

# Times Series Methods for Modeling and Simulations of Terrain Profiles

**Jinfeng Wei**

Department of Mathematics  
Maryville University of St. Louis  
650 Maryville University Dr, St. Louis, MO 63141, USA

**T. C. Sun**

Department of Mathematics  
Wayne State University  
42 W Warren Ave, Detroit, MI 48202, USA

## Abstract

*In this paper time series methods are used to model two terrain profiles obtained by the US Army at the Aberdeen Proving Grounds, Maryland and the Yuma Test Facility, Arizona. One profile has nonstationary behavior but has a linear structure. We can model it quite well by using the uniformly modulated processes introduced by Priestley. The other profile is also nonstationary but has a nonlinear structure. We model it with 3 different methods: ARMA-GARCH, TAR and EMD methods of which the EMD method may not be familiar to most statisticians. They all seem to fit it well. We use the linear damage theory in mechanics to make a more appealing fitness check of the 3 methods. We believe the experience we gained in this study may provide some insight into other nonstationary and nonlinear time series model builders.*

## 1 Introduction

To fit a non-stationary and non-linear time series model, based on an observed sample, is a very delicate task which requires not only the finding of a suitable method but also a deeper understanding of the data and the need of a diagnostic checking method in addition to the usual way of just looking at the auto-covariance function of the residues. In general, we may find several models which all fit the data reasonable well, but the question is whether we can select one which is the best in some sense. In this paper, we shall study these modeling problems by working with two sets of real data of terrain profiles and we shall also consider the fitness question in the study of the second set of data. We shall refer to the references [9], [11], [15], [18] and [19] for the basic non-stationary and non-linear models, and we would like also to mention the paper [20] for an in-depth discussion of the difficulty of non-linear modeling. In this paper we shall study two sets of data, collected from the US Army vehicle test courses. One set of data is collected from Belgian Block test track, at Aberdeen, Maryland, which is a man-made course constructed of bricks and is relatively stable (Fig.1). The data are of the form

$$X_t, \quad t=1, 2, \dots, 11,000,$$

Where  $t$  represents the distance from the starting point with each unit being 4 inches and  $X_t$  is the height of the terrain profile at  $t$ . The other set of data is from Perryman3 off road track, at Yuma, Arizona, which is a course that emulates the characteristics of travelling off road and is therefore very rough (Fig.1). The data are given in a similar form

$$X_t, \quad t=1, 2, \dots, 55,000.$$

We are interested in finding computer models to simulate these terrain profiles. The simulated terrain profiles can then be used as input data to vehicle shakers in laboratories for testing the durability of vehicles or as input to computer analysis of vehicle dynamics. Another use of the simulated profiles is the study and quantification of the roughness of each test course. We decide to use statistical models to fit the data for the following reason. Every time we measure the terrain profiles we would get a different set of data due to the measuring errors and due to the fact that the linear tracks on which the measuring vehicle travels cannot be exactly the same each time. However the data collected from each course at different times should share similar intrinsic properties. Therefore each measured data profile is like a sample from a stochastic process.

In this paper we shall use time series models with time being the distance from the starting point. If a time series is linear, Gaussian and stationary, a well-studied autoregressive moving average (ARMA (p, q)) model can be used with proper orders p and q ([3]). For non-stationary data, one type of non-stationary time series which is structurally very close to a stationary process is called the uniformly modulated process ([15]). If the series is nonlinear, various ad hoc nonlinear methods must be considered.



**Figure 1: A Snake View of Belgian Block**



**Figure 2: A Snake View of Perryman3 Terrain**

	Test Statistic Value	Critical Value	Conclusion
Stationarity Test	876.4331	16.9190	Non-stationary
Linearity Test	16.4979	166.4153	Linear
Gaussianity Test	54.3644	266.3781	Gaussian

Table 1: the Statistical Tests for One Segment of Belgian Block Profile

	Test Statistic Value	Critical Value	Conclusion
Stationarity Test	1.9482e+003	16.9190	Non-stationary
Linearity Test	400.9567	166.4153	Non-linear
Gaussianity Test	2.3693e+003	266.3781	Non-Gaussian

Table 2: the Statistical Tests for One Segment of Perryman3 Profile

We have used Keenan’s test for linearity ([12]), Kolmogorov-Smirnov’s test for Gaussianity and we tested for stationarity by comparing the mean and variance for different segment of the data. The study in [4] concludes that Belgian Block data is linear, Gaussian, but not stationary, and Perryman3 data are neither linear, nor Gaussian, nor stationary. The Belgian Block test track is much better-behaved than the Perryman3 off-track terrain.

Based on the statistical test results for these two testing courses, we are proposing the following time series models for them respectively. In Section 2, a uniformly modulated process is used to model and to simulate the Belgian Block profile. In Section 3, three different non-linear models are proposed for the Perryman3 data profile: (1) ARMA-GARCH model, (2) TAR model, and (3) Empirical Mode Decomposition (EMD) method. They all seem to model the data well. In Section 4, we use the linear damage theory in fatigue mechanics and compare the rain flow cycles of the simulation results for these three nonlinear models in order to make a better selection. The EMD decomposition method is not well known among statisticians. We believe it may be of interest for further study in more details by people that are interested in time series.

## 2 Modeling of Belgian Block Profile

In [4], the Belgian Block profile is tested and found to be linear, Gaussian, but not stationary. To model and simulate a non-stationary time series, we propose to use the uniformly modulated process (UMP) ([15]).

### 2.1 Uniformly Modulated Process (UMP)

Let  $\{X_t, t=0, \pm 1, \pm 2, \dots\}$  be a discrete parameter process. It is a uniformly modulated process if it is of the form

$$X_t = c(t) X_{0,t}$$

where  $c(t)$  is a deterministic function and  $X_{0,t}$  is a stationary process, or its evolutionary spectral density functions  $h_t(\omega)$  is of the form

$$h_t(\omega) = c_t^2 h(\omega)$$

where  $h(\omega)$  is a spectral density function. When  $h_t(\omega)$  is independent of  $t$  or  $c(t)$  is a constant, then the time series is stationary. To test whether  $X_t$  is a uniformly modulated process, we follow the procedures given in Priestley’s book ([15]), pp174-183.

### 2.2 UMP Modeling of Belgian Block Profile

To illustrate the method, we only use a segment of the data  $X_t, t = 2001, 2002, \dots, 4000$  of 2000 points. We did not use the first 2000 points because we thought that the data at the beginning of the measurement may be subjected to more measurement errors. The Priestley test, applied to the above data, shows that the data come from a UMP. Following the procedures given in [15] we first estimate the values of the evolutionary spectrum at a few selected points and the Priestley test converts them into an ANOVA problem. Fig.3 is the ANOVA table the test results.

The column sum of squares (SS) is the variation over different frequencies, the row sum of squares (SS) is the variation over different time values and the error sum of squares (SS) is the sum of errors and interactions between the time and the frequency. From the table we can see that the interaction is very small and hence the data should come from a uniformly modulated process. Therefore, the process  $X_t$   $t = 2001, 2002, \dots, 3999, 4000$  can be written as

$$X_t = d(t) X_{0,t} \tag{2.1}$$

Where  $d(t)$  is a deterministic function of  $t$  and  $X_{0,t}$  is a stationary time series. To obtain  $X_{0,t}$ , we estimate the local variance of each point  $t$  to approximate  $d(t)$ . To compute the variance at  $X_t$  at each  $t$ . For each  $t$ , we use 100 points

$$x(t-50), x(t-49), \dots, x(t-1), x(t), x(t+1), \dots, x(t+49), x(t+50)$$

centered at  $t$ , then our estimate  $\hat{d}(t)^2$  is the sample variance of  $X$ . We can then obtain  $X_{0,t}$  by dividing  $X_t$  by  $d(t)$ . The remaining work is to build a model for the stationary process  $X_{0,t}$ .

Since it is linear, Gaussian and stationary,  $X_{0,t}$  can be fitted by an ARMA (p, q) model. We choose ARMA (9, 10) because it has the smallest AICC number ([1]). The estimation of the ARMA model is given in Table 3.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	174.307	6	29.0512	1397.34	0
Rows	3.349	9	0.3722	17.9	0
Error	1.123	54	0.0208		
Total	178.779	69			

Figure 3: UMP test for Belgian Block data

AR Coeff.	Estimate	Standard Error	MA Coeff.	Estimate	Standard Error
$\alpha_1$	1.1480	0.3488	$\beta_1$	-0.0452	0.3478
$\alpha_2$	-1.3129	0.5139	$\beta_2$	0.9457	0.1459
$\alpha_3$	1.8893	0.6158	$\beta_3$	-0.6732	0.3639
$\alpha_4$	-1.5127	0.8002	$\beta_4$	0.4608	0.3039
$\alpha_5$	1.4296	0.7123	$\beta_5$	-0.7623	0.2415
$\alpha_6$	-0.9084	0.6281	$\beta_6$	-0.0251	0.2727
$\alpha_7$	0.7265	0.4025	$\beta_7$	-0.5716	0.0648
$\alpha_8$	-0.4357	0.3100	$\beta_8$	0.0756	0.1923
$\alpha_9$	-0.1387	0.2073	$\beta_9$	0.1216	0.0365
			$\beta_{10}$	0.1314	0.0546

Table 3: The Estimation of ARMA(9, 10) Model for Belgian Block Profile in UMP

The ACF values of the residuals (Fig.4.c) show that the residuals are uncorrelated. This indicates that the residuals are *i. i. d.* noise. So the ARMA (9, 10) model fits  $X_{0,t}$  very well. Then we can simulate  $X_{0,t}$  by using the ARMA (9, 10) model. We denote the simulation of  $X_{0,t}$  by  $\hat{X}_{0,t}$  and  $\hat{X}_{0,t}$  are given in Fig.4. The simulation of  $X_t$ , denoted by  $\hat{X}_t$ , is the product of  $\hat{X}_{0,t}$  and the local sample variance  $\hat{d}(t)$ . Fig.5 shows part of the Belgian Block data  $X_t$  ( $t = 2001, 2002, \dots, 4000$ ) and its simulation  $\hat{X}_t$ .

### 3. Modeling of Perryman3 Profile

The Perryman3 off-track terrain is unpaved. It is a dirt road and so it is very rough. Its statistical properties are not as good as those of Belgian Block profile. In [4], the statistical tests have shown that the Perryman3 profile data is non-linear, non-Gaussian and non-stationary. To analyze a non-linear time series, ad hoc methods must be used. For the non-linear Perryman3 profile data, we shall apply three methods to model and simulate it. In section 3.1,

we describe briefly the ARMA-GARCH residuals method and give the modeling result. In section 3.2 the Threshold Autoregressive method is introduced and applied, and in section 3.3 we combine the Empirical Mode Decomposition method and the Threshold Autoregressive method to build a model for the Perryman3 profile. All the three methods can represent the data profile reasonably well.

### 3.1 ARMA-GARCH Method and Modeling for Perryman3 Profile

#### 3.1.1 ARMA-GARCH Model

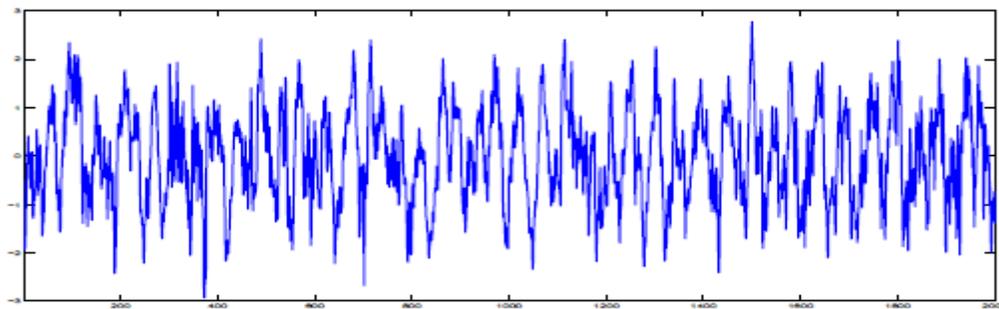
The Autoregressive Conditional Heteroscedasticity (ARCH) model was first proposed by Engle in [8], and later Bollerslev proposed the Generalized ARCH model (GARCH) in [2]. A GARCH process ([18]) is defined by the following formulas,

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim i.i.d N(0,1)$$

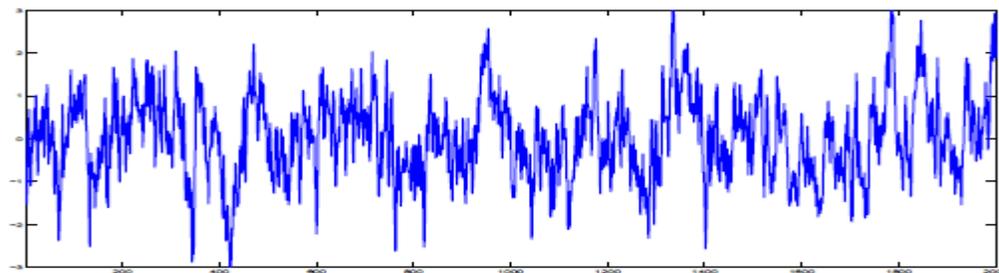
$$\sigma_t^2 = \varphi_0 + \sum_{i=1}^{p'} \varphi_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q'} \beta_j \sigma_{t-j}^2$$

We can see that a GARCH process is a process depending non-linearly on its noise. It has the property that large changes follow large changes and small changes follow small changes. This phenomenon is called conditional heteroscedasticity, which is also known as the GARCH effect.

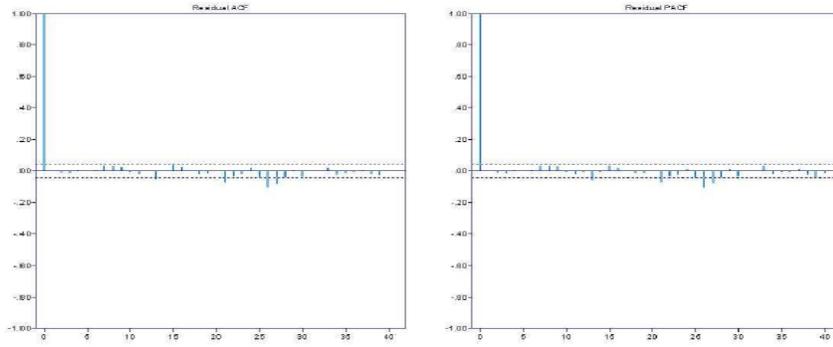
Before applying a GARCH model, we have to make sure whether there is a GARCH effect in the time series. McLeod and Li in [14] proposed a formal test for the GARCH effect based on the Ljung-Box test.



(a)  $X_{0,t}$

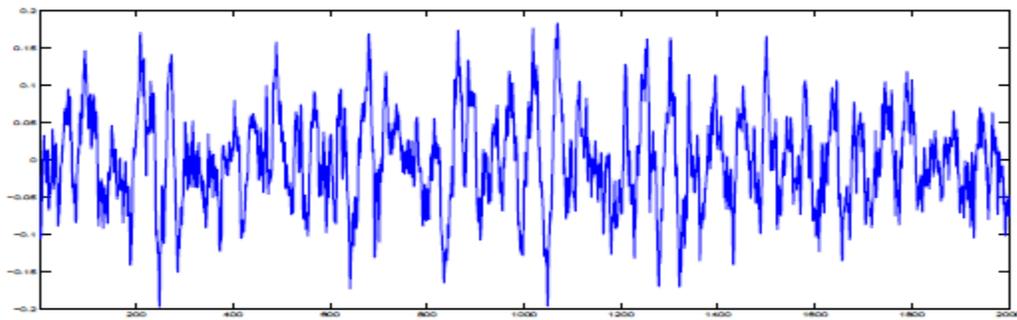


(b)  $\hat{X}_{0,t}$

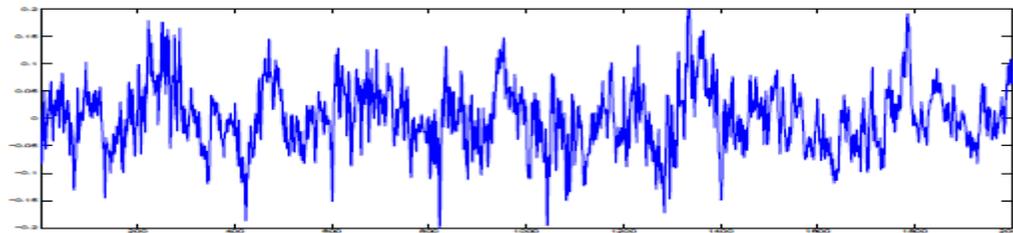


(c) ACF and PACF of Residuals of ARMA(9, 10) model for  $X_{0,t}$

Figure 4:  $X_{0,t}$  and its simulation  $\hat{X}_{0,t}$  by ARMA(9, 10) in UMP model for Belgian Block Profile



(a) Original Belgian Block Profile



(b) Simulated Belgian Block Profile

Figure 5: Simulation of UMP model for Belgian Block Profile

They considered the autocorrelation function of the squares of the time series, and tested whether the first  $L$  autocorrelations for the squared residuals are collectively small in magnitude. For a fixed large enough  $L$ , the Ljung-Box Q-statistic of McLeod-Li test is given by

$$Q = T(T + 2) \sum_{t=1}^L \frac{\hat{r}_t^2(\varepsilon^2)}{T - t}$$

Where  $T$  is the sample size, and  $\hat{r}_t^2$  is the squared sample autocorrelation of squared residual series at lag  $t$ . Under the null hypothesis of a linear generating mechanism of the data, i.e. no ARCH effect in the data, the test statistic is asymptotically  $\chi^2(L)$  distributed.

The ARMA-GARCH model may be interpreted as a combination of an ARMA model which is used to remove the linear dependence, and a GARCH model which is used to model the GARCH effect in the residual series from the ARMA model. The general ARMA ( $p, q$ ) – GARCH ( $p', q'$ ) model has the following form

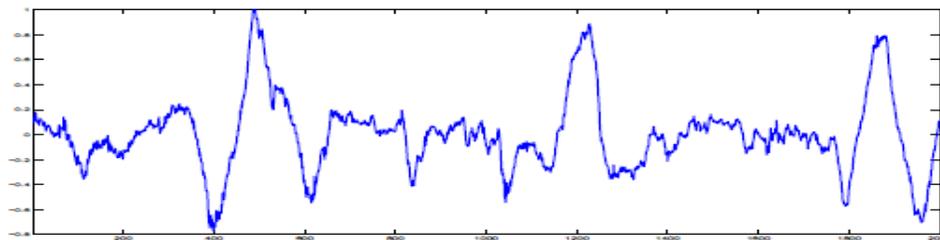
$$x_t = a_0 + \sum_{i=1}^p \alpha_i x_{t-i} + \sum_{j=1}^q \beta_j \varepsilon_{t-j}$$

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim i.i.d N(0,1)$$

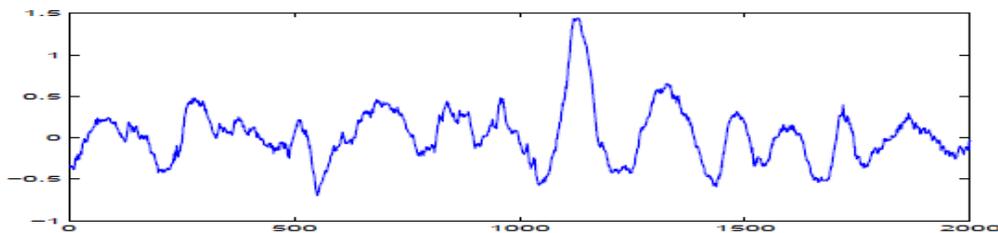
$$\sigma_t^2 = \varphi_0 + \sum_{i=1}^{p'} \varphi_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q'} \eta_j \sigma_{t-j}^2$$

### 3.1.2 ARMA-GARCH Modeling for Perryman3 Profile

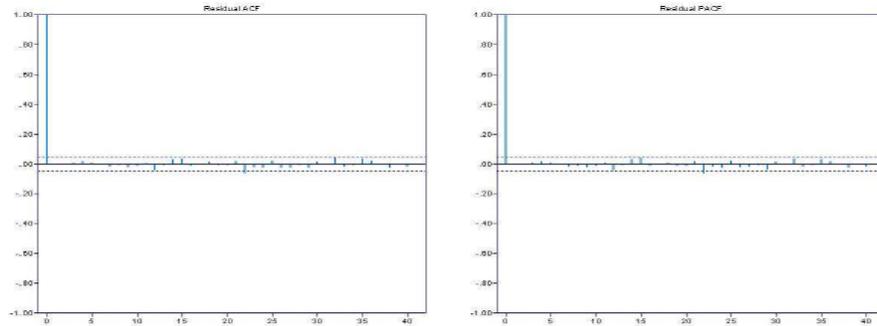
The time series plot of the nonlinear Perryman3 profile  $X_t$ , is given in Fig.6.a. From the graph of the Perryman3 profile, we can see that there are some big bumps and big hollows. And a big bump is usually followed by a big hollow. So it appears that there is a GARCH effect in the Perryman3 profile. First, we choose an ARMA (4, 3) model with the smallest AIC to remove the linear component of the file and denote the fitted value by  $u_t$ . Since the original Perryman3 profile data are nonlinear, the residuals of the linear ARMA (4, 3) model must also be nonlinear. Here we consider applying the GARCH model for the residuals. We first use the McLeod-Li GARCH effect test for  $e_t$ , the residuals of the ARMA (4, 3) model, with the lag  $L = 1, 2, \dots, 50$ . We find that all of the  $\chi^2$  test statistics values are much greater than the corresponding critical values in Table 4. Therefore, we can reject the null hypothesis. That is there is a GARCH effect in the residuals  $e_t$ . After some searching we decide to use GARCH (1, 1) model for  $e_t$ . The estimations of ARMA (4, 3) – GARCH (1, 1) model are given in Table 5, where  $\alpha_i$ ,  $i = 1, 2, 3, 4$  and  $\beta_j$ ,  $j = 1, 2, 3$  are the coefficients of the ARMA (4, 3) model,  $\varphi_i$ ,  $i = 0, 1$  and  $\eta_1$  are the coefficients of GARCH (1, 1). The ACF and PACF of the residuals of GARCH (1, 1) model (Fig.6.c) show that the residuals are i. i. d. It confirms our belief that the ARMA (4, 3) – GARCH (1, 1) model can be used to fit the nonlinear and nonstationary Perryman3 profile. The simulation results are given in Fig.6.



(a) Original Perryman3 Profile



(b) Simulation of ARMA(4,3) – GARCH(1,1) model for Perryman3 Profile



(c) ACF and PACF of Residuals of  $ARMA(4, 3) - GARCH(1, 1)$  model for Perryman3 Profile

Figure 6: the Simulation of  $ARMA(4, 3) - GARCH(1, 1)$  model for Perryman3 Profile

Lag	Mcleod-Li Test Statistic Value $\chi^2$	Critical value
1	146.2881	3.8415
2	159.3292	5.9915
3	170.9507	7.8147
10	232.1785	18.3070
20	272.6118	31.4104
30	338.2048	43.7730
50	353.3548	67.5048

Table 4: Mcleod-Li Test for GARCH Effect

Coeff.	Estimate	Standard Error	Coeff.	Estimate	Standard Error
$\alpha_1$	.989977	.235034	$\beta_1$	-.084673	.237596
$\alpha_2$	.602967	.437326	$\beta_2$	-.488125	.241840
$\alpha_3$	-.250928	.263279	$\beta_3$	-.199568	.120232
$\alpha_4$	-.345773	.113981			
$\phi_0$	.000103	.000016	$\eta_1$	.688334	.034203
$\phi_1$	.176680	.022246			

Table 5: The Estimation of  $ARMA(4, 3) - GARCH(1, 1)$  Model for Perryman3 Profile

### 3.2 Threshold Autoregressive Method and Modeling of Perryman3 Profile

#### 3.2.1 Threshold Autoregressive (TAR) Method

The Threshold Autoregressive (TAR) model was first used by H. Tong [18]. This method models nonlinear processes based on a “piecewise” linear approximation via partitioning a state-space into several subspaces. And the stationarity and Gaussianity may be preserved in each subspace. A threshold autoregressive (TAR) model with  $k$  ( $k \geq 2$ ) regimes is defined as

$$x_t = \sum_{i=1}^p \{b_{i0} + b_{i1}x_{t-1} + b_{i2}x_{t-2} + \dots + b_{i,p_i}x_{t-p_i} + \sigma_i \varepsilon_t\} I(x_{t-d} \in A_i)$$

where  $\varepsilon_t \sim i. i. d.(0, 1)$ , and  $d, p_1, \dots, p_k$  are some unknown positive integers,  $\sigma_i > 0$ ,  $b_{i,j}$  are unknown parameters, and  $\{A_i\}$  forms a partition of  $(-\infty, \infty)$  in the sense that  $\cup_{i=1, \dots, k} A_i = (-\infty, \infty)$  and  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ . Here  $I$  is the indicator function such that  $I(F) = 1$  if the statement  $F$  is true and  $I(F) = 0$  if  $F$  is false. The sets  $A_i$  represent the “thresholds”.

Based on the observations, we estimate the parameters  $b_{i,j}$ ,  $\sigma_i$ ,  $d$  and determine the order's  $p_i$  and the partition  $\{A_i\}$ . Since there are many unknown parameters to be determined in a TAR model, we first assume that the partition  $\{A_i\}$  and the orders  $p_i$  are known. As usual, the delay parameter  $d$  is not too big. Here we choose  $d$  from the set  $\{0, 1, 2, 3, 4, 5\}$ . For each fixed  $d$ , we minimize the least squares estimator to obtain  $\hat{b}_{i,j}$ .

$$L(b_i, d; A_i) = \sum_{x_{t-d} \in A_i} \{x_t - (b_{i0} + b_{i1}x_{t-1} + b_{i2}x_{t-2} + \dots + b_{i,p_i}x_{t-p_i})\}^2$$

And then we choose the best  $\hat{d}$  such that the sum of  $L(b_i, d; A_i)$  is minimized. Finally, we determine the best autoregressive order  $p_i$  by minimizing the generalized

$$AIC\{p_i\} = \sum_{i=1}^k [T_i \hat{\sigma}^2(p_i)] + 2(p_i + 1)$$

where  $\hat{\sigma}^2(p_i) = \frac{1}{T_i} L(\hat{b}_i, \hat{d}; A_i)$  and  $T_i$  is the number of elements in the set  $\{t: p_i < t \leq T \text{ and } x_{t-\hat{d}} \in A_i\}$ .

### 3.2.2 TAR Modeling for Perryman3 Profile

The time plot of the Perryman3 profile (Fig.6.a) consists of a rather flat profile with a few bumps coming up intermittently. It is mostly those bumps which make the whole profile non-linear. In Fig.6.a, it seems that the bumps come up periodically, but looking carefully one can see that there is no fixed period between the bumps. So we cannot use any seasonal models for the Perryman3 profile data. To describe the bumps more precisely, we apply a TAR model for the Perryman3 profile. We model the flat part and bumps separately by autoregressive models and used a threshold to switch the process back and forth between the flat part and the bump part. So we partition the state-space into two regimes: bump part and flat part.

To choose the best threshold for the bump, we use a set of thresholds as  $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ . For each given threshold, we set  $p_1, p_2 \leq 10$  and  $0 \leq d \leq 4$ . The AIC selected the following TAR model as the best TAR model for Perryman3 profile. That is the best threshold is 0.4 and the delay parameter  $d = 1$  and the autoregressive orders  $p_1 = 7, p_2 = 6$ . The estimated TAR model is

$$x_t = \begin{cases} -0.0084 + 0.8972x_{t-1} + 0.1954x_{t-2} + 0.0201x_{t-3} - 0.1101x_{t-4} \\ \quad + 0.1310x_{t-5} - 0.1604x_{t-6} + 0.0312\epsilon_t x_{t-1} \leq 0.4 \\ 3.9324x_{t-1} - 6.4820x_{t-2} + 5.9222x_{t-3} - 3.5120x_{t-4} + 1.6406x_{t-5} \\ \quad - 0.6322x_{t-6} + 0.1308x_{t-7} + 0.0669\epsilon_t x_{t-1} > 0.4 \end{cases}$$

In this TAR model, if the process is in the flat part and we detect  $x_{t-1} > 0.4$ , then the process will be switched to the bump part from time  $t$ . Similarly if the process is in the bump part and we detect  $x_{t-1} \leq 0.4$ , then the process will be switched to the flat part from time  $t$ . A simulation of the Perryman3 profile by TAR model is given in Fig.7. The ACF of the residuals in Fig.7.c shows that this TAR model fits the data. The simulation in Fig.7 tells us that this method models the bumps part reasonably well.

### 3.3 Empirical Mode Decomposition Method and Modeling for Perrman3 Profile

#### 3.3.1 Empirical Mode Decomposition Method

The Empirical Mode Decomposition (EMD) method was developed by Huang ([11]) for analyzing non-stationary and non-linear data. The EMD method decomposes the data into a collection of Intrinsic Mode Functions (IMF) and a residue. The essence of the method is to identify the intrinsic oscillatory modes by their characteristic time scales in the data empirically, and then decompose the data accordingly. This decomposition method will use the envelopes defined by the local maxima and local minima of the data profile  $X_t$  separately. Once the extrema are identified, all the local maxima are connected by a cubic spline line as the upper envelope. Repeating the procedure for the minima produces the lower envelope. The upper and lower envelopes should cover almost all the data between them. The mean of the upper and the lower envelopes is designated as  $m_{10}$ , and the difference between the data  $X_t$  and the mean  $m_{10}$  is denoted by  $h_{10}$ , i.e.

$$X_t - m_{10} = h_{10} \tag{3.1}$$

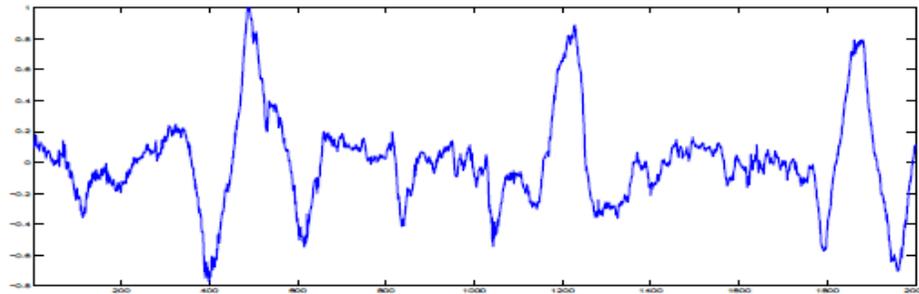
This is called a sifting process. The Fig.8 demonstrates the sifting process visually. Then taking  $h_{10}$  as new data and repeating the procedure, we get  $h_{11}$ . We repeat this sifting process  $k$  times until  $h_{1k}$  is an IMF. An intrinsic mode function (IMF) is defined as a function each of whose oscillations crosses the zero axis and the mean of whose upper and lower envelopes is zero.

$$\begin{aligned}
 h_{10} - m_{11} &= h_{11} \\
 h_{11} - m_{12} &= h_{12} \\
 &\dots \dots \dots \\
 h_{1(k-1)} - m_{1k} &= h_{1k}
 \end{aligned}
 \tag{3.2}$$

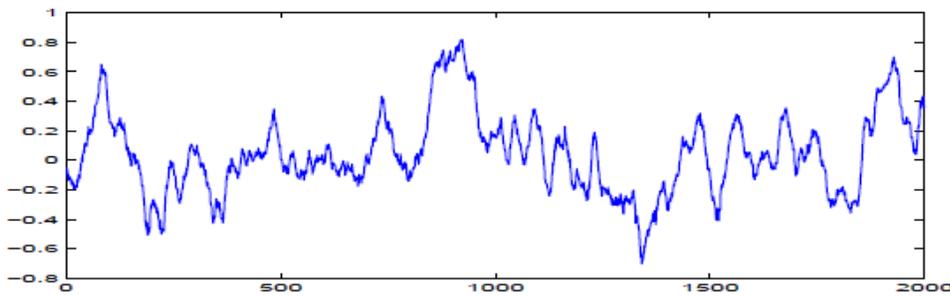
Let  $C_1 = h_{1k}$  be the first IMF, then  $C_1$  should contain the finest scale or the shortest period component of the signal. We can separate  $C_1$  from the data by

$$X_t - C_1 = r_1
 \tag{3.3}$$

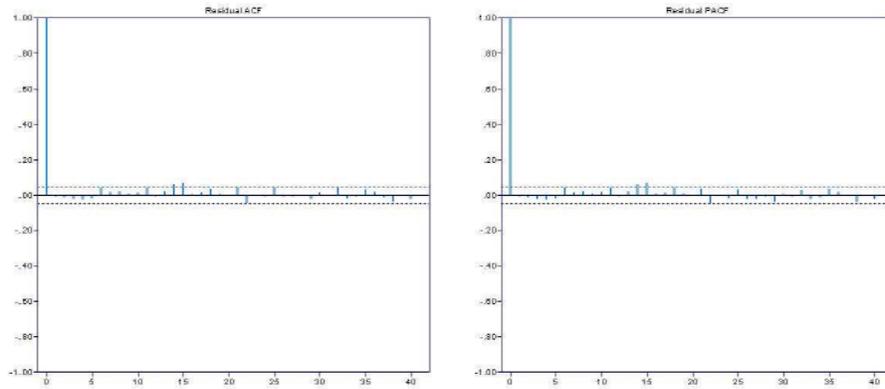
Where  $r_1 = m_{10} + m_{11} + m_{12} + \dots + m_{1k}$ . Since the residuals  $r_1$  still contains information of longer period components, it is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated on all the subsequent  $r_j$ , and the result is



(a) Original Perryman3 Profile

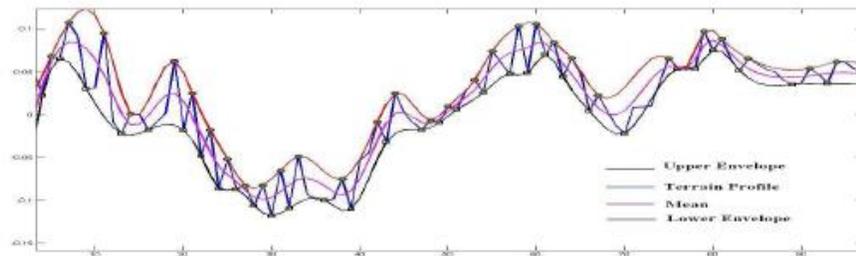


(b) Simulation of TAR model for Perryman3 Profile

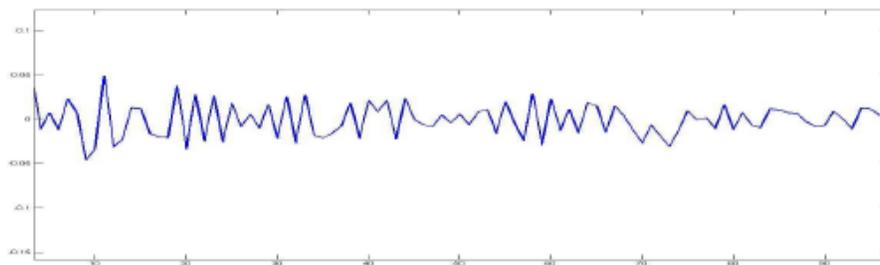


(c) ACF and PACF of Residuals of TAR model for Perryman3 Profile

Figure 7: the Simulation of TAR model for Perryman3 Profile



(a) Decomposition



(b)  $h_{10}$

Figure 8: Sifting Process

$$\begin{aligned}
 X_t - C_1 &= r_1 \\
 r_1 - C_2 &= r_2 \\
 &\dots \\
 r_{n-1} - C_n &= r_n
 \end{aligned}
 \tag{3.4}$$

This procedure can be stopped any time as the problem demands or by any of the following predetermined criteria; either when the component  $C_n$  or the residual  $r_n$  becomes so small that it is less than a predetermined value of substantial consequence, or when the residual  $r_n$  becomes a monotonic function from which no more IMF can be extracted. By summing up equations (3.4), we finally obtain

$$X_t = \sum_{i=1}^n C_i + r_n
 \tag{3.5}$$

Where  $C_i(t)$  is the  $i$ th IMF and  $r_n(t)$  is residual after  $n$  decompositions.

As described above, the process is indeed like sifting: to separate the finest local mode from the data first based only on the characteristic time scale. The sifting process, however, has two effects: (a) to eliminate riding waves; and (b) to smooth uneven amplitudes. Therefore every oscillation of each IMF crosses the zero axis and the

	Linearity Test Statistic	Critical Value	Conclusion
$C_1(t)$	240.2541	166.4153	Non-Linear
$C_2(t)$	24.5864	166.4153	Non-Linear
$r_2(t)$	429.4308	166.4153	Non-Linear

Table 6: Linearity Test for IMF's and the Residuals

	Stationarity Test Statistic	Critical Value	Conclusion
$C_1(t)$	875.5837	16.9190	Non-Stationary
$C_2(t)$	1.1339e+003	16.9190	Non-Stationary
$r_2(t)$	1.9556e+003	16.9190	Non-Stationary

Table 7: Stationarity Test for IMF's and the Residuals

oscillations are all symmetric with respect to the zero axis. The first IMF  $C_1(t)$  captures oscillations of  $\{X_t\}$  with the highest frequencies. The second IMF  $C_2(t)$  captures oscillations with the second highest frequencies etc. The frequency ranges of the IMF's  $C_1(t), \dots, C_n(t)$  are almost disjoint, hence they are practically uncorrelated. For a nonlinear and non-Gaussian data, the linearity and Gaussianity properties of the  $C_i(t)$ 's are, in general, getting worse as  $i$  increases. Hence, we have to stop at some  $n$  for which the  $C_i(t), i = 1, \dots, n$  can still pass the linearity test. The residual  $r_n(t)$  is non-linear and we have to use one of the methods described above to model it or just treat it as a deterministic function since it is much smoother than the original data.

### 3.3.2 EMD Modeling of the Perryman3 Profile

We apply the IMF sifting process to the Perryman3 data and we stop at  $n = 2$ . Then we have

$$X_t = C_1(t) + C_2(t) + r_2(t).$$

Before we build a model for the three IMF components  $C_i(t), i = 1, 2$  and the residuals  $r_2(t)$ , we first have to test their linearity, stationarity, and Gaussianity. The test results are given in Table 6, Table 7, and Table 8. We could find that the first IMF component  $C_1(t)$  is non-linear, non-stationary, and non-Gaussian; the second IMF component  $C_2(t)$  is non-stationary, but linear and Gaussian; the residuals  $r_2(t)$  is non-linear, non-stationary, and non-Gaussian as well as  $C_1(t)$ .

Table 9 shows that the correlation coefficients between the two IMF's  $C_i(t), i = 1, 2$  and the residuals  $r_2(t)$  are all quite small. So they are uncorrelated. Hence we may add the models for the three components together as a model for the Perryman3 profile.

We take ARMA-GARCH model for the first IMF component  $C_1(t)$  which is non-linear, non-stationary, and non-Gaussian. The coefficients and standard errors of the model are listed in the Table 10.

	Gaussianity Test Statistic	Critical Value	Conclusion
$C_1(t)$	493.2715	266.3781	Non-Gaussian
$C_2(t)$	70.6241	266.3781	Gaussian
$r_2(t)$	2.8351e+003	266.3781	Non-Gaussian

Table 8: Gaussianity Test for IMF's and the Residuals

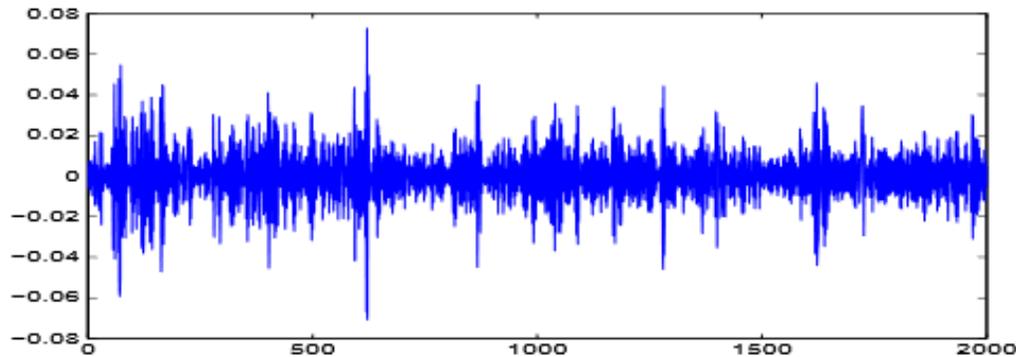
Correlation Coefficient	$C_1(t)$	$C_2(t)$	$r_2(t)$
$C_1(t)$	1.0000	-0.0202	0.0110
$C_2(t)$	-0.0202	1.0000	-0.0135
$r_2(t)$	0.0110	-0.0135	1.0000

Table 9: Correlation of IMF's and the Residuals

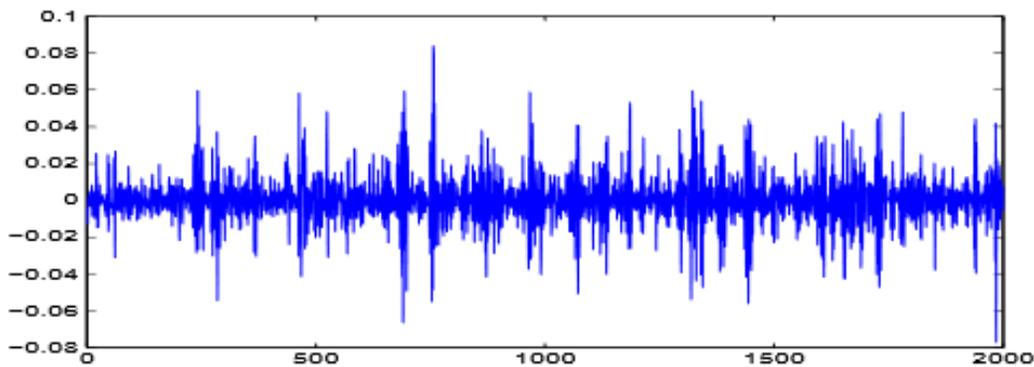
Fig.9 demonstrates the simulation result of ARMA (2, 4)-GARCH (1, 1) model. The ACF and PACF of the residuals (Fig.9) are all bounded in the 95% confidence interval. It means that the ARMA (2, 4)-GARCH (1, 1) model fits the first IMF component  $C_1(t)$ . Since the second IMF component  $C_2(t)$  is non-stationary, linear and Gaussian which has the similar property with the Belgian Block profile, it is easy for us to take the UMP test for it. The ANOVA table in Fig.10 indicates that the interaction between frequency and time is very small, therefore, we take  $C_2(t)$  as a uniformly modulated process (UMP). i.e.  $C_2(t)=D(t) x_{0,t}$ , where  $D(t)$  is a deterministic function of  $t$  and  $x_{0,t}$  is a stationary time series. As what we did in Section 2.2 for Belgian Block profile, we apply ARMA(4, 5) model for  $x_{0,t}$ . Table 11 lists the coefficients and the standard errors of the ARMA(4, 5) model for  $x_{0,t}$ . The Fig.11 is the simulation result for  $C_2(t)$  from a UMP model.

Coeff.	Estimate	Standard Error	Coeff.	Estimate	Standard Error
$\alpha_1$	1.484220	.006830	$\beta_1$	-1.802332	.022954
$\alpha_2$	-.969044	.006690	$\beta_2$	1.266481	.046900
			$\beta_3$	-.036268	.046674
			$\beta_4$	-.16405	.022642
$\phi_0$	.000017	000003	$\eta_1$	.403754	.040812
$\phi_1$	.539377	.056516			

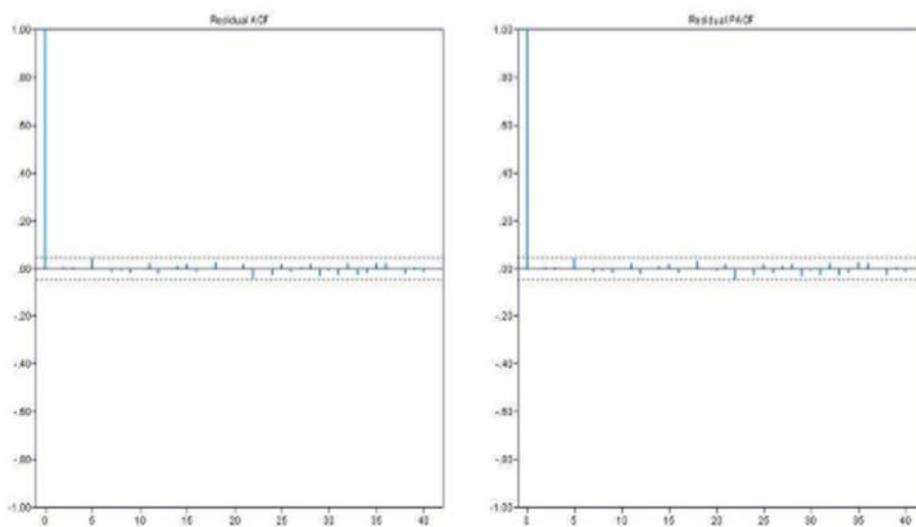
Table 10: The Estimation of ARMA(2, 4) – GARCH(1, 1) Model for IMF  $C_1(t)$



(a)  $C_1(t)$



(b)  $\hat{C}_1(t)$



(c) ACF and PACF of Residuals of  $ARMA(2, 4) - GARCH(1, 1)$  model for  $C_1(t)$

Figure 9: the Simulation of  $ARMA(2, 4) - GARCH(1, 1)$  model for  $C_1(t)$

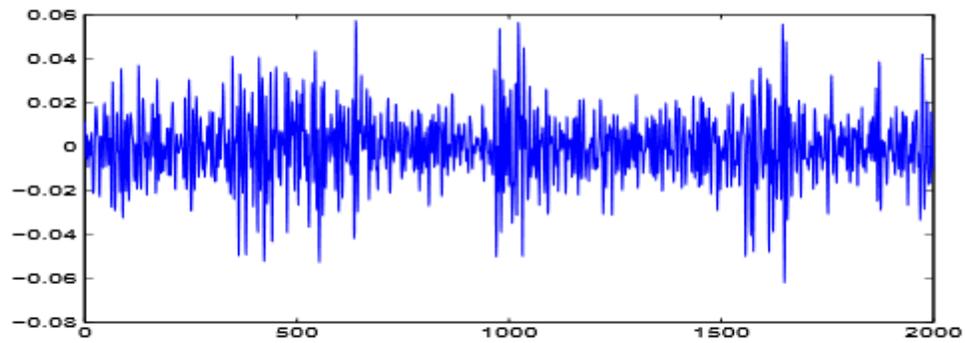
ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	195.632	6	32.6053	284.1	0
Rows	18.204	9	2.0227	17.62	0
Error	6.197	54	0.1148		
Total	220.033	69			

Figure 10: UMP test for IMF  $C_2(t)$

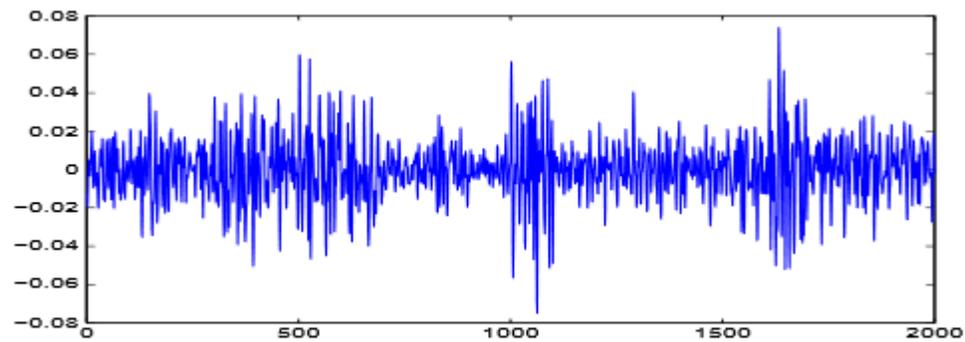


AR Coeff.	Estimate	Standard Error	MA Coeff.	Estimate	Standard Error
$\alpha_1$	1.83278	0.19121	$\beta_1$	0.26935	0.18995
$\alpha_2$	-1.53703	0.29341	$\beta_2$	-0.49610	0.13050
$\alpha_3$	0.68818	0.20464	$\beta_3$	-0.39838	0.08485
$\alpha_4$	-0.13384	0.06935	$\beta_4$	0.09448	0.11216
			$\beta_5$	0.04201	0.06326

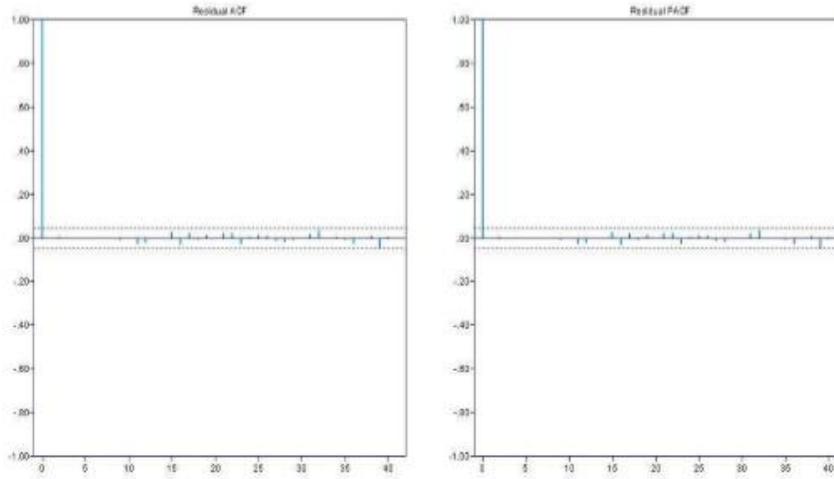
Table 11: The Estimation of  $ARMA(4, 5)$  Model for IMF  $C_2(t)$  in UMP



(a)  $C_2(t)$



(b)  $\hat{C}_2(t)$



(c) ACF of Residuals of the UMP model for  $C_2(t)$

Figure 11: the Simulation of a UMP model for  $C_2(t)$

After building proper models for the two IMF components, we next need to create a model for the residuals  $r_2(t)$ . Although the residuals  $r_2(t)$  is non-linear, nonstationary, and non-Gaussian as well as  $C_1(t)$ , the bumps are still there. But the residuals  $r_2(t)$  are much smoother than the original Perryman3 profile because all the high frequencies oscillations are removed to the IMF components. We can consider a TAR model for  $r_2(t)$ . We have already found that the best threshold is 0.4 for the original Perryman profile in Section 4.2. Since the residuals have the same trend and bumps as the original Perryman profile, we also set the threshold as 0.4 for the residuals  $r_2(t)$ . The AIC shows that the best delay parameter  $d = 1$ . And the autoregressive order  $p_1 = 5$  in the low regime,  $p_2 = 7$  in the high regime. The details of the TAR model are given below.

$$r_2(t) = \begin{cases} -0.0002 + 3.5066r_{t-1} - 4.9570r_{t-2} + 3.4945r_{t-3} - 1.1599r_{t-4} \\ \quad + 0.1144r_{t-5} + 0.0312\epsilon_t r_{t-1} \leq 0.4 \\ 1.0024r_{t-1} - 2.2230r_{t-2} + 4.9022r_{t-3} - 3.1320r_{t-4} + 1.3416r_{t-5} \\ \quad - 0.7322r_{t-6} + 0.2318r + 0.1629\epsilon_t r_{t-1} > 0.4 \end{cases}$$

Fig.12 shows the residuals and its simulation by the TAR model. The ACF of the residuals shows that the residuals of the TAR model are independent. It means that the TAR model fits  $r_2(t)$  very well. Finally, Fig.13 gives a simulation result for Perryman profile by the EMD method based on the simulations for the two IMF components  $C_1(t)$ ,  $C_2(t)$  and the residuals  $r_2(t)$ .

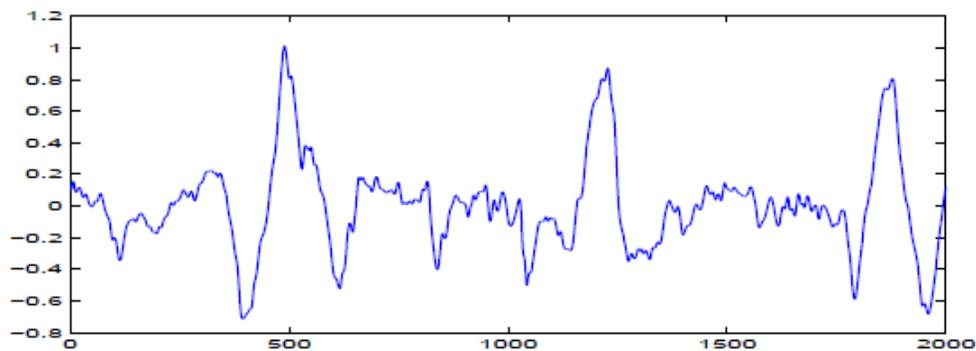
#### 4 A Criterion of Model Fitness from Damage Theory

In the last section we have applied 3 different methods to model the same Perryman3 terrain profile. All the 3 methods seem to model it well as judged by the autocorrelation functions of their residues. Therefore it compels us to find another criterion to differentiate the fitness of these models. In the following we are going to use the linear theory of damage and the idea of rain flow count to define a new criterion for this purpose.

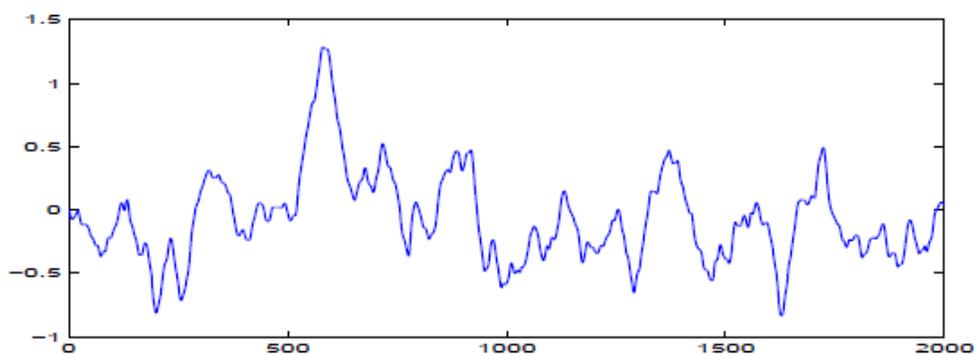
1) Linear Damage Theory for Material Failure due to Fatigue. Suppose we subject a piece of material specimen under repeated tension and compression stress cycles of the same amplitude  $S$  and the material starts to break down after  $N$  such tension and compression cycles. According to the experiments  $S$  and  $N$  have the following relation

$$N * (S^b) = C,$$

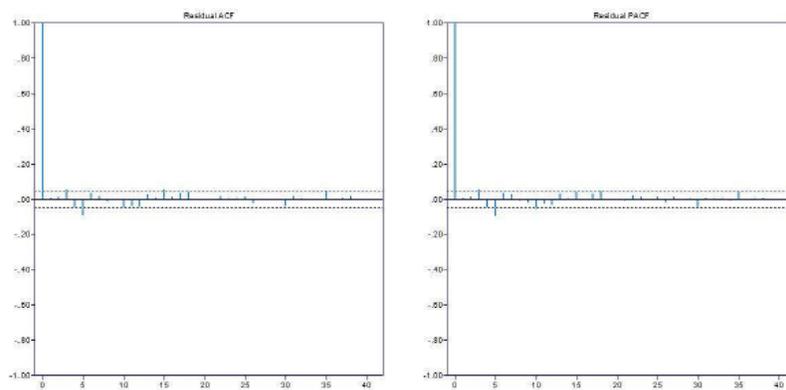
Where  $b$  and  $C$  are constants depending on the material. The relation is called an S-N curve (Fig.14). So for each  $S$  value there is a corresponding  $N$  value according to the S-N curve.



(a)  $r_2(t)$

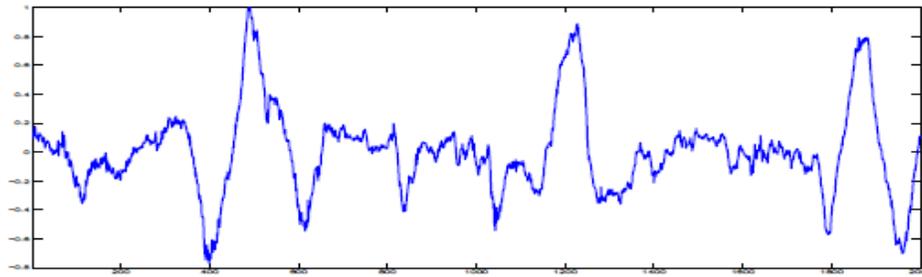


(b)  $\hat{r}_2(t)$

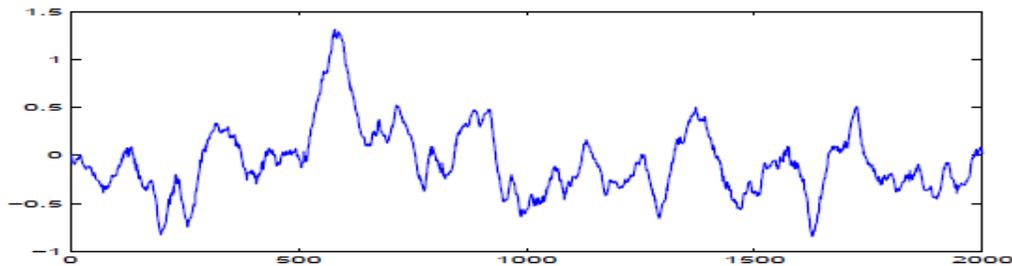


(c) ACF of Residuals of TAR model for  $r_2(t)$

Figure 12: Simulation of TAR model for  $r_2(t)$



(a) Original Perryman Profile



(b) Simulation of EMD model for Perryman Profile

Figure 13: the Simulation of EMD model for Perryman Profile

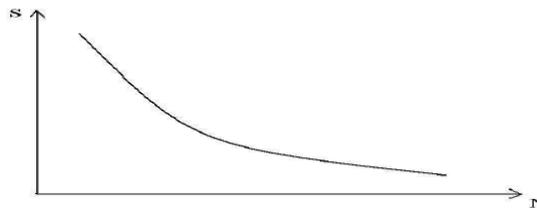


Figure 14: S-N Curve

The linear damage rule (or Palmgren-Miner Rule, [6] or [5]) can be stated as the following: In an experiment with  $n_i$  cycles of stress amplitude  $S_i$ ,  $i = 1, 2, \dots, k$ , the total cumulative damage fraction done to the specimen is

$$D = \sum_{i=1}^k \frac{n_i}{N_i}$$

Where  $N_i$  is the number of cycles until fracture corresponding to  $S_i$  from its S-N curve. The specimen has fatigue failure when  $D = 1$ .

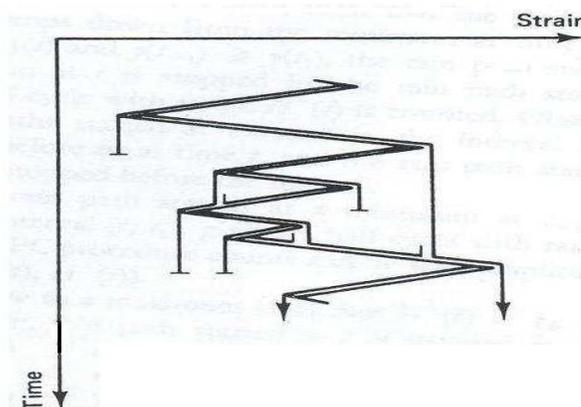
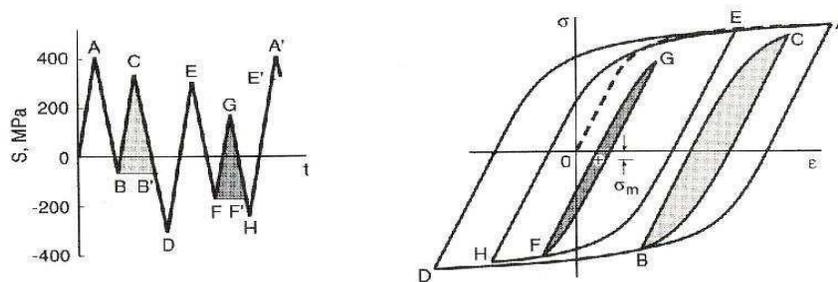
**2) Rain flow Counts of Cycles.**

The graph on the left part of the Fig.15 represents a portion of a stress history. The question is what should be the amplitudes corresponding to the stress oscillations in that graph. The right part of the Fig.15 shows the corresponding stress-strain relation in a hysteresis loop curve. If you follow the stress history OABCDEFGHA' in the left figure, the stress-strain history follows the hysteresis curve in the same order OABCDEFGHA' in the right figure. One can see clearly in the hysteresis curve that there are exactly four and one-half cycles corresponding to the stress history in the left figure. They are cycles ADA', BCB', FGF', EHE' and a half cycle OA. This method of counting the stress cycles is called the rain flow count of cycles ([7]).

The name rain flow count comes from an idea as shown in the Fig.16, because it looks like rain flow from the side of a pagoda. In [16] Rychlik gave a mathematical definition of rain flow counts which we have used to write computer programs to do the counting of rain flow cycles and to compute the amplitudes for each history of stress oscillations.

3) The Histogram of Rain flow Counts of Cycles of Different Amplitudes. Consider each terrain profile as a sequence of rain flow cycles of stress oscillations. According to the linear damage theory the cumulative damage caused by the terrain profile is characterized by its histogram of the rain flow cycles at different amplitudes in combination with a certain S-N curve. Thus, to determine which model fits better, we can just compare the rain flow cycle histogram of each model with that of the terrain profile.

The simulation will change every time because of the random noise. To calculate the histogram stably and unbiasedly, we run the simulation 1000 times for each model and consider the average histogram of each model. Since the amplitude of most rain flow cycles is less than 0.1, we mainly focus on the comparison of the number of rain flow cycles with amplitude less than 0.1. Fig.17 shows us that the histograms of the rain flow cycles of the simulation profiles and the original profile. We further figure out the number of rain flow cycles with amplitude less than 0.1 are 448, 453, 464 and 358 respectively for the original Perryman3 profile, the simulated profile by ARMA- GARCH model, the simulated profile by TAR model and the simulated profile by



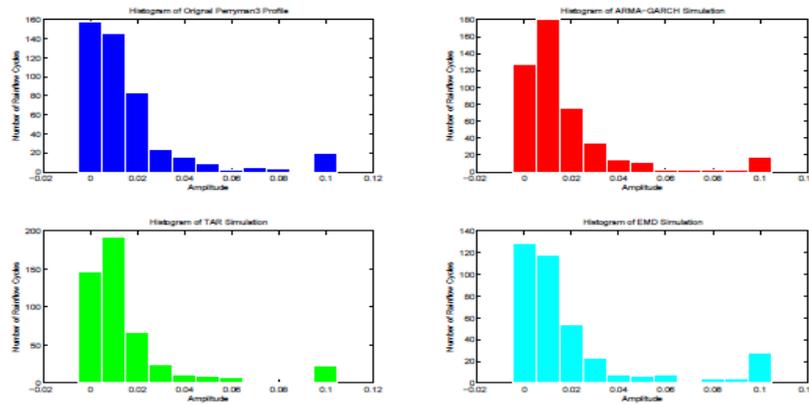


Figure 17: Histograms of Rainflow Cycles of the Simulations for Perryman3 Profile

EMD model. Though the numbers of rain flow cycles from the simulations of ARMA-GARCH and TAR models are closer to that of the original Perryman3 profile, the shape of the histogram of EMD simulation is more like that of the original Perryman3 profile. To compare the histograms better, we make the cubic spline interpolation for each histograms and get Fig.18. Then we find that the smoothed histogram curve of EMD model is almost the same as that of the original Perryman3 profile. Therefore using this damage criterion for model fit, EMD fits the original profile better than the TAR and ARMA-GARCH model.

**5 Conclusion**

After a new vehicle has been designed and manufactured, it and its various parts have to be tested on a testing track for durability due to fatigue from vibrations caused by driving along the track. However it is equally important to have computer models to simulate these test tracks so one can do an initial test of the vehicle on a vehicle tester or a shaker in a laboratory using the simulated tracks as input and, with this simulated track, one can also run computer based simulations of the vehicle during its designing stage. In this paper we have found a model (UMP) for the Belgium Block test track and 3 models (ARMA-GARCH, TAR, EMD) for the Perryman3 off track terrain. All the 3 models seem to fit well by looking at the autocorrelation functions of their error residues. Since the goal of the driving test is to study material failure due to fatigue caused by vibrations, we propose another criterion for model fitness based on the

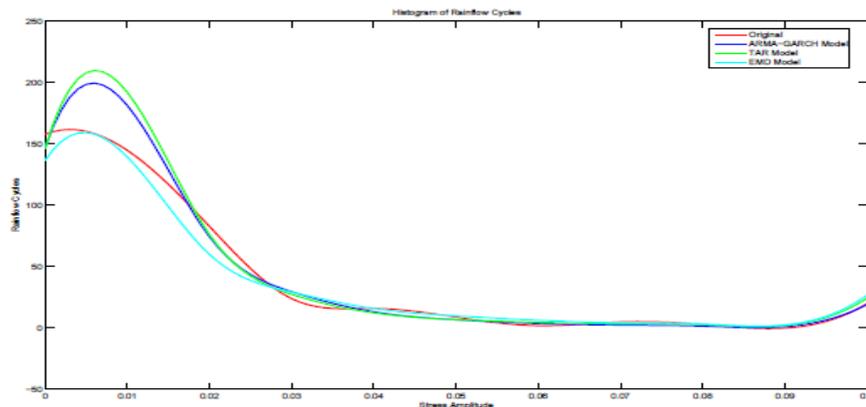


Figure 18: Spline Interpolated Histograms of Rainflow Cycles of the Simulations for Perryman3 Profile

Figure 18: Spline Interpolated Histograms of Rainflow Cycles of the Simulations for Perryman3 Profile linear damage theory. Using this criterion it seems that the EMD model fits the best.

## **Acknowledgement**

This work is partially supported by a research grant from the Automotive Research Center (ARC).

## **References**

- Bickel, P. J. and Doksum, K. A. *Mathematical Statistics*, Prentice Hall, 1997.
- Bollerslev, T., Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics*, 31, (1986), 307-327.
- Box, G. E. P., Jenkins, G. and Reinsel, G., *Time Series Analysis Forecasting and Control*, Prentice Hall, 1994.
- Chaika, M., Gorsich, D. and Sun, T. C., Some statistical tests in the study of terrain modelling, *Int. J. Vehicle Design*, 36, (2004), 132-148.
- Crandall, S. H. and Mark, W. D., *Random Vibration*, Academic Press, 1963.
- Dowling, N. E., *Mechanical Behavior of Materials*, Prentice Hall, 1999.
- Downing, S. D. and Socie, D. F., Simplified Rainflow counting Algorithm. *Int. J. Fatigue* 4 1, 31-40.
- Engle, R., Autoregressive conditional heteroscedasticity with estimates of the variance of Uk inflation, *Econometrica*, 50, (1982), 987-1008.
- Fan, J. and Yao, Q., *Nonlinear Time Series*, Springer Verlag, 2003.
- Granger, C. W. J. and Andersen, A. P., *An introduction to bilinear time series models*, Vandenhoeck and Ruprecht, 1978.
- Huang, N. E., Shen, Z., Long, S. R. and et al., The Empirical Mode Decomposition and the Hilbert Spectrum for nonlinear and non-stationary time series analysis, *Pro. R. Soc. Lond. A* 454, (1998), 903-995.
- Keenan, D. M., A Tukey nonadditivity-type test for time series nonlinearity, *Biometrika*, bf 72 (1985) 39-44.
- Ljung, G. M. and Box, G. E. P., On a measure of lack of fit in time series models, *Biometrika*, 65, (1978), 297-303.
- McLeod, A. I. and Li, W. K., Diagnostic checking ARMA time series models using squared residual autocorrelations, *Journal of Time series Analysis*, 4, (1983), 269-273.
- Priestley, M. B., *Non-linear and Non-stationary Time Series Analysis*, Academic Press, 1988.
- Rychlik, I., A new definition of the rainflow cycle counting method, *Int. Journal of Fatigue*, 4 (1987), 119-121.
- Sun, T. C., Gorsich, D., Chaika, M., Alyass, K., Wei, Jinfeng and Ferris, J., Time series modeling of terrain profiles, *SAE 2005 Transactions, Journal of Commercial Vehicles*, pp.221-227.
- Tong, H., *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford University Press, Oxford, 1990.
- Tsay, R. S., *Analysis of Financial Time Series*, John Wiley & Sons, 2005.
- Tsay, R. S., Non-linear time series analysis of blowfly population. *Journal of Time Series Analysis*, 9, (1988), 247-263.
- U. S. Army Test Operation Command Test Operating Procedure, TOP 1-1-010, 1987.