Unmixing: A New Direction from Classical Tidal Harmonic Analysis?

Yu Li, Shengdong Xie, Bikong Wang, Shijun He, Dongmei Huang
College of Information
Shanghai Ocean University
No. 999, Hu Cheng Ring Rd, Lin Gang Xin Cheng
Shanghai, China
heshijun6@163.com

Haiyang He
Information Science and Engineering
East China University of Science and Technology
Shanghai, China
Hairong Qi
Electrical Engineering and Computer Science
University of Tennessee
Knoxville, TN 37996
hqi@utk.edu

***Abstract*—Harmonic analysis has been the conventional method for tidal prediction where seven (or more) most important constituents/harmonics are modeled and separately predicted whose aggregates have shown to be able to accurately reconstruct the observed tidal data. However, harmonic analysis also has its potential drawbacks. First, it can only accurately predict the astronomical tide. Non-sinusoidal fluctuations over time scales on the order of hours have been observed but cannot be reconstructed through harmonic analysis. Second, local astronomical tides are best predicted when based on data collected for at least 18.6 years, which is rather a rigid requirement at places without such long period observation data. Third, for tides along a complex coastline, more than 100 components must be considered in order to predict accurately. Finally, it has been shown that harmonic analysis is not effective in extreme weather conditions, e.g., those with tropical storms. In this paper, we propose a deviation from conventional tidal analysis approaches and investigate into the problem from the perspective of "signal unmixing", where we interpret the observed data as a linear combination of constituents and apply robust *unsupervised* unmixing algorithm, referred to as the minimum-volume-constrained nonnegative matrix factorization (MVC-NMF), to decompose the observation into a set of source signals (i.e., the harmonics). The unmixing–based tidal analysis is fundamentally different from harmonic-based analyses by not limiting itself to just astronomical tide, and being unsupervised thus not requiring long-term training. Preliminary experiments are conducted to study the feasibility of the proposed approach for tidal analysis.**

Keywords—signal unmixing; harmonic analysis; tidal analysis; unsupervised unmixing

I. Introduction

Tidal analysis and prediction is essential from a number of aspects, including safe navigation of ships, coastal development, and fishing and recreational boating and surfing [11]. In addition, since tides are more predictable than wind energy and solar power, tidal power has potential for future electricity generation [12]. In [6], the authors also nicely summarized recent research findings related to the significant impact of tides to the marine life. It has been proved that the ocean micro-structure has high correlation with the turbulent dissipation and the tidal cycles [7][8]. In coastal areas, tides contribute significantly with the vertical mixing, redistributing nutrients and oxygen, that impact marine life and also, with the flow of CO_2 between sea and air [9][10]. More importantly, with accurate and in-time prediction of wave height during storm surge, many damages of the surge to the human society can be alleviated.

For a long time, tidal prediction has been adopting the harmonic analysis method where the complex tidal curve is decomposed into a number of harmonic terms and the changes of the harmonic constants and astronomical effects are calculated for prediction purpose.

Although popular, harmonic analysis-based tidal prediction does carry a few deficiencies. First, the prediction accuracy is tightly coupled with the number of harmonics (or constituents) used. The more harmonics used, the higher the accuracy. However, as the number of harmonics increases, the computational complexity of the prediction process also increases. And the increase rate of computational complexity probably is much faster than that of the number of harmonics. This is why many times only seven most important constituents are adopted. However, the joint effect of the neglected constituents may result in large prediction error. In [6], the authors also point out that existing numerical models for coastal and shallow waters still are not fully satisfactory, mainly due to the reduced number of tidal constituents used.

Second, harmonic analysis cannot predict the impact of non-cyclical factors [2]. Some non-cyclical factors, such as typhoon and precipitation, have great influence on the prediction of the tides. Conventional tidal prediction only works for the astronomical tide after removing the storm surge (tsunami), where only the relationship between astronomical movement in day, the earth and the moon and the regularity of the tidal rise, are taken into consideration. Therefore, the predicted values naturally cannot correspond to the observed water level in the tide gauge station, which leads to the uncertainty of the forecast results.

Finally, in order to obtain the right set of parameters for the harmonics, at least 18.6 years of data need to be collected [3], which becomes inconvenient for places that have not yet collected such long period observation data.

978-0-933957-43-5 ©2015 MTS

In this paper, we exploit the potential of the "unmixing" technique where we interpret the observed data as a linear combination of constituent elements and apply robust *unsupervised* unmixing algorithm to extract these elements as well as the mixing coefficients. This unmixing-based methodology is the key enabler to go beyond what are immediately detectable, separable, and extractable in the data, bringing unprecedented performance levels to tidal analysis.

The proposed tidal unmixing approach is fundamentally different from the harmonic analysis in three aspects. First, it does not differentiate between astronomical tide and non-astronomical tide, but instead, treat the observed data as a whole, and thus does not require separate handling of different types of tides. Second, the approach is unsupervised in the sense that it does not require prior knowledge of the source signatures. For normal weather conditions, we can imagine that the source signatures would align very well with the seven most important harmonics used in harmonic analysis. However, the unmixing method, due to its formulation, is flexible enough to encompass other source signatures that also play an important role in reconstruction of the observation data. This is extremely powerful when little prior information is known under extremely weather conditions. Third, the unmixing approach does not need long-term observation data as in conventional harmonic analysis that facilitates online tidal prediction.

The rest of the paper is organized as follows: Section II describes the conventional harmonic analysis technique. Section III discusses the unsupervised unmixing algorithm adopted. Section IV describes experiments conducted, data sets used, as well as results obtained. Section V concludes the paper.

II. Harmonic Analysis

The fundamental of the harmonic analysis is to establish a tidal model for the pure harmonic expression of the tidal model. By the least square adjustment principle, the amplitude and the late phase of the sub tidal wave of the model are obtained.

Tidal changes depend on the relative position of Earth, the Moon and the Sun. According to the law of universal gravitation, the expression of the tide, after the decomposition of the Moon's balance, can be obtained in the following form:

$$\zeta_{月}=(3/4)(M/E)(a/D)^3 a \{[1/2-(3/2)\sin^2\varphi] (2/3-2\sin^2\delta) + \sin2\varphi\sin2\delta\cos T_1 + \cos^2\varphi\cos^2\delta\cos2T_1\} \quad (1)$$

where E, M are the mass of the Moon and the Earth, respectively, a is the average radius of the Earth, D is the center distance between land and Moon, φ is the geographic latitude, δ is the lunar bare, T_1 is the angle of the moon.

For the variables in Eq. 1, when the angle of the Sun and the Moon is replaced by the longitude and the solar time, and import the assistant spring equinox, we can obtain a number of major harmonic terms, that are referred to as the sub-tides. For these harmonic terms, the coefficient determines the amplitude and phase determines the tidal cycle, In reality, due to the existence of the flow inertia and friction, the peak often appears behind when it is supposed to. Therefore, simplifying Eq.1, we have:

$$\zeta=fH\cos(\sigma t+v_0+u-K) \quad (2)$$

where H is the average amplitude, K is the local late angle. Since variables H, K are determined by the geographical location, to a fixed point, these variables are approximately constant, and therefore referred to as the tidal harmonic constants.

The general expression for the sub-tide is

$$\zeta=A\cos\sigma t+B\sin\sigma t \quad (3)$$

Since $A=R\cos\theta_0$, $B=R\sin\theta_0$, $R=(A^2+B^2)^{1/2}$, $tg\theta_0=B/A$, so $H=R/f$, $K=tg^{-1}(v_0+u-\theta_0)$. It can be seen that the solution of tidal harmonic constant can be converted into solving the A, B coefficient through astronomical variables.

III. Unsupervised Signal Unmixing

Over the years, many approaches have been proposed to combat the deficiencies of conventional harmonic analysis, including, for example, neural network-based prediction for non-astronomical tides while keeping harmonic analysis for the astronomical tide prediction [1], support vector machine-based analysis to take advantage of its non-linear representation capability of data [5], adapting the classical tidal harmonic analysis to non-stationary tides like river tides [4], etc.

In this paper, we propose a deviation from conventional tidal analysis approaches and investigate into the problem from the perspective of "unmixing".

Mixed measurements are frequently encountered in real-world applications. Due to the resolution issue associated with discrete sampling and the effect of unknown sources, the measurements can rarely be pure. The existence of mixed measurements has brought the decomposition or unmixing technique to a wide array of applications. For example, in the famous cocktail party problem, the listening attention should be focused on a single talker among a mixture of conversations and background noises [13]. In the field of remote sensing, due to the large footprint, a single pixel usually covers more than one type of ground constituent. Thus, the measured spectrum at a single pixel is a mixture of several ground cover spectra, where pixel unmixing has been applied to subpixel object quantification [14], mineral identification [16], area estimation [15][17], etc.

A. The Linear Mixing Model

In tidal unmixing, we interpret the observed data as a linear combination of constituent elements and adopt robust *unsupervised* unmixing algorithms to extract these elements as well as the mixing coefficients.

In the linear mixing model (LMM), the acquired "mixed" measurement data is modeled as linear combination of each constituent source signal or simply "sources",

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \quad \text{subject to} \quad \mathbf{1}^T\mathbf{s} = 1;\ s_i \geq 0,\ i = 1,\cdots,c \quad (4)$$

where **x** denotes the *N*-dimensional **observation tidal signal** (or mixture), **A** is an (*N* x *c*)-dimensional **source-signature**

matrix with each column representing the signature of the source, and c represents the number of sources. **s** is the c-dimensional **abundance vector** showing the amount of contributions or proportions of each source signature in forming the mixture **x**. The last term, **n**, takes into account possible errors and noises. The constraints listed in Eq. 4 are to enforce that each element in the abundance vector should sum to one and be non-negative.

The success in the proposed tidal unmixing approach is to find an effective constraint such that the solution to the cost function would be unique. We plan to investigate two different constraints (the baseline sum-to-one and non-negative constraint, as well as the minimum volume constraint) and compare the performance with those of harmonic analysis, especially under extreme weather condition for accurate tidal prediction.

B. *Minimum-Volume-Constrained Non-Negative Matrix Factorization (MVC-NMF)*

We adopt the MVC-NMF [18] unmixing algorithm which was originally designed for spectral unmixing purpose. MVC-NMF is based on non-negative matrix factorization (NMF). Given a non-negative matrix $X \in R^{l \times n}$ and a positive integer $c < \min(l, n)$, the task of NMF is to find two matrices $A \in R^{l \times c}$ (the source signature) and $S \in R^{c \times n}$ (the abundance) with non-negative elements such that $X \approx AS$.

One immediate and critical question would be what is the criterion for the optimal solution. In the convex geometry-based unmixing algorithms, the optimal solution is either defined as the one that circumscribes the data cloud and at the same time has the minimum volume, or defined as the one that inscribes the data cloud with the maximum volume. In order to find the constituents (or source signals or subtides) from the complex tidal data (i.e., the mixture), we have to extend the search space outside the given data cloud. In the meanwhile, we should keep the simplex that circumscribes the data as compact as possible. Combining the goal of minimum approximation error with the volume constraint, we arrive at the following constrained optimization problem,

$$\begin{aligned} &\text{minimize} && f(\mathbf{A},\mathbf{S}) = \frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|_F^2 + \lambda J(\mathbf{A}) \\ &\text{subject to} && \mathbf{A} \geq \mathbf{0},\ \ \mathbf{S} \geq \mathbf{0},\ \ \mathbf{1}_{\mathbf{c}}^{\mathbf{T}}\mathbf{S} = \mathbf{1}_N^T \end{aligned} \tag{5}$$

where $\mathbf{1}c$ ($\mathbf{1}_N$) is a c (N)-dimensional column vector of all 1's, and $J(\mathbf{A})$ is the penalty function, calculating the simplex volume determined by the estimated source signatures. The regularization parameter $\lambda \in \mathbf{R}$ is used to control the tradeoff between the accurate reconstruction and the volume constraint. The first term serves as the external force to drive the search to move outward, so that the generated simplex contains all data points with relatively small errors. The second term serves as the internal force, which constrains the simplex volume to be small. A unique solution is found when these two forces balance each other.

In this work, we plan to apply MVC-NMF to obtain the constituents of the tidal data and study its effectiveness as compared to the ones extracted from harmonic analysis.

IV. EXPERIMENT

A. *Data Description*

The dataset used in this work is from Shell Beach Station, LA. The period of interest is August 18-30, 2012 when the tropical storm, Isaac, hit the coastline. The sampling rate is a typical one sample per hour. There are 7 features in total collected in the set, including celestial factors, meteorological factors, and physical oceanographic factor (i.e., wind speed, wind direction, gusts, barometer, water temperature, and water level).

B. *Experimental Setup*

In the preliminary tests, we have extracted only the water level feature and applied MVC-NMF to see if the algorithm will achieve stable performance even if the number of constituents is selected differently. We have constructed a mixture matrix X by truncating the original time-series of dimension 320x1 into half-day sequences of 12x26.

C. *Results and Discussion*

Figures 1-5 illustrate the extracted endmembers (or sources) as the number of sources goes from 4 to 8. By comparing between the figures, we make two observations.

First, as the number of sources increases, there are more unique constituents appear. However, between Figures 4 and 5, the last endmember is just a noisy version of the 2nd second endmember, showing the number of appropriate constituent can be set at 7.

Second, the constituents are not all sinusoidal, showing its potential to model non-cyclical waves.

More extensive evaluates are needed to fully assess the potential of unmixing-based tidal analysis. First of all, correspondence between the constituents extracted from the unmixing algorithm as well as those from harmonic analysis need to be studied. Second, the evaluation should be carried all the way through tidal prediction. After all, the purpose of tidal analysis is primarily in forecasting. Therefore, as long as the prediction accuracy is competitive, the exact matching with the harmonics is not necessary. Third, in the preliminary study, we only considered one feature, water level. It is interesting to see how performance can be improved if a subset of the features are used if not all. From this perspective, unmixing-based analysis provides a flexible framework for multi-dimensional study. Fourth, we have chosen, arbitrarily, the length of the segment to form the mixture matrix. It is also of interest to see the effect of varying segment length.

V. CONCLUSIONS

In this paper, we proposed to apply the concept of "unmixing" to the problem of tidal data analysis, where instead of the traditional harmonic analysis, the endmembers (or constituents) as well as its abundances (or mixing coefficients) are obtained through a linear unmixing process. The unmixing process is basically a constrained optimization where the nonnegativity and sum-to-one constraints are imposed on the abundance vector. In addition, the minimum volume constraint is also added to ensure the uniqueness of the solution.

The effectiveness of the unmixing-based tidal analysis is yet to be validated by more comprehensive experiments. However, preliminary results had shown promise of the proposed approach.

REFERENCES

[1] Anran Zhang, Jianchuan Yin, Jiangqiang Hu, Chao Yu, "Modular tidal level short-term forecasting based on BP neural networks," *33rd Chinese Control Conference (CCC)*, Nanjing, China, July 2014.

[2] B.L. Polagye, J. Epler, J. Thomson, "Limits to the predictability of tidal current energy," *OCEANS*, Seattle, WA, September 2010.

[3] NASA, "Ocean Motion: Observing and Predicting Tides – Background." http://oceanmotion.org/html/background/tides-observing.htm

[4] P. Matte, D.A. Jay, E.D. Zaron, "Adaptation of classical tidal harmonic analysis to nonstationary tides, with application to river tides," *Journal of Atmospheric and Oceanic Technology*, 30(3):569-589, 2013.

[5] B. Wang, S. He, "Optimized tides prediction method based on support vector regression for extreme weather condition," Technical Report, Shanghai Ocean University, 2014.

[6] E. Marone, F. Raicich, R. Mosetti, "Harmonic tidal analysis methods on time and frequency domains: similarities and differences for the Gulf of Trieste, Italy, and Paranagua Bay, Brazil." http://www.mares.io.usp.br/aagn/aagn14/BGTA_Final.pdf

[7] K.L. Polzin, J.M. Toole, J.R. Ledwell, R.W. Schmitt, "Spatial variability of turbulent mixing in the abyssal ocean." *Science*, 276, pp 93-96, 1997.

[8] J.L. Ledwell, E.T. Montgomery, K.L. Polzin, L.C. St. Laurent, R.W. Schmitt, J.M. Toole, "Evidence for enhanced mixing over rough topography in the abyssal ocean." *Nature*, 403, pp 179-182, 2000.

[9] S.I. Romero, A.R. Piola, M. Charo, C.A.E. Garcia, "Chlorophyll-a variability off paragonia based on SeaWiFS data." *Journal of Geophysical Research*, 111, C05021, 2006.

[10] A.A. Bianchi, L. Bianucci, A.R. Piola, D.R. Pino, I. Schloss, A. Poisson, C.F. Balestrini, "Vertical stratification and air-sea CO2 fluxes in the Patagonian shelf," *Journal of Geophysical Research*, 110, C08003, 2005.

[11] "Why do we study tides?" National Ocean Service, http://oceanservice.noaa.gov/facts/tidestudy.html

[12] "Tidal power", Wikipedia, https://en.wikipedia.org/wiki/Tidal_power

[13] B. Arons. A review of the cocktail party effect. Technical report, MIT Media Lab, 1992.

[14] C.-I. Chang and D. C. Heinz. Constrained subpixel target detection for remotely sensed imagery. *IEEE Trans. Geosci. Remote Sensing*, 38(3):1144–1159, May 2000.

[15] N. Quarmby, J. Townshend, J. Settle, K. White, M. Milnes, T. Hindle, and N. Silleos. Linear mixture modelling applied to avhrr data for crop area estimation. *Int. J. of Remote Sensing*, 13(3):415–425, 1992.

[16] M. O. Smith, P. E. Johnson, and J. B. Adams. Quantitative determination of mineral types and abundances from reflectance spectral using principal components analysis. *J. Geophys. Res.*, (90):C797–C804, 1985.

[17] G. Thomas, S. Hobbs, and M. Dufour. Woodland area estimation by spectral mixing: applying a goodness-of-fit solution method. *Int. J. Remote Sensing*, 17(2):291–301, 1996.

[18] L. Miao, H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained non-negative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765-777, March 2007.

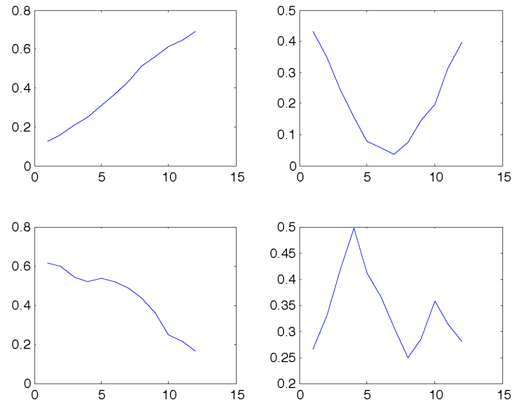

Figure 1: 4 endmembers

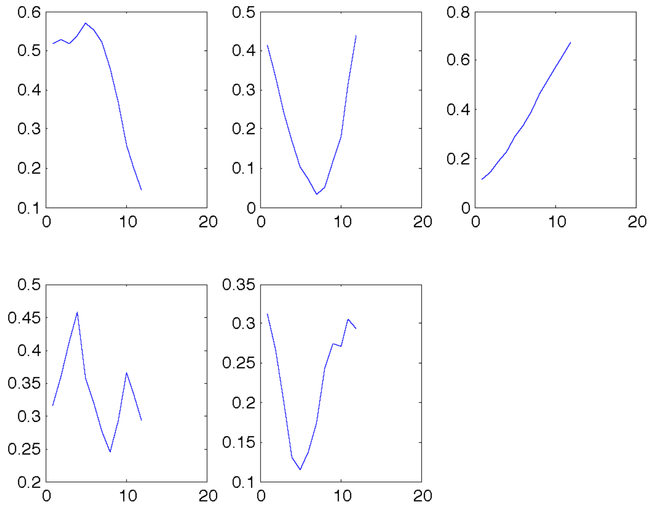

Figure 2: 5 endmembers

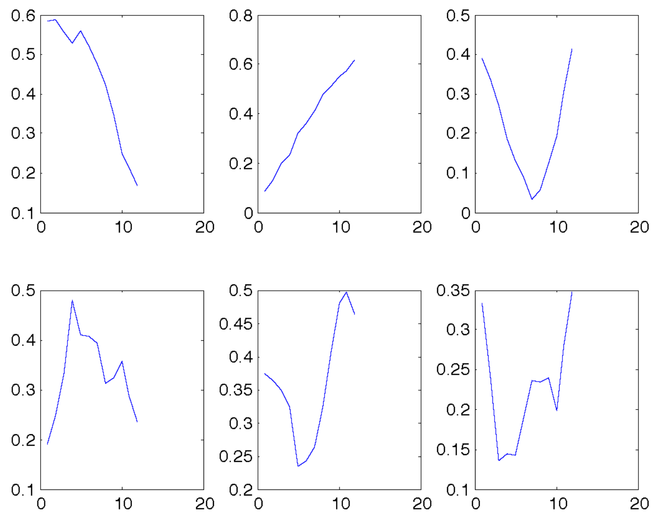

Figure 3: 6 endmembers

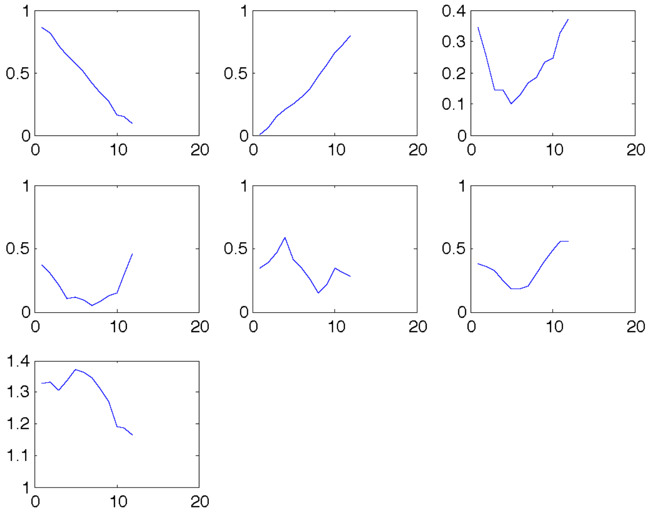

Figure 4: 7 endmembers

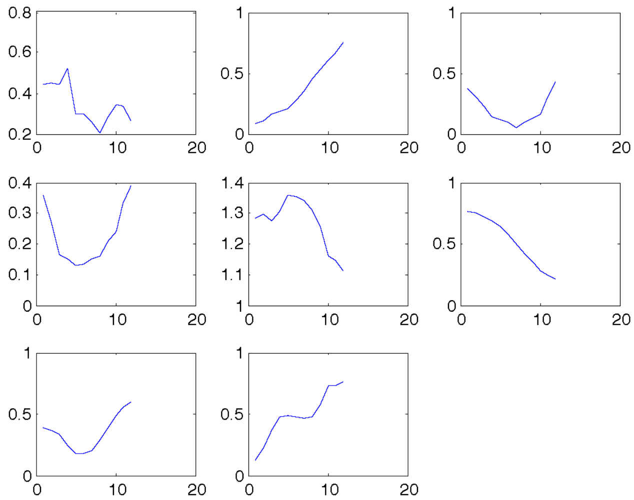

Figure 5: 8 endmembers