Scientific Research

Average Power Function of Noise and Its Applications in Seasonal Time Series Modeling and Forecasting

Qiang Song

RedPrairie Corporation, Alpharetta, USA
E-mail: *qiang.song@redprairie.com, qsong3@hotmail.com*

Abstract

This paper presents a new method of detecting multi-periodicities in a seasonal time series. Conventional methods such as the average power spectrum or the autocorrelation function plot have been used in detecting multiple periodicities. However, there are numerous cases where those methods either fail, or lead to incorrectly detected periods. This, in turn in applications, produces improper models and results in larger forecasting errors. There is a strong need for a new approach to detecting multi-periodicities. This paper tends to fill this gap by proposing a new method which relies on a mathematical instrument, called the ***Average Power Function of Noise*** (**APFN**) of a time series. APFN has a prominent property that it has a strict local minimum at each period of the time series. This characteristic helps one in detecting periods in time series. Unlike the power spectrum method where it is assumed that the time series is composed of sinusoidal functions of different frequencies, in APFN it is assumed that the time series is periodic, the unique and a much weaker assumption. Therefore, this new instrument is expected to be more powerful in multi-periodicity detection than both the autocorrelation function plot and the average power spectrum. Properties of APFN and applications of the new method in periodicity detection and in forecasting are presented.

Keywords: Seasonal Time Series, Forecasting, Seasonality Detection, Average Power Function of Noise,
              Average Power Spectrum, Autocorrelation Functions

1. Introduction

Recently, modeling and forecasting seasonal time series with multiple periodicities has regained attentions from researchers [1-8]. There is an abundance of literature on modeling and forecasting time series with a single period [9,10]. However, with multiple periodicities existing in a seasonal time series, modeling and forecasting such a time series becomes quite complicated. An important task in modeling such time series is the determination of the cycle length of each period. In the literature of modeling and forecasting seasonal time series with multi-periodicities, the lengths of periods are determined by human beings, based on either their experience or their specific domain knowledge of the time series under study. For example, in determining the cycle lengths of periods in modeling and forecasting daily electricity load [7] two distinct periods are determined manually based on power spectrum and autocorrelation plots, one being the intraday period which corresponds to 24 hours, and

the other being the intraweek period which corresponds to 168 hours. When the plurality of time series to work with is small, it may not be a significant problem for human beings to determine the length of each period. Nevertheless, it will become a forbidding task if the number of time series is in the range of tens of thousands, not to mention the possibility that the periodicities may change over time. In this case, one has to resort to automated algorithms to detect the periodicities in a time series.

Traditionally, the average power spectrum and the autocorrelation functions have been used in detecting multiple periods [9-11]. Usually, if the number of periods M in a seasonal time series is pre-determined, then the periods corresponding to the largest N peaks of the average power spectrum will be the periods in the time series [11, p. 161]. With ACF plot, one looks at the repeating peaks and decides proper periods [9, p. 342]. However, there is a certain degree of subjectivity with those two approaches. This is reflected in applications

where not all the *M* largest peaks of the average power spectrum are used in estimating the periods and not all the repeating peaks on the ACF plot are used as the periods [5], and there could be conflicting outcomes from those two approaches. For instance in [5], two periods, 224 and 32, are determined for the five Central Telephony series. Period 224 corresponds to the largest peak on the power spectrum plot while the peak at lag 224 on the ACF plot is even not significant at the confidence level of 0.95. Period 32 corresponds to a very small and even ignorable peak on the power spectrum; however, it concurs with a significant peak (at the confidence level 0.95) at lag 32 on the ACF plot (see Figure 4 and Figure 5 of [5] for details). It couldn't be an easy job to manually choose 224 and 32 as the periods before actually running the forecast. This apparent contradiction lies in the subjectivity in detecting periods using the average power spectrum and the ACF plots, and for this reason it requires human intervention, modeling experience, and even domain knowledge.

For the past decades, the average power spectrum has been utilized as a powerful apparatus in analyzing the distribution of the energy of a signal over different frequencies. It helps detect frequency components in a signal. Undoubtedly, this tool is extremely helpful in signal analysis, communications, circuitry design, and many other areas. Nevertheless, the average power spectrum may not be as effective as one would expect in determining periods in seasonal time series and forecasting simply because the frequencies or the periods with the strongest energy may not be the right periods needed in forecasting. **Figure 11** in this paper presents a time series of the daily sales at a retail store for over one year period of time. This time series exhibits 3 major different periods: 7, 28 and 364 due to the operational business cycles. On the sample ACF plot in **Figure 12**, there are significant peaks at lag 7 and at lag 28, and both are significant at the significance level of 0.95. However, the peak at lag 364 is not significant at the significance level of 0.95. At lag 28, the ACF plot has the largest peak. At lag 7, the ACF plot has the second largest peak. On the sample power spectrum in **Figure 13**, period 7 has the largest peak, and period 31 has the second largest peak. Now, let us suppose we need to determine the periods of this time series, and we pick the highest peaks on either the ACF plot or the power spectrum which is common practice in the literature. Here is the dilemma with the power spectrum and the ACF methods: If only two periods are chosen, those two methods would generate different results.

Seasonal time series repeats itself after a certain time lag [9], in spite that due to the noise in the time series the degree to which the time series repeats varies. *The right period should be the one at which the time series repeats*

itself with the smallest discrepancy, *not necessarily the one repeating the most frequently but with a larger discrepancy*. The largest peak on the power spectrum is the period in which the time series repeats the most frequently. For example, in the illustrative time series in **Figure 11**, the sample power spectrum has the largest peak at period 7. This is because the business activity of the store repeats week after week, and this pattern repeats the most frequently. However, the state of the business of the store in different weeks of the same month could be quite different due to different patterns of customer spending behaviors. Thus, it is hard to say that the business repeats itself very well week after week. Looking at the time series plot, it seems that the business of the store repeats much better month after month than week after week. However, the sample power spectrum does not have the largest peak at period of 28. Instead, it has a peak at period 31. The sample ACF plot of this time series has the largest peak at lag 28; but, this is not consistent with the sample power spectrum. A deadlock has been encountered here in selecting a period for this seasonal time series based on the ACF plot and the power spectrum plot. Therefore, it is necessary to search for a new and different apparatus that does not depend on the frequencies in the time series, but *depends on the degree to which the time series repeats itself*. The goal of this paper is to present the construction and properties of such an instrument, namely, *Average Power Function of Noise*.

This paper is organized as follows. In Section 2, the definition and properties of Average Power Function of Noise are presented. Distribution of the local minimum of the sample APFN is discussed in Section 3, followed by a discussion of measure of seasonality in Section 4. Three illustration examples are presented in Section 5, and conclusions and discussions are given in Section 6.

2. The Average Power Function of Noise and Its Properties

In general, a seasonal time series exhibits periodic behavior with period *p*, or a seasonal time series repeats itself after *p* basic time units [9, p. 327]. However, due to the noise in the time series, a seasonal time series rarely repeats itself in a perfect fashion. There is always discrepancy when comparing the observations of a seasonal time series at two different times over exactly *p* time unit span. Mathematically, a seasonal time series can be described as $x(t+p) = x(t) + \varepsilon(t)$ where $\varepsilon(t)$ is the discrepancy and *p* is the period. Here, *p* should be understood as the minimum integer value that can be found and is greater than 1. *This discrepancy*, *when processed properly*, *could be an indicator of periodicity of the sea-*

sonal time series. This is because if the time span between two observations is indeed the period of the seasonal time series, the discrepancy should be smaller than if the time span is not the period of the seasonal time series. If the seasonal time series possesses multiple periods, then the discrepancies at different periods may be quite different. A smaller discrepancy indicates a closer and better similarity between the observations at different times with zero discrepancy being the perfect and ideal case. We will explore and extend this idea in this section.

Without loss of generality, we will work with stochastic processes and extend the obtained results to time series in the sequel. In literature, discrepancies can be in various forms, such as absolute deviations, square of errors, and even absolute relative errors, just to name a few. Therefore, it is quite likely to obtain certain results parallel to what will be reported in this paper, if different forms of discrepancies are adopted. In this paper, we choose the square of errors as the form of discrepancy simply because of its good analytical properties.

First, we present the definition of Average Power Function of Noise, the corner stone of this paper.

Definition 1 (Average Power Function of Noise): Let $x(t)$ be a stochastic process. Then, for any given real numbers τ, if the following limit exists, it will be called the **Average Power Function of Noise** (APFN) of $x(t)$ at lag τ:

$$\text{APFN}(\tau) = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(x(t+\tau)-x(t)\right)^2}{2T} \, dt \right] \quad (1)$$

Obviously, $\text{APFN}(\tau)$ is an even function of τ if $x(t)$ is periodic. First, we consider the case where $x(t)$ has a unique period p. Then, we have the following major result.

Theorem 1. Let $x(t)$ be a stochastic process and be expressed as $x(t) = a(t) + \varepsilon(t)$ where $a(t)$ is the mean process of $x(t)$, *i.e.*, $a(t) = E(x(t))$, and is a deterministic function of t, and $\varepsilon(t)$ is a white noise process, *i.e.*, $\varepsilon(t)$ follows $N(0, \sigma_\varepsilon^2)$. If $a(t)$ is a periodic function of t with a unique period p so that $a(t) = a(t+p)$, then $\text{APFN}(\tau)$ has a strict local minimum when $\tau = p$.

Proof: As $x(t+p) = a(t+p) + \varepsilon(t+p)$, if setting $\tau = p$ in (1), then (1) can be written as

$$\text{APFN}(p) = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(a(t+p)+\varepsilon(t+p)-a(t)-\varepsilon(t)\right)^2}{2T} \, dt \right] = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(\varepsilon(t+p)-\varepsilon(t)\right)^2}{2T} \, dt \right]$$

$$= \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\varepsilon^2(t+p)-2\varepsilon(t+p)\varepsilon(t)+\varepsilon^2(t)}{2T} \, dt \right]$$

Exchanging the order of integration and expectation above and noticing that $\varepsilon(t+p)$ and $\varepsilon(t)$ are independent, we obtain

$$\text{APFN}(p) = \lim_{T \to \infty} \left[\int_{-T}^{T} \frac{E\left[\varepsilon^2(t+p)-2\varepsilon(t+p)\varepsilon(t)+\varepsilon^2(t) \right]}{2T} \, dt \right] = \lim_{T \to \infty} \left[\int_{-T}^{T} \frac{2\sigma_\varepsilon^2}{2T} \, dt \right] = 2\sigma_\varepsilon^2 \quad (2)$$

When $\tau = p' \in N(p, \delta)$ where $\delta > 0$ and $N(p, \delta)$ is a neighborhood of p, we have

$$\text{APFN}(p') = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(x(t+p')-x(t)\right)^2}{2T} \, dt \right] = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(a(t+p')+\varepsilon(t+p')-a(t)-\varepsilon(t)\right)^2}{2T} \, dt \right]$$

$$= \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(a(t+p')-a(t)\right)^2 + 2\left(a(t+p')-a(t)\right)\left(\varepsilon(t+p')-\varepsilon(t)\right)+\left(\varepsilon(t+p')-\varepsilon(t)\right)^2}{2T} \, dt \right]$$

Exchanging the order of integration and mean calculation and noticing that $\varepsilon(t)$ is white noise and $a(t+p') - a(t) \neq 0$, we will get

$$\text{APFN}(p') = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(a(t+p')-a(t)\right)^2 + 2\sigma_\varepsilon^2}{2T} \, dt \right]$$

$$= \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(a(t+p')-a(t)\right)^2}{2T} \, dt \right] + 2\sigma_\varepsilon^2 > 2\sigma_\varepsilon^2$$

That is, we have

$$\text{APFN}(p') > \text{APFN}(p). \quad (3)$$

Therefore, $\text{APFN}(\tau)$ has a strict local minimum at p. This finishes the proof.

One of the forecasting methods, called naïve method, is to use the most recent history data as the forecast of the future. Here, if we use the history data exactly τ time units ago as the forecast, then $x(t+\tau) - x(t)$ can

be interpreted as the forecasting error of $x(t+\tau)$ using $x(t)$. Thus, (1) may be interpreted as the average power of forecasting errors using the naïve method. Theorem 1 implies that if a stochastic process is seasonal and if we use the right period in the naïve method to predict its future, the mean square function of the forecasting errors will have a local minimum. Intuitively, this is quite natural as for seasonal stochastic process, observations with the same periodicity are quite close to each other and therefore the differences are minimal. One of the merits of (1) is that it relates periods in a seasonal stochastic process to a type of forecasting errors in a natural way. This gives one an instrument helpful in detecting periods from multiple candidates: When detecting periods for a seasonal stochastic process, select those periods that correspond to the smallest APFN values in (1). **Figure 14** presents the plot of the sample APFN of the seasonal time series of daily sales at a retail store. It can be seen that the APFN plots have a local minimum at period 7, 28, and 364, and the values of APFN at those three periods are different.

With Theorem 1, it is easy to prove the following corollaries which assume that $x(t)$ has a unique period.

Corollary 1. $APFN(\tau)$ has the same periods as $x(t)$ if $x(t)$ itself is periodic.

Corollary 2. If p is a period of $x(t)$, then $APFN(\tau)$ has a strict local minimum at any integral multiples np where n is a nonzero integer. That is, if p is a period of $x(t)$, then the corresponding local minima will repeat with the same period.

Corollary 3. If $x(t)$ is a white noise, then $APFN(\tau)$ is a constant with respect to τ.

Corollary 4. Let $APFN^*$ be the global minimal value of the average power function of noise. Then, we have

$$APFN^* \geq 2\sigma_\varepsilon^2, \text{ or } \frac{APFN^*}{2} \geq \sigma_\varepsilon^2.$$

Corollary 4 implies that the global minimal value of APFN, when divided by 2, can be used as an upper bound of the variance of the noise in the stochastic process. This will be seen very useful in estimating the forecasting error variance.

A seasonal stochastic process may have multiple periods. Now, let us consider the potential sources that contribute to the error or discrepancy when forecasting a seasonal stochastic process, including processes with a unique period and multiple periods. Of course, the major source is the noise in the data. This noise will always produce error in forecasting. There is a second source, however, that is any improper time lags chosen in forecasting $x(t+\tau)$ when multiple periods exist. For different values of τ, the errors could be quite different. This is reflected in the APFN plot which exhibits multiple local minima with different values. We intend to interpret this phenomenon by means of the conditional variance of the noise defined as a deterministic periodic function of τ. With this said, we could have the following theorem.

Theorem 2. Let $x(t)$ be a stochastic process and be expressed as $x(t) = a(t) + \varepsilon(t)$ where $a(t)$ is the mean process of $x(t)$ and is a deterministic periodic function of t, and $\varepsilon(t)$ is a noise process which follows $N(0, \sigma_\varepsilon^2)$ and the conditional variance $\Phi(\tau) = E\left[\varepsilon^2(t+\tau)\big|\varepsilon(t)\right]$ is a deterministic periodic function of τ and $E\left[\varepsilon(t+p_i)\big|\varepsilon(t)\right] = 0$. If both $a(t)$ and $\Phi(\tau)$ have the same periods p_i so that $a(t) = a(t+p_i)$, $\Phi(\tau)$ has a local minimum when $\tau = p_i$ and $\Phi(\tau) = \Phi(\tau + p_i)$ where $i = 1, 2, \cdots, M$, then there exists a j such that $APFN(\tau)$ has a global minimum when $\tau = p_i$ where $1 \leq j \leq M$.

Proof: Let p_i be one period. It can be shown that

$$APFN(p_i) = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\varepsilon^2(t+p_i) - 2\varepsilon(t+p_i)\varepsilon(t) + \varepsilon^2(t)}{2T} dt\right]$$

$$= \lim_{T \to \infty} \int_{-T}^{T} \frac{E\left[\varepsilon^2(t+p_i) - 2\varepsilon(t+p_i)\varepsilon(t) + \varepsilon^2(t)\right]}{2T} dt$$

In the above, it is assumed that the orders of the integration and the expectation operations can be exchanged.

Conditioning the expectation on $\varepsilon(t)$, we obtain

$$APFN(p_i) = \lim_{T \to \infty} \int_{-T}^{T} \frac{E\left[E\left[\varepsilon^2(t+p_i) - 2\varepsilon(t+p_i)\varepsilon(t) + \varepsilon^2(t)\right]\big|\varepsilon(t)\right]}{2T} dt$$

$$= \lim_{T \to \infty} \int_{-T}^{T} \frac{E\left[E\left[\varepsilon^2(t+p_i)\big|\varepsilon(t)\right]\right] - 2E\left[\varepsilon(t)E\left[\varepsilon(t+p_i)\big|\varepsilon(t)\right]\right] + E\left[\varepsilon^2(t)\right]}{2T} dt$$

As $E\left[\varepsilon(t+\tau)^2\big|\varepsilon(t)\right]$ is a deterministic periodic function of τ and $E\left[\varepsilon(t+p_i)\big|\varepsilon(t)\right] = 0$, we have

$$\text{APFN}(p_i) = \lim_{T \to \infty} \int_{-T}^{T} \frac{E\left[E\left[\varepsilon^2(t+p_i)|\varepsilon(t)\right]\right] + \sigma_\varepsilon^2}{2T} dt$$

$$= \Phi(p_i) + \sigma_\varepsilon^2$$

When $\tau \neq p_i$,

$$\text{APFN}(\tau) = \lim_{T \to \infty} E\left[\int_{-T}^{T} \frac{\left(a(t+\tau)-a(t)\right)^2 + 2\left(a(t+\tau)-a(t)\right)\left(\varepsilon(t+\tau)-\varepsilon(t)\right) + \left(\varepsilon(t+\tau)-\varepsilon(t)\right)^2}{2T} dt\right]$$

$$= \lim_{T \to \infty} \int_{-T}^{T} \frac{E\left[E\left[\left(a(t+\tau)-a(t)\right)^2 + 2\left(a(t+\tau)-a(t)\right)\left(\varepsilon(t+\tau)-\varepsilon(t)\right) + \left(\varepsilon(t+\tau)-\varepsilon(t)\right)^2 \Big| \varepsilon(t)\right]\right]}{2T} dt$$

$$= \lim_{T \to \infty} \int_{-T}^{T} \frac{\left(a(t+\tau)-a(t)\right)^2 + 2E\left\{E\left[\left(a(t+\tau)-a(t)\right)\left(\varepsilon(t+\tau)-\varepsilon(t)\right)\big|\varepsilon(t)\right]\right\} + E\left\{E\left[\left(\varepsilon(t+\tau)-\varepsilon(t)\right)^2\big|\varepsilon(t)\right]\right\}}{2T} dt$$

$$> \lim_{T \to \infty} \int_{-T}^{T} \frac{E\left[E\left[\left(\varepsilon(t+p_i)-\varepsilon(t)\right)^2\big|\varepsilon(t)\right]\right]}{2T} dt = \text{APFN}(p_i)$$

for in the above $\lim_{T \to \infty} \int_{-T}^{T} \frac{\left(a(t+\tau)-a(t)\right)^2}{2T} dt > 0$,

$E\left(\varepsilon(t+\tau)-\varepsilon(t)\right) = 0$, and $E\left(\varepsilon(t)\right) = 0$

Therefore, $\text{APFN}(p_i)$ is a strict local minimum. Among the M local minima, there must be one that is the smallest, and that minimum must correspond to a certain period p_j. This finishes the proof.

Theorem 2 explains why the APFN could have multiple local minima with different values in the case where multiple periods exist: it is due to the conditional variance function of $\varepsilon(t)$ which has different values at different periods.

From Corollary 2, we can prove that if p_i is a period of $x(t)$ for $i = 1, 2, \cdots, M$, then $\text{APFN}(\tau)$ will have a local minimum at any multiples of p_i, *i.e.*, $\text{APFN}(np_i)$ is also a local minimum. Hence, we have the following conclusion.

Corollary 5. If $x(t)$ has multiple distinct periods $p_1, p_2, \cdots, p_m$, then $\text{APFN}(\tau)$ has a strict local minimum at each period np_i where $i = 1, 2, \cdots, M$.

From Theorem 2, we may conclude that not all periods have equal APFN values. The period that yields the global minimum APFN value is the most important one and is the one that must be found in modeling as it may produce the smallest forecasting error using the naïve method. For this reason, we give the following definition.

Definition 2. Suppose $x(t)$ is a stochastic process and can be expressed as $x(t) = a(t) + \varepsilon(t)$ where $a(t)$ is a deterministic periodic function of t and $\varepsilon(t)$ is a noise process which follows $N(0, \sigma_\varepsilon^2)$ and the conditional variance $E\left\{\varepsilon(t+\tau)^2|\varepsilon(t)\right\}$ is a deterministic periodic function of τ. Suppose also that both $a(t)$

and $E\left\{\varepsilon(t+\tau)^2|\varepsilon(t)\right\}$ have the same periods p_i where $i = 1, 2, \cdots, M$, then the period that yields the smallest APFN value is called the primary period while all others secondary periods.

Hence, if only one period is chosen for a seasonal stochastic process, it should be the primary period. However, one may not be able to find the primary period if using the power spectrum or ACF plot.

The APFN plot can also be used to identify if a deterministic trend exists in a process.

Corollary 6. If $x(t) = bt + \varepsilon(t)$ where b is a non-zero constant and $\varepsilon(t)$ is a white noise process, then $\text{APFN}(\tau)$ is an increasing function of $|\tau|$.

To apply APFN in detecting the cycle lengths of multiple periods for a seasonal stochastic process, we recommend the following outline which can be used as a basis in designing detailed algorithms.

Step 1. Determine the number of periods, *M,* to be detected for a given seasonal time series;

Step 2. Calculate APFN(*t*) using formula (4) to be given in the next section for different integer values of *t*;

Step 3. Detect all local minima of APFN(*t*) and find the corresponding values of *t*;

Step 4. Sort all the local minima from the least to the greatest, and rearrange the corresponding *t* values accordingly. Then, the *t* values of the first *M* local minima are the cycle lengths, and the first *t* value is the primary period.

Some comments are in order regarding the number of data points required in calculating the value of APFN(*t*). In applications, one uses the sample APFN function in detecting cycle lengths. Evidently, the more data used in calculation, the more reliable the results. Although it is not clear at this stage what the minimum number of data points should be needed in calculating APFN, our em-

pirical results suggest that it is reasonable to have at least 50 data points in calculating each value of APFN in order to get good results. Nevertheless, this is still an open problem that requires more research.

3. Distribution of Local Minima of the Sample Average Power Functions of Noise

In applications and especially when working with time series, we use the following form of definition of APFN:

$$\text{APFN}(p) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \left(x(t+p) - x(t) \right)^2 \quad (4)$$

As local minima of the *APFN* are the only interest to us, and at each local minimum $a(t+p) = a(t)$, denote $\delta_p(t) = \varepsilon(t+p) - \varepsilon(t)$. Then, the above Equation (4) can be written as

$$\text{APFN}(p) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \delta_p^2(t).$$

Now, we want to derive the distribution of APFN(p) for any finite values of *N*. As $\varepsilon(t)$ follows $N(0, \sigma_\varepsilon^2)$

by assumption, we know that $\delta_p(t) = \varepsilon(t+p) - \varepsilon(t)$ follows $N(0, 2\sigma_\varepsilon^2)$. Furthermore, we can infer that $\delta_p(t)$ is independent of $\delta_p(t+m)$ where $m \neq 0$, and $\delta_p^2(t)$ follows $\chi^2(1)$. It is easy to see that $E(\delta_p^2(t)) = 2\sigma_\varepsilon^2$. To find the variance of $\delta_p^2(t)$, notice that $\text{cov}(\varepsilon^2(t), \varepsilon^2(t+p)) = 0$,

$\text{cov}(\varepsilon^2(t), \varepsilon(t)\varepsilon(t+p)) = 0$, and

$\text{cov}(\varepsilon^2(t+p), \varepsilon(t)\varepsilon(t+p)) = 0$. Then, we would have,

$$\text{var}(\delta_p^2(t)) = \text{var}(\varepsilon^2(t+p) - 2\varepsilon(t)\varepsilon(t+p) + \varepsilon^2(t))$$
$$= \text{var}(\varepsilon^2(t+p)) + \text{var}(\varepsilon^2(t)) + 4\text{var}(\varepsilon(t)\varepsilon(t+p))$$

As

$$\text{var}(\varepsilon^2(t)) = E(\varepsilon^4(t)) - \left(E(\varepsilon^2(t))\right)^2$$
$$= E(\varepsilon^4(t)) - \sigma_\varepsilon^4 \quad \text{as } E(\varepsilon(t)) = 0$$
$$= 3\sigma_\varepsilon^4 - \sigma_\varepsilon^4 \quad \text{as } \varepsilon(t) \text{ follows } N(0, \sigma_\varepsilon^2)$$
$$= 2\sigma_\varepsilon^4 = \text{var}(\varepsilon^2(t+p)).$$

$$\text{var}(\varepsilon(t+p)\varepsilon(t)) = E(\varepsilon^2(t+p)\varepsilon^2(t)) - \left(E(\varepsilon(t+p)\varepsilon(t))\right)^2$$
$$= E(\varepsilon^2(t+p)\varepsilon^2(t)) \quad \text{as } E(\varepsilon(t+p)\varepsilon(t)) = 0$$
$$= E\left\{ E(\varepsilon^2(t+p)\varepsilon^2(t)|\varepsilon(t)) \right\} = E\left\{ \varepsilon^2(t) E(\varepsilon^2(t+p)|\varepsilon(t)) \right\}$$
$$= E(\varepsilon^2(t)) E(\varepsilon^2(t+p)) \quad \text{as } \varepsilon(t+p) \text{ and } \varepsilon(t) \text{ are independent}$$
$$= \sigma_\varepsilon^2 \sigma_\varepsilon^2 = \sigma_\varepsilon^4.$$

Hence, $\text{var}(\delta_p^2(t)) = 8\sigma_\varepsilon^4$.

Thus, we have proved the following theorem.

Theorem 3: Under the assumptions in Theorem 1, for any finite values of $N > 0$, any local minimum APFN(p) follows $\chi^2(N)$ with degree of freedom N, mean $2\sigma_\varepsilon^2$ and variance $\frac{8}{\sqrt{N}}\sigma_\varepsilon^4$ where p is the unique period of the time series and N is the sample size of the time series.

When multiple distinct periods exist, the distributions of local minima can be treated similarly, but they are more involved, and for this reason will be studied later.

4. Measure of Seasonality of Seasonal Time Series

It is not uncommon to hear people ask such a question: "How seasonal is a time series?" This question seems not easy to answer. Some time series is obviously very seasonal whereas others are not so. To answer this question, one needs a number from 0 to 1 as the measure of how seasonal a time series is. This measure must be very low

(close to zero) if a time series is almost a white noise and must be very high (close to 1) if a time series is generated by a sinusoidal process. We intend to provide an approach to measuring the seasonality of a time series by means of APFN.

Suppose a seasonal time series has a unique period *P* (>1). Then, this time series will repeat after any periods *K* which is a multiple of *P*, *i.e.*, $K = mP$ where *m* is a positive integer. On the APFN plot, a local minimum would appear at each lag of *K*. The lags of all the local minima are multiples of *P*. In other word, 100% of the lags of all local minima can be expressed as a multiple of *P*. Due to noise in the data, or due to a different period in the time series, lags of the local minima of the *APFN* may not be all multiples of *P*. Some lags of the local minima may be multiples of a period different from *P*. In this case, less than 100% of the lags of local minima are multiples of *P*. Such a seasonal time series is less seasonal than the one with 100% of lags of local minima being multiples of *P*. When a time series is pseudo-white noise, a much smaller portion of the lags of local minima

can be expressed as a multiple of *P*. Thus, the percentage of local minima that can be expressed as a multiple of a certain positive integer *P* (>1) can be used as a measure or indicator of how seasonal a time series could be. Based on this argument, we propose the following definition of measure of seasonality.

Definition (Measure of Seasonality) Let $\{x_t\}$ be a time series and assume P_i are co-prime positive integers greater than 1 where $i = 1, 2, \cdots, M$. Denote per(P_i) as the percentage of local minima on *APFN* plot which can be expressed as multiples of P_i. Then, max{per(P_i) for $i = 1, 2, \cdots, M$ } is defined as the measure of seasonality of $\{x_t\}$.

With the above definition, the seasonality measure of the illustrative time series in **Figure 11** is 0.60.

Comments are in order to interpret measure of seasonality: If this measure is 100%, one knows that the periods of the time series has a common factor, *P*. Then, if *P* is used in a seasonal model, errors due to the inaccuracy in *P* can be ignored. However, if this measure is less than 100%, say, 60%, then 40% of the local minima of the *APFN* do not share the same common factor as the rest 60% of the local minima. This indicates that the period of the time series changes over time. If *P* is the period shared by most of the data, it could happen that when *P* is used in modeling the part of the data that have a different period, then large modeling errors could be produced simply because the period *P* is not properly used.

5. Applications

In this section, we provide numerical examples to illustrate how to use the Average Power Function of Noise in determining cycle lengths of different periods in seasonal time series forecasting. For this purpose, we will work with three time series, and apply the sample power spectrum, the sample ACF, and the APFN to identify the periods. Then, for each of the three methods, we implement the identified periods in different models to produce forecast in order to compare different forecasting models. The algorithms used in detecting periods are as follows: To utilize sample ACF plot, the lags corresponding to the highest peaks are used as the periods; to utilize the power spectrum plot, the periods corresponding to the highest peaks are treated as the periods; to utilize the APFN, the discrete form of the definition (4) is used in calculation, and the periods corresponding to the lowest values are used as the periods. For simplicity and to be consistent to literature [5-7], only two periods are used in modeling and forecasting.

5.1. Example 1

In this example, we generate a time series of 251 data

points with this formula:

$$x(t) = 500\{\sin(0.4t) + \sin(0.9t) + \sin(0.1t) + \varepsilon(t)\} + 2000$$

where $t = 1, 2, 3, \cdots, 251$, and $\varepsilon(t)$ is a random number sampled from uniform distribution U(0,1). **Figure 1** illustrates this time series. **Figure 2** illustrates the sample ACF, and **Figure 3** illustrates the power spectrum of this time series. Out of the 251 data points, the first 218 data points are used in modeling while the last 33 data points are used in calculating forecast error and are therefore not used in modeling. The sample ACF plot indicates 4 major peaks at lags of 63, 14, 49 and 77. All the peaks are significant at the significance level of 0.95. On the power spectrum plot, three major peaks are seen at periods of 64, 7, and 16. **Figure 4** illustrates the APFN plot which exhibits 4 major local minima at periods of 63, 126, 189, and 140. For reasons mentioned above, we choose the first two major periods in modeling and forecasting. To model the time series, an additive seasonal autoregressive model with 2 different periods is selected while the order of the model is determined by an algorithm to ensure the modeling error is white noise. The additive seasonal autoregressive model is given by this equation [12]

$$x_t = \sum_{i=1}^{M} \left(a_i x_{t-i} + b_i x_{t-i\,p^1} + c_i x_{t-i\,p^2} \right)$$

where p^1, and p^2 are the two different periods in the model. Once the periods are determined, the order of each model is determined by an automated algorithm. A Seasonal AR(1) model is identified with the periods determined by both the ACF and the APFN methods while a Seasonal AR(4) model is identified with the periods detected by the Power Spectrum method. **Table 1** lists the models with periods detected using different methods and model parameters. Numbers within the parentheses are the standard errors of the corresponding parameters.
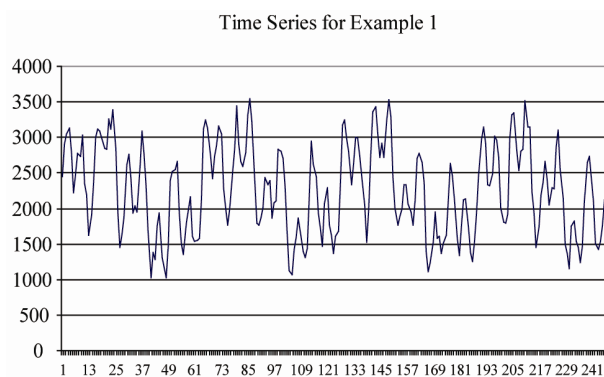

Figure 1. A seasonal time series exhibiting multiple seasonalities.
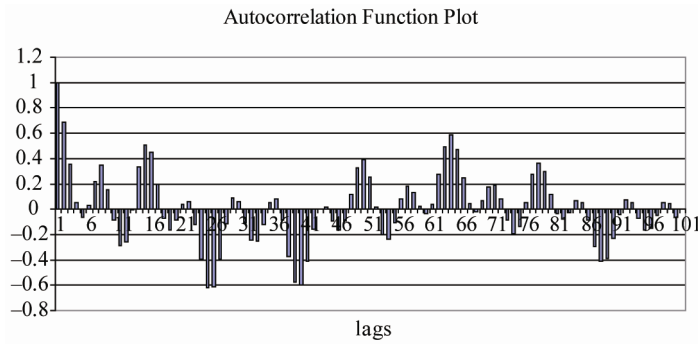
AJOR

Autocorrelation Function Plot

Figure 2. Autocorrelation function plot of Example 1.

Table 1. Models and the forecasting MAPE for Example 1.

Method Used	Model	MAPE
ACF	$x_t = \underset{(0.0066)}{0.065}x_{t-1} + \underset{(0.0066)}{0.854}x_{t-63} + \underset{(0.0068)}{0.0766}x_{t-14}$	8.49%
Power Spectrum	$x_t = \underset{(0.0305)}{0.7349}x_{t-1} + \underset{(0.0304)}{0.0077}x_{t-2} - \underset{(0.0304)}{0.3816}x_{t-3} + \underset{(0.0302)}{0.3215}x_{t-4} + \underset{(0.0299)}{0.5281}x_{t-64} - \underset{(0.0298)}{0.2353}x_{t-128}$ $- \underset{(0.0297)}{0.09279}x_{t-192} - \underset{(0.0297)}{0.0802}x_{t-256} + \underset{(0.0303)}{0.3699}x_{t-7} - \underset{(0.0305)}{0.3250}x_{t-14} - \underset{(0.0305)}{0.3647}x_{t-21} + \underset{(0.0306)}{0.3528}x_{t-28}$	54.4.1%
APFN	$x_t = -\underset{(0.0060)}{0.0973}x_{t-1} + \underset{(0.0059)}{0.7494}x_{t-63} + \underset{(0.0059)}{0.3292}x_{t-126}$	8.30%

Power Spectrum Plot for Example 1

Figure 3. Power spectrum plot of Example 1.

Average Power Function of Noise for Example 1
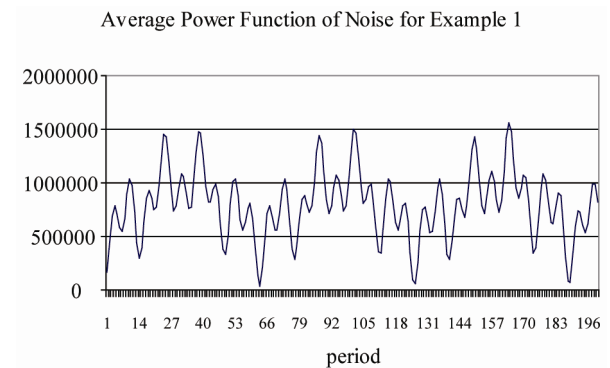
Figure 4. Sample average power function plot indicating multiple local minima.

Figure 5 illustrates the forecasts of different models. It can be seen that the model whose periods are identified using power spectrum fails to produce good forecasts. Although the two models whose periods are detected by ACF and by APFN produce about the same forecasting errors, the latter produces slightly better forecasts.

5.2. Example 2

In this example, 200 data points are generated using this formula: $x_t = \sin(12t) + \sin(400t)$ where $t = 1, 2, 3, \cdots, 200$. In modeling and analysis, the first 167 data points are used while the last 33 data points are used only in forecasting error calculation. **Figure 6** presents the time plot of this time series. **Figure 7** exhibits the sample ACF plot, **Figure 8** shows the empirical power spectrum plot and **Figure 9** represents the APFN plot. From the ACF plot in **Figure 7**, it can be seen that two significant peaks appear at lags of 12 and 21, both being significant at the significance level of 0.95. Hence, periods of 12 and 21 are detected with ACF plot for this time series; from the power spectrum plot in **Figure 8**, peaks are found at periods of 11, and 3. Therefore, periods of 11 and 3 are determined from the power spectrum plot. Finally, from the APFN plot in **Figure 9**, two local minima are seen at periods 77 and 56 which have the lowest values. Thus, with the APFN method the periods
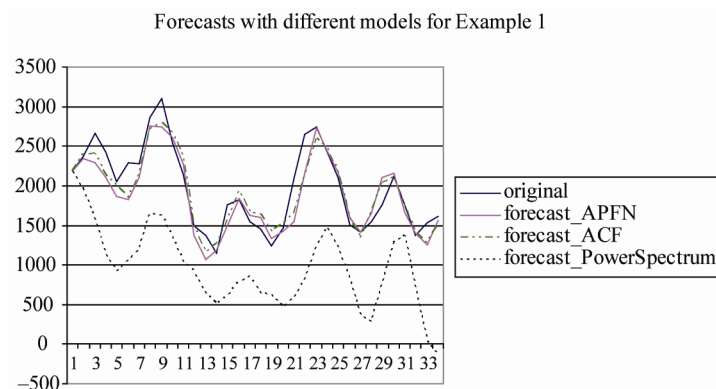
Forecasts with different models for Example 1

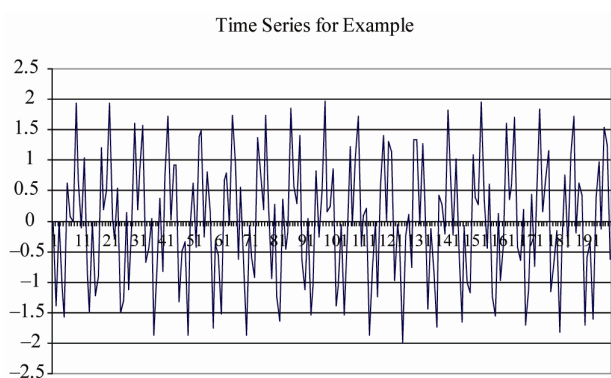**Figure 5. Forecasts using different models for Example 1.**

Time Series for Example

Figure 6. A seasonal time series with multiple seasonal patterns for Example 2.

Power Spectrum Plot for Example 2
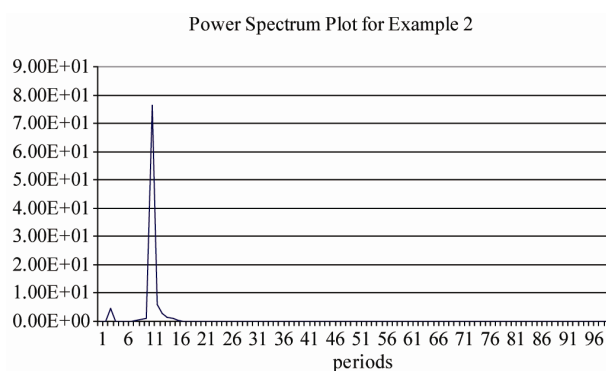
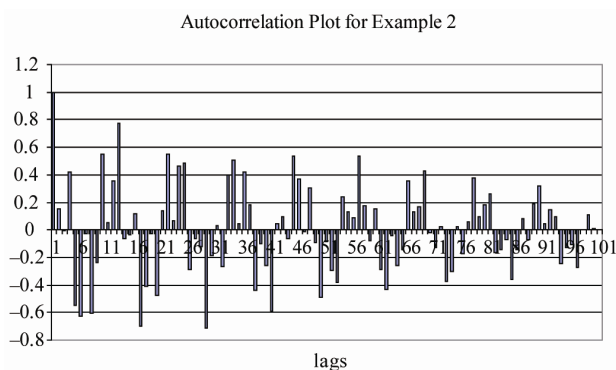**Figure 8. Power spectrum plot for Example 2.**

Autocorrelation Plot for Example 2

Figure 7. Autocorrelation function plot for time series in Example 2.

Average Power Function of Noise Plot
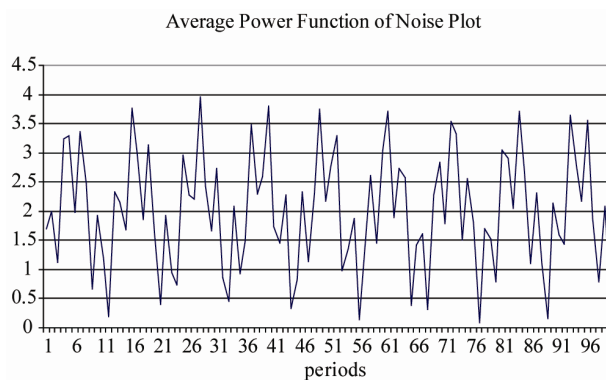
Figure 9. Plot of average power function of noise for Example 2.

are detected as 77 and 56 for this time series. Once the periods are determined, the order of a Seasonal AR model is determined by an automated algorithm. The result is that a Seasonal AR(2) model is recommended with the periods determined by all the three different methods. **Table 2** lists the models and parameters with standard errors of the parameters listed in the corresponding parentheses. For this example, all three models produce good forecasts whereas the model of the power

spectrum produces the smallest forecasting error and the ACF produces the largest forecasting error. **Figure 10** plots the forecasts of all the three models. The reason why all three methods produce good forecasts is that there is no noise introduced in the data.

5.3. Example 3

In this example, 466 retail daily sales data points are used. The first 433 data points are used in modeling and

Table 2. Models and parameters for Example 2.

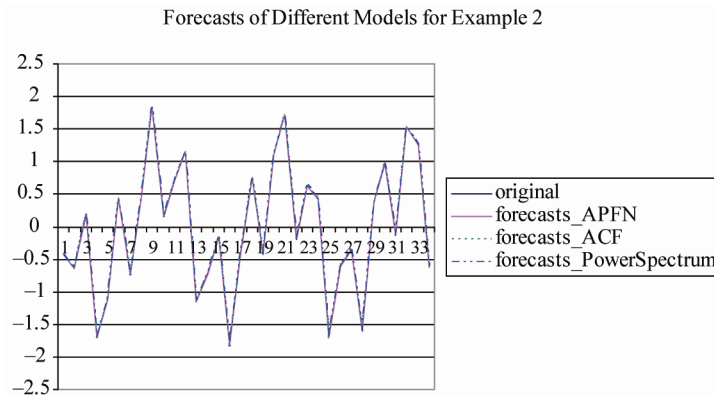| Method Used	Model	MAPE
ACF	$x_t = \underset{(0.00019)}{-0.0747}x_{t-1} + \underset{(0.00019)}{0.3651}x_{t-2} + \underset{(0.00019)}{0.5941}x_{t-12} - \underset{(0.00019)}{0.2645}x_{t-24} + \underset{(0.00019)}{0.5424}x_{t-21} + \underset{(0.00019)}{0.0896}x_{t-42}$	0.82%
Power Spectrum	$x_t = \underset{(0.00015)}{0.1220}x_{t-1} - \underset{(0.00015)}{0.2479}x_{t-2} + \underset{(0.00015)}{0.2346}x_{t-12} + \underset{(0.00015)}{0.6097}x_{t-24} + \underset{(0.00015)}{0.1859}x_{t-4} - \underset{(0.00015)}{0.40471}x_{t-8}$	0.646%
APFN	$x_t = \underset{(0.00032)}{-0.003}x_{t-1} - \underset{(0.00032)}{0.0337}x_{t-2} + \underset{(0.00032)}{0.5385}x_{t-77} - \underset{(0.00032)}{0.0257}x_{t-154} + \underset{(0.00032)}{0.493}x_{t-56} + \underset{(0.00032)}{0.1098}x_{t-112}$	0.652%

Figure 10. Forecasts obtained with different models in Example 2.

the last 33 data points are used in forecasting accuracy analysis. **Figure 11** shows the plot of the time series. **Figure 12** illustrates the sample ACF plot which exhibits two peaks at lags 7 and 28, both being significant at the significance level of 0.95. **Figure 13** shows the sample power spectrum and shows two peaks at periods 7 and 30, and **Figure 14** shows the Average Power Function of Noise plot of the time series which has two major local minima at periods of 364 and 336. Again, we pick 2 periods for the time series. An automated algorithm is used to determine the order and the parameters for each model.
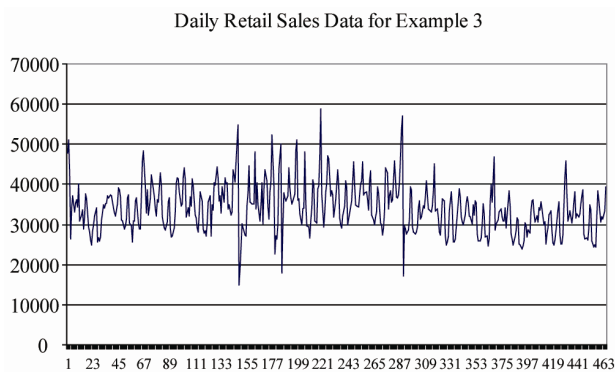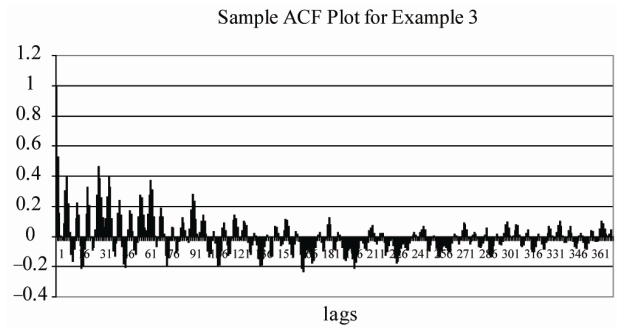

Figure 12. Sample Autocorrelation function plot for Example 3.

For the ACF model, a Seasonal AR(5) model is identified; for the Power Spectrum model, a Seasonal AR(6) model is identified, and for the APFN model, a Seasonal AR(1) model is identified. **Table 3** lists the models, the parameters and the standard errors, and the forecasting MAPE. As the model corresponding to the power spectrum method is too complicated, it is not listed in the table. It can be seen that the APFN model produces the best forecasts while the Power Spectrum model fails to produce meaningful forecasts. **Figure 15** illustrates the forecasts of all the 3 models. The power spectrum method
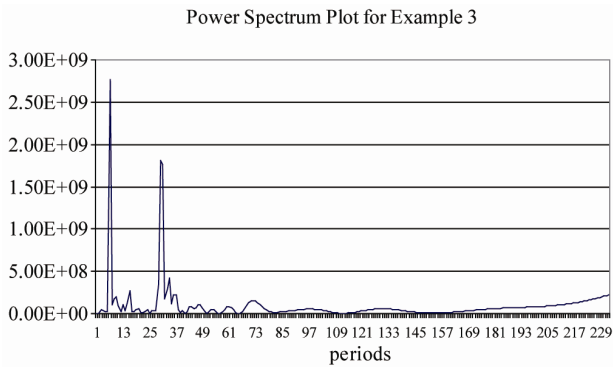

Figure 11. Daily retail sales plot for Example 3.

Power Spectrum Plot for Example 3

Figure 13. Plot of power spectrum for Example 3.

evidently detects an incorrect period, 30, and this incorrect period causes the modeling algorithm to produce a 6th order model to fit the data, and fails to produce good forecasts in the end with a MAPE of 33.08%. This illustrative example indicates the significance of using the right periods in modeling and forecasting seasonal time series.

6. Conclusions and Discussions

In this paper, the definition of Average Power Function

of Noise (APFN) has been proposed, and properties of such a function have been discussed. The most important property of APFN is that it has a local minimum at the time lags which are the periods of a seasonal time series. The numerical examples have exhibited the merits and the power of APFN in detecting periods in seasonal time series. ACF and power spectrum are proven to be powerful instruments in signal analysis and many other arenas. They may not be the best instruments in detecting periods for seasonal time series because, as pointed out
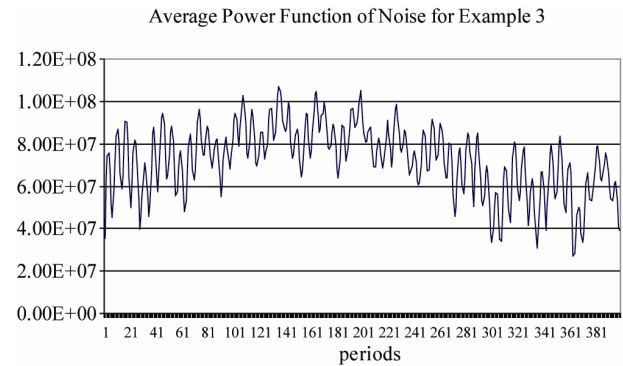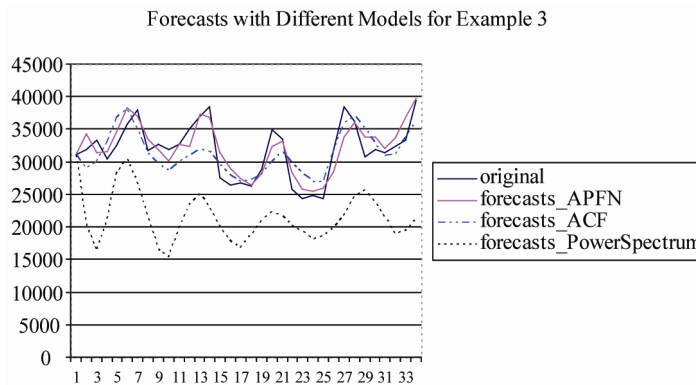
Average Power Function of Noise for Example 3

Figure 14. Plot of average power function of noise for Example 3.

Forecasts with Different Models for Example 3

Figure 15. Forecasts obtained with different models for Example 3.

Table 3. Models and parameters for Example 3.

Method Used	Model	MAPE
ACF	$x_t = 0.332x_{t-1} - 0.094x_{t-2} - 0.035x_{t-3} - 0.0302x_{t-4} + 0.108x_{t-5} + 0.234x_{t-7} - 0.065x_{t-14}$ $(0.0062)\ (0.0062)\ (0.0062)\ (0.0062)\ (0.0062)\ (0.0062)\ (0.0062)$ $- 0.0167x_{t-21} + 0.2589x_{t-28} - 0.029x_{t-35} + 0.072x_{t-56} + 0.139x_{t-84} + 0.065x_{t-112} + 0.059x_{t-140}$ $(0.0062)\ (0.00889)\ (0.0062)\ (0.0062)\ (0.0062)\ (0.0062)\ (0.0062)$	7.89%
Power Spectrum	Too complicated to list	33.08%
APFN	$x_t = 0.247x_{t-1} + 0.4304x_{t-364} + 0.2623x_{t-336}$ $(0.0111)\ (0.0103)\ (0.0101)$	5.36%

at the beginning of the paper, the detected periods of a seasonal time series should be the ones at which the time series repeats with the smallest discrepancy, not necessarily the ones at which it repeats the most frequently. The results of seasonal time series forecasting depend on not only the model but the periods of the time series as well. Forecasts are very sensitive to the periods detected and used in the model. If periods are not properly chosen, one may end up with high order models and inferior forecast results. From the empirical examples, it is quite evident that properly detected periods of a seasonal time series almost always accompany low order model and superior forecasts, and APFN consistently produces the lowest order models in the three illustrative examples. Therefore, the significance of properly detected periods cannot be overemphasized in modeling and forecasting seasonal time series.

As discussed at the beginning of this paper, ACF and power spectrum methods could lead to conflicting results in detecting periods for the same data set. For that reason, it will be difficult to automate the detection algorithm with ACF and power spectrum plots. However, it is relatively simple and easy to automate the period detection algorithm using APFN. Such an algorithm will involve selecting the first few lowest local minima, by sorting the local minima in an ascent order.

The APFN is in the form of square of discrepancy. Different forms of discrepancy are also possible and could be adopted. For example, the following two different forms can be considered in lieu of APFN:

$$\text{MAD}(\tau) = \lim_{T \to \infty} E \left\{ \int_{-T}^{T} \frac{|x(t) - x(t+\tau)|}{2T} \, dt \right\}$$

$$\text{MAPE}(\tau) = \lim_{T \to \infty} E \left\{ \int_{-T}^{T} \frac{\left| (x(t) - x(t+\tau)) / x(t+\tau) \right|}{2T} \, dt \right\}$$

However, it might be quite difficult to conduct theoretic analysis based on those two forms.

It must be admitted that the results presented in this paper are quite rudimentary. Further efforts should be made in order to have a better understanding of the properties of APFN. One of the issues around APFN is its connection to ACF and power spectrum. In other words, in the frequency domain, what do we know about APFN?

7. References

[1] P. Cortez, M. Rio, M. Rocha and P. Sousa, "Internet Traffic Forecasting Using Neural Networks," 2006 *International Joint Conference on Neural Networks*, Vancouver, 16-21 July 2006, pp. 2635-2642.

[2] A. M. De Livera and R. J. Hyndman, "Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing," Working Paper, Department of Econometrics and Business Statistics, Monash University, 2009.

[3] P. G. Gould, A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman and F. Vahid-Araghi, "Forecasting Time Series with Multiple Seasonal Patterns," *European Journal of Operational Research*, Vol. 191, No. 1, 2008, pp. 207-222. doi:10.1016/j.ejor.2007.08.024

[4] B. J. Morzuch and P. G. Allen, "Forecasting Hospital Emergency Department Arrivals," 26*th Annual Symposium on Forecasting*, Santander, June 11-14 2006.

[5] J. W. Taylor, "A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center," *Management Science*, Vol. 54, No. 2, 2008, pp. 253-265. doi:10.1287/mnsc.1070.0786

[6] J. W. Taylor, L. M. de Menezes and P. E. McSharry, "A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day Ahead," *International Journal of Forecasting*, Vol. 22, No. 1, 2006, pp. 1-16. doi:10.1016/j.ijforecast.2005.06.006

[7] J. W. Taylor, "Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing," *Journal of Operational Research Society*, Vol. 54, No. 8, 2003, pp. 799-805. doi:10.1057/palgrave.jors.2601589

[8] J. W. Taylor, "Triple Seasonal Methods for Short-Term Electricity Demand Forecasting," *European Journal of Operational Research*, Vol. 204, No. 1, 2010, pp. 139-152. doi:10.1016/j.ejor.2009.10.003

[9] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, "Time Series Analysis," 3rd Edition, Prentice Hall, New Jersey, 1994.

[10] D. Pena, G. C. Tiao and R. S. Tsay, "A Course in Time Series Analysis," John Wiley & Sons, Hoboken, 2001.

[11] P. Stoica and R. Moses, "Spectral Analysis of Signals," Pearson Prentice Hall, New Jersey, 2005.

[12] Q. Song and A. O. Esogbue, "A New Algorithm for Automated Modeling of Seasonal Time Series Using Box-Jenkins Techniques," *International Journal of Industrial Engineering and Management Systems*, Vol. 7, No.1, 2008, pp. 9-22.