

**DETC2017-67483**

## AN EVALUATION OF ARFIMA PROGRAMS

**Kai Liu and Xi Zhang**

School of Mechanical Electronic and Information Engineering  
China University of Mining and Technology, Beijing  
Beijing, 100083, China  
kliu\_cmutb@163.com, zhangxi6681@163.com

**YangQuan Chen\***

School of Engineering  
University of California, Merced  
Merced, CA 95343, USA  
ychen53@ucmerced.edu

### ABSTRACT

*Strong coupling between values at different time that exhibit properties of long range dependence, non-stationary, spiky signals cannot be processed by the conventional time series analysis. The ARFIMA model, which employs the fractional order signal processing techniques, is the generalization of the conventional integer order models—ARIMA and ARMA model. Therefore, it has much wider applications since it could capture both short-range dependence and long range dependence. For now, several software have developed functions dealing with ARFIMA processes. However, it could be a big difference, if using different numerical tools for time series analysis. Time to time, being asked about which tool is suitable for a specific application, the authors decide to carry out this survey to present recapitulative information of the available tools in the literature, in hope of benefiting researchers with different academic backgrounds. In this paper, 4 primary functions concerning simulation, fractional order difference filter, estimation and forecast are compared and evaluated respectively in the different software and informative comments are also provided for selection.*

### INTRODUCTION

Humans are obsessed about their future so much that they worry more about their future than enjoying the present. Time series modelling and analysis are scientific ways to predict the future. When dealing with empirical time series data, it usually

comes to the classic book of Box and Jenkins's methodology for time series models in the 1970s, in which it introduced the autoregressive integrated moving average (ARIMA) models to forecast and predict the future behavior [1]. However, the conventional integer order models can only capture short-range dependence (SRD); for example, Poisson processes, Markov processes, autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and ARIMA processes [2].

In time series analysis, another traditional assumption is that the coupling between values at different time instants decreases rapidly as the time difference or distance increases. Long-range dependence (LRD), also called long memory or long-range persistence, is a phenomenon that may arise in the analysis of spatial or time series data. LRD was first highlighted in hydrological data by the British hydrologist H. E. Hurst, and then the other fields in econometrics, network traffic, linguistics and the earth sciences. LRD, which is characterized by the Hurst parameter means that there is a strong coupling between values at different time. This indicates that the decay of the autocorrelation function is hyperbolic and decays slower than exponential decay, and that the area under the function curve is infinite. We can also say that their autocorrelation functions (ACF) are power-law distributed.

Due to the increasing demand on modeling and analysis of LRD and self-similarity in time series, such as financial data, communications networks traffic data and underwater noise, the fractional order signal processing (FOSP) technique is becoming a booming research area. Moreover, fractional Fourier transform (FrFT), which is the generalization of the FFT, has become

---

\*Address all correspondence to this author.

one of the most valuable and frequently used techniques in the frequency domain analysis of the fractional order systems.

Compared to conventional integer order models, the ARFIMA model gives a better fit and result when dealing with the data which possess the LRD property. Sun *et al.* applied the ARFIMA model to analyze the data and predict the future levels of the elevation of Great Salt Lake (GSL) [3]. The results showed that the prediction results have a better performance compared to the conventional ARMA models. Li *et al.* examined 4 models for the GSL water level forecasting: ARMA, ARFIMA, GARCH and FIGARCH. They found that FIGARCH offers best performance indicating that conditional heteroscedasticity should be included in time series with high volatility [4]. Sheng and Chen proposed a new ARFIMA model with stable innovations to analyze the GSL data, and predicted the future levels. They also compared accuracy with previously published results [5]. Contreras-Reyes and Palma developed the statistical tools “afmtools” package in R for analyzing ARFIMA models. In addition, the implemented methods are illustrated with applications to some numerical examples and tree ring data base [6]. Baillie and Chung considered the estimation of both univariate and multivariate trend-stationary ARFIMA models, which generated a long memory autocorrelated process around a deterministic time trend. The model was found to be remarkably successful at representing annual temperature and width of tree ring time series data [7]. Dr. Jurgen Doornik is a research fellow at the University of Oxford and a director of OxMetrics Technologies Ltd. OxMetrics is an econometric software including the Ox programming language for econometrics and statistics, developed by Doornik and Hendry. Several papers and manuals are available for the ARFIMA model with OxMetrics [8–10].

Nowadays, there are lots of numerical tools available for the analysis of the ARFIMA processes. Since applications are developed by different groups based on different understandings of algorithms and definitions of accuracies and procedures. As a consequence, the estimation and prediction results may be different or even conflicting with others. For the scholars or engineers who are going to do the modeling work of the ARFIMA processes, they might get confused that which tool is more suitable to choose. Thus, we have evaluated techniques concerning ARFIMA process analysis so as to provide some guidelines when choosing appropriate methods to do the analysis. With this motivation, this paper briefly introduces their usage and algorithms, evaluates the accuracy, compares the performance, and provides informative comments for selection. Through such efforts, it is hoped that an informative guidance is provided to the readers when they face the problem of selecting a numerical tool for a specific application.

The rest of the paper is organized as the follows: Section 2 introduces the basic mathematics of LRD and the ARFIMA model. Section 3 gives a brief review and description on the software commonly used for the analysis of the ARFIMA processes.

In Section 3, the quantitative performances of the tools are evaluated and compared in four primary categories—simulation, processing, estimation and prediction in the ARFIMA process. Conclusions are given in Section 4.

## LRD AND ARFIMA MODEL

When the hydrologist Hurst, H.E. spent many years analyzing the records of elevation of the Nile river in the 1950s, he found a strange phenomena: the long-range recording of the elevation of the Nile river has much stronger coupling, and the ACF decays slower than exponentially [11]. In order to quantify the level of the coupling, the rescaled range (R/S) analysis method was provided to estimate the coupling level, which is now called the Hurst parameter. Many valuable Hurst parameter estimators were provided to more accurately characterize the LRD time series. Since then the LRD or long memory phenomenon has attracted numerous research studies. Based on Hurst’s analysis, more suitable models, such as ARFIMA and fractional integral generalized autoregressive conditional heteroscedasticity (FIGARCH) were built to accurately analyze LRD processes.

Autocorrelation analysis is a very useful technique for identifying trends and periodicities in the data, in a manner that is often more precise than can be obtained with simple visual inspection.

$$\rho_k = \frac{\text{Cov}(X_t, X_{t-k})}{\text{Var}(X_t)}. \quad (1)$$

The correlations function  $\rho(k)$  of the stationary SRD stochastic models, such as the ARMA processes and Markov processes, is absolutely summable, while the correlations function  $\rho_k$  is not absolutely summable for the processes with long-range dependence. Signals with long-range correlations, which are characterized by power-law decaying autocorrelation function, occur ubiquitously in nature and many man-made systems. Because of the strong coupling and the slow decaying autocorrelation, these processes are also said to be long memory processes. Typical examples of LRD signals include financial time series, underwater noise, electroencephalography (EEG) signal, etc. The level of the dependence or coupling of LRD processes can be indicated or measured by the estimated Hurst parameter, or Hurst exponent  $H \in (0, 1)$  [11]. If  $0 < H < 0.5$ , the time series is a negatively correlated process, or an anti-persistent process. If  $0.5 < H < 1$ , the time series is a positively correlated process. If  $H = 0.5$ , the time series has no statistical dependence. The LRD processes are also closely related to fractional calculus. In order to capture the property of coupling or power-law decaying autocorrelation, fractional calculus based LRD models have been suggested, such as ARFIMA and FIGARCH models [12, 13]. The ARFIMA model is a generalization of ARMA model, which is a typical fractional order system.

## Autoregressive (AR) model

The notation  $AR(p)$  refers to the autoregressive model of order  $p$ . The  $AR(p)$  model is written [1]:

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t, \quad (2)$$

where  $\phi_1, \dots, \phi_p$  are autoregressive parameters,  $c$  is a constant, and the random variable  $\varepsilon_t$  is white noise. Some constraints are necessary on the values of the parameters so that the model remains stationary. For example, processes in the  $AR(1)$  model with  $|\phi_1| \geq 1$  are not stationary. In statistics and signal processing, an autoregressive (AR) model is a representation of a type of random process; as such, it describes certain time-varying processes in nature, economics, etc.

## Moving Average (MA) model

The notation  $MA(q)$  refers to the moving average model of order  $q$  [1]:

$$X_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (3)$$

where the  $\theta_1, \dots, \theta_q$  are the moving average parameters of the model,  $\mu$  is the expectation of  $X_t$  (often assumed to equal 0), and the  $\varepsilon_t, \varepsilon_{t-1}, \dots$  are again, white noise error terms. The moving average (MA) smooths a time series, which can produce cyclic and a trend like plots even when the original data are themselves independent random events with fixed mean. This characteristic lessens its usefulness as a control mechanism.

## ARIMA and ARFIMA model

The above AR and MA models can be generalized as follows:

$$(1 - \sum_{i=1}^p \phi_i B^i)(1 - B)^d (X_t - \mu) = (1 + \sum_{i=1}^q \theta_i B^i) \varepsilon_t. \quad (4)$$

The above  $(1 - B)^d$  is called a difference operator  $\nabla^d$ . The ARMA or ARIMA models can only capture the SRD property since  $d$  is confined in the range of integer order. Therefore, in order to capture the LRD property of the fractional system, ARFIMA( $p, d, q$ ) model was proposed. In fact, the operator can be defined in a natural way by using binomial expansion for any real number  $d$  with Gamma function.

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d+1-k)} (-B)^k. \quad (5)$$

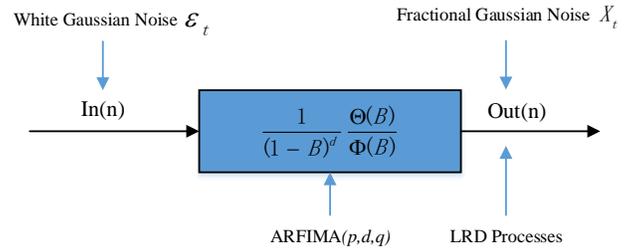


FIGURE 1. ARFIMA MODEL

Many authors suggested the use of the fractionally ARIMA model by using a fractional difference operator rather than an integer one could better take into account this phenomenon of LRD [14]. Hosking *et al.* defined an extension of the ARIMA model which allows for the possibility of stationary long-memory models [15]. Thus, the general form of ARIMA( $p, q, d$ ) process  $X_t$  in the Eq. 4—the ARFIMA( $p, d, q$ ) process is defined as:

$$\Phi(B)(1 - B)^d X_t = \Theta(B) \varepsilon_t, \quad (6)$$

where  $d \in (-0.5, 0.5)$ , and  $(1 - B)^d$  is defined as the fractional differencing operator in the Eq. 5. ARFIMA( $p, d, q$ ) processes are widely used in modeling LRD time series, where  $p$  is the autoregressive order,  $q$  is the moving average order and  $d$  is the level of differencing [16]. The larger the value of  $d$ , the more closely it approximates a simple integrated series, and it may approximate a general integrated series better than a mixed fractional difference and ARMA model.

Fig. 1 presents the discrete ARFIMA process which can be described as the output of the fractional-order system driven by a discrete wGn. The ARFIMA( $p, d, q$ ) process is the natural generalization of the standard ARIMA or ARMA processes. In a fractionally differenced model, the difference coefficient  $d$  is a parameter to be estimated first. The intensity of self-similar of ARFIMA is measured by a parameter  $d$  [17]. For the finite variance process with fractional Gaussian noise,  $d$  has a closed relation with Hurst parameter  $H$  [2, 17]:

$$d = H - 1/2. \quad (7)$$

In this way, the parameter  $d$  may be chosen to model long-time effects, whereas  $p$  and  $q$  may be selected to model relatively short-time effects.

## REVIEW AND EVALUATION

ARFIMA( $p, d, q$ ) processes are widely used in modeling LRD time series, especially for the high frequency trading data, network traffic and hydrology dataset etc. In practice, several time series exhibit LRD in their observations, leading to the development of a number of estimation and prediction methodologies to account for the slowly decaying autocorrelations. The ARFIMA process is one of the best-known classes of long-memory models. As introduced in Section 1, most statistical analysis software are embedded with ARIMA models. A review of the current software dealing with fractional ARIMA model analysis are as follows:

### 1. MATLAB applications

MATLAB<sup>®</sup> (Matrix laboratory) is a multi-paradigm numerical computing environment and fourth-generation programming language developed by MathWorks. The MATLAB applications are interactive applications written to perform technical computing tasks with the MATLAB scripting language from MATLAB File Exchange, through additional MATLAB products, and by users.

### 2. SAS software

SAS (Statistical Analysis System), is a software suite developed by SAS Institute for advanced analytics, multivariate analyses, business intelligence, data management, and predictive analytics.

### 3. R packages

The R Studio team contributes code to many R packages and projects. R users are doing some of the most innovative and important work in science, education, and industry. It's a daily inspiration and challenge to keep up with the community and all it is accomplishing.

### 4. Ox Metrics

Ox<sup>TM</sup> is an object-oriented matrix language with a comprehensive mathematical and statistical function library. Many packages were written for Ox including software mainly for econometric modelling. The Ox packages for time series analysis and forecasting

MATLAB codes are open-source applications where we could download, view and revise the code if possible while other three are packaged and embedded in the software modules. In the following evaluation parts, we could clearly see the differences between them even with the same input.

As introduced in the Section 1 and 2, four primary embedded functions concerning simulation, fractional difference filter, parameter estimation and forecast, are tested and evaluated for the ARFIMA processes in Tab. 1. It should be noted that the first two functions can be regarded as the forward problem solving system, while the latter two are developed for the backward problem solving systems which is much more significant. In view of the above, this section can be divided into 4 parts.

TABLE 1. NUMERICAL TOOLS FOR THE ARFIMA PROCESS

	MATLAB	R	SAS	OxMetrics
Simulation	✓*	✓	✓	✗
Fractional Difference	✓	✓	✓	✗
Parameter Estimation	✓	✓	✓	✓
Forecast	✓	✓	✗	✓

Note: ✓\* means it can simulate ARFIMA processes, but cannot choose or define the initial seeds.

## Simulation

On the website of MATLAB Central, there are two files that can simulate ARFIMA processes. They are developed by Faticchi [18] and Caballero [19]. But user cannot choose initial random seed, that is, it can only simulate one certain series of ARFIMA process. ARFIMA( $p, d, q$ ) estimator is developed by Inzelt, which is used for linear stationary ARFIMA( $p, d, q$ ) process [20].

R is a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, statistical tests, time series analysis, classification, clustering, etc. Like the MATLAB Central, CRAN is a platform that store identical, up-to-date, versions of code and documentation for R. There are several major packages concerning ARFIMA process according to the authors' survey in the Fig. 2.

The first two packages are used for the processing of the ARFIMA processes, including Hurst fitting, calculation and fractional order differencing etc., while the latter two are mainly used for the parameter estimation of ARFIMA. The last package 'arfima' is the most comprehensive tool that could simulate, estimate and predict the results of ARFIMA processes. In the paper, we use the last one package to compare with the other software.

SAS and R could also generate the ARFIMA process by defining the order of AR( $p$ ) and MA( $q$ ), setting the parameters  $\phi$ ,  $\theta$  and  $d$ , respectively. Besides, the number of the initial random seeds could/should be set for the stochastic process. Random seeds are defined by the internal algorithm, which make the initial stochastic process a difference. However, it may be a big difference if picking arbitrary seeds. In order to illustrate the above problems, we have generated the ARFIMA(1, 0.4, 1) process with the  $d = 0.4$ ,  $\phi = 0.5$ ,  $\theta = -0.1$  and  $\sigma = 1$ . Then we set 100 different initial random seeds with 3000 sampling points, and do the same estimation. It should be kept in mind that even with the same simulation software that generates the processes, the estimation results could be a big difference in the Fig. 2. However, from the perspective of the sample-path analysis for the

**TABLE 2.** COMPARISON OF ARFIMA PACKAGES IN R

Package	Author	Release Date	Typical Functions	Requirements
fractal	William Constantine <i>et al.</i> [21]	2016-05-21	hurstSpec	R >= 3.0.2
fracdiff	Martin Maechler <i>et al.</i> [22]	2012-12-02	fracdiff	longmemo, urca
afmtools	Javier E. Contreras-Reyes <i>et al.</i> [23]	2012-12-28	arfima.whittle	R (>= 2.6.0), polynom, fracdiff, hypergeo, sandwich, longmemo
ArfimaMLM	Patrick Kraft <i>et al.</i> [24]	2015-01-21	arfimaMLM	R (>= 3.0.0), lme4, fractal
arfima	Justin Q. Veenstra <i>et al.</i> [25]	2015-12-31	arfima	R (>= 2.14.0), ltsa

stochastic processes, this could be the advantage compared to the MATLAB ARFIMA applications, which can only generate one certain series (path). Besides, we have also found that the SAS software is somewhat better or “conservative” while R software is more “aggressive”. We could check the comparisons below with dashed lines showing true values of parameters.

It should be also noted that, even with a certain series of the initial starting random seeds, the estimation results could also have quite a few variations. For example, we have set the fractional order  $d$  from 0 to 0.5, and do the simulation and estimation accordingly in MATLAB. It can be seen that the estimation  $\hat{d}$  are jumping up and down around the true values (the red line) in the Fig. 3.

Here are some comments of this subsection:

I. Estimation results are also depended on the of the initial random seeds, even series are from their own simulations.

II. The test results may be different if not enough points are generated.

III. Estimation results may not be accurate if only use one method.

### Fractional Order Difference Filter

Many time series contain trends, i.e. they are non-stationary. Trends may be linear, or have some more complex form, such as polynomial or logistic. Whatever the form of the trend it is usually preferable to remove and/or specify the trend explicitly prior to further analysis and modeling. By differencing the trend if necessary, the smoothed data should stay stationary. Based on the theory of Box-Jenkins, an ARIMA model can be viewed as a “filter” that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts. Since the beginning of the 1980s, long memory ARFIMA model has been well investigated, especially many papers that based on the parameter estimation problem. Shumway and Stoffer gave a brief overview of “long memory ARMA” models in [26]. This

type of model might possibly be used when the ACF of the series tapers slowly to 0 and spectral densities are unbounded at  $f = 0$ .

In some instances, however, we may see a persistent pattern of non-zero correlations that begins with a first lag correlation that is not close to 1. In these cases, models that incorporate “fractional differencing” may be useful. Difference the time series data using the approximated binomial expression of the long-memory filter is a prerequisite to estimate of the memory parameter in the ARFIMA( $p, d, q$ ) model. The user should not only put numeric vector or univariate time series, but also specify the order of the fractional difference filter. By passing through fractional order difference filter, the ARFIMA series will yield residuals which are uncorrelated and normally distributed with constant variance in the Fig. 4.

The sample ACF and partial autocorrelation function (PACF) are useful qualitative tools to assess the presence of autocorrelation at individual lags. The Ljung-Box Q-test is a more quantitative way to test for autocorrelation at multiple lags jointly [27]. The Ljung-Box test statistic is given by:

$$Q(m) = N(N+2) \sum_{h=1}^m \frac{\hat{\rho}_h^2}{N-h}, \quad (8)$$

where  $N$  is the sample size,  $m$  is the number of autocorrelation lags, and  $\rho_h$  is the sample autocorrelation at lag  $m$ . Under the null hypothesis, the asymptotic distribution of  $Q$  is chi-square with  $m$  degrees of freedom. If use `lbqtest` function in the MATLAB Econometrics Toolbox, it returns the rejection decision and  $p$ -value for the hypothesis test.  $P$ -values indicate the strength at which the test rejects the null hypothesis. If all the  $p$ -values are larger than 0.01, there is strong evidence to accept the hypothesis that the residuals are not autocorrelated.

Thus, we have generated ARFIMA(0,0.4,0) processes in Fig. 5 and use fractional order difference filter with the order  $d = 0.4$  to filter the LRD property in the Fig. 6. It is obvious that

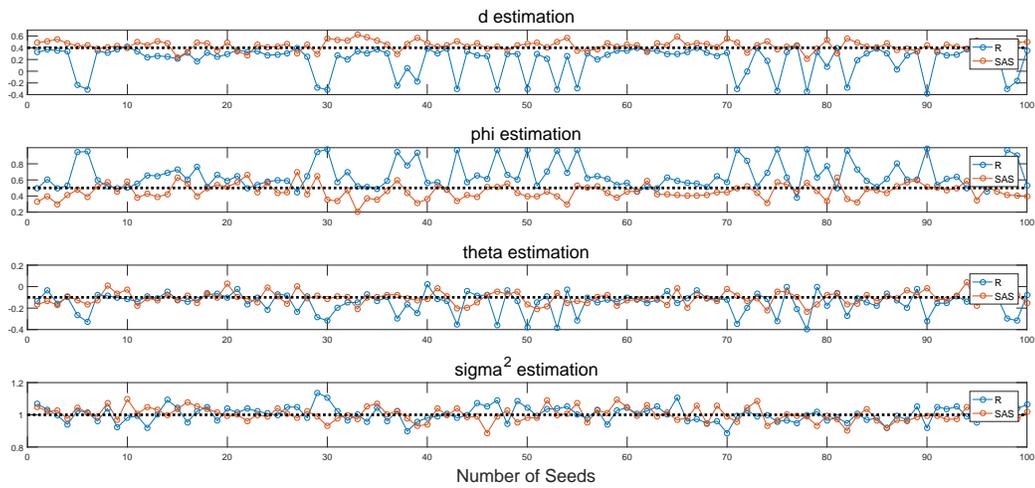


FIGURE 2. ESTIMATION RESULTS OF SAS AND R

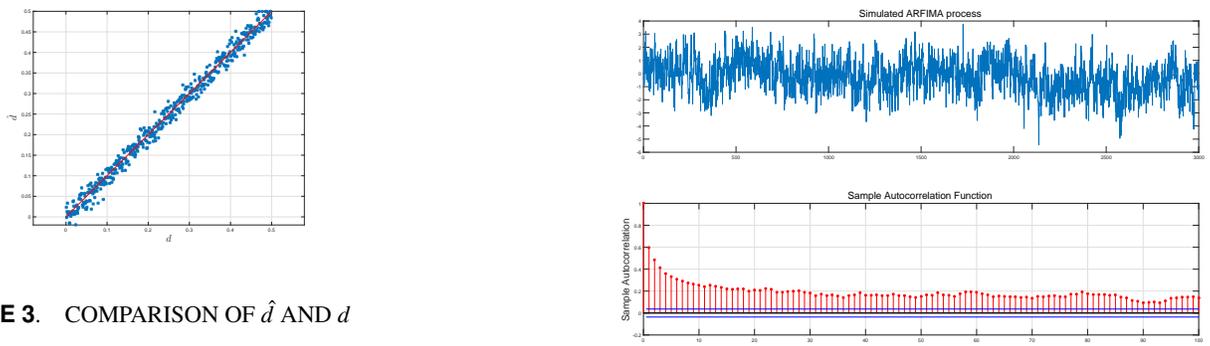


FIGURE 3. COMPARISON OF  $\hat{d}$  AND  $d$

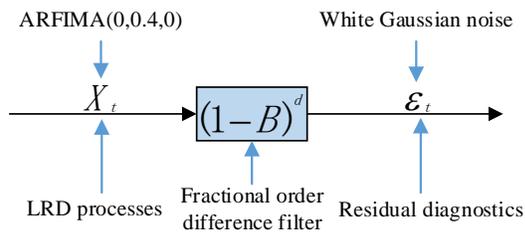


FIGURE 4. FRACTIONAL ORDER DIFFERENCE OF THE ARFIMA(0,0.4,0) PROCESS

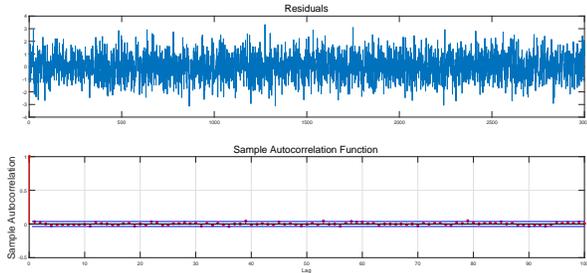
by passing through the fractional order difference filter, the slow decaying property of LRD has been eliminated.

In order to evaluate the residuals, we choose to use  $p$ -values to quantify the goodness of fitting in the Tab. 3.

FIGURE 5. SIMULATION OF ARFIMA(0,0.4,0) PROCESS

Comments: The above four software all have fractional order operator to filter the LRD process successfully. In general,  $d$  is the parameter to estimate first. If use the calculation defined by the Hurst method,  $d$  could probably be the fractional one. Therefore, the fractional order filter would be the primary tool to eliminate the LRD property or the heavy-tailedness in order to get the stationary series.

Meanwhile, however, the fractional order  $d$  is closely related to the Hurst parameter in Eq. 7. There are more than ten methods to estimate Hurst parameters, R/S method, aggregated variance method, absolute value method, periodogram method, whittle method, Higuchi's method etc. These methods are mainly useful as simple diagnostic tools for LRD time series. These Hurst estimators have been introduced to analyze the LRD time series in [28–30]. Therefore, the results of Hurst estimators can be dif-



**FIGURE 6.** ARFIMA(0,0.4,0) PROCESS PASSING THROUGH THE FRACTIONAL DIFFERENCE FILTER

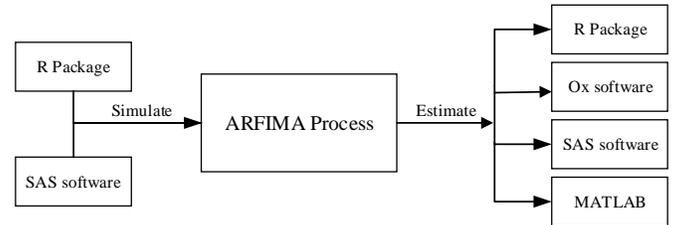
**TABLE 3.** FRACTIONAL FILTERS

	MATLAB	R	SAS	Ox
Function	d_filter	diffseries	fdif	N/A
	0.0710	0.09998	0.1062	N/A
<i>p</i> -values with	0.2253	0.2395	0.2414	N/A
1,5,10,15 lags	0.5850	0.5320	0.5198	N/A
	0.5330	0.4571	0.4473	N/A

ferent if applying different methods. In addition, from Eq. 5, it is interesting to note that, there are infinite factorial series in the expansion of binomial expansion. In practice, we usually take the first three factorials for approximation. That is to say, the accuracy of differencing is also determined by how many factorials are used for approximation. As a result, the different methods make the subsequent estimation differ from each other in the following section.

### Parameter Estimation

From Fig. 2, we could see that even though R and SAS can both simulate the ARFIMA processes, the properties of these processes are not the same mainly because of the distinctive random seeds defined by different software. Therefore, when reviewing and evaluating the accuracy of above software, the same ARFIMA series should be guaranteed first. Herein, we have proposed to use the following steps to compare the results in the Fig. 7. In addition, OxMetrics is a software that cannot generate ARFIMA simulation, but it can estimate and forecast ARFIMA-FIGARCH processes. MATLAB cannot generate multiple ARFIMA series for the same parameter combinations. So we have used R and SAS to provide the ARFIMA series for the inputs of



**FIGURE 7.** SIMULATION AND ESTIMATION OF THE ARFIMA PROCESS

estimations. Since we have got the simulation results, the parameters of the ARFIMA processes can be estimated and compared with true values (parameter setting values). First, we have used the simulation data from R software, and then use these three software to do the estimation in the Fig. 8. Second, we have used SAS to do the same simulation, and then use these three software to do the estimation in the Fig. 9. Without loss of generality, we pick 10 groups of 3000 sampling points to see who could capture the accuracy.

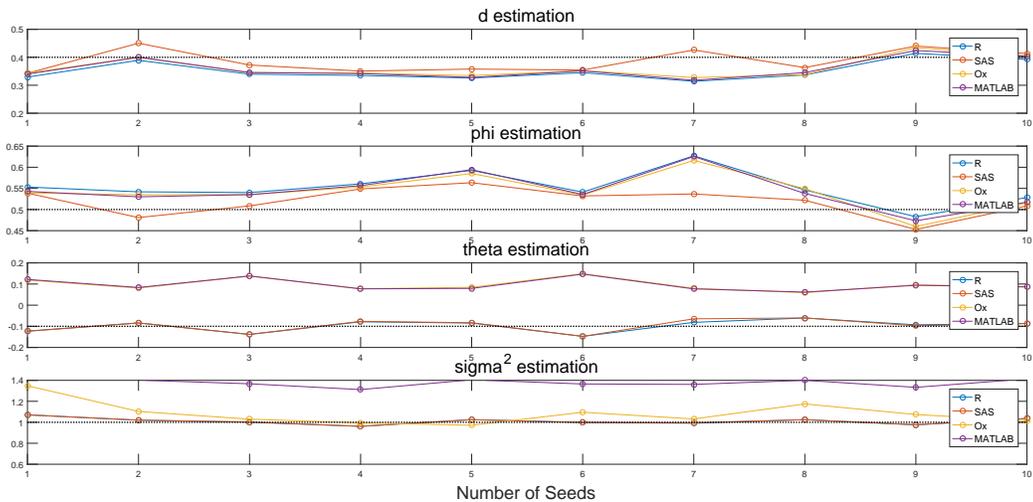
Comments: From above plots, it is very interesting to find that the estimation results of ARFIMA simulations are relatively accurate when they come from the same simulation data set. But Ox Metrics and MATLAB estimated the negative values of  $\theta$ .

### Forecast

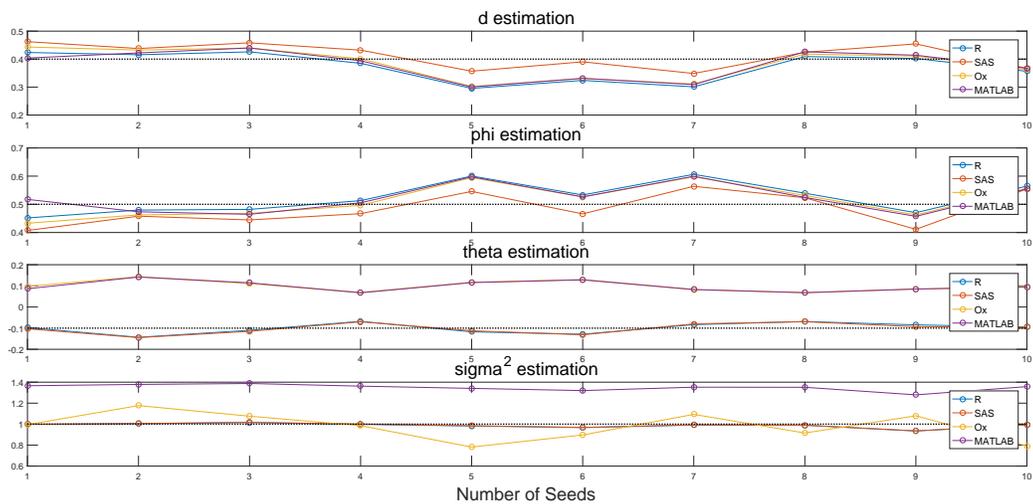
Simulation data could only be the auxiliary part of the software, it can never be a powerful and useful tool for the ARFIMA process analysis if the software cannot retrieve the estimation parameters from the real data with LRD. What's more, the last and the most significant part of the ARFIMA process is to forecast thereby predict the future behavior. Therefore, mean absolute percentage error (MAPE) values are used for the evaluation of the forecast results for the data from real life. The error square of the prediction results from different methods with the increasing number of predictions are illustrated in the Fig. 10.

$$\sigma_t^2 = \sum_{i=1}^n (\hat{y}_i - y_i)^2. \quad (9)$$

Data description: Centered annual pinus longaeva tree ring width measurements at Mammoth Creek, Utah, from 0 A.D to 1989 A.D with 1990 sampling points in time series [31, 32]. The data can be divided into two part, the first part with 1900 points are used to estimate ARFIMA parameters, and the second part with 90 points are used to compared with the prediction results from the fitted ARFIMA models. Finally, the results with the imple-



**FIGURE 8.** PARAMETER ESTIMATION WITH DIFFERENT METHODS (R SIMULATION INPUTS)



**FIGURE 9.** PARAMETER ESTIMATION WITH DIFFERENT METHODS (SAS SIMULATION INPUTS)

mented methods which are applied to real-life time series are summarized in the Tab. 4.

Here are some comments of this subsection:

I.  $d$  is the parameter to be estimated first when doing AR-FIMA model fitting. Therefore, if the estimation of  $d$  is different for a certain time series, the following estimations for AR( $\Phi$ ) and MA( $\Theta$ ) will be of difference.

II. The ideal length (horizon) of predictions is within 30 steps. With the increasing steps of forecast, prediction errors are adding up. If a long range prediction series is required, R and MATLAB should be priorities for their smaller prediction errors.

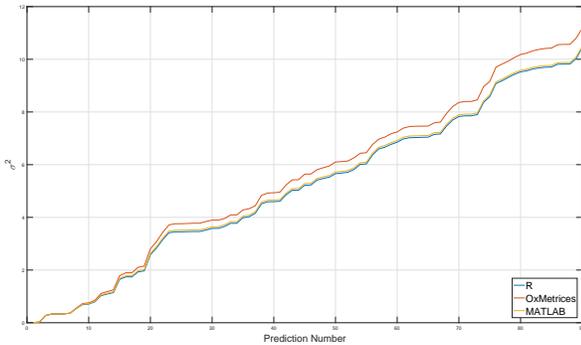
III. Compared with other forecast result with true values in 4, R produces the minimum prediction errors and MAPE.

## CONCLUSIONS

Compared to the conventional integer order models which can only capture SRD, the ARFIMA model gives a better fit and result when dealing with the data which possess the LRD property. Nowadays, several software have integrated with AR-FIMA solutions. However, the final results of estimation and prediction could be different or conflicting if choosing different

**TABLE 4.** PARAMETER ESTIMATIONS AND FORECAST COMPARISONS

	Parameters	R	SAS	OxMetrics	MATLAB
1	mu	0.9833	N/A	0.98799	0.9878
2	d	0.1670	0.1479624	0.282087	0.2313
3	ar	0.9070119	0.8939677	-0.254265	0.6473
4	ma	0.8603811	0.8318787	0.18698	0.6393
5	sigma	0.1078173	0.1073417	0.1246	0.1163
6	pvalue Lag1	0.9195	N/A	0.7709458	0.9101
7	pvalue Lag5	0.6369	N/A	0.341324	0.6959
8	pvalue Lag10	0.8659	N/A	0.4367925	0.9037
9	pvalue Lag15	0.6491	N/A	0.6229542	0.6776
10	LogLikelihood	2117.224	1851.5512	-570.599	1162.527
11	MAPE	28.95	N/A	29.36	29.02

**FIGURE 10.** PREDICTION COMPARISON WITH DIFFERENT METHODS

methods. Therefore, a comprehensive review and evaluation of the numerical tools for ARFIMA processes is presented in the paper so as to provide some guidelines when choosing appropriate methods to do the analysis. Qualitative analysis as well as quantitative evaluations of the selected ARFIMA tools are conducted. Through such efforts, it is hoped that an informative guidance is provided to the readers when they face the problem of selecting a numerical tool for a specific application.

#### ACKNOWLEDGMENT

This work is supported by a scholarship from the China Scholarship Council (CSC).

#### Disclosure statement

No potential conflict of interest was reported by the authors. All results presented are reproducible and available upon request.

#### REFERENCES

- [1] Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M., 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- [2] Sheng, H., Chen, Y., and Qiu, T., 2011. *Fractional processes and fractional-order signal processing: techniques and applications*. Springer Science & Business Media.
- [3] Sun, R., Chen, Y., and Li, Q., 2007. "Modeling and prediction of Great Salt Lake elevation time series based on ARFIMA". In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 1349–1359.
- [4] Li, Q., Tricaud, C., Sun, R., and Chen, Y., 2007. "Great Salt Lake surface level forecasting using FIGARCH model". In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 1361–1370.
- [5] Sheng, H., and Chen, Y., 2009. "The modeling of Great Salt Lake elevation time series based on ARFIMA with stable innovations". In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 1137–1145.

- [6] Contreras-Reyes, J. E., and Palma, W., 2013. “Statistical analysis of autoregressive fractionally integrated moving average models in R”. *Computational Statistics*, **28**(5), pp. 2309–2331.
- [7] Baillie, R. T., and Chung, S., 2002. “Modeling and forecasting from trend-stationary long memory models with applications to climatology”. *International Journal of Forecasting*, **18**(2), pp. 215–226.
- [8] Doornik, J. A., and Ooms, M., 2003. “Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models”. *Computational Statistics & Data Analysis*, **42**(3), pp. 333–348.
- [9] Doornik, J. A., and Ooms, M., 1999. “A package for estimating, forecasting and simulating ARFIMA models: Arfima package 1.0 for Ox”. Preprint, Erasmus University.
- [10] Doornik, J. A., Ooms, M., et al., 2004. “Inference and forecasting for ARFIMA models with an application to US and UK inflation”. *Studies in Nonlinear Dynamics and Econometrics*, **8**(2), pp. 1208–1218.
- [11] Hurst, H. E., 1951. “Long-term storage capacity of reservoirs”. *Transactions of the American Society of Civil Engineers*, **116**, pp. 770–808.
- [12] Karasaridis, A., and Hatzinakos, D., 2001. “Network heavy traffic modeling using  $\alpha$ -stable self-similar processes”. *IEEE Transactions on Communications*, **49**(7), pp. 1203–1214.
- [13] Sheng, H., and Chen, Y., 2011. “FARIMA with stable innovations model of Great Salt Lake elevation time series”. *Signal Processing*, **91**(3), pp. 553–561.
- [14] Granger, C. W., and Joyeux, R., 1980. “An introduction to long-memory time series models and fractional differencing”. *Journal of Time Series Analysis*, **1**(1), pp. 15–29.
- [15] Hosking, J. R., 1981. “Fractional differencing”. *Biometrika*, **68**(1), pp. 165–176.
- [16] Brockwell, P. J., and Davis, R. A., 2013. *Time series: theory and methods*. Springer Science & Business Media.
- [17] Reisen, V., Abraham, B., and Lopes, S., 2001. “Estimation of parameters in ARFIMA processes: A simulation study”. *Communications in Statistics-Simulation and Computation*, **30**(4), pp. 787–803.
- [18] Fatichi, S., 2009. ARFIMA simulations. See also URL: [https://www.mathworks.com/matlabcentral/fileexchange/25611-arfima-simulations/content/ARFIMA\\_SIM.m](https://www.mathworks.com/matlabcentral/fileexchange/25611-arfima-simulations/content/ARFIMA_SIM.m), Oct.
- [19] Caballero, C. V. R., 2015. ARFIMA ( $p, d, q$ ). See also URL: [https://www.mathworks.com/matlabcentral/fileexchange/53301-arfima-p-d-q-/content/dgp\\_arfima.m](https://www.mathworks.com/matlabcentral/fileexchange/53301-arfima-p-d-q-/content/dgp_arfima.m), Sep.
- [20] Inzelt, G., 2011. ARFIMA ( $p, d, q$ ) estimator. See also URL: <https://www.mathworks.com/matlabcentral/fileexchange/30238-arfima-p-d-q--estimator>, Jun.
- [21] Constantine, W., Percival, D., Constantine, M. W., and Percival, D. B., 2016. The fractal package for R. See also URL: <https://cran.r-project.org/web/packages/fractal/fractal.pdf>, May.
- [22] Maechler, M., Fraley, C., and Leisch, F., 2012. The fracdiff package for R. See also URL: <https://cran.r-project.org/web/packages/fracdiff/fracdiff.pdf>, Dec.
- [23] Contreras-Reyes, J. E., Goerg, G. M., and Palma, W., 2012. The afmtools package for R. See also URL: <http://www2.uaem.mx/r-mirror/web/packages/afmtools/afmtools.pdf>, Dec.
- [24] Kraft, P., Weber, C., and Lebo, M., 2015. The ArfimaMLM package for R. See also URL: <https://cran.r-project.org/web/packages/ArfimaMLM/ArfimaMLM.pdf>, Jan.
- [25] Veenstra, J. Q., and McLeod, A., 2015. The arfima package for R. See also URL: <https://cran.r-project.org/web/packages/arfima/arfima.pdf>, Dec.
- [26] Shumway, R. H., and Stoffer, D. S., 2010. *Time Series Analysis and Its Applications: With R Examples*. Springer Science & Business Media.
- [27] Ljung, G. M., and Box, G. E., 1978. “On a measure of lack of fit in time series models”. *Biometrika*, **65**(2), pp. 297–303.
- [28] Sheng, H., Chen, Y., and Qiu, T., 2011. “On the robustness of Hurst estimators”. *IET Signal Processing*, **5**(2), pp. 209–225.
- [29] Ye, X., Xia, X., Zhang, J., and Chen, Y., 2012. “Effects of trends and seasonalities on robustness of the Hurst parameter estimators”. *IET Signal Processing*, **6**(9), pp. 849–856.
- [30] Chen, Y., Sun, R., and Zhou, A., 2007. “An improved Hurst parameter estimator based on fractional Fourier transform”. In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 1223–1233.
- [31] Palma, W., 2007. *Long-memory time series: theory and methods*, Vol. 662. John Wiley & Sons.
- [32] Palma, W., and Olea, R., 2010. “An efficient estimator for locally stationary gaussian long-memory processes”. *The Annals of Statistics*, **38**(5), pp. 2958–2997.