Quest Journals
Journal of Research in Applied Mathematics
Volume 3 ~ Issue 4 (2017) pp: 01-07
ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

KZ Spatial Waves Separations

Ming Luo, Igor Zurbenko*
Department Of Epidemiology and Biostatistics, State University of New York At Albany, Rensselaer, USA

Received 22 Dec, 2016; **A**ccepted 06 Jan, 2017 © The author(s) 2017. **P**ublished with open access at **www.questjournals.org**

ABSTRACT: *In this paper, we proposed a new identification algorithm based on Kolmogorov–Zurbenko Periodogram (KZP) to separate motions in spatial motion image data. The concept of directional periodogram is utilized to sample the wave field and collect information of motion scales and directions. KZ Periodogram enables us detecting precise dominate frequency information of spatial waves covered by highly background noises. The computation of directional periodogram filters out most of the noise effects, and the procedure is robust for missing and fraud spikes caused by noise and measurement errors. This design is critical for the closure-based clustering method to find cluster structures of potential parameter solutions in the parameter space. An example based on simulation data is given to demonstrate the four steps in the procedure of this method. Related functions are implemented in our recent published R package {kzfs}.*

Keywords: *KZ Periodogram, directional periodogram, parameter identification, spatial wave separations, closure-based clustering, parameter clusters, inverse problem.*

I. INTRODUCTION

Motion image identification in different types of data is very important subject in many applications. Those images may depend on time and contain different scales. The simplest example is waves in the ocean coming from two different directions. One wave can be strong long scale, and another is shorter scale wave propagating in different direction. When both are covered by strong noise, data realization could be very noisy 3D structure. Similar examples can be found in engineering, acoustics, astronomy, design of audio halls, climate control, oceans waves alarm systems, Tsunami-waves prediction, and many other fields.

This paper aims to the separation of motion scales in 2D motion images on different directions. To this end, we utilize Kolmogorov–Zurbenko Periodogram (KZP) [1-3] as the tool to detect precise spectral signals from noise-covered spatial/temporary data. The concept of directional periodogram is introduced based on KZP and used for recording the direction and frequency information of spatial waves. In the third section, we will discuss a novel motion scale parameter identification algorithm based on directional periodogram, the closure-based clustering method. A simulation example is exhibited to show the procedure of this method. The summary section discusses the advantage and limit of this approach.

II. KOLMOGOROV–ZURBENKO PERIODOGRAM

Kolmogorov–Zurbenko Periodogram (KZP) is designed to detect periodic signals or seasonality covered by heavy noise. It has a sharp frequency resolution for capturing frequency of interest, and provides practically no spectral leakage from side lobes. In fact, KZP had the nearest to the optimal mean square error in the estimation of power spectrum [1, 2]. It can stable the variance of the periodogram, and permits the separation of two signals on the edge of a theoretically smallest distance.

Definition 1: For a sample of series $\{X(t)\}$, t = 0,1, ... , N - 1, the KZ Periodogram is:

$$KZP(t, m, k, \nu_0) = \frac{1}{S\rho_0} \sum_{\tau=-S\rho_0}^{S\rho_0} |\, KZFT_{m,k,\nu_0}[\mathrm{X}(\tau + t)]\,|^2$$

where $KZFT_{m,k,\nu_0}[X(t)]$ is given by

$$KZFT_{m,k,\nu_0}[X(t)] = \sum_{s=-k(m-1)/2}^{k(m-1)/2} X(t+s) \times a_s^{m,k} \times e^{-i(2\pi\nu_0)s}$$

*Corresponding Author: Igor Zurbenko, E-mail: izurbenko@albany.edu
Department Of Epidemiology and Biostatistics, State University of New York At Albany, Rensselaer, USA

$$a_s^{m,k} = \frac{c_s^{k,m}}{m^k}, s = \frac{-k(m-1)}{2}, ..., \frac{k(m-1)}{2},$$
$$\sum_{r=0}^{k(m-1)} z^r c_{r-k(m-1)/2}^{k,m} = (1+z+\cdots+z^{m-1})^k, \rho_0 = 1/\nu_0.$$

KZP can be viewed as the iteration of average for the regular periodogram in time over S periods ρ_0 based on ($2S\rho_0 + 1$) observations around moment (or position) t. Here S is roughly half of the pre-selected window size in periods for KZP; therefore the averaging window includes more points in the low frequency region. Please note that $KZP\ (t, m, k, v_0)$ is for power periodogram value on a specific frequency v_0. The power distribution for an frequency range, i.e. $KZP\ (t, m, k)$, is the aggregation of a series of $KZP\ (t, m, k, v_i)$ for v_i in the frequency series $\{v_1, \ldots, v_n\}$ covered this range.

KZP is based on small window Fourier analyses. Theoretically, their major advantage is its suitability for scenarios of non-stationary process, or data with long series. However, recent applications [4-6] found that it can also be applied with relative short series under highly noise and missing values. For short data series, the accuracy of periodogram is limited by the length of the available data. One "work-around" method is to sample n times in the spectral analysis, but it may lead to oscillations in the regular periodogram. KZP suppresses the oscillation and stabilize the periodogram in a large extent.

We can further improve the accuracy of detected dominant frequencies $\{\nu_d\}$ (or their modes) by searching for the local maximum KZP values. Based on the definition of KZP, the large magnitude values of periodogram usually are around the dominant frequencies $\{\nu_d\}$ of the input data series. The dominant frequencies not necessarily are sharply cut single spikes; they could be energy distributed in a narrow frequency range. But usually there is a local maximum (or mode) for such energy clusters. In practice, we can first compute the KZP for a series of frequency $\{v_s\}$ with relative large interval; then search around the dominate frequencies for higher resolution result. Generally speaking, this is a one-dimensional optimization process. The initial values can be set as the dominate frequencies detected by multiple-sampling KZP. After optimization, in many cases, the accuracy of KZP is less related to available data length but rather the measurement errors in the data series.

III. DIRECTIONAL PERIODOGRAM

Suppose we are interested in checking the spectral behavior for data series along a given direction θ in a wave field. In the general situation, a data series $\{d_i\}$ on a line along angle θ is a sample of the wave field. Then the periodogram of arbitrary $\{d_i\}$ can be represented by a function $KZP_\theta\,(t, v)$, where v is the frequency, and T = $\{t\}$ is a finite indexed set for the sampling space of the wave field. Usually T can be taken as a series of points on the x- or y-axis. Respectively, for any fixed t_0 and v_0, $KZP_\theta\,(t_0, v_0)$ is the estimation of spectral density on frequency v_0 for data series passing through the point $(t_0, 0)$ or $(0, t_0)$.

Definition 2: For spatial wave $w(x, y)$ on a wave field $\mathbf{D} = \{(x, y)\}$, $|\ x\ | \le dx$, $|\ y\ | \le dy$, its directional periodogram for a given direction angle θ is the ensemble average of $KZP_\theta\,(t, v_0)$ on frequency v_0 for all data series $\{d_i\}_t$ on the parallel lines along direction θ and projected on x- or y-axis, where $t \in$ T, $\|\mathrm{T}\| < \infty$, $\|\{d_i\}_t\| < \infty$. That is,

$$KZP_\theta\,(v_0) = \mathrm{E}\,[KZP_\theta\,(t, v_0)] \approx \frac{1}{\|T\|} \sum_{t \in T} KZP_\theta(t, v_0) \tag{1}$$

Here T is the index set for parallel lines of the sampling space. In practice, the length of sample series on the line indexed by t, i.e. $\|\{d_i\}_t\|$, should be larger than a minimum value m_{ln}. Usually, m_{ln} is given as a percentage of the largest series length for a specific direction.

Proposition 1: For signal $w(x, y)$ propagated along direction β with frequency f, v_d is the dominated frequency on directional periodogram of direction θ, then we have

$$\mathrm{E}\,(v_d) = f \cdot |\cos(\theta - \beta) / \cos(\theta)|, \text{ for data series projected on } x\text{-axis} \tag{2a}$$
$$\mathrm{E}\,(v_d) = f \cdot |\cos(\theta - \beta) / \sin(\theta)|, \text{ for data series projected on } y\text{-axis} \tag{2b}$$

The proof is given in the Appendix. In practice, to avoid infinite values of v_d, the following protocol is adopted as default: for $\theta \le$ p/4 or $\theta \ge 3\pi/4$, project data series $\{d_i\}_t$ on x-axis; otherwise, project $\{d_i\}_t$ on y-axis. For wave signals with homoscedastic Gaussian noises, since the noises spectrum are uniformly distributed on the whole frequency range, it is easy to see that Proposition 1 is still true if the noise is less than a certain level.

Please note that $KZP_\theta(v_0)$ is applicable for continuous and discrete data (regular- or irregular-sampled). In the following, we will focus on the situation in which data are collected (or aggregated) on a pre-defined grid on the wave field. Don't lost generality, we denote the grid as $\mathbf{G} = \{(x, y)\}$, where $x = 0, 1, ..., dx\text{-}1$, $y = 0, 1, ..., dy\text{-}1$, dx and dy are integers. We can also set $T \subset \{0, 1, \ldots, dx\text{-}1\}$ or $T \subset \{0, 1, \ldots, dy\text{-}1\}$, depending on sampled data series $\{d_i\}_t$ are projected on x- or y-axis. This actually requires that each sample line passes through (at least) one grid point on x- or y-axis. Here T is a subset of all possible parallel lines for a given sampling direction; it is selected to reflect the spectral feature of wave signals.

IV. IDENTIFY WAVE PARAMETERS WITH DIRECTIONAL PERIODOGRAM

Suppose a group of wave $\{X_1, \ldots, X_n\}$ propagated along unknown direction $\{\beta_1, \ldots, \beta_n\}$ with unknown frequency $\{f_1, \ldots, f_n\}$ in the wave field, $i = 1, \ldots, n$. The dominated frequency on directional periodograms of direction angle θ are observed as $v_1, \ldots, v_n$. Then we have a group of n equations in the form of eq. 2a or 2b. This n equation system contains $2n$ unknown parameters in pairs of (f_i, β_i). We may want to introduce more directional periodograms and make it "overdetermined", then find the solution with optimization. However, since the projection of mapping spectral spikes onto different directions doesn't keep the frequency order, we lost the information to match the observed spikes with frequency parameters. Additional sampling will add n equations and n new unknown variables. The traditional approach for inverse problems is not feasible.

As a basic fact, we have n^2 possible combination of (v_i, f_i), $i = 1, 2, ..., n$, for directional periodogram observations from each pair of sampling directions, and it leads to $2n^2$ possible solutions of (f_i, β_i); moreover, it probability needs to include some potential parameters caused by aliasing. However, for k pairs of sampling directions, the "real" wave parameters should appear in almost all k potential solution groups. The only exception is for the case with sampling direction orthogonal to the wave's direction of propagation, in which dominate frequency $v_d = 0$ and therefore couldn't be detected. In the following, we will develop this intuitive idea into a new procedure to identify spatial wave parameters.

On the wave parameter plane, if two points (f_i, β_i) and (f_j, β_j) are close enough, i.e., suppose $| f_i - f_j | < t_f$ and $| \beta_i - \beta_j | < t_\beta$, then they are called to be in the same cluster. Points in cluster m ($m = 1, ..., n$) are viewed as different measurements of the same parameters (f^m, β^m), i.e., they are practically equivalent in the tolerance range. Assuming Gaussian measurement errors, we introduced two random variables for each cluster.

$$f_i^m = f^m + e_f, \quad e_f \sim N(0, \sigma_f^2) \tag{3a}$$

$$\beta_i^m = \beta^m + e_\beta, \quad e_\beta \sim N(0, \sigma_\beta^2) \tag{3b}$$

Corresponding to n spatial waves, there are n parameter clusters, for which each have about k points inside. Outside of the clusters, more than $2kn^2 - kn$ points are separately distributed on the frequency-direction plane. Our task is to identify these n clusters based on this model.

We developed a new procedure for this unusual clustering problem. The idea is that, in the general conditions, the k points in the same cluster will form a closure of k-nearest neighbour with a large probability. Even if only a part of the n clusters can be identified with k-nearest neighbour closure, the tolerances of these identified clusters can be used as reference for other clusters. This is actually the estimation of σ_f and σ_β, and the cluster tolerance can be set as $t_f = c_f \cdot \sigma_f$ and $t_\beta = c_\beta \cdot \sigma_\beta$, where c_f and c_β are constants. Once the estimation of tolerance has been given, identification of clusters is straight forward.

There are 4 steps in our procedure of wave parameter identification.

1. **Sampling on orthogonal direction pairs**
   - Calculate directional periodogram on the discrete frequency series
   - Record periodogram spikes for each direction on discrete frequency series
   - Search for local maximum periodogram values on aperiodic spectrum
   - Find the potential solutions for each orthogonal directional periodogram pairs
2. **Identify parameter clusters on the frequency-direction plane**
   - Estimate n and k, and find the closure of k-nearest neighbor
   - Estimate the tolerance level of identified clusters
   - Check potential clusters based on estimated tolerance level
3. **Estimate wave parameters**
   - Exclude unlikely points from cluster and clusters with low supports
   - Estimate parameters, output plots and suggestion
4. **Validation of results**
   - Check consistency for estimations of different tolerance levels
   - Cross-validation by excluding one or more periodogram observations

The first step is collection of directional periodogram observations based on discrete frequency series. This step is time-consuming, especially when the wave field is large. Here, the condition of orthogonal direction is required for the accuracy of detected location of intersection point of two sampling lines. The dominate frequencies for periodogram observations based on discrete frequency series will be recorded. You need at least 3 orthogonal direction pairs to go to further steps. Then we search for the dominate spikes in aperiodic spectrum by maximizing the power periodogram. We applied golden section search and successive parabolic interpolation in this step [7]. It is critical for the accuracy of the directional periodogram and the wave parameters.

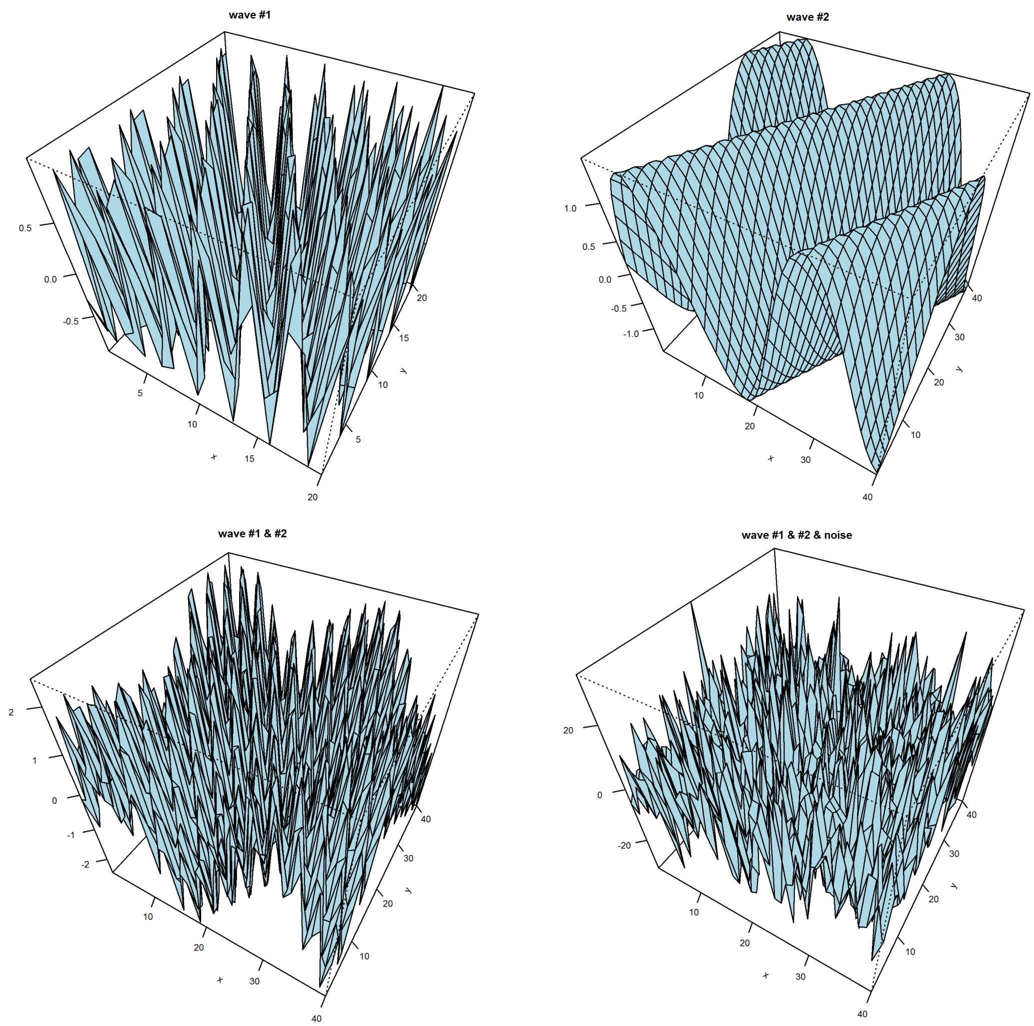

Figure 1. Wave shapes of two signals (partial)

Upper left: wave 1 without noise; upper right: wave 2 w/o noise; Bottom left: wave 1 + 2 w/o noise; bottom right: wave 1 + 2 + noise; Signals: *sin* $(2\pi\cdot 0.4\cdot t)$ along 30°; 1.5*sin* $(2\pi\cdot 0.05\cdot t)$ along -30°; noise: N (0,100)

The second step is to identify parameter clusters on the frequency-direction plane. The first work of this step is to estimate k and n: estimation of n is based on the mode of numbers of frequency spikes on each directional periodogram; k is the number of orthogonal pairs of sampling directions. Then the algorithm will list all k-nearest neighbors for each point of possible solutions, and search for the closure structures: the set union of the k-nearest neighbor for a set of k points contains nothing but themselves. Each k-nearest neighbor closure is a cluster; its tolerance will be utilized to search for other parameter clusters.

The point number in a cluster is called the support of this cluster. These supports must come from different orthogonal pairs. The expected support for a cluster should be k in general, or $k - 1$ if there is sampling direction that is orthogonal to the wave direction of this cluster. Clusters with low supports will be excluded, too. When k is small, say $k < 10$, the wave parameter is estimated with median; otherwise average value is used.

The last step will check on different tolerance levels for validation. The cross-validation procedure will exclude one or more pairs of orthogonal directional periodogram and check changes in the results. It's designed for the occasionally missed spectral spikes caused by background noises. If validation shows inconsistent results, go back to step 1 and increase number *k*. The estimation will become stable with increasing of *k*.

V. EXAMPLE OF WAVE PARAMETER IDENTIFICATION

Suppose there are 2 waves propagate along direction 30° and -30° in a 400×400 wave field. Their frequencies are 0.4 and 0.05; amplitudes are 1 and 1.5, respectively. The noise is $N(0, 10^2)$ (see Figure 1).

Sampling Directions (°)	Discrete Directional Periodogram		Optimized Directional Periodogram	
	Frequency	Periodogram	Frequency	Periodogram
15	0.0375	224.6917	0.03653	325.7852
15	0.4000	201.5241	0.40000	201.5241
-15	0.0500	297.8649	0.05000	336.3946
-15	0.2925	181.2478	0.29277	193.8835
75	0.0125	234.7062	0.01347	325.7361
75	0.2925	165.9792	0.29272	175.0699
-75	0.0375	242.9188	0.03658	321.1320
-75	0.1075	171.4818	0.10714	186.5533
-45	0.0700	181.9683	0.06824	272.3969
-45	0.1475	148.9511	0.14632	164.1086
45	0.0175	270.0084	0.01818	297.9743
45	0.4550	174.7875	0.45340	195.8976
90	0.0250	305.2746	0.02500	305.2746
90	0.2000	216.6582	0.20000	216.6582
0	0.0425	242.7422	0.04240	338.4799
0	0.3475	159.6071	0.34637	189.0890

Table 1. Frequency spikes of 8 directional periodograms of 4 orthogonal direction pairs

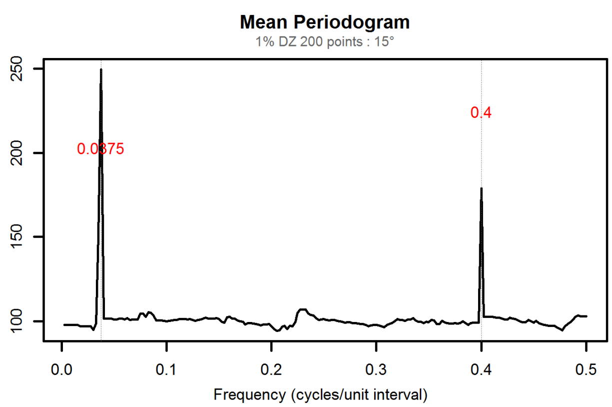

Figure 2. Directional Periodogram along 15°
Signals: $sin(2\pi\cdot 0.4\cdot t)$ along 30°; $1.5sin(2\pi\cdot 0.05\cdot t)$ along -30°; noise ~ $N(0, 100)$;

We sampled on directions pairs of (0°, 90°), (-45°, 45°), (-15°, 75°), (-75°, 15°). Figure 2 is an example of these directional periodograms. For all available 8 directional periodogram (see table 1), the expected number of wave parameter clusters is 2. Table 1 also lists the dominate frequencies and their periodogram values based on discrete series and optimized dominate frequencies. Most of the noise effects have been filtered out by the periodogram. For the coarse frequency records, the algorithm suggests f_1 = 0.0504 and β_1 = -30.23°, f_2 = 0.3998 and β_2 = 29.96°; while based on the optimized values, the results are f_1 = 0.0499 and β_1 = -30.08°, f_2 = 0.4000 and β_2 = 30.00°. The accuracy is improved after optimization.

The wave parameter clusters is visualized with Figure 3. It shows 2 clusters with 4 supports, and 2 clusters with 2 supports. If the expected cluster number is known, it may be good enough to identify the 2 clusters with 4 supports as the wave parameters. But if the cluster number is unknown and we want to put the identification in the frame of statistical hypothesis test, we may need more data to make the decision.

Figure 4 is for another run of the parameter identification procedure on the same two wave signals. This time we include 13 sampling pairs. As we can see, the gap between the support numbers of identified clusters and the other potential clusters are enlarged to 12 – 6 = 6. Therefore, we have more confidence to decide if the detected clusters are from the real wave parameters. The estimations are consistent with previous result: before optimization, we get f_1 = 0.0499, β_1 = -30.22°, f_2 = 0.4004, β_2 = 30.04°; after optimization, the result is f_1 = 0.0499, β_1 = -30.06°, f_2 = 0.4000, β_2 = 30.00°.

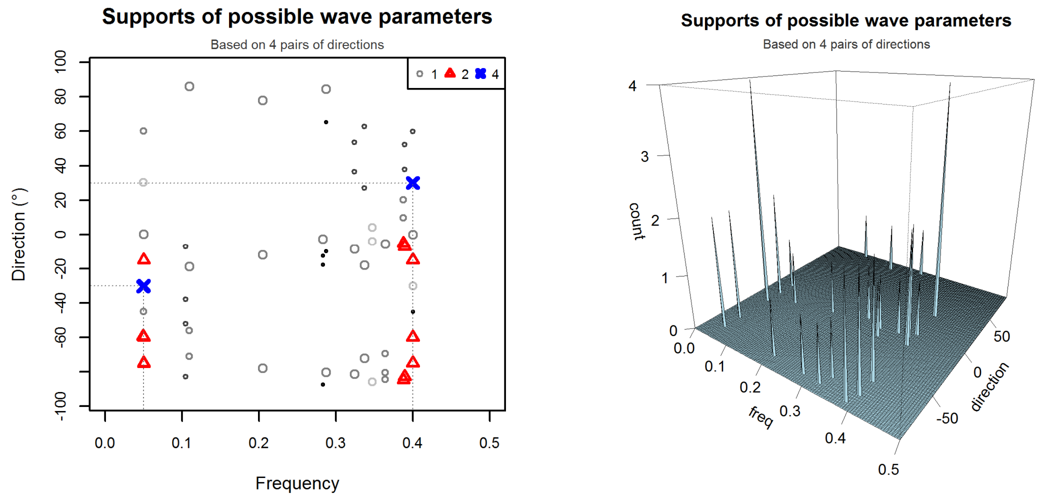

Figure 3. Supports for identified wave parameter clusters based on 4 orthogonal sampling pairs

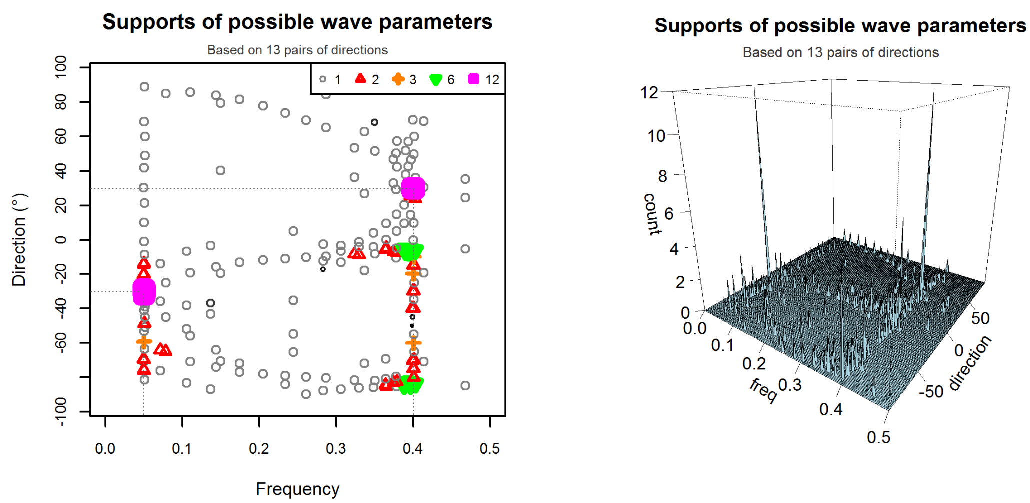

Figure 4. Supports for wave parameter clusters based on 13 orthogonal sampling pairs

VI. SUMMARY

This paper introduced directional periodogram based on the definition of KZP. Briefly, directional periodograms are averaging of KZP for all parallel sampling lines along a given direction. Averaging helps to stable the variance of periodogram and filter out a large part of noise effects. It has been showed that the dominate spikes on directional periodogram are the function of the sampling direction, the wave frequency and direction. This is the base for KZ spatial wave separation.

For the task of wave parameter identification, we proposed the algorithm of closure-based clustering plus tolerance-based clustering method. The algorithm is designed to resist incorrectly identified or missed periodogram signals caused by noises, and it gives consistent estimations when the number of sampling directions increases. It works well for spatial waves with sinusoidal signals or single-mode spectrum, and usually requires that the wave signals are “stationary” in the wave field.

This algorithm has been realized in our R package {kzfs}. This package is designed for the separation and reconstruction of motion scales in 2D motion images on different directions based on KZP and KZFT. For the wave parameter identification, {kzfs} provides functions to check directional periodograms for spatial waves in the wave field. It helps to automatically identify and mark prominent spectrum spikes of periodograms. Functions are provided to support all four steps of the identification process, and can be used combing with the support of KZ adaptive filters {kza} [9]. For signal reconstruction, {kzfs} utilizes KZFT to provide accurate recovered signals under several times of noises with correlation coefficients > 80% for mixed multiple signals, and > 90% for separation of dominated wave patterns.

REFERENCE

[1]. I. Zurbenko. *The Spectral Analysis of Time Series*. North-Holland Series in Statistics and Probability, 1986.

[2]. W. Yang and I. Zurbenko. Kolmogorov–Zurbenko filters. *WIREs Comp Stat*, 2:340–351, 2010.

[3]. A. DiRienzo, I. Zurbenko. Semi-adaptive nonparametric spectral estimation. *Journal of Computational and Graphical Statistics* 8(1): 41-59, 1998.

[4]. I. Zurbenko, M. Luo. Restoration of Time-Spatial Scales in Global Temperature Data. *American Journal of Climate Change*, 1(3): 154-163, 2012. doi:10.4236/ajcc.2012.

[5]. I. Zurbenko and A. Potrzeba-Macrina. Tides in the atmosphere. *Air Quality, Atmosphere and Health*, 6(1), 39–46, 2013.

[6]. I. G. Zurbenko, M. Luo. Surface Humidity Changes in Different Temporal Scales. *American Journal of Climate Change*, 4(3): 226-238, 2015. doi:10.4236/ajcc.2015.43018

[7]. R. Brent. *Algorithms for Minimization without Derivatives*. Englewood Cliffs N.J.: Prentice-Hall. 1973.

[8]. M. Luo, I. Zurbenko. Multi-Scale Motions Separation with Kolmogorov-Zurbenko Periodogram Signals. 2016.9.23. R package *kzfs*. https://CRAN.R-project.org/package=kzfs

[9]. B. Close, I. Zurbenko, Kolmogorov-Zurbenko adaptive filter. 2016.3.24. R package *kza*. https://CRAN.R-project.org/package=kza

Appendix: Proof of Proposition 1

Proof: Don't lose any generality, suppose wave $w(x, y)$ propagates along direction angle β, as showed in Figure A. The wave front starts from the dash line passed through point O to the dash line passed through point A, B, and C; the coordinate of O is (0, 0). The segment between the two dash lines is OB = d; and the angle of OBC is 90°. Assume the directional periodogram is checked along direction of angle θ. The length of line segment OA is $d / cos(\theta - \beta)$. Denote the coordinate of point A as (x_A, y_A), then we have:

$$x_A = d \cdot cos(\theta) / cos(\theta - \beta) \tag{A.1a}$$

$$y_A = d \cdot sin(\theta) / cos(\theta - \beta) \tag{A.1b}$$

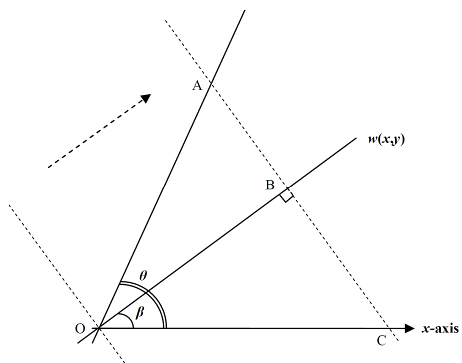

Figure A. Wave signal $w(x, y)$ between two dash lines

From the perspective of x-axis, there are $|x_A|$ unit intervals along the line segment of OA. In other words, $|x_A|$ is the segment length when project OA onto the x-axis. Since wave signal $w(x, y)$ has the same phase at point A, B, and C, line segment OA correspond to $d{\cdot}f$ wave periods, as it is on the line segment of OB. This means that the wave frequency is

$$f_{dx} = d{\cdot}f / |x_A| = f \cdot |cos(\theta - \beta) / cos(\theta)|.$$

Similarly, from perspective of y-axis, there are $|y_A|$ unit intervals corresponding to $d{\cdot}f$ periods of wave, and the frequency is

$$f_{dy} = d{\cdot}f / |y_A| = f \cdot |cos(\theta - \beta) / sin(\theta)|.$$

It is easy to see that this relationship holds for any continuous wave signal $w(x, y)$ along direction θ. For any sampling series on a parallel line along direction θ, its KZP is the consistent estimator of $w(x, y)$ spectrum along this direction when the sampling frequency is larger than the Nyquist frequency [1]. Then,

$$E(v_{dx}^{t}) = f \cdot |\cos(\theta - \beta) / \cos(\theta)|, \text{ for data series projected on } x\text{-axis}$$

$$E(v_{dy}^{t}) = f \cdot |\cos(\theta - \beta) / \sin(\theta)|, \text{ for data series projected on } y\text{-axis}$$

where $t \in T$, T is the index set for the parallel lines along direction θ in the sampling space. This equation is also true for the ensemble average of KZP of data series indexed by a finite set T, which is defined as the directional periodogram on direction θ for signal $w(x, y)$. Then we finish the proof of Proposition 1.

*Corresponding Author: Igor Zurbenko