Arabian Journal of Geosciences (2021) 14:797
https://doi.org/10.1007/s12517-021-07135-x

ORIGINAL PAPER

Modeling and forecasting of principal minerals production

Sunila Saadat[1] · Ijaz Hussain[1] · Muhammad Faisal[2,3]

Received: 29 July 2019 / Accepted: 22 April 2021

Abstract

Although coal reserves are abundant in Pakistan, still share of gas and oil is about 65% in the energy mix. Pakistan, despite being a mineral-enriched country, is facing an alarming situation as its power generation is based on foreign exchange. The mineral sector of Pakistan is dominated by four principal minerals which are gas, oil, gypsum, and coal, while gypsum being a source of reclamation and poverty reduction. There is a strong need to analyze and forecast the production of these four principal minerals to cope up with emerging challenges. Data contains 114 observations from the period of July 2005 to December 2014, measured in terms of metric tonnes (MT). In parametric models, Box-Jenkins (B-J) methodology, a regression model with auto-regressive errors (ARAR), and Holt-Winter (HW) method are used to model. In nonparametric models, univariate singular spectrum analysis (SSA) and multivariate SSA (MSSA) modeling approach are applied. Data is split into train and test data in order to specify a suitable model for forecasting. Root Mean Square Error (RMSE), Mean Absolute Percentage Error, and Theil's U statistic are utilized as the measure of accuracy. For gas and coal, HW model is a suitable model to forecast. For gypsum and oil, Auto-regressive Integrated Moving Average (Box Jenkins ARIMA) and MSSA provide more accurate predictions, respectively. The forecasts for gas and gypsum as compared to 2014 are expected to be approximately 11 % and 45 %, respectively, more in 2020. In 2020, the forecasts of oil are expected to be eight times more than in 2014. The production of coal in 2020 is expected to decrease 12 % than in 2014. There is a strong need to optimize the production of coal by providing incentives for exploration and mining. The stakeholders should make serious efforts to bring the production of coal at an optimum level such as by providing modern equipment and high incentives to promote coal mining and exploration.

Keywords Forecasting principal mineral · Box Jenkins (B-J) methodology · Holt's method · ARAR model · Singular spectrum analysis · Multivariate SSA

Introduction

Economic growth and natural mineral resources wealth are positively correlated (Dollar and Kraay 2002). The infrastructure of any country is an indication of its richness and optimum utilization of natural resources in the development of the economy. The difference between developing and the developed economy is due to the lack of mineral resources, its exploitation, and utilization. The major sources for generating energy are natural gas, crude oil, and coal in the whole world. The contribution of crude oil is 39% followed by natural gas 33% and coal 28 % in the world's energy (Hannah Ritchie 2018). Today, in the whole world, coal consumption in energy generation has decreased by less than 30 % from 96% in the nineteenth century. The relative mix of these energy generation sources may vary among countries according to their economic condition. In China, 70% of fossil fuel consumption source is coal, whereas in Argentina, coal contribution is less than 2 % and gas accounting for 60 %. Global coal production

Responsible Editor: Domenico M. Doronzo

✉ Ijaz Hussain
ijaz@qau.edu.pk

Sunila Saadat
sunilasaadat1@gmail.com

Muhammad Faisal
M.Faisal1@bradford.ac.uk

1 Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan

2 Faculty of Health Studies, University of Bradford, Bradford BD7 1DP, UK

3 Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

Published online: 30 April 2021

has peaked in 2013–2014 being a first source of fossil fuel energy. These three major minerals natural gas, crude oil, and coal are substitutes and complements for each other and also rivals in production. So, their production may also be dependent.

Pakistan has the 7th largest coal reserves in total primary coal production (Outlook et al. 2010). Gas and oil resources are depleting but contributing more than 65% in power generation. The share of coal in the overall energy mix is continuously declining by 68% in 1948 to 5% in 2002 (Hussain 2003). Natural gas is depleting in Pakistan, and demand is rising. Demand for 8 to 10 billion cubic feet per day (*bcfd*) is expected by 2030 against the current production level of 3.8 *bcfd*[1]. Oil is found in raw form and fulfills only 15 % of its total requirement, while the rest is imported. In 2013, Pakistan import oil 372,800 *bbl/day* and ranked 25th in the list of oil import countries. Gypsum is a great source of poverty reduction, development of backward areas, reclamation of sodic soil, and in construction of low-cost cheap houses (Zami and Lee 2010). Annually, gypsum required by agriculture sector is 28 million tons, while total deposits of gypsum are 5 to 6 billion tons. The annual production of gypsum is 0.359 million tons, but demand is 55.3 million tons/annum respectively, which indicating huge difference between demand and supply. Pakistan's coal reserves meet only 11% of its total requirement. About 80% of the cement industry has now switched over to indigenous coal from furnace oil that has saved considerable foreign exchange being spent on the import of furnace oil (Satti et al. 2014). The mining industry tendency needs to be assessed for making effective policies for technical and financial management. The positive and negative consequences of the mining sector can be quickly addressed by transmitting internationally. Due to an increase in exports and prices of minerals, various low- and middle-income mineral enrich countries have attained development. Pakistan being mineral enriched developing country is facing the worst situation as its power generation is relying mainly on foreign exchange and its sources are not utilizing properly. The contribution of the mineral sector is 0.5% of GDP and is expected to increase considerably on development of mineral projects.

Schindler and Zittel (2008) aimed to forecast future supply of crude oil until 2030. They concluded that production is maximum in 2006 and then starts declining. Also, global oil production will be much lower and hard to cover the supply gaps by other substitutes in 2030. Mohr and Evans (2009) developed a model which is capable of forecasting mineral resources. The model considers supply and demand interactions applied to all coal producing countries. They concluded that global coal production will peak between 2010 and 2048 on a mass basis while peak between 2011 and 2047 on an energy basis. The peak of ultimately recoverable resources will be in 2034 on mass and 2036 on an energy basis. Mohr and Evans (2011) analyzed the production of global natural gas and concluded that peak of world natural gas production lies between 2025 and 2066 by assuming six scenarios, three static (demand and supply do not interact) and three dynamic. Rutledge (2011) analyzed global coal production by curve fitting on production history. They concluded that total coal production will be 680 metric ton. It also indicates 90% of production will take place by 2070, so alternatives should be considered. McMahon and Moreira (2014) concluded that in addition to economic growth, countries rich in minerals other than oil have experienced significant improvements in their Human Development Index (HDI) scores that are on average better than those experienced by countries without minerals.

Kumar and Shahbaz (2012) examined the association between coal consumption and economic growth for Pakistan over the period 1971 to 2009. They declared that coal consumption is positively related to economic growth which implies that energy conservation policies tend to lower economic growth and recommended to stimulate coal mining. They also suggested adopting appropriate technology to convert coal into natural gas to limit carbon dioxide emissions. They also suggested adopting Fischer-Tropsch technology to save a huge amount on oil imports. Shahbaz et al. (2013) studied the relationship between natural gas consumption and economic growth of Pakistan. They also included capital, labor, and exports using Auto-regressive Distributive Lag (ARDL) approach and innovative accounting approach to examine cointegration and causality among variables. Their found natural gas consumption, capital, labor, and exports are positively affecting economic growth in Pakistan. They concluded that natural gas conservation policies may perform as obstacles for economic growth. Ahmed et al. (2017) analyzed effects on oil price shocks on industrial level production and concluded the negative impacts of oil price shocks using the vector autoregression (VAR) model. They recommended forecasting the price of oil to control its impact on industrial production. Aized et al. (2018) analyzed renewable energy policy that ensures a sustainable supply of natural gas and electricity in Pakistan using Long-Range Energy Alternate Planning (LEAP) system. They assumed four scenarios and concluded green Pakistan scenario as the most suitable option that has minimum operation and externality cost.

Ahmed et al. (2016) reviewed energy sectors and possible sharing opportunities among China, India, and Pakistan (CIP). They discussed Renewable Energy Sources (RES), i.e., hydro, solar, wind, biomass, nuclear, and geothermal

[1] https://tribune.com.pk/story/1741692/2-pakistan-needs-double-gas-production-5-7-years/

subject to energy requirements. They recommended that the exploitation of CIP energy potential and energy sharing opportunities will support global peace and prosperity. Rehman et al. (2017b) presented simulations of future production of natural gas, crude oil, and coal in Pakistan. For this purpose, they analyzed production of past 45 years primary energy resources data by using a generic Systems Thinking, Experimental Learning Laboratory with Animation (STELLA) model. Their results depicted that Hubbert peak of crude oil has been achieved in 2013 with 4.52 million toe, the peak production of natural gas is expected in 2024 with 32.70 million toe, and the peak production of coal is expected in 2080 with an estimated production of 134.06 million. They recommended energy security policy for sustained supply of energy resources. Rehman et al. (2017a) conducted a study to forecast electricity, natural gas, oil, coal, and liquid petroleum gas (LPG) for all sectors of Pakistan economy. They applied different forecasting methodologies: Auto-regressive Integrated Moving Average (ARIMA), Holt-Winter, and LEAP. They compared their demand forecasts with actual annual demand and found as the most appropriate technique for energy demand forecasting. They concluded that in 2035, the contribution of different sources in energy demand will be as 38.16 % for oil, natural gas 36.57 %, electricity 16.22 %, coal 7.52 %, and LPG 1.52 %. Akhtar et al. (2018) reviewed the energy potential from different sources and claimed that Pakistan can produce 100,000MW of electricity from Thar coal for 20 years, and 11,500 MW energy is expected to be produced by coal fired plants in 2017 to 2019. Ejaz et al. (2018) analyzed and forecast existing energy projects and CPEC using LEAP model by assuming three scenarios from 2013 to 2020. Coal scenario is concluded as suitable for economic and environmental effects in order to achieve high electrical energy in the future. Rehman and Deyuan (2018) analyzed and forecast consumption and supply of commercial energy sources that includes petroleum and oil, natural gas, coal, and electricity. They concluded that oil and natural gas production policies are required for economic development. They also suggest greenhouse gas mitigation technologies and consideration of alternative energy sources to increase coal production.

Nawaz et al. (2019) concluded cointegration between financial development and economic growth using the production function from 1972 to 2017. Their results revealed a bidirectional causality between financial development and economic growth. In this paper, we analyzed monthly production of natural gas, oil, gypsum, and coal for Pakistan and their relationship. Also, the interest of the study lies in how the production of observed minerals changes over time and what will be their expected production in the next years? Our study is based on modeling and forecasting monthly production of gas, oil gypsum, and coal for Pakistan to get them answered.

Data and methodology

Description of data

The secondary monthly production data of four major minerals, natural gas, crude oil, gypsum, and coal, are considered. There are 114 observations comprising of monthly production of observed minerals measured in terms of metric tons (MT) for the period of July 2005 to December 2014. It is available on the official website of Pakistan Bureau of Statistics (http://www.pbs.gov.pk/energy-and-mining-tables). The data is collected from:

1. Provincial Directorates of Mines and Minerals Punjab, Sindh, KPK, Balochistan, and FATA
2. Directorate General of Petroleum Concession, M/O Petroleum, and Natural Resources Islamabad

Models

Seasonal Auto-regressive Integrated Moving Average (SARIMA) model, Holt's model, a regression model with auto-regressive errors (ARAR) algorithm and singular spectrum analysis (SSA), and multivariate singular spectrum analysis (MSSA) approach are considered to forecast production of observed mineral resources. The brief description along with merit and demerits of each model is presented in Table 1. The detailed description of the models is described as follows.

Seasonal and nonseasonal ARIMA model

Seasonality is a regular pattern repeated over a fixed interval of time (such as monthly or quarterly). Seasonal data is analyzed by taking seasonal difference to remove seasonal trend. Seasonal Auto-regressive Integrated Moving Average (SARIMA) models are a linear combination of past values, integrated terms, and error terms. It allows taking previous values and the seasonal lag values for forecasting next values. It comprises of two parts, i.e., seasonal and nonseasonal. It acts as a filter that separates noise from the signal and then utilize signal to make a prediction. It is a regression-type model in which seasonal lag of the dependent variable and/or seasonal lag of forecast error is used as predictors. If predictor only contains seasonal lag values of the variable, then it is a "seasonal auto-regressive model." If predictor contains seasonal lag values of error terms, then it is not a linear model as error terms cannot be interpreted as an independent variable. It is denoted as $SARIMA\ (p,d,q)(P,D,Q)$, where p,q represents the number of auto-regressive and moving average terms, P and Q represent the number of seasonal auto-regressive terms and seasonal moving average terms, and d

Table 1 Summarized time series models with merits and demerits

Characteristics	Parametric models			Nonparametric model	
Names	Box-Jenkins	Holt-Winters	ARAR algorithm	SSA	MSSA
Functional form	Linear combination of past values, integrated terms, and error terms	Exponential method for forecasting time series	Applying the AR model on memory shortening transformations	Model is fitted to noise reduced data	SSA in a multivariate framework
Main components	Seasonal and nonseasonal components	Estimation of forecasting, three equations of level (α_t), trend (β_t), and seasonal (S_t) utilizing smoothing parameters to forecast the series	Memory shortening process then fitting AR process to mean corrected series and combining both to get forecast equation	Decomposition and reconstruction	Decomposition and reconstruction Single modification between SSA and MSSA is of trajectory matrix which is block Hankel
Weights	Forecast error is used as predictors	More weights to the recent values than historic values so more precision is attained	Accounts for shock effect in long memory	"Separability," which refers to the intensity of extracting the components from time series	Understand explained patterns in the multivariate framework
Seasonality	Seasonal lag values of error terms are involved, and then it is not a linear model	Seasonal value is a weighted average of seasonal index of current the time and of last year	Seasonal effects are easily handled	Accounts for every component of filtered time series	Analysis of interrelation and co-movements of several time series variables in the same duration
Mathematical expression	$\phi(B^s)\varphi(B)\Delta_s{}^D\Delta^d Y_t = \vartheta(B_s)\theta(B)\varepsilon_t$	$Y_{t+h} = \alpha_t + \beta_t h + s_{t-m+h^*}$	$\varpi(\mathrm{B})Y_t = \pi_1(\overline{M}) + \varepsilon_t$	$Y = Y^{(1)} + Y^{(2)} + \ldots + Y^{(M)}$	$Y = Y^{(1)} + Y^{(2)} + \ldots + Y^{(M)}$
Merit	Basic approach to analyze univariate series by Box Jenkins methodology	Being able to adapt changes in trend with seasonal and cyclic fluctuations	Parsimonious, less restrictive, and has simple inference	Appropriate technique to overcome the issue of small sample size	It helps to reveal ghost limit cycles with higher confidence than univariate analysis
Demerit	Modeling approach includes many parameters	Holt's only forecast about past information	It depends on cyclic variations of past data	Difficult to extract weak impact components in a strong noise	M-SSA analysis suffers from a degeneracy problem
Utilization of data	Don't utilize full information of the data due to difference operator	Utilizes exponential smoothing to encode a lot of data	Utilizes all the information of the data	Extracts signal and then information is utilized to model	Extracts signal and then information is utilized to model

and D represent the number of difference and seasonal difference taken to be stationary.

$$\phi(B^s)\varphi(B)\Delta_s{}^D\Delta^d Y_t = \vartheta(B_s)\theta(B)\varepsilon_t \tag{1}$$

where $\phi(B^s)$ is a vector of seasonal auto-regressive coefficients $\phi(B^s) = (1 - \phi_1 B^s - \phi_2 B^{2s} - \cdots - \phi_P B^{Ps})$.

$\varphi(B)$ is a vector of auto-regressive coefficients $\varphi(B) = (1 - \varphi_1 \mathrm{B} - \varphi_2 B^2 - \cdots - \varphi_P B^P)$.

$\Delta_s{}^D$ is a seasonal difference transformation $\Delta_s{}^D = (1 - B^s)^D$.

Δ^d is difference transformation $\Delta^d = (1 - \mathrm{B})^d$.

$\vartheta(B_s)$ is a vector of seasonal moving average coefficients $\vartheta(B_s) = (1 + \vartheta_1 B^s + \vartheta_2 B^{2s} + \cdots + \vartheta_Q B^{Qs})$.

$\theta(B)$ is a vector of moving average coefficients $\theta(B) = (1 + \theta_1 \mathrm{B} + \theta_2 B^2 + \cdots + \theta_Q B^Q)$.

Holt's method

Box-Jenkins modeling approach includes many parameters; Holt (2004) modified simple exponential method for forecasting time series. Holt's method is simple and generally performs better. Results attained by Holt's method are more precise as its dependence is low on historic data and depends on all previous values by giving more weights to the recent values. Holt's method outperforms *SARIMA* when cyclic variations also exist in seasonal data. It considers level (α_t) and trend (β_t) smoothing values to forecast the series.

$$Y_{t+h} = \alpha_t + \beta_t h, \tag{2}$$

where,

$$\beta_t = \mu(\alpha_t - \alpha_{t-1}) + (1-\mu)(\beta_{t-1})$$
$$\alpha_t = \lambda(Y_t) + (1-\lambda)(\alpha_{t-1} + \beta_{t-1})$$

Winters (1960) extended Holt's method for forecasting seasonal time series by considering two types of seasonality depending on the variations of series, i.e., additive and multiplicative. The additive method is implied when variations in the series are almost the same, whereas multiplicative method is implied when variations of the series are proportional to time. Holt-Winters (HW) method considers estimation of forecasting and three equations of level, trend, and seasonality utilizing smoothing parameters.

$$Y_{t+h} = \alpha_t + \beta_t h + s_{t-m+h^*} \tag{3}$$

$$\text{where} \quad h^* = (\mathrm{h}-1)|\mathrm{m}| + 1$$
$$\beta_t = \mu(\alpha_t - \alpha_{t-1}) + (1-\mu)(\beta_{t-1})$$
$$\alpha_t = \lambda(Y_t - S_{t-m}) + (1-\lambda)(\alpha_{t-1} + \beta_{t-1})$$
$$S_t = \nu(Y_t - \alpha_{t-1} - \beta_{t-1}) + (1-\nu)(S_{t-m})$$

h^* indicates the seasonal indices of last year that is utilized for forecasting values. Trend (β_t), level (α_t), and seasonal (S_t) are smoothing equations, and μ, λ and ν are their smoothing functions, respectively. Trend value is the same as in Holt's method where level for forecasting period is a weighted average of current seasonal adjusted value and current nonseasonal value. Seasonal value is a weighted average of seasonal index of the current time and of last year.

Regression model with auto-regressive errors (ARAR model)

SARIMA utilizes differencing in which information contained in the data is not fully utilized. For this purpose, Parzen (1982) developed *ARARMA* model to overcome this issue. The methodology is then adapted to develop ARAR algorithm developed to forecast the time series. It is parsimonious and less restrictive and has simple inference used to forecast. It accounts for shock effect in long memory than to eradicate its effect. ARAR algorithm is based on applying the AR model on memory shortening transformations.

ARAR algorithm Mathematically, the ARAR algorithm is defined in Brockwell and Davis (2016) as:

Step 1. The time series is identified on the basis of long memory or a short memory. For long memory time series data, difference transformations are applied to shortening memory. The short-term memory process can be expressed as:

$$M_t = Y_t + \pi_1 Y_{t-1} + \cdots + \pi_i Y_{t-i} = \pi(B) Y_t \tag{4}$$

where $\pi_i{}'s$ coefficients are memory shortening process. M_t is memory-shortened series.

Step 2. The AR model is applied on $Z_t = M_t - \overline{M}$ and can be expressed as:

$$Z_t = \psi_1 Z_{t-1} + \psi_{l_1} Z_{t-l_1} \quad + \psi_{l_2} Z_{t-l_2} + \psi_{l_i} Z_{t-l_i} + \varepsilon_t \tag{5}$$
$$\psi(\mathrm{B}) \mathrm{Z}_t = \varepsilon_t$$

where the error term (ε_t) is normally distributed as $N(0, \sigma^2)$.

Step 3. Forecasts are made by combining both models of Eqs. (4) and (5).

$$\varpi(\mathrm{B}) = \pi(\mathrm{B})\psi(\mathrm{B})$$

The final model utilized for forecasting can be represented as follows:

$$\varpi(\mathrm{B}) Y_t = \pi_1 \left(\overline{M}\right) + \varepsilon_t \tag{6}$$

Singular spectrum analysis

SSA is a nonparametric technique for decomposing time series and extracting trend, seasonality, and periodicity. In earlier approach, model is fitted directly on noisy series, whereas in SSA approach, noise component is filtered, and then the model is fitted to noise reduced data. It provides better results if the observed series is very noisy. It is an appropriate technique to overcome the issue of small sample size. It involves two main steps decomposition and reconstruction. Each of the main steps has two sub-steps. In decomposition, time series is decomposed into its components and then again reconstruct the decomposed components to forecast. The main focus of SSA is "separability," which refers to the intensity of extracting the components from time series. SSA does not employ any assumption of stationary or any assumption of residuals. It considers the bootstrapping method for making a confidence interval. SSA approach depends on historical data and performs well when the data is large. SSA algorithm considers the following steps:

- Decomposition: It involves two sub-steps to map and decompose the observed series.

1. Decomposition into trajectory matrix:

Time series is mapped into L lagged vectors, and L is the parameter to be selected for the embedding of the original series and ranges between $2 \leq L \leq \frac{T}{2}$. A single series is decomposed into K multidimensional series each of lag L, where $K = T-(L-1)$. Trajectory matrix is represented as:

$$\overline{T_y} = \begin{bmatrix} Y_1 & Y_2\cdots & Y_K \\ Y_2 & Y_3\cdots & Y_{K+1} \\ \vdots & \vdots & \vdots \\ Y_L & Y_{L+1} & \cdots Y_T \end{bmatrix}$$

Each column of $\overline{T_y}$ is a subset of original series. Elements are the same at cross diagonal, so it is Hankel matrix.

2. Singular value decomposition (SVD)

Trajectory matrix is decomposed by computing eigenvalues and eigenvectors of $\overline{T_y}\,\overline{T_y}'$, so that it can be represented as:

$$\overline{T_y}\,\overline{T_y}' = \mathrm{A}\Lambda\mathrm{B}'$$

Here, Λ is diagonal matrix of eigenvalues in decreasing order $\sqrt{\lambda_1} > \sqrt{\lambda_2} > \cdots > \sqrt{\lambda_L}$, where

A and B′ are orthonormal matrix. Vectors of matrix A are eigenvectors, and B′ are of factor vectors which can be represented as:

$$B = \frac{\overline{T_y}'\mathrm{A}}{\lambda}$$

Each element of $\overline{T_y}$ can be represented as eigen triples and decomposed into "m" values where

$$m = \max\left[i, \sqrt{\lambda_i} > 0\right]$$
$$\overline{T_y} = \overline{T_{1y}} + \overline{T_{2y}} + \cdots + \overline{T_{my}}$$
$$\overline{T_y} = A_1\sqrt{\lambda_1 B_1'} + A_2\sqrt{\lambda_2 B_2'} + \cdots + A_m\sqrt{\lambda_m B_m'}$$
$$\overline{T_y} = \sum_{i=1}^{m} A_i\sqrt{\lambda_i B_i'}$$

- Reconstruction: The series is reconstructed in two steps ignoring noise component.

1. Eigen triple grouping

In grouping, the trajectory matrix is grouped into trend, seasonality, periodicity, and noise. After attaining SVD of trajectory matrix, the decomposed subsets are grouped in "s" disjoint subsets by considering separability. Each group can be represented as:

$$\overline{T_{iy}} = \overline{T_{1iy}} + \overline{T_{2iy}} + \cdots + \overline{T_{siy}}$$

Trajectory matrix can be represented as:

$$\overline{T_y} = \overline{T_{1y}} + \overline{T_{2y}} + \cdots + \overline{T_{sy}}$$

2. Diagonal averaging

The averaging over diagonal of each grouped matrix yield to series of the original length. It transforms the grouped matrix into a Hankel matrix. The obtained values of the series are additive components of the original series. The resultant series can be represented as:

$$Y = Y^{(1)} + Y^{(2)} + \ldots + Y^{(M)}$$

where $Y^{(1)}$ always becomes a trend value, and the rest of the elements are not fixed, but they represent periodicity and seasonality.

Multivariate singular spectrum analysis

MSSA is a generalization of SSA in a multivariate framework. If the observed series are not independent, then it is required to observe and analyze them simultaneously. Multivariate time series analysis is the analysis of interrelation and co-movements of several time series variables in the same duration. MSSA extracts information of observed series to understand explained patterns in the multivariate framework (Groth and Ghil 2011). It is a nonparametric technique and useful approach to control the problem of sample size. Like SSA, it has two main phases: decomposition and reconstruction. The observed time series is disintegrated into its constituents in decomposition and then restructure by ignoring noise part of the disintegrated components to projection. The single modification between SSA and MSSA is of trajectory matrix which is block Hankel (Hassani et al. 2013).

- Embedding

In embedding, block Hankel trajectory matrix of order LxM where M = N − L + 1 is formed by subdivided Z_t into L dimensional embedding where $0 \leq L \leq \frac{N}{2}$. Trajectory matrix is formed for K variate time series. This can be expressed as vertical or horizontal MSSA.

$T = [T_1, T_2, \ldots, T_K]$ where, $T_i = T = [T_{1i}, T_{2i}, \ldots, T_{Ki}]$

- Singular value decomposition (SVD)

In this step, the SVD of T matrix is formed by eigenvalues and eigenvectors. let $\Lambda = \text{diag}\ [\lambda_1, \lambda_2, \cdots, \lambda_s]$ be eigenvalues and $V = [V_1, V_2, \cdots, V_s]$ are corresponding eigenvectors of TT', where $TT'' = V\ \Lambda V'$, and s denotes the rank of the matrix T.

$$T = V\ V' T = \sum_{i=1}^{s} V_i V_i' T = T_1 + T_2 + \ldots + T_s$$

where $T_i = V_i V_i' T$ is unitary matrix corresponding to i^{th} largest singular value $(\sqrt{\lambda_i})$.

- Grouping

In grouping, the focus is to separate signal and noise component as follows:

$$T = \widehat{S} + \widehat{\eta}$$

where $\widehat{S} = T_1 + T_2 + \ldots + T_r$ and r is the number of signal components $(r < s)$.

- Diagonal averaging

It is performed to transform reconstructed matrix T into the form of Hankel matrix that can be converted into time series. The arithmetic averaging $\widetilde{t_{i(p,q)}}$ over all (p,q) i.e., p + q−1 = j leads to j^{th} term of reconstructed series $\widetilde{T}_j$.

Results and discussion

Exploratory analysis

The scale of the production varies, and nonstationary behavior is clearly depicted by observed minerals (see Fig. 1). The highest and lowest value of gas production takes place in the first quarter of 2012 (i.e., March 2012 about 137017.9 MT) and third quarter of 2005 (i.e., September 2005 about 107865.4 MT), respectively. The production of gas series is seasonal with increasing trend until 2012. The positive trend discontinues after the highest level of production in 2012. This discontinuity is a result of an increase in wellhead prices of natural gas launched by the Government of Pakistan in 2012. The implementation of this policy resulted into an increase in gas prices for future discoveries and for fields not ready for production as consequences can be seen in September 2012. In February 2011, the Pakistan government imposed more taxes on gas production due to a substantial increase in gas consumption as compared to oil and coal in the energy sector.

One of the biggest barriers is that exploratory prices of gas are linked with crude oil. Producer (well head) prices of gas do not encourage the more exploration and production.

The production of oil increases in constant trend but possess an increasing trend after 2012. In competing with energy deficit, subsequent policies were not enough to provide an incentive for exploration and production. The Government of Pakistan offered a suitable price in 2012 petroleum policy

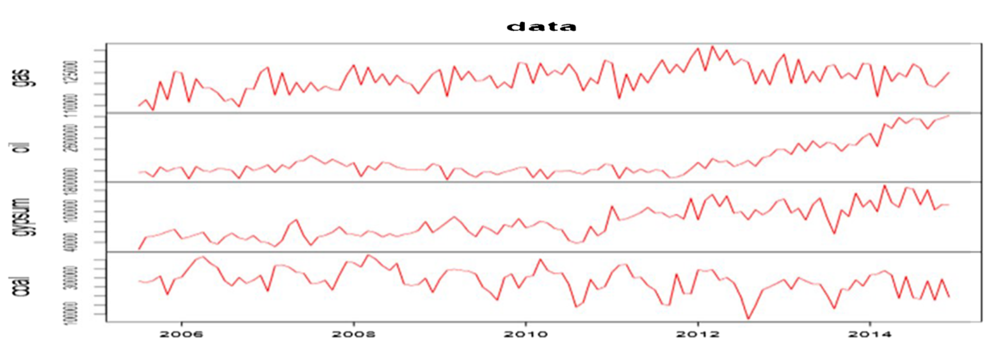

Fig. 1 Time series plot of monthly production for gas, oil, gypsum, and coal in Pakistan

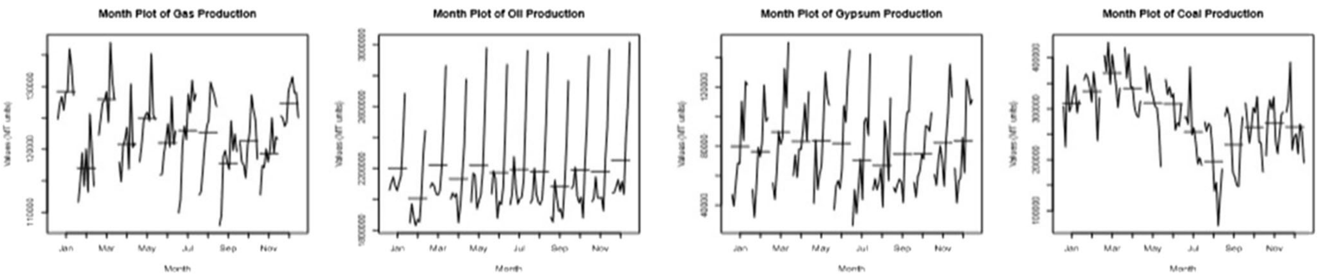

Fig. 2 Month-wise plot for production of gas, oil, gypsum, and coal in Pakistan

aimed to enhance exploration and production to overcome energy deficit and thus resulted in increasing production.

The production of gypsum increases with short-term nonseasonal fluctuations. In 2012, the price of gypsum increased about US 1.50 $/t for local consumers after the opening of Wagha border because of exporting gypsum to neighboring India in massive quantities.

The coal production has also seasonal fluctuations with constant trend until 2008. Then, trend decreases with a sudden drop at a minimum level of production in August 2012 (i.e., 70884 MT) and continues in constant trend. The maximum level of coal production lies in March 2008 (i.e., 430430 MT). The present rate of excavation which averages 3.2 million tons annually, where the estimated coal reserves in Pakistan are about 183 billion tons and the drill-proven reserves are estimated at 579 million tons, are expected to last for 180 years[2]. There are no incentives provided to enhance coal exploration and production. The technology used for coal extraction is even worse than at the time of independence. According to the annual report of the State Bank of Pakistan (2010–2011), the short-term fluctuations in mining productions are caused by floods, technical faults, and field maintenance which can take 30 days at a time.

Gas The most left part of Fig. 2 shows that the minimum average production of gas occurs in February, whereas the highest average production takes place in January, March, and December. Gas production follows increasing and decreasing pattern alternatively except for July and August. The consumption of gas has usually two seasonal peaks: in summer and winter. To balance these seasonal swings, gas is placed in storage that has the highest level between October and November.

Oil The second left subfigure of Fig. 2 shows that the highest average production of oil takes place in January, March, May, and December, whereas the lowest average production takes place in February and then in September. The production pattern follows a seasonal pattern like gas except for June, July, and August.

[2] http://reviewandanalysis.blogspot.com/2007/10/pakistan-is-saudi-arabia-of-coal.html

Gypsum The second right subfigure of Fig. 2 shows that the average production of gypsum is highest in March, whereas the lowest production takes place in August and then in July. The average production of gypsum is almost equal in April, May, June, November, and December. It is also equal in September and October which is quite less than January and then in February. The pattern of average production depicts nonseasonality in production.

Coal The right subfigure of Fig. 2 shows that the average production of coal increases from January to highest in March and then starts decreasing with lowest in August. After August, the average production increases until November and then slightly decreases in December.

SARIMA modeling through Box-Jenkins methodology

Box and Jenkins developed the methodology to model the observed series through the SARIMA model by considering the following steps:

Identifying stationary and model parameters

Diebold and Kilian (2000) suggested that unit root pretesting is preferable for increasing forecasting accuracy. The test takes nonstationary or unit root as null hypothesis except for Kwiatkowski Phillips Schmidt Shin (KPSS) test which accounts for stationary as the null hypothesis. The most frequent strategy is to apply different tests and to conclude the same results. By the unit root pretesting, there is no evidence for rejecting the null hypothesis. Seasonal and/or nonseasonal difference transformations are applied to confirm stationary of series. The difference order should not exceed 2 including both seasonal and nonseasonal difference.

Unit root test statistics A hypothesis that the series contains unit root is rejected and concluded that transformed series are free from the unit root (see Table 2). As all series for unit root Dickey-Fuller (ur.df) test have a smaller value of test statistics than the critical value at 1 % which is −2.58 leading the rejection of the null hypothesis. The test statistic value of Phillips-Perron unit root (ur.pp) test is smaller than the critical value at

Table 2 The unit root test-statistic for stationary transformed series of gas, oil, gypsum, and coal

Series	Difference	Unit Root Dickey-Fuller (ur.df)	Unit Root Phillips Perron (ur.pp)	Unit Root Kwiatkowski Phillips Schmidt Shin (ur.kpss)
Gas	Seasonal	−3.0859	−4.5518	0.2996
Oil	First and seasonal	−9.5162	−12.6447	0.0953
Gypsum	First	−11.0516	−20.2827	0.0301
Coal	Seasonal	−5.4096	−7.7985	0.0882

1 % that is −3.4957 leading rejection of the null hypothesis. The value of the test statistic for unit root KPSS test is smaller than the critical value at 10 % which is 0.347 leading failure to reject the null hypothesis. From unit root testing, strong evidence exists against the presence of unit root. The series of gas and coal are transformed into stationary series by applying seasonal difference, whereas oil is transformed by taking difference at both seasonal and nonseasonal lag. Gypsum is transformed into stationary series by taking first-order difference only due to nonseasonal fluctuations.

Autocorrelation function (ACF) and partial autocorrelation function (PACF) In Fig. 3, ACF of gas depicts stationary behavior, and the alternative pattern suggests that the ARMA model is appropriate for the nonseasonal component where PACF of gas cuts off after lag 1 indicating $q - p = 1$. There is an alternative pattern that exists for the seasonal lag in ACF, whereas PACF cuts off at lag 12 indicating $Q-P = 0$. So, *SARIMA* of order $(1, 0, 2)(1, 0, 1)_{12}$ is an appropriate model for gas series. For oil, there is a significant negative spike at lag 1 and at lag 12 in ACF, whereas PACF dies down after specific lags indicating MA and SMA of order 1. So, *SARIMA* of order $(0, 0, 1)(0, 0, 1)_{12}$ is appropriate for oil series. For gypsum, ACF cuts off, and PACF dies down after lag 2 indicating MA model of lag 2 for nonseasonal component, and there is no significant spike at seasonal lags indicating that there is no seasonality in series. So, *ARIMA* (0,0,2) is a suitable order for modeling gypsum series. For coal, there is no specific pattern for the nonseasonal component. In PACF, there are negative spikes at seasonal lags, whereas in ACF, there is a significant negative spike at lag 12 that indicates to include SMA of order 1 in the seasonal component. So, *SARIMA* with a nonseasonal component of order (0,0,0) and a seasonal component having an order $(0, 0, 1)_{12}$ is suggested an order for modeling coal series.

Modeling

The suggested models are applied for each original series. The difference order required for stationary series is directly applied to original data along with the suggested order of the model. The significance of estimated coefficients is checked through *z*-test. The seasonal period is taken as 12, as data is monthly.

- Gas:

$$\begin{aligned} Y_t = {} & 87.1893\text{t} + 0.5179Y_{t-1} \\ & + 1.1159Y_{t-12} - 0.5779Y_{t-13} - 0.1159Y_{t-24} \\ & + 0.06Y_{t-25} + 0.0081\epsilon_{t-1} \\ & + 0.3883\epsilon_{t-2} - 0.7569\epsilon_{t-12} - 0.06\epsilon_{t-13} \\ & - 0.2939\epsilon_{t-24} \end{aligned}$$

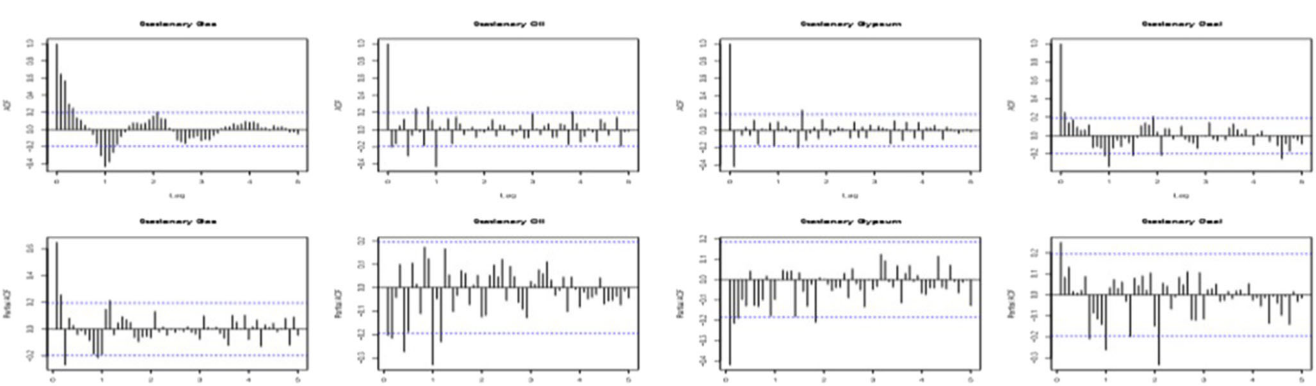

Fig. 3 Auto-correlation function (ACF) and partial auto-correlation function (PACF) of stationary series of gas, oil, gypsum, and coal where upper plots are of ACF and lower plots are of PACF

- Oil:

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} - 0.2374\epsilon_{t-1} - 0.6284\epsilon_{t-12} + 0.1491\epsilon_{t-13}$$

- Gypsum:

$$Y_t = 690.8422t + Y_{t-1} - 0.6969\epsilon_{t-1} - 0.1867\epsilon_{t-2}$$

- Coal:

$$Y_t = Y_{t-12} - 767.2925t - 0.5522\epsilon_{t-12}$$

SARIMA $(1,0,2)(1,1,1)_{12}$ with drift is estimated for gas. All estimated coefficients are significant except for the first order moving average term (ma 1) and first seasonal autoregressive (sar 1) term, but second order moving average term (ma 2) is significant. For oil, SARIMA $(0,1,1)(1,0,1)_{12}$ is estimated, and all estimated coefficients are significant. ARIMA(0,1,2) with drift is estimated for gypsum, and all estimated coefficients are significant at 5% significance level. SARIMA $(0,0,0)(0,1,1)_{12}$ with drift is estimated for coal, and both coefficients are significant.

Holt-Winters method

As observed series are seasonal except gypsum, Holt-Winter's additive seasonal model Eq. (3) is utilized to forecast seasonal series, and Holt's linear trend model Eq. (2) is used to forecast gypsum. Additive seasonal model is implied because it outperforms the multiplicative method, and seasonality in the series is the same throughout the year.

- Gas:

$$\alpha_t = 0.5619(Y_t - s_{t-m}) + 0.4381(\alpha_{t-1} + \beta_{t-1}), \quad where\ \alpha_0 = 115882.75$$

$$\beta_t = 0.0001(\alpha_t - \alpha_{t-1}) + 0.9999\beta_{t-1}, \quad where\ \beta_0 = 180.049$$

$$s_t = 0.0001(Y_t - \alpha_{t-1} - \beta_{t-1}) + 0.9999 s_{t-m}, where\ s_0 = -2027.56, 1920.70, -2183.75,$$

$$5101.63, -5692.10, 6576.85, 4422.90, -3240.92, -1397.48, -4318.93, 161.744, 676.92$$

- Oil:

$$\alpha_t = 0.7246(Y_t - s_{t-m}) + 0.2754(\alpha_{t-1} + \beta_{t-1}), \quad where\ \alpha_0 = 1997854.69$$

$$\beta_t = 0.0333(\alpha_t - \alpha_{t-1}) + 0.9667\beta_{t-1}, \quad where\ \beta_0 = 5492.4457$$

$$s_t = 0.0001(Y_t - \alpha_{t-1} - \beta_{t-1}) + 0.9999 s_{t-m}, \quad where\ s_0 = -491.86, 54905.96, -23631.39,$$

$$72754.39, -133349, 69016.56, 44243.99, -20434.53, -1073.59, -89533.44, 4568.07, 23024.84$$

- Gypsum:

$$\alpha_t = 0.3179 Y_t + 0.6821(\alpha_{t-1} + \beta_{t-1}), \quad where\ \alpha_0 = 46221.10$$

$$\beta_t = 0.0047(\alpha_t - \alpha_{t-1}) + 0.9953\beta_{t-1}, \quad where\ \beta_0 = 1866.4815$$

- Coal:

$$\alpha_t = 0.0273(Y_t - s_{t-m}) + 0.9727(\alpha_{t-1} + \beta_{t-1}), \quad where\ \alpha_0 = 342061.26$$

$$\beta_t = 0.0001(\alpha_t - \alpha_{t-1}) + 0.9999\beta_{t-1}, \quad where\ \beta_0 = -748.9019$$

$$s_t = 0.0001(Y_t - \alpha_{t-1} - \beta_{t-1}) + 0.9999 s_{t-m},$$

$$where\ s_0 = 22728.95, 24678.58, 53028.76, 82012.19, 46261.09, 22098.42, -16111.57,$$

$$-9925.2, -19399.06, -67961.07, -101236, -36175.04$$

ARAR modeling

The assumption of stationary is not needed in ARAR modeling. It used an automated algorithm to model observed series, so seasonal and nonseasonal difference is not taken. It utilizes the AR model in algorithm and provides adjustments. It can be observed that the ARAR algorithm utilizes seasonal lag for gas and coal. It considers lag 1 for all observed series. For gas, lag 1, lag 2, lag 12, and lag 14 are considered as optimum lag and highest weight assign to lag 12. For oil, lag 1, lag 2, lag 7, and lag 12 is considered to be optimum, and lag 12 has the highest optimum coefficient. As gypsum is nonseasonal series, no seasonal lag is utilized lag 1, lag 2, and lag 3, and lag 17 is considered as optimum lag, and the coefficient of lag 1 is highest. For coal, lag 1, lag 12, lag 25, and lag 26 are considered to be optimum, and lag 12 has the highest optimum coefficient.

- Gas:

$$Y_t = 0.3906Y_{t-1} + 0.3626Y_{t-2} + 0.6018Y_{t-12} - 0.3927\,Y_{t-13} - 0.0978\,Y_{t-14} + 0.4058\,Y_{t-24} - 0.2682Y_{t-26}$$

- Oil:

$$\mathrm{Y_t} = 0.1168\mathrm{Y}_{t-1} + 0.7995\mathrm{Y}_{t-2} - 0.1178\mathrm{Y}_{t-3} + 0.2105\mathrm{Y}_{t-4} + 0.2287\mathrm{Y}_{t-7} - 0.2306\mathrm{Y}_{t-9} + 0.4260\mathrm{Y}_{t-12} - 0.4296\mathrm{Y}_{t-14}$$

- Gypsum:

$$\mathrm{Y_t} = 0.4356\mathrm{Y}_{t-1} + 0.2276\mathrm{Y}_{t-2} + 0.1245\mathrm{Y}_{t-3} + 0.174\mathrm{Y}_{t-4} - 0.1893\mathrm{Y}_{t-17} + 0.1858\mathrm{Y}_{t-18}$$

- Coal:

$$\mathrm{Y_t} = 0.2285\mathrm{Y}_{t-1} + 0.6386\mathrm{Y}_{t-12} - 0.2194\mathrm{Y}_{t-13} + 0.3082\mathrm{Y}_{t-24} - 0.3175\mathrm{Y}_{t-25} + 0.1868\mathrm{Y}_{t-26} + 0.3048\mathrm{Y}_{t-37} - 0.1793\mathrm{Y}_{t-38}$$

Singular spectrum analysis

SSA is a nonparametric approach and does not require any assumption of time series analysis, so it is directly applied on observed series.

Decomposition analysis of observed series

As window length "L" should be less than equal to half of the series length, $L = 45$ is selected for the decomposition of each series. It is recommended that if a periodic component of series is known, then L should be chosen as a multiple of that integer. The actual objective of decomposition analysis is to identify the pair of eigenvalues which characterize observed series into its components.

In Fig. 4, if the paired components depict any identifiable pattern, then these paired eigenvectors contribute to separating patterns from residuals of the observed series. For gas, pair 2 vs 3 represents star-shaped polygon indicating period 2.4. For oil, 4 vs 5 represents star-shaped polygon that depicts the result of a combination of modulated sine waves. Both gas and oil depict star structure polygon indicating a period of 2.4. For gypsum, 6 vs 7 pair represents five vertex polygon. The number of vertex of polygon represents the period of the sine wave. For coal, 2 vs 3 and 4 vs 5 represent regular polygon of vertex 12 and 4, respectively.

Reconstruction of observed series

As the observed series is decomposed into 45 eigen triple values, it implies that these components also contain noise component. It is important not to include many eigenvalues to model series without noise component. SSA is model-free approach; however, the performance of the reconstructed series can be compared with original series graphically. Trend component of series can be extracted by grouping slowly varying eigenvalues, whereas periodicity can be extracted by plotting paired eigenvector having polygon. If the length of the series increases, the separability of components also increases asymptotically.

In Fig. 5, the bottom line in each plot represents residuals for the corresponding series. Each series is reconstructed by pairing eigen triples as pair (1,2) and (2,3) for gas, (1,2) and (4,5) for oil, (1,2) and (6,7) for gypsum, and, 1, (2,3), (4,5) are selected for coal. The variations in SSA for gas, oil, and gypsum are much less as compared to original series, whereas for coal, SSA captures the variations of the observed series.

Multivariate singular spectrum analysis

MSSA is a useful nonparametric technique for filtering noise and signal components of any multivariate time series data. Now, all observed minerals are analyzed in a single system to determine the predicted series with the contribution of itself and also by other variables.

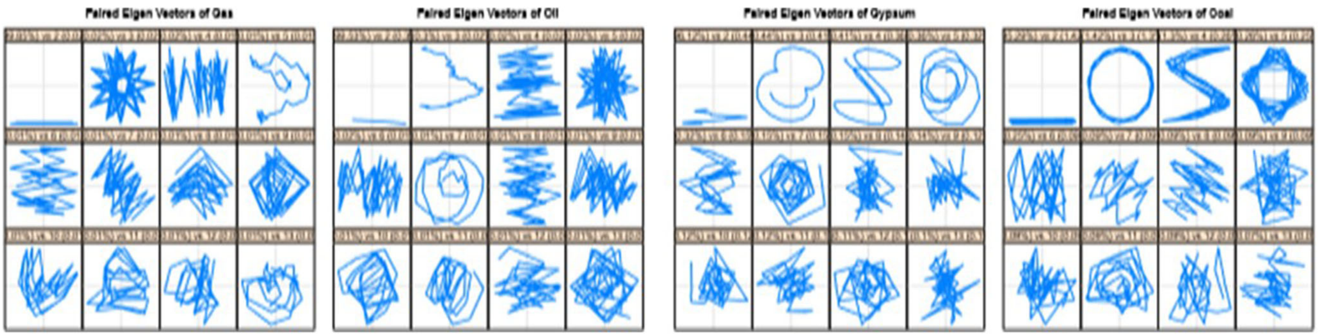

Fig. 4 Paired eigenvector plot in identifying systematic and irregular component

Granger causality test

From Table 3, results indicate that the prediction of gas can be improved by oil and gypsum. For oil, all three variables are helpful in improving prediction. For gypsum, gas improves the predictability. For coal, prediction cannot be improved by including any other variable. Also, there is bidirectional causality between (gas, oil) and in (gas, gypsum). It can be concluded that the production of observed minerals are helpful in improving the prediction of other observed mineral production except for coal, where coal is helpful in improving the prediction of oil.

Decomposition analysis of observed series

The parameter of decomposition, that is, window length “L” should be less than or equal to half of the series length, $L = 57$ is selected for the decomposition of each series. For MSSA, “L” is kept larger to improve separability of the observed system. The plot of the decomposition phase is analyzed for the whole system to select the optimum parameters for the reconstruction phase. The actual objective of decomposition analysis is to identify the pair of eigenvalues which characterize observed series into its components. The original observed data is used in MSSA as it has the ability to analyze any kind of data.

In Fig. 6, (1st, 2nd) and (2nd, 3rd) pair of eigenvectors are helpful in explaining trend structure, whereas the regular shape formed by (3rd, 4th) and (6th, 7th) paired eigenvectors represents the contribution in explaining cyclical and seasonal fluctuations. The same selection can be made from weighted correlation plot.

Reconstructed series

In Fig. 7, production of oil series is highest among observed series, so lines corresponding to the highest level of production represents the original and reconstructed series of oil. The level of coal production is lower than oil but higher than gas and gypsum. Reconstructed series captures trend and cycle variations of coal production. Gas and gypsum series do not show much fluctuations as these have much less production level as compared to oil and coal. The level of production for gas is slightly higher than gypsum, but after 2012, the level of gas and gypsum production is almost at the same level.

Diagnostic error of modeling techniques

Data is divided into two parts, train data and test data. For each series, train data contains 90 observations, i.e., from July 2005 to December 2012. Test data is held to predict by applied models to check the performance of every model. It contains 24 observations, i.e., from January 2013 to December 2014. Hyndman and Koehler (2006) suggested various measures for diagnostic model performance to select the most appropriate model for forecasting. These measures can be classified as scale-dependent and percentage errors. Percentage errors are the ratio of error to original values (i.e., MPE and MAPE).

- Root Mean Square Error ($\boldsymbol{RMSE}$) $= \sqrt{mean\left(Y_i - \widehat{Y_i}\right)^2}$

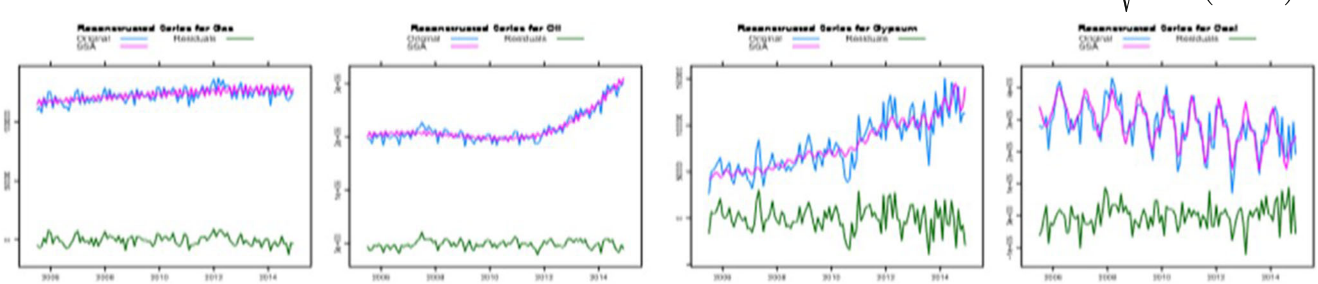

Fig. 5 A reconstructed series plot by pairing selected eigenvectors for each observed series

Table 3 Granger causality test for difference log transformed observed variables

Series	Gas	Oil	Gypsum	Coal
Gas	-	0.04	0.06	0.14
Oil	0.00	-	0.05	0.01
Gypsum	0.07	0.75	-	0.18
Coal	0.16	0.78	0.93	-

It measures the standard deviation of predictions to observed values.

- Mean Absolute Percentage Error $(\boldsymbol{MAPE}) = mean\left|\left(\frac{Y_i-\widehat{Y_i}}{Y_i}\right)\right|*100$

It measures the size of the error in percentage terms.

$$\boldsymbol{Theil's\ U} = \sum_{i=1}^{t-1}\left(\frac{Y_{i+1}-\widehat{Y_{i+1}}}{Y_i}\right)^2 \Big/ \sum_{i=1}^{t-1}\left(\frac{Y_{i+1}-Y_i}{Y_i}\right)^2$$

It measures the relative accuracy of the forecast from the naive model.

In Table 4, RMSE is highest among estimated errors for all the estimated models. For gas, the error measures of SSA and MSSA are comparatively much larger than the other estimated models**,** whereas ARAR and SARIMA have comparatively lower errors but still higher as compared to HW. So, error measures of HW model are minimum for gas among applied models. So, HW is considered being the most appropriate model to forecast natural gas production. For oil, HW, ARAR, and SARIMA have higher errors as compared to SSA and MSSA. Error measures of MSSA model are minimum among applied models for oil series. So, MSSA is considered being the most suitable model for forecasting crude oil production. MSSA has highest error measures except for Theil's U statistic for gypsum. The error measures of ARIMA and Holt are much close, whereas RMSE and MAPE are minimum for ARIMA. So, ARIMA is the most suitable model for forecasting gypsum production. The error measures of SSA are highest among estimated models for coal. The performance of SARIMA and ARAR is close but less than MSSA. The RMSE and MAPE of HW are minimum for coal**,** so HW model is appropriate to forecast the coal production.

Forecasting

In Fi. 8, forecasts are made for the next 6 years, i.e., from January 2015 to December 2020. The forecasts of gas are expected to increase in the future with seasonal fluctuations until 2020 at a constant rate and are expected to be approximately 139,279.7 MT units that are 11 % more than the production in 2014. The gray-shaded region represents confidence interval (CI), whereas dark-shaded region represents 80 % and light-shaded region indicates 95 % CI. The forecasts of oil are expected to increase exponentially until 2020 and are expected to be 23,477,650 MT. This is approximately eight times higher than the value of production in 2014. The forecasts of gypsum are also expected to be 45 % more than the production value of gypsum in 2014 approximately 163112.1 MT until 2020. The forecast productions of coal are expected to decrease with periodic fluctuations in the future. The production in 2020 is expected to be 170,969.76 MT units which are 12 % less than the value of production in 2014.

Discussion

Since each mineral has different characteristics, therefore the discussion about each mineral is provided as follows.

Gas Holt-Winters additive seasonal model is used to forecast the production of natural gas based on accuracy. The nature of seasonal variation in observed series is roughly constant throughout the series; that is the reason of applying seasonal additive model. Ribeiro et al. (2019) also utilized Holt-Winters model to forecast the production of Brazilian natural gas for the year 2018 based on the observations from year

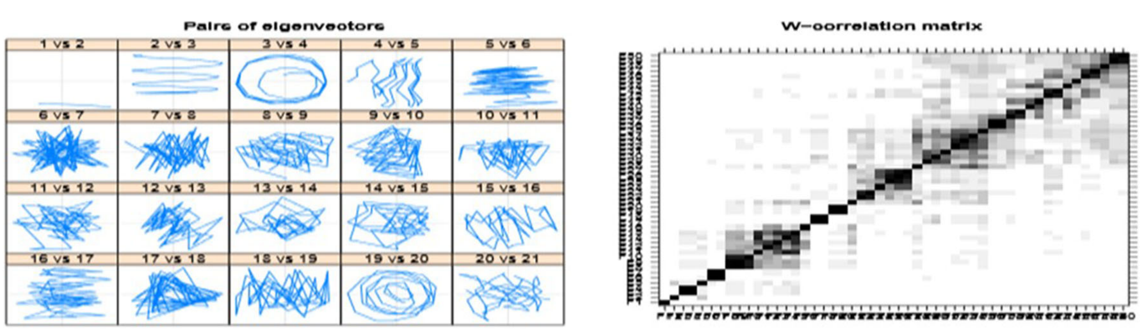

Fig. 6 Paired eigenvector and weighted correlation through MSSA decomposition

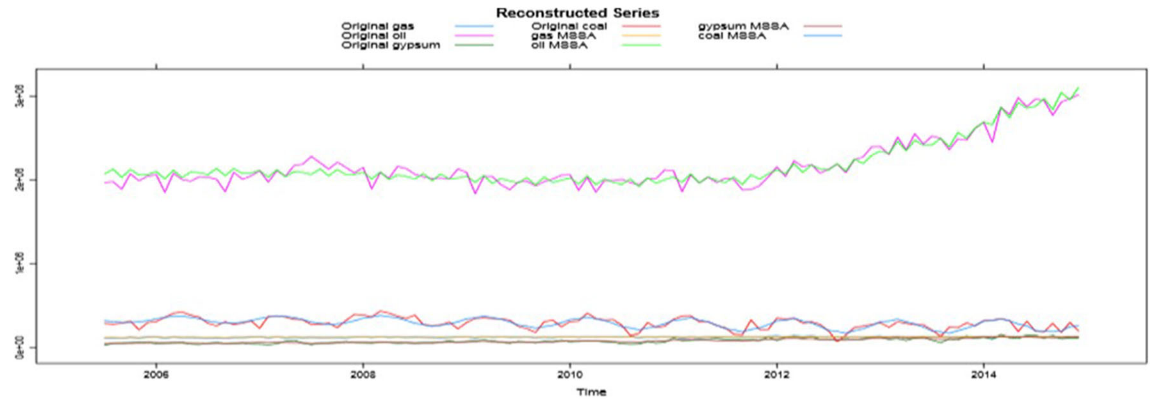

Fig. 7 MSSA reconstruction; x and y axes represent years and production (MT), respectively

2010 to 2017. They compared the performance of both additive and multiplicative seasonal method and declared that multiplicative method is more suitable to forecast Brazilian natural. It can be observed from Table 4 that the test error for natural gas for SSA and MSSA is significantly larger as compared to other models because these techniques ignore irregular component while modeling. ARAR could not perform better due to lack of strong cyclic pattern in the series. SARIMA contains auto-regressive component to capture relationship with previous data and moving average component to capture relationship with forecasting error. It requires difference transformation to make stationary series and is incapable of capturing turning points due to the loss of information while differencing. The Holt-Winters predicted value depends on almost all the data by capturing level, trend, and seasonality effectively.

Table 4 Comparison of test error measures for gas, oil, gypsum, and coal series

Series	Model	RMSE	MAPE	Theil's U
Gas	SARIMA	6803.2	4.58	1.05
	HW	3477.3	2.24	0.58
	ARAR	7565.8	5.01	1.08
	SSA	13323.5	7.48	0.60
	MSSA	55731.9	28.44	14.54
Oil	SARIMA	302056.3	10.05	3.10
	HW	312528.9	10.44	3.25
	ARAR	309646.1	9.97	4.65
	SSA	134050.1	4.09	1.81
	MSSA	119123	3.80	2.43
Gypsum	ARIMA	20575.6	14.50	23.45
	Holt	20687.2	14.65	28.81
	ARAR	22190.6	15.39	5.73
	SSA	24557.2	15.54	5.16
	MSSA	38323.9	21.08	9.45
Coal	SARIMA	58720.7	24.56	1.39
	HW	53070.1	19.59	1.75
	ARAR	56580.1	22.41	1.27
	SSA	151423.7	312.62	4.25
	MSSA	80132.6	20.12	13.09

Oil Multivariate singular spectrum analysis performs better for oil as compared to other applied models that clearly indicates that predicted values of crude oil production are not only function of its previous values but also production of other observed minerals. In univariate analysis, SSA performs much better than SARIMA, ARAR, and Holt-Winters indicating that production of crude oil contains weak structure of irregular component and is worth ignoring. As the production of crude oil is dependent on the production of all observed variables, the drastic increase rate of oil production is an indication of ignorance from another mineral sector. The production analysis indicates that production policies and strategies tend to maximize the production of crude oil and provide incentives to increase its exploration and mining. The implementation of policies imposed by Pakistan Government of more taxes on gas production to decrease its consumption in the energy sector. The biggest barrier is that exploratory prices of gas are linked with crude oil. The decrease in exploration prices of crude oil results to increase in exploration prices and taxes of other alternative sources thus resulting in non-increasing production.

Gypsum The test error measure of MSSA is highest for Gypsum as compared to other applied models indicating that the predicted values of gypsum are function of its lagged values. The performance of SSA is also worst as compared to other applied univariate models indicating that irregular component of the observed series is worth including in the model. The performance of ARAR is worst as compared to Holt's and ARIMA model depicting that the impact of cyclical variation in the observed series is also not strong. ARIMA performs better as compared to Holt's model indicating that predicted values of gypsum are linear function of lagged values of itself and forecast error.

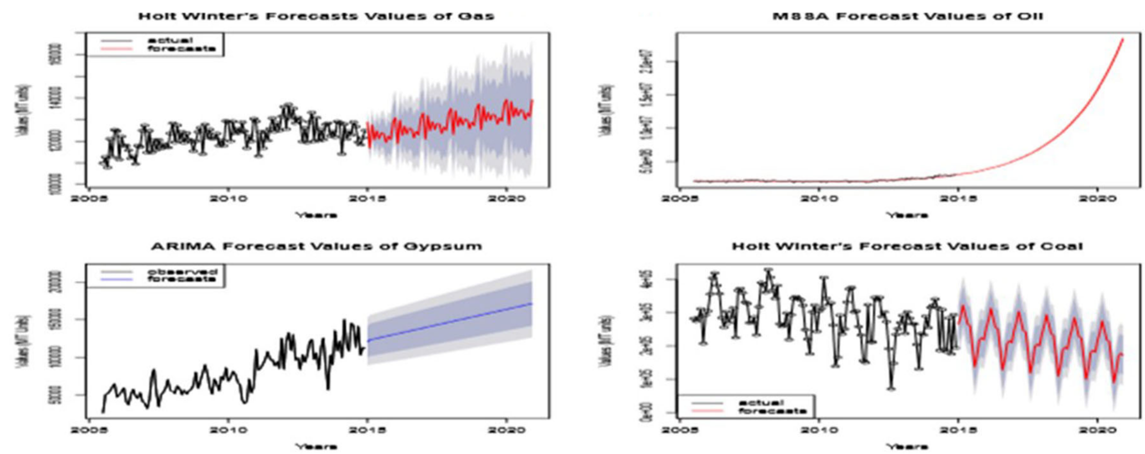

Fig. 8 Forecasting of observed series by suitable selected model

Coal The test error of SSA is highest for predicting coal as compared to other applied models indicating that SSA is unable to capture explained variations of the observed series, while MSSA is more flexible in capturing variations of coal production than SSA. The performance of SARIMA is worst as compared to ARAR, and Holt-Winters model indicates that the modeling procedure lacks amount of information contained in the data, and also the predicted values are not linear function of past values and forecast error. Holt-Winters seasonal additive model captures the variation of coal production well as compared to ARAR indicating that it is capable of estimating level, trend, and seasonality observed more effectively than other applied models.

The future projection of gas and gypsum depicts the increase in production at a constant rate, while production of oil depicts increase at a higher rate. The production of coal is expected to decrease at a constant rate indicating lower production in the future. This indicates serious negligence towards coal production as the reserves of coal are abundant, but the production level of coal is not optimum. The solution of demand-supply gaps of observed variables lies only in maximizing production of coal by using it as an alternative source. Coal can be utilized as the substitute for natural gas and crude oil to save the heavy amount from importing these sources. Also, it can be used to produce synthetic gypsum as an alternative to fulfill its demand.

Conclusion

The mineral sector of Pakistan is dominated by four principal minerals which are gas, oil, gypsum, and coal. Box-Jenkins methodology, a regression model with auto-regressive errors and Holt-Winters method, univariate singular spectrum analysis, and multivariate singular spectrum analysis modeling approach are applied to select the most appropriate model for forecasting of gas, oil, gypsum, and coal (Fig. 8). It is concluded that Holt-Winters additive seasonal model, multivariate singular spectrum analysis, Auto-regressive Integrated Moving Average, and regression model with auto-regressive errors performed better as compared with other models for gas, oil, gypsum, and coal, respectively.

The need of the hour is to focus on maximizing production of coal and by utilizing it optimally to solve the challenges of demand-supply gaps. The expected reserves of coal are much more than reserves of oil in Saudi Arabia measured up to 850 trillion cubic feet. There is a strong need to put efforts towards mineral sector so that demand can be fulfilled without depending on others. Coal is considered a savior of Pakistan's economy. Various projects are ongoing to resolve the demand gap such as the Thar Coal project. There is strong need to bring reforms in energy generation principles and making coal as a major source of generating energy as well as a proper substitute for gas and oil in the industrial sector to reduce high consumption of gas and oil. The concerned authorities, Government, and mining sector should make serious efforts to bring the production of coal at an optimum level. The coal mining sector should be facilitated with modern equipment, and high incentives should be given to promote coal mining and exploration. The current projects for coal mining are appreciable, but more strategies and planning are required to acquire the optimum utilization and allocation of coal through maximum production.

Acknowledgements The authors extend their appreciation to the Deanship of Scientific Research at the King Saud University for funding this work through research group no RG-1437-027.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Ahmed S, Mahmood A, Hasan A, Sidhu GAS, Butt MFU (2016) A comparative review of china, India and Pakistan renewable energy sectors and sharing opportunities. Renew Sust Energ Rev 57:216–225

Ahmed F, Osama Daudpota, M, and Kashif M (2017). Oil price shocks and industry level production using vector autoregression: empirical evidence from Pakistan.

Aized T, Shahid M, Bhatti AA, Saleem M, Anandarajah G (2018) Energy security and renewable energy policy analysis of Pakistan. Renew Sust Energ Rev 84:155–169

Akhtar J, Yaqub MI, Iqbal J, Sheikh N, Saba T (2018) Way forward in meeting energy challenges in Pakistan. Int J Ambient Energy 39(8): 904–908

Brockwell, P. J. and Davis, R. A. (2016). Introduction to time series and forecasting. springer.

Diebold FX, Kilian L (2000) Unit-root tests are useful for selecting forecasting models. J Bus Econ Stat 18(3):265–273

Dollar D, Kraay A (2002) Growth is good for the poor. J Econ Growth 7(3):195–225

Ejaz S, Aamir M, Khan MA, and Ashfaq B. (2018). Modeling and analysis of CPEC energy power projects using leap model. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pages 1–8. IEEE.

Groth A, Ghil M (2011) Multivariate singular spectrum analysis and the road to phase synchronization. Phys Rev E 84(3):036206

Hannah Ritchie MR (2018). Fossil fuels. Retrieved from: 'https://ourworldindata.org/fossilfuels' [Online Resource]. Accessed 13 July 2019

Hassani H, Heravi S, Zhigljavsky A (2013) Forecasting UK industrial production with multivariate singular spectrum analysis. J Forecast 32(5):395–408

Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast 20(1):5–10

Hussain T (2003). Planning for sustainable development. Commission on Science and Technology for Sustainable Development in the South, page 72.

Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. Int J Forecast 22(4):679–688

Kumar S, Shahbaz M (2012) Coal consumption and economic growth revisited: structural breaks, cointegration and causality tests for Pakistan. Energy Explor Exploit 30(3):499–521

McMahon G and Moreira S (2014). The contribution of the mining sector to socioeconomic and human development.

Mohr SH, Evans GM (2009) Forecasting coal production until 2100. Fuel 88(11):2059–2067

Mohr S, Evans G (2011) Long term forecasting of natural gas production. Energy Policy 39(9):5550–5560

Nawaz K, Lahiani A, Roubaud D (2019) Natural resources as blessings and finance growth nexus: a bootstrap ARDL approach in an emerging economy. Res Policy 60:277–287

Outlook AE et al (2010) Energy information administration. Dept Energy 92010(9):1–15

Parzen E (1982) ARARMA models for time series analysis and forecasting. J Forecast 1(1):67–82

Rehman A, Deyuan Z (2018) Pakistan's energy scenario: a forecast of commercial energy consumption and supply from different sources through 2030. Energy Sustain Soc 8(1):26

Rehman SAU, Cai Y, Fazal R, Das Walasai G, Mirjat NH (2017a) An integrated modeling approach for forecasting long-term energy demand in Pakistan. Energies 10(11):1868

Rehman SAU, Cai Y, Mirjat NH, Walasai GD, Shah IA, Ali S (2017b) The future of sustainable energy production in Pakistan: a system dynamics-based approach for estimating hubbert peaks. Energies 10(11):1858

Ribeiro RCM, Marques GT, dos Santos PC (2019). Holt-Winters Forecasting for Brazilian Natural Gas Production. Int J Innov Educ and Res 7(6):119–129

Rutledge D (2011) Estimating long-term world coal production with logit and probit transforms. Int J Coal Geol 85(1):23–33

Satti SL, Hassan MS, Mahmood H, Shahbaz M (2014) Coal consumption: an alternate energy resource to fuel economic growth in Pakistan. Econ Model 36:282–287

Schindler J and Zittel W (2008). Crude oil–the supply outlook. Energy watch group.

Shahbaz M, Lean HH, Farooq A (2013) Natural gas consumption and economic growth in Pakistan. Renew Sust Energ Rev 18:87–94

Winters PR (1960) Forecasting sales by exponentially weighted moving averages. Manag Sci 6(3):324–342

Zami MS, Lee A (2010) Economic benefits of contemporary earth construction in low-cost urban housing–state-of-the-art review. J Build Apprais 5(3):259–271