REVIEW ARTICLE

Autoregressive models in environmental forecasting time series: a theoretical and application review

Jatinder Kaur[1,2] · Kulwinder Singh Parmar[2] · Sarbjit Singh[3]

Abstract

Though globalization, industrialization, and urbanization have escalated the economic growth of nations, these activities have played foul on the environment. Better understanding of ill effects of these activities on environment and human health and taking appropriate control measures in advance are the need of the hour. Time series analysis can be a great tool in this direction. ARIMA model is the most popular accepted time series model. It has numerous applications in various domains due its high mathematical precision, flexible nature, and greater reliable results. ARIMA and environment are highly correlated. Though there are many research papers on application of ARIMA in various fields including environment, there is no substantial work that reviews the building stages of ARIMA. In this regard, the present work attempts to present three different stages through which ARIMA was evolved. More than 100 papers are reviewed in this study to discuss the application part based on pure ARIMA and its hybrid modeling with special focus in the field of environment/health/air quality. Forecasting in this field can be a great contributor to governments and public at large in taking all the required precautionary steps in advance. After such a massive review of ARIMA and hybrid modeling involving ARIMA in the fields including or excluding environment/health/atmosphere, it can be concluded that the combined models are more robust and have higher ability to capture all the patterns of the series uniformly. Thus, combining several models or using hybrid model has emerged as a routinized custom.

Keywords Time series analysis · Air Quality Index (AQI) · Autoregressive models · Statistical modeling · Forecasting

Introduction

Though globalization, industrialization, and urbanization have escalated the economic growth of nations, these activities have played foul on the environment. Better understanding of ill effects of these activities on environment and human health and taking appropriate control measures in advance are the need of the hour. For this the relevant data could be analyzed through time series. Time series forecasting is an extensive quantitative technique involving collection and analyzation of historical observations for the development of an appropriate model. Theoretically, analysis of time series basically contains three steps: characterization, modeling, and forecasting. While forecasting calculates short-term progression of the system, the modeling component establishes long-term behavioral features of the system. The first step determines the fundamental properties such as measure of randomness or degrees of freedom. This analysis can be employed to build predictive models with minimum errors for forewarning. It is important to underline that the modeling choice in any temporal, or more generally, in any spatio-temporal prediction, is relevant and must be suitably faced (De Iaco and Posa 2018; Cappello et al. 2018; De Iaco et al. 2013 2015).

Autoregressive integrated moving average (ARIMA) model is the most popular accepted time series models (Shahwan and Odening 2007; Singh et al. 2020b). The Box-Jenkins methodology (Box and Jenkins 1970), high

✉ Kulwinder Singh Parmar
  kulmaths@gmail.com

1 Department of Mathematics, Guru Nanak Dev University College Verka, Amritsar, Punjab, India 143501

2 Department of Mathematics, I.K. Gujral Punjab Technical University, Jalandhar, Punjab, India 144603

3 Department of Mathematics, Guru Nanak Dev University College, Narot Jaimal Singh, Pathankot Amritsar, Punjab, India 145026

mathematical precision, and reliability are what makes ARIMA models very popular (Singh et al. 2021b). ARIMA models have large number of applications. The model is applied to forecast various things like commodity prices (Weiss 2000); for load forecasting in the power system (Nicolaisen et al. 2000; Hippert et al. 2001); future of energy resources, such as oil (Morana 2001) or natural gas (Buchananan et al. 2001); daily environmental factors such as ozone levels (Robeson and Steyn 1990; Prybutok et al. 2000); forecasting various air pollutant (Kulkarni et al. 2018; Chaudhuri and Dutta 2014), noise pollution data (Garg et al. 2015), water quality time series data (Faruk 2010), for water quality management (Parmar and Bhardwaj 2014); etc. Thus, we find large applications of ARIMA almost in every field.

ARIMA and environment are highly correlated. Forecasting is needed in all the fields of environment such as air pollution, noise pollution, fossil fuels, rainfall data, and underground waters, as all these factors have direct relationship with health. Forecasting in these fields can result in forewarning which can be highly beneficial. Though there are many research papers on application of ARIMA in various fields including environment, there is no substantial work that reviews the building stages of ARIMA. In this regard, the present work attempts to present three different stages through which ARIMA was evolved. It needs to be highlighted that detailed information about the process of evolution is equally significant as knowledge about its application. In this review paper, we have tried to review more than 100 papers based on pure ARIMA and its hybrid modeling. The first part of this paper deals with general introduction to ARIMA, followed by its historical overview, and the last part deals with its application review. The last section is divided into two categories, where major emphasis is laid on application of ARIMA in the field of atmosphere/environment/health or factors influencing air quality such as air pollutants, noise pollution, and rainfall, and the last section deals with application of ARIMA in other fields such as financial data and load forecasting data. The last categories are again divided into two further subsections: one involving purely ARIMA models and the other development of hybrid models with ARIMA.

ARIMA model analysis

The classes of autoregressive moving average (ARMA) models are frequently used while modeling linear and stationary time series due to their outstanding results and effectiveness (Al-Saba and El-Amin 1999). In 1921, Yule presented pure moving average process, whereas he introduced pure autoregressive process in 1927. Box and Jenkins (1970), Hannan (1970), and Anderson (1971) have been pioneers in building various techniques using autoregressive

moving average models. Hipel et al. (1977) have given the theoretical and practical approaches for the model building stages of ARIMA.

ARIMA processes are a kind of stochastic techniques which are used to investigate behavioral pattern of time series. ARIMA models are quite flexible in nature as they can represent pure autoregressive (AR), pure moving average (MA), and mixed AR and MA (ARMA) series. Unfortunately, ARIMA models are unable to capture nonlinear pattern of the series and thus are not suitable for approximating complex real-world problems. Time series which has either a trend or seasonal patterns do not exhibit stationary behaviors. ARIMA (p, d, q) model only captures trends and not seasonal behavior of time series. To model a seasonal pattern, we have *ARIMA* $(p,d,q)(P,D,Q)$ model.

Mathematical formulation

Moving average (*MA*) process

A process $\{z_t\}$ is said to be a moving average process of order q if.

$z_t = a_t - \theta_{t-1} - \ldots - \theta_q a_{t-q}$ where θ_i; $i = 1,2,3\ldots,q$ are constants and $\{a_t\}$ is a purely random process with mean zero and variance σ^2 (Box et al. 1994).

Autoregressive (*AR*) process

A process $\{z_t\}$ is regarded as an autoregressive process of order p if.

$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t$ where ϕ_j $(j = 1,\ldots,p)$ are constants and $\{a_t\}$ is a purely random process with mean zero and variance σ^2 (Box et al. 1994). This model works like multiple regression model.

Autoregressive moving average (*ARMA*) process

Both the autoregressive (*AR*) and the moving average (*MA*) are combined to build the autoregressive moving average (*ARMA*) model. The (*ARMA*)(p, q) model for a time series which contains p (*AR*) terms and q (*MA*) terms can be expressed as.

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t \ ; t = 1, \ldots, T$$

Here, a_t is known as normal white noise process. It has zero mean and variance σ^2. T is the amount of data in the time series (Box et al. 1994). *AR* parameters should satisfy the condition for stationarity and *MA* parameters should satisfy the conditions for invertibility.

Seasonal ARIMA

ARIMA as such does not support seasonal data, i.e., time series with repeating cycles. However, such time series are expressly modeled by ARIMA extension known as seasonal ARIMA, i.e., SARIMA. Seasonal ARIMA models are defined by 7 parameters $p, d, q, P, Q, D, and\ s$ and mathematically defined as

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)\left(1 - \beta_1 B^s - \beta_2 B^{2s} - \cdots - \beta_P B^{Ps}\right)(1 - B)^d (1 - B^s)^D y_t = c$$
$$+ \left(1 - \Psi_1 B - \Psi_2 B^2 - \cdots - \Psi_q B^q\right)\left(1 - \theta_1 B^s - \theta_2 B^{2s} - \cdots - \theta_Q B^{Qs}\right)\epsilon_t$$

- p and P are non-seasonal and seasonal autoregressive polynomial orders, respectively.
- q and Q are non-seasonal and seasonal moving average polynomial orders, respectively.
- d and D are order of normal and seasonal differencing, respectively.
- s is the period of the seasonal pattern appearing.

The *ARMA* models work only on stationary data, but in reality, data of the various fields is non-stationary, making *ARMA* unfit for these problems. With the help of differencing, the data can be made stationary, and this step leads to the development of *ARMA*. *ARMA* is in fact generalization of *ARMA* processes.

Sometimes by using visualization tools such as ACF and PACF, it is not possible to identify the parameters p, d, q and P, D, Q. In that case, the model with the lowest BIC (Bayesian information criterion) or AIC (Akaike information criterion) is selected. Based on different information theoretic techniques such as AIC or minimum description length (MDL), various methods have been proposed by researchers Hurvich and Tsai (1989), Ljung (1987), and Shibata (1976) for order selection.

Astrom and Eykhoff (1971), Van Boom and Enden (1974), and Unbehauen and Göhring (1974) acquainted on time series model identification in the engineering field. Since the beginning of the 1970s, various estimation-type identification methods have come into limelight. Akaike (1969) introduced final

Historical overview

Based on the principle of parsimony, statisticians Box and Jenkins (1970) gave a practical approach to build *ARMA* model. The method uses a three-step iterative approach of model identification, parameter estimation, and diagnostic checking to build the best parsimonious model from a general class of *ARMA* models (Fig. 1). The process is repeated until a satisfactory model is obtained which can be then used for prediction (Singh et al. 2020a, b, 2021a, b). In this section, historical evolution of various steps involved in *ARMA* modeling is discussed in detail.

Model identification

Model identification is a herculean task in any *ARMA* modeling. In this regard, autocorrelation function (ACF) and partial autocorrelation function (PACF) are vital statistics in determining the order of the model. While ACF explains correlation, PACF describes partial correlation between the series and lags of itself. Box and Jenkins (1970) introduced the concept of degree of differencing "d" also. Generally, d is 0, 1, or 2. After d is selected, p and q are calculated from the overall trend of ACF and PACF of the appropriately differenced series. Cleveland (1972) presented inverse autocorrelation function (IACF) for this step. But the method was not suitable for mixed models.
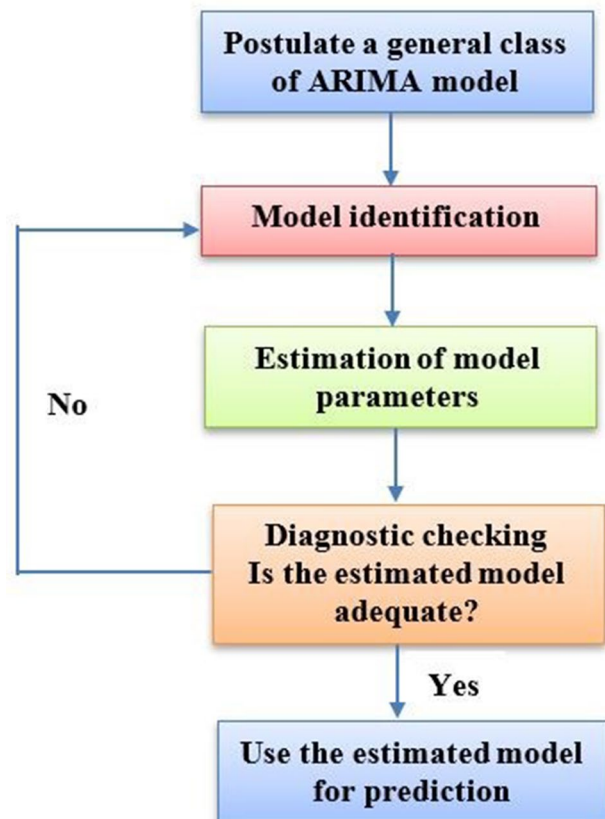

Fig. 1 Methodology of *ARIMA* model

prediction error (FPE) to determine the order p of the AR models which is defined as

$$FPE_{(P)} = (1 + \frac{2p}{n})\sigma_p^2 \qquad (1)$$

where n is the number of observations and $\hat{\sigma}^2_{(P)}$ is an estimate of white noise variance.

For the mixed model, the criteria are expressed as

$$\delta(p, q) = log(\hat{\sigma}^2) + \frac{(p + q)g(n)}{n} \qquad (2)$$

Here, $\hat{\sigma}^2$ represents the maximum likelihood estimate for the residual variance σ^2. The values $\hat{p}$ and $\hat{q}$ minimizing δ (p, q) are the best approximations for p and q. Taking different values of g(n) leads to different criteria (Table 1).

Another procedure known as criterion of autoregressive transfer function (CAT) was introduced by Parzen (1974) where the actual model is presumed to have an AR ($\propto$) representation. The order selected $\hat{p}$ is interpreted as best finite order AR approximation to the true AR ($\propto$) process. Extending the work of Woodward and Gray (1981), Glasbey (1982) introduced generalization of partial autocorrelations (GPAC) as crucial techniques in ARMA model identification. To estimate d, Janacek (1982) proposed a method using log of the power spectrum. For order estimation of a finite moving average process, Bhansali (1983) gave autoregressive and window estimates of the inverse correlation function. Monahan (1983) used Bayesian approach for determining the order (*p, q*) of the *ARMA* model. Pattern identification techniques using extended Yule-Walker equations for *ARMA* order identification was employed during early 1980s. Tucker (1982) replaced R and S arrays with RS array for the *ARMA* model identification. By integrating order identification approach for mixed stationary and non-stationary *ARMA* models, Tsay and Tiao (1984) eliminated the need of differencing required for producing stationarity in the time series. Augmenting their study, Tsay and Tiao (1985) introduced the smallest canonical correlation (SCAN) method. Broersen (1985) presented weak parameter criterion (WPC) for model order selection. Poskitt (1987) modified Hannan and Rissanen (1982) criterion of order selection. The order of *ARMA* model was determined using white noise test by Pukkila and Krishnaiah

(1988). Koehler and Murphree (1988) preferred Schwarz information criterion (SIC) over Akaike information criterion (AIC) for order selection. Hurvich and Tsai (1989) developed a bias-corrected method AlCc for better model order choices. Koreisha and Pukkila (1993) introduced iterative procedures for determining the degree of differencing required to make time series data stationary. Zhang and Zhang (1993) developed an algorithm involving only autocorrelations for determining order of MA processes. Liang et al. (1993) gave a new approach based on the Eigen values of covariance matrix for *ARMA* model order determination. Sreenivasan and Sumathi (1997) formulated a new generalized parameters technique for both seasonal and non-seasonal *ARMA* model identification.

Model estimation

The estimation of AR parameters is very crucial in time series analysis for the adequate information about the model. Maximum likelihood methods, ordinary least squares (OLS), and method of moments are some of the extensively used techniques for parameter estimation in time series analysis. Pure autoregressive models are either estimated by OLS method or by using Levinson (1947)-Durbin (1960) algorithm. Durbin (1960) showed that method of OLS leads to optimum estimates of the model provided errors are normally distributed. Burg (1975) came with an algorithm similar to forward backward prediction especially for short records and Huzii (1981) proposed a method based on higher order moment for estimating AR coefficients. Proposing an alternative to the Burg's estimates Pukkila and Krishnaiah (1988) calculated true correlation matrix of the lagged variables. Basu and Das (1992) consider optimality of the maximum likelihood estimator under a general set-up of roots of the characteristic equation of the p th order of autoregressive process.

In comparison to the parameter estimation methods available for AR model, the number of techniques to approximate MA parameters is fewer. Durbin (1959) is pioneer in the estimation of MA parameters with a simple estimation procedure. A quadratically convergent algorithm was formulated by Wilson (1969) to estimate the parameters of a MA process. Whereas Gauss–Newton method was used

Table 1 Model identification criterions depending upon values of *g*(n) along with their mathematical expressions

Value of g(n)	Criterion	Mathematical expression
g(*n*) = 2	Akaike information criterion (AIC) given by Akaike (1974)	$AIC(p, q) = \log(\sigma^2) + \frac{2(p+q)}{n}$
g(*n*) = log*n*	Bayesian information criterion (BIC) given by Schwarz (1978)	$BIC(p, q) = \log(\sigma^2) + \frac{(p+q)\log n}{n}$
g(*n*) = *c*log*n*	Hannan and Quinn's criterion (HQ) given by Hannan and Quinn 1979	$HQ(p, q) = \log(\sigma^2) + \frac{(p+q)\log\log n}{n}$ where c is a constant to be specified

to estimate the parameters of a non-linear function, Fuller (1976) utilized it for the MA process. Godolphin (1977, 1978) proposed an alluring computational procedure augmenting the previous studies.

For *ARMA* cases, myriad studies are available. Wilson (1969) and Marquardt (1963) have developed an algorithm to estimate the *ARMA* parameters which was employed by Box and Jenkins (1970) to lay the foundation of all the *ARMA*, AR and MA processes. The research done by the latter is regarded to be a milestone in the field of modeling. McLeod (1977) gave an easier implementing modified version of Box and Jenkins (1970) which approximated exact ML estimators very precisely. Tuan (1984) derived many recursive relations which worked as a tool in identifying the order and parameters of *ARMA* at its preliminary stages. Using autocovariance function, Choi (1986) presented an algorithm for the parameter estimation of stationary *ARMA* process. The iterative algorithm proposed by Choi in 1986 was not convergent; therefore, he developed a convergent Newton–Raphson solution for MA parameters in the subsequent year. Saikkonen (1986) derived two-step estimators which were asymptotically efficient.

Recently, different studies have focused on maximum likelihood (ML) procedures as a tool to estimate *ARMA* models. Basu and Das (1991) analyzed the asymptotic properties of least-squares estimation procedures of the parameters of an *ARMA* (p, q) process in the stable case. The SCA System can be used to estimate the parameters of the model. For best results, a conditional likelihood function is selected along with the detection and adjustment of outliers. Mikosch et al. (1995) derived the asymptotic properties of the estimators. The major drawback of ML estimates is that it often lies outside the invertibility region. However, the generalized least-squares method for the estimation of pure MA and *ARMA* model presented by Pukkila et al. (1990) has succeeded in overcoming it.

Model diagnostic checking

Before the remarkable advent of Box and Jenkins (1970) procedure, hypothesis testing methods were largely used for *ARMA* model identification. McLeod (1974) examined the need to check whiteness and homoscedasticity of the residuals in the diagnostic checking step. Godfrey (1979) proposed a new approach based upon Silvey's (1959) Lagrange multiplier method to check the adequacy of the *ARMA* model. Hokstad (1983) has proposed a diagnostic test with the estimated cross correlation function (CCF) between the observed values and the residuals. The CCF can also be used as an indicator for the required improvement in the model.

Model forecasting

Among the wide applications of time series, the most popular is forecasting. Forecasting is easily attainable with state space framework of which *ARIMA* models are special case. With the state space framework, observational vectors are brought into a system with one element at a time (Durbin and Koopman 2012). State space models have great contribution in the science of environment (Harvey et al. 2004) and also play a pivotal in the analytical handling of time series models (Harvey 1989).

For the practical computation of forecasting, the simplest and most elegant method is of difference equation from which minimum mean square error forecasts can be directly generated (Box and Jenkins 1970). Also, the probability limits for the forecasts can be obtained by solving recursively. Makridakis and Wheelwright (1977) concluded that the adaptive filtering technique can be applied to time series forecasting dealing with real data. Cartwright (1985) has assessed the forecasting performance of Priestley's model and concluded that forecasting errors can be significantly reduced with the use of broader classes of time series. Ray (1988) concluded that forecasting performance of Bilinear model is more than Box-Jenkins model and threshold autoregression model. Tiao and Tsay (1994) studied developments of time series in both the linear and non-linear domain. They were of the opinion that when the parameters involved are estimated adaptively, linear models provide more accurate forecasts. Thus, we can see the model building process is quite laborious and needs great human expertise.

Wrapping up all the above steps, the general statistical methodology scheme of *ARIMA* is as presented by Fig. 1.

Modeling and forecasting work flow of *ARIMA*

As observed *ARIMA* finds large applications in various fields due its high mathematical precision and great reliability. The next section of the paper deals with review of papers employing *ARIMA* exclusively and hybrid *ARIMA* models published in world's best-class journals. The rising disastrous threats to the atmosphere or air quality are turning to be great concerns of the twenty-first century. Its hazardous effects are felt by human beings as well as by the entire wildlife. Thus, forecasting in this field can be a great contributor to governments and public at large in taking all the required precautionary steps in advance. There are various environmental factors affecting the air quality such as air pollutants, water pollutants, noise pollution due to traffic congestions, rainfalls, and surface erosions. Also, the environment exhibits close connections with energy resources. Thus, our coming section deals with all these

environmental factors first forecasted with *ARIMA* alone followed by hybrid models involving *ARIMA*. The data set undertaken, methodology adopted, and results obtained, all are discussed elaborately for the complete understanding of the readers.

Survey of air quality/environment/health data involving application of *ARIMA* models

Kulkarni et al. (2018) studied the variation of SO_2, RSPM, NOx, and SPM parameters present in the atmosphere of Nanded City (Maharashtra, India). The randomness, trend, and seasonality present in the data were also analyzed. The forecasted result revealed that RSPM and SPM are exceeding the permitted limits. Jaiswal et al. (2018) observed the statistical trend of CO, NO_2, SO_2, $PM_{2.5}$, and PM_{10} concentrations for the time span January 2013 to December 2016 for the city Varanasi (India) using Mann–Kendall and Sen's slope estimator approach. Different *ARIMA* models, namely, *ARIMA* (1,0,0), (1,0,1), and (1,1,1), were fitted on the three data sets of summer, monsoon, and winter season and their results compared. *ARIMA* (1,1,1) was chosen as the best fit model for forecasting all the pollutants. Pohoata and Lungu (2017) tried to analyze the air quality of the city Ploiesti in Romania for the pollutants O_3, CO, NO_2, NOx, and PM_{10}. While *ARIMA* (3,1,3) provided good results for NOx, NO_2 and O_3, it failed to give satisfactory results for CO and PM_{10}. Kumar et al. (2004) forecasted one-day advance O_3 concentration in Brunei Darussalam using *ARIMA* modeling approach. *ARIMA* (1,0,1) was the best fitted method. Liu et al. (2018) used *ARIMA* with numerical forecasts (ARIMAX) with the vision to improve forecast of O_3, $PM_{2.5}$, and NO_2 for Xingtai (China). Significant reduction in RMSE, viz., 47.8–49.7%,14.3–21.0%, and 41.2–46.3%, respectively, for O_3, $PM_{2.5}$, and NO_2, was seen by employing CMAQ-ARIMA for the daily 1-h and 8-h forecasting values at all the stations. Dynamic hourly forecast shows that ARIMAX can also be successfully applied to forecast of 7- to 72-h $PM_{2.5}$, 4- to 72-h NO_2, and 4- to 6-h O_3. Zhang et al. (2018) forecasted $PM_{2.5}$ concentrations using AQI and meteorological parametric data for Fuzhou (China) with the help of *ARIMA*. It was observed that $PM_{2.5}$ concentrations were positively correlated with PM_{10}, NO_2, and SO_2 concentrations and negatively correlated with meteorological parameters. Average $PM_{2.5}$ concentrations were 52% higher in cold periods in comparison to warm periods. *ARIMA* (6,1,1) was found to be best model for the data.

This is the first applicability of *ARIMA* model on the dataset of VC. Prior to model simulation of VC over the region of Delhi, trends and variations in the data set were analyzed. 12 *ARIMA* models *ARIMA* (0,0,1), (1,0,2), (0,0,5), (1,0,0), (0,0,1), (1,0,0), (0,0,1), (0,0,1), (1,0,0), (0,0,2), (0,0,1), and (0,0,3), respectively, for each month from January to

December were developed separately. The result revealed that past continued to impact future values of VC.

Kumar and Goyal (2011) thrived to build a forecasting AQI model. Three models are assessed, i.e., *ARIMA* (1,1,1), PCR, and ARIMA combined with PCR with respect to 4 seasons of the year in Delhi. The hybrid model performed better in predicting AQI one-day advance. The hourly and monthly concentrations of CO spread over 7 years of data of Hong Kong were analyzed by Lau et al. (2009) using *ARIMA* modeling. Association of hourly concentration of CO with different days of the week was examined. The hourly data of CO was like traffic data. This strong association was proven by SARIMA $(0, 1, 1) (0, 1, 1)_{24}$ model. It was also shown that data possessed long-term memory features. Kumar and De Ridder (2010) took the daily maximum O_3 concentration data of four sites of London and Brussels for study. They studied GARCH modeling technique in association with FFT-ARIMA to make forecasts of ozone episodes at these sites. In the study of Slini et al. (2002), *ARIMA* (1,1,1) and *ARIMA* (1,1,0) were used over the data of maximum daily O_3 concentration data of Athens (Greece) from 1990 to 1998. The forecasting performance of these two models were observed under the three categories of alarm limit greater than 180,170 and 160 μg/m^3. Robeson and Steyn (1990) tried to develop three predictive models, i.e., D/S model, *ARIMA* (1,1,0), and TEMPER model for the temporal variability of ozone in the lower Fraser Valley of British Columbia by taking 8 years data for the period 1978–1985 from two monitoring stations T9 and T11. On comparison of forecasting ability for ozone concentrations, TEMPER model outperformed the other two models. Siew et al. (2008) compared the performance of ARMA (3, 1, 3) and integrated ARFIMA (0, − 0.5, 2) models for the data of O_3, PM_{10}, NO_2, SO_2, and CO concentration from March 1998 to December 2003 of Selangor Malaysia. Though both models could not forecast all the values completely, ARFIMA performed slightly better than *ARIMA*.

Ahn (2000) applied second order differencing *ARIMA* models to daily groundwater head time series (1985–1990) from 7 monitoring wells located in Collier County, Florida. Variance and autocovariance equations were derived for the second-order time series models using *ARIMA* (0,2,1), (0,2,2), (1,2,0), (2,2,0), and (1,2,1) as function of parameters of the model under study. Mirzavand et al. (2014) worked on AR, MA, ARMA, *ARIMA*, and SARIMA models in analogous to forecast groundwater levels up to 60 mo in plain expanses of Kashan aquifer in Iran. The seasonality and stationarity of the data were checked. Taheri Tizro et al. (2014) analyzed several water parameters of Hor Rood River at Kakareza station with the help of *ARIMA* modeling. Based on R^2, AIC, RMSE, and MAPE, *ARIMA* (2,1,3), (2,1,3), (1,1,3), (1,1,3), (2,1,1), (2,1,1), (1,1,1), and (2,1,2) were found to be best suitable for parameters TDS, EC, HCO3-,

SO_4^{2-}, Ca^+, Na^+, ph, and SAR generation, respectively. The increasing trend of majority of parameters showed a picture of deteriorating water quality conditions in the region.

Garg et al. (2015) in their paper simulated daily mean LDay (06–22 h) and LNight (22–06 h) in A- and C-weightings in conjunction with single-noise metrics, day-night average sound level (DNL) for a period of 6 months for the station East Arjun Nagar in Delhi using *ARIMA* methodology. *ARIMA* (0,0,14), (0,1,1), (7,0,0), (1,0,0), and (0,1,14) are chosen to fit models for LDay dBA, LNight dBA, LDay DB(C), LNight dBC, and DNL dBA, respectively. Augmenting their work, the authors in 2016 compared *ARIMA* and ANN on the same problem and found ANN model gave better results than *ARIMA* model. Guarnaccia et al. (2017) presented two methods for the acoustic data set of Nice (France) international airport for the year 2000. The two methods utilized were DD-TSA (deterministic decomposition model) and SARIMA (0,1,1) (0,1,1)$_{24}$. While the former method captured long-term behavior of the data set, the latter captures short-term behavior. To quantify the forecasting errors, residual analysis was carried out. The DD-TSA gave slightly better results in terms of low standard deviation. Williams and Hoel (2003) applied SARIMA (1,0,1) (0,1,1)$_{672}$,(1,0,2)(0,1,1)$_{672}$, (2,0,1)(0,1,1)$_{672}$ for M25 station and $(1, 0, 1)(0, 1, 1)_{672}, (3, 0, 0)(0, 1, 1)_{672}, (1, 0, 2)(0, 1, 1)_{672}$ for I-75 stations on the vehicular traffic data. The authors concluded that one-step seasonal *ARIMA* predictions constantly outperformed *ARIMA* and random walk forecast results.

Ab Razak et al. (2018) using Mann–Kendall trend analysis found *ARIMA* (0, 1, 2) is the best suitable model for the daily and monthly rainfall and stream flow data of stations of Malaysia for the period 2000–2010. Zakaria et al. (2012) developed four *ARIMA* models (3,0,2)(2,1,1)$_{30}$, (1,0,1) (1,1,3)$_{30}$, (1,1,2)(3,0,1)$_{30}$, and (1,1,1)(0,0,1)$_{30}$ for the weekly rainfall data for the stations Sinjar, Mosul, Rabeaa, and Talafar in north west Iraq for the period 1990–2011 stations.

Benvenuto et al. (2020) found *ARIMA* (1, 0, 4) and *ARIMA* (1, 0, 3) to be the best model for determining the prevalence and incidence of COVID-19, respectively, from January 20, 2020, to February 10, 2020. Logarithmic transformation was done out to check the seasonality influence on the prediction.

Suresh and Priya (2011) took 57 years data from (1950–195 l) to (2007–2008) of sugarcane area, production, and productivity for Tamil Nadu for analyzing. Various *ARIMA* models with *p* and *q* varying from 0, 1, and 2 were fitted. While *ARIMA* (1, 1, 1) model was found suitable for sugarcane area and productivity, *ARIMA* (2,1, 2) was appropriate for sugarcane production.

Li and Li (2017) applied GM (1,1), *ARIMA* (1,2,0), metabolism GM (1,1), and GM-*ARIMA* to forecast future energy needs of Shandong province using energy data

for the span 1995–2015. Employing histogram judgment method, it was shown that data contains non-linear sequencing. This led to the development of GM (1,1) model on the data. Unfortunately, this method was also precluded because of 21 data points in the data set. The optimal requirement of data points for application of gray model is 5–10. In the next step, gray metabolic forecast model was tried. After the successful application of gray metabolic model on the basis of relative error, a series of residuals was obtained. Residual corrections are done with the help of *ARIMA* model. This process enhanced the prediction accuracy significantly, and this was the major innovation of this study. Aamir and Shabri (2016) employed *ARIMA*, GARCH, and *ARIMA*-Kalman to predict crude oil rates in Pakistan by undertaking average monthly prices of crude oil for the time February 1986 to March 2015. The *ARIMA* Kalman filter technique proved to be best approach as MAE and RMSE were minimum in this case as compared to *ARIMA* and GARCH.

It is very much evident from the above discussion that *ARIMA* is highly significant in the field of environment and health. *ARIMA* being simple and reliable has tremendous potential to forecast in these areas. Thus, it was important to review contribution of *ARIMA* in these domains.

Hybrid modeling

The real-world problems are usually complex rather than being simple. Thus, linear predictive methods do not perform as desired when used to process data from a non-linear system (Patil 1990). To find out whether the series is linear or non-linear requires enormous efforts of the forecasters. The researchers try different combination of models based on different theoretical and practical approaches and various other factors such as sampling variation, model structure and uncertainty to develop a model which yields more accurate results and enhanced forecasting (Jenkins 1982; Makridakis 1989). Bates and Granger (1969) are considered to be pioneer in introducing combining forecasts as an alternative to use one single forecast. The literature indicates that performance of time series forecasting increases through combining forecasts (Makridakis et al. 1982).

The next part of our study deals with review of papers based on *ARIMA* hybrid modeling in the field of environment/atmosphere/health.

Analyzing *ARIMA* hybrid-based studies

Mani and Volety (2021) forecasted three air pollutants, namely, CO, NH_3, and O_3 of Vijayawada station by employing both *ARIMA* and the LSTM models. Kalman filters are also used

to enhance the performance and forecasting abilities. The latter model proved to have higher accuracy using RMSE and MAE as performance indices. Wang et al. (2017) came up with hybrid-GARCH model to overcome conditional heteroscedasticity almost present in every hybrid model. The authors used *ARIMA* and SVM models to explain the linear and non-linear components, respectively, of the AQI data comprising $PM_{2.5}$, SO_2, NO_2, CO, and O_3 concentrations for six stations from the Shenzhen air quality monitoring network (China) for the time period 01 Sep 2013 to 10 Sep 2013. For estimating the coefficients of individual models, GARCH model is introduced. The accuracy level of hybrid-GARCH model in terms of metrices MAE and RMSE were higher than the individual models. Samia et al. (2012) attempted to foresee one day in advance the max 24-h ma PM concentrations in the region of Sfax Southern Suburbs using MLP, ARIMAX, and the hybrid model. The study revealed that hybrid ARIMAX-ANN outperformed the individual models since it can explore both linear and nonlinear patterns. Zhu et al. (2017) stated that the most challenging problem while forecasting AQI is of data being highly complex and non-stationary. Thus, they presented two hybrid models EMD-SVR-Hybrid and EMD-IMFs-Hybrid for AQI data of Xingtai, China, collected from June 2014 to August 2015. To obtain smooth IMF, EMD technique is employed. Then, SVR is used to predict total sum of IMF's and finally S-*ARIMA* is applied for analyzing residual sequences obtained from the two proposed models. In this paper where *ARIMA* is used for modeling AQI data, S-*ARIMA* is kept for forecasting IMF5 as well as for analysis of the residues. The predicted outcomes of AQI are sum of EMD-IMFs and S-*ARIMA*. The IMF4, IMF5, IMF6 and IMF7 chosen from IMFs are forecasted by Holt-Winters (0.9,0.2,0.3), S-*ARIMA* (1,1,1) (0,1,1), Holt Winters (0.1,0.2,0.5), and GM (1,1) respectively. Chelani and Devotta (2006) examined the effectivity of their proposed hybrid model based on chaos theory with *ARIMA* and nonlinear models for the time series data of NO_2 concentration present in the air from 1999 to 2003 at a site in Delhi. The prediction performance based on MAPE, RMSE, and RE confirmed that the hybrid modeling is more effective than individual models in forecasting the air pollutant concentrations. Díaz-Robles et al. (2008) studied hourly and daily time series of PM_{10} and meteorological data during 2000–2006 at the Las Encinas monitoring station in Temuco. Hybrid model outperformed ARIMAX and ANN in terms of RMSE, MAE, and BIC. Prybutok et al. (2000) studied a neural network model for forecasting daily maximum ozone levels and showed that the neural network model is superior to the two conventional statistical models, regression and Box-Jenkins *ARIMA* models. The maximum ozone data for the year 1994 from Houston was the research object.

Ji et al. (2019) forecasted the future carbon prices using the hybrid *ARIMA*-CNN-LSTM model. The linear features of the data set were modeled by *ARIMA*. CNN model extracted the spatial features of the residual of *ARIMA* model. And finally long-term dependencies of these features were captured by LSTM. It was evident from the experimental analysis that the hybrid *ARIMA*-CNN-LSTM outperformed the individual models. The dataset of carbon future prices from 7 April 2008 to 6 May 2019 was taken. Koutroumanidis et al. (2009) showed that the hybrid *ARIMA*-ANN model has a better adaptability and can make better predictions of future selling prices of the fuelwood produced by Greek state forest farm as compared to both the *ARIMA* model and the simple ANN model.

Faruk (2010) observed that the best fit models for water temperature (0^C), boron (mgl^{-1}), and DO (mgl^{-1}) are SARIMA $(1,1,1) (0,0,1)_{12}$, $(1,1,1) (0,0,1)_{12}$, and $(1,1,1) (0,0,1)_{12}$, respectively. They further found hybrid modeling approach of combining SARIMA with NNBP can give more reliable predictions for these parameters of a river than the Neural Network and traditional baseline *ARIMA* modeling approach individually.

Chattopadhyay and Chattopadhyay (2010) and Somvanshi et al. (2006) examined the rainfall data of India and found that hybrid models outperformed individual models in terms of forecasting efficiencies. Wei et al. (2016) found that the most appropriate model for the morbidity data for hepatitis from the Heng County CDC from January 2005 to December 2012 is *ARIMA* $(0,1,2) (1,1,1)_{12}$. (Table 2).

From Table 3, it can be clearly seen that errors are reduced when hybrid models are implemented on the various data sets. Thus, both theoretical and empirical findings are in strong favor of combining different methods to achieve effective and efficient forecasts (Newbold and Granger 1974). Consequently, hybridizing different models has become the latest research objectives in the field of modeling and forecasting. Table 2 provides concise results of sections "Survey of air quality/environment/health data involving application of *ARIMA* models," "Analyzing *ARIMA* hybrid-based studies," "Reviewing application of ARIMA alone," and "Investigating ARIMA with hybrid methodology" and some of the other papers involving ARIMA in tabular form which would be easier to comprehend.

Examining role of ARIMA in studies other than air quality/environment/health concerns

ARIMA and its hybrid models have got such wide utilization that they cannot be restricted to merely one or two arenas of study. Therefore, in this section, few research studies in the field of financial data or load forecasting or some other fields are briefly reviewed to provide a glimpse of its varied applications. Following the previous chronology, we review papers with ARIMA alone and then followed by review of papers involving ARIMA and its hybrid methodology.

Table 2 Concise results of application of *ARIMA* and its hybrid methodology in the field of environment/health/air quality

Author	Methodology	Data set	Findings
Kulkarni et al. (2018)	*ARIMA*	SO_2, RSPM, NOx, and SPM concentration data set of Nanded City (Maharashtra, India) from 2011 to 2016 with the frequency of 2 days in a week	1. The randomness, trend, and seasonality present in the data were analyzed using 95% confidence interval 2. The forecasted result revealed that RSPM and SPM are exceeding the permitted limits and thus requires immediate attention of the govt. officials
Jaiswal et al. (2018)	*ARIMA* (1,0,0), (1,0,1), (1,1,1)	The AQI data in terms of CO, NO_2, SO_2, $PM_{2.5}$, and PM_{10} for the time span January 2013 to December 2016 for the city of Varanasi (India) is taken	1. Mann–Kendall and Sen's slope estimator approach were employed 2. Different *ARIMA* models, namely, *ARIMA* (1,0,0), (1,0,1), (1,1,1) were fitted on the three data sets of summer, monsoon, and winter seasons and their results compared 3. *ARIMA* (1,1,1) was chosen as the best fit model for forecasting
Pohoata and Lungu (2017)	*ARIMA*	Ambient concentrations of SO, NO_2, NOx, CO, O_3, PM_{10}, and $PM_{2.5}$ from 6 monitoring stations of Ploiesti (Romania) for the year 2013	1. Principal component analysis was carried out 2. While *ARIMA* (3,1,3) provided good results for NOx, NO_2, and O_3, it failed to give satisfactory results for CO and PM_{10}. This showed limitations of using *ARIMA*
Chaudhuri and Dutta (2014)	*ARIMA*	The daily data is collected for the parameters SO_2, NO_2, CO, PM_{10}, CO, O_3, surface temperature (°C), and surface relative humidity (%) from 2002 to 2012 over Kolkata, India, from WBPCB	1. To identify the trends in data, nonparametric test Mann–Kendall (MK) is employed. The test depicted seasonality and monthly variability in the data 2. Box plot for each of the pollutants SO_2, NO_2, CO, PM_{10}, CO, and O_3 from 2002 to 2014 is plotted. *ARIMA* (1,1,1), *ARIMA* (0,2,2) and *ARIMA* (2,1,2) are employed, and their white noise (WN) variances are calculated 3. *ARIMA* (0, 2, 2) is the best fit model for the daily forecasting of these pollutants with AIC, BIC, MAPE, FPE, and AICc as evaluation criterions
Guarnaccia et al. (2017)	Seasonal *ARIMA*	Hourly CO concentration (ppm) data values for the year 2012 of San Nicolas de Garza	SARIMA $(0,1,1) \times (0,1,1)_{24}$ and SARIMA $(1,0,1) \times (2,0,1)_{24}$ models were applied to the hourly CO concentrations data for statistical and time series analysis. The models showed good predictive results for CO concentrations
Kumar et al. (2004)	*ARIMA*	Hourly concentration data of O_3 at the airport in Brunei Darussalam from July 8, 1998, to March 15, 1999	*ARIMA* (1,0,1) was the best suitable method. NMSE, Fractional Bias and MAPE were used as deciding performance criterion. χ^2 goodness fit test indicated the model is most appropriately fit

Table 2 (continued)

Author	Methodology	Data set	Findings
Liu et al. (2018)	ARIMAX, CMAQ-ARIMA, hybrid model	O_3, $PM_{2.5}$, and NO_2 concentration data values from October 2014 to September 2015 for three stations of Hong Kong	1. Employing CMAQ-ARIMA led to significant reduction in RMSE for the daily 1-h and 8-h forecasting values at all the stations 47.8–49.7%, 14.3–21.0%, and 41.2–46.3%, respectively, for O_3, $PM_{2.5}$, and NO_2. For the 72 hourly advance forecasts, RMSE values are slashed respectively by 18.2%, 32.1%, and 36.7% for $PM_{2.5}$, NO_2, and O_3 by employing ARIMAX using CMAQ 2. A great reduction in RMSE values is observed for rural NO_2 and nonrural $PM_{2.5}$ when ARIMAX is employed with three numerical models 3. Dynamic hourly forecast shows that ARIMAX can also be successfully applied to forecast of 7- to 72-h $PM_{2.5}$, 4- to 72-h NO_2, and 4- to 6-h O_3 4. Augmenting the advantages of ARIMA and numerical modeling, a more efficient and consistent hybrid ARIMAX model is developed for real-time air quality forecasting
Zhang et al. (2018)	ARIMA	AQI data and meteorological parametric data for Fuzhou (China) from August 2014 to July 2016	1. $PM_{2.5}$ concentrations were positively correlated with PM_{10}, NO_2, and SO_2 concentrations and negatively correlated with meteorological parameters. Also, high seasonal fluctuations were observed during cold periods as compared to warm periods 2. ARIMA (6,1,1) best fitted the data
Slini et al. (2002)	ARIMA	The data of maximum daily O_3 concentration of Athens (Greece) from 1990 to 1998	The forecasting performance of ARIMA (1,1,1) and ARIMA (1,1,0) were compared and the limitations of model over the false alarms also explained
Kumar and Goyal (2011)	ARIMA	Daily AQI data of RSPM, SO_2, NO_2, and SPM for a period of 7 years ranging from 2000 to 2006 at ITO station, Delhi	Three models are assessed, i.e., ARIMA (1,1,1), PCR and ARIMA combined with PCR with respect to 4 seasons of the year in Delhi. The hybrid model performed better in predicting AQI one-day advance
Saha et al. 2019	ARIMA	Monthly average and seasonal variation of ventilation coefficient values over Delhi from Jan 2006 to Feb 2014	This is the first applicability of ARIMA model on the dataset of VC. 12 ARIMA models for each month of the year were developed using Box-Jenkins methodology and used for forecasting. The average long-term value of VC was calculated as 1249 ± 236 m^2/s
Lau et al. (2009)	ARIMA, SARIMA	CO concentration data of district of Hong Kong, between January 1999 and December 2005	1. The hourly data of CO was found to behave like traffic data, and this strong association was proven by SARIMA $(0, 1, 1) (0, 1, 1)_{24}$ model 2. Long-term memory were features found in data

Table 2 (continued)

Author	Methodology	Data set	Findings
Díaz-Robles et al. (2008)	ARIMA, ANN, hybrid model	Hourly data of PM_{10} for the time span 2000–2006 of the Las Encinas monitoring station situated in Temuco	Hybrid model outperformed ARIMAX, ANN, and MLR on the basis of experimental results
Kumar and Jain (2010)	ARIMA, ARMA	Daily mean concentration O_3, CO, NO, and NO_2 at an urban traffic site (ITO) of Delhi, India, from August 2006 to July 2007	ARIMA is a successful tool in modeling and forecasting O_3, CO, NO, and NO_2 and thus can be effectively utilized for the forewarning purposes
Kumar and Ridder (2010)	FFT-ARIMA and FFT-ARIMA-GARCH	Daily maximum O_3 concentration at Brussels B1 (1 Jan 1998 to 31 July 2007), Brussels (B2) (1 Jan 2002 to 30 Jun 2007), London (L1) (1 May 1996 to 30 Sep 2007) and London (L2) (1 Jan 2000 to 30 Sep 2008)	Modeling heteroskedastic effects using GARCH in O_3 time series helps in making a) More accurate forecast b) Improving confidence intervals c) Reduction in false alarms at the given sites
Siew et al. (2008)	ARIMA, ARFIMA	Air pollution index from March 1998 to December 2003 in Shah Alam, Selangor	The MAPE, RMSE, and MAE values of the ARMA (3,1,3) and integrated ARFIMA (0, − 0.5, 2) models were compared. ARFIMA gave slightly better performance than ARIMA
Abhilash et al. (2018)	ARIMA	Pollution data of NO_2, PM_{10}, and SO_2 from January 2013 to March 2016 of Bengaluru (India)	RSPM is the paramount contributor to air pollution, while SO_2 is the most submissive one in this study. Separate analysis is done for each pollutant and results analyzed. ARIMA model is apt for short-term predictions
Gocheva-Ilieva et al. (2014)	SARIMA, ARIMA	Data for concentrations of PM_{10}, SO, NO_2, NO, NOx, and O_3 in Blagoevgrad from 1 September 2011 to 31 August 2012	Factor analysis and SARIMA approach are suitable techniques in examining air pollution levels for small urban areas
Chelani and Devotta (2006)	ARIMA, nonlinear model, hybrid model on the basis of chaos theory	NO_2 concentration data set of Delhi from 1999 to 2003	Hybrid models are better substitute to individual models in forecasting air pollutants
Kim (2010)	AR, ARX, T-TARX, P-TARX	Daily ozone concentration data (in ppb) for the time period January 2001 to December 2003 in San Bernardino County	The proposed model could capture non-linearity as well as non-stationarity more efficiently and also forecast with one-day advance more accurately
Ediger and Akar (2007)	ARIMA, SARIMA (of seasonal order 1, 2, 3, 4, and 5)	Total primary energy data of Turkey (hard coal, asphaltite lignite, Petro coke, wood, animal and plant remains, oil, natural gas, geothermal, hydro-power, heat, electricity, and solar) from 2005 to 2020 was used	ARIMA and SARIMA are an efficient tool in forecasting energy demand of Turkey. The accurateness of the fitted model is checked with MSE
Ediger et al. (2006)	ARIMA, SARIMA, regression model	Fossil fuel reserves of Turkey from 1950 to 2003	1. Various forecasting models were developed for different fossil fuel types 2. Fossil fuel production of Turkey will end in 2038

Table 2 (continued)

Author	Methodology	Data set	Findings
Li and Li (2017)	GM (1,1), ARIMA (1,2,0), GM-ARIMA	Energy demand of Shandong province from 1995 to 2015	1. GM- ARIMA model has higher accuracy in terms of average relative error in comparison to GM (1,1). ARIMA (1,2,0) and metabolism GM (1,1) 2. Series of residuals are obtained with gray metabolic model. Residual corrections are done with the help of ARIMA model. This process enhanced the prediction accuracy significantly, and this was the major innovation of this study
Aamir and Shabri (2016)	ARIMA, GARCH and ARIMA-Kalman	The average monthly prices of Pakistan crude oil from February 1986 to March 2015	The ARIMAKalman filter technique proved to best approach as MAE and RMSE were minimum as compared to ARIMAand GARCH
Zhu and Wei (2013)	ARIMA, LSVM, ANN, ARIMALSSVM	Two main carbon future prices one maturing on December 2010 (April 22, 2005, to December 20, 2010) and other on December 2012 (April 22, 2005, to June 27, 2011)	ARIMALSSVM is a better forecasting model than the other models. To calculate the optimal parameters, PSO technique is used with an aim of better forecasting accuracy
Koutroumanidis et al. (2009)	ARIMA, ANN, ARIMA -ANN hybrid model	Mean annual prices data for coniferous fuelwood for the period 1964–2006 in Greece	Result revealed better adaptability of ARIMA–ANN as compared to ARIMA model
Narayanan et al. (2013)	ARIMA	Monthly rainfall data for March, April, and May for six stations Abu, Ahmedabad, Ajmer, Amritsar, Bikaner, and Jodhpur for a period of 60 years were analyzed from 1949 to 1999	1. To investigate the rainfall trend, pre-whitened Mann Kendall test was employed 2. Forecast results for 2000 to 2009 were compared using RMSE 3. Univariate ARIMA modeling showed there is major increase in the pre-monsoon rainfall over the north-west part of the country
Ab Razak et al. (2018)	ARIMA	Daily and monthly rainfall and stream flow data for Bandar Segamat, Kemelah, and Ladang Sungai Labis stations (Malaysia) for the period 2000 to 2010	ARIMA (0, 1, 2) is the best fit model. Mann–Kendall trend analysis was used to observe the trend in the series
Geetha and Nasira (2016)	ARIMA	Five years rainfall data of Trivandrum from 2009 to 2014	For the given time series, ARIMA is the best suited model with forecast accuracy above 80%
Zakaria et al. (2012)	Seasonal ARIMA (3,0,2)x(2,1,1)30, (1,0,1)x(1,1,3)30, (1,1,2) x(3,0,1)30 and (1,1,1)x(0,0,1)30	Rainfall data for Sinjar, Mosul, Rabeaa, and Talafar for the time period 1990–2011	Box-Jenkins (ARIMA) model methodology could be used as a suitable tool to forecast the weekly rainfall in semi-arid region like North West of Iraq for the up-coming 5 years (2012–2016)
Chattopadhyay and Chattopadhyay (2010)	ARIMA, ARNN	Average rainfall data over India for the time period from 1871 to 1999	1. ARIMA (0,1,1) is the most suitable method 2. ARNN is superior to ARIMA (0,1,1) based on Willmott's index and Pearson correlation
Somvanshi et al. (2006)	ARIMA, ANN	Average annual rainfall data of Hyderabad for the period of 1901 to 2003	ANN is an apt technique for modeling the rainfall data of Hyderabad based on RMSE, MAE, AIC, BIC

Table 2 (continued)

Author	Methodology	Data set	Findings
Wang et al. (2018)	ARIMA model and grey model GM (1,1)	Monthly incidence of hepatitis B in China from March 2010 to October 2017	1. Incidence of hepatitis B in China has seasonal variation. AIC is lowest in ARIMA $(3,1,1)$ $(0,1,2)_{12}$ 2. RMSE, MAE, and MAPE of ARIMA $(3,1,1)$ $(0,1,2)_{12}$ model was lower than GM (1,1) model so ARIMA model showed better hepatitis B fitting and forecasting performance than GM (1,1) model
Duan and Zhang (2020)	ARIMA	Daily new confirmed cases of the COVID-19 infections in Japan and South Korea from January 20, 2020, to April 26, 2020	ARIMA (6,1,7) and ARIMA (2,1,3) was found to be best fit model for Japan and South Korea, respectively
Wei et al. (2016)	ARIMA, GRNN	The monthly morbidity data for hepatitis from the Heng County CDC from January 2005 to December 2012	1. The most appropriate model for the data under consideration is ARIMA $(0,1,2)$ $(1,1,1)_{12}$ 2. The hybrid ARIMA-GRNN model showed better forecasting performance than the individual model in terms of lower MAPE, MAE, MSE, and RMSE
Singh et al. (2020a, b)	ARIMA, Hybrid Wavelet – ARIMA	The dataset consisting of death cases by COVID-19 of Italy, Spain, France, the UK, and the USA from 21 January 2020 to 11 April 2020	The performance of hybrid model is nearly 80% better than ARIMA model for the countries Italy, Spain, and the UK. However, it is approximately 50% improved for the countries France and the USA
Ispriyanti (2018)	ARIMA, GARCH	Cayenne production monthly data from January 2003 to November 2015 in Central Java	For the given data ARIMA ([1,3],0,0)-GARCH (1,0) is the best model
Singh et al. (2013)	ARIMA	Annual paddy production and its area under cultivation from 1974–1975 to 2010–2011of Chhattisgarh (India)	ARIMA (2, 1, 2) was found the best fit model for Paddy Area while ARIMA (2, 1, 0) was the most appropriate model for Paddy production in Chhattisgarh (India) with AIC as selection criterion
Assis et al. (2010)	ARIMA (3,1,2), GARCH (1,1), mixed ARIMA (3,1,2)/GARCH (1,1) exponential smoothing,	Monthly average data from January 1992—December 2006 of Bagan Datoh cocoa bean prices	The GARCH model gave better results than the exponential smoothing, ARIMA and the mixed ARIMA/GARCH in forecasting given data using RMSE, MAE, and MAPE
Ali (2013)	EGARCH, GJR-GARCH, TGARCH, AVGARCH, NGARCH, IGARCH, and APARCH	FIB time series from 2006 to 2008	Keeping Akaike, Bayes, Log Likelihood information, and Hannan-Quinn as criterions TGARCH proved to be slightly better than the other models in capturing response of the pathogen indicator variable. Nyblom Statistic is also used for each model
Williams and Hoel (2003)	ARIMA, SARIMA $(1,0,1)(0,1,1)_{672}, (1,0,2)$ $(0,1,1)_{672}, (2,0,1)(0,1,1)_{672}$ for M25 station and SARIMA $(1,0,1)(0,1,1)_{672}, (3,0,0)(0,1,1)_{672}, (1,0,2)(0,1,1)_{672}$ for I-75 stations, random walk	Vehicular traffic Data of M25 motorway of London and I-75 freeway of Atlanta	Since MAD and MAPE values were lowest in case of SARIMA as compared to ARIMA and Random Walk therefore the experimental results supported the theory of SARIMA as the best model for modeling and forecasting traffic data
Nie et al. (2012)	ARIMA, SVM, ARIMA-SVM	The electric load data of power company in Heilongjiang of China from March 1 to May 31, 1999	On the basis of MAPE and RMSE hybrid, ARIMA-SVM model is better than the two separate models alone

Table 2 (continued)

Author	Methodology	Data set	Findings
Zhang (2003)	*ARIMA*, ANN, Hybrid	The Wolf's sunspot data, the Canadian lynx dat,a and the British pound and US dollar exchange rate data from 1980 to 1993	Combined model has greater forecasting accuracy on the basis of MAD, MSE
Chen and Wang (2007)	SVM, SARIMA, SARIMASVM (integration of SARIMA and SVM)	Production data of the Taiwanese machinery industry from January 1991 to December 1996	When forecasted errors were calculated, the hybrid model SARIMASVM2 gave better results
Khashei and Bijari (2011)	*ARIMA*, ANN, Zhang's hybrid models, novel hybrid model	Wolf's sunspot data (1700–1987), Canadian lynx data (1821–1934), and the British pound/US dollar exchange rate data (1980–1993)	The new model developed outperformed *ARIMA*, ANN, and Zhang's hybrid models across varied time horizons
Tseng et al. (2002)	SARIMA and SARIMABP	Total production revenues time series data of Taiwan machinery trade for the time span 1991 to 1996	SARIMABP outperformed SARIMA model
Wang (2011)	*ARIMA* and fuzzy time series by heuristic models	Taiwan exports data for the time period January 1990 and 30 March 2002	The heuristic fuzzy time series model is an appropriate tool when information is lacking and vital decisions are needed
Conejo et al. (2005)	Wavelet ARIMA	Electricity market data of mainland Spain for the year 2002	Wavelet-ARIMA outperforms ARIMA alone in terms of weekly error (%)
Pai and Lin (2005)	*ARIMA*, SVM, hybrid model	Closing price of 10 stocks from Oct. 21, 2002, to Dec. 31, 2002	By calculating MAE, MSE, MAPE, RMSE for each of the ten companies under consideration and each model, it was observed that prediction performance of single ARIMA and SVM was less than the hybrid model. Hybridizing two different models with optimal parameters resulted in reduction of the forecasting errors
Wang et al. (2012)	ESM, *ARIMA*, BPNN, the equal weight hybrid model (EWH), and the random walk model (RWM)	Monthly SZII closing index and DJIAI opening index from China and the USA, respectively, from January 1993 to December 2010	Improved results were obtained with the help of hybrid models with RMSE, MAPE, and ME as measuring errors in both the cases
Ho et al. (2002)	*ARIMA*, RNN, MFNN	The failure time data for a repairable compressor system at a Norwegian process plant was investigated in the time period between 1968 and 1989	RNN at the optimal weighting factor gives satisfactory performances compared to the ARIMA model. MAD and MSE were calculated
Aslanargun et al. (2007)	*ARIMA*, Neural Network, hybrid model	Number of monthly tourist arrivals to Turkey between January 1984 and December 2003	(12:1:1) MLP& (5:1:1:1) MLP and (12:1:1) MLP& (6:9:1) RBFN hybrid gave better results
Faruk (2010)	*ARIMA*, ANN, Hybrid model	The dataset of Aydin measurement, comprising three water quality parameters monitored over 9 years (1996–2004), was analyzed	Hybrid modeling approach can give more reliable predictions of water temperature, boron, and dissolved oxygen time series of a river than the ANN and ARIMA modeling approach individually

Table 3 Application of performance evaluation measures on various ARIMA and its hybrid models

Authors	Research object area/selection criteria/research Data	Model used	Performance measures
Mani and Volety (2021)	Air pollutant CO	ARIMA	RMSE = 0.13 MAE = 0.12
		LSTM	RMSE = 0.13 MAE = 0.09
	NH$_3$	ARIMA	RMSE = 0.44 MAE = 0.39
		LSTM	RMSE = 0.14 MAE = 0.13
	O$_3$	ARIMA	RMSE = 13.6 MAE = 13.13
		LSTM	RMSE = 2.06 MAE = 1.70
Wang et al. (2017)	Station 1. Fuyong	1. ARIMA	RMSE = 8.1242 MAE = 6.0771
		2. SVM	RMSE = 8.2779 MAE = 5.8432
		3. Hybrid GARCH	RMSE = 7.8458 MAE = 5.5375
	2. Longhua	1. ARIMA	RMSE = 5.2084 MAE = 4.4061
		2. SVM	RMSE = 4.7392 MAE = 3.8072
		3. Hybrid GARCH	RMSE = 4.6373 MAE = 3.7256
	3. Henggang	1. ARIMA	RMSE = 7.1915 MAE = 6.1382
		2. SVM	RMSE = 7.1910 MAE = 5.8290
		3. Hybrid GARCH	RMSE = 6.4584 MAE = 5.2359
	4. Pingshan	1. ARIMA	RMSE = 7.1943 MAE = 5.5786
		2. SVM	RMSE = 7.2744 MAE = 5.4968
		3. Hybrid GARCH	RMSE = 7.0182 MAE = 5.1594
Samia et al. (2012)	PM10	ARIMAX	RMSE = 51.094 R^2 = 0.675 MSE = 2610 MAE = 25.677
		ANN	RMSE = 29.939 R^2 = 0.888 MSE = 896.2 MAE = 20.573
		ARIMAX-ANN	RMSE = 11.656 R^2 = 0.983 MSE = 135.828 MAE = 8.268

Table 3 (continued)

Authors	Research object area/selection criteria/research Data	Model used	Performance measures
Zhu et al. (2017)	AQI	ARIMA	MAPE = 186.400 RMSE = 208.6240 MAE = 198.545
		SVR	MAPE = 30.200 RMSE = 42.7060 MAE = 35.812
		Wavelet-SVR	MAPE = 30.890 RMSE = 42.7388 MAE = 36.2368
		GRNN	MAPE = 25.91 RMSE = 36.0930 MAE = 26.7732
		Wavelet-GRNN	MAPE = 24.24 RMSE = 33.0731 MAE = 25.9814
		EMD-GRNN	MAPE = 21.86 RMSE = 27.6400 MAE = 22.3637
		EMD-IMFs-Hybrid	MAPE = 15.600 RMSE = 25.767 MAE = 17.240
		EMD-SVR-Hybrid	MAPE = 15.600 RMSE = 24.462 MAE = 18.101
Kumar and Goyal (2011)	Summer	ARIMA	RMSE = 35.30 NMSE = 0.0106
		PCR	RMSE = 35.70 NMSE = 0.0113
		MODEL3 (ARIMA + PCR)	RMSE = 31.99 NMSE = 0.0086
	Monsoon	ARIMA	RMSE = 62.55 NMSE = 0.0537
		PCR	RMSE = 68.93 NMSE = 0.0694
		MODEL3 (ARIMA + PCR)	RMSE = 61.94 NMSE = 0.0531
	Post-monsoon	ARIMA	RMSE = 57.70 NMSE = 0.0354
		PCR	RMSE = 62.93 NMSE = 0.0416
		MODEL3 (ARIMA + PCR)	RMSE = 56.55 NMSE = 0.0341
	Winter	ARIMA	RMSE = 26.74 NMSE = 0.0054
		PCR	RMSE = 58.39 NMSE = 0.0301
		MODEL3 (ARIMA + PCR)	RMSE = 65.76 NMSE = 0.0390

Table 3 (continued)

Authors	Research object area/selection criteria/research Data	Model used	Performance measures
Ji et al. (2019)	Carbon prices	CNN	MAPE = 0.0623 RMSE = 0.8616
		LSTM	MAPE = 0.0456 RMSE = 0.7251
		ARIMA	MAPE = 0.0423 RMSE = 0.7015
		ARIMA-CNN-LSTM	MAPE = 0.0421 RMSE = 0.6940
Chelani and Devotta (2006)	NO_2 data	ARIMA	MAPE = 17.3 RMSE = 58.78 RE = 0.24
		Nonlinear prediction	MAPE = 11.6 RMSE = 55.37 RE = 0.23
		Hybrid model	MAPE = 5.37 RMSE = 13.93 RE = 0.19

Table 3 (continued)

Authors	Research object area/selection criteria/research Data	Model used	Performance measures
Kim (2010)	**AR** at location 1. Crestline	AR	RMSE = 15.32 R^2=0.74
	2. Fontana	AR	RMSE = 19.45 R^2=0.66
	3. Redlands	AR	RMSE = 17.51 R^2=0.70
	4.SB-4th St	AR	RMSE = 17.22 R^2=0.68
	5. Upland	AR	RMSE = 18.05 R^2=0.64
	ARX at location 1. Crestline	ARX	RMSE = 13.37 R^2=0.80
	2. Fontana	ARX	RMSE = 14.89 R^2=0.80
	3. Redlands	ARX	RMSE = 13.96 R^2=0.81
	4. SB-4th St	ARX	RMSE = 13.96 R^2=0.81
	5. Upland	ARX	RMSE = 13.11 R^2=0.81
	P-TARX at location 1. Crestline	P-TARX	RMSE = 12.92 R^2=0.82
	2. Fontana	P-TARX	RMSE = 13.91 R^2=0.83
	3. Redlands	P-TARX	RMSE = 13.41 R^2=0.83
	4. SB-4th St	P-TARX	RMSE = 12.28 R^2=0.84
	5. Upland	P-TARX	RMSE = 12.37 R^2=0.83
	T-TARX at location 1. Crestline	T-TARX	RMSE = 12.56 R^2=0.83
	2. Fontana	T-TARX	RMSE = 12.25 R^2=0.85
	3. Redlands	T-TARX	RMSE = 13.10 R^2=0.84
	4. SB-4th St	T-TARX	RMSE = 11.60 R^2=0.86
	5. Upland	T-TARX	RMSE = 11.89 R^2=0.85
Somvanshi et al. (2006)	Rainfall data	ANN (4,2,1)	RMSE = 5.5186 R^2=0.9841 MAE = 0.9841
		ARIMA (4,0,0)	RMSE = 35.882 R^2=0.953 MAE = 24.388

Table 3 (continued)

Authors	Research object area/selection criteria/research Data	Model used	Performance measures
Wei et al. (2016)	Hepatitis data	ARIMA	MAPE = 0.1115 RMSE = 1.5561 MSE = 2.4215
		GRNN	MAPE = 0.0.0150 RMSE = 0.4726 MSE = 0.2233
		ARIMA-GRNN	MAPE = 0.0878 RMSE = 0.9391 MSE = 0.8820
Zhu and Wei (2013)	1.Carbon prices Dec 10	ARIMA	RMSE = 0.2474
		LSSVM	RMSE = 0.2473
		ANN	RMSE = 0.2568
		ARIMALSSVM2	RMSE = 0.0311
		ARIMASVM	RMSE = 0.0636
		ARIMAANN	RMSE = 0.0967
	2.Carbon prices Dec12	ARIMA	RMSE = 0.2678
		LSSVM	RMSE = 0.2676
		ANN	RMSE = 0.2712
		ARIMASSVM	RMSE = 0.0628
		ARIMAANN	RMSE = 0.0963
		ARIMALSSVM2	RMSE = 0.0309
Koutroumanidis et al. (2009)	1. Coniferous species	ARIMA	MAPE = 16.9133 RMSE = 0.051308
		ANN	MAPE = 14.583646 RMSE = 0.05233171
		ARIMA-ANN hybrid model	MAPE = 14.0036172 RMSE = 0.050151798
	2. Broadleaved species data	ARIMA	MAPE = 16.6856 RMSE = 0.147595
		ANN	MAPE = 13.016263 RMSE = 0.0966550
Ho et al (2002)	1. Short term	RNN(α = 0.3)	MAE = 58.7
		RNN(α = 0.6)	MAE = 58.5
		RNN(α = 0.8)	MAE = 161.2
		MFNN	MAE = 297.6
		ARIMA	MAE = 54.1
	2. 1. Long term	RNN(α = 0.3)	MAE = 89.5
		RNN(α = 0.6)	MAE = 88.7
		RNN(α = 0.8)	MAE = 177.6
		MFNN	MAE = 187.2
		ARIMA	MAE = 86.7
Díaz-Robles et al. (2008)	PM_{10} data	ARIMAX	RMSE = 28.46 MAE = 19.87
		ANN	RMSE = 28.57 MAE = 20.65
		Hybrid	RMSE = 8.80 MAE = 6.74

Table 3 (continued)

Authors	Research object area/selection criteria/research Data	Model used	Performance measures
Faruk ([2010](#))	Water temperature (°C)	ARIMA	MAPE = 42.076 RMSE = 0.102
		ANN	MAPE = 21.114 RMSE = 0.048
		Hybrid	MAPE = 18.282 RMSE = 0.039
	DO (mg l^{-1})	ARIMA	MAPE = 47.10 RMSE = 0.113
		ANN	MAPE = 29.612 RMSE = 0.061
		Hybrid	MAPE = 26.985 RMSE = 0.051
	Boron (mg l^{-1})	ARIMA	MAPE = 57.512 RMSE = 0.165
		ANN	MAPE = 36.813 RMSE = 0.074
		Hybrid	MAPE = 33.611 RMSE = 0.063
Singh et al. ([2020a, b](#))	Italy	ARIMA	MSE = 2.565 $\times 10^6$ MAE = 1.243 $\times 10^3$
		Wavelet-ARIMA hybrid model	MSE = 0.398 $\times 10^6$ MAE = 0.464 $\times 10^3$
	Spain	ARIMA	MSE = 0.782 $\times 10^6$ MSE = 0.782 $\times 10^6$
		Wavelet-ARIMA hybrid model	MSE = 0.028 $\times 10^6$ MAE = 0.136 $\times 10^3$
	France	ARIMA	MSE = 11.006 $\times 10^6$ MAE = 24.640 $\times 10^3$
		Wavelet-ARIMA hybrid model	MSE = 5.245 $\times 10^6$ MAE = 1.627 $\times 10^3$
	USA	ARIMA	MSE = 16.540 $\times 10^6$ MAE = 2.822 $\times 10^3$
		Wavelet-ARIMA hybrid model	MSE = 3.900 $\times 10^6$ MAE = 1.341 $\times 10^3$
	UK	ARIMA	MSE = 29.742 $\times 10^6$ MAE = 1.316 $\times 10^3$
		Wavelet-ARIMA hybrid model	MSE = 0.064 $\times 10^6$ MAE = 0.193 $\times 10^3$

Reviewing application of ARIMA alone

CONTRERAS et al. ([2003](#)) proposed two ARIMA models to predict hourly prices in the electricity markets of Spain and California. The Spanish data showed volatility. Jadevicius and Huston ([2015](#)) generated around twenty ARIMA models ranging from ARIMA (1, 0, 0) to ARIMA (4, 0, 4). ARIMA (3, 0, 3) model outperformed all the models to model Lithuanian house price index. Garcia et al. ([2005](#)) applied ARIMA and GARCH model to forecast one-day ahead electricity for mainland Spain and California with the conclusion that GARCH model is more accurate than ARIMA. Yaziz et al. ([2016](#)) surveyed the modeling and forecasting performances of gold prices using ARIMA-TGARCH with Gaussian, Student's t skewed, Student's t, GED, and skewed GED innovations. Hybrid ARIMA (0,1,0)-TGARCH (1,1) with t-innovation was chosen as the best model.

Ariyo et al. ([2014](#)) and Wadi et al. ([2018](#)) applied various ARIMA models on the stock exchange data sets to find ARIMA (2,1,0), ARIMA (1,0,1), and ARIMA (2,1,1) best

fits Nokia stock index, Zenith bank stock index, and Amman Stock Exchange, respectively. Siregar et al. (2017) based on ACF and PACF concluded ARIMA (3.0, 2) is the most appropriate model for predicting the sales of the factory in the city of Bandung. Wabomba et al. (2016) found ARIMA (2,2,2) as the best model for Kenyan GDP data. ARIMA (1,1,1) was found the best suitable method on the monthly gold price data by Guha and Bandyopadhyay (2016). In his research work, Lin (2018) gave an extensive review on GARCH models as well as detailed analysis of stock markets in general and particularly in China.

Investigating ARIMA with hybrid methodology

To capture the volatility present in the data, Babu and Reddy (2014) decomposed the series into two sets using MA filter. The proposed hybrid model resulted in improved one-step and multi-step ahead prediction accuracy. Nie et al. (2012) compared hybrid of ARIMA model and SVMs for short-term load forecasting with individual models via simulation for the electric load data of power company in Heilongjiang of China from March 1 to May 31, 1999. Hybrid ARIMA-SVM model performed better than the two separate models alone. Chen and Wang (2007) observed that the values of NMSE, R (correlation coefficient), and MAPE were lowest for hybrid model (SARIMASVM2) for the production data of the Taiwanese machinery industry from January 1991 to December 1996. Zhang (2003) compared ARIMA and ANN for the Wolf's sunspot data from 1700 to 1987, the Canadian lynx data, and the British pound spanning from 1821 to 1934 and the US dollar exchange rate data extending from 1980 to 1993 and concluded combined model has greater forecasting accuracy. Khashei and Bijari analyzed the same data as used by Zhang (2003) in the year 2011. A new hybrid model better than Zhang's model, ARIMA and ANN alone, was developed. Tseng et al. (2002) concluded that SARIMABP model outperformed SARIMA models for the total production revenues of Taiwan machinery industry. Wang (2011) compared ARIMA and fuzzy time series by heuristic models with Taiwan exports data from January 1990 and 30 March 2002. For longer analyzing time period, the MSE of the time series ARIMA model are lower, while for shorter analyzing period, the MSE are more. The heuristic fuzzy time series model is an appropriate tool when information is lacking and an urgent decision is needed. Wang and Leu (1996) used a hybrid model to conclude that ANN provides better results with differenced data than raw data in case of Taiwan stock exchange.

Wang et al. (2012) observed data for monthly closing index of SZII and opening index of DJIAI from China and the USA, respectively, from January 1993 to December 2010. It was seen that hybrid model could more effectively capture various relationships in the data. Ho et al. (2002) applied ARIMA, RNN,

and MFNN to the failure time data for a repairable compressor system at a Norwegian process plant. RNN at the optimal weighting factor gives satisfactory performances compared to the ARIMA model. The simple and wide use of ARIMA models led Mondal et al. (2014) to study 56 Indian stock markets spread over different sectors. The high lightening feature of this research was analysis of sector-based ARIMA models, thus covering larger portion of Indian stocks. Different ARIMA models were generated and their AIC compared.

All the above narration can be summarized that in recent past hybrid modeling techniques have substituted single modeling processes. Also, these emerging nonlinear soft computing techniques are robust, parsimonious in their data requirements and provide good long-term forecasting. Though with the advent of new soft computing methods many difficulties while implementing ARIMA have been overcome, still ARIMA remains the benchmark in the field of modeling and forecasting due its high level of simplicity and great level of reliability.

Comprehension of ARIMA models along with its hybrid requires knowledge of performance evaluation criteria also. Thus, our coming section deals with various performance evaluation metrices.

Performance evaluation of hybrid models

Different statistical metrices such as MAD (mean absolute deviation), SSE (sum squared error), RMSE (root mean squared error), MSE (mean squared error), MAPE (mean absolute percentage error) are employed while evaluating the performance of the proposed model. MAD, RMSE and MAPE are defined by

$$\left.\begin{array}{l} \text{MAD} = \frac{1}{n} \sum_{t=1}^{n} \left| y(t) - \widehat{y}(t) \right| \\ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left| y(t) - \widehat{y}(t) \right|^2} \\ \text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y(t) - \widehat{y}(t)}{y(t)} \right| \times 100 \end{array}\right\}$$

where $\widehat{y}(t)$ denotes the predicted value of $y(t)$ and n is the number of points of the training and testing data sets. It is very much evident from Table 3 that errors are reduced when hybrid models are implemented on the various data sets.

Table 3 represents the various models employed and the criteria chosen for performance evaluation of these various models by various authors.

Conclusion

Though ARIMA finds its application extensively in the field of forecasting, it too has shortcomings. ARMA models are linear models, but the time series involving environmental/atmosphere/air quality/financial data, etc. are rarely pure

linear combinations. A great competence, expertise, and experience of a researcher is required while implementing results of the Box-Jenkins procedure. Moreover, results are strongly affected by the path chosen and hence are path dependent (Weigend and Gershenfeld 1994). But almost all the models though different from each other have similar estimated correlation patterns, resulting in the arbitrary choice of the model (Box et al. 1994). Another drawback of Box-Jenkins procedure is that it is sometimes highly time consuming in the model identification step. ARMA models are highly unsuitable for running policy simulations.

After such a massive review of ARIMA and hybrid modeling involving ARIMA in the fields including or excluding environment/health/atmosphere, it can be concluded that the combined models are more robust and have higher ability to capture all the patterns of the series uniformly. Thus, combining several models or using hybrid model has emerged as a routinized custom, though ARIMA still remains the benchmark of many baseline models.

Declarations

Research ethics/ethics approval and consent to participate We further confirm that there is no aspect of the work covered in this manuscript that has involved human patients.

Consent for publication The authors give consent for the publication of this manuscript.

Competing interests The authors declare no competing interests.

References

Aamir M, Shabri A (2016) Modelling and forecasting monthly crude oil price of Pakistan: a comparative study of ARIMA, GARCH and ARIMA Kalman model. In AIP Conf Proc 1750(1):060015

Ab Razak NH, Aris AZ, Ramli MF, Looi LJ, Juahir H (2018) Temporal flood incidence forecasting for Segamat River (Malaysia) using autoregressive integrated moving average modelling. J Flood Risk Managt 11:794–804

Abhilash MSK, Thakur A, Gupta D, Sreevidya B (2018) Time series analysis of air pollution in Bengaluru using ARIMA model. Ambient Commun Comp Sys. Springer, Singapore, pp 413–426

Ahn H (2000) Modeling of groundwater heads based on second-order difference time series models. J Hydrol 234(1–2):82–94

Akaike H (1969) Fitting autoregressive models for prediction. Ann Inst Stat Math 21(1):243–247

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723

Ali G (2013) EGARCH, GJR-GARCH, TGARCH, AVGARCH, NGARCH, IGARCH and APARCH models for pathogens at marine recreational sites. J Stat Econ Methods 2(3):57–73

Al-Saba T, El-Amin I (1999) Artificial neural networks as applied to long-term demand forecasting. Artific Intell Eng 13:189–197

Anderson TW (1971) The stationarity of an estimated autoregressive process. STANFORD UNIV CA DEPT OF STATISTICS

Ariyo AA, Adewumi AO, Ayo CK (2014) Stock price prediction using the ARIMA model. In: UKSim-AMSS 16th Int Conf Comp Modelling Simu: 106–112

Aslanargun A, Mammadov M, Yazici B, Yolacan S (2007) Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. J Stat Comput Simul 77(1):29–53

Assis K, Amran A, Remali Y (2010) Forecasting cocoa bean prices using univariate time series models. Res World 1(1):71

Astrom KJ, Eykhoff P (1971) System identification—a survey. Automatica 7(2):123–162

Babu CN, Reddy BE (2014) A moving-average filter-based hybrid ARIMA–ANN model for forecasting time series data. Appl Soft Comput 23:27–38

Basu AK, Das JK (1991) On estimation and asymptotic properties of the parameters of ARMA (p, q) process in the stable case. Calcutta Statist Assoc Bull 41(1–4):45–64

Basu AK, Das JK (1992) Optimality of the maximum likelihood estimator in AR (p) model under a general set-up of the roots. Calcutta Statist Assoc Bull 42(1–2):1–18

Bates JM, Granger CW (1969) The combination of forecasts. Journal of the Operational Research Society 20(4):451–468

Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M (2020) Application of the ARIMA model on the COVID-2019 epidemic dataset. Data Brief 29:105340

Bhansali RJ (1983) A simulation study of autoregressive and window estimators of the inverse correlation function. J Roy Stat Soc: Ser C (appl Stat) 32(2):141–149

Box GEP, Jenkins G (1970) Time series analysis, forecasting and control. Holden-Day, San Francisco, CA

Box GEP, Jenkins GM, Reinsel GC (1994) Time series analysis: forecasting and control, 3rd edn. Prentice-Hall, Englewood Cliffs, NJ

Broersen P (1985) Selecting the order of autoregressive models from small samples. IEEE Trans Acoust Speech Signal Process 33(4):874–879

Buchananan WK, Hodges P, Theis J (2001) Which way the natural gas price: an attempt to predict the direction of natural gas spot price movements using trader positions. Energ Eco 23(3):279–293

Burg JP (1975) Maximum entropy spectral analysis. Stanford University

Cappello C, De Iaco S, Posa D (2018) Testing the type of non-separability and some classes of space-time covariance function models. Stoch Env Res Risk Assess 32(1):17–35

Cartwright PA (1985) Forecasting time series: a comparative analysis of alternative classes of time series models. J Time Ser Anal 6(4):203–211

Chattopadhyay S, Chattopadhyay G (2010) Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. CR Geosci 342(2):100–107

Chaudhuri S, Dutta D (2014) Mann-Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models. Environtl Monitng Assesst 186(8):4719–4742

Chelani AB, Devotta S (2006) Air quality forecasting using a hybrid autoregressive and nonlinear model. Atmos Environ 40(10):1774–1780

Chen KY, Wang CH (2007) A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Sys App 32(1):254–264

Choi BS (1986) An algorithm for solving the extended Yule-Walker equations of an autoregressive moving-average time series (Corresp.). IEEE Transac Inform Theor 32(3):417–419

Cleveland WS (1972) The inverse autocorrelations of a time series and their applications. Technometrics 14(2):277–293

Conejo AJ, Plazas MA, Espinola R, Molina AB (2005) Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Transac Power Sys 20(2):1035–1042

Contreras J, Espinola R, Nogales FJ, Conejo AJ (2003) ARIMA models to predict next-day electricity prices. IEEE Trans Power Sys 18(3):1014–1020

De Iaco S, Myers DE, Palma M, Posa D (2013) Using simultaneous diagonalization to identify a space–time linear coregionalization model. Math Geosci 45(1):69–86

De Iaco S, Palma M, Posa D (2015) Spatio-temporal geostatistical modeling for French fertility predictions. Spatial Stat 14:546–562

De Iaco S, Posa D (2018) Strict positive definiteness in geostatistics. Stoch Env Res Risk Assess 32(3):577–590

Díaz-Robles LA, Ortega JC, Fu JS, Reed GD, Chow JC, Watson JG, Moncada-Herrera JA (2008) A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco. Chile Atmos Environ 42(35):8331–8340

Duan X, Zhang X (2020) ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. Data Brief 31:105779

Durbin J (1959) Efficient estimation of parameters in moving-average models. Biometrika 46(3/4):306–316

Durbin J (1960) Estimation of parameters in time-series regression models. J Roy Stat Soc: Ser B (methodol) 22(1):139–153

Ediger VŞ, Akar S (2007) ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy 35(3):1701–1708

Ediger VŞ, Akar S, Uğurlu B (2006) Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. Ener Policy 34(18):3836–3846

Faruk DO (2010) A hybrid neural network and ARIMA model for water quality time series prediction. Engg Appln Artif Intell 23:586–594

Fuller WA (1976) Introduction to statistical time series, new york: Johnwiley. FullerIntroduction to Statistical Time Series1976

Garcia RC, Contreras J, Van Akkeren M, Garcia JBC (2005) A GARCH forecasting model to predict day-ahead electricity prices. IEEE Trans Power Sys 20(2):867–874

Garg N, Soni K, Saxena TK, Maji S (2015) Applications of Autoregressive integrated moving average (ARIMA) approach in time-series prediction of traffic noise pollution. Noise Control Engg J 63(2):182–194

Geetha A, Nasira GM (2016) Time-series modelling and forecasting: modelling of rainfall prediction using ARIMA model. Int J Soc Sys Sci 8(4):361–372

Glasbey CA (1982) A generalization of partial autocorrelations useful in identifying ARMA models. Technometrics 24(3):223–228

Gocheva-Ilieva SG, Ivanov AV, Voynikova DS, Boyadzhiev DT (2014) Time series analysis and forecasting for air pollution in small urban area: a SARIMA and factor analysis approach. Stoch Env Res Risk Assess 28(4):1045–1060

Godfrey LG (1979) Testing the adequacy of a time series model. Biometrika 66(1):67–72

Godolphin EJ (1977) A direct representation for the maximum likelihood estimator of a Gaussian moving average process. Biometrika 64(2):375–384

Godolphin EJ (1978) Modified maximum likelihood estimation of Gaussian moving averages using a pseudoquadratic convergence criterion. Biometrika 65:203–206

Guarnaccia C, Quartieri J (1836) Tepedino C (2017) Deterministic decomposition and seasonal ARIMA time series models applied to airport noise forecasting. In: AIP Conf Proc 1:020079

Guha B, Bandyopadhyay G (2016) Gold price forecasting using ARIMA model. J Adv Manage Sci 4(2)

Hannan EJ (1970) Multiple time series Wiley. New York

Hannan EJ, Rissanen J (1982) Recursive estimation of mixed autoregressive-moving average order. Biometrika 69(1):81–94

Harvey A (1989) Forecasting, structural time series models and the Kalman filter. Cambridge University Press

Harvey A, Koopman SJ, Shephard N (2004) State space and unobserved component models. Cambridge University Press

Hipel KW, McLeod AI, Lennox WC (1977) Advances in Box-Jenkins modeling: 1. Model Construction. Water Resour Res 13(3):567–575

Hippert HS, Pedreira CE, Souza RC (2001) Neural networks for short-term load forecasting: a review and evaluation. IEEE Trans Power Syst 16:44–55

Ho SL, Xie M, Goh TN (2002) A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Comp Industl Engg 42(2–4):371–375

Hokstad P (1983) A method for diagnostic checking of time series models. J Time Series Anal 4(3):177–183

Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. Biometrica 76(2):297–307

Huzii M (1981) Estimation of coefficients of an autoregressive process by using a higher order moment. J Time Ser Anal 2(2):87–93

Ispriyanti D (2018) Modelling of cayenne production in Central Java using ARIMA-GARCH. Int J Phy: Conference Series 1025(1):012120

Durbin J, Koopman SJ (2012) Time series analysis by state space methods, 2nd edn. Oxford University Press

Jadevicius A, Huston S (2015) ARIMA modelling of Lithuanian house price index. Int J Housing Markets Anal.

Jaiswal A, Samuel C, Kadabgaon VM (2018) Statistical trend analysis and forecast modeling of air pollutants. Global J Environtl Sci Managt 4(4):427–438

Janacek GJ (1982) Determining the degree of differencing for time series via the log spectrum. J Time Ser Anal 3(3):177–183

Jenkins GM (1982) Some practical aspects of forecasting in organizations. J Forecast 1:3–21

Ji L, Zou Y, He K, Zhu B (2019) Carbon futures price forecasting based with ARIMA-CNN-LSTM model. Procedia Comp Sci 162:33–38

Khashei M, Bijari M (2011) A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Appl Soft Comput 11(2):2664–2675

Kim SE (2010) Tree-based threshold modeling for short-term forecast of daily maximum ozone level. Stoch Environ Res Risk Assess 24(1):19–28

Koehler AB, Murphree ES (1988) A comparison of the Akaike and Schwarz criteria for selecting model order. J Roy Stat Soc: Ser C (Appl Stat) 37(2):187–195

Koreisha SG, Pukkila TM (1993) New approaches for determining the degree of differencing necessary to induce stationarity in ARIMA models. J Stat Plan Infer 36(2–3):399–412

Koutroumanidis T, Ioannou K, Arabatzis G (2009) Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model. Energ Policy 37(9):3627–3634

Kulkarni GE, Muley AA, Deshmukh NK, Bhalchandra PU (2018) Autoregressive integrated moving average time series model

for forecasting air pollution in Nanded city, Maharashtra, India. Model Earth Syst Environ 4(4):1435–1444

Kumar A, Goyal P (2011) Forecasting of daily air quality index in Delhi. Sci Total Environ 409(24):5517–5523

Kumar K, Yadav AK, Singh MP, Hassan H, Jain VK (2004) Forecasting daily maximum surface ozone concentrations in Brunei Darussalam-an ARIMA modeling approach. J Air Waste Managt Assoc 54(7):809–814

Kumar U, De Ridder K (2010) GARCH modelling in association with FFT–ARIMA to forecast ozone episodes. Atmos Environ 44(34):4252–4265

Kumar U, Jain VK (2010) ARIMA forecasting of ambient air pollutants (O 3, NO, NO 2 and CO). Stoch Environ Res Risk Assess 24(5):751–760

Lau JC, Hung WT, Yuen DD, Cheung CS (2009) Long-memory characteristics of urban roadside air quality. Transp Res Part d: Transp Environ 14(5):353–359

Li S, Li R (2017) Comparison of forecasting energy consumption in Shandong, China Using the ARIMA model, GM model, and ARIMA-GM model. Sustainability 9(7):1181

Liang G, Wilkes DM, Cadzow JA (1993) ARMA model order estimation based on the eigenvalues of the covariance matrix. IEEE Trans Signal Process 41(10):3003–3009

Lin Z (2018) Modelling and forecasting the stock market volatility of SSE Composite Index using GARCH models. Future Gen Comp Sys 79:960–972

Liu T, Lau AK, Sandbrink K, Fung JC (2018) Time series forecasting of air quality based on regional numerical modeling in Hong Kong. J Geophysical Res Atmos 123(8):4175–4196

Ljung L (1987) System Identification Theory for the User. Prentice-Hall, Englewood Cliffs, NJ

Makridakis S (1989) Why combining works? Int J Forecast 5:601–603

Makridakis S, Andersen A, Carbone R, Fildes R, Hibon M, Lewandowski R, Newton J, Winkler R (1982) The accuracy of extrapolation (time series) methods: results of a forecasting competition. J Forecast 1:111–153

Makridakis S, Wheelwright SC (1977) Forecasting: issues & challenges for marketing management. J Mark 41(4):24–38

Mani G, Volety R (2021) A comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant levels forecasting using sensor fusion with ground station data. Cogent Engg 8(1):1936886

Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11(2):431–441

McLeod I (1974) Contributions to applied time series, master's thesis. Univ. of Waterloo, Ont

McLeod I (1977) Derivation of the theoretical autocovariance function of autoregressive-moving average time series. J Roy Stat Soc: Ser C (appl Stat) 26(2):194–194

Mikosch T, Gadrich T, Kluppelberg C, Adler RJ (1995) Parameter estimation for ARMA models with infinite variance innovations. Ann Stat 305–326

Mirzavand M, Sadatinejad S J, Ghasemieh H, Imani R, Motlagh M S (2014) Prediction of ground water level in arid environment using a non-deterministic model. J Water Res Protect 2014

Monahan JF (1983) Fully Bayesian analysis of ARMA time series models. J Econ 21(3):307–331

Mondal P, Shit L, Goswami S (2014) Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. Int J Comp Sci Engg Appl 4(2):13

Morana C (2001) A semiparametric approach to short-term oil price forecasting. Energ Eco 23(3):325–338

Narayanan P, Basistha A, Sarkar S, Kamna S (2013) Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. CR Geosci 345(1):22–27

Newbold P, Granger CWJ (1974) Experience with forecasting univariate time series and the combination of forecasts. J R Stat Soc Ser A 137:131–164

Nicolaisen JD, Richter Jr, CW, Sheblé GB (2000) Price signal analysis for competitive electric generation companies. In Proc Conf Elect Utility Deregulation and Restructuring and Power Technol, London, U.K. 66–71

Nie H, Liu G, Liu X, Wang Y (2012) Hybrid of ARIMA and SVMs for short-term load forecasting. Ener Procedia 16:1455–1460

Pai PF, Lin CS (2005) A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 33(6):497–505

Parmar KS, Bhardwaj R (2014) Water quality management using statistical analysis and time-series prediction model. Appl Water Sci 4:425–434

Parzen E (1974) Some recent advances in time series modeling. IEEE Trans Autom Control 19(6):723–730

Patil RB (1990) Neural networks as forecasting experts: test of dynamic modeling over time series data. Thesis, Oklahoma State University, M.Sc

Pohoata A, Lungu E (2017) A complex analysis employing ARIMA model and statistical methods on air pollutants recorded in Ploiesti. Romania Rev Chim 68(4):818–823

Poskitt DS (1987) A modified Hannan—Rissanen strategy for mixed autoregressive-moving average order determination. Biometrika 74(4):781–790

Prybutok VR, Yi J, Mitchell D (2000) Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. European J Oper Res 122:31–40

Pukkila T, Koreisha S, Kallinen A (1990) The identification of ARMA models. Biometrika 77(3):537–548

Pukkila TM, Krishnaiah PR (1988) On the use of autoregressive order determination criteria in univariate white noise tests. IEEE Trans Acoust Speech Signal Process 36(5):764–774

Ray D (1988) Comparison of forecasts: an empirical investigation Sankhyā 50B:258-277

Robeson SM, Steyn DG (1990) Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. Atmos Environ 24(2):303–312

Saikkonen P (1986) Asymptotic properties of some preliminary estimators for autoregressive moving average time series models. J Time Ser Anal 7(2):133–155

Samia A, Kaouther N, Abdelwahed T (2012) A hybrid ARIMA and artificial neural networks model to forecast air quality in urban areas: case of Tunisia. Adv Mater Res 518:2969–2979

Schwarz G (1978) Estimating the dimension of a model. The annals of statistics, 461–464.

Shahwan T, Odening M (2007) Computational Intelligence in Economics and Finance. Springer, Berlin Heidelberg, New York, pp 63–74

Shibata R (1976) Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika AC-63(1):117–126

Siew LY, Chin LY, Wee PMJ (2008) ARIMA and integrated ARFIMA models for forecasting air pollution index in Shah Alam Selangor. Malaysian J Analytical Sci 12(1):257–263

Silvey SD (1959) The Lagrangian multiplier test. Ann Math Stat 30(2):389–407

Singh DP, Kumar P, Prabakaran K (2013) Application of ARIMA model for forecasting Paddy production in Bastar division of Chhattisgarh. Amer Int J Res Sci Technol Engg Math 5:82–87

Singh S, Parmar KS, Kaur J, Kumar J, Makkhan SJS (2021) Prediction of COVID-19 pervasiveness in six major affected states of India and two-stage variation with temperature. Air Qual Atmos Health 14(12):2079–2090

Singh S, Parmar KS, Kumar J, Makkhan SJS (2020) Development of new hybrid model of discrete wavelet decomposition and

autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. Chaos Solitons Fractals 135:109866

Singh S, Parmar KS, Kumar J (2021) Soft computing model coupled with statistical models to estimate future of stock market. Neural Comput Appl 33(13):7629–7647

Singh S, Parmar KS, Makkhan SJS, Kaur J, Peshoria S, Kumar J (2020) Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. Chaos Solitons Fractals 139:110086

Siregar B, Nababan EB, Yap A, Andayani U (2017) Forecasting of raw material needed for plastic products based in income data using ARIMA method. In: 5th Int Conf Electl, Electron Infor Engg (ICEEIE) 135–139

Slini T, Karatzas K, Moussiopoulos N (2002) Statistical analysis of environmental data as the basis of forecasting: an air quality application. Sci Total Environ 288(3):227–237

Somvanshi VK, Pandey OP, Agrawal PK, Kalanker NV, Prakash MR, Chand R (2006) Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. J Ind Geophys Union 10(2):141–151

Sreenivasan M, Sumathi K (1997) Generalised parameters technique for identification of seasonal ARMA (SARMA) and non-seasonal ARMA (NSARMA) models. Korean J Comput Appl Math 4(1):135–146

Suresh KK, Priya SK (2011) Forecasting sugarcane yield of Tamilnadu using ARIMA models. Sugar Tech 13(1):23–26

Taheri Tizro A, Ghashghaie M, Georgiou P, Voudouris K (2014) Time series analysis of water quality parameters. J Appl Res Water Wastewater 1(1):40–50

Tiao GC, Tsay RS (1994) Some advances in non-linear and adaptive modelling in time-series. J Forecast 13(2):109–131

Tsay RS, Tiao GC (1984) Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. J Am Stat Assoc 79(385):84–96

Tsay RS, Tiao GC (1985) Use of canonical analysis in time series model identification. Biometrika 72(2):299–315

Tseng FM, Yu H, Tzeng G (2002) Combining neural network model with seasonal time series ARMA model. Technol Forecast Soc Change 69:71–87

Tuan PD (1984) The estimation of parameters for autoregressive moving average models. J Time Ser Anal 5(1):53–68

Tucker WT (1982) On the pade tablé and its relationship to the r and s arrays and arm a modeling. Commun Stat-Theory Methods 11(12):1335–1379

Unbehauen H, Göhring B (1974) Tests for determining model order in parameter estimation. Automatica 10(3):233–244

Van den Boom AJW, Van Den Enden AWM (1974) The determination of the orders of process-and noise dynamics. Automatica 10(3):245–256

Wabomba MS, Mutwiri MP, Fredrick M (2016) Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (ARIMA) models. Sci J Appl Math Stat 4(2):64–73

Wadi SAL, Almasarweh M, Alsaraireh AA, Aqaba J (2018) Predicting closed price time series data using ARIMA Model. Modern Appl Sci 12(11):181–185

Wang CC (2011) A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export. Expert Syst Appl 38(8):9296–9304

Wang JH, Leu JY (1996) Stock market trend prediction using ARIMA-based neural networks. IEEE Int Conf Neural Networks 4(6):2160–2165

Wang JJ, Wang JZ, Zhang ZG, Guo SP (2012) Stock index forecasting based on a hybrid model. Omega 40(6):758–766

Wang P, Zhang H, Qin Z, Zhang G (2017) A novel hybrid-Garch model based on ARIMA and SVM for PM2. 5 concentrations forecasting. Atmos Poll Res 8(5):850–860

Wang YW, Shen ZZ, Jiang Y (2018) Comparison of ARIMA and GM (1, 1) models for prediction of hepatitis B in China. PLoS ONE 13(9):e0201987

Wei W, Jiang J, Liang H, Gao L, Liang B, Huang J, ... Chen H (2016) Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng County, China. PloS one, 11(6):e0156768

Weigend AS, Gershenfeld NA (1994) Time Series Prediction: Forecasting the Future and Understanding the Past Addison-Wesley. Reading

Weiss E (2000) Forecasting commodity prices using ARIMA. Techl Anal Stocks Commod 18(1):18–19

Williams BM, Hoel LA (2003) Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. J Transpo Engg 129(6):664–672

Wilson G (1969) Factorization of the covariance generating function of a pure moving average process. SIAM J Numer Anal 6(1):1–7

Woodward WA, Gray HL (1981) On the relationship between the S array and the Box-Jenkins method of ARMA model identification. J Am Stat Assoc 76(375):579–587

Yaziz SR, Azizan NA, Ahmad MH, Zakaria R (2016) Modelling gold price using ARIMA-TGARCH. Appl Mathl Sci 10(28):1391–1402

Yule GU (1921) On the time-correlation problem, with especial reference to the variate-difference correlation method. J Roy Stat Soc 84(4):497–537

Yule GU (1927) On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 226(636–646), 267–298

Zakaria S, Al-Ansari N, Knutsson S, Al-Badrany T (2012) ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. J Earth Sci Geotech Eng 2(3)

Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175

Zhang L, Lin J, Qiu R, Hu X, Zhang H, Chen Q, Wang J (2018) Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model. Ecol Indicat 95:702–710

Zhang XD, Zhang YS (1993) Singular value decomposition-based MA order determination of non-Gaussian ARMA models. IEEE Trans Signal Process 41(8):2657–2664

Zhu B, Wei Y (2013) Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega 41(3):517–524

Zhu S, Lian X, Liu H, Hu J, Wang Y, Che J (2017) Daily air quality index forecasting with hybrid models: a case in China. Environtl Poll 231:1232–1244