

# Analysis of trends and variability of toxic concentrations in the Niagara River using the Hilbert-Huang transform method



Christina W. Tsai<sup>a,\*</sup>, Huntington Treadwell<sup>b</sup>

<sup>a</sup> Department of Civil Engineering, National Taiwan University, Taipei, Taiwan

<sup>b</sup> 207 Jarvis Hall, Civil, Structural and Environmental Engineering, State University of New York at Buffalo, Buffalo, NY 14260, United States of America

## ARTICLE INFO

### Keywords:

Contaminant concentrations  
Time series analysis  
Hilbert Huang Transform  
Trend identification  
Variability

## ABSTRACT

This study introduces a more recent data analysis method, Hilbert Huang Transform method (HHT), to describe contaminant concentration data of a non-stationary and non-linear nature. In order to improve the modeling of the contaminant concentrations, it is proposed to first process the data using the Empirical mode decomposition (EMD) method from HHT to obtain a collection of intrinsic mode functions (IMFs) which can then be modeled separately using either autoregressive moving average (ARMA) models expanded with a seasonal term, or linear regression analysis, depending on the nature of the IMF. Three priority contaminants measured at Niagara-on-the-Lakes are selected for this study. It is found that the trend of fluoranthene concentrations from April of 1986 to March of 1997 is decreasing and then beginning to increase; the 1,2,4-trichlorobenzene concentrations are decreasing; while the dieldrin concentrations are decreasing. With HHT, appropriate time series models can be identified and constructed for the studied contaminant concentrations to better illustrate the variability of each IMF (and thus the contaminant concentrations) for the studied period. For all data sets modeled in this study, pre-processing the data with HHT allowed for higher  $R^2$  values, correlation coefficients and lower sum of squared errors when compared to modeling without HHT. It is thus confirmed that pre-processing the data with HHT and modeling with time series analysis will provide a more effective means of the studied data sets when identifying and analyzing the trends and variability of studied contaminant concentrations in the Niagara River.

## 1. Introduction

### 1.1. Problem description

The field of water quality modeling is important for providing information and guidance for water quality management and decision-making, including managing chemical discharges into lakes and rivers. To make educated decisions regarding water quality, it is important to be able to understand and predict the behavior of the pollutants in the water resources of concern, particularly on their trends and variability. It is a manageable task to measure toxic concentrations in these water bodies; however, it can be difficult to model the concentrations when they are non-stationary and non-linear in nature. For a long period of time, the time series analysis has been demonstrated to be a viable tool to identify trends and variability of complex data (Bernal et al., 2017; Desai and Rifai, 2010; Elgar et al., 2015; Gedikli et al., 2008; Lohani and Wang, 1987; Safari et al., 2018; Shao et al., 2010). The Hilbert

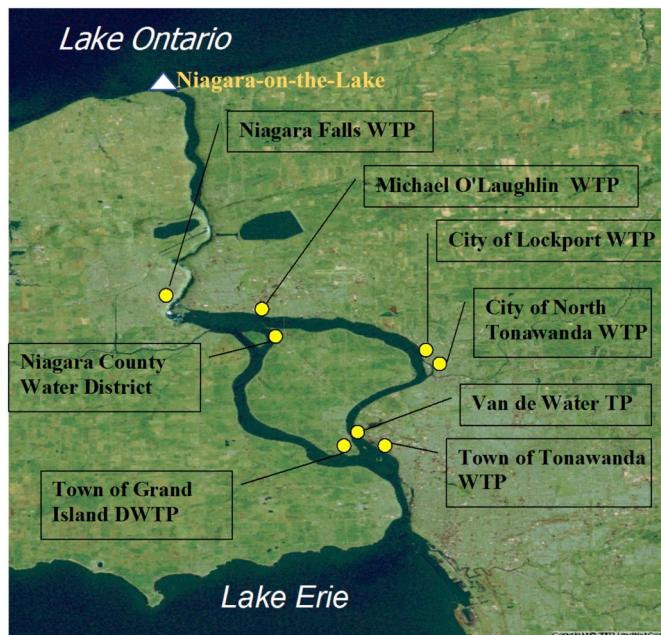
Huang Transform (HHT) has only recently been introduced to the geophysical field for its ability to analyze non-stationary and non-linear time series such as these (Franceschini, 2007; Huang and Wu, 2008; Gairola and Chandrasekhar, 2017). Previously, the HHT method has been used in multiple disciplines including aeronautic, biology, physics, and medical research fields. Wu et al. (2007) adopted HHT to analyze the time-varying trend in long term global-mean surface temperature. Franceschini and Tsai (2010) applied HHT for analysis and forecasting of toxic concentrations in the Niagara River. In their study, they coupled the empirical mode decomposition (EMD) method of HHT with autoregressive analysis to model and forecast contaminant concentrations. Kuai and Tsai (2012) have utilized HHT to identify multiple time scales embedded in a complex sediment transport dynamic process. Recently Tsai et al. (2018) also employed HHT to study the impacts of climate change induced extreme weather events on dengue fever.

The primary motivation for this study is the need for a way to identify and analyze the contaminant concentration data of a non-

\* Corresponding author.

E-mail addresses: [cwtsai@ntu.edu.tw](mailto:cwtsai@ntu.edu.tw) (C.W. Tsai), [hrt@buffalo.edu](mailto:hrt@buffalo.edu) (H. Treadwell).

<sup>1</sup> Formerly, Associate Professor, 223 Jarvis Hall, Civil, Structural and Environmental Engineering, State University of New York at Buffalo, Buffalo, NY 14260, United States of America.



**Fig. 1.** Area of study.

stationary and non-linear nature. Better characterization of data would facilitate the analysis of uncertainty associated with model predictions (Melching and Anmangandla, 1992; Tsai and Franceschini, 2005; Tsai and Li, 2014). This study aims at expanding upon the available modeling methodologies, creating a systematic modeling approach and emphasizing the advantage of using HHT in the field of water quality modeling. In order to improve the modeling of the contaminant concentrations, it is proposed to first process the data using the EMD method from HHT to obtain a collection of intrinsic mode functions (IMFs) which can then be modeled separately using an appropriate time series analysis model to better describe the variability of studied contaminant concentrations. It is hypothesized in this study that pre-processing the data with HHT and modeling with time series analysis will provide a more accurate representation (by  $R^2$  value, correlation coefficients and sum of squared errors) of the studied data sets when compared to modeling with time series analysis without HHT.

## 1.2. Niagara River background

The Niagara River is a natural channel connecting Lake Erie and Lake Ontario, and home to the historic Niagara Falls, as shown in Fig. 1. Historically, the Niagara River has served as a source for both clean drinking water and hydroelectric power (from Niagara Falls). However, since the development of industry, the quality of the water has been degraded due to contaminant discharges from various industries and wastewater treatment plants. Although regulations have been improved, these past actions have left the Niagara River to be classified as an Area of Concern in the Great Lakes region. Therefore, it is pertinent to monitor the status of the river and keep the contaminant concentrations in check. This study proposes an improved method of analyzing the contaminant concentration data in order to better

understand the trend and variability of the river contaminants.

### 1.3. Description of data

The contaminant concentration data sets analyzed in this paper were selected from a group of 18 priority toxic pollutants that have been sampled from the Niagara River since 1986, under the Niagara River Toxics Management Plan. The water samples were collected at the head (Fort Erie, Canada) and mouth (Niagara on the Lakes, Canada) of the Niagara River on a weekly/biweekly basis. A total of three contaminants were selected for this study. The list of selected contaminants and the reason for their selection is presented in Table 1.

For each contaminant in Table 1, their measured concentrations at Niagara-on-the-Lakes were analyzed in this study. This resulted in a total of three contaminant concentration data sets.

The sampling period of all the contaminants spans from April of 1986 to March of 1997. As it is common to have some incompleteness in data with a long temporal monitoring scale, steps were taken to fill in the missing data points in order to gain a consistent time step for time-series analysis. The fluoranthene data set is used as an example here to explain how the data were formatted to achieve a consistent time step.

The raw fluoranthene data set spanning the entire sampling period (from April 1986 to March 1997) is contained in Fig. 2(a). Some points (concentrations  $> 10 \text{ ng/L}$ ) were considered outliers and omitted from the data set (Fig. 2(b)). To start, a section of approximately 4.5 years (from 5/11/89 to 11/11/93) of weekly fluoranthene concentration data (215 points) was chosen from the whole data set to optimize completeness. Within this shortened data set, some of the samples were not collected exactly one week apart (i.e. every 6 or 8 days instead of 7). In these cases, the available measured concentration is assumed to be representative within the chosen time step (i.e., concentration is assumed uniformly distributed within the selected time step). In this way, a forecast could be created at equal, weekly intervals. Linear regression was used to fill in any other data point gaps. In total, 21 new points were created to make a data set of 236 points (Fig. 2(c)). The created points are well distributed (with no more than three points in a row) and therefore believed to have an insignificant impact on the originality of the data.

All three figures show that the concentrations are non-linear by appearance, with many varying frequencies throughout the data. Nonlinearity refers to data series possessing nonstandard features such as non-normality, asymmetric cycles, bimodality, nonlinear relationships between lagged variables, variation of prediction performance over the state-space, time irreversibility, sensitivity to initial conditions, and etc. (Fan and Yao, 2003). Fig. 2(c) also shows the non-stationarity of the selected data as the mean and variance are not constant throughout the data collection period.

## 2. Methodology

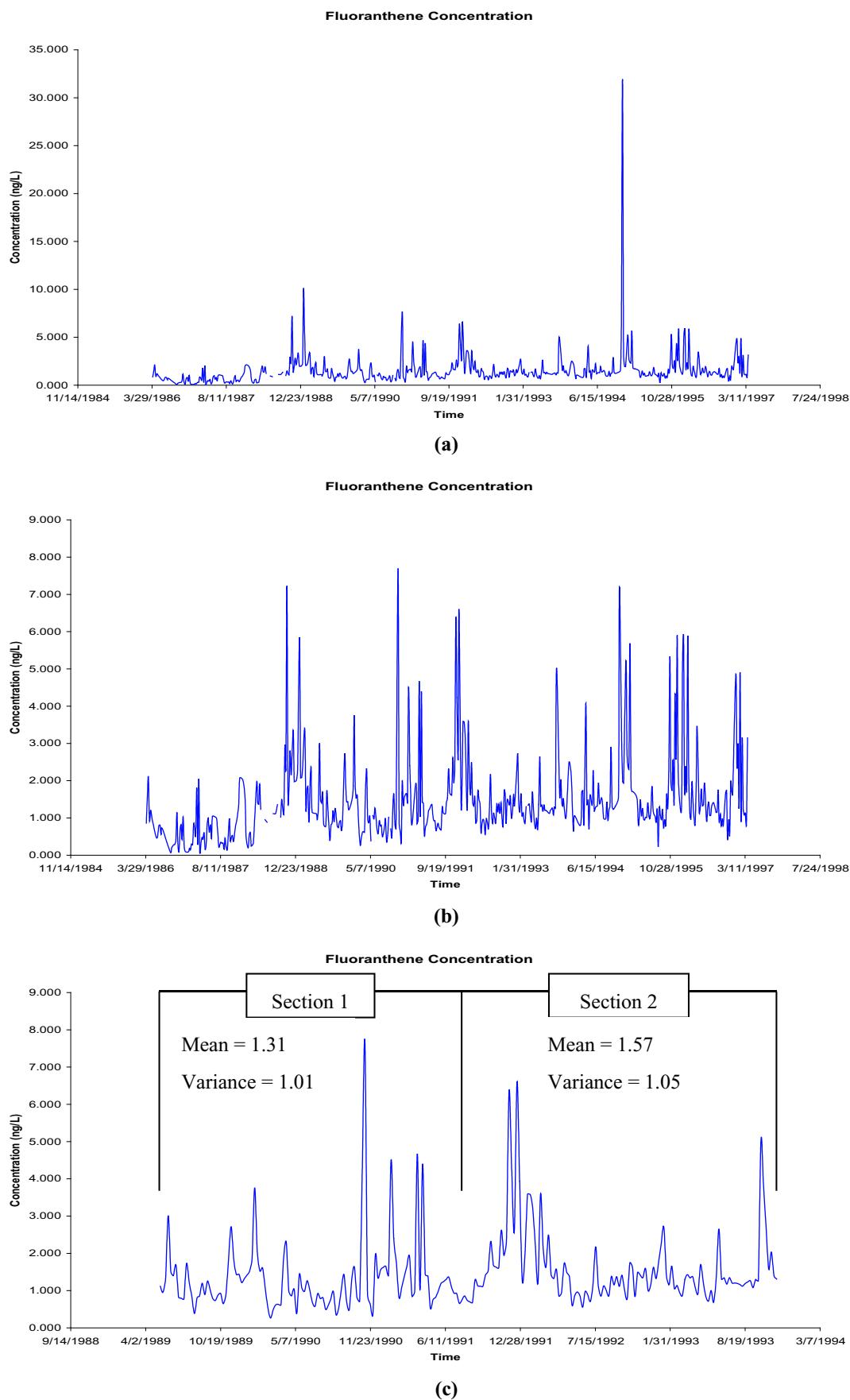
### 2.1. Hilbert Huang Transform

The Hilbert Huang Transform is composed of two methods: empirical mode decomposition (EMD) and the Hilbert Spectral Analysis (HSA). Empirical mode decomposition is a mathematical process that reduces non-stationary and non-linear data into intrinsic mode

**Table 1**  
Criteria for Contaminant Selection.

Contaminant	Importance <sup>a</sup>
Fluoranthene	Fluoranthene is a carcinogen that is suspected to be a gastrointestinal and liver toxin.
Dieldrin	Dieldrin has been associated with Parkinson's Disease, breast cancer, and damages to the immune, reproductive and nervous systems.
1,2,4 – Trichlorobenzene	1,2,4 – Trichlorobenzene is a suspected carcinogen, neurotoxin, and is also suspected to adversely affect child development.

<sup>a</sup> As well as the importance of the contaminants to human health, the selected contaminants also have more complete data sets.



**Fig. 2.** (a) Raw Fluoranthene Data Set with Missing Sampling Points and Dates. (b) Fluoranthene Data Set After Removing Outliers (the two highest data points from Fig. 2(a)). (c) Selected Fluoranthene Data Set with Reformatting (note the time scale).

**Table 2**  
Summary of IMF Characteristics.

Frequency range	Intrinsic mode function	Description	Generally modeled by <sup>a</sup>
High to Intermediate Frequency	IMF 1	Very noisy. No discernible trends. A short seasonal period may be incorporated into the ARMA model.	ARMA
	IMF 2	Less noise than IMF 1. A short seasonal period can be selected from the autocorrelation function.	ARMA
	IMF 3	Non-linear, non-sinusoidal. Increased seasonal periods from IMF 2.	ARMA
	IMF 4	May or may not be sinusoidal in nature.	ARMA
	IMF 5	More sinusoidal in nature. Longer seasonal periods.	ARMA
	IMF 6	More sinusoidal in nature. When very sinusoidal, linear regression analysis may be used. Otherwise, use ARMA model.	ARMA or Linear Regression Analysis
Lower Frequency	Residual Trend	Simple curve with increasing or decreasing slope. Linear regression is used to continue the trend.	Linear Regression Analysis

<sup>a</sup> Specific to this study. Length of data set will dictate the number of IMFs produced from EMD.

functions (IMFs). EMD uses the sifting process to extract IMFs, which represent the analyzed signal at different time scales. Sifting is completed when the last component is monotonic (i.e., with only one maximum and one minimum). An IMF is defined as a function that satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero (Huang et al., 1998). Beginning with the raw data set, the EMD process is implemented to produce IMF 1. The process continues by decomposing IMF 1, resulting in IMF 2. The pattern continues until only a residual trend remains. In this study, six IMFs and a residual trend were produced by the EMD process for each data set. Table 2 contains a descriptive summary of the characteristics of each IMF. Each IMF is generally categorized by its apparent frequency. IMF 1 is generally very noisy and has a very high frequency (great variability over a short period of time). The frequency decreases with each consecutive IMF. In this study, IMFs 1–3 are generally of a higher frequency with short seasonal periods (time between consecutive maxima or minima). IMFs 4 and 5 continue to decrease in frequency (with longer seasonal periods) and begin to approach a sinusoidal likeness (intermediate frequency). IMF 6 is generally very sinusoidal in nature, which is then decomposed, leaving a residual trend.

The purpose of the EMD process is to decompose a data series into a collection of Intrinsic Mode Functions (IMFs). The EMD procedure separates the original time series into polynomial and sinusoidal components and a residual. Eq. (1) represents the original time series function as the sum of the n IMFs and the residual produced from the EMD process.

$$X(t) = \sum_{k=1}^n c_k(t) + r_n(t) \quad (1)$$

where  $c_k(t)$  is the intrinsic mode function,  $r_n(t)$  is the residual,  $k$  denotes the number of each sequential IMF (i.e.  $c_3(t)$  is the third IMF resulting from the sifting process) and  $n$  is the number of IMFs.

The Hilbert Transform can be coupled with the EMD process to obtain the Hilbert Spectrum ( $H(\theta, t)$ ). Taking the Hilbert Transform of the IMFs in Eq. (1) and expressing it in the analytical function form allows for the development of Eq. (1) into Eq. (2).

$$X(t) = \operatorname{Re} \left\{ \sum_{k=1}^n A_k(t) \exp[i\theta_k(t)] \right\} + r_n(t) \quad (2)$$

where  $\operatorname{Re}\{\dots\}$  represents the real part of the function,  $A(t)$  is the envelope of  $X(t)$ , and  $\theta(t)$  is the instantaneous or local phase of  $X(t)$ . Eq. (2) can be used to plot the Hilbert Spectrum, which is a three-dimensional graphical representation of the energy of the signal with respect to time and frequency (Franceschini and Tsai, 2010). Analyzing and interpreting the Hilbert Spectrum can be insightful as to the properties

of the original time series. The Hilbert Spectrum also provides a quantitative measure of the degree of stationarity (Huang et al., 1998).

## 2.2. Random number generation

Early discussions included the possibility of modeling IMF 1 of each data set with random number generation. When viewing IMF 1 of each data set, the signal is very noisy and random in appearance. Random number generation is based on a normal distribution with density function (Eq. (3)),

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)} \quad -\infty < y < \infty \quad (3)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation, respectively (Mendenhall and Sincich, 2007). The mean and standard deviation may be calculated from the IMF and input into the random number generator to create a model and forecast. However, it was found that a random “model” would not fit to the data well because the model is randomly generated and does not fit to any trend. Creating a forecast with random number generation may provide a better argument as data are generated with the same mean and standard deviation as the original function.

### 2.2.1. Autoregressive moving average modeling and analysis

Autoregressive moving average (ARMA) modeling was used to model intrinsic mode functions with high to intermediate frequency (primarily IMFs 1–5). An ARMA model, as it is named, is composed of an autoregressive model, AR(p), and a moving average model, MA(q). The ARMA model is described in Eq. (4).

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (4)$$

where  $c$  is a constant,  $\varepsilon_t$  is white noise,  $\varphi$  are the p parameters of the model and  $\theta$  are the q parameters of the model. The parameter, “p”, accounts for the number of terms in the autoregressive model. Similarly, the parameter, “q”, accounts for the number of terms in the moving average model. Throughout the analysis, ARMA models were generally used on IMFs 2 and 3.

The selection of the ARMA model parameters, p and q, required the implementation of an iterating program. The values of p and q were determined by minimizing the Akaike information criterion (AIC) value. The AIC value is a measure of goodness of fit of a model and a nice tool for model selection.

The ARMA models created in this study contain a separate seasonal term to capture the oscillating nature of the data. The seasonal period incorporated into the model is estimated from the autocorrelation function (ACF) of the particular IMF. The seasonal term parameters used in the model were p, d, and q, assuming fixed values of 1, 1, and 0, respectively. During the initial stages of analysis, the parameter values

of the seasonal term were varied. Due to the insignificant resulting effect, it was decided that the parameters would assume fixed values for simplicity.

### 2.2.2. Linear regression analysis

IMFs that appeared to be very sinusoidal in nature with lower frequency and longer seasonal periods were modeled using linear regression analysis. A general form of the linear models used in this study is shown in Eq. (5).

$$y = \gamma \cdot \sin\left(\frac{2\pi t}{T_1}\right) + \beta \cdot \cos\left(\frac{2\pi t}{T_2}\right) \quad (5)$$

where  $t$  is time,  $T$  is the seasonal period,  $\gamma$  is the sine term fitting coefficient and  $\beta$  is the cosine term fitting coefficient. The seasonal period,  $T$ , is selected by estimation from the autocorrelation function of the intrinsic mode function. The model is then fit to the data by adjusting the  $\gamma$  and  $\beta$  coefficients. The resulting model is used to create a forecast of the IMF. Throughout the analysis, linear models were used on IMF 6 (when very sinusoidal in nature) and the residual trend.

## 3. Modeling and results

The following section will discuss the procedure and effort put forth in modeling the three contaminant concentration data sets previously described: fluoranthene, dieldrin, and 1,2,4-trichlorobenzene measured at Niagara-on-the-Lakes. Fluoranthene was focused on as the primary data set. The modeling effort of the fluoranthene data set is described in detail. The two remaining data sets were modeled in a very similar fashion and are summarized in the following sections. All three data sets were also modeled using an AR model without HHT to test the hypothesis.

For modeling each IMFs, first, the data set is modified to keep a consistent time step. Then a log-transform is applied to the data following by performing the empirical mode decomposition to produce a set of independent intrinsic mode functions. Then we evaluate each IMF by visual inspection and analyzing the autocorrelation function. At this juncture, data will be classified into two groups of high to intermediate or lower frequency. Next, a forecast of each IMF can be obtained with respect to each classification. At final steps, each IMF prediction is combined and by taking exponential data are transformed to the original state. The general proposed procedure for modeling each IMF is described in Fig. 3.

### 3.1.1. Fluoranthene concentration (measured at Niagara on the lakes)

The empirical mode decomposition process of HHT was performed on the fluoranthene concentration time series to create the six IMFs and the residual found in Fig. 4(a).

### 3.1.2. Intrinsic mode functions (IMFs)

Each IMF was modeled separately based on the nature of its contents and by referring to the procedural guidelines of Fig. 3. IMF 1 contains much of the noise of the original data set. When looking at the autocorrelation function for IMF 1 (Fig. 4(b)), no discernible trend can be made from the figure.

Referring back to the procedural guidelines of Fig. 3, when the data appear noisy and have a high frequency, an ARMA model with a short seasonal period may be used to model and forecast the data. It was found that increasing the parameter  $p$  would provide a better fit to the IMF 1 function. A seasonal period was selected which would produce a forecast that carried on the natural frequency of the data. Values for model parameters  $p$  and  $q$  were selected to be 22 and 3, respectively. A seasonal period of 3 was selected to reproduce the frequency of the data. The variance and Akaike information criterion (AIC) value of the

model are found in Table 3(a).

To further investigate the goodness of fit of the model, time series diagnostic tests were performed (see Fig. 4(c)). These tests involved inspecting the standardized residuals, the ACF of the residuals, and the p-values from the Ljung-Box statistic test (not to be confused with the number of  $p$  parameters for the ARMA models). The Ljung-Box test is a version of the Portmanteau goodness-of-fit test, where higher p-values represent a better goodness-of-fit (Brockwell and Davis, 2002). Higher p-values represent that the residuals are not significant, which means the model was able to fit the data more significantly.

The fit of the model and the associated forecast are presented in Fig. 4(d). The model fits the data with an  $R^2$  value of 0.265 and a correlation coefficient (CC) of 0.515. The forecast carries on a high frequency which is consistent with the original data, though the magnitude of the original function is not represented in the forecast.

IMF 2 can be seen as the second plot from the top found in Fig. 4(a). By visual inspection, there is much less noise in IMF 2 after being decomposed from IMF 1. The autocorrelation function for IMF 2 is displayed in Fig. 5(a).

The IMF 2 autocorrelation function appears less noisy than the ACF of IMF 1. From this figure, it is possible to make an estimate of the seasonal period (the distance between two sequential, local maxima or minima) of IMF 2. Similarly, it can be concluded that IMF 2 is a higher frequency data set with shorter seasonal periods and therefore can be further analyzed using an autoregressive moving average (ARMA) model.

The first step to creating an ARMA model for IMF 2 was to estimate a seasonal period from the ACF found in Fig. 5(a). This estimation can be made by visual inspection or by tabulating the values to get a closer look at the lag between the chosen sequential, local maxima or minima. In this study, two points were selected which appeared to represent an average spacing throughout the function. A period of 13 points (or weeks, as the data are weekly) was estimated as the seasonal period for IMF 2 and input into the seasonal term of the model.

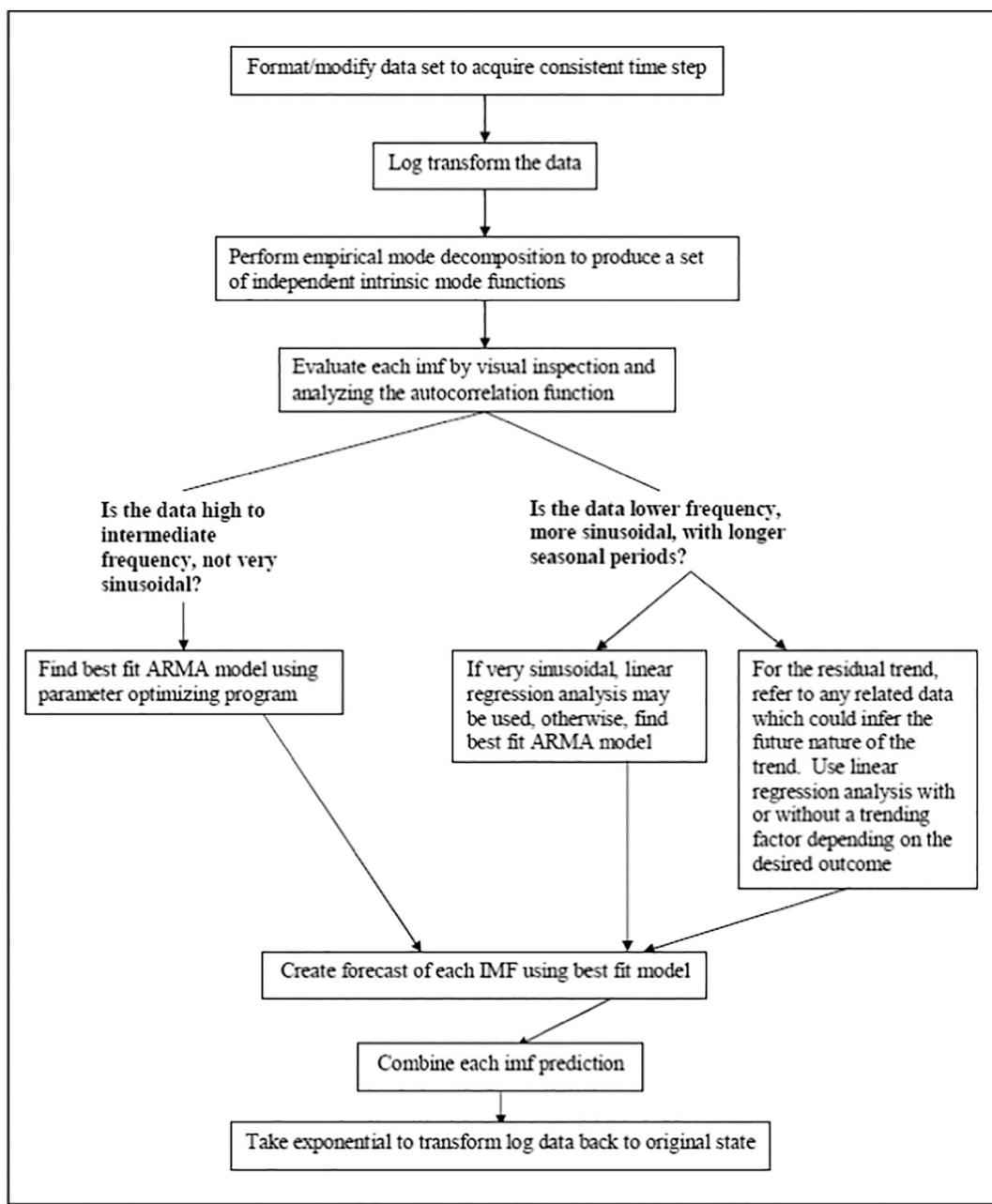
The second step to the model was to select the parameter values for  $p$  and  $q$ . The parameters were each given a range of possible values to be input into the ARMA model. Through iteration, the program evaluated all of the possible parameter combinations and selected the combination which resulted in a minimized AIC value. For IMF 2, the values for  $p$  and  $q$  were selected to be 4 and 3, respectively. Therefore, the selected ARMA model will have 4 autoregressive coefficients and 3 moving average coefficients. The ARMA model was also expanded with a seasonal term with parameter  $p$  assuming a fixed value of 1. The value and standard error for each coefficient, as well as the variance and AIC value of the model, are found in Table 3(b). Model fit diagnostics are presented in Fig. 5(b).

Fig. 5(b) shows that the expanded ARMA model for IMF 2 has high p-values across the board and therefore is a good fit to the data. The randomness of the standardized residuals in the figure shows that the trend of the data was picked up in the model, as there is no discernible trend in the residuals. The fit of the model and associated forecast are presented in Fig. 5(c). The model fits the data well with an  $R^2$  value of 0.897 and a correlation coefficient of 0.947. The forecast appears to be consistent in nature with the original data.

IMF 3 was similar to IMF 2, though generally lower in frequency (see IMF 3 in Fig. 5(a)). The ACF for IMF 3 is found in Fig. 6(a).

Similarly, IMF 3 can be considered a higher frequency data set with shorter seasonal periods, as it is certainly not sinusoidal in nature. After taking a closer look at the ACF for IMF 3, a seasonal lag of 15 was selected. Executing the parameter optimization sequence resulted in parameter  $p$  and  $q$  values of 4 and 3, respectively. The details of the parameters are found in Table 3(c). The results of the diagnostic tests for IMF 3 are found in Fig. 6(b).

The time series diagnostic tests show high p-values and random standardized residuals. Therefore, the model appears to be a good statistical fit to the data. As presented in Fig. 6(c), the model fits the



**Fig. 3.** General Procedure for Modeling Intrinsic Mode Functions (IMFs).

data well with an  $R^2$  value of 0.997 and a correlation coefficient of 0.998. The resulting forecast is a very logical representation of the data it was based on.

The fourth IMF of the fluoranthene data set presented a small problem in the modeling. A closer look at IMF 4 is presented in Fig. 7(a).

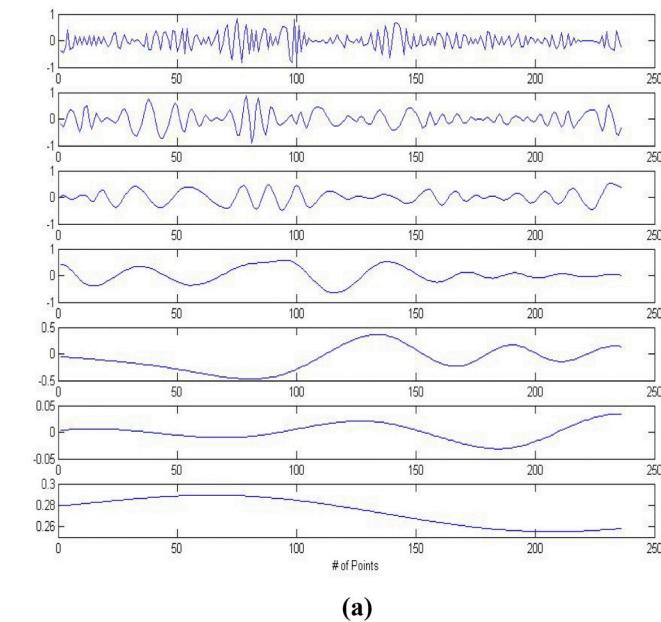
The challenge of modeling IMF 4 was that the nature of the IMF is inconsistent with itself. A very apparent change in the nature of the function takes place around 150 weeks. The IMF starts out with a larger amplitude and lower frequency pattern, but then changes to a lower amplitude and higher frequency pattern. This behavior can also be seen in the ACF of IMF 4, found in Fig. 7(b).

The ACF shows two different sets of seasonal periods. The first section (the larger mounds) shows consecutive maximums and minimums occurring approximately every 50 weeks (around 1 year). The last part of the function shows seasonal periods of about 25 weeks (or around 6 months). In order to model and forecast this function, it was decided to focus on the later section and create an ARMA model with a

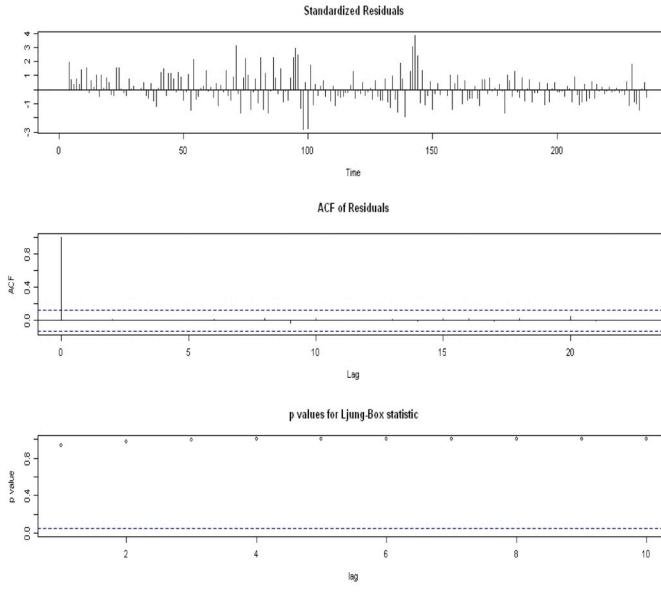
seasonal period of 25 weeks. Focusing on the later section of data should provide a forecast which continues the function in a more logical manner. When comparing models using a seasonal period of 25 weeks versus 50 weeks, it was found that the model using a 25-week seasonal period had higher p-values and was therefore statistically more significant. The details of the model are found in Table 3(d). The model fit diagnostic test results can be seen in Fig. 7(c).

Fig. 7(c) shows generally high p-values and random standardized residuals, suggesting the model is a good statistical fit to the data, as shown in Fig. 7(d). The model fits the data with an  $R^2$  value and correlation coefficient of 0.999. The forecast emulates the higher frequency portion of the function, which is consistent with the selected seasonal period of 25 weeks.

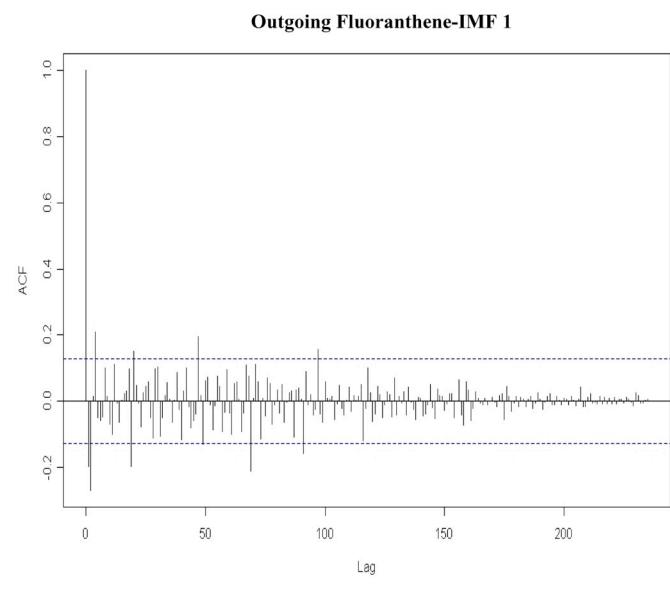
When looking at IMF 5 in Fig. 4(a), it can be observed that the function has a much lower frequency than its previous components, and longer periods between consecutive maxima or minima. It also appears to be somewhat sinusoidal in nature. Therefore, similarly, this function



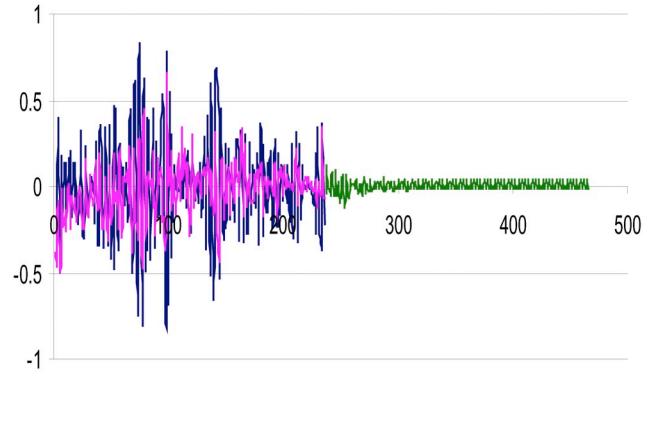
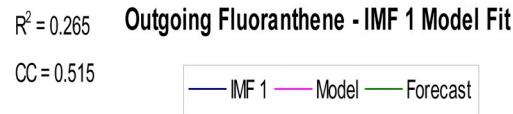
(a)



(b)



(c)



**Fig. 4.** (a) IMFs 1 (top) through 6 and residual (bottom) of the Log Transformed Fluoranthene Data. (b) Autocorrelation Function of IMF 1. (c) Model Fit Diagnostics for IMF 1. (d) IMF 1 Model and Forecast.

should be modeled using an ARMA model. Fig. 8(a) contains the ACF for IMF 5.

A period of 45 weeks was estimated by looking at the spacing of the later section of IMF 5. This period was selected in order to produce a forecast that would carry out the later part of the function and appear more logical. The details of the model are found in Table 3(e). The model fit diagnostics are found in Fig. 8(b).

Fig. 8(b) generally random standardized residuals and moderate p-values. Fig. 8(c) confirms that the model has no trouble fitting the data. The  $R^2$  value and correlation coefficient are both 0.999. The forecasted portion of the model is a good representation of the later part of the function.

When referring back to Fig. 4(a), IMF 6 can be seen as the second plot from the bottom of the figure. The function appears sinusoidal in its steady spacing of maxima and minima; however, the magnitude of each local max and min is increasing with time. Therefore, an ARMA

model was selected for modeling IMF 6. The ACF of IMF 6 was analyzed in order to incorporate the appropriate period into the model. The autocorrelation function of IMF 6 is displayed in Fig. 9(a). A period of 112 weeks was estimated from the figure. The model fit diagnostics are found in Fig. 9(a). A summary of the model details is presented in Table 3(f).

Fig. 9(b) contains the time series diagnostics of the model fit. Based on p-values of the residual, only a portion of the model may be statistically significant to the data. However, Fig. 9(c) displays the model fitting the data very well with an  $R^2$  value and correlation coefficient of 0.999. The forecast continues the increasing trend of each cycle of the function.

### 3.1.3. Residual trend

The final component to model and forecast is the residual trend of the fluoranthene time series, which can be seen as the last plot in

**Table 3** ARMA Model Coefficients; (b) IMF 2 ARMA Model Coefficients; (c) IMF 3 ARMA Model Coefficients; (d) IMF 4 ARMA Model Coefficients; (e) IMF 5 ARMA Model Coefficients; (f) IMF 6 ARMA Model Summary; (g) Fluoranthene Residual Model Summary.

(a)	
Model coefficients	
Value	22
sigma <sup>2</sup> estimated as 0.06387; log likelihood = -16.95, aic = 87.91	
P	q
ar1	sr1
ar2	
ar3	
ar4	
ma1	ma2
ma3	sr1

(b)	
Model Coefficients	
Value	1.9980
S.E.	0.1257
sigma <sup>2</sup> estimated as 0.09387; log likelihood = 198.6, aic = -379.21	
ar1	-1.9861
ar2	0.2552
ar3	1.1253
ar4	0.2411
ma1	-0.2721
ma2	0.1116
ma3	0.1237
sr1	0.0497

(c)	
Model coefficients	
Value	3.4189
S.E.	0.0547
sigma <sup>2</sup> estimated as 0.0001857; log likelihood = 623.94, aic = -1229.89	
ar1	-4.6757
ar2	0.1505
ar3	3.0222
ar4	0.1499
ma1	-0.7824
ma2	0.0542
ma3	0.0885
sr1	-0.1334

(d)	
Model Coefficients	
Value	3.7996
S.E.	0.0149
sigma <sup>2</sup> estimated as 5.399e-07; log likelihood = 1212.45, aic = -2406.9	
ar1	-5.4996
ar2	0.0425
ar3	3.9533
ar4	0.0416
ma1	-0.8947
ma2	0.0140
ma3	0.0422
sr1	0.0159

(e)	
Model coefficients	
Value	3.7996
S.E.	0.0149
sigma <sup>2</sup> estimated as 5.399e-07; log likelihood = 1212.45, aic = -2406.9	
P	q
ar1	
ar2	
ar3	
ar4	
ma1	
ma2	
ma3	
sr1	

(continued on next page)

Table 3 (continued)

	ar1	ar2	ar3	ar4	mal	ma2	sr1
Value	3.6453	-4.9638	2.9988	-0.6754	-0.7813	0.6276	-0.4579
S.E.	0.0481	0.1436	0.1434	0.0478	0.0561	0.0412	0.0734
sigma^2 estimated as 3.121e-08; log likelihood = 1368.09, aic = -2720.18							
(f)							
Model coefficients							
	p			q			
	ar1	ar2	ar3	ar4	mal	ma2	sr1
Value	0.9228	0.6104	-0.0101	-0.5315	0.3228	0.0368	-0.0707
S.E.	0.0157	0.0256	0.0202	0.0128	0.1109	0.0789	0.0955
sigma^2 estimated as 4.394e-09; log likelihood = 988.4, aic = -1958.81							
(g)							
Coefficients	Estimate				t value		Pt(>   t  )
Intercept	0.2789				5.555e-05		< 2e-16 ***
T		-4.832e-05			4.523e-07		< 2e-16 ***
$\gamma \rightarrow \sin$		1.383e-02			3.938e-05		< 2e-16 ***
$\beta \rightarrow \cos$			-7.334e-04		2.506e-05		< 2e-16 ***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1							
Residual standard error: 0.0002197 on 232 degrees of freedom							
Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997							
F-statistic: 2.764e+05 on 3 and 232 DF, p-value: < 2.2e-16							

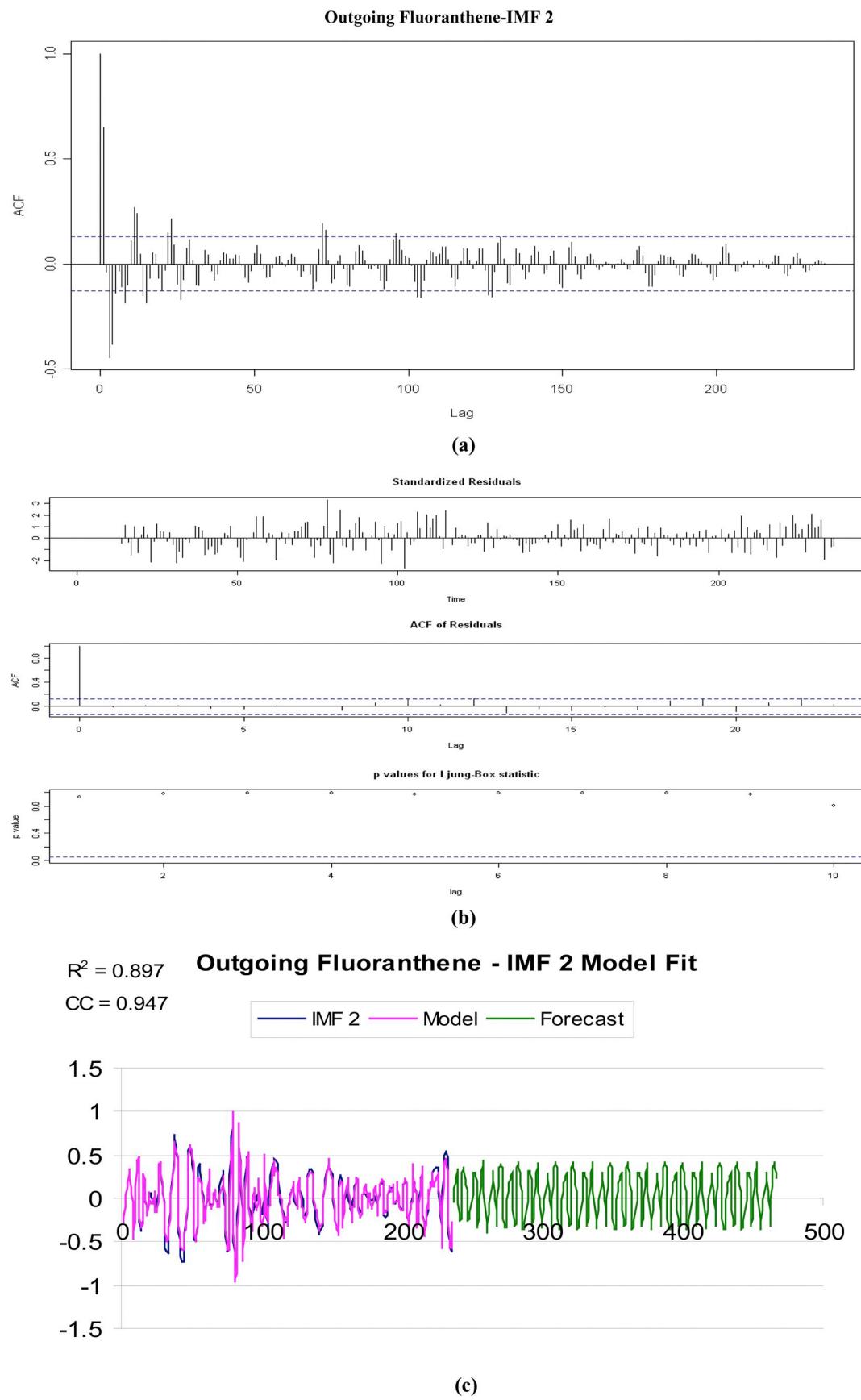
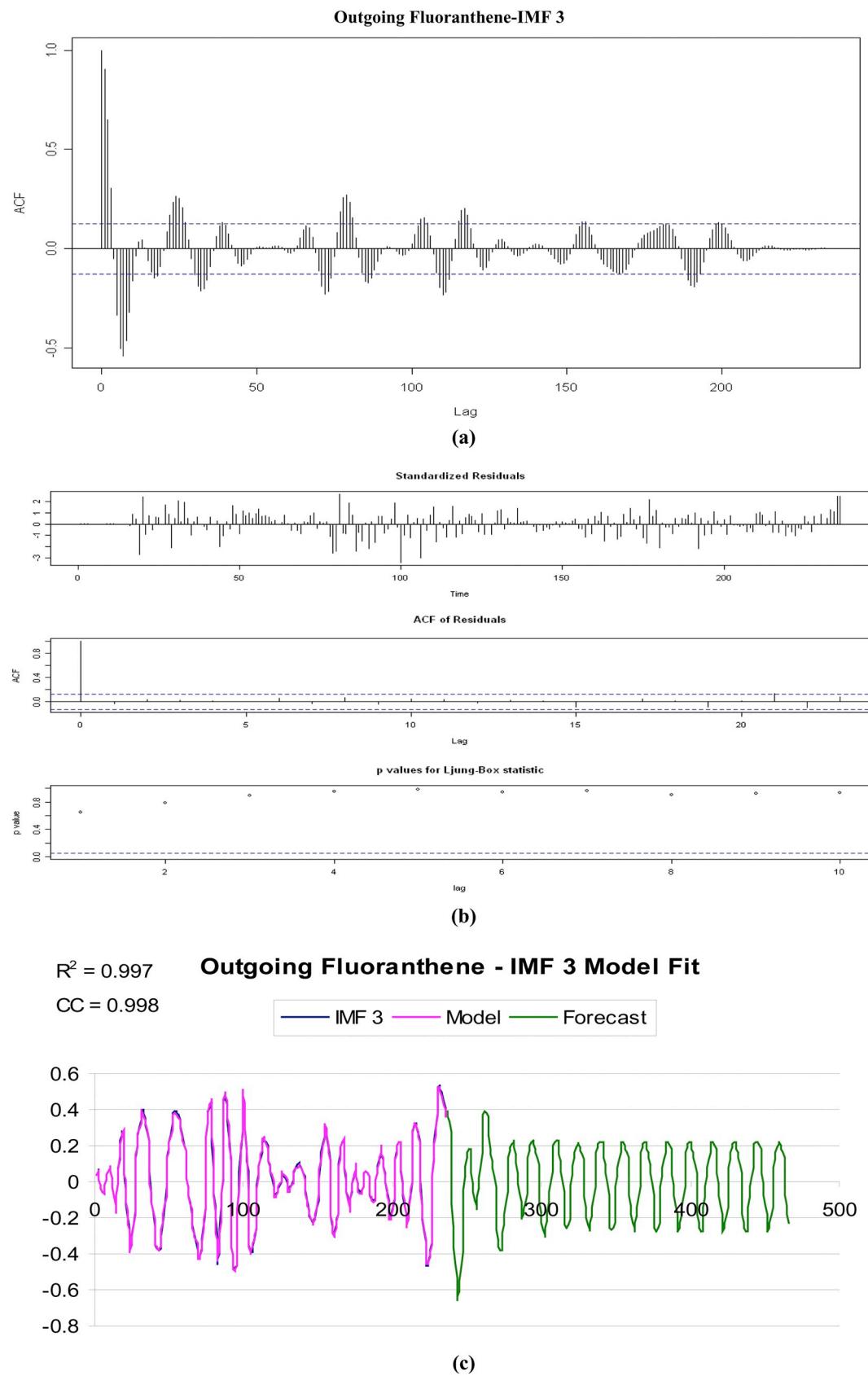
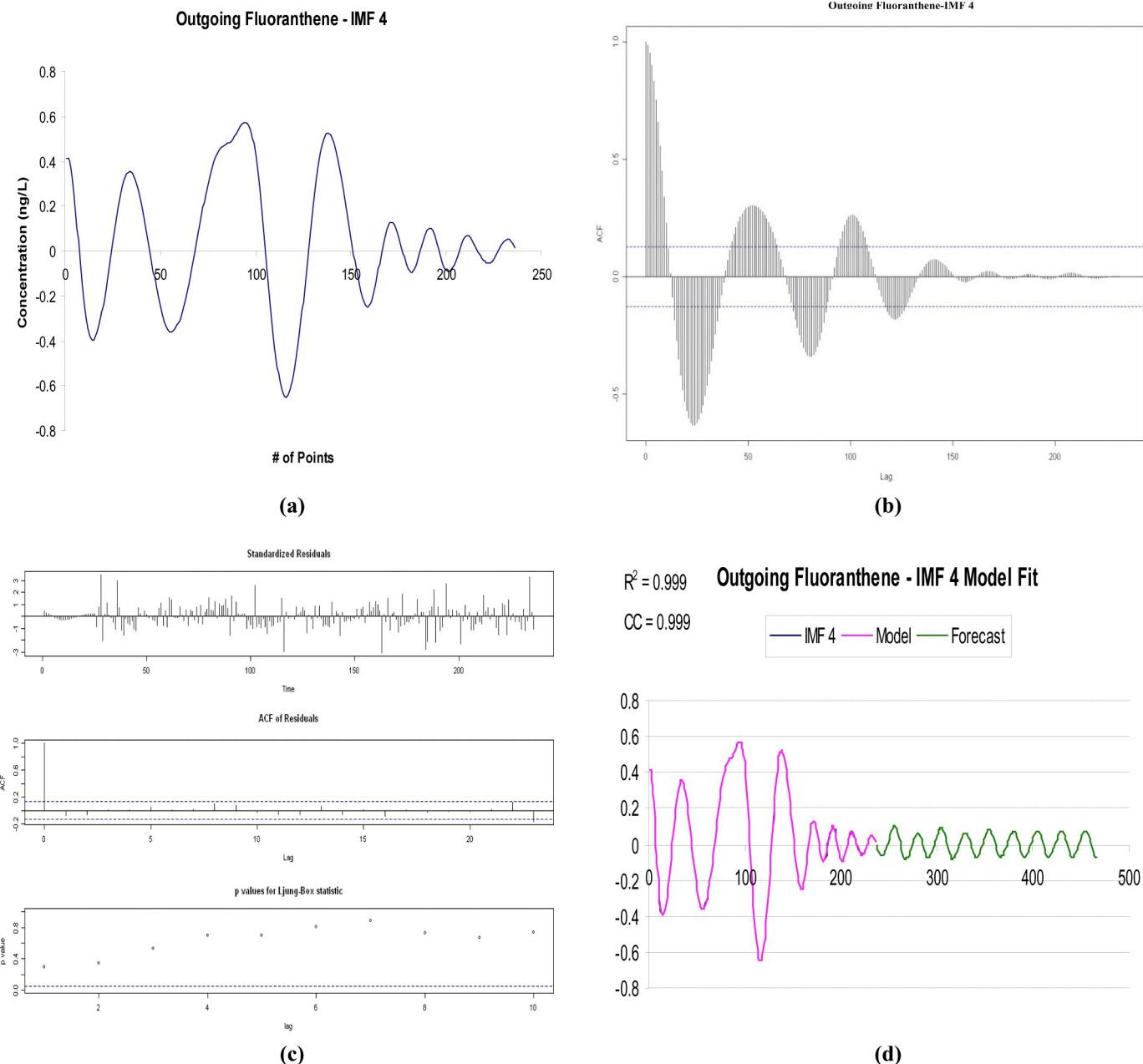


Fig. 5. (a) Autocorrelation Function of IMF 2. (b) Model Fit Diagnostics for IMF 2. (c) IMF 2 Model and Forecast.



**Fig. 6.** (a) Autocorrelation Function of IMF 3. (b) Model Fit Diagnostics for IMF 3. (c) IMF 3 Model and Forecast.



**Fig. 7.** (a) IMF 4 of Fluoranthene Concentration. (b) Autocorrelation Function of IMF 4. (c) Model Fit Diagnostics for IMF 4. (d) IMF 4 Model and Forecast.

**Fig. 4(a).** The decision of how to model the residual trend was based on the nature of the trend itself. The residual starts with an increasing trend, and then decreases until it begins to increase again toward the end of the series. It was decided to use the linear regression analysis method, and use a linear model with a sine and cosine term to carry out the increasing trend at the end of the residual. Once again, a period was needed as an input to the model. This was estimated from the ACF of IMF 7 found in Fig. 10(a).

An assumption was made that the second maximum would be equally spaced from the minimum. Therefore, by doubling the location of the minimum (about 150 weeks), an approximate seasonal period was found to be 300 weeks. The resulting fit of the model is found in Fig. 10(b). Fig. 10(b) shows that the model is able to fit the simple curve very well. A summary of the model details is found in Table 3(g).

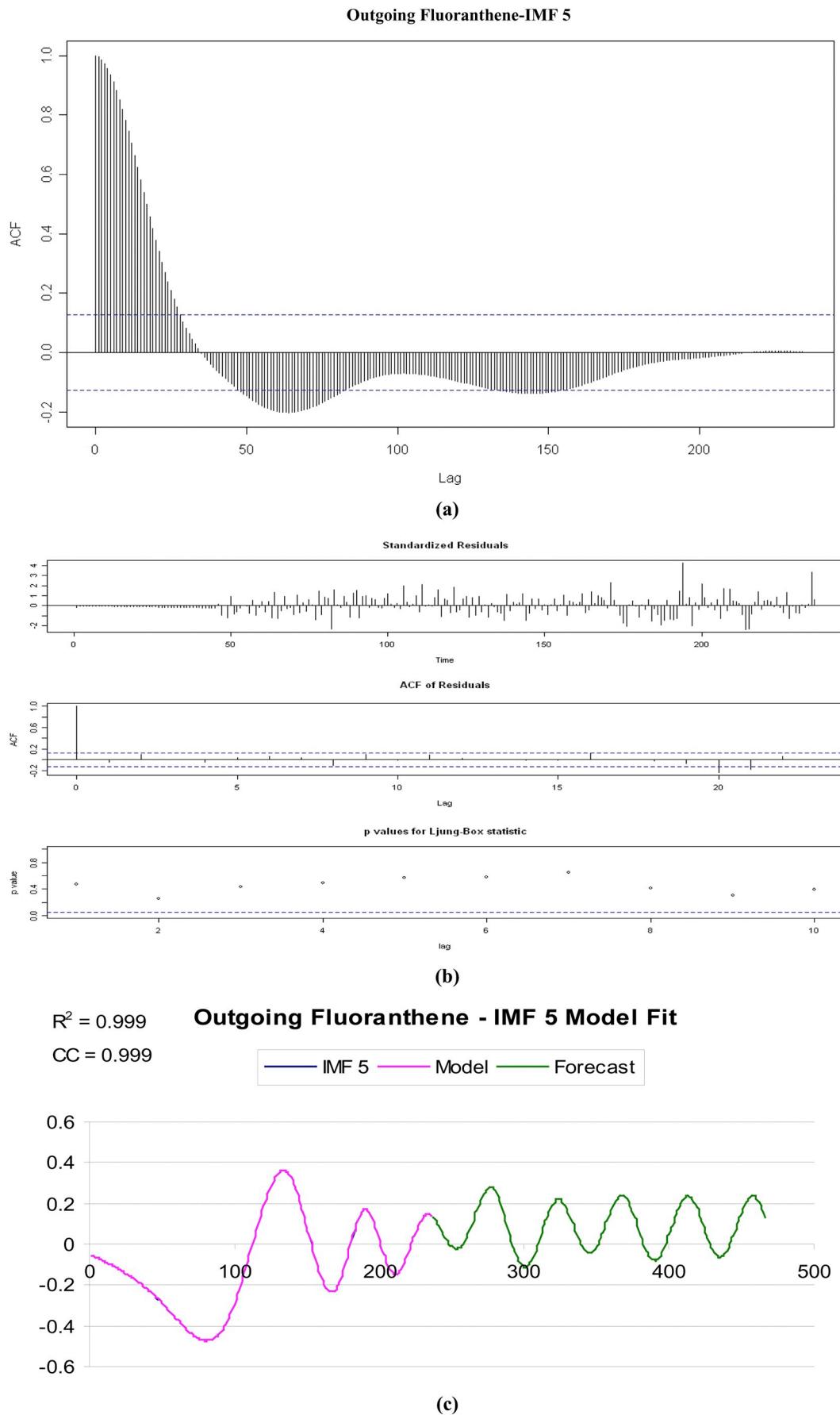
Table 3(f) confirms the fit of the model, showing both multiple and adjusted R-squared values of 0.999. It also shows that the sine and cosine terms are very significant. The forecast based on this model is

found in Fig. 10(c). The forecast in Fig. 10(c) performs well in terms of carrying out the last part of the original trend, and then appears to start the cycle over again.

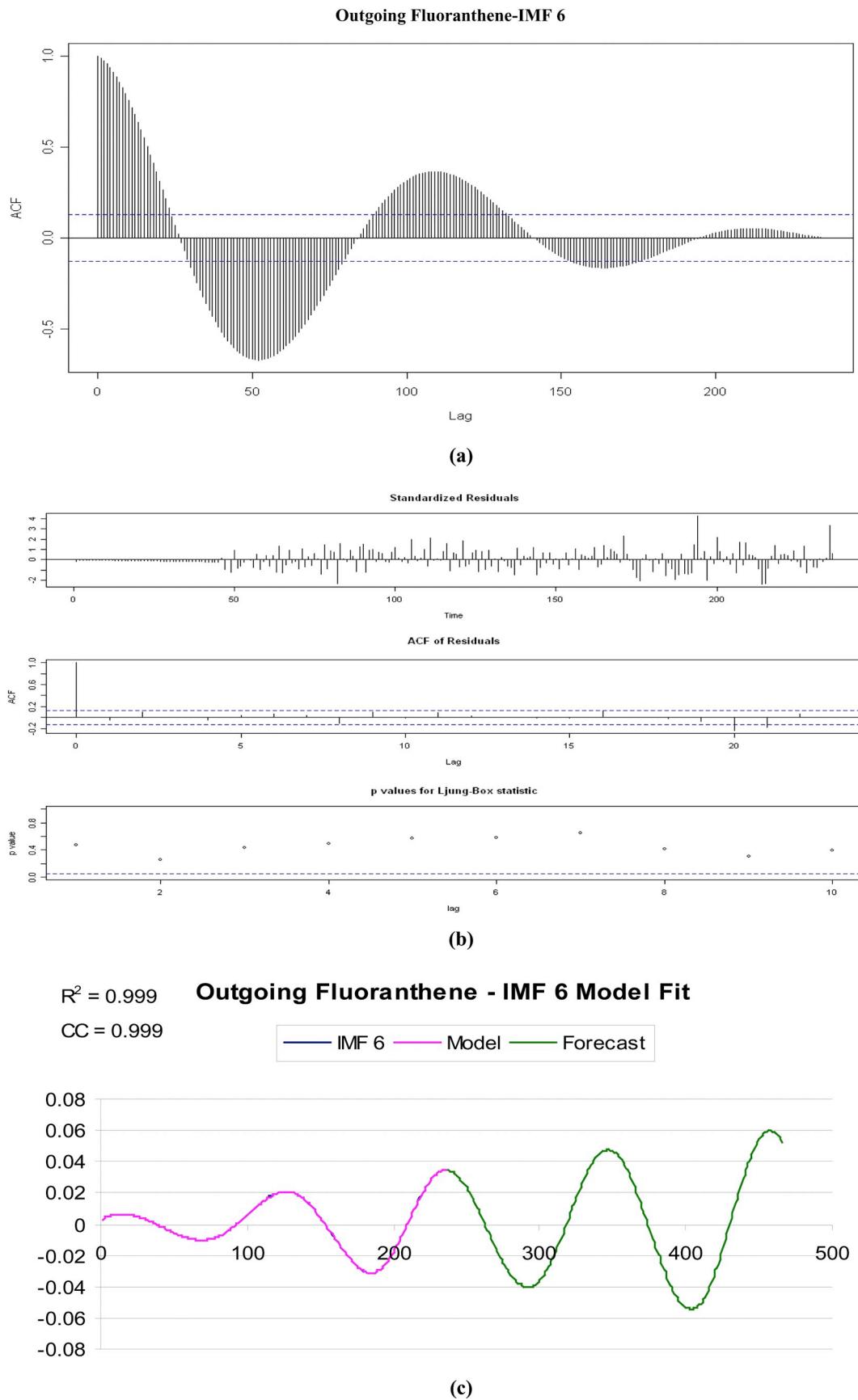
### 3.1.4. Hypothesis discussion – modeling w/ HHT vs. Modeling w/o HHT

Now that each component of the decomposed fluoranthene data has been modeled and forecasted, it is a simple task to add them back together and reverse the log transformation (by taking the exponential) to obtain a complete model of the original fluoranthene data set. The final result of the fluoranthene modeling effort is presented in Fig. 11(a). The model was able to fit the data with an  $R^2$  value of 0.779 and has a strong correlation coefficient of 0.882.

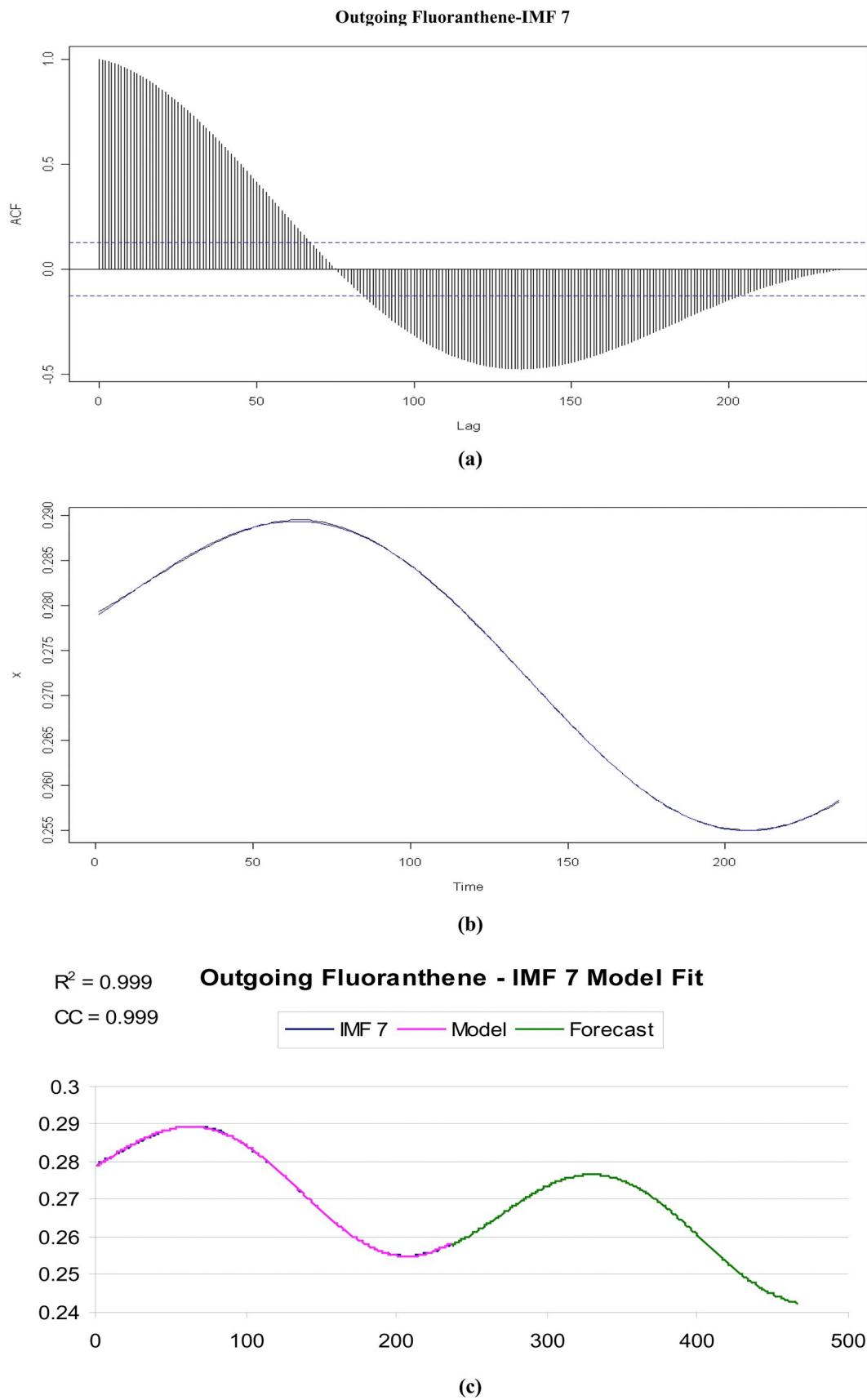
The hypothesis of this study proposed that pre-processing the contaminant time series data with HHT, coupled with ARMA modeling and linear regression analysis, would produce a model with a higher  $R^2$  value than modeling the data without HHT. The following figure (Fig. 11(b)) contains the same fluoranthene data set, but this time,



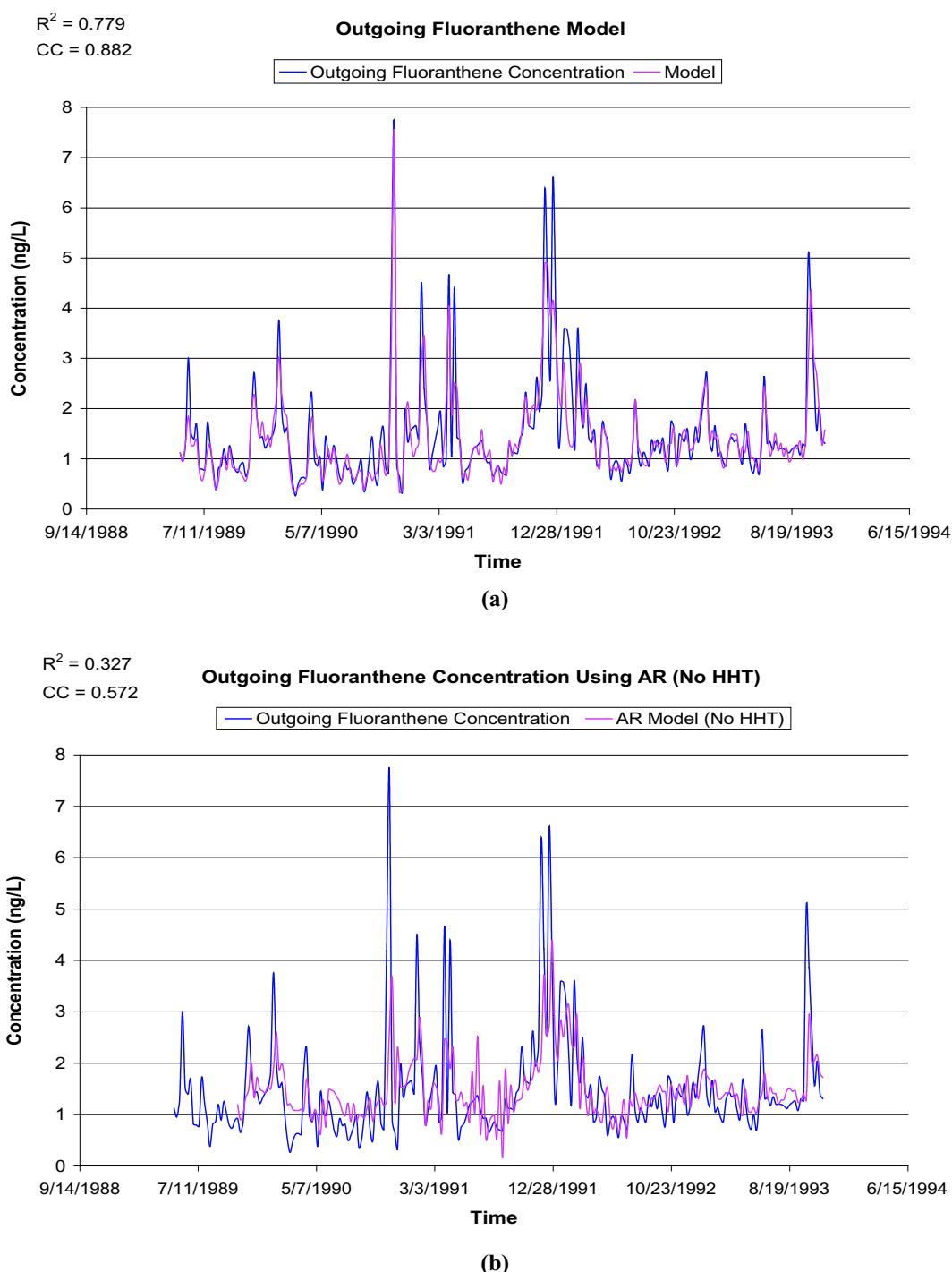
**Fig. 8.** (a) Autocorrelation Function of IMF 5. (b) Model Fit Diagnostics for IMF 5. (c) IMF 5 Model and Forecast.



**Fig. 9.** (a) Autocorrelation Function of IMF 6. (b) Model Fit Diagnostics for IMF 6. (c) IMF 6 Model and Forecast.



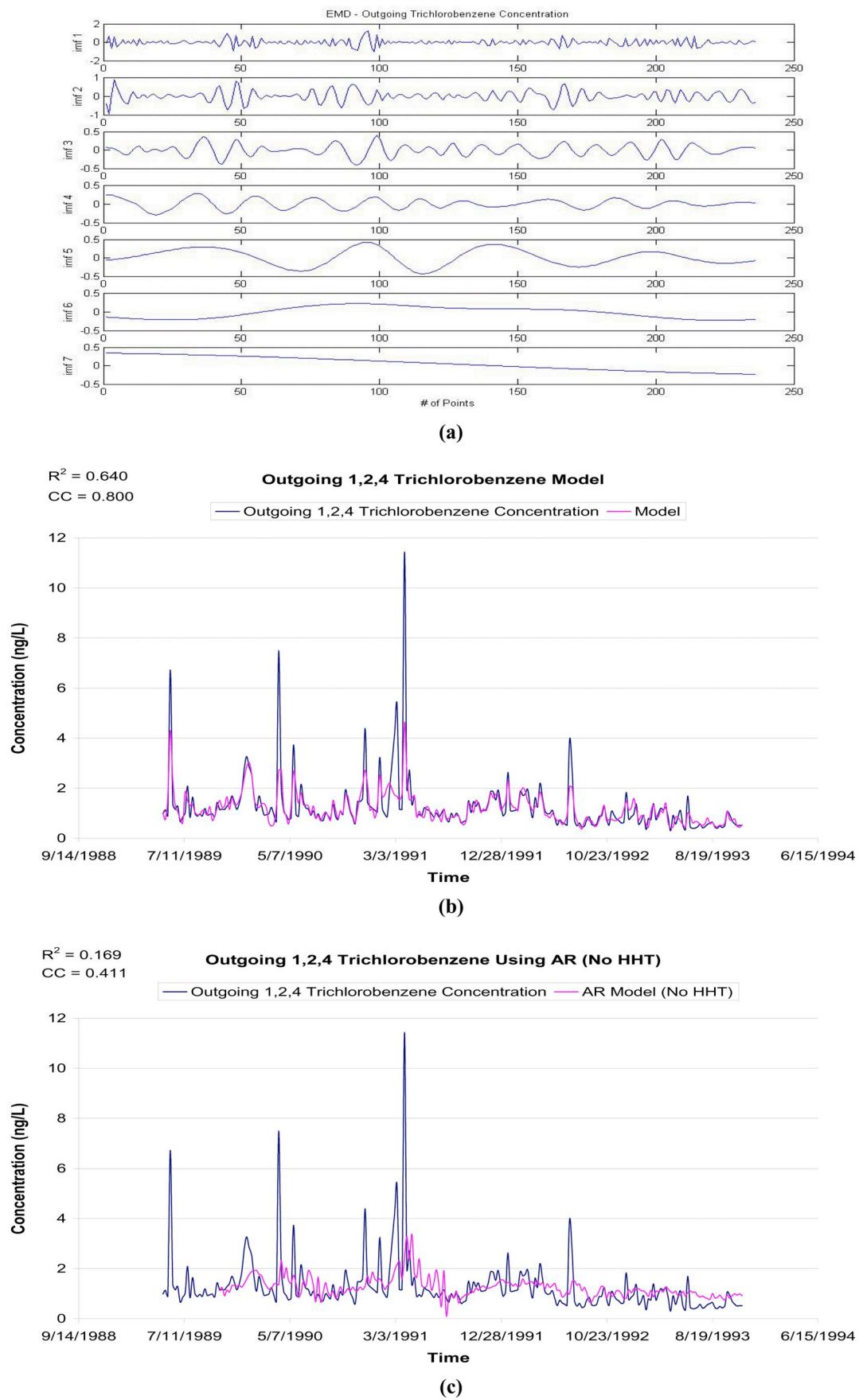
**Fig. 10.** (a) Autocorrelation Function of Fluoranthene Residual. (b) Fluoranthene Residual Model Fit. (c) Fluoranthene Residual Trend Model and Forecast.



**Fig. 11.** (a) Fluoranthene Concentration Model ARMA with HHT. (b) Fluoranthene Concentration AR Model without HHT.

**Table 4**  
Model performance summary (ARMA w/ HHT vs AR w/o HHT).

Contaminant concentration	$R^2$ Value		Correlation coefficient		Sum of squared errors	
	HHT	No HHT	HHT	No HHT	HHT	No HHT
Fluoranthene	0.779	0.327	0.882	0.572	55.48	195.25
Dieldrin	0.721	0.534	0.849	0.731	0.16	2.22
1,2,4 Trichlorobenzene	0.640	0.169	0.800	0.411	124.20	288.65



**Fig. 12.** (a) IMFs of Log Transformed 1,2,4, Trichlorobenzene Concentration Data. (b) 1,2,4 Trichlorobenzene Concentration Model. (c) 1,2,4 Trichlorobenzene Concentration AR Model w/o HHT.

**Table 5**  
1,2,4 Trichlorobenzene IMF Model Summary.

IMF	Modeling method	Parameter selection	Seasonal period	R <sup>2</sup> -Value	Correlation coefficient
1	ARMA	p = 23, q = 3	1	0.221	0.470
2	ARMA	p = 5, q = 3	7	0.869	0.932
3	ARMA	p = 4, q = 1	15	0.995	0.997
4	ARMA	p = 5, q = 3	24	0.999	0.999
5	ARMA	p = 4, q = 3	54	0.999	0.999
6	ARMA	p = 4, q = 3	100	0.999	0.999
Residual trend	Linear regression	$\gamma = 0.0392$ $\beta = 0.0296$	322	0.999	0.999

modeled with an AR model without pre-processing the data with the empirical mode decomposition (EMD) process of HHT. When comparing Figs. 11(a) and 10(b), the advantage of coupling EMD with ARMA modeling becomes apparent. The AR model in the latter image is unable to fit the peaks of the measured data. The model appears to follow the moving average of the data without matching the frequency with great accuracy. The AR model without HHT matches the data with an R<sup>2</sup> value of 0.327 compared to 0.779 with the proposed procedure. The results of hypothesis testing for concentrations of fluoranthene, 1,2,4 trichlorobenzene and dieldrin are summarized in the following sections.

The hypothesis was tested on all data sets, comparing R<sup>2</sup> values of models created with HHT to models created without HHT. The results are presented in Table 4. Sum of the squared errors is shown in Table 4 to compare the models created with and without HHT. The lower values of the squared errors show the better performance of the models created with HHT compared to the models created without HHT.

### 3.1.5. 1,2,4 Trichlorobenzene concentration (measured at Niagara on the lakes)

The log-transformed 1,2,4 trichlorobenzene concentration data was processed using empirical mode decomposition to create Fig. 12(a), which contains six intrinsic mode functions and a residual. By analyzing each function, its associated autocorrelation function and referring to the procedural guidelines of Fig. 2, each IMF was modeled accordingly. The details of each modeled IMF and residual trend is summarized in Table 5.

The resulting 1,2,4 trichlorobenzene concentration model is presented in Fig. 12(b). The model fits the data with an R<sup>2</sup> value of 0.64 and a correlation coefficient of 0.80. To test the significance of pre-processing the data with HHT, the 1,2,4 trichlorobenzene data was modeled with an AR model without HHT. The resulting model is presented in Fig. 12(c). Without pre-processing the data with the EMD process of HHT, the AR model fits the data with an R<sup>2</sup> value of 0.17 and a correlation coefficient of 0.41. Pre-processing the data with HHT produces a better fit model.

### 3.1.6. Dieldrin concentration (measured at Niagara on the lakes)

The log-transformed dieldrin concentration data were processed using the empirical mode decomposition method of HHT. The resulting intrinsic mode functions and residual are contained in Fig. 13(a). By analyzing each function, its associated autocorrelation function and referring to the procedural guidelines of Fig. 3, each IMF was modeled accordingly. The details of each modeled IMF and residual trend is summarized in Table 6.

The resulting dieldrin concentration model is presented in Fig. 13(b). The model fits the data with an R<sup>2</sup> value of 0.72 and a correlation coefficient of 0.85. To test the significance of pre-processing the data with HHT, the dieldrin data were modeled with an AR model without HHT. The resulting model is presented in Fig. 13(c). Without pre-processing the data with the EMD process of HHT, the AR model fits

the data with an R<sup>2</sup> value of 0.53 and a correlation coefficient of 0.73. Pre-processing the data with HHT produces a better fit model.

## 4. Discussion

### 4.1. Discussion of IMFs

The EMD process of HHT allows for the decomposition of nonlinear and non-stationary data sets. In general, IMF 1 contains the majority of the noise of the function. Modeling IMF 1 is the most difficult and has the poorest model performance. Due to poorer performing models, IMF 1 is the limiting factor (inhibits a higher R<sup>2</sup> value) to the overall model performance of each contaminant concentration.

IMFs 2 through 5 were found to be high to intermediate in frequency (decreasing with each consecutive IMF). IMF 6 was found to be sinusoidal in nature. When the frequency and amplitude of the IMF 6 signals were generally constant, linear modeling with sine and cosine predictors was used.

IMF 7 contained the residual trend of each data set. The trend of the contaminant concentrations over time is not always clear from a plot of the original time series due to the high fluctuations and variability in the data (nonlinearity and non-stationarity). Decomposing the data completely allows for the overall trend of the data to be very clear, revealing to the user the general increasing or decreasing nature of the signal over time. Revealing the trend of each concentration is likely the most important contribution of the EMD process. The plots of the concentrations of 1,2,4 trichlorobenzene and dieldrin elude to the trend of the data without considerable efforts. However, the trends of the fluoranthene data sets are not obvious without the implementation of the EMD process. The trend of each contaminant data set analyzed in this study is as follows:

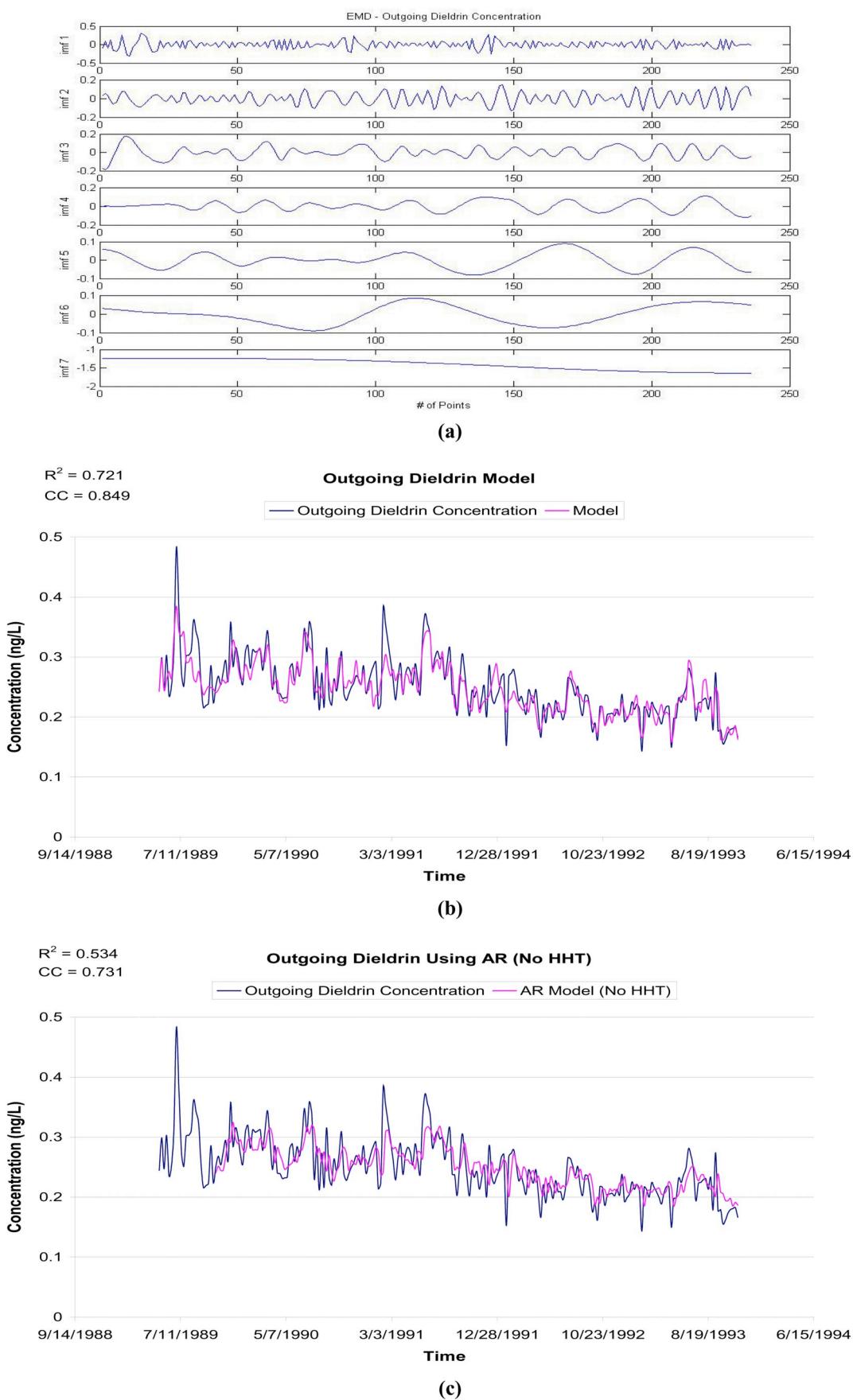
- (1) Fluoranthene – Decreasing, then beginning to increase
- (2) 1,2,4 Trichlorobenzene – Decreasing
- (3) Dieldrin – Decreasing.

### 4.2. Discussion of model performance

Combining the EMD process of HHT with ARMA modeling and linear regression analysis, three contaminant concentrations were analyzed and modeled. The resulting models fit the data with R<sup>2</sup> values ranging from 0.64 (1,2,4 trichlorobenzene) to 0.779 (fluoranthene). Modeling the data sets without HHT resulted in R<sup>2</sup> values ranging from 0.17 (1,2,4 trichlorobenzene) to 0.327 (fluoranthene). In all cases, pre-processing the data with HHT resulted in a better fit model.

Aside from creating a more accurate model, HHT provides other advantages to the field of water quality modeling. One of the most important advantages is the ability of HHT to decompose contaminant concentrations and reveal the overall trend of the data. For multiple time series (particularly the fluoranthene series), the trend of the data is not obvious to decipher by visual inspection alone. The EMD process of HHT provides the trend of the data which could be very important when evaluating the status of contaminant levels in a water body. This knowledge would be especially useful when discussing regulatory issues. HHT could help provide insight by revealing contaminants of increasing concentration, perhaps leading to stricter permits on municipal or industrial discharges.

Previous work by Franceschini and Tsai (2010) modeled fluoranthene and chrysene concentrations with HHT and AR models. The resulting R<sup>2</sup> values of the models were 0.81 and 0.63, respectively. No obvious advantage in modeling performance can be claimed by the procedure proposed in this study. However, utilizing ARMA and linear models with sine and cosine predictors does enhance the forecasted concentrations by including an approximation of the natural frequencies of the original data. This could be useful for estimating future maximum or minimum concentrations, rather than average values from



**Fig. 13.** (a) IMFs of Log Transformed Dieldrin Concentration Data. (b) Dieldrin Concentration Model. (c) Dieldrin Concentration AR Model w/o HHT.

**Table 6**  
Dieldrin IMF model summary.

IMF	Modeling method	Parameter selection	Seasonal period	R <sup>2</sup> -Value	Correlation coefficient
1	ARMA	p = 23, q = 2	3	0.114	0.338
2	ARMA	p = 5, q = 3	6	0.813	0.902
3	ARMA	p = 4, q = 1	16	0.996	0.998
4	ARMA	p = 5, q = 2	25	0.999	0.999
5	ARMA	p = 5, q = 3	53	0.999	0.999
6	ARMA	p = 3, q = 2	97	0.999	0.999
Residual trend	Linear regression	$\gamma = 0.0969$ $\beta = 0.0103$	297	0.999	0.999

a “trend only” forecast.

#### 4.3. HHT vs. FFT

In the past, the Fast Fourier Transform (FFT) has been a very common method for analyzing time series. However, the FFT can only be properly applied under certain conditions. Two such conditions are: 1) the data must be linear and 2) the data must be strictly periodic or stationary (Huang et al., 1998). In this study, the contaminant concentration data can be classified as both nonlinear and non-stationary. Fig. 14 contains the Hilbert-Huang spectrum of the dieldrin data set. The dieldrin concentration is dominated by intermittent frequency which suggests nonlinearity. The high frequency modulation indicates that the time series is non-stationary. Therefore, it is not preferred to use the Fourier transform as its applicability is restricted. Figs. 14 and 15 provide a comparison of the FFT and HHT spectrums for the dieldrin concentration data. The differences between the two methods are generally consistent throughout all of the data sets.

Figs. 14 and 15 show multiple differences in the two methods for analyzing the dieldrin concentration. The most apparent advantage of HHT versus FFT is the inclusion of a time scale. The HHT spectrum

(Fig. 14) shows the varying strengths (or energies) of at least four different time scales from the data. The frequencies listed on the y-axis represent different interval time scales: 0.001–0.003 Hz (one to three-year variation), 0.03 Hz (monthly variation) and 0.05 Hz (biweekly variation). The dieldrin concentration has stronger energies in the lower frequencies (0.002–0.02 Hz). This suggests that the dieldrin concentration is the result of natural or anthropogenic activities that may repeat at a seasonal or yearly periodicity. The FFT image (Fig. 15) shows only the relative signal strength (amplitude) at varying frequencies. The FFT also shows that the dominant signal is at the zero frequency, which is physically incorrect.

## 5. Conclusions and recommendations

### 5.1. Conclusions

The Hilbert Huang Transform (HHT) has only recently been introduced to the field of water quality analysis for its ability to analyze non-stationary and nonlinear time series. In this study, HHT has been combined with the time series analysis (autoregressive moving average (ARMA) modeling) and linear regression analysis to model and forecast three contaminant concentration data sets which have been shown to be both non-stationary and non-linear in nature. The hypothesis “combining HHT with a time series analysis in general can better treat datasets with nonlinear and non-stationary features compared to modeling without HHT” was presented and validated in this study.

The EMD process of HHT allows for the decomposition of nonlinear and non-stationary data sets into more stationary functions called IMFs. In general, IMF 1 contained the majority of the noise of each data set. IMF 1 is the limiting factor to the overall model performance due to its highly variable nature. IMFs 2 through 5 contained high to intermediate frequencies and were modeled using ARMA models with an expanded seasonal term. IMF 6 was found to be sinusoidal in nature and was modeled using ARMA modeling or linear regression analysis when the frequency and amplitude of the function were consistent. IMF 7

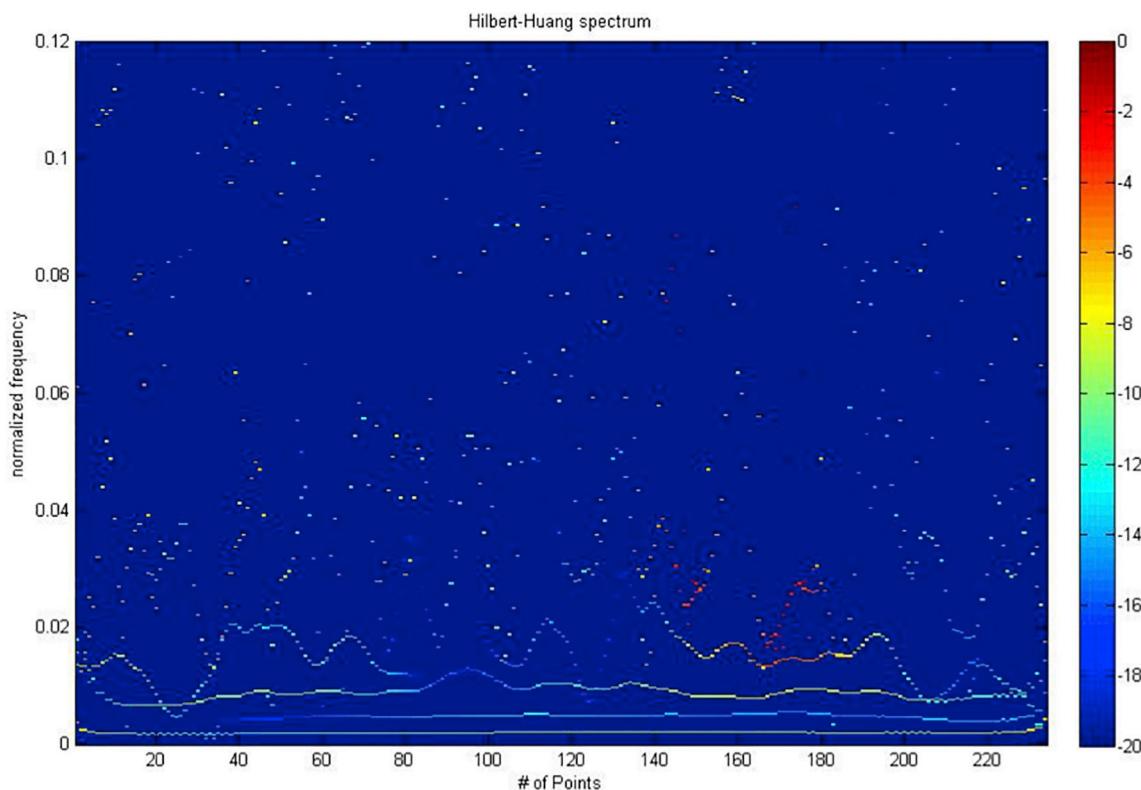
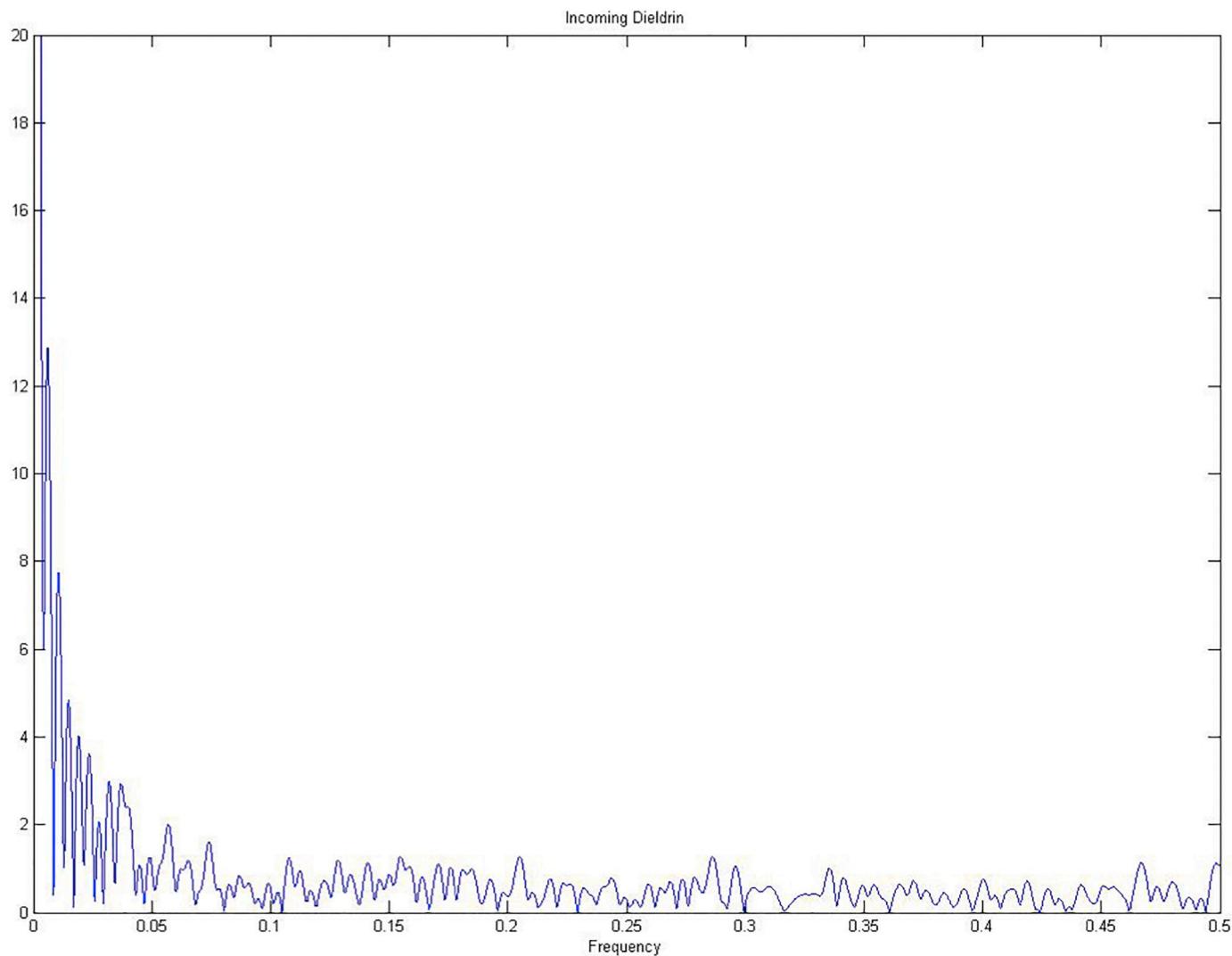


Fig. 14. HHT Spectrum of Dieldrin Concentration.



**Fig. 15.** FFT Spectrum of Dieldrin Concentration.

contained the trend of each data set and was modeled using linear regression analysis.

It can be concluded that the trend of fluoranthene concentrations from April of 1986 to March of 1997 is decreasing and then beginning to increase; the 1,2,4 Trichlorobenzene concentrations are decreasing; while the dieldrin concentrations are decreasing. With HHT, appropriate time series models can be identified and constructed for the studied contaminant concentrations to better illustrate the variability of each IMF (and thus the contaminant concentrations) for the studied period. For all data sets modeled in this study, pre-processing the data with HHT allowed for higher  $R^2$  values, correlation coefficients and lower sum of squared errors when compared to modeling without HHT. As such, it can be suggested pre-processing the data with HHT would enhance the effectiveness and accuracy of the data interpretation and model predictability.

In conclusion, this study aims at better illustrating the contaminant concentration data of a non-stationary and non-linear nature. Three priority contaminants measured at Niagara-on-the-Lakes are selected for this study. With HHT, appropriate time series models can be identified and constructed for the studied contaminant concentrations.

Results of this study, in general, can assist in contamination source identification at various temporal scales (such as potential sources at weekly, seasonally, and yearly level etc.). Information on the

contamination at different characteristic frequencies, if available, can provide some guidance for the planning and management of the contamination remediation efforts at various levels. Residual trends of the contamination from the HHT analysis can provide a quantitative indicator for assessing the effectiveness of river remediation and ecological restoration for river sustainability.

## 5.2. Recommendations

Although the results of the tested hypotheses presented in this study had overall positive outcomes, the procedure is not without limitations. For example, the seasonal period incorporated into each model was selected by visual inspection of the ACF of the IMF or IMF signal itself. Applying a “best fit” method of selecting the seasonal period would help remove the subjectivity of selecting the parameter and may improve model performance and robustness.

Regarding model performance, IMF 1 appears to be the limiting factor in creating an overall model to fit the raw data. The challenge with modeling IMF 1 is its inherent noisy nature. Improving the fit of the IMF 1 model could greatly improve the fit of the overall model, which may translate into an improved forecast. Modeling IMF 1 with random number generation may be an alternative to ARMA modeling.

## Acknowledgement

The authors gratefully acknowledge the financial support from National Science Foundation under grant contract number EAR-0748787. The authors would also like to thank Mr. S. M. Mousavi at National Taiwan University for his assistance in preparing the revised manuscript.

## References

- Bernal, J.L., Cummins, S., Gasparrini, A., 2017. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int. J. Epidemiol.* 46 (1), 348–355.
- Brockwell, P.J., Davis, R.A., 2002. *Introduction to Time Series and Forecasting*. Springer-Verlag, NY.
- Desai, A.M., Rifai, H.S., 2010. Variability of Escherichia coli concentrations in an urban watershed in Texas. *ASCE J. Environ. Eng.* 136 (12), 1347–1359.
- Elgar, F.J., Pförtner, T.-K., Moor, I., De Clercq, B., Stevens, G.W.J.M., Currie, C., 2015. Socioeconomic inequalities in adolescent health 2002–2010: a time-series analysis of 34 countries participating in the Health Behaviour in School-aged Children study. *Lancet* 385 (9982), 2088–2095.
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Verlag.
- Franceschini, S., 2007. Development of a Framework to Account for Uncertainty Sources in Modeling Toxic Concentrations in the Niagara River. Ph.D. Dissertation. SUNY at Buffalo, Buffalo, NY.
- Franceschini, S., Tsai, C., 2010. Application of Hilbert-Huang Transform method for analyzing toxic concentrations in the Niagara River. *J. Hydrol. Eng. ASCE* 15 (90), 90–96.
- Gairola, G.S., Chandrasekhar, E., 2017. Heterogeneity analysis of geophysical well-log data using Hilbert-Huang transform. *Physica A* 478, 131–142.
- Gedikli, A., Aksoy, H., Unal, N.E., 2008. Segmentation algorithm for long time series analysis. *Stoch. Env. Res. Risk A.* 22 (3), 291–302.
- Huang, N.E., Wu, Z., 2008. A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Rev. Geophys.* 46, RG2006. <https://doi.org/10.1029/2007RG000228>.
- Huang, N.E., Zheng, S., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A454*, 903–995.
- Kuai, K.Z., Tsai, C., 2012. Identification of varying time scales in sediment transport using the Hilbert-Huang Transform method. *J. Hydrol.* 401–402, 245–254. <https://doi.org/10.1016/j.jhydrol.2011.12.007>.
- Lohani, B.N., Wang, M.M., 1987. Water quality data analysis in Chung Kang River. *ASCE J. Environ. Eng.* 113 (1), 186–195.
- Melching, C.S., Anmangandla, S., 1992. Improved first-order uncertainty method for water-quality modeling. *ASCE J. Environ. Eng.* 118 (5), 791–805.
- Mendenhall, W., Sincich, T., 2007. *Statistics for Engineering and the Sciences*. Pearson Prentice-Hall, Upper Saddle River, N.J.
- Safari, N., Chung, C.Y., Price, G.C.D., 2018. Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis. *IEEE Trans. Power Syst.* 33 (1), 590–601.
- Shao, Q., Li, Z., Xu, Z., 2010. Trend detection in hydrological time series by segment regression with application to Shiyang River Basin. *Stoch. Env. Res. Risk A.* 24 (2), 221–233.
- Tsai, C.W., Franceschini, S., 2005. Evaluation of probabilistic point estimate methods in uncertainty analysis for environmental engineering applications. *ASCE J. Environ. Eng.* 131 (3), 387–395.
- Tsai, C.W.\*., Li, M., 2014. Uncertainty analysis and risk assessment of DO concentrations in the Buffalo River using the Perturbation moments method. *ASCE J. Hydrol. Eng.* 19 (12), 04014032 (12 pp.). [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000985](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000985).
- Tsai, C.W., Yeh, T.-G., Hsiao, Y.-R., 2018. Evaluation of hydrologic and meteorological impacts on dengue fever incidences in southern Taiwan using time-frequency analysis methods. *Ecol. Informat.* 46, 166–178.
- Wu, Z., Huang, N.E., Wallace, J.M., Smoliak, B.V., Chen, X., 2007. On the time-varying trend in global-mean surface temperature. *Clim. Dyn.* 37 (3), 759–773. <https://doi.org/10.1007/s00382-011-1128-8>.