Author's Final Version (10 December 2023).

Why Are the ARIMA and SARIMA not Sufficient?

Shixiong Wang[a], Chongshou Li[b,c], Andrew Lim[d]

[a]Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 117576
[b]SWJTU-Leeds Joint School, Southwest Jiaotong University, Chengdu 611756, China
[c]School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
[d]Red Jasper Holdings, Singapore 128424

Abstract

The autoregressive moving average (ARMA) model takes a significant position in time series analysis for a wide-sense stationary time series. The difference operator and seasonal difference operator, which are bases of ARIMA and SARIMA (Seasonal ARIMA), respectively, were introduced to remove the trend and seasonal component(s) so that the original non-stationary time series can be transformed into a wide-sense stationary one, which can then be handled by the Box-Jenkins methodology. However, such difference operators are more practical experiences than exact theories by now. In this paper, we investigate the power of the (seasonal) difference operator from the perspective of spectral analysis, linear system theory, and digital filtering, and point out the characteristics and limitations of the (seasonal) difference operator. Besides, a general method that transforms a non-stationary (the non-stationarity in the mean sense) stochastic process to be wide-sense stationary is presented. Intensive experiments validate the claims in this paper, and all the source data and codes are available online at GitHub: https://github.com/Spratm-Asleaf/ARIMA.

Keywords: Time Series Analysis, Difference Operator, Spectral Analysis, Digital Filtering, Linear System.

1. Introduction

One of the intriguing topics of time series analysis is to physically analyse the internal mechanism/dynamics of a system generating the focused time series, and subsequently build a proper mathematical model to describe the dynamics of this system so that we can predict the future with satisfying accuracy. Generally, such a time series is a stochastic process rather than a deterministic one which makes the problem more complex.

When it comes to the stochastic process modeling, the reputed Box-Jenkins methodology [1], also known as ARMA and ARIMA model, stands out. The philosophy of the ARMA model is from the Wold's Decomposition theorem [9, p. 517]. The theorem supports that the ARMA model is mathematically sufficient to describe a **regular** wide-sense stationary (WSS) stochastic process. After modeling, the least square method and maximum likelihood method can be utilised to estimate the parameters of the model based on the collected time series samples. As a result, we can make use of the past information (i.e., collected samples) to reconstruct the underlying dynamics of the focused stochastic process, and further make satisfying prediction. As complements to ARMA, the ARIMA (resp. SARIMA) aims to transform the focused non-stationary (in the mean sense) stochastic process to be stationary by difference (resp. seasonal difference) operator with proper orders so that the resulted time series can be fed into an ARMA model. Therefore, in Box-Jenkins methodology, the predictable component of a WSS process in one realization is treated as the trend of the regular component of the same WSS process which can be eliminated by the difference or seasonal difference operator. For notation breifness, we collectively refer to ARIMA and SARIMA as S-ARIMA.

Email addresses: `s.wang@u.nus.edu` (Shixiong Wang), `lics@swjtu.edu.cn` (Chongshou Li), `i@limandrew.org` (Andrew Lim)

However, such difference operators are more empirical experiences than exact theories. Therefore, we do not know why they work well somewhere and ineffectively elsewhere. To this end, in this paper, we aim to investigate the power of the (seasonal) difference operator from the perspective of spectral analysis, linear system theory, and digital filtering, and point out the characteristics and limitations that (seasonal) difference operator and S-ARIMA hold. Besides, a general operator that works for transforming a non-stationary (in the mean sense) stochastic process to be wide-sense stationary will be presented. At last, we will show that the overall methodology for predicting a non-stationary stochastic process, which is the generalization of S-ARIMA and termed as ARMA-SIN. For preliminaries on linear system theory, spectral analysis, and digital filtering, refer to [4, 10, 3].

Section 2 defines some notations. In Section 3, we review some preliminaries on stochastic process and explain the rationale behind the Box-Jenkins methodology (i.e., the S-ARIMA method). In Section 4, some examples are given to show the insufficiencies of the S-ARIMA and the advantages of ARMA-SIN over S-ARIMA, just as intuitive understandings for readers. In Section 5, we explain the mathematical reasons why S-ARIMA makes sense and point out its theoretical insufficiency. Section 6 presents the general ARMA-SIN methodology, and Section 7 explains and derives the methods we presented in the warming-up Section 4. Finally, a real-world example is provided in Section 8 to demonstrate the value of the ARMA-SIN. Glossaries of interest in this paper are placed in Appendix. All the source data and codes generating experimental results in this paper are available online at GitHub: `https://github.com/Spratm-Asleaf/ARIMA`.

2. Notations

1. Let $\boldsymbol{v} = a : l : b$ define a vector with the lower bound a, upper bound b, and step length l. For example, $\boldsymbol{v} = 1 : 1 : 3$ implies $a = 1$, $b = 3$, and $l = 1$. As a result, $\boldsymbol{v} = [1, 2, 3]^T$.
2. Let the function $length(\boldsymbol{x})$ return the length of the vector $\boldsymbol{x}$; if $\boldsymbol{x} = [0.1, 0.2, 0.3]^T$, $length(\boldsymbol{x}) = 3$.
3. Let $\boldsymbol{t}$ denote the continuous time, and $\boldsymbol{n}$ the corresponding discrete time. Therefore, if $\boldsymbol{t} = 0 : 0.1 : 10$ (i.e., the time span is $10s$, and the sampling time is $T_s = 0.1s$), we will have $\boldsymbol{n} = \boldsymbol{t}/T_s = 0 : 1 : length(t) - 1 = 0 : 1 : 100$; let $N = length(\boldsymbol{n})$.
4. Let $\boldsymbol{x}$ denote a time series of interest.
5. Let $randn(N)$ return a Gaussian white series (that has zero-mean and unit-variance) with length of N.
6. Let $ARMA(p, q|\boldsymbol{\varphi}, \boldsymbol{\theta})$ define an ARMA process with autoregressive order of p and moving average order of q. Besides, the coefficient vectors $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ are for the autoregressive part and moving average part, respectively.
7. Let the operator $\boldsymbol{y} = H(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{b})$ define a difference equation (a.k.a., a linear system) as follows

$$a_0 y_k + a_1 y_{k-1} + a_2 y_{k-2} + ... + a_p y_{k-p} = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2} + ... + b_q x_{k-q}, \tag{1}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$ are vectors and $length(\boldsymbol{a}) = p + 1$, $length(\boldsymbol{b}) = q + 1$.

3. Rationale Behind the S-ARIMA Method

Definition 1 (WSS Stochastic Process [9]). A real-valued stochastic process $x(t)$ is WSS if it satisfies:

- **Invariant Mean**: $E\{x(t)\} = \eta$, where η is a constant;
- **Invariant Autocorrelation**: $E\{x(t_1)x(t_2)\} = E\{x(t_1 + \tau)x(t_1)\} = R(\tau)$, meaning it only depends on the time difference $\tau := t_2 - t_1$, irrelevant to t_1 and t_2.

Invariant autocorrelation immediately admits the **Invariant Variance**, since $E\{x(t)\}^2 = R(0)$.

Theorem 1 (Wold's Decomposition Theorem [9]). *Any WSS stochastic process $x(n)$ can be decomposed into two sub-processes: (a) regular process, and (b) predictable process. Namely*

$$x(n) = x_r(n) + x_p(n), \tag{2}$$

where $x_r(n)$ is a regular process and $x_p(n)$ is a predictable process. Furthermore, the two processes are orthogonal (implying uncorrelated): $E\{x_r(n+\tau)x_p(n)\} = 0, \forall \tau \in \boldsymbol{n}$.

The detailed concepts of regular process (a.k.a. rational-spectra process from the spectral analysis perspective), and predictable process (a.k.a. line-spectra process) can be found in [9]. Intuitively, a regular process is mathematically as $x_r(n) = ARMA(p,q)$, and a predictable process is as $x_p(n) = \sum_{i=0}^{m} c_i e^{jw_i n}$ for some non-negative integer m where w_i are non-random discrete frequencies[1] with $w_0 = 0$; $Ec_i = 0$, $Ec_i c_j^* = 0$ when $i \neq j$, and $Ec_i^2 = \alpha_i > 0$. Note that the random coefficients c_i of the complex-valued exponentials $e^{jw_i n}$ are irrelevant to the time $\boldsymbol{n}$. It means that c_i are determined prior to $n = 0$ [i.e., c_i do not vary over time but they change in every different realization of $x(n)$]. Therefore, in one realization of $x(n)$, the predictable component $x_p(n)$ can be treated as a non-zero-mean but deterministic trend [which can be eliminated by the (seasonal) difference operator] of the regular component $x_r(n)$. This explains why the S-ARIMA method is rational for any WSS stochastic process: if this WSS stochastic process is regular, we directly use the ARMA model to fit; if it contains a predictable component, we alternatively use the S-ARIMA. In the spectral domain, the power spectra (i.e., squares of Fourier transform [9, Eq. (9-207)]) of a regular process is continuous and does not contain outliers, while the power spectra of a predictable process only contains impulses/outliers [9, pp. 517-518]. Therefore, the essence of stationarizing a time series is to remove large-valued outliers in its power spectra (or equivalently in its Fourier transform). In other words, a natural qualitative stationarity test criterion is whether the spectra contain outlier(s).

4. Scenarios of Warming-up

As intuitive understandings for readers, we in this section provide some simulation scenarios to illustrate the theoretical insufficiency that S-ARIMA holds. Together with, the counterpart solutions given by ARMA-SIN will be also demonstrated. Following this warming-up, in the subsequent sections, we will progressively detail the motivations, philosophies, mathematics, and methodologies behind the ARMA-SIN.

4.1. ARMA Series of Ground Truth

We generate a ARMA series for analysis later. Without loss of generality, we arbitrarily set

$$\begin{aligned} \boldsymbol{\theta} &= [13, 5, 6]^T \\ \boldsymbol{\varphi} &= [40, 2, 3, 6, 9]^T, \end{aligned} \tag{3}$$

meaning the time series generated from a Gaussian white series ϵ_k is given as

$$40x_k^0 + 2x_{k-1}^0 + 3x_{k-2}^0 + 6x_{k-3}^0 + 9x_{k-4}^0 = 13\epsilon_k + 5\epsilon_{k-1} + 6\epsilon_{k-2}, \tag{4}$$

where k denotes the discrete time index (namely $k \in \boldsymbol{n}$) and $x_k^0 := 0, \epsilon_k := 0$ if $k < 0$. However, this discrete linear system is guaranteed to be minimum-phase stable [4]. That is, it is stable and inversely stable.

Therefore, if the focused raw time series is as $\boldsymbol{x} = f(\boldsymbol{x}^0)$ (for example $\boldsymbol{x} = \boldsymbol{x}^0 + \boldsymbol{n}$, i.e., linear trend), the operator that exactly transforms $\boldsymbol{x}$ to its wide-sense stationary component $\boldsymbol{x}^0$ should be the best. This is because we finally aim to use the ARMA model to fit the transformed series. Let $\hat{\boldsymbol{x}}^0$ be the transformed series from $\boldsymbol{x}$. The nearer between $\hat{\boldsymbol{x}}^0$ and $\boldsymbol{x}^0$, the better.

[1] For the concept of discrete frequency, see [4].

4.2. The Case of Variant Mean

In this subsection, we investigate a scenario being with variant mean, implying it is non-stationary in the mean sense. Let $\boldsymbol{t} = 0 : 0.5 : 100$ (viz., $T_s = 0.5$), $\boldsymbol{n} = \boldsymbol{t}/T_s$, and $\boldsymbol{x} = \boldsymbol{x}^0 + 0.1\boldsymbol{t}$. It means the trend component is a linear function.

If we follow the standard modeling procedure with Box-Jenkins (ARIMA) method [6, 2, 1, 5], we have the estimated ARIMA(4,1,2) model to handle this problem, meaning the operator used is the first order differencing. Instead, if we use ARMA-SIN method, we have

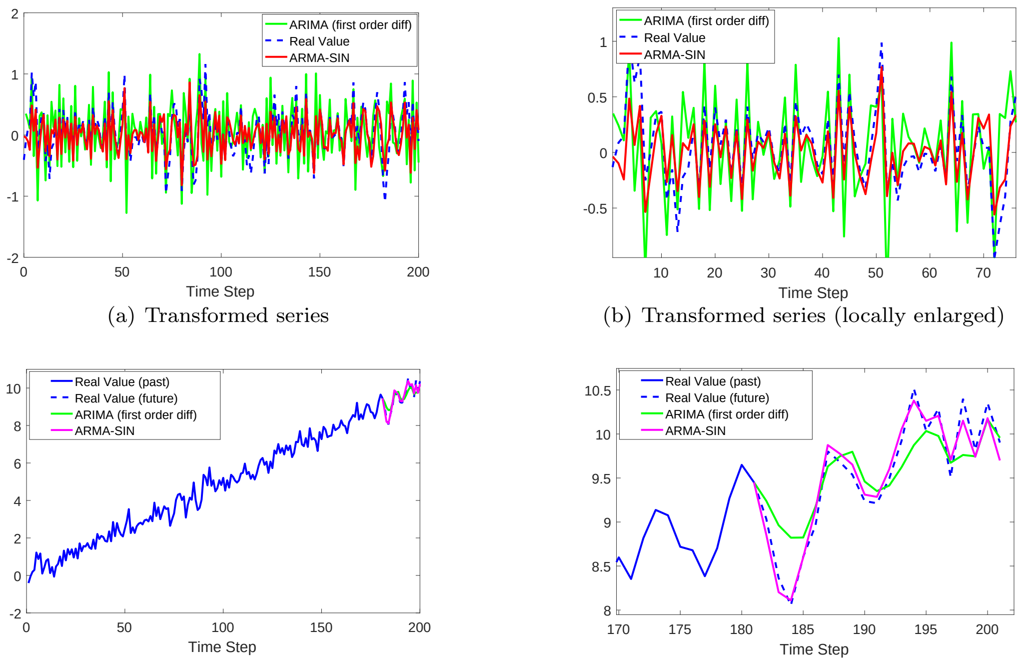
$$\hat{\boldsymbol{x}}^0 = H(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{b}), \tag{5}$$

where

$$\boldsymbol{a} = [1, -1.7101, 1.3712, -0.3152]^T,$$

$$\boldsymbol{b} = [0.6226, -1.5757, 1.5757, -0.6226]^T,$$

and the ARMA part to model $\boldsymbol{x}^0$ is ARMA(4,1). The results of two methods are shown in Figure 1.



(a) Transformed series

(b) Transformed series (locally enlarged)

(c) Prediction results

(d) Prediction results (locally enlarged)

Figure 1: Transformed series and prediction results for variant mean (diff: difference).

Besides, we have 100 times of monte carlo simulation and the averaged prediction MSE is given in Table 1. Clearly, the transformed series $\hat{\boldsymbol{x}}^0$ of ARMA-SIN is nearer to its ground truth $\boldsymbol{x}^0$ than that of ARIMA. Thus the prediction accuracy is more satisfactory.

Table 1: Averaged prediction MSE of ARIMA and ARMA-SIN for variant mean

	ARIMA	ARMA-SIN
MSE	0.0977	0.0203

4.3. The Case of Using SARIMA

In this subsection, we investigate a scenario that is suitable for using SARIMA. Let $\boldsymbol{t} = 0 : 0.1 : 50$ (viz., $T_s = 0.1$), $\boldsymbol{n} = \boldsymbol{t}/T_s$, and $\boldsymbol{x} = \boldsymbol{x}^0 + sin(5\boldsymbol{t})$. It means the trend component is a sine function.

If we follow the standard modeling procedure with Box-Jenkins (ARIMA) method, we have the estimated SARIMA(4,1,1)(12,1,0,0) model to handle this problem, meaning the operator used is the 12-lag seasonal differencing. Instead, if we use ARMA-SIN method, we have

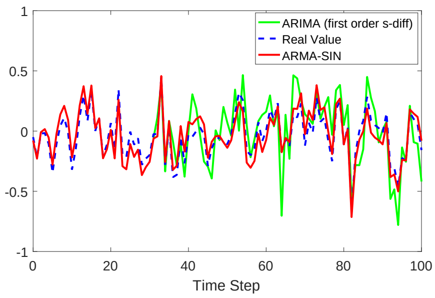
$$\hat{\boldsymbol{x}}^0 = H(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{b}), \tag{6}$$

where

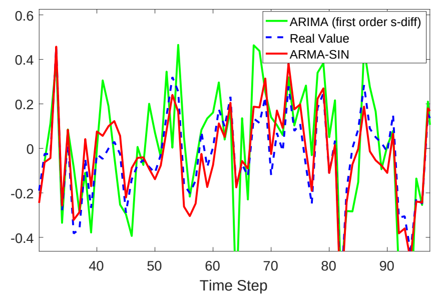
$$\boldsymbol{a} = [1, -5.1801, 11.8864, -15.30778, 11.6563, -4.9815, 0.9430]^T,$$

$$\boldsymbol{b} = [0.9713, -5.0804, 11.7716, -15.3086, 11.7716, -5.0804, 0.9713]^T,$$

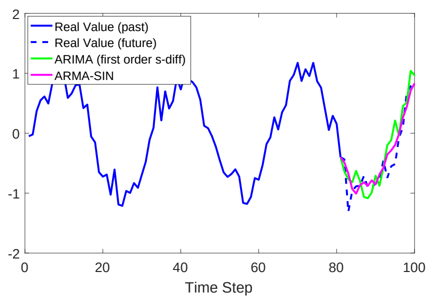
and the ARMA part to model $\boldsymbol{x}^0$ is ARMA(4,1). The results of two methods are shown in Figure 2.



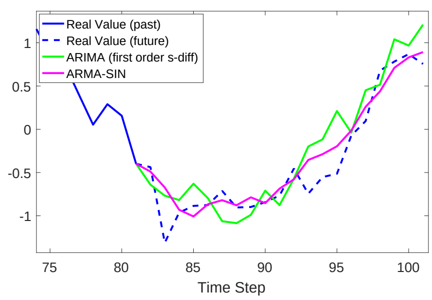
(a) Transformed series



(b) Transformed series (locally enlarged)



(c) Prediction results



(d) Prediction results (locally enlarged)

Figure 2: Transformed series and prediction results of SARIMA and ARMA-SIN (s-diff: seasonal difference).

Besides, we have 100 times of monte carlo simulation and the averaged prediction MSE is given in Table 2. Clearly, the transformed series $\hat{\boldsymbol{x}}^0$ of ARMA-SIN is nearer to its ground truth $\boldsymbol{x}^0$ than that of SARIMA.

Table 2: Averaged prediction MSE of SARIMA and ARMA-SIN

	ARIMA	ARMA-SIN
MSE	0.1022	0.0516

Thus the prediction accuracy is more satisfactory.

4.4. The Case of Directly Estimating the Seasonal Component

In this subsection, we investigate a scenario that we can directly estimate the seasonal component. Let $\boldsymbol{t} = 0 : 0.1 : 10$ (viz., $T_s = 0.1$), $\boldsymbol{n} = \boldsymbol{t}/T_s$, and $\boldsymbol{x} = \boldsymbol{x}^0 + f(\boldsymbol{t})$, where $f(\boldsymbol{t}) = sin(2\boldsymbol{t})$. Our purpose is to estimate out the function $sin(2t)$ directly from the collected history data.

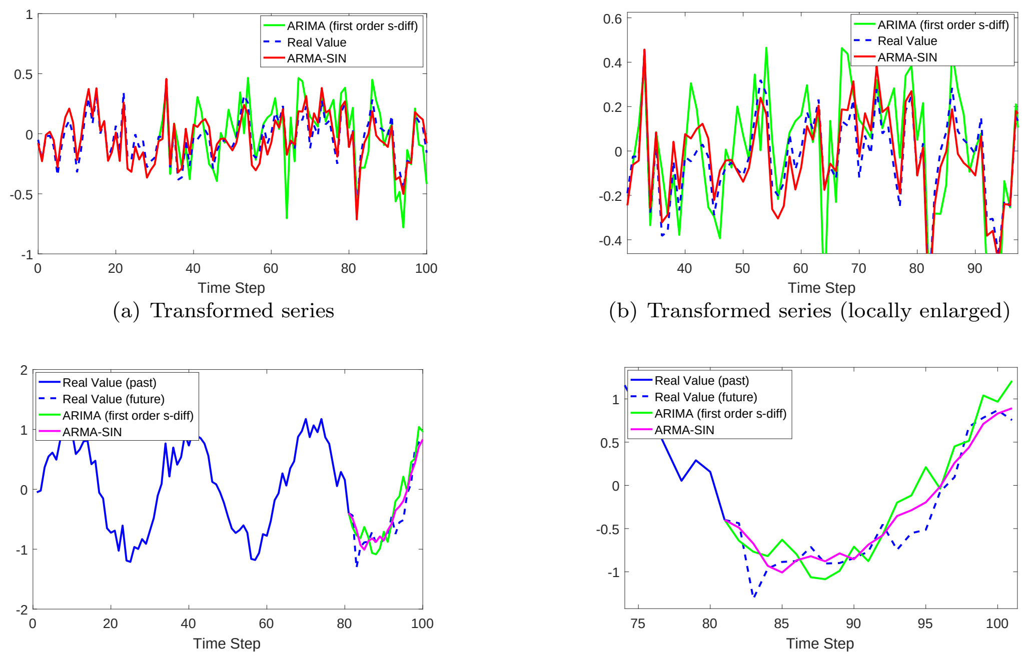
If we use our ARMA-SIN method, we can know that $\hat{f}(x)$ is with the form as

$$\begin{aligned}\hat{f}(t) &= 0.9721cos(0.2001n - 1.5440) \\ &= 0.9721cos(0.2001t/T_s - 1.5440) \\ &= 0.9721cos(2.001t - 0.4945\pi)\end{aligned} \tag{7}$$

It is amazingly close to its ground truth of $f(t) = sin(2t) = cos(2t - \pi/2) = cos(2t - 0.5\pi)$.

If we follow the standard modeling procedure with Box-Jenkins methodology, we have the estimated SARIMA(4,0,1)(31,0,0,0) model to handle this problem, meaning the operator used is the 31-lag seasonal differencing.

The results of two methods are shown in Figure 3.



(a) Transformed series (b) Transformed series (locally enlarged)

(c) Prediction results (d) Prediction results (locally enlarged)

Figure 3: Transformed series and prediction results of direct estimation.

Besides, we have 100 times of monte carlo simulation and the averaged prediction MSE is given in Table 3.

Table 3: Averaged prediction MSE of direct estimation

	SARIMA	ARMA-SIN
MSE	0.0844	0.0355

Clearly, the transformed series $\hat{\boldsymbol{x}}^0$ of ARMA-SIN is nearer to its ground truth $\boldsymbol{x}^0$ than that of SARIMA. Thus, the prediction accuracy is more satisfactory.

Intuitively, we can see from the three examples above that the ARMA-SIN method is indeed interesting over the S-ARIMA method. In the following sections, we progressively explain the philosophy behind and detail the ARMA-SIN method.

5. Secret Behind the ARIMA and SARIMA

5.1. Nature of Difference Operator and Seasonal Difference Operator

As mentioned in the previous section, the ARIMA and SARIMA attempt to make stationary a stochastic process by difference operator and seasonal difference operator. In this section, we investigate the nature of ARIMA and SARIMA from the perspective of SADFA (see Appendix). Specifically, we focus on the nature of the difference operator and seasonal difference operator.

Theorem 2. *The nature of the d-order difference operator is actually a high-pass digital filter that denies the low-frequency components of a time series. Since the trend of a time series is generally the low-frequency components (see Discrete Fourier Transform [4]), the d-order difference operator takes effect to make stationary a non-stationary (in the mean sense) stochastic process.*

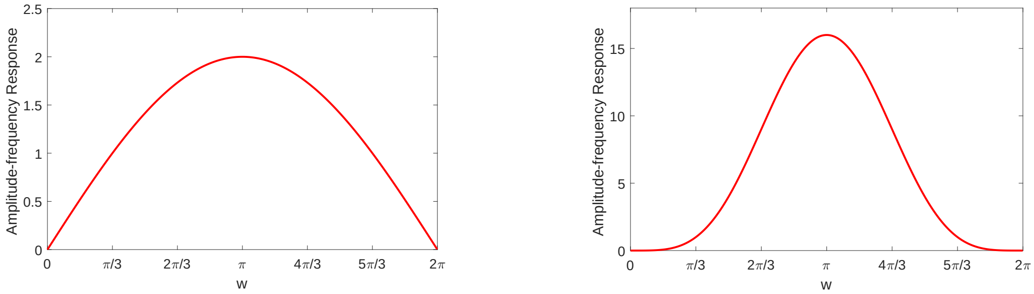
Proof. The transfer function of the d-order difference operator is given as

$$H(z) = (1 - z^{-1})^d, \tag{8}$$

and the amplitude-frequency response is as

$$|H(e^{jw})| = |(1 - e^{-jw})^d| = [\sqrt{2 - 2\cos(w)}]^d. \tag{9}$$

Eq. (9) immediately admits the theorem, since in the interval $[0,\ \pi]$, $|H(e^{jw})|$ is increasing from zero. Intuitively, see Figure 4.



(a) 1-order difference operator

(b) 4-order difference operator

Figure 4: Amplitude-frequency responses of the difference operator.

Theorem 3. *The nature of the L-lag seasonal difference operator (L-SDO) is actually a comb digital filter that denies some frequency components with $w = 2k\pi/L, k = 0, 1, 2, ..., L-1$. Since the seasonal components (namely, periodic components) with period L has period L in its spectra as well (meaning the frequencies of seasonal components are $w = 2k\pi/L, k = 0, 1, 2, ..., L-1$; see Discrete Fourier Series [4]), the L-lag difference operator takes effect to make stationary a non-stationary (in the mean sense) stochastic process.*

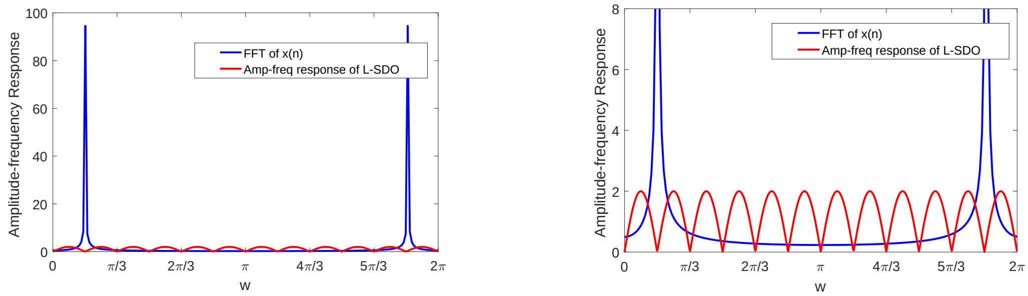
Proof. The transfer function of the L-lag seasonal difference operator is given as

$$H(z) = 1 - z^{-L}, \tag{10}$$

and the amplitude-frequency response is as

$$\begin{aligned} |H(e^{jw})| &= |1 - e^{-jLw}| \\ &= |[1 - \cos(Lw)] + j\sin(Lw)| \\ &= \sqrt{2 - 2\cos(Lw)}. \end{aligned} \tag{11}$$

Eq. (11) immediately admits the theorem, since in the interval $[0,\ \pi]$, $|H(e^{jw})|$ are zero-valued at $w = 2k\pi/L, k = 0, 1, 2, 3, ..., L-1$. Intuitively, we consider a sine series $x(n)$ with period $L = 12$ (i.e., $t = 0 : 2\pi/L : 100$) and take 12-lag seasonal difference over it. The results are illustrated in Figure 5.



(a) Full picture (b) Locally enlarged

Figure 5: The FFT of $x(n)$ and the amplitude-frequency responses of the L-lag seasonal difference operator. In the frequency domain, the FFT of $x(n)$ is multiplied by zero and therefore eliminated. As a result, in the time domain, the seasonal/periodic component $x(n)$ is eliminated.

Remark 1. Recall that in the discrete Fourier frequency domain, every $|H(e^{jw})|$ is periodic over w and the period is 2π. Since the highest frequency is indicated by $w = \pi$ and lowest is by $w = 0$ and $w = 2\pi$, we should only pay attention to the interval $[0,\ 2\pi]$. The trend of a time series is usually in low-frequency interval (around $w = 0$ and $w = 2\pi$) with large values [e.g., see Figure 6 (a) when w is below $\pi/6$]. Besides, according to Discrete Fourier Series [4], the seasonal component has impulses uniformly distributed (uniformly spaced) in the whole frequency domain $w \in [0,\ \pi]$ (e.g., see Figure 7 (a), around $w = \pi/6$). Therefore, in the frequency domain, we expect to remove all these outliers (large values) so that the transformed time series can become wide-sense stationary.

Remark 2. Note that if the seasonal component of a time series is perfectly a sine function, it only has one impulse in its spectra because the other impulses are zero-valued. This is why Figure 7 (a) only has one impulse other than many. For a general periodic series, it can be decomposed into a sum of sines. Thus, the spectra of a general periodic series have uniformly spaced impulses [e.g., see blue lines in Figure 10 (b)]. For more, see Discrete Fourier Series [4].

5.2. Why Are the ARIMA and SARIMA Not Sufficient

By now, Theorem 2 and Theorem 3 support the effectiveness of S-ARIMA somewhere. However, it is obvious that except the unwanted frequencies, the d-order difference operator of ARIMA and the L-lag seasonal difference operator of SARIMA also negatively impact the desired frequencies which should remain unchanged, for example, the frequencies higher than $w = \pi/3$ in Figure 4 have been significantly and unwantedly amplified, and the frequencies at and around $w = 2k\pi/12, k = 0, 2, 3, ...11$, (i.e., $k \neq 1$) in Figure 5 have also been unwantedly wiped away. This raises the theoretical insufficiency of S-ARIMA.

Theorem 4. *The ARIMA and SARIMA model are theoretically insufficient since they also negatively impact (distort or eliminate) the innocent frequency components (points or intervals).*

Proof. Due to Theorem 2, Theorem 3, and the statement itself in this theorem, the conclusion stands.

Remark 3. When we mention the insufficiency of S-ARIMA, we actually mean its theoretical insufficiency rather than the prediction accuracy in some practical applications. This is because the exact model only outperforms other models when the problem is exact. For example, if the data is generated from a linear function with sufficiently small Gaussian white noise, then the linear regression model should be better than any other high-order polynomial regression model. However, we cannot claim that the linear model is best all the time. Note that the focused time series, namely the exact problem we study in this paper, is a wide-sense stationary stochastic process. Thus, the ARMA model, according to the Wold's Decomposition theorem [9, p. 517], is the corresponding exact model. It means the operator that makes the original time series exact as wide-sense stationary is better, because the ultimate issue is to train an ARMA model. It is in this sense that we assert the S-ARIMA model is insufficient.

Remark 4. Although the ARIMA and SARIMA are theoretically insufficient, they are easiest methods to handle the non-stationary problems in practice. It means instead of designing some more proper (but more complicated) operators to make stationary a time series, we can for simplicity choose (seasonal) difference operator, if the performances are satisfactory for some specific problems in engineering. It is simplicity of such difference operators that makes them popular in engineering over years.

6. The General ARMA-SIN Model

According to Theorem 1, any WSS time series has two components: an ARMA part and a complex-valued exponential sum part (i.e., sum of sine time functions in the time domain). Therefore, the ARMA-SIN **model** is for a general non-stationary time series where SIN means the sum of SINe functions. Besides, our analysis in this paper is based on SADFA (spectral analysis and digital filtering approach) and Box-Jenkins methodology. In detail, when we have a general time series, we first use SADFA to transform it to be wide-sense stationary and then use Box-Jenkins methodology to model the transformed series. Thus, the term ARMA-SIN in this paper is also referred to a time series analysis **method** where SIN is for SADFA and ARMA is for Box-Jenkins methodology.

In Algorithm 1, we detail the methodology of ARMA-SIN. In nature, ARMA-SIN first decomposes a time series and then treats the decomposed components respectively.

Algorithm 1 ARMA-SIN

1: **Spectral Analysis:** transform the focused time series into Fourier frequency domain and identify the impulses frequency points (or interval) in frequency domain. Usually, points for the seasonal components and interval (low-frequency region around $w = 0$) for the trend;
2: **Digital Filtering:** design a proper digital filter with proper cut-off frequencies such that the focused time series can be made stationary. Ideally, such proper digital filters should only remove the unwanted frequency points or interval while keeping the rest unchanged;
3: **ARMA:** follow the standard Box-Jenkins methodology to model the stationary remainder (transformed time series) as a ARMA one;
4: **Forecast:** predict the future with two independent process, ARMA part and the trend (and/or seasonal trend) part, respectively, and then integrate the results together. Note that the raw time series subtract the ARMA remainder gives the trend (or seasonal trend) part.

Remark 5. As we can see, the ARIMA and SARIMA are special cases of Algorithm 1. For ARIMA, the digital filter we use is just the difference operator (i.e., a special case of high-frequency-pass filters, that is, a low-frequency-stop filter) to remove (i.e., stop) the low-frequency trend part. For SARIMA, the digital filter we use is just the seasonal difference operator (i.e., a special case of comb filters, that is, a fixed-point-frequency-stop filter) to remove (i.e., stop) the frequency impulses in frequency domain.

7. Derive the Solutions for Scenarios of Warming-up

In this section, we detail the derivation for solutions of scenarios of warming-up given in Section 4. For briefness, we will not repeat the standard modeling procedures of Box-Jenkins methodology [1]. All the filters that this paper designed are based on IIR (infinite impulse response) model and Elliptic method [4].

Remark 6. Usually it is hard to decide the best parameters like cut-off frequencies at the first glance for a general problem. However, we can try some times with plausible parameters and choose the best among them. This seems tedious but unavoidable in practice because there is no free lunch: advanced methods mean complex coding, complex parameters selecting, and/or complex calculation burden. In this sense, the use of simple (seasonal) difference operator seems really delightful for all of us. Nevertheless, the cost is the lacking of the interpretability.

7.1. The Case of Variant Mean

The time series is shown in Figure 1 (c). The trend is linear. According to Discrete Fourier Transform theory, the linear trend can be represented by a finite sum of sine functions with acceptable approximation error. In order to separate (extract) the linear trend, we should design a high-pass filter which denies the low-frequency component (trend) to pass and allows the high-frequency components to go through without any changes. Since we use the Elliptic method to design a IIR filter (the function *ellipord* in MATLAB, for more, see its reference page in [8]), we set the parameters in Table 4.

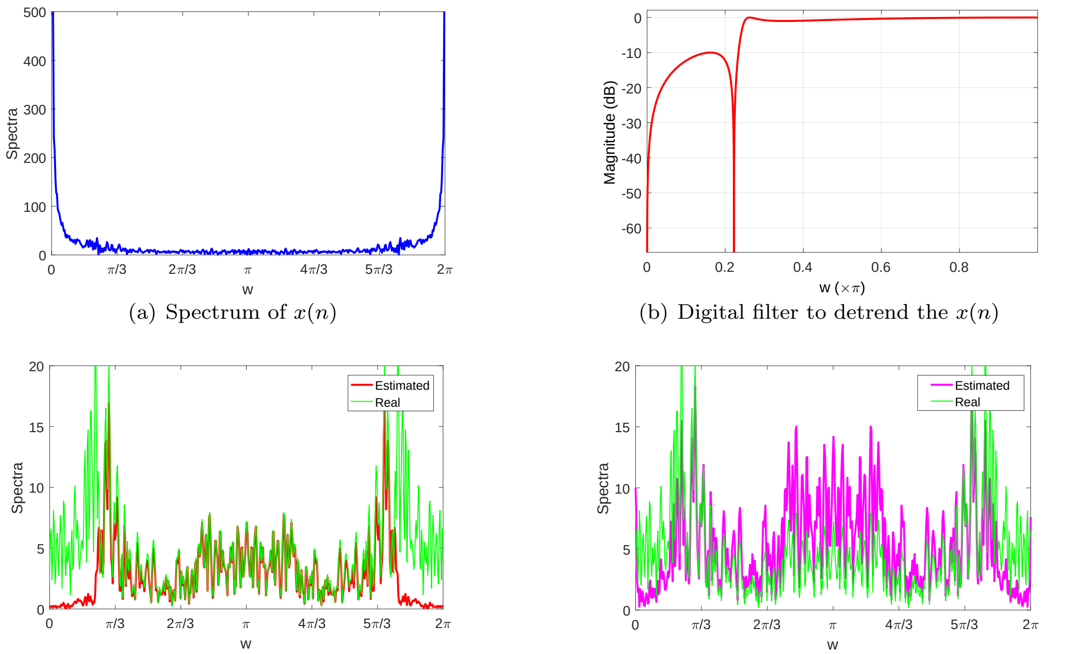
Table 4: The filter parameters for the case of variant mean

	Wp	Ws	Rp	Rs
Value	0.25	0.2	1	10

The specific meanings and detailed usage of parameters in Table 4 should be found in [4] or the reference page of the function *ellipord* in [8].

This high-pass filter actually defines the system (5).

The spectral analysis results are given in Figure 6.



(a) Spectrum of $x(n)$

(b) Digital filter to detrend the $x(n)$

(c) Estimated spectra of x^0 using ARMA-SIN

(d) Estimated spectra of x^0 using 1-order difference

Figure 6: Spectra analysis for the case of variant mean.

From Figure 6, we can see the 1-order difference operator significantly distorts (amplifies) the spectra in high frequency interval, although it is powerful to detrend. However, our ARMA-SIN keeps 1 [n.b., $20\log(1) = 0dB$] in higher frequency area, making no changes to innocent components. This is why our ARMA-SIN outperforms ARIMA. Note that Figure 6 (b) is given in logarithmic unit of decibel (dB). For more on decibel, see [4].

7.2. The Case of Using SARIMA

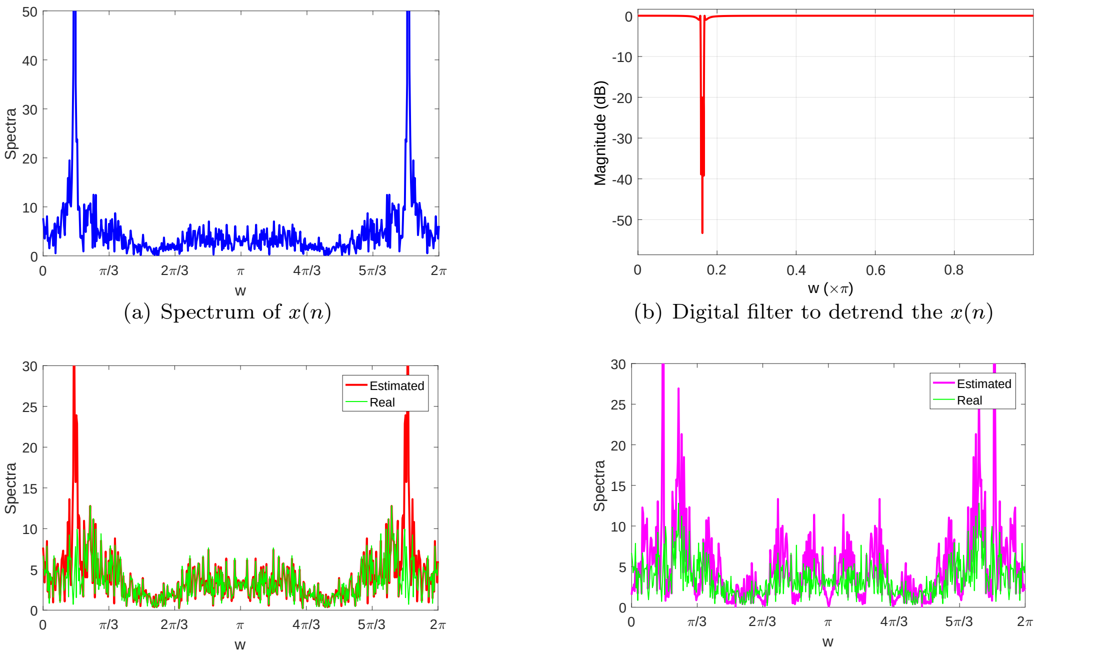
This case is concerned with the periodic time series which is suitable for SARIMA. The time series is shown in Figure 2 (c). In Figure 7, we can check the spectra of the time series. We can see that there is an outstanding line-spectra (i.e., impulse, outlier) around $w = 0.5\pi$. Thus we want to design a band-stop digital filter to deny the outstanding frequency component. Since we use the Elliptic method to design a IIR filter (the function *fdesign.bandstop* in MATLAB, for more, see its reference page in [7]), we set the parameters in Table 5.

Table 5: The filter parameters for the case of using SARIMA

	Value	Comments
Fpass1	0.158	First Passband Frequency ($\times\pi$)
Fstop1	0.16	First Stopband Frequency ($\times\pi$)
Fstop2	0.165	Second Stopband Frequency ($\times\pi$)
Fpass2	0.168	Second Passband Frequency ($\times\pi$)
Apass1	1	First Passband Ripple (dB)
Astop	20	Stopband Attenuation (dB)
Apass2	1	Second Passband Ripple (dB)

This band-stop filter actually defines the system (6).

The spectral analysis results are given in Figure 7.



(a) Spectrum of $x(n)$

(b) Digital filter to detrend the $x(n)$

(c) Estimated spectra of x^0 using ARMA-SIN

(d) Estimated spectra of x^0 using 12-lag seasonal difference

Figure 7: Spectra analysis for the case of using SARIMA.

The spectra of the 12-lag seasonal difference operator can be found in Figure 5 as a comparison to Figure 7.

From Figure 7, we can see that the 12-lag seasonal difference operator significantly distorts/eliminates the spectra at and around $w = 2k\pi/12, k = 0, 2, 3, ..., 11$, although it is powerful to wipe out the frequency at $w = 2\pi/12$ $(k = 1)$. This is why our ARMA-SIN outperforms ARIMA.

7.3. The Case of Directly Estimating the Seasonal Component

In this section we use the information presented in spectra of a time series to directly estimate the seasonal (impulses in frequency domain) components. The philosophy is based on DFT (discrete Fourier transform) and Inverse DFT [4, 10, 3].

Suppose the outstanding value of FFT of $x(n)$ is at $n = K$ ($K \in \{1, 2, 3, ..., length(\boldsymbol{n})\}$) and its value is H_K, then we have $w = 2\pi K/N$. Besides, by DFT (and/or FFT), the amplitude in time domain is given as $A = 2|H_K|/N$ and phase is as $\varphi = \varphi(H_k)$, where $|H_K|$ denotes the modulus of H_K and $\varphi(H_k)$ denotes the argument (angle) of H_K. Note that H_K is complex-valued. Therefore the line-spectra (impulse in spectra) component of $x(n)$ with frequency of $w = 2\pi K/N$ has its expression in time domain as

$$x_p(n) = A\cos(wn + \varphi). \tag{12}$$

Specifically in this section, the spectra of the interested time series is as Figure 8.

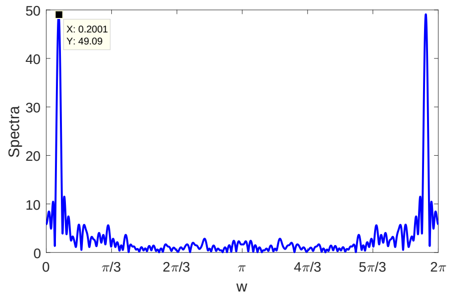


Figure 8: Spectra analysis for the case of direct estimation.

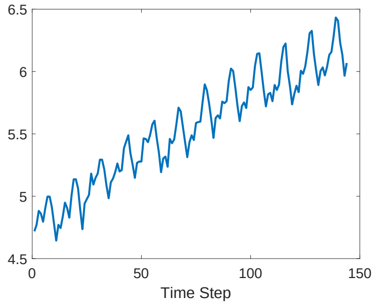
From the figure we can directly know that $w = 0.2001$, and $A = 2 \times 49.09/N = 0.9721$ (n.b., $N = 101$). Besides, by checking the corresponding phase at this frequency $w = 0.2001$, we have $\varphi = -1.5440$. Thus (12) admits (7).

8. Real-World Example: Air Passengers Forecasting

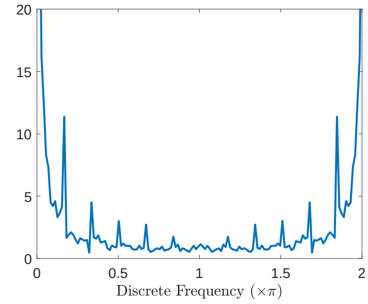
In this section, we study the classic AirPassengers dataset: number of air passengers per month. Technical manipulation details and exhaustive explanations are available in the disclosed source data and codes. For the demonstration purpose and in view of necessity, we do not repeat them here and only display visual results. The plot, spectra, sample autocorrelation function, and sample partial autocorrelation function of the log-transformed AirPassengers series are shown in Figure 9. The spectra contain a great number of large-valued outliers; it implies that both a low-frequency trend and several seasonal components exist in the log-transformed AirPassengers series.

Therefore, the log-transformed AirPassengers series must not be wide-sense stationary, and we need to design digital filters to de-trend and de-season. After applying the digital filters that remove the low-frequency trend and seasonal components over the log-transformed series, the residual series is shown in Figure 10.

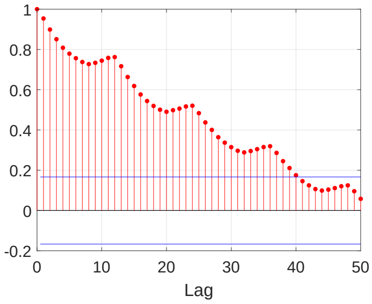
As we can see, there no longer exist outliers and the spectra become continuous;i.e., the digitally filtered residual series is wide-sense stationary. Therefore, the ARMA model can be used to fit the residual series. Overall, the forecasting performance of SARIMA and ARMA-SIN is shown in Figure 11. For long-term forecasting, the proposed ARMA-SIN outperforms SARIMA. This is because the nature of ARMA-SIN is to decompose a time series, and forecast the decomposed components respectively. Hence, a long-term forecasting procedure is admitted, especially when forecasting the trend and seasonal components. In contrast, SARIMA does not conduct such decomposition and forecast the series as a whole: the long-term forecasting ability is lacking. This reminds us the facts in Remark 3: the insufficiency of S-ARIMA is just theoretical. In practice, for short-term (especially rolling-step) forecasting, S-ARIMA possibly has highly effective performance.



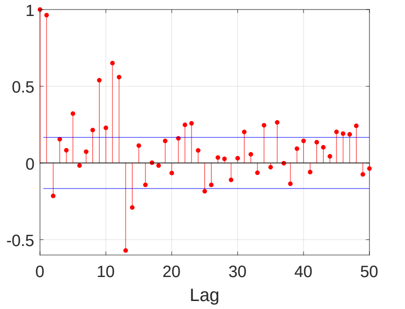
(a) Log-transformed series



(b) Spectra of log-transformed series

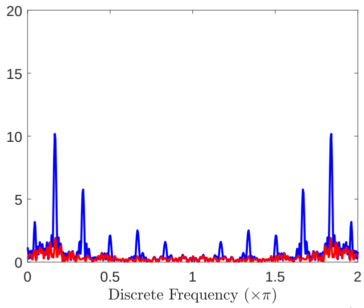


(c) Sample autocorrelation function

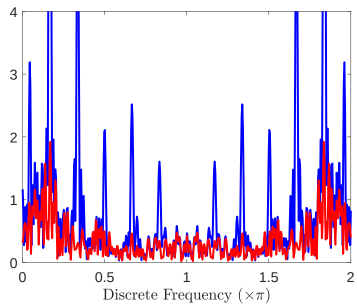


(d) Sample partial autocorrelation function

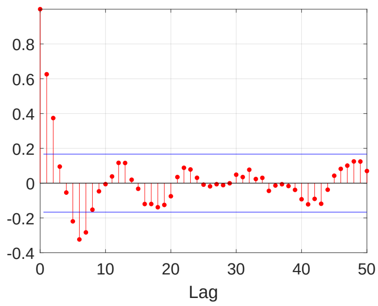
Figure 9: Properties of the log-transformed series.



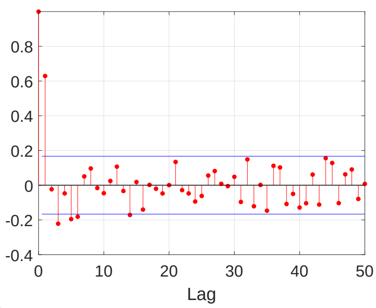
(a) Spectra of digitally filtered series (red)



(b) Spectra of digitally filtered series (red; closeup)

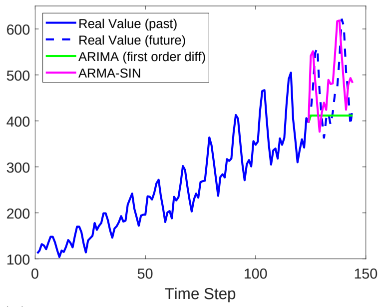


(c) Sample autocorrelation function

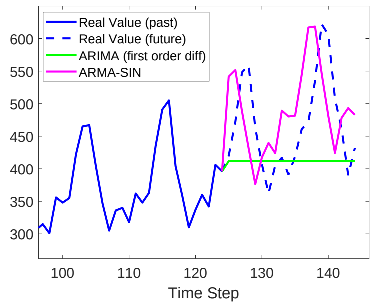


(d) Sample partial autocorrelation function

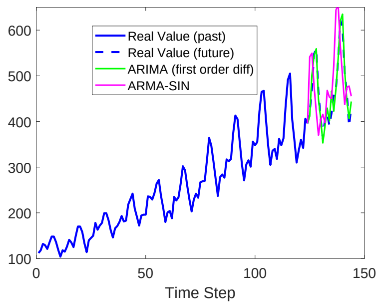
Figure 10: Properties of the digitally filtered series. In (a) and (b), red lines are of the digitally filtered series (both de-trended and de-seasoned), while blue lines are of the only de-trended series (for comparison). Since already de-trended, there exist no outliers near 0π in blue lines.



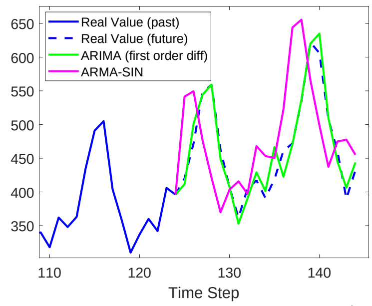
(a) Long-term forecasting performance



(b) Long-term forecasting performance (closeup)



(c) Short-term forecasting performance



(d) Short-term forecasting performance (closeup)

Figure 11: Forecasting performance of SARIMA and ARMA-SIN. In (a) and (b), the long-term (20-step-ahead) forecasting performance is shown, while in (c) and (d), the short-term (rolling 1-step-ahead) forecasting performance of SARIMA and ARMA-SIN is shown.

9. Conclusions

We have in this paper discussed the nature, philosophies, effectiveness, insufficiency, and improvements of ARIMA and SARIMA from the perspectives of Linear System Analysis, Spectra Analysis, and Digital Filtering. We show that S-ARIMA also impacts innocent frequencies when making stationary a time series. In contrast, ARMA-SIN can remain the innocent frequencies unchanged. However, it is admitted that although the displayed ARMA-SIN is general and interesting, it is relatively complex to be applied in practice, compared to the simple ARIMA, SARIMA, and the like. Using this method requires enough experiences to identify the spectral components of interest and setup proper parameters like cut-off frequencies to extract or discard them by designing a proper digital filter. Although depressive mentioning this, we should make clear that it is still a bright future if we can design some powerful algorithms to help us automatically select proper parameters to design digital filters, just like the cross-validation method in training a machine learning model. However, this challenging work should be jointly handled with inspired scholars in this area.

Appendix

In order not to confuse readers from different communities, we mention some basic glossaries of interest in this paper.

- ARMA: Autoregressive Moving Average;
- ARIMA: Autoregressive Integrated Moving Average;
- SARIMA: Seasonal Autoregressive Integrated Moving Average;
- Spectrum: The spectrum of a time function is its Fourier transform (or the squares of the Fourier transform; cf. power spectra) which describes a time series in the Fourier frequency domain;

- Spectral Analysis: The analysis and processing for a time series in Fourier frequency domain. The Fourier transform and inverse Fourier transform connect the time domain and Fourier frequency domain;
- (Discrete) Linear System: A discrete linear operator, defined by a difference equation [see (1)] in an autoregressive moving average form, that transforms a time series into another time series (for example, the difference operator);
- Digital Filter: A linear system that takes effect of transforming a time series in the Fourier frequency domain. However, it works for a time series in the time domain. For example, the moving average method, although it is directly applied in the time domain, in essence, it removes the low-frequency component (i.e., the trend) in the Fourier frequency domain;
- SADFA: Spectral Analysis and Digital Filtering Approach used to analyze and process a time series;
- DFT/DFS: Discrete Fourier Transform/Series;
- FFT: Fast Fourier Transform;
- DTFT: Discrete Time Fourier Transform;

Besides, two terms Time Series and Stochastic Process are used interchangeably in this paper.

Declarations of Interest

The authors declare that there is no any potential competing interests.

References

[1] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[2] Rodrigo N Calheiros, Enayat Masoumi, Rajiv Ranjan, and Rajkumar Buyya. Workload prediction using arima model and its impact on cloud applications' qos. *IEEE Transactions on Cloud Computing*, 3(4):449–458, 2015.

[3] Luis Chaparro and Aydin Akan. *Signals and Systems using MATLAB*. Academic Press, 2018.

[4] Paulo SR Diniz, Eduardo AB Da Silva, and Sergio L Netto. *Digital Signal Processing: System Analysis and Design*. Cambridge University Press, 2010.

[5] James D Hamilton. Time series analysis. *Economic Theory. II, Princeton University Press, USA*, pages 625–630, 1995.

[6] Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2018.

[7] MathWorks. Bandstop filter specification object. `https://www.mathworks.com/help/dsp/ref/fdesign.bandstop.html`, 2018. Accessed on March 14, 2019.

[8] MathWorks. Minimum order for elliptic filters. `https://www.mathworks.com/help/signal/ref/ellipord.html?searchHighlight=ellipord&s_tid=doc_srchtitle`, 2018. Accessed on March 14, 2019.

[9] Athanasios Papoulis and S Unnikrishna Pillai. *Probability, Random Variables, and Stochastic Processes*. Tata McGraw-Hill Education, 2002.

[10] Won Young Yang. *Signals and Systems with MATLAB*. Springer Science & Business Media, 2009.