Quest Journals
Journal of Research in Applied Mathematics
Volume 3 ~ Issue 7 (2017) pp: 29-37
ISSN(Online) : 2394-0743 ISSN (Print):2394-0735
www.questjournals.org

Research Paper

Separation of Spatio-Temporal Frequencies

Edward Valachovic[1*], Igor Zurbenko[1]
[1]*Department Of Epidemiology And Biostatistics, State University Of New York At Albany, Rensselaer, USA*
*Corresponding Author: Edward Valachovic**

Received 09 October, 2017; **A**ccepted 28 October, 2017 © The author(s) 2017. **P**ublished with open access at **www.questjournals.org**

ABSTRACT: *Time series data are often composed of principal components with features that operate in independent time scales. Statistical analysis absent the necessary separation of uncorrelated frequencies can confound factors, obscure true relationships, and diminish measures of associative model strength. With guided selection of filter parameters, Kolmogorov-Zurbenko filters and their extensions provide targeted tools to filter bandwidth of the frequency domain and can smooth variation, interpolate missing data, and provide the separation of uncorrelated frequencies to enable appropriate analysis of factors within each independent time scale. This study first proves that under certain conditions any two frequencies are separable, and then derives formulas for the minimum number of observations necessary to separate two frequencies and the minimum spectral distance between frequencies that may be separable given a set of data. Finally, this study simulates the separation and reconstruction of component frequencies from raw noisy signals with missing observations to demonstrate and evaluate the data requirements, investigatory limits, and performance of these filters to separate frequencies of interest in both temporal and spatial data.*

Keywords: *Frequency Domain, Kolmogorov-Zurbenko Filter, Parameter Selection, Signal Separation, Spatio – Temporal, Time Series.*

I. INTRODUCTION

Time Series Analysis, or the observation of data across time, commonly involves variation that exhibits periodicities, or cyclic fluctuations. These cycles can result from natural phenomenon, such as seasonal or daily rhythms, to manmade processes such as work weeks. The variation associated with one cycle may be smaller than that associated with different component factors such as the overall mean trend across time, random variation, exogenous shocks, or other cycles. In the time domain, each observation is the collective sum of all factors at that time point. Smaller factors can be obscured. However, these sources of variation may operate on different time scales, such as the case with cycles operating at an associated frequency. Therefore, within the frequency domain, the spectral representation of a time series provides an opportunity to separate and investigate different time scales without the entanglement in the time domain.

Kolmogorov-Zurbenko (KZ) filters and their extensions are able to separate portions of the frequency domain to exclude interfering frequencies [1, 2]. These filters are used to isolate frequencies in a variety of fields such as the environment, meteorology, and climate [3, 4, 5]. They have also been used to separate and model pollution and public health [6, 7]. They can be applied to time series of higher dimensionality [8]. Recently their use was extended to epidemiological surveillance data in a multivariate analysis of the frequency separated uncorrelated components of variables thereby greatly improving model fit [9]. Many of these examples highlight the use of Kolmogorov-Zurbenko filters to smooth data, reduce random variation, interpolate missing observations, and necessarily separate portions of the frequency domain prior to analysis [9,10].

Some of the prior examples address filtering only one portion of the spectrum, or widely separated components. However, with the increasing use of KZ filters in various research fields and the use of filters to split closely adjacent frequencies or where data is scarce, guidance for minimum data requirements is necessary to guarantee filter performance. First this study proves that under certain circumstances for any two frequencies, and by extension any number of frequencies, there exists pairs of KZ filters that retain one frequency while excluding the other. Next, and more useful in practice, this study derives an expression for what number of observations is necessary to separate two given frequencies as well as the closest that two different frequencies may be in an analysis and still be separated given a fixed set of observations. These expressions are then extended in spatial frameworks to work with the spatial equivalent of frequency. Computer simulations of component signals, combined with strong random errors and missing data are used to model real world spatio-temporal raw data. Then programming within statistical analysis software demonstrates and assesses the applications, limitations, and outcomes of KZ filters to isolate, separate, and reconstruct the two original frequencies from the raw data signal.

II. METHODS

1.1 Statistical Analysis Tools

The Kolmogorov-Zurbenko (KZ) filter is an iteration of a moving average of length m, where m is a positive odd integer [1, 2]. It is a filter with two parameters. The parameter m is the filter window size and 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:

Definition 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 a moving average filter of m time points for k iterations the Kolmogorov-Zurbenko filter can be produced:

Definition 2: Kolmogorov-Zurbenko filter as an iterated algorithm

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

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

…

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

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:

Definition 3: Kolmogorov-Zurbenko energy transfer function

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

The cutoff frequency is a limit or boundary at which the energy transferred through a filter is suppressed or diminished rather than allowed to pass through. A cutoff frequency is used in many fields such as physics, communications, and electrical engineering, and selection depends upon the application. One common boundary is the point where output power is one half that of the input, a power ratio in $10{*}\log_{10}$ of -3 decibels units. Here the power ratio is left variable, expressed as α. The cutoff frequency, where the transfer function takes the value $\alpha \in (0, 1)$ for a KZ filter is:

Definition 4: Kolmogorov-Zurbenko cut off frequency

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

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 $\{X(t):\ t \in \mathbb{Z}\}$ that has parameters m time points, and k iterations but is shifted to center at a frequency ν and is defined:

Definition 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:

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

Where the KZ filter is symmetric around 0, 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.

Definition 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:

Definition 7: Kolmogorov-Zurbenko Fourier Transform cut off frequency

$$|\lambda-\nu| \approx \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(\alpha)^{\frac{1}{2k}}}{m^2-(\alpha)^{\frac{1}{2k}}}}, \alpha\epsilon(0,1)$$

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

1.2 Statistical Theory

With sufficiently large numbers of observations, this study first proves that any two frequencies can be separated by Kolmogorov-Zurbenko (KZ) filters with appropriate chosen filter parameters, so that each frequency is outside of the filter cutoff from the other frequency. In practice this does not mean that different frequencies are separable for any set of data. However, the cutoff 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. Next, this study derives how many observations may be necessary in order to separate two given frequencies and then it details what separation is possible given a certain quantity of data. Finally, these derivations are extended to separation of spatial frequencies.

Proceeding with the outline of the first proof, it is through the choice of filter parameters that control is exercised over the KZ filters, and their extensions such as KZFT. With the goal to separate and filter each of two different given frequencies and control the range of the cutoff frequency, it is possible to create two filters that center or pass one frequency while selecting window size, m, so that 1/m is less than or equal to one half the separation range or bandwidth between the two given frequencies. The cutoff frequency would then be closer to the central target frequency, thereby attenuating the other target frequency enough so that interference is kept below a predetermined arbitrarily small level controlled by the choice of α ∈ (0, 1). For simplicity, the proofs use the KZFT filter which can center the bandpass filter over a given frequency, and attenuate other frequencies based on the choice of the window size m and number of iterations.

Proposition 1: If λ_1 and λ_2 are frequencies where $\lambda_1 \neq \lambda_2$, then there exists an n ∈ ℕ, time points, and KZFT filters with parameters m_1, k_1, λ_1 and m_2, k_2, λ_2 so that $\left|\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_i}}{m_i^2-\alpha^{1/2k_i}}}\right| \le \frac{|\lambda_1-\lambda_2|}{2}, \alpha \in (0,1), i = 1,2.$

Proof of Proposition 1: Let $\frac{|\lambda_1-\lambda_2|}{2} = d$. There exists an m ∈ ℕ time points such that m = ceiling(min(1/*d*, $1/\lambda_1$, $1/\lambda_2$)). Now take n, the number of observations, so that n ≫ m, hence m is a viable choice for window size. Then 1/*d* ≤ m ≤ n and *d* ≥ 1/m ≥ 1/n. It follows: $\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k}}{m^2-\alpha^{1/2k}}} \le 1 * \sqrt{\frac{1-\alpha^{1/2k}}{m^2-\alpha^{1/2k}}} \le \sqrt{\frac{1}{m^2}} = \frac{1}{m} \le d$. This is true for any k ∈ ℕ. Take $m_1 = m_2 = m$, and $k_1 = k_2 = 1$, and from the solutions above it follows $\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_i}}{m_i^2-\alpha^{1/2k_i}}} \le d = \frac{|\lambda_1-\lambda_2|}{2}, i = 1,2$, QED.

This proposition only required that there is some sufficiently large number of observations. It does not indicate that this is in some way a wise choice for n, or for that matter a practical choice. In practice it is unlikely someone is able to choose any large number n of observations with which to separate frequencies. For this reason it is interesting to know what a lower limit of n observations that would be necessary to again be certain that given two frequencies, they can be separated outside of filter cutoff boundaries. This necessitates adjusting filter parameters so that bandwidth is not wasted with unnecessarily larger choices of m or k. The target frequencies are unchanged, thus the only parameters remaining are the filter windows size and iterations. Larger numbers of iterations narrows the bandpass filter, but requires higher numbers of observations because time points are discarded from the beginning and end of the available data due missing data outside the filter window time range. The only remaining parameter to adjust is the filter window size. Adjusting the window size of two filters centered at different frequencies so that the respective cutoff frequencies approach but do not overlap should separate with the minimum number of observations required. Fig.1 illustrates KZFT filters centered over different frequencies and how lowering the choice of m for each filter should decrease the number of observations required while still separating the frequencies. As the filters widen, bandpass regions do not overlap up to the point that cutoff boundaries equal. This attenuates the interference caused by the other frequency with the minimum number of observations.

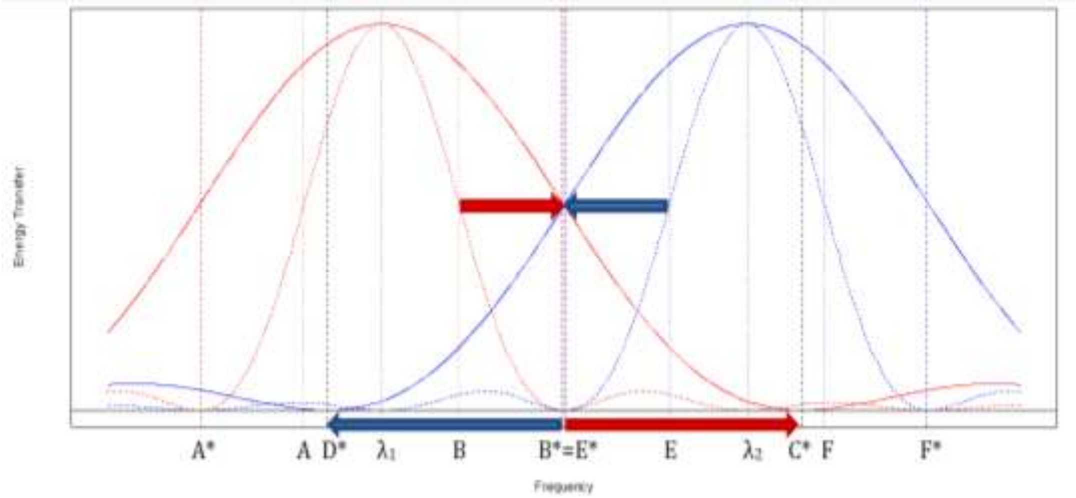


Figure 1: Illustration of 2 different frequencies λ_1 and λ_2, and the reduction of filter window sizes so A and B cutoffs shift to A* and B* from a KZFT filter centered at λ_1 and E and F cutoffs shift to E* and F* from a KZFT filter centered at λ_2. Window size is reduced until cutoff B* equals cutoff E*. C* and D* provide the new frequencies to set window size in the respective KZFT filters.

With this motivation we proceed to the next proposition describing the theoretical lower limit of observations necessary to separate two given frequencies so that the cutoff frequencies of KZFT filters are near equal and attenuated regions do not overlap.

Proposition 2: Given λ_1 and λ_2 are frequencies where $\lambda_1 \neq \lambda_2$, and k_1 and k_2 are given parameters of KZFT filters, if $m_1 = m_2 = m^* = max\left(ceiling\left(\sqrt{\alpha^{1/2k_1} + \frac{1-\alpha^{1/2k_1}}{\frac{\pi^2}{6}\left(\frac{|\lambda_1-\lambda_2|}{2}\right)^2}} \right), ceiling\left(\sqrt{\alpha^{1/2k_2} + \frac{1-\alpha^{1/2k_2}}{\frac{\pi^2}{6}\left(\frac{|\lambda_1-\lambda_2|}{2}\right)^2}} \right) \right)$ and $n \geq \max(m^*, \text{ceiling}(1/\lambda_1), \text{ceiling}(1/\lambda_2))$ then $\left| \frac{\sqrt{6}}{\pi} \sqrt{\frac{1-\alpha^{1/2k_i}}{m_i^2-\alpha^{1/2k_i}}} \right| \leq \frac{|\lambda_1-\lambda_2|}{2}$, $\alpha \in (0,1), i = 1,2.$

Proof of Proposition 2: Let $\frac{|\lambda_1-\lambda_2|}{2} = d$. It follows that

$$m = \sqrt{\alpha^{1/2k} + \frac{1-\alpha^{1/2k}}{\frac{\pi^2}{6}d^2}} \underset{c=\frac{\pi^2}{6}d^2}{\Longleftrightarrow} m = \sqrt{\alpha^{1/2k} + \frac{1-\alpha^{1/2k}}{c}} \Leftrightarrow$$

$$\Leftrightarrow m^2 = \alpha^{1/2k} + \frac{1-\alpha^{1/2k}}{c} \Leftrightarrow m^2 = \frac{c\alpha^{1/2k} + 1 - \alpha^{1/2k}}{c} \Leftrightarrow cm^2 = c\alpha^{1/2k} + 1 - \alpha^{1/2k}$$

$$\Leftrightarrow cm^2 - c\alpha^{1/2k} = 1 - \alpha^{1/2k} \Leftrightarrow c = \frac{1-\alpha^{1/2k}}{m^2 - \alpha^{1/2k}} \underset{c=\frac{\pi^2}{6}d^2}{\Longleftrightarrow} \frac{\pi^2}{6}d^2 = \frac{1-\alpha^{1/2k}}{m^2-\alpha^{1/2k}} \Leftrightarrow d$$

$$= \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k}}{m^2-\alpha^{1/2k}}}$$

This is true for any k. With $m_1 = m_2 = m^* =$

$$max\left(ceiling\left(\sqrt{\alpha^{1/2k_1} + \frac{1-\alpha^{1/2k_1}}{\frac{\pi^2}{6}\left(\frac{|\lambda_1-\lambda_2|}{2}\right)^2}}\right), ceiling\left(\sqrt{\alpha^{1/2k_2} + \frac{1-\alpha^{1/2k_2}}{\frac{\pi^2}{6}\left(\frac{|\lambda_1-\lambda_2|}{2}\right)^2}}\right)\right)$$

and $n \geq$ max(m^*, ceiling($1/\lambda_1$), ceiling($1/\lambda_2$)) it follows $\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_i}}{m^{*2}-\alpha^{1/2k_i}}} \leq \frac{|\lambda_1-\lambda_2|}{2}$ $, i = 1{,}2$, QED.

Proposition 1 proved that any two different frequencies may be separated by the cutoffs of appropriate KZ filters as the number of observations goes to infinity. Proposition 2 derives the smallest n possible to separate two given frequencies with these filters. Generally much larger numbers of observations are desired to more accurately represent the patterns in data over time. In practice, n is not chosen but is fixed with the data available. Waiting for additional future observations to be recorded to extend the dataset may not be practical or possible. A final question is therefore, with a fixed n, what is the closest that two frequencies may be and still be separated with KZ filters.

Proposition 3: If n is the number of observations, and λ_1 and λ_2 are two frequencies so that $\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_i}}{m_i^2-\alpha^{1/2k_i}}} \leq \frac{|\lambda_1-\lambda_2|}{2}$ $, \alpha \in (0{,}1), i = 1{,}2,$ where m_1, k_1, λ_1 and m_2, k_2, λ_2 are parameters of KZFT filters,

then $|\lambda_1 - \lambda_2| \geq \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_1}}{n^2-\alpha^{1/2k_1}}} + \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_2}}{n^2-\alpha^{1/2k_2}}}$.

Proof of Proposition 3: Assume the given statements. By definition of KZFT filters, m_i □ n, *i* = 1, 2, so it follows $m_i{}^2 - \alpha^{1/2k_i} \leq n^2 - \alpha^{1/2k_i}$ and the cutoff frequency $\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_i}}{n^2-\alpha^{1/2k_i}}} \leq \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_i}}{m_i^2-\alpha^{1/2k_i}}}$. Then,

$$\frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_1}}{n^2-\alpha^{1/2k_1}}} + \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_2}}{n^2-\alpha^{1/2k_2}}} \leq \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_1}}{m_1^2-\alpha^{1/2k_1}}} + \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-\alpha^{1/2k_2}}{m_2^2-\alpha^{1/2k_2}}} \leq \frac{|\lambda_1-\lambda_2|}{2} + \frac{|\lambda_1-\lambda_2|}{2} = |\lambda_1 - \lambda_2|,$$

QED.

The previous three propositions extend immediately from purely temporal data into spatial and higher dimensioned mixed spatio-temporal frameworks. Spatial frequency, or the reciprocal of distance (1/distance), may substitute in place of the frequency in the prior propositions. This provides equivalent statements and proofs, omitted here, of spatial corollaries to the respective propositions above. The results of these propositions provide for the robust application of Kolmogorov-Zurbenko filters and their extensions to separate spatio-temporal components in multidimensional time series data. In real world datasets, only with sufficient observations and an appropriate choice of KZ parameters can the separation between different frequencies be effective so that time scale components can treated as independent.

1.3 Simulation

The theoretical conclusions of this study are supported by the use of simulations under assumed conditions and settings comparable to real world spatio-temporal data analysis. Simulations help understand and illustrate the performance of multidimensional Kolmogorov-Zurbenko filters to recover signals from original observed raw data, conditions that may require the separation, isolation and recovery of signals of different frequency, from a high degree of noise and missing data rates.

Analysis is performed in R version 3.1.1 statistical software using the KZA and KZFT packages [11, 12] with datasets in arrays with two spatial dimensions and one time dimension. Arrays are constructed with 50 x-axis and 50 y-axis spatial units, and 100 time units. These arrays are populated by the sum of two spatio-temporally dependent sin wave signals with different frequencies and spatial patterns, where time and a combination of x and y coordinates determine the phase of the sin wave. The result is a motion picture in time of moving and interacting waves entangled in the time domain. Next, random variation is introduced by generating equal size arrays of elements randomly selected from a uniform distribution spanning ±ε where ε is five times the amplitude of the original signals. These arrays of random variations are then combined with the array of the original pure signals. Finally, each (x, y) coordinate within the array is assigned a uniformly distributed randomly generated number from which a fixed percentage are selected and discarded as missing. This simulates the geographic scarcity of often present in data. In this example the chance of being selected missing is 50 percent. The resulting arrays of data are composed of the pure signals combined with noise, and then with randomly selected missing observations discarded to form the final raw data to be processed. One frame, or time point, of the final result of the simulated data array can be seen in Fig. 2.

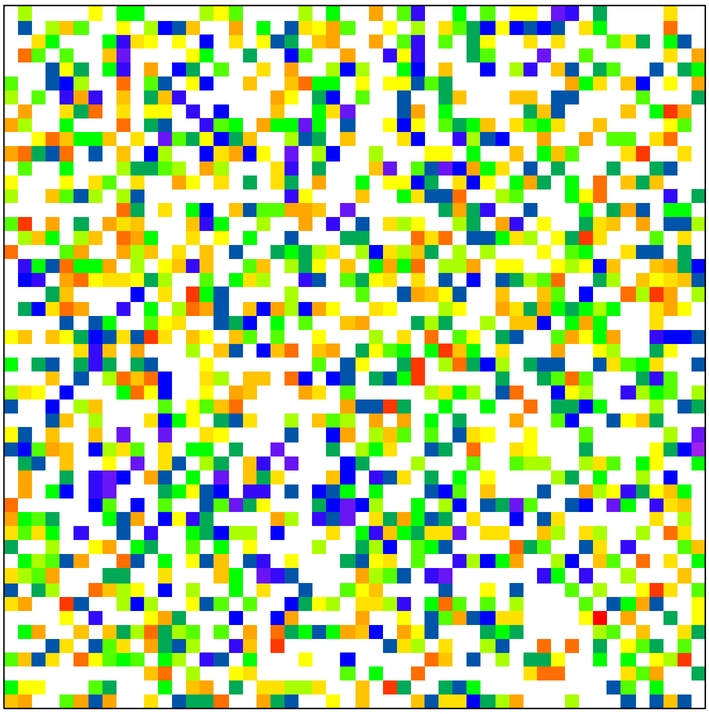


Figure 2: Simulated raw data of two signals, with noise and missing observations at one time point.

This simulation design is used in two simulation scenarios to demonstrate the formulas described above for separating frequencies. In each scenario, KZFT filters are centered above the original signal frequency, while choosing parameters to exclude the other frequency outside the cutoff boundary for that filter. A combination of KZFT filters removes the longer period, lower frequency, signal to reconstruct the shorter period, higher frequency, signal. The resulting reconstructed high frequency signal can be compared to the original true high frequency signal initially used in construction of the data at the same time point. The comparison is made at the same time point in Fig. 3 and Fig. 4. The separation, filtering, and signal reproduction is then repeated with the role of high and low frequency signals reversed. The signals in scenario 1 have frequencies 0.020 and 0.025, a separation of 0.005. In the second scenario the frequencies are farther apart at frequencies 0.020 and 0.050, a separation of 0.03. With this data, KZ filters are used to smooth observations, interpolate missing values, reduce random variation, and separate the two component signals.

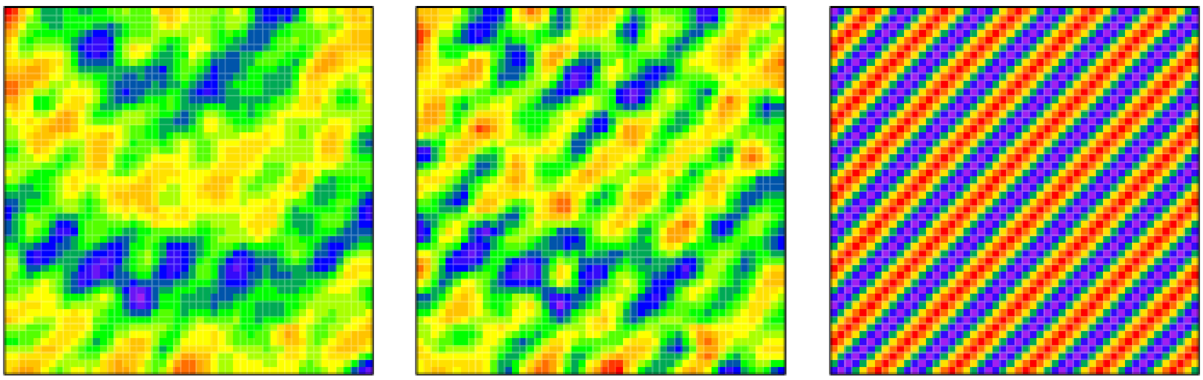
1.4 Assessment of Quality of Fit

Correlation is a normalized measure of the association between random processes. Correlation measures the similarity in how one random process varies in time relative to a different process. It assumes a value between -1, implying perfect negative correlation, and positive 1, implying perfect positive correlation between random variables. Zero implies that the two random variables are not correlated. The coefficient of determination, calculated as the square of correlation, is a measure that indicates the quality of fit of one time series to another by the fraction of the variance of one that is explained by another. In classical statistics, particularly linear regression through ordinary least squares, a typical assumption is that observations are independent and identically distributed. Time series processes are unlikely to be independent, violating these assumptions, but the use of R^2 for time series does not require the assumption of independence of observations and is mathematically identical in calculation to that in classical statistics. This means that functions for calculating the R^2 provided in statistical software can be used in time series, with care to interpret it as a measure of the goodness of fit between two time series and the percentage of variance of one time series that is explained by another. Here correlation and coefficient of determination are used after simulation between one reconstructed component against another, and against the known true original signal to assess the fit, revealing the ability to both separate and to reconstruct, respectively.

III. RESULTS

According to Proposition 3, in a time series with 100 observations and with $\alpha = 0.5$, or half power, the closest two frequencies may be is approximately 0.0062, with Kolmogorov-Zurbenko (KZ) filters having parameters k_1, $k_2 = 2$. We note here one KZ iteration does not completely interpolate all missing data, and more than two iterations require filter windows with wider support than the number of observations given, making two the natural choice. As may be the case, there are times when some parameters are dictated by the particular application or research design. A minimum frequency separation of 0.0062 is more than the frequency separation in scenario one, 0.005. This indicates that in the first scenario the frequencies are too close to each other for 100 observations to sufficiently separate them. A minimum frequency separation of 0.0062 is less than that in scenario two, 0.03, indicating 100 observations is a least sufficient. Indeed, according to Proposition 2, a frequency separation of 0.005 should require 125 observations at a minimum. A frequency separation of 0.03 should require 50 observations at a minimum.

What results in the following figures after filtration and signal reconstruction are images that have smoothed noise and interpolated missing observations, but in scenario one where signal separation is 0.005, the two signals are not well separated (Fig. 3a and Fig. 4a). Both high and low frequencies are still present and somewhat visible, looking like a mix of the true high and low frequency component (Fig. 3c and Fig. 4c). The filters left the two signals entangled. In the images corresponding to the second scenario, there is improved signal separation (Fig. 3b and Fig. 4b). When the frequencies are farther apart the reconstructed higher frequency signal looks increasing like the true high frequency signal and exhibits less remnants of the low frequency signal. When the signal frequencies are very close, given limited data, the filters capture more of the adjacent frequencies including the other interfering signal, resulting in a reconstructed image that is more of a blend of the higher and lower frequency signals. When the signals are close there can be confusion as to what the reconstructed signal indicates is the true pattern at that frequency. In a real world scenario, where the true component signals are not known, the reconstructed signals can easily be mistaken as arising from the other component, or without indicating a given pattern at all.



a) b) c)

Figure 3: Reconstructed high frequency signals when separation is (a) 0.005, (b) 0.03, and (c) the true high frequency component signal.

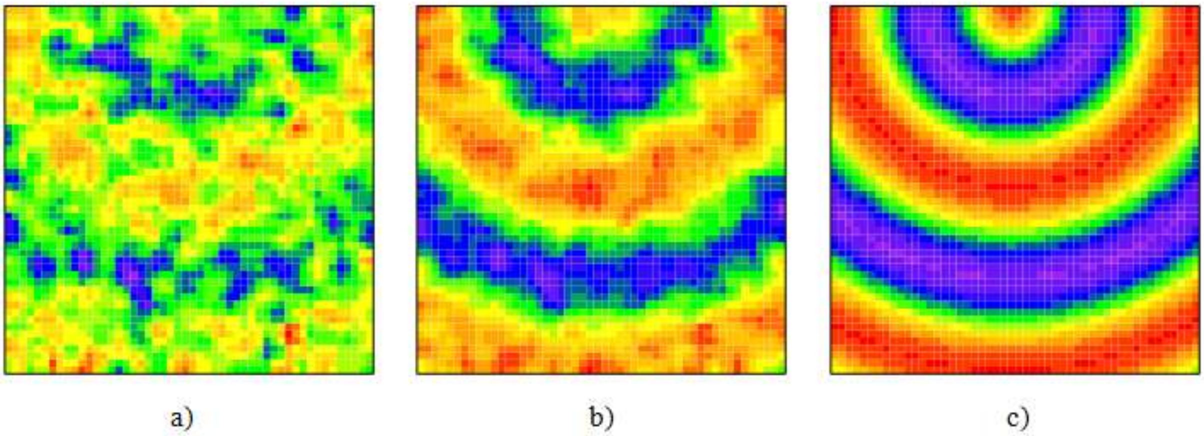


Figure 4: Reconstructed low frequency signals when separation is (a) 0.005, (b) 0.03, and (c) the true low frequency component signal.

Table 1 displays calculated fit statistics for the models of reconstructed signals produced. In the first scenario where there is a small separation between signals, the reconstructed high frequency signal, reconstructed low frequency signal, and original raw data with noise and missing observations are fit against the true component high and low signal. The same is done for the second scenario when the signal separation is larger. Correlation and coefficient of determination are provided.

Table 1: Correlation and Coefficient of Determination between raw data, recovered signals, and true signals.

Simulation Scenario 1, Frequency Separation 0.005		
Correlation (R^2)	True High	True Low
Raw Data	0.236 (0.056)	0.232 (0.053)
Recovered High	0.463 (0.214)	0.453 (0.205)
Recovered Low	0.097 (0.009)	0.463 (0.214)
Simulation Scenario 2, Frequency Separation 0.030		
Raw Data	0.229 (0.052)	0.231 (0.053)
Recovered High	0.567 (0.321)	0.391 (0.153)
Recovered Low	0.016 (<0.001)	0.910 (0.828)

Results indicate that in both scenarios, the raw data did not fit either the high or low frequency component well, with correlations of approximately 0.23 and explaining approximately 5% of the variation of each of the component signals. This is not surprising given that the raw data was composed of both signals but only in the presence of severe noise and missing observations. This illustrates the challenge in the analysis. The recovered signals were an improvement in fit to their respective targeted true signals in both scenarios. However, the recovered high and low signals modeled the true high and low frequency signals better when the signal separation increased. The recovered high frequency signal correlation improved from 0.463 to 0.567, and explained more than ten percent additional variation in the true high frequency signal. The success of the recovered low frequency signal was even more striking, with correlation improving from 0.463 to 0.910, and explaining over eighty percent of the variation in the true low frequency signal. The disadvantage in scenario one, where frequency separation is below the Proposition 2 minimum, is that the recovered signal for a given targeted frequency had higher fit statistics to the wrong component frequency when the signal separation was low. The recovered low frequency had higher correlation with the true high frequency component in scenario one, and the recovered high frequency had much higher correlation with the true low frequency component, 0.453, almost as high as the correlation with the true high frequency component that was the target. This indicates that signal separation was poor and prone to model misspecification.

IV. CONCLUSION

This study illustrates the importance of understanding the applications and limitations of Kolmogorov-Zurbenko (KZ) filters. It extends the theory of separating component signals by proving that under suitable circumstances, KZ filters can separate any two or more frequencies of interest. The study develops propositions to guide 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. This helps to understand what questions these tools can answer retrospectively given a set of data, and assists the design of future research with an increasing reliance on this class of filters as a primary investigatory tool.

Noise exceeding many times the strength of component signals as well as missing data rates of half or more of all observations may seem impossible obstacles with other statistical analysis techniques. The simulations in this study were not only intended to illustrate the use of KZ filters to handle these difficulties, but were intentionally designed with scenarios chosen to stress KZ methods for signal separation. The purposely low number of simulated observations coupled with smaller and then greater signal separation, as guided by the previous propositions, illustrated signal reconstruction where data was theoretically inadequate verses minimally sufficient. Correlation analysis between the reconstructed signals and the original component signal in both scenarios indicates the superiority of KZ filters where there is sufficient data to more effectively separate frequencies. In practice, far greater numbers of observations are desirable, several multiples of the longest signal period. Still, in these challenging simulated conditions the successful signal reconstruction and quality of fit was visibly and measurably evident.

The use of KZ filters has increased as demand increases to meet statistical analysis challenges such as multidimensional spatial and temporal data analysis, large random errors, high rates of missing observations, signal interference, and situations where other statistical analysis methods are inadequate. As these statistical analysis tools find use in a variety of scientific fields the theoretical results developed here are necessary to ensure performance of Kolmogorov-Zurbenko filters.

REFERENCES

[1] Zurbenko, I. (1986). The Spectral Analysis of Time Series. North-Holland. 248 pages

[2] Yang, W., & Zurbenko, I. (2010). Kolmogorov-Zurbenko Filters. Computational Statistics, 2(3), 340-351. DOI: 10.1002/wics.71

[3] Wise, E. K., & Comrie, A. C. (2005). Extending the Kolmogorov–Zurbenko filter: application to ozone, particulate matter, and meteorological trends. Journal of the Air & Waste Management Association, 55(8), 1208-1216. DOI: 10.1080/10473289.2005.10464718

[4] 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

[5] Anh, V., Duc, H., & Azzi, M. (1997). Modeling anthropogenic trends in air quality data. Journal of the Air & Waste Management Association, 47(1), 66-71. DOI: 10.1080/10473289.1997.10464406

[6] Valachovic E, Zurbenko I (2014) Skin Cancer, Irradiation, and Sunspots: The Solar Cycle Effect. Biomed Res Int 2014:1-9. doi:10.1155/2014/538574.

[7] Gupta, I., Salunkhe, A., & Kumar, R. (2010). Modelling 10-year trends of PM 10 and related toxic heavy metal concentrations in four cities in India. Journal of hazardous materials, 179(1), 1084-1095. https://doi.org/10.1016/j.jhazmat.2010.03.117

[8] Zurbenko, I., & Ming, L. (2017). KZ Spatial Waves Separations. Journal of Research in Applied Mathematics, 3(4), 1-7.

[9] Valachovic, E. & Zurbenko, I. Multivariate analysis of spatial–temporal scales in melanoma prevalence. Cancer Causes Control (2017) 28: 733. https://doi.org/10.1007/s10552-017-0895-x

[10] Tsakiri, K.G. & Zurbenko, I.G. Explanation of fluctuations in water use time series. Environ Ecol Stat (2013) 20: 399. https://doi.org/10.1007/s10651-012-0225-0.

[11] 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/.

[12] Close, B., & Zurbenko, I. (2013). Kolmogorov-Zurbenko adaptive filters. (Version 3). Retrieved from http://cran.r-project.org/web/packages/kza/index.html.

Edward Valachovic*. "Separation of Spatio-Temporal Frequencies." Quest Journals Journal of Research in Applied Mathematics , vol. 03, no. 07, 2017, pp. 29–37.