
A plea for adaptive data analysis for understanding physical phenomena

by Norden E. Huang*

Collecting and analyzing data, synthesizing and theorizing the results are the core of scientific activities. Therefore, data analysis should be a critical link of this chain; its purpose is to reveal the underlying control mechanisms and the physical meaning of data. Indeed, data analysis is indispensable to every science and engineering endeavor; it is a critical step to converting the cold numbers to physical insights and understandings of a system's evolution and variation. For this reason, data analysis methods should focus more on revealing physically meaningful perspectives rather than on simply obtaining representative mathematical parameters. As data from natural phenomena are highly variable, their analysis poses a severe challenge to traditional data analysis methods. The main reason is that many of the standard methods are developed with limitations, whether implicit or explicit. For example, if we try to define a distribution based on probability theory, the implicit assumption is to have a homogeneous population; for time series data this means stationarity. In the Fourier-based spectral analysis, we face not only stationarity (all frequency or wave number values are constants), but also linearity, for Fourier transform converts multiplication to addition. Consequently, the Fourier expansion results might exist mathematically, but their physical interpretation is problematical.

Lack of proper method notwithstanding, data stream and its increasing complexity never stops. In fact, ever since the arrival of Information Technology, we have faced data everywhere. Now, we are literally drowning in data, yet we are still thirsty for knowledge. As Poincaré had famously put it:

Science is built up of facts,
as a house is built of stones;
but an accumulation of facts is no more a science
than a heap of stones is a house.

Facts here are actually data. Unfortunately, data analysis, as practiced now based on traditional methods, had been relegated to data processing in most cases. There is a whole world of difference between analyzing and processing data. Processing implies almost no thinking: one just do something according to given algorithms. Yet, analyzing requires separating the whole dataset into parts and examining the properties of each part and the inter-relationships among them. In most cases, at the end of processing, one would get a set of mathematical parameters based on the algorithm; while, at the end of analyzing, one should get a full understanding of the physics involved.

I presume that I have "processed" many a set of data, written the results and published papers accordingly like many of the research scientists have done till one day a seemingly trivial problem confronted me: the problem of understanding how ocean waves grow from small ripples to gigantic waves. This is a classic and practical problem. The very way the problem is posed involves an anthropomorphic view: waves just grow naturally. No one has ever given a picture on how this process could happen in waves physically. The detailed mechanism has been thought to entail very complicated nonlinear wave-wave interaction processes, which had been formulated in terms of a six-fold Boltzmann integral covering page after page of mathematical formulae. After laborious computations, the results indeed indicate that the waves would grow gradually, confirming the anthropomorphic analogy: like human beings growing from ba-

* Research Center for Adaptive Data Analysis, National Central University, Chungli, Taiwan
E-mail: norden@ncu.edu.tw

bies to adults naturally. The wave research community seems to be satisfied, but not me. The reason is very simple, for it would violate physical limitations.

If one thinks slightly further, one would find that such an analogy is totally inadequate; it actually is physically impossible. The waves could not have grown gradually. As the growth of human beings requires room to accommodate the process, so does waves. When a train of waves of uniform size is growing, as each wave should grow of its own accord, we do need room to accommodate the growth. If every wave in the train grows a small amount, the sum of the small amounts would no longer be small at the end of a wave train. To make room, the last wave in the train would have to move further than the other ones. But water waves are dispersive: the phase and group velocity of a wave depends on its wave length and period. The shorter wave would propagate slower than the longer ones. Wave of the same length should propagate at the same speed; therefore, no wave could move out further to make room for others to grow. This seems to conflict with the rigorous mathematical result. What is wrong?

At the time, I was working in NASA and I had a well-equipped wave tank built to study the statistical properties of the ocean surface in order to understand the radar backscattering from satellites as a means to infer ocean surface conditions such as sea surface wind velocity, ocean current and wave height, for example. I contrived a simple experiment to see how waves actually evolve as a given uniform wave train propagates down a long channel. Once I got the data, I “processed” them using the time-honored Fourier analysis, and I could not get any cue from the details. The spectra at each station indeed evolved gradually. The spectrum at the end of the wave tank showed a slight downshift of the peak frequency, indicating that the waves indeed “grew” slightly just as the mathematical formulation dictated. I was puzzled. I actually counted the waves and found the number of the waves in the total data length indeed decreased, another indication that the wave had gotten longer. But how could this happen? After considerable struggle, I decided to look closer and tried a different transform that would not just give me a mean but would allow me to examine wave by wave. Suddenly, the Hilbert transform came into my mind.

The Hilbert transform allows one to find the analytic pair of a given signal through converting a set of data in real numbers into complex numbers. Then, one can construct and examine the phase variation and see how exactly each individual wave has evolved. If the waves are growing gradually, the phase function should be smooth. The Hilbert transform results shocked me: I found that the phase function experienced discrete jumps. Each jump was exactly

2π , indicating loss of one whole cycle of oscillation, which means that a whole wave had simply disappeared. This is equally impossible, for the conservation law of wave motion strictly disallows annihilation of any waves. I poured over the data and concentrated on the part where the missing wave occurred and found the reason: the waves actually had undergone a fusion process, two waves merging into one locally. This newly formed wave became twice as long as their prior state. As a result, it propagated faster and caught up with the wave ahead. This time, they shared the space, having evolved from three waves into two. This process would go on until n -waves become $(n - 1)$ waves. Thus, the wave growth process actually happens locally, discretely and abruptly, when waves fuse together like particles. Although we still treat the waves as a continuum, the actual wave evolution even at this macro-scale also retains the wave particle duality. This shocking phenomenon appears in the traditional Fourier analysis as a mundane gradual sideband evolution and spectral peak shift. Indeed, as the Fourier spectrum is expressed in terms of the mean frequency, it is totally insensitive to such local changes (Huang et al., 1999).

I was both shocked and excited by these revelations: shocked because of the discrete nature of the waves, and excited because of the newly found tool of instantaneous frequency (Huang et al., 2009). The latter reveals a mechanism that explains the detailed physics without violating any physical limitations. I was particularly impressed by the power of the instantaneous frequency in revealing the details of the wave evolution process. In fact all theoretical wave motions require the existence of a phase function so that the temporal derivative is the negative frequency and spatial derivative is the wave number. Therefore, the sum of the temporal derivative of wave number and the spatial derivative of frequency should always be zero. This is known as the wave conservation law. The very existence of the wave conservation law requires that the frequency and wave number to be differentiable. Even though Fourier analysis does not violate the wave conservation law, the constant-valued wave number and frequency in Fourier analysis reduces the wave conservation law to the trivial case of “zero plus zero equals zero”. Fourier analysis would produce such a result for any problem. Thus the application of Fourier analysis could drain all the physics from the result just as in the present wave evolution process.

The realization of the power of instantaneous frequency indicates that all physically meaningful time-frequency analysis has to be expressed in term of it. In order to have instantaneous frequency, we have to express the data in adaptive basis, not as the *a priori* simple harmonics components as in Fourier anal-

ysis or wavelet basis. The reason is simple: The natural phenomena are high variable. In most cases, the processes involved are neither linear nor stationary. *A Priori* basis would most likely to fail, for it could not be expected to fit all the ever-changing waveforms distorted by the nonlinear and nonstationary processes. Even if a particular basis could be selected to fit the given data within some time period, the nonstationarity would cause the controlling process to vary with time. Consequently, the fixed *a priori* basis would not fit the data all the time. On the contrary, instantaneous frequency representation overcomes both nonlinearity and nonstationarity limitations. To obtain the instantaneous frequency, one would have to use adaptive basis. Such a basis would adapt to the characteristics of the data and fit any given form of the data all the time. The difficulty encountered in using an adaptive basis is that the dictionary for the basis would have to contain non-denumerably infinitely many basis functions. On the other hand, the advantage of adaptive basis is that it gives an extremely sparse representation of the data from the vast basis space. With the proposed adaptive Empirical Mode Decomposition (EMD, Huang et al., 1998), for example, the number of representative components is only $\ln_2 N$, with N as the number of data points, which is considerably less than the $N/2$ components used in the time-honored Fourier expansion.

It should be pointed out that the term, adaptive data analysis, is not new. Windrow and Stearns (1985) have published a book on adaptive data processing. But the method is primarily on feedback loop in a circuitry, which is hard to apply in general type of data analysis. What we want to emphasize is the expansion of the data on an adaptive basis so that the instantaneous frequency could be computed. In this aspect, the adaptive wavelet analysis seems to fit the requirement (see for example, Wickerhauser, 1996 and Chang, et al., 2000). But in detailed examination, one would find that adaptive wavelet does not exactly fit the definition of adaptivity, for once a wavelet is selected it would be required to fit the all the data. Dynamic process changes in the time span would make application of the adaptive wavelet problematic. Furthermore, the pre-selected wavelet basis would require integration transform to find the frequency. Then, we have to suffer the consequence imposed by the limitation of uncertainty principle. A more recent review on the related issues can be found in Rehman et al., (2014).

There are more fundamental differences between the results from adaptive basis and *a priori* basis. For *a priori* basis, we have to prove the properties of the basis first to make sure that expansions in terms of the basis are convergent, complete, orthogonal and unique. Once a basis satisfies all these conditions, all

we need do is to convolve our data with the given basis to obtain a set of mathematical parameters with respect to this particular basis. This is equivalent to establishing a coordinate system with each basis function as a unit vector, and the expansion serves as a projection of the data in this coordinate system. Once the basis is established, all datasets have to use the same coordinate system. For adaptive data analysis, we have to derive the basis from the data and based on the data. There would be infinitely many basis functions to choose from. In selecting the basis function, the implicit assumption is that it not only would have to fit the data best, but also to impart all the necessary properties required such as convergence, completeness, orthogonality and uniqueness. Of course, we could only check the properties of the basis *a posteriori*, for the basis functions are selected specifically for the given data set only. A change of the value at even a single point would call for a totally different set of bases.

The algorithm of the selection could be the EMD method in which an iterative upper and lower envelopes averaging separates the data into a collection of mono-component functions symmetric with respect to zero known as the Intrinsic Mode Functions (IMF). This decomposition is implemented in the time domain. Alternative approaches using basis pursue would also work. Once we have an IMF basis, we can compute the Instantaneous Frequency directly through quadrature.

The distribution of frequency is known as the spectrum of the data. It is an extremely useful quantifier for the data. But the importance of having instantaneous frequency reaches far beyond having a high fidelity spectrum, we can use it to help us to quantify nonlinearity and nonstationarity, the two subjects whose description in the past can only be handled from a qualitative standpoint. Now let us ponder what Bill Gates had once stated: "My plan to fix the World's biggest problems: Measure them." Unfortunately, our current definitions for either nonlinearity or nonstationarity are all qualitative. The proffered answer is either yes or no, with no nuances or quantifiers. This state of affairs has for long led to the loose use of the term, "nonlinear effects," not as an answer to clarify but as a fig leaf to hide our ignorance. The Instantaneous Frequency offers a glimpse of hope here. If a phenomenon can be classified as oscillatory (in fact all time series could be regarded as undulations with respect to time; therefore oscillatory) and as a frequency spectrum could shed some light on the phenomenon, we could use instantaneous frequency to quantify both of its nonlinearity and nonstationarity using the variation of frequency as function of time.

There are two types of variations in frequency: inter-wave and intra-wave modulations. Traditionally, data are expanded in Fourier series in which the frequency values for all the basis functions are constant. Therefore, we have to say the expansion is meaningful only for linear and stationary processes. For the new Instantaneous Frequency, we would have different frequency values for each IMF at every data point. This should be the case: For example, when we listen to the performance of a symphony orchestra, we should hear different instruments at any given moment, with each IMF representing one kind of instrument. As the music sound is ever changing, it is natural to classify the frequency variations from one cycle of sound wave oscillation to the next in order to represent the nonstationarity. But there is an additional subtle variation of frequency: the intra-wave modulation representing the frequency within one cycle of oscillation.

To quantify nonlinearity, we have to use the intra-wave frequency modulations. This is not as direct as nonstationarity. Nevertheless, it does not need a great stretch of our imagination to rationalize the comparison between a simple harmonic wave with one having nonlinear distortions that would end up with harmonics. Traditionally, harmonic distortion is already identified as nonlinear effects. Due to the interference of noise, which invariably resides in the higher frequency range, no one before has succeeded in using harmonics to quantify nonlinearity. The intra-wave modulation, on the other hand, requires that frequency variations all happen near the fundamental frequency. Therefore, it is very easily quantifiable. Both quantification of nonlinearity and nonstationarity would have great practical applications in the health monitoring of machinery, structures and even the health of human beings. A machine is designed to operate smoothly; therefore, any nonstationary behavior bodes problems. The degree of nonstationarity could be used to establish criteria for the safe operation of machinery. Structures such as building, bridge or the frame of an airplane, are designed based on elastic strength of materials. Therefore, any deviation from the linear stress-strain relationship could bode deficiency. Consequently, we could use the degree of nonlinearity to establish criteria for safety too.

As for the health of a human being or any life system, the situation is much more complicated. If we view a living organism holistically as a complex system, then its behavior could be regarded as the result of interactions amongst the subsystems in the organism. The process is extremely complicated for high level organisms such as human beings. A universal rule has been proposed, by a group at Harvard Medical School, to quantify the health and disease state of living organisms in terms of the degree of

complexity. Health organisms would respond to challenges from the outside environment; therefore, the system would have to be able to occupy a large part of the living variable phase space. As the organism becomes sick or old, the system would degenerate and the region in the phase space occupied would shrink. As a result, measurement of any dynamic biomarkers, such as heart beats or movements, could be used to quantify the health state of the organism through a multi-scale entropy computation. Such evaluation though does not need Instantaneous Frequency per se, but the entropy calculation is a statistical measure requiring mean operation. Here, again, adaptive data analysis offers a critical service, for any calculation of mean requires that the data be detrended.

Trend and detrend are in great need in data analysis, especially in financial and economics research, yet there is no clear definition. In a casual Internet search, for example, there are presently more than 12 million items related to trend and detrend. Among these millions of usages, however, none gave a rigorous and satisfactory scientific definition of either trend or detrend other than a best fitted straight line. As there is a lack of precise definition for trend, determinations of trend and detrend are totally *ad hoc* operations mostly based on regression analysis. This has led to an interesting situation: One economist's trend is another economist's cycle.

Trend determination could be treated as regression analysis, which has long been a core topic in the field of statistics and econometrics. It has also been used extensively in scientific research to define causal relationships. Traditionally, there are two approaches to regression analysis: parametric and non-parametric regression. For parametric regression, the analysis is limited to the determination of the a number of coefficients or parameters of the pre-selected function, typically through least square or the maximum likelihood methods. The idea behind this approach is simple and definitive, but the justification for adopting it is rather thin: How could one defend the selection of one fixed functional form over others? As the final result of the regression depends solely on the pre-selected functional form, the *ab initial* determined functional form would not only be the most important step of the analysis, it could also totally prejudice the final result. Of critical importance is the fact that the selection of the functional form is totally arbitrary.

With the difficulties of parametric regression unresolved, increasingly recent developments have concentrated on the non-parametric or semi-parametric approaches. The techniques used in the non-parametric are much richer: there are studies of the asymptotic form of the functions assuming infinitely

many data, as well as smoothing with kernel estimators and spline, for example. Essentially, in the kernel smoothing approach, a smoothing function is selected for a given window. Through repeated applications of the smoothing operations on the scattered observational data, a final regression form will merge. Here the smoothing function is not a major problem; even a box-car filter works well enough. The crucial factor here is determination of the window size, which would determine the final smoothness of the regression. Selection of the window size is particularly problematic when the process is nonstationary. The spline approach seems to have overcome this difficulty, but the selection of the controlling points for the spline presents another crucial problem, for the number of the controlling point would determine the smoothness of the result. And also, there remains the crucial problem of deciding which of the numerous spline functions should be adopted.

All these difficulties can be traced to the lack of definition for trend. This philosophical difficulty was anticipated by Confucius (551 to 479 BC) some 2500 years ago:

名不正則言不順, 言不順則事不成

Without a proper definition,
logical discourse would be impossible.
Without logical discourse,
nothing can be accomplished.

A definitive definition for trend was proposed by us (Wu et al., 2007): Within the given data span, the trend is an intrinsically fitted monotonic function, or a function in which there can be at most one extremum. In fact, trend should be the residue after all cycles are removed. The definition of a cycle in the data refers to the existence of two extrema. Removal of cycles can be easily done by envelopes. Indeed, the trend should be determined through EMD. Comparing the current practice, I believe the trend determined through EMD makes eminent sense as has been repeatedly demonstrated in different applications. In the application of quantification of complexity, systematic combination of the IMF from the shortest period gradually to the longest ones would give an Intrinsic scale stratification. As the all IMFs have zero mean, the partial sums would not contain any trend to complicate the mean operation. The intrinsic multi-scale entropy result would reveal a richer structure of the data.

All the above results are possible due to adaptive decomposition. To be adaptive, the method has to be independent of any *a priori* basis. The EMD is a time domain scale separation method that is independent of any basis. Yet in mathematical theory, the mainstream approach is still based on *a priori* basis. Our experience has increasingly pointed out that the general approach of adaptive data analysis should be

examined seriously. As technology becomes more sophisticated, data both in quantity and quality would increase. We are overwhelmed by the volume of data, yet at the same time we are also underserved by the information we could extract from the data. In order to be rigorous, all traditional data analysis methods are limited by the necessary, yet artificial, assumptions of stationarity and linearity. The only way to break through this bottleneck is to use adaptive data analysis. EMD is an adaptive data analysis method, which enables us to define instantaneous frequency, the true and physically meaningful way to represent frequency. With the frequency definition, we can also quantify the degree and order of nonlinearity, and also determine the trend. A new day for adaptive data analysis seems to have dawned. Our experience indicates that the combination of EMD and the spectral analysis based on Instantaneous Frequency in terms of a Hilbert Spectral Analysis (Huang et al., 2010) could help us to clarify and provide insight to the underlying physical processes for many scientific, engineering and biomedical problems. This combination has been designated as Hilbert-Huang Transform (HHT) by NASA. Unfortunately, a firm theoretical foundation for HHT is still elusive. The recent results by Hou's group (Hou et al., 2009; Hou and Shi, 2011 and 2012) at Caltech have shown great promise in this respect. It is time for us to pay attention to the theoretical foundation of this crucial approach to data analysis in order to put the versatile data analysis method on firm ground.

References

- Chang S. G., Yu B. and Vetterli M., 2000: Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Proc.* **9**, 1532-1546.
- Hou T. Y., Yan M. P. and Wu Z., 2009: A variant of the EMD method for multi-scale data. *Adv. Adap. Data Anal.* **1**, 483-516.
- Hou T. Y. and Shi Z., 2011: Adaptive data analysis via sparse time-frequency representation. *Adv. Adapt. Data Anal* **03**, 1-28. DOI: 10.1142/S1793536911000647
- Hou T. Y. and Shi Z. Q., 2013: Data driven time-Frequency analysis. *Applied and computational Harmonic Analysis* **35**, 284-308. <http://dx.doi.org/10.1016/j.acha.2012.10.001>.
- Huang N. E., Shen Z., Long S. R., Wu M. C., Shih S. H., Zheng Q., Tung C. C. and Liu H. H., 1998: The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis. *Proc. Roy. Soc. London, A* **454**, 903-995.
- Huang N. E., Shen Z. and Long R. S., 1999: A New View of Nonlinear Water Waves - The Hilbert Spectrum. *Ann. Rev. Fluid Mech.* **31**, 417-457.
- Huang N. E., Wu Z., Long S. R., Arnold K. C., Blank K. and Liu T. W. 2009: On instantaneous frequency. *Adv. Adap. Data Anal.* **1**, 177-229.
- Huang N. E., Chen X., Wu Z. and Lo M., 2010: On Hilbert Spectral Representation: A true time-frequency representation for nonlinear and nonstationary data. *Adv. Adap. Data Anal.* **3**, 63-93.

ur Rehman N., Looney D., Park C. and Mandic D. P., 2014: Adaptive Multiscale Time-Frequency Analysis of Bio-/Neurinformatics, N. Kasabov ed. Springer Handbook, Springer, Berlin-Heidelberg, pp. 745-762.

Wickerhauser M. V., 1996: Adapted Wavelet Analysis: From Theory to Software. A K Peters/CRC Press, Wellesley, MA, p. 504.

Widrow B. and Stearns S. D., 1985: Adaptive Signal Processing. Prentice Hall, Upper Saddle River, NJ.

Wu Z., Huang N. E., Long S. R. and Peng C.-K., 2007: On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc. Natl. Acad. Sci.* **104**, 14889-14894.