then $|\lambda_1 - \lambda_2| \geq \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{1/2k_1}}{n^2-(1/2)^{1/2k_1}}} + \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{1/2k_2}}{n^2-(1/2)^{1/2k_2}}}$.

Subject to these constraints, it is possible to separate two or more given PC components in a MPC time series using KZ filters, KZFT filter extensions, and combinations of these with difference filters.

To better understand the design of the KZ and KZFT filters that will be used to separate MPC time series prior to block bootstrapping PC components, we consider each filter parameter. Recall that the KZ filter has two parameters, m and k, and the KZFT filter adds a third parameter, ν, which is the easiest to interpret since it is the frequency around which the filter will be centered. Where the KZ filter is symmetric around zero, the KZFT is a symmetric band pass filter around frequency ν, and the KZFT filter reduces to a KZ filter when $\nu = 0$.

The KZ filter parameter m defines the width of the moving average filter window. It can be interpreted as defining the endpoints of the filter width, where the time series energy is eliminated. The effect of varying the KZ filter parameter m is illustrated in the following figure.

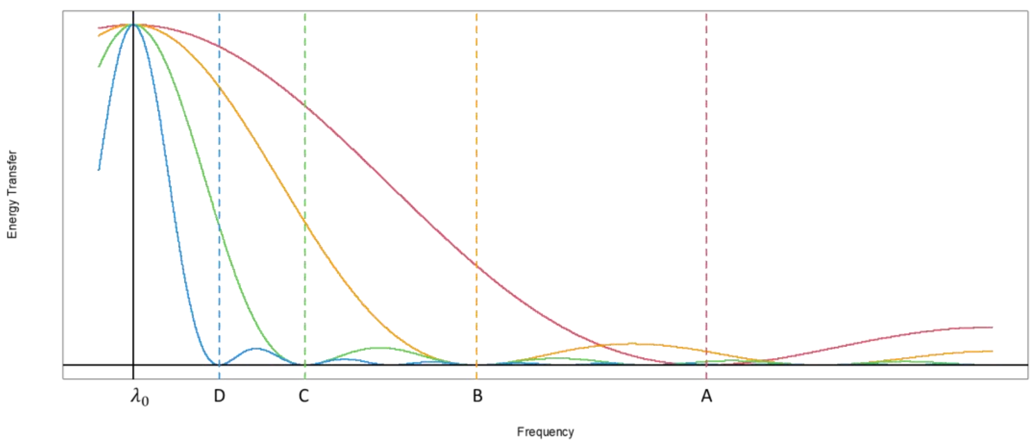

Figure 6: One side of the symmetry energy transfer functions for Kolmogorov-Zurbenko filters centered at frequency λ_0 with parameters $k = 1$ and A) $m = 3$ in red, (B) $m = 5$ in green, (C) $m = 10$ in blue, and (D) $m = 100$ in purple.

The KZ filter parameter k defines the number of iterations of the moving average filter that is performed. It can be interpreted as defining the sharpness of the filter, and it can move the filter cut-off even though it does not change the filter endpoints where the time series energy is completely attenuated. The effect of varying the KZ filter parameter k, for a constant fixed m is illustrated in the following figure.

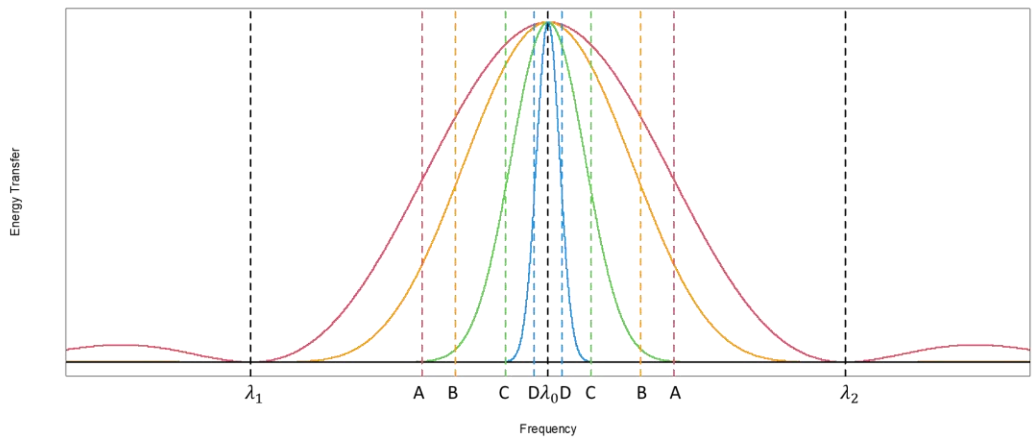

Figure 7: Energy transfer functions and KZFT filters centered at λ_0 for parameters m, such that $\frac{1}{m} = \lambda_0 - \lambda_1 = \lambda_2 - \lambda_0$ and $k = 1$ in red with cut-off (A), $k = 2$ in orange with cut-off (B), $k = 10$ in green with cut-off (C), $k = 100$ in blue with cut-off (D).

The next figure below shows how two KZFT filters centered at respective frequencies can filter and separate one bandwidth containing one frequency, from another bandwidth containing the other frequency.

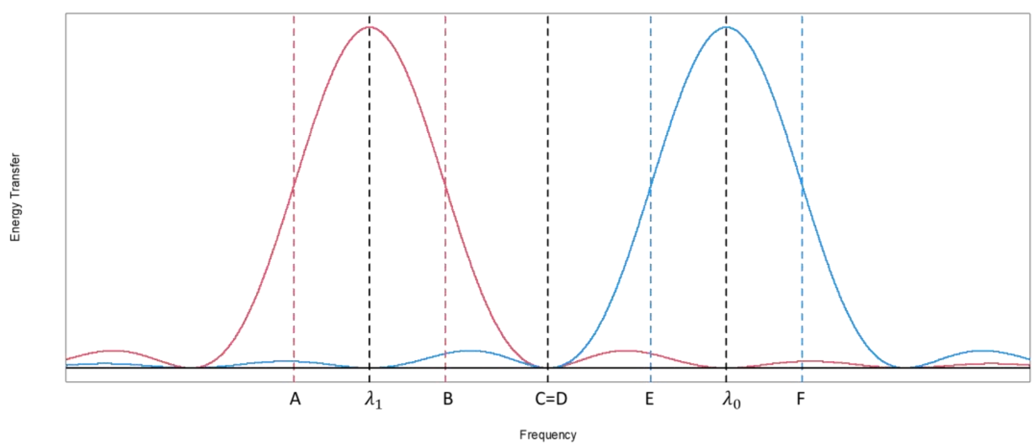

Figure 8: Illustration of the energy transfer functions from a KZFT filter centered at λ_1 with cut-offs A and B and a KZFT filter centered at λ_2 with cut-offs E and F. C and D are the frequencies excluded by the choice of window size in the respective KZFT filters, set here to be halfway between frequencies λ_1 and λ_2.

Understanding the control provided by the parameter choice of KZ and KZFT filters enables these tools to take a MPC time series, and filter all other frequencies, particularly any other PC components, that lie outside of a narrow band around a first specific PC component frequency. This filter passes only the first PC component. A difference filter between the original MPC time series and the first PC component time series results in a time series of the original MPC time series excluding that first PC component. Now proceeding to the second PC component, and filtering a narrow band around that frequency, excluding any others, provides a time series of just the second PC component. This process can be repeated, resulting in separated individual PC component time series. Each PC component time series has only one PC correlation structure. When there are no further PC components left to filter, what remains is the MPC time series absent all PC components. Provided that there is a sufficiently lengthy spatio-temporal MPC time series dataset to support implementing the filters, KZ filters offer the flexibility to separate any PC component from MPC time series.

After PC component separation, FSPBB then block bootstraps each individual PC component time series. Using methods outlined earlier, such as using block lengths equal to the period of a particular PC component, FSPBB will preserve each PC structure. This allows the investigation of PC components individually in the MPC time series. Finally, bootstrapping the full MPC time series, or any set of the PC components simultaneously, is accomplished by block bootstrapping each PC component separately with an appropriate block length, and then summing each resample. Summing a resample from every PC component to be included, will produce one resample of the intended set of components. Provided an equal number of resamples are produced for each PC component, this will produce the same number of resamples the set of PC components including the full MPC time series.

3. Simulations

The use of FSPBB for an MPC time series is supported by simulations under assumed conditions and settings comparable to real world MPC time series data analysis. Simulations validate and demonstrate the performance of Kolmogorov-Zurbenko filters to recover PC components as compared to block bootstrapping an MPC time series without frequency of components. These simulations show that standard block bootstrapping strategies from the original time series data under the presence of MPC will destroy some PC structures, while FSPBB can preserve every PC component correlation structure.

3.1 Simulation Methods

Analysis is performed in R version 4.1.1 (2013) statistical software using the KZFT function in the KZA package, see Close and Zurbenko (2013) for more detail, with datasets as a time series measured on an ordered interval dimension, in this case time. All time series are constructed with 300-time units. First, two periodically correlated sine wave signals with different periods, or frequencies, where the time coordinate determines the phase of the sine waves, are summed. The result is a multiple periodically correlated time series of interacting waves entangled in the time domain. Next, random variation is introduced by generating equal length vectors of elements randomly selected from a standard normal distribution. These random variations are then combined with the summed PC components. The final MPC time series of data is composed of the two PC components obscured by noise, seen in the figure below, and this would represent the data ordinarily available at the time of analysis.

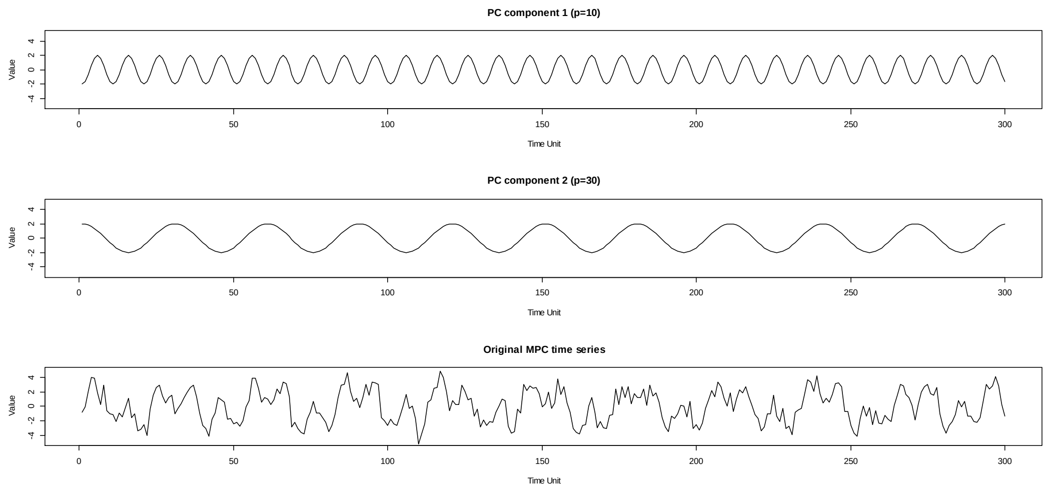

Figure 9: Simulated data of two PC components, and their sum with noise forming the MPC time series.

In these simulations, PC1 has period $p_1 = 10$ and PC2 has period $p_2 = 30$ and both have an amplitude of 2. The simulation design is used in two scenarios, first to block bootstrap the MPC time series without separating PC component frequencies by blocking to preserve the PC correlation structure for PC1 by fixing the block size to 10. The second scenario performs FSPBB by separating the PC components using KZFT filters, and block bootstrapping each component according to the described strategy, with fixed block size 10 for PC1, and 30 for PC2. In the second scenario, KZFT filters are centered above the PC component frequencies, while choosing parameters to exclude the other PC component outside of the cut-off boundary for that filter. The PC components in both simulations have frequencies 1/10 and 1/30, respectively, a frequency separation of approximately 0.067. In each scenario, all block bootstrapping is performed with 10,000 resamples.

3.2 Assessment of the Block Bootstrap Strategy

To assess the performance of the ordinary periodic block bootstrapping of the MPC time series verses the FSPBB strategy, we look at how well each strategy has replicated the correlation structure of the MPC time series as well as how well each strategy has replicated the PC correlation of individual components. While the sampling distribution of many different statistics can be replicated by block bootstrapping, with PC or MPC time series the periodic mean, sometimes referred to as seasonal mean when measured in months, are of primary interest. Here, the performance can be visualized in terms of replicating the periodic mean. After 10,000 replications, 95% confidence intervals (CIs) for the periodic means are created. These CIs are plotted for scenario one without frequency separation of the PC components in the MPC time series, and scenario two, with PC component separation prior to bootstrapping the individual components (FSPBB). Finally, the two strategies are compared by comparing their differences in CI size and coverage.

4. Results

This first step in using the FSPBB on a MPC time series is to determine if there is sufficient data available to apply KZ filters to separate adjacent frequencies. According to Proposition 2, in a time series with *n*=300 data points and choosing $\alpha = 0.5$, or the half power point for the filter cut-offs, the closest two frequencies may be and still be separated with Kolmogorov-Zurbenko (KZ) filters having parameters $k_1, = k_2 = 1$ is approximately

0.0028. A minimum frequency separation of approximately 0.0028 is less than that in the scenarios, 0.067, indicating 300 observations is sufficient to use KZ filters to separate these PC components.

After performing direct block bootstrapping, without frequency separation, on the simulated MPC time series, we can visually inspect the resulting 95% CI for the periodic mean which spans the red region in the following figure. While this region does contain the true periodic mean, represented by the black line, we notice this region representing 95% confidence is relatively wide. Furthermore, we notice a periodicity in this region, that appears to match the PC1 component period, $p_1 = 10$. So direct block bootstrapping, without frequency separation, on the simulated MPC time series with block size equal to $p_1 = 10$ has preserved the PC1 correlation structure. We also notice that within the 95% CI there appears no residual periodic structure associated with the PC2 component, at $p_2 = 30$. The red region does not change in any periodic way with a period of 30. This results in regions that substantially misalign with the true periodic mean, observed where the most red is visible in the figure, on a periodic cycle of period $p_2 = 30$. The correlation structure of the PC2 component has been destroyed.

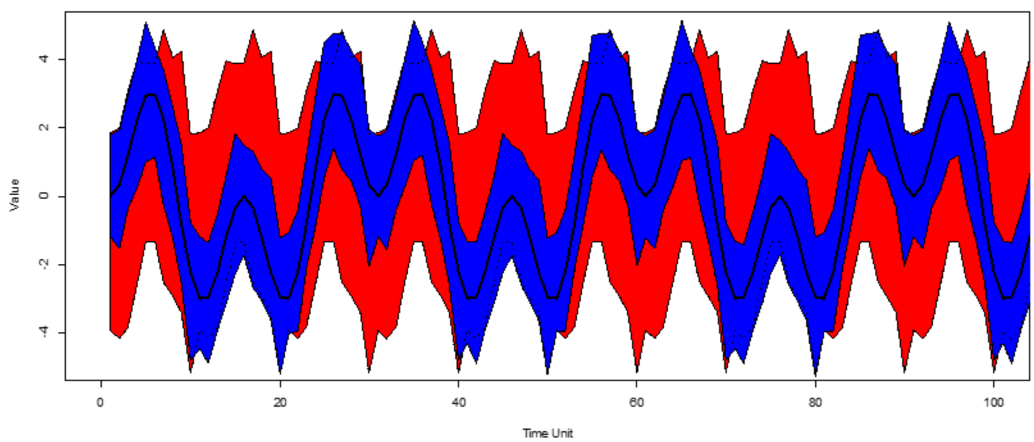

Figure 9: A piece of the MPC time series periodic mean, seen as the black line. 95% CIs for the periodic mean of the MPC time series using direct block bootstrapping of the MPC ($p = 10$) without frequency separation in red and block bootstrapping the MPC by block bootstrapping the frequency separated components PC1 ($p_1 = 10$) and PC2 ($p_2 = 30$) separately in blue.

If we contrast this with the blue region in the figure, representing block bootstrapping the MPC time series after frequency separation, or FSPBB, by separately block bootstrapping the PC components, we observe relatively narrow 95% CIs for the periodic mean. Again, the periodic mean of the MPC time series is contained within the blue 95% CI, but this CI aligns well with the true periodic mean. We notice that the blue 95% CI changes with both a period of 10 and 30, so the PC1 and PC2 correlation structures have been preserved.

While this is only the results from one illustrative example, to quantify the comparison, we note that block bootstrapping of the MPC time series without frequency separation of the PC components, resulted in 95% CIs for the periodic mean that ranged from approximately 77% larger to 115% larger than the CIs produced when block bootstrapping the separated PC components of the MPC time series using FSPBB.

5. Discussion

To gain insight about a population from which a sample is available, bootstrapping, or resampling with replacement from a given set of data, is a useful tool with few assumptions to estimate the sampling distribution of statistics on that population. Simple random resampling, however, does not preserve correlation structures between successive data points, and in general is not useful for time series or more broadly spatio-temporal data. Block bootstrapping methods attempt to replicate and preserve correlations in spatio-temporal data. Periodically correlated (PC) components of a given period, p, exhibit a unique type of correlation structure between observations in spatio-temporal data. Many block bootstrapping strategies are not suitable for preserving PC time series. While there are certain block bootstrapping strategies that account for the period, p, and bootstrap to preserve the PC correlation structure, these strategies are not suitable time series with two or more different PC components, called multiple periodically correlated (MPC) time series. Any choice of block length for resampling directly from MPC time series with two or more PC component will destroy at least one PC component's correlation structure.

This study suggests the better strategy of FSPBB, which bootstraps MPC time series by frequency separating PC components, block bootstrapping these components individually, accounting for their individual PC correlation structure, and then combining resamples to produce bootstraps of the original MPC time series. Simulations showed more accurate and more precise confidence intervals can result from performing FSPBB on MPC time series.

There are limitations in this strategy. This study illustrates the importance of understanding the applications and limitations of Kolmogorov-Zurbenko (KZ) filters and spatio-temporal analysis by frequency separation. There are limitations to what frequencies are detectable and how closely two frequencies can be and still be separated by Kolmogov-Zurbenko filters. Separating components by frequency must follow well defined guidelines given by Valachovic (2020). With this guide it is possible to determine what separation may be expected given a set of data, and similarly what data is required to investigate the separation of two or more targeted signals. In addition, the new FSPBB strategy introduced here needs further investigation to determine consistency under a variety of different factors including but not limited to relative PC component strength, frequency variation, additional PC components, non-PC components, and level of noise.

The conclusions of this study and the advantages of this new strategy are clear. The results show that the presence of multiple PC components prevent the accurate estimation of the sampling distribution of MPC time series. Multiple PC components can even prevent accurate estimation of the sampling distribution of any single specific PC component, even if only one is of interest. The presence of other PC components in a MPC time series had a negative impact on replicating any one PC structure of interest, even if correctly block bootstrapping for just that component. Frequency separation of the PC components becomes a necessary step before block bootstrapping, and FSPBB used on MPC time series preserves the correlation structures of multiple PC components.

References

1. Arndorfer, S. and Zurbenko, I. (2017). Time Series Analysis on Diabetes Mortality in the United States, 1999- 2015 by Kolmogorov-Zurbenko Filter. *Journal of Biometrics & Biostatistics*. 08. 10.4172/2155-6180.1000384.
2. Chan, V., Lahiri, S.N., and Meeker, W.Q (2004), 'Block bootstrap estimation of the distribution of cumulative outdoor degradation', *Technometrics*, 46, 215–224.
3. Close, B., and Zurbenko, I. (2013). Kolmogorov-Zurbenko adaptive filters. (Version 3). Retrieved from http://cran.r-project.org/web/packages/kza/index.html
4. Dudek, A. E., Leśkow, J., Paparoditis, E., & Politis, D. N. (2014). A generalized block bootstrap for seasonal time series. *Journal of Time Series Analysis*, 35(2), 89-114.
5. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*. 7 (1): 1–26. doi:10.1214/aos/1176344552.
6. Kang, D., Hogrefe, C., Foley, K. L., Napelenok, S. L., Mathur, R., & Rao, S. T. (2013). Application of the Kolmogorov–Zurbenko filter and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model. *Atmospheric environment*, 80, 58-69. https://doi.org/10.1016/j.atmosenv.2013.04.046
7. Kunsh, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics*. 17 (3): 1217-1241. Doi: 10.1214/aos/1176347265
8. Ming, L. and Zurbenko, I. (2012). Restoration of Time-Spatial Scales in Global Temperature Data. *American Journal of Climate Change*, 1(3), 154-163.
9. Politis, D.N. (2001), 'Resampling time series with seasonal components', In Frontiers in Data Mining and Bioinformatics: Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics, Orange County, California, June 13-17, 2001, pp. 619–621.
10. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
11. Tsakiri, K. G., and Zurbenko, I. G. (2010). Determining the main atmospheric factor on ozone concentrations. *Meteorol Atmos Phys*, 109(3), 129-137.
12. Valachovic, E. (2020). Spatio-Temporal Analysis by Frequency Separation: Approach and Research Design. In JSM Proceedings, Section on Statistics and the Environment. Alexandria, VA: American Statistical Association. 1931-1946.
13. Valachovic, E. & Zurbenko, I. (2013). Skin Cancer and the Solar Cycle: An Application of Kolmogorov-Zurbenko Filters. In JSM Proceedings, Section on Statistics in Epidemiology. Alexandria, VA: American Statistical Association. 2165-2178.
14. Valachovic, E. and Zurbenko, I. (2014). Skin Cancer, Irradiation, and Sunspots: The Solar Cycle Effect. *Biomed Res Int*, 2014, 1-9. doi:10.1155/2014/538574
15. Valachovic, E. and Zurbenko, I. (2017). Multivariate analysis of spatial–temporal scales in melanoma prevalence. *Cancer Causes Control*, 28: 733. https://doi.org/10.1007/s10552-017-0895-x
16. Wei, W. (1989). Time Series Analysis: Univariate and Multivariate Methods. Redwood City, California: Addison-Wesley.
17. Yang, W., and Zurbenko, I. (2010). Kolmogorov-Zurbenko Filters. *Computational Statistics*, 2(3), 340-351. DOI: 10.1002/wics.71
18. Zurbenko, I. (1986). The Spectral Analysis of Time Series. North-Holland, Amsterdam. 248 pages.
19. Zurbenko, I. and Cyr, D. (2011). Climate fluctuations in time and space. *Clim Res*, 46(1), 67-76.

20. Zurbenko, I., Potrzeba, A. (2010). Tidal waves in the atmosphere and their effects. *Acta Geophys*. 58, 356–373. https://doi.org/10.2478/s11600-009-0049-y