Journal of the Society of Korea Industrial and Systems Engineering
Vol. 33, No. 2, pp.112－126, June 2010.

ARMA-PL : 시계열 데이터에 나타나는 중첩된 주기 및 선형추세에 대한 고찰

서정열* · 이세재* · 오현승** · 구자활* · 임 택* · 조진형*†

*금오공과대학교 산업시스템공학과
**한남대학교 공과대학 산업경영공학과

ARMA-PL : Tackling Nested Periods and Linear Trend in Time Series Data

Jung-Yul Suh* · Sae-jae Lee* · Hyun-Seung Oh** · Ja-Hwal Koo* · Lim Taek* · Jin-Hyung Cho*†

*Department of Industrial and Systems Engineering, Kumoh National Institute of Technology
**Department of Industrial and Systems Management Engineering, Hannam University

시계열데이터는 ARMA 분석에 적합지 않은 요소를 내재하고 있는 경우가 있다. 특히 선형성과 주기성을 가진 요소가 확률적인 분포와 자주 혼재되어 있다. 이 논문에서는 이런 선형적 주기적 요소를 찾아내고 분석하는 방법을 제시한다. 특히 주기적 요소는 여러 주기가 층층이 겹쳐져서 나타난다. 주기 간에는 서로 일정 정수비율을 유지하며, 한 주기 안에 다른 주기가 내포되어 있는 경우(nested periods)가 많다. 시간규모(time-scale)개념을 도입하여 이러한 주기적 요소를 개념적으로 정립하고자 했다. 선형적 요소와 주기적 요소가 제거된 후 추출된 데이터는 MA-approximation 이라는 방법을 사용하여 가장 데이터에 근접한 ARMA 모델을 찾아낸다. 마지막으로 선형적 주기적 요소와 ARMA 추정결과를 종합하여 control boundary를 결정하는 방법을 제시한다.

Keywords[1] : Nested Periods, PL-ARMA, MA-approximation, Time-scale, PWF

1. Overview

In a manufacturing process, readings from sensors placed at various locations of facilities are monitored to determine if the process is working normally. Each reading is assigned a parameter and the stream of data collected over time coming from each reading constitute a time-series for the corresponding parameter. Of the main interest is whether there is abnormally high or low value in the time-series data. Such a data value is an indication that something is wrong with the process, which should be corrected. Many time-series data thus obtained have very high autocorrelation among neighboring data values. Frequently, there are significant long-distance autocorrelation. In such a case, we cannot assume that all data are *iid* (identically independently distributed), and using Shewhart Chart Method will not work. This problem has been studied previously by using ARIMA model [1, 3, 4] However, our investigations revealed following problems

논문접수일 : 2010년 03월 24일 논문수정일 : 2010년 04월 04일 게재확정일 : 2010년 04월 06일
† 교신저자 joy@kumoh.ac.kr
※ This research was supported by Kumoh National Institute of Technology.
1) Brief description of keywords is given in the Glossary of Keywords at the end of this paper.

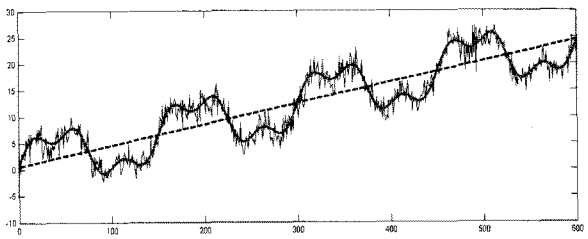
were not addresse.

Many real-life time-series data can contain many levels of dynamic operating on different time-scales. That is, time-series can be an aggregation of components, whose rate of change can range from "very fast" to "very slow." Slow changing ones tend to create significant level of long-distance autocorrelation. A rather simple example is time-series data showing multiple periods of varying length. We need more than ARIMA model to address this problem.

In this paper, we will show how to identify, estimate and extract these components with varying time-scales, and subtract them from the original time-series data. The resulting data can be then modeled with ARIMA (in our case, ARMA). We will show how to **identify the presence of different time-scales** in a time-series. In estimating time-scales, we use a crucial assumption. That is, time-scales in a series can be arranged in an hierarchical order, ranging from the shortest one to the longest one. In this hierarchy, each time- scale is a 'simple' fraction or multiple of another time-scale right above or below. We essentially have a 'nested time- scale' with specific ratio pattern. Many time-series in nature show such a relationship and this also turns out to be the case in the set of time series data we used. Wavelet analysis is also based on this observation. That is, frequency resolution in low frequency region is coarse (low), while compact (high) in high frequency region [11-13]. **Estimation of nested time-scales** makes up the major part of this paper.

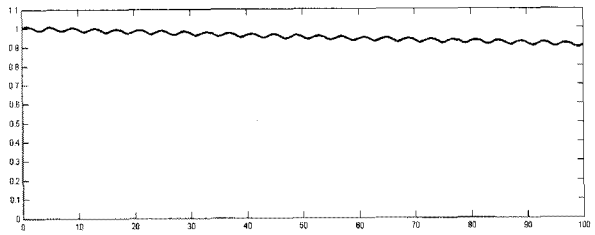
Data we used contain multiple components with different periods each obeying the ratio constraint mentioned above. It turns out that such periodic components can account for significant portion of data value, and with the contribution of those components subtracted from each data value, the resulting time-series can be modeled better with ARIMA, producing a much closer fit.

We now have a time-series model with combination of ARIMA model and nested time-scale model (or nested periodic model in a narrower application). Next, we need to **determine the control boundaries of the time-series**. That is, upper and lower boundary (upper control limit (UCL) and lower control limit (LCL)). While it is straightforward to estimate the boundary for pure ARIMA model, this new combined model requires more qualified approach. It depends on how we interpret the nature of nested time-scale components. That is, would they represent acceptable change or unwanted distortion which needs to be eliminated? Based on this interpretation, different control boundary needs to be estimated.



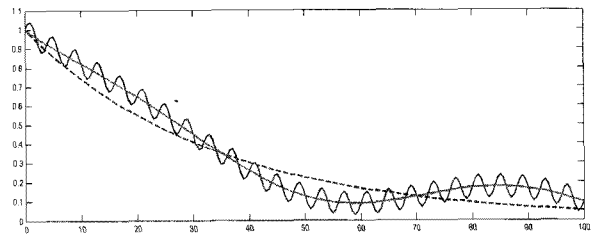
typical time series with
strong linear component (dotted line)
and periodic component (smooth curve)

<Figure 1.1>



autocorrelation with strong linear component

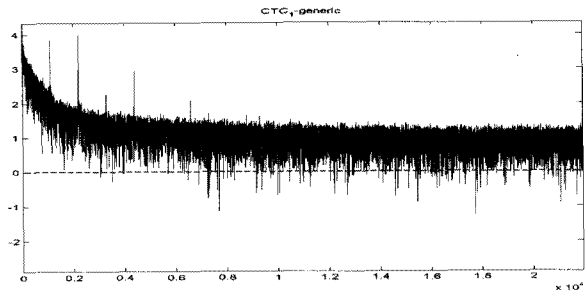
<Figure 1.2>



autocorrelation with linear component removed

solid line(dark) : ρ original autocorrelation for time series with short and long periods
solid line(light) : autocorrelation after short period removed
broken line : autocorrelation after long period removed, shows overall trend

<Figure 1.3>



FFT of time series with periodic components

<Figure 1.4>

2. Description of the Problem

The data we are investigating are temperature readings measured from 9 different locations of a glass furnace. Each location produces time series data of temperature readings sampled at one minute interval spanning an entire month.

There are two features of the times series we need to pay attention to.

First, the temperature measurement is subject to steadily widening bias. It is measured by a thermometer embedded inside one of heat-resisting bricks which make up the wall of the furnace. The brick is subject to steady erosion by molten glass (silicon) which comes in contact with the brick. As time goes by, the brick becomes thinner and the thermometer is exposed to more heat. Naturally, even if the temperature of the furnace stays constant, the reading would become higher. A brief look at the data confirms that the data have a clear upward linear trend in many of 9 time series sets. Second, the temperature reading is also subject to periodic changes. Two sources of periodic pattern are already known.

The first one arises from the design of the furnace.

It reverses the flow every 20 minutes, resulting in corresponding temperature change with the same period. The other is due to nature. Temperature goes up during the day and down at night. It affects the temperature inside glass furnace, which also changes in the same direction.

These problems manifest themselves in autocorrelation chart. The presence of a linear component leads to an autocorrelation whose value converges to non-zero value, which cannot be dismissed as negligible. In our case, the autocorrelation is solidly stay above zero, never reaching zero even for very long distance autocorrelation. Of course, this could be also the indication of very long-distance periodic components.

In our study, time series typically has a strong linear component as can be seen in <Figure 1.1>. This leads to the situation where its autocorrelation stays near 1 regardless of time lag. See <Figure 1.2>.

The presence of periodic components can manifest itself as recurring peaks in the autocorrelation charts. In <Figure 1.2>, we can observe small recurring wave patterns in the autocorrelation plot, which is the contribution of periodic components. In <Figure 1.3>, we can see the autocorrelation plot after linear component has been removed from the time series. While the autocorrelation value is now decreasing overall as the distance (time lag) increases (overall trend : broken line), its value is fluctuating up and down in regular periodic pattern (dark-colored solid line.) By looking for such pattern in the chart, we can detect the presence of short periods. For the longer periods, it may be more difficult to detect the pattern because it is less pronounced (light-colored solid line). The pattern may cover fairly long period and it could escape the inspection of the chart by naked eye.

The use of Fast Fourier Transform (FFT) can reveal the existence of periods more directly as you can see in <Figure 1.4>. At certain periods, its amplitude from FFT result could be unusually high compared with neighboring periods. Those periods with significantly high local peaks in amplitude are due to periodic components. In our data from glass furnace, 20 minute period produces sharply high peaks in the region of FFT dominated by low amplitude. These are from 20 minute period and its harmonics. As for longer periods such as 24 hours or longer, their amplitude is not prominent enough to be readily noticeable. This is because they show up on the backdrop of high but 'rapidly decreasing' amplitude in the beginning segment of FFT, where long periods dominate. They can be better detected after some highlighting and filtering are done.

A device or a natural system tends to exhibit characteristic periodic behavior inherent in its structure. Such a behavior could be frequently made up of several layers of nested periods with different lengths. The data we are studying are one of such cases. A single time series has multiple dynamics overlaid in a hierarchical order, and which are moving at a different speed, hence having a different time-scale.

3. Theoretical Characterization

3.1 Normal distribution of data from stationary ARMA model

If we would use an old method such as Shewhart Chart for a time-series z_t, the series has to satisfy the constraint that each z_t is iid. This implies that the sampling interval of z_t should not be short enough to introduce significant autocorrelation between adjacent samples. Consequently, we can only take limited number of samples in a given duration.

It would be nice if we can take as many samples as we want and still carry out something similar to Shewhart Chart analysis. It depends on whether we can get around the requirement of the data being iid. In the following, we will

demonstrate that we can in case of stationary $ARMA(p, q)$ model.

For the following analysis, we will use $\tilde{z}_t = z_t - \mu$ where μ is a mean of all $\tilde{z}_t$'s.

That is, every $\tilde{z}_t$ has an identical mean 0. $\tilde{z}_t$ represents the deviation of z_t from its mean μ.

Then, $ARMA(p,q)$ model can be represented as follows :

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

That is,

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\tilde{z}_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t$$

or

$\phi(B)\tilde{z}_t = \theta(B) a_t$, where

$\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$

$\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t$

Here, a_t is iid with $a_t \sim N(0, \sigma_a^2)$ for all t, that is, a_t is a noise with normal distribution.

$\phi(B)$ can be decomposed as follows : $\phi(B) = (1 - G_1 B)(1 - G_2 B) \cdots (1 - G_p B)$.

$G_1^{-1}, \cdots, G_p^{-1}$ are roots of $\phi(B) = 0$. Then $\phi^{-1}(B)$ can be expressed as follows: $\phi^{-1}(B) = \sum_{i=1}^{p} \frac{K_i}{1 - G_i B}$ where K_i's are appropriate constants.

SinceARMA(p, q) is stationary, AR(p) part is also stationary. Let $\psi(B) = \phi^{-1}(B)$

Then, $\psi_m = \sum_{i=0}^{p} K_i G_i^m$, where $\Psi(B) = \sum_{m=0}^{\infty} \psi_m B^m$.

Therefore, the coefficients $\{\psi_m\}$ of $\psi(B)$ are absolutely summable. Then,

$$\begin{aligned}
\tilde{z}_t &= \phi(B)^{-1}\theta(B) a_t \\
&= \sum_{i=1}^{p} \frac{K_i}{1 - G_i B} \theta(B) a_t = \sum_{i=1}^{p} K_i \Big(\sum_{m=0}^{\infty} G_i^m B^m \Big)\ \theta(B) a_t \\
&= \sum_{i=1}^{p} K_i \Big(\sum_{m=0}^{\infty} G_i^m B^m \Big)\ (1 - \theta_1 B - \cdots - \theta_q B^q) a_t \\
&= \sum_{i=1}^{p} K_i \Big(\sum_{n=0}^{\infty} (G_i^n - \theta_1 G_i^{n-1} - \cdots - \theta_q G_i^{n-q}) \Big) B^n a_t \\
&= \sum_{n=0}^{\infty} K_i \sum_{i=1}^{p} (G_i^n - \sum_{j=1}^{\min(n,q)} G_i^{n-j}) a_{t-n}
\end{aligned}$$

Let $\alpha_n = \sum_{i=1}^{p} K_i (G_i^n - \sum_{j=1}^{\min(n,q)} G_i^{n-j})$.

Then

$$E(\tilde{z}_t) = \sum_{n=0}^{\infty} \alpha_n E(a_{t-n}) = \sum_{n=0}^{\infty} \alpha_n \times 0 = 0,$$

$$Var(\tilde{z}_t) = \sum_{n=0}^{\infty} \alpha_n^2 E(a_{t-n}^2) = \sigma_a^2 \sum_{n=0}^{\infty} \alpha_n^2$$

since all a_t's are iid.

What we can find out from this calculation is that $\tilde{z}_t$'s are not independent, but they all have the same normal distribution with mean 0 and variance $\sigma_a^2 \sum_{n=0}^{\infty} \alpha_n^2$.

Since $\alpha_n = (1 - \sum_{j=1}^{\min(n,q)} G_i^{-j}) \sum_{i=1}^{p} K_i G_i^n$

$= (1 - \sum_{j=1}^{\min(n,q)} G_i^{-j})\, \psi_n$,

α_n's are absolutely summable, and $\sum_{n=0}^{\infty} \alpha_n^2$ converges.

$Var(\tilde{z}_t) = \sigma_a^2 \sum_{n=0}^{\infty} \alpha_n^2$ converges as well.

We can conclude that, by calculating the sample mean and the standard deviation, we can set up the boundaries for acceptable parameters, with the upper boundary being the mean plus some constant multiple C of the standard deviation, and the lower boundary being the mean minus the said value. Any value falling outside the boundaries is considered 'out-of-control' event. This is the same process we used in Shewhart Chart.

3.2 Extending to PL-ARMA Model

PL-ARMA is a special case of ARIMA model. It has linear components and periodic components overlaid on top of stationary ARMA. AR part of PL-ARMA can be expressed as follows:

$\Phi(B) = L(B) P(B) G(B)$ where

(1) $L(B) = 1 - B$, a linear component

(2) $P(B) = (1 - 2B \cos\frac{2\pi}{n_1} + B^2) \times (1 - 2B \cos\frac{2\pi}{n_2} + B^2) \times \cdots \times (1 - 2B \cos\frac{2\pi}{n_p} + B^2)$

a periodic component

(3) $G(B)=(1-G_1B)(1-G_2B)\cdots(1-G_mB)$,
a stationary AR component

In $P(B)$, $(1-2B\cos\frac{2\pi}{n_k}+B^2)=(1-\omega_kB)(1-\overline{\omega_k}B)$

where $\omega_k=\cos\frac{2\pi}{n_k}+i\sin\frac{2\pi}{n_k}$,

$\overline{\omega_k}=\cos\frac{2\pi}{n_k}-i\sin\frac{2\pi}{n_k}$, $|\omega_k|=|\overline{\omega_k}|=1$ for all $k=1\cdots p$

In $G(B)$, $|G_i|<1$ for all $i=1\cdots m$.

As you can see, $L(B)$ and $P(B)$ are not stationary terms because B's coefficient in $L(B)$ is 1 and the ones in $P(B)$ are ω_k, and $\overline{\omega_k}$ whose absolute value is also 1. With these components added, AR component $\Phi(B)$ is no longer stationary. MA part of PL-ARMA is no different from stationary ARMA. We can use the same $\theta(B)$ from the discussion of ARMA model above. Then the entire model is expressed in the following equation:

$$\Phi(B)\,z_t=\theta(B)\,a_t$$

While the equation may look not much different from the one for stationary ARMA model except more added terms, the estimation of the PL-ARMA model is actually more complicated. If there is a linear trend, it would take the form of $z_t=a+bt$. The value of b can be estimated from $(1-B)$ transformation, that is, from $\{z_t-z_{t-1}\}$, the difference between two neighboring data value of time series. However, the value of a should be estimated again by plugging bt back to z_t.

If there is a periodic trend, it will pose further challenges. Extracting periodic component may not be achieved with conventional estimation method(or algorithm based on it.) Especially, if the component has a fairly long period spanning time steps, the estimated value n_k of period in $\omega_k=\cos\frac{2\pi}{n_k}+i\sin\frac{2\pi}{n_k}$ or $\overline{\omega_k}=\cos\frac{2\pi}{n_k}-i\sin\frac{2\pi}{n_k}$ can be fairly large as well. if n_k and n_l have large values, ω_k and ω_l (or $\overline{\omega_k}$ and $\overline{\omega_l}$) would be virtually indistinguishable from each other and also to 1.

This means the estimation requires extreme accuracy which may be practically impossible. A tiny error in estimation of ω_k and ω_l (or $\overline{\omega_k}$ and $\overline{\omega_l}$) could lead to wildly inaccurate estimate of n_k and n_l.

3.3 Three Stage Estimation Process for PL-ARMA

We propose that the estimation process is divided into three stages for a time series $\{x_t\}$:

stage 1 : Extracting a linear component

Perform a linear regression on $\{x_t\}$,
find a best fit $\hat{x_t}=a_0+b_0t$, and
let $y_t=x_t-\hat{x_t}$.
Note : $\overline{y_t}=0$ where $\overline{y_t}$ is the mean of y_t.

The dotted line in <Figure 2.1> is the best fit $\hat{x_t}=a_0+b_0t$. After removing it from the time series $\{x_t\}$, we get $\{y_t\}$ in <Figure 2.2>. The autocorrelation of $\{x_t\}$ in <Figure 2.4> is changed to that of $\{y_t\}$ in <Figure 2.5>.

stage 2 : Extracting a periodic component

a) Perform Fast Fourier Transform on $\{y_t\}$ and select periods with usually high amplitude.
Let $\mathbf{P}=\{n_1, n_2, \cdots, n_p\}$ be the set of such periods The process to find such amplitudes is not simple and would be described in detail later in the paper.

b) Perform multiple regression on $\{y_t\}$ using periodic terms with the selected periods and find the best fit:

$$\hat{y_t}=c+\sum_{i=1}^{p}(a_i\cos\frac{2\pi t}{n_i}+b_i\sin\frac{2\pi t}{n_i})$$

The regression is performed as follows :

$Y=\{y_t\}$, $E=\{e_t\}$, $C_i=\left\{\cos\frac{2\pi t}{n_i}\right\}$,

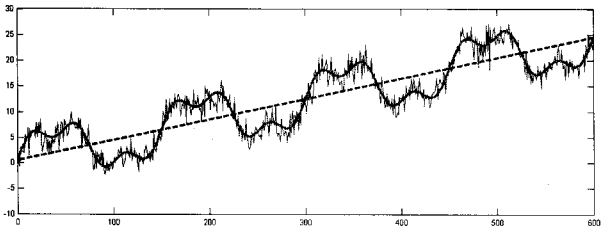
$S_i=\left\{\sin\frac{2\pi t}{n_i}\right\}$, for $i=1\cdots p$

$Y=E+c+\sum_{i=1}^{p}(a_iC_i+b_iS_i)$

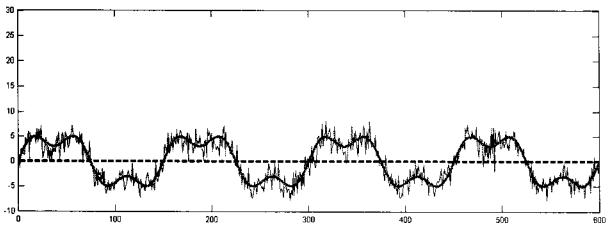
That is, $e_t=y_t-(c+\sum_{i=1}^{p}(a_i\cos\frac{2\pi t}{n_i}+b_i\sin\frac{2\pi t}{n_i}))$

Calculate c, a_i, and b_i which minimize $\sum_{t=1}^{T}e_t^2$

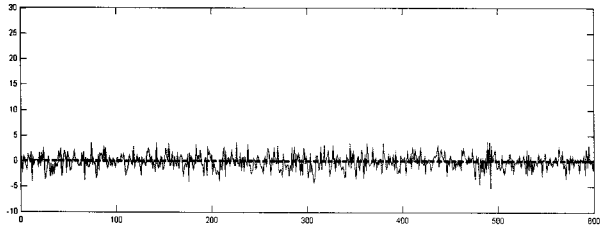
Note : T is the length of time series.



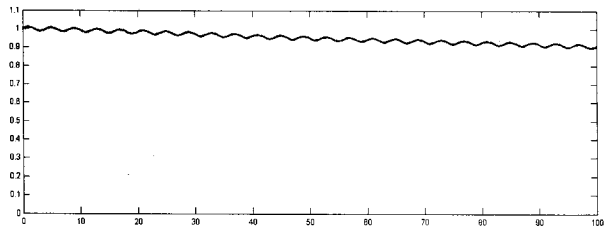
<Figure 2.1>



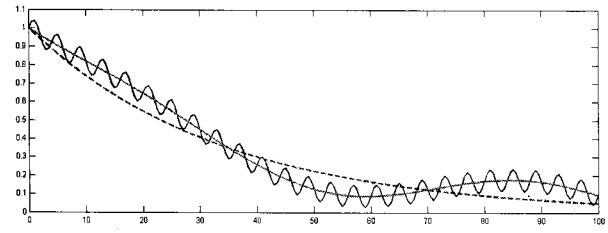
<Figure 2.2>



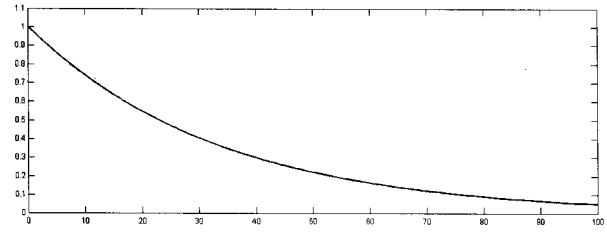
<Figure 2.3>



<Figure 2.4>



<Figure 2.5>



<Figure 2.1>

Let $z_t = y_t - \hat{y}_t$.

Expanding the right side of the equation, we get:

$$z_t = x_t - (\hat{x}_t + \hat{y}_t)$$
$$= x_t - (a_0 + c + b_0 t + \sum_{i=1}^{p} (a_i \cos\frac{2\pi t}{n_i} + b_i \sin\frac{2\pi t}{n_i})$$

The smooth curve in <Figure 2.2> represents the periodic component, $\hat{y}_t$. After it is removed from $\{y_t\}$, we get time series $\{z_t\}$, which is shown in <Figure 2.3>. With periodic component removed, autocorrelation of $\{y_t\}$ shown in <Figure 2.5> (dark-colored solid line) is changed to that of $\{z_t\}$ in <Figure 2.6>.

With a linear component $\hat{x}_t$ and a periodic component $\hat{y}_t$ removed from the original time series value x_t, resulting new time series $\{z_t\}$ is now more amenable to ARMA model estimation method.

stage 3 : Finding the best-fit ARMA model for $\{z_t\}$

This can be done using many existing programs from SAS or MATLAB. While linear component $\hat{x}_t$ or periodic component $\hat{y}_t$ generates high autocorrelation over fairly long distance, thus creating long distance dependency, we assume that the ARMA model for $\{z_t\}$ has dependency over relatively short distance. That is, AR and MA component for $\{z_t\}$ contain terms from the relatively recent past, not inordinate number of time steps away.

There is still a problem to address. Even $\{z_t\}$ may not be perfect instance of ARMA model, even though it could be quite close. Some data segment in $\{z_t\}$ could still distort the estimation outcome.

As we have seen in previous sections, ARMA model implies that data from $\{z_t\}$ should make normal distribution. Practically speaking, $\{z_t\}$ should be very close-fit of normal distribution.

We devised our own test to determine how closely the data from $\{z_t\}$ resembles to normal distribution. This is done by making use of three known measure of normality fitness : Kolmogorov-Smirnov Statistic Λ, skewness ϑ, and kurtosis κ. For data from $\{z_t\}$, we calculate three value. The our normality fit N can be defined as follows :

$$N = 10|\Lambda| + |\vartheta| + |\kappa|.$$

In our study, we set the cut-off line to be 1. That is, if $N \leq 1$ then $\{z_t\}$ is judged to be a good fit of normal distribution.
if $N > 1$ then $\{z_t\}$ is not a good fit of normal distribution.

For $\{z_t\}$ which is possibly not a good fit for ARMA model, we still have to compute the best estimate of ARMA model for them. It is a question of how we find a way to discount non-ARMA aspect of the time series and find a decent estimate of ARMA model for $\{z_t\}$.

This can be addressed by MA-approximation method. It will be described in full in the following sections. With MA-approximation, we can finally get adequate ARMA model for $\{z_t\}$.

The discussion on setting up the proper control boundary, UCL and LCL for $\{x_t\}$ will follow later. We briefly mention here that it will involve the use of average deviation of $\{\hat{y}_t\}$, standard deviation and R-bar value of $\{z_t\}$ along with the linear component $\hat{x_t} = a_0 + b_0 t$

4. Detection of periodic components

4.1 Time-scale

Time-scale can be seen as a time span during which later data can have high positive correlation with the earlier data in the span. Beyond it, the correlation value is very low or even negative. That is, beyond the time span, the autocorrelation between the earlier data and the later data becomes close to zero.

This kind of characterization may appear to present complication in case of time-series data of period p with no random components, because it is possible that data can have high positive correlation with those far back in the time series, if the distance between them is a multiple of p, their correlation can reach 1. Besides, the series is made up of data following deterministic pattern. However, it turns out that $\rho(\frac{p}{2}) = 0$ even in such a case. So if we choose the minimum time span during which autocorrelation between data is above zero, the time-scale for a periodic time series can be also defined using autocorrelation, even though the time series is not random but observing simple deterministic pattern. The only caveat is that the time span is not identical to the period, but the half of its length.

It is not enough to use ARIMA model for the time series with long-term dependency and multiple time scale, including nested multiple periods. Data we analyze frequently has long-term dependency. While theoretically possible, estimating ARIMA model to account for long-distance dependency is usually time-consuming and not accurate. As an ad hoc method, one can produce shortened time-series made up of data sampled with particular interval apart. However, this tends to be arbitrary. In addition, there is a problem of multiple time-scales. Some preliminary attempts have been made by Box and othersl [4, 5] to tackle time series with periodic components. However, they have not been fully explored.

Before moving on, we need to define what 'multiple time-scales' means. In many real time-series data, we can find multiple factors changing at a different rate. For example, time-series data of temperature inside a large furnace shows that there is a very short recurring pattern with period of 20 minutes, another pattern with 24 hour period, another with 1 year period (seasonal change). There are other intermediate lengths of periods in-between ranging from a few days to a month. Some of these periods are pronounced while others are not. Each of such periods are overlaid one over another to make a trend which accounts for large portion of given time-series data. With this in place, the remaining portion can be modeled after ARIMA or stationary ARMA model.

Multiple time-scale patterns do not always have to be periodic.

1) Frequently, such patterns are not exact. They are subject to minor changes.
2) More importantly, we do see multiple time-scales in time-series data with no noticeable periodic components.

One may wonder how do you define time-scales in such a case. Time-scale is related to how fast data value changes. The pace of increase and decrease determines it. If the rate of change is faster, data has shorter time scale, while slower rate of change, longer time scale. However, in the model discussed here, we assume that well-defined periods exist for time-series data. More general treatment of time-scale will be done in future research.

In our model, there is a particular relation among these time-scales. We assume that time-scales obey certain restriction. That is, the relative ratio between them should be gen-

erated from a handful of simple ratios. They are 1 : 2, 1 : 3 or in some cases, 2 : 3. Based on these ratios, one can generate general ratio of $1 : 2^m 3^n$ (m, n are non-negative integers and at least one of them is positive). Or in some cases, $2^m : 3^n$ is possible for small m, n. So time scales can constitute a hierarchy with certain kind of rations among them. As we can see later, this has the effect of restricting the possible choices of time scale factors.

One may ask why we should impose such an assumption. It could be unnecessarily restrictive and may turn out to be unrealistic.

However, this kind of ratio emerges naturally in many real phenomena. Periods or frequencies are coupled in such a ratio. It is called locking of periods (frequencies.) This also provides us with the simplification of our task. It is not that we arbitrarily chose this restriction. Nature imposes such a simplified order in physical phenomena.

4.2 Extraction of Periods

In this section, we describe the method to extract periods for periodic component $\hat{y_t}$. After applying Fast Fourier Transform to time series data $\{y_t\}$, we want to select those periods (or frequencies) whose amplitude is unusually high compared with other periods in its neighborhood. The amplitude should be salient in certain local segment of FFT chart. It is similar to picking a salient local peak in a single variable function. As mentioned previously, we assume salient periods make the nested period hierarchy (in terms of period length) where each period is simple integer multiples (2, 3, etc) or its inverses (1/2, 1/3) of one above or below in the hierarchy. This means that salient periods would occur in exponentially longer intervals as frequency becomes higher (or period gets shorter). As a result, in the area of long periods (low frequencies), a peak is picked from relatively narrow neighborhood, while in the area of short periods (high frequencies), it is picked from fairly wide neighborhood.

The extraction of periods proceeds as follows:

1) Apply Fast Fourier Transform on time series data and create a list whose entry is a log value of each amplitude in FFT result.
2) Subtract the below-trend amplitude from the original amplitude after finding a trend curve and sometimes additionally apply some threshold for further suppression of below-average amplitude.
3) Smoothen the result in (3) using Polynomial Weight Filter (PWF).

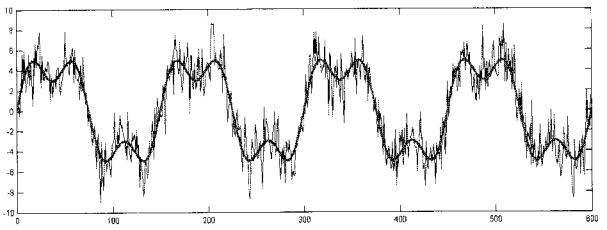
4.2.1 FFT and log Transformation

Apply Fast Fourier Transform on times series $\{y_t\}$ and obtain:

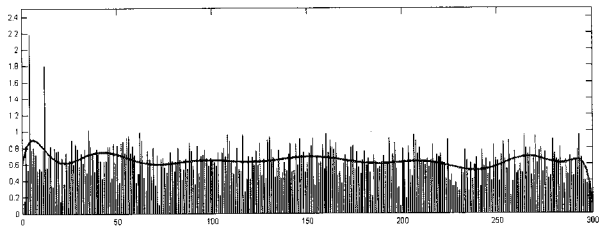
$F = (f_1, f_2, \cdots, f_T)$ where T is the length of $\{y_t\}$.

Transform each value f_i of F using function $\zeta(x) = \log_{10}(|x|+10) - 1$ and obtain L:

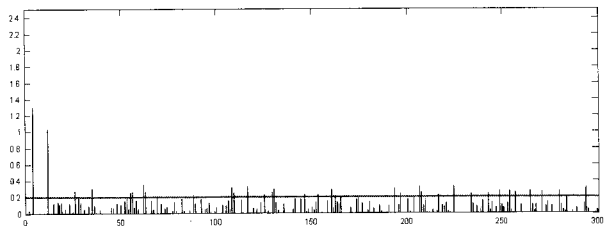
$$L = (l_1, l_2, \cdots, l_T),\ l_i = \zeta(f_i) \text{ for } i = 1, \cdots, T.$$



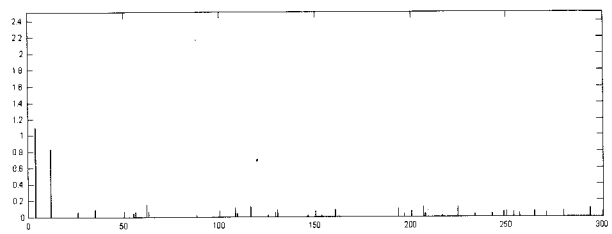
<Figure 3.1>



<Figure 3.2>



<Figure 3.2>



<Figure 3.2>

The reason we use this transformation is that the value of $|f_i|$ can wildly vary, that is, the difference between ampli-

tudes $|f_i|$ and $|f_j|$ of two period(or frequency) can be better measured in terms of ratio $|f_i|/|f_j|$ than in terms of $|f_i|-|f_j|$. In short. the difference can be exponential Hence, it is more sensible to take log value of each amplitude for our processing. The next factor we need to consider is that we are only interested in finding periods(frequencies) of unusually high amplitude. Those with lower amplitudes are not much of our concern. While we need to preserve the difference among the higher values, we do not need to highlight the difference among many lower values, especially excessively big negative log values for $|f_i|$ quite close to 0.

The function $\zeta(x)$ would serve this purpose. $\zeta(x) \geq 0$ and it will not exaggerate the difference among small values(especially those close to 0) of $|x|$. For x with large $|x|$, the difference between $\log_{10}|x|$ and $\zeta(x)$ converges asymptotically to 1, which means that for two large positive values x and y, $|\log_{10}|x|-\log_{10}|y||$ is almost equal to $|\zeta(x)-\zeta(y)|$, effectively preserving the difference. In <Figure 3.1>, we have a typical time series data with periodic component in it. <Figure 3.2> shows FFT plot after applying $\zeta(x)$ on its values.

4.2.2 Pruning Lower Part of Amplitude

Even after completing the transformation above, the graph of L would show that the amplitude is taking the shape of generally descending curve. Using a least square method, we compute the best fit polynomial curve of degree k, $\ell(i) = \sum_{m=1}^{k} a_m i^m$, for $L=(l_1, l_2, \cdots, l_T)$.

$\ell(i)$ can be seen as a general trend of L. Now subtract $\ell(i)$ from L, and ignores any negative values, filling them with zero instead. Then we have:

$$D=(d_1, d_2, \cdots, d_T) \text{ where } d_i=\max(0, l_i-\ell(i)).$$

This will better highlight salient local peaks, especially the ones in the region of long periods(or low frequencies.) The curve in <Figure 3.2> shows $\ell(i)$. There are two high peaks on the left-end side along with some low amplitude. In this simple case, the curve does not show definite descending trend, but in general it is indeed the case (See the discussion on the final results of analysis later).

The plot of D is shown in <Figure 3.3>.

Now general trend curve of D would be almost horizontal. Instead of using trend curve, we will compute the mean value of D and subtract it and some more from D, creating :

$$D^{(1)}=(d_1^1, d_2^1, \cdots, d_T^1) \text{ where}$$

$d_i^1=\max(0, d_i-(\bar{d}+s(d)))$ and $\bar{d}$ is D's mean and $s(d)$, D's sample standard deviation. In <Figure 3.3>, the horizontal line is the cut-off line $\bar{d}+s(d)$ and <Figure 3.4> shows $D^{(1)}$.

If needed, this operation can be repeated further. After i increases, $D^{(i)}$ will have smaller and smaller mean and sample standard deviation. Only those with unusually high amplitudes will remain having non-trivial amplitude values, while the rest having 0.

4.2.3 Smoothing Using Polynomial Weight Filter (PWF)

Here, we will discuss how to apply smoothing filter on $D^{(m)}$ to locate salient peaks with unusually high amplitude.

Before moving further, we examine the property of two well-known filters, moving average filter, exponential decay filter, and later compare them with polynomial weight filter(PWF) which we will define.

(1) Moving Average Filter (MAF)

The following is the typical moving average filter. It has fixed window width of $2l+1$. The window width can be an even number, too. However, we chose an odd number for simplicity. Every value within the window will be given equal weight and f_n will be the average of $2l+1$ values inside the window.

$$f_n=\frac{1}{2l+1}(d_{n-l}+\cdots+d_{n+l}), \text{ width } 2l+1.$$

definition of extended values on either boundary.

$$f_{-l+1}=\cdots=f_0=f_1 \text{ and } f_T=f_{T+1}=\cdots=f_{T+l}.$$

(2) Exponential Decay Filter (EDF)

This is the filter with exponential decay. It can be defined as follows :

$$f_1=d_1 \text{ and } \rho<1$$

$$f_n=(1-\frac{\rho(1-\rho^{n-1})}{1-\rho^n})d_n+\frac{\rho(1-\rho^{n-1})}{1-\rho^n}f_{n-1},$$

$$\simeq(1-\rho)d_n+\rho f_{n-1}$$

In general,

$$f_n = (\sum_{i=0}^{n-1} d_{n-i}\rho^i)/\sum_{i=0}^{n-1}\rho^i = \frac{d_1+\rho d_2+\cdots+\rho^{n-1}d_n}{1+\rho+\cdots+\rho^{n-1}}$$

There is no fixed window width for this filter. All n values will contribute to the weighted average. However, values beyond a certain distance l from n are given so little weight that their entire sum would be negligible. It can be called an **effective window** of length l, and it depends on threshold ϵ, the total sum of residual weights which can be safely dismissed. It can be computed as follows:

$S(m, n) = \sum_{i=m}^{n}\rho^i$. Then, $S(0,\infty) = S(0, l-1) + S(l,\infty)$,

$$S(l,\infty) = \rho^l S(0,\infty) = \frac{\rho^l}{1-\rho}$$

$l(\epsilon)$ is defined as the smallest l for which

$\frac{S(l,\infty)}{S(0,l)} = \frac{\rho^l}{1-\rho^l} < \epsilon$. That is, $l(\epsilon) = \lceil \log_\rho(\frac{\epsilon}{1+\epsilon}) \rceil$.

$l(\epsilon)$ is the width of effective window with threshold ϵ.

(3) Polynomial Weight Filter(PWF)

The above two weight filters are commonly used but as we have shown, their (effective) window length is fixed. This is not what we want. We want a filter whose effective window is progressively widening as we move up in the direction of shorter period (higher frequency). As we said before, in our study, we assume salient periods make the nested period hierarchy (in terms of period length) where each period is simple integer multiples (2, 3, etc) or its inverses (1/2, 1/3, etc) of one above or below in the hierarchy.

That is, the time series has a periodic component which has periodic term for each period n_i in the following set P:

$P = \{n_1, n_2, \cdots, n_p\}$ where $n_1 < n_2 < \cdots < n_p$

and $n_{i+1}/n_i = 2^{j_i}3^{k_i}$ for $i = 1, \cdots, p-1$.

In this case, the distribution of salient periods (or frequencies) over period (frequency) axis would be exponentially sparse as we move up from longer periods (low frequencies) to shorter periods (higher frequencies).

To pick up salient periods in the range of shorter periods(higher frequencies,) we should have proportionally wider window for smoothing operation to weed out insignificant peaks over the proportionally wider area, leaving only the one with a salient amplitude peak.

For this purpose, we define Polynomial Weight Filter of Order α as follows:

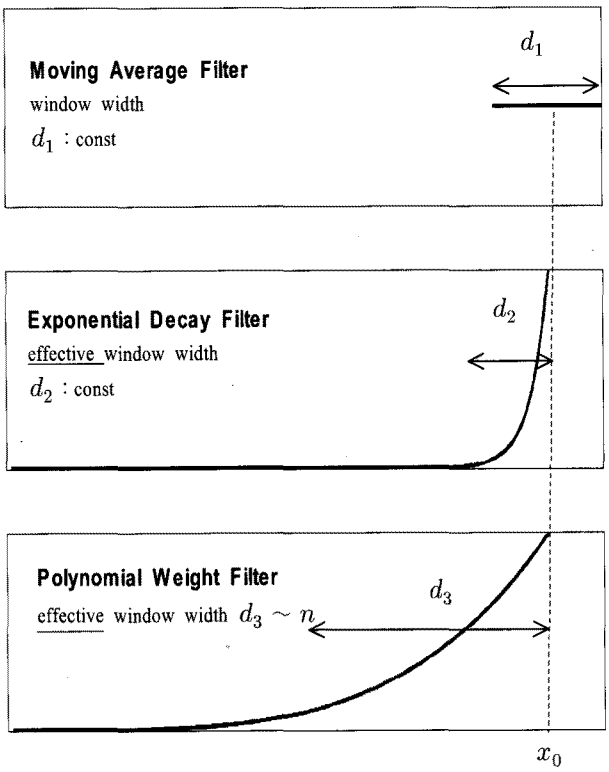
$$f_n = (1-(\frac{n-1}{n})^\alpha)d_n + (\frac{n-1}{n})^\alpha f_{n-1},$$
$$= \frac{1}{n^\alpha}((n^\alpha-(n-1)^\alpha)d_n + (n-1)^\alpha f_{n-1})$$
$$f_1 = d_1$$

In the case of $\alpha = 1$, f_n is a simple average of $d_1, \cdots, d_n$, $\frac{1}{n}(d_1+\cdots+d_n)$

In general,

$$f_n = \frac{1d_1 + (2^\alpha-1)d_2 + \cdots + (k^\alpha-(k-1)^\alpha)d_3 + \cdots + (n^\alpha-(n-1)^\alpha)d_n}{n^\alpha}$$

Now, given a threshold ϵ, let us calculate the width of effective window as we did in Exponentially Decaying Filter. Let l be the effective window width with threshold ϵ. Then the total weight allocated to $d_1, \cdots, d_{n-l}$ is:



<Figure 4>

$$\sum_{k=1}^{n-l}\frac{(k^\alpha-(k-1)^\alpha)}{n^\alpha}=\frac{(n-l)^\alpha}{n^\alpha}$$

Since $\frac{(n-l)^\alpha}{n^\alpha}=(\frac{n-l}{n})^\alpha<\epsilon,\ l>(1-\epsilon^{\frac{1}{\alpha}})n.$

So $l=\lceil (1-\epsilon^{\frac{1}{\alpha}})n \rceil$

Therefore, $l\sim n$, i.e., l is proportional to n.

If $\epsilon=1/8$ and $\alpha=3$, we can see that $l=\lceil \frac{n}{2} \rceil$.

<Figure 4> shows the difference among three filters. While the (effective) window width of both Moving Average Filter d_1 and Exponential Decay Filter d_2 are constant in length, the effective window width of Polynomial Weight Filter is proportional to n.

5. MA-Approximation

As we have said previously, for time series $\{z_t\}$ which may not be a perfect instance for ARMA model, we may still have to compute the best estimate of ARMA model for them. It is a question of how we find a way to discount non-ARMA aspect of the time series and find a decent estimate of ARMA model for $\{z_t\}$.

One of the most unsettling outcome we frequently encounter in estimating ARMA model for $\{z_t\}$ with some non-ARMA aspect is that such an anomaly could push mean derived from estimated ARMA model too far away from the sample mean of $\{z_t\}$. The sample mean of $\{z_t\}$ and the mean from its estimate ARMA model doesn't have to agree, but in some cases, the difference between the two become too large. In extreme cases, the difference is more than sample standard deviation of a time series. Especially, AR part of the model is usually responsible for such an outcome. On the one hand, there is a need to minimize residual errors in estimating ARMA model. On the other hand, we need to prevent the resulting model from being too skewed in one direction, because of some anomaly present inside the time series data. It would be best if we can identify the nature of anomaly and have method to systematically factor it out, as we did with linear and periodic components. However, despite our best efforts, there would be always some anomaly we could not identify. All we can do is to remove as many non-ARMA components as possible, and produce the time series data most amenable to ARMA model estimation.

To do this, we need to create a new criteria to measure the fitness of given ARMA model for some time series $\{z_t\}$.

5.1 Theoretical Framework

Recall our typical ARMA(p, q) model:

$$x_t=\phi_1x_{t-1}+\cdots+\phi_px_{t-p}+a_t-\theta_1a_{t-1}-\cdots-\theta_qa_{t-q}$$

That is,

$$(1-\phi_1B-\cdots-\phi_pB^p)x_t=(1-\theta_1B-\cdots-\theta_qB^q)a_t,$$
$$\text{or } \phi(B)x_t=\theta(B)a_t,$$

where

$$\phi(B)=(1-\phi_1B-\cdots-\phi_pB^p),$$
$$\theta(B)=(1-\theta_1B-\cdots-\theta_qB^q)a_t$$

Here, a_t is iid with $a_t\sim N(0,\sigma_a^2)$ for all t.

Then, we have shown that z_t can be transformed into a following form:

$$x_t=\alpha_0a_t+\alpha_1a_{t-1}+\cdots+\alpha_ia_{t-i}+\cdots+\cdots=\sum_{n=0}^{\infty}\alpha_na_{t-n}$$

where α_n's are absolutely summable. In particular, $\lim_{n\to\infty}\alpha_n=0$.

That is, z_t can be expressed as infinite sum of MA terms. All ARMA model can be converted into MA model of infinite degree where coefficients α_n's are absolutely summable.

Let $z^{(k)}$ is the best fit for time series $\{z_t\}$ among all possible MA models of degree k (note : MA model of degree k = ARMA(0, k)).

If $\{z_t\}$ is indeed an instance of ARMA model, as k grows, $z^{(k)}$ converges to the best fit ARMA model. Even when $\{z_t\}$ is not a perfect instance of ARMA model, $z^{(k)}$ could converges to the best possible ARMA approximation of $\{z_t\}$.

It should be noted that residual error r_k^2 of $z^{(k)}$ is a monotonically decreasing sequence. As k grows, r_k^2 gets smaller.

On the other hand, μ_k, the mean of estimated $z^{(k)}$, could diverge from $\bar{z}$, the sample mean of $\{z_t\}$, as k grows, if

$\{z_t\}$ is not a perfect instance of ARMA model. This does not always happen, but it does happen once in a while. For $\{z_t\}$, initially the difference is zero, that is, $\mu_1 - \bar{z} = 0$. It generally increases but not monotonically, our study shows.

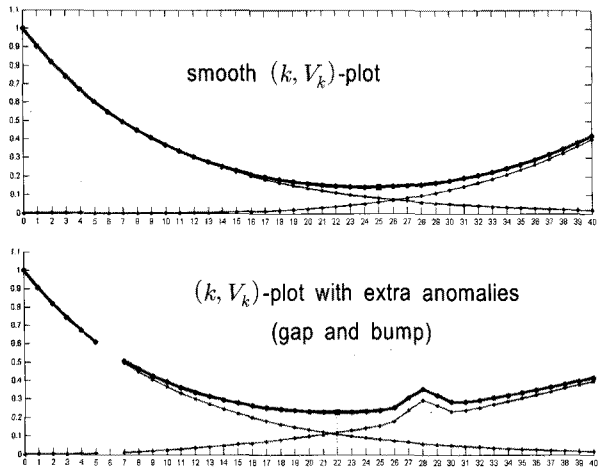
Now we define the fitness measure V_k for $z^{(k)}$:

$V_k = \dfrac{r_k + |\mu_k - \bar{z}|}{s_z}$ where s_z is the sample standard deviation of $\{z_t\}$ and serves as a normalizing factor. Now, we can see that, on the right side of equation, r_k is declining but $|\mu_k - \bar{z}|$ is rising overall, as k increases.

Now find $z^{(k)}$ for successive values of k, and plot (k, V_k) graph.

(k, V_k)-plot is shown in <Figure 5>. In the top graph, the dark solid line is for V_k, and the (increasing) light solid line is for $|\mu_k - \bar{z}|/s_z$, while (decreasing) light solid line is for r_k / s_z. As you can see, the value of V_k generally drops and slowly go back up. The pace of rise is frequently quite slow compared with the initial drop. The little square in the chart marks the minimum of V_k. Theoretically V_k can take lower value for some larger value of k beyond values plotted in the graph, but practically speaking, it would be quite unlikely.

In the second graph, one can notice that there is some gap in the graph. For certain value of k, the estimation of $z^{(k)}$ by standard time-series software package simply fails. The graph of $|\mu_k - \bar{z}|/s_z$ shows a bump in the latter part of slope. It does not rise smoothly when the time series contains some non-ARMA feature.



dark solid line : V_k
light solid line(decreasing) : r_k / s_z
light solid line(increasing) : $|\mu_k - \bar{z}| / s_z$

<Figure 5>

In addition to large gap between μ_k and $\bar{z}$ for higher k value, we can add two more indicators to the criteria for how far $\{z_t\}$ is from real ARMA model:

(1) $|\mu_k - \bar{z}|/s_z$ does not rise smoothly
(2) gaps are found in (k, V_k) graph

Whether $\{z_t\}$ turns out to be close to real ARMA model or not, best ARMA fit can be found. Let $(k_{\min}, V_{k_{\min}})$ is the point marked by a red square. $V_{k_{Min}}$ would be the minimum value of V_k. Then, $z^{(k_{\min})}$ is the best fit MA(or ARMA) model for $\{z_t\}$.

6. Setting up Control Boundary

In preceding sections, we presented how to separate linear $\{\hat{x}_t\}$ and periodic components $\{\hat{y}_t\}$ from time series $\{x_n\}$, creating $\{z_n\}$ for which we can find the best fit $z^{(k_{\min})}$ of MA model. Now we turn our attention to setting up the control boundary to detect out-of-control event. Setting up the control boundary has been studied by many researchers [1, 3, 9] for non-periodic or iid cases.

6.1 Control boundary for $\{z_n\}$

Since we decided to regard $\{z_n\}$ as ARMA model time series, we compute $z^{(k_{\min})}$, the MA model which best fits $\{z_n\}$. We get estimated mean μ_k from $z^{(k_{\min})}$ as the baseline of our control boundary. Now there are two ways to set upper control limit(UCL) and lower control limit(LCL.)

Using s_z : We define UCL$= \mu_{k_{\min}} + 3s_z$, LCL$= \mu_{k_{\min}} - 3s_z$ where s_z is the sample standard deviation of $\{z_n\}$. The justification of using s_z in defining the two control limits is that values of $\{z_n\}$ constitute a normal distribution since it is assumed to be an ARMA model.

Using $\overline{R_z}(q)$: Apply the technique borrowed from Range Chart ($\overline{R}$-chart.) First, find the shortest time lag $d_{\min}$ at which autocorrelation goes 'near zero' in $\{z_n\}$ (criteria for 'near zero' to be described in the next paragraph).

For a group $Z_m(q)=\{z_m, z_{m+d_{\min}}, \cdots, z_{m+(q-1)d_{\min}}\}$, define

$$R_z(m,q)=\frac{|\max(Z_m(q))-\min(Z_m(q))|}{d_R(q)},$$

and

$$\overline{R_z}(q)=\frac{1}{T-(q-1)d_{\min}}\sum_{m=1}^{T-(q-1)d_{\min}} R_z(m,q)$$

(Note : $2\le q\le 10$ and $d_R(q)$ can be found in <Table 1>)

<Table 1> dR-Table

k (sample size)	$d_R(k)$ (divisor)	k	$d_R(k)$	k	$d_R(k)$
2	1.128	5	2.326	8	2.847
3	1.693	6	2.534	9	2.970
4	2.059	7	2.704	10	3.078

The we define :

$$\text{UCL}=\mu_{k_{\min}}+3\overline{R_z}(q),\ \text{LCL}=\mu_{k_{\min}}-3\overline{R_z}(q).$$

In our study, we set $q=6$.

6.1.1 How to calculate $d_{\min}$

Plot the autocorrelation chart of $\{z_n\}$. Define near-zero value δ.

Find the shortest time lag $d_{\min}$ for which $|\rho_{d_{\min}}|<\delta$. In our study, we set $\delta=0.1$.

So any pair of values in $\{z_n\}$ which are $d_{\min}$ apart, their autocorrelation is quite low, close to zero. This means that such pair of values are statistically 'nearly independent.' So practically speaking, this would satisfy the statistical condition for applying Range Chart method. One can set δ to a different value. It all depends on the nature of application.

6.2 Control boundary for $\{x_n\}$

As you remember, $x_n=z_n+\hat{y_n}+\hat{x_n}$, where $\hat{x_t}$ is a linear component, $\hat{y_n}$, a periodic component. We already know how to set up the control boundary for $\{z_n\}$.

To set up the control boundary for $\{x_n\}$, we will add contributions from $\hat{x_t}$ and $\hat{y_n}$ to the control boundary for $\{z_n\}$. The question is how to do it. For $\hat{y_n}$, we need to have some way to measure its deviation away from its mean. Since $\hat{y_n}$ is a periodic pattern, its values would not make up a normal distribution. They may be distributed more evenly than values from normal distribution which tend concentrate more around its mean. So the better measure of deviation could be average deviation $a_{\hat{y}}$, which tend to discount the contribution of extreme values compared with standard deviation. It can be defined as follows :

$$a_{\hat{y}}=\frac{1}{T}\sum_{t=1}^{T}|\hat{y_t}-\bar{\hat{y}}| \text{ where } \bar{\hat{y}}=\frac{1}{T}\sum_{t=1}^{T}\hat{y_t}, \text{ the mean of } \hat{y_t}$$

For linear component $\hat{x_t}$, we won't measure its deviation in any way. Linear component would be directly incorporated into control boundary of $\{x_n\}$. Let me explain this further.

If $\hat{x_t}=a_0+b_0t$, UCL and LCL for $\{x_n\}$ are :

(1) Using s_z :

$$\begin{aligned}\text{UCL}&=\hat{x_t}+(\bar{\hat{y}}+Aa_{\hat{y}})+(\mu_{k_{\min}}+Ds_z)\\&=(a_0+\bar{\hat{y}}+\mu_{k_{\min}})+b_0t+Aa_{\hat{y}}+Ds_z\end{aligned}$$

(A and D are positive real numbers.)

$$\begin{aligned}\text{LCL}&=\hat{x_t}+(\bar{\hat{y}}-Aa_{\hat{y}})+(\mu_{k_{\min}}-Ds_z)\\&=(a_0+\bar{\hat{y}}+\mu_{k_{\min}})+b_0t-Aa_{\hat{y}}-Ds_z\end{aligned}$$

(2) Using $\overline{R_z}(q)$:

$$\begin{aligned}\text{UCL}&=\hat{x_t}+(\bar{\hat{y}}+Aa_{\hat{y}})+(\mu_{k_{\min}}+D\overline{R_z}(q))\\&=(a_0+\bar{\hat{y}}+\mu_{k_{\min}})+b_0t+Aa_{\hat{y}}+D\overline{R_z}(q)\end{aligned}$$

(A and D are positive real numbers)

$$\begin{aligned}\text{LCL}&=\hat{x_t}+(\bar{\hat{y}}-Aa_{\hat{y}})+(\mu_{k_{\min}}-D\overline{R_z}(q))\\&=(a_0+\bar{\hat{y}}+\mu_{k_{\min}})+b_0t-Aa_{\hat{y}}-D\overline{R_z}(q)\end{aligned}$$

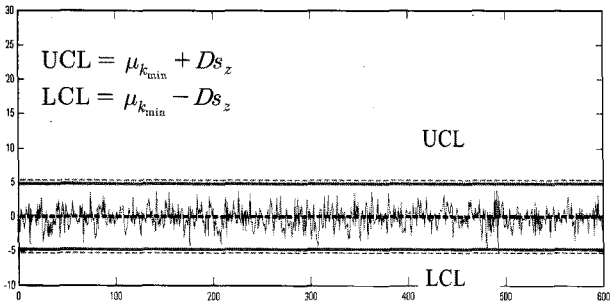
Most of the time, $D=3$. As for A, it may vary depending on the situation. In our case, $A=2$. $\bar{\hat{y}}$ and $\mu_{k_{\min}}$ are often close to zero, and the above equations may turn out to be no different from:

(1) $\text{UCL}=a_0+b_0t+Aa_{\hat{y}}+Ds_z$,
$\text{LCL}=a_0+b_0t-Aa_{\hat{y}}-Ds_z$

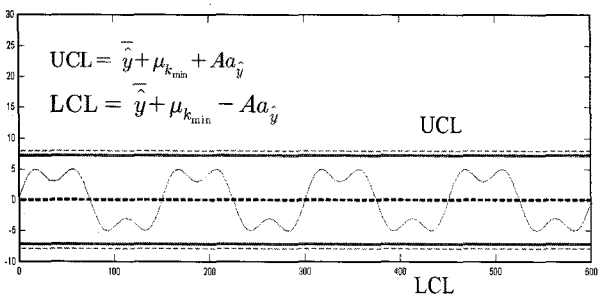
(2) $\text{UCL}=a_0+b_0t+Aa_{\hat{y}}+D\overline{R_z}(q)$,
$\text{LCL}=a_0+b_0t-Aa_{\hat{y}}-D\overline{R_z}(q)$

As you can see, UCL and LCL are no longer just horizontal lines, but could be slanted linear trend. The control boundary of $\{x_n\}$ are a pair of trend lines. <Figure 6> shows the whole process.

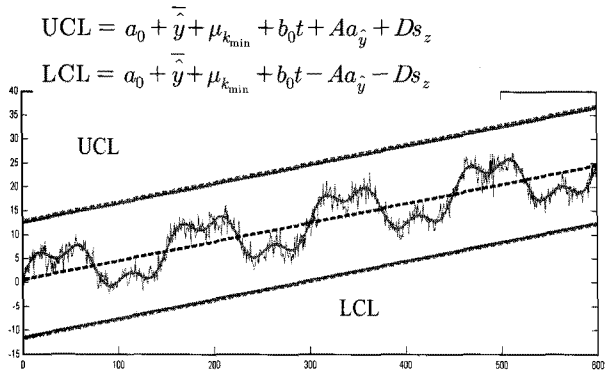
Finally, there is some room for debate regarding the nature of periodic components. In our study, we regard periodic variation as something undesirable, hence we included its average deviation $a_{\hat{y}}$ in setting up the control boundary of (original) time series $\{x_n\}$. However, in some cases, we could regard periodic variation as acceptable change. If that is the case, we only need to set up the control boundary for $\{z_n\}$ to detect out-of-control event. As for linear change trend, in our cases, it stems from measuring problem due to erosion in heat-resistant bricks in which sensors are embedded, not the actual change of temperature inside glass furnace itself. Naturally, the linear change should be regarded as acceptable. In different circumstances, linear trend change itself could be a problem. Then we may have to find a way to set the control boundary for linear trend.



$\{z_t\}$ ARMA via MA-approximation



$\hat{y}_t$ periodic component



x_t original time series

<Figure 6>

7. Conclusion

In this research, we investigated the way to handle linear and periodic components of a time series, which we often encounter among time series obtained from the real world(glass furnace.) First, we extract linear trend and periodic components from original time series $\{x_t\}$. Periods are extracted using FFT and Polynomial Weight Filter(PWF.) Then, after the two components are removed, we apply MA-approximation method to find the best fit MA fit for the resulting time series $\{z_t\}$. Then we set up the new kind of control boundary incorporating average deviation of periodic component $a_{\hat{y}}$.

This study requires some further investigation using more cases of time series. The proper boundary value for UCL and LCL is one of the things which needs further empirical investigation. At the same time, we may need to put some efforts to simplify the whole process. Actual results of applying this method to real-world time-series data have been compiled. However, the discussion on the results would take up too much additional space, and are not included in this paper. It will be described in some future publication on this topic.

In addition to improvements that may be required for our results, we should explore the way to expand the scope of our investigation to multiple concurrent time series data. In the real-life situations, we get multiple time series data from various sensors located in target facilities. They are usually correlated in some way, and frequently have causative relations. In order to properly diagnose the problem, focusing on each time series separately could be a limited approach.

At the same time, the idea of time scale merits more research, too. Each time series has its own unique dynamic, represented in the presence of multiple time scale. Further variables can be introduced for a factor corresponding to each time scale. Such ideas are now merely at incubating

stage. We would like to see this angle pursued in future research.

All these further investigation would help better and reliable control of facilities to be managed.

References

[1] 구자활; "자기상관 있는 공정에서 공정진단과 관리방법에 관한 연구", 석사학위논문, 금오공대, 1999.

[2] 장도수; "장치공정에서 설비 주변 환경관리에 관한 연구", 박사학위논문, 금오공대, 2006.

[3] 조진형 등; "A Diagnostic Method of Control-in/out in the Glass Furnace", 산업경영시스템학회지, 30(1), 2007.

[4] Box, George E. P., Jenkins, Gwilym M., and Reinsel, Gregory C.; Time Series Analysis Forecasting and Control, Prentice Hall, 1994.

[5] Box, George E. and Luceno, Alberto; Statistical Control by Monitoring and Feedback Adjustment, 1st edition, John Wiley and Sons, Inc. New York, NY, USA, 1997.

[6] Chang, Ih, Tiao, George C., and Chen, Chung; "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30(2), 1988.

[7] Alwan, Layth C. and Roberts, Harry V.; "Time-Series Modeling for Statistical Process Control," *Journal of Business and Economic Statistics*, 6(1), 1988.

[8] Atienza, O. O., Tang, L. C., and Ang, B. W.; "A SPC Procedure for Detecting Level Shifts of Autocorrelated Processes," *Journal of Quality Technology*, 30(4), 1998.

[9] Cho, Jin-Hyung, Chang, Sung-Ho, Lee, Sae-Jae, Jang, Do-Soo, Suh, Jung-Yul, and Oh, Hyun-Seung; "Using Ambient Control to Prevent External Disturbances in Large-scale Furnace," *Journal of Society of Korea Industrial and Systems Engineering*, 29(2), 2006.

[10] Hogan, Jeffrey A. and Lakey, Joseph D.; "Time-Frequency and Time-Scale Methods : Adaptive Decompositions, Uncertainty Principles, and Sampling," Springer Verlag, 2004.

[11] Percival, Donald B. and Walden, Andrew T.; Wavelet Methods for Time Series Analysis, Cambridge University Press, 2006.

[12] Walker, James S.; Fast Fourier Transforms, 2nd Ed., CRC Press, 1996.

[13] Walnut, David F.; An Introduction to Wavelet Analysis, Springer Verlag, 2001.

Glossary of Keywords

(a) Time-scale : specification of divisions (scale) of time or rhythm.

(b) PL-ARMA : ARMA model with periodic and linear components superimposed.

(c) MA-approximation : approximation of ARMA model with MA-only model : analogous to Taylor expansion of a function in its general approach.

(d) Nested periods : an ordered list of periods in which a period is an integer multiple of another one right before it.

(e) PWF : polynomial weight filter, which produces weighted average of data, in which each weight is specified in a polynomial form.

(f) Near-independent lag : time lag for which autocorrelation is nearly zero. 'Nearly zero' is determined by a preset threshold.