Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

İSTANBUL UNIVERSITY PRESS

RESEARCH ARTICLE

Forecasting Coal Production in India: A Time Series Approach

Avni GANGWAR[1], Diksha RATHOR[1], Praveen Kumar TRIPATHI[1]

ABSTRACT

This article is intended to produce the forecasts for coal production in India through some time series models. This study describes the component-based and correlation-based time series models for its purpose. The separate analyses were performed by applying Naïve, Holt's and ARIMA models on a real data set based on the coal production in India between 1980 and 2022. On the basis of the retrospective predictions and accuracy measure results, an ARIMA (2,2,2) model was selected as a good choice for the data in hand. A particular ARIMA (2,2,2) model was selected by using the AIC and BIC of model selection. For the validity of the finally selected ARIMA (2,2,2) model, a residual diagnostics check has been performed; and the future predictions have been made for the next 5 years. Such an analysis is expected to add some new approaches in the literature of forecasting the energy sources, especially with reference to India.

Keywords: Time Series, Coal Production, Model Selection Criteria, Residual Diagnostic, Prediction

DOI: 10.26650/JODA.1557949

[1]Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India

ORCID: A.G. 0009-0001-5376-1925; D.R. 0009-0009-5078-365X; P.K.T. 0000-0002-5652-8301

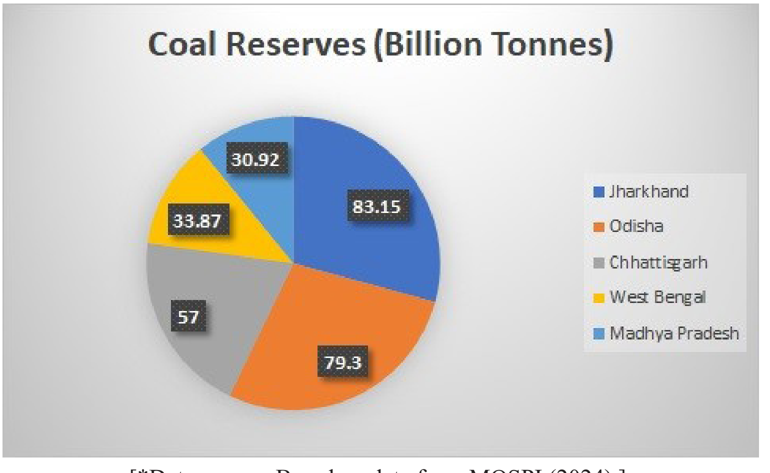
Corresponding author: Praveen Kumar TRIPATHI, Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India
E-mail: pkt2088@gmail.com

Submitted: 05.10.2024
Revision Requested: 27.11.2024
Last Revision Received: 30.11.2024
Accepted: 09.12.2024
Published Online: 20.12.2024

Citation: Gangwar, A., Rathor, D., & Tripathii P. K.. (2024). Forecasting Coal Production in India: A Time Series Approach. *Journal of Data Applications, 3*, 17-32. https://doi.org/10.26650/JODA.1557949

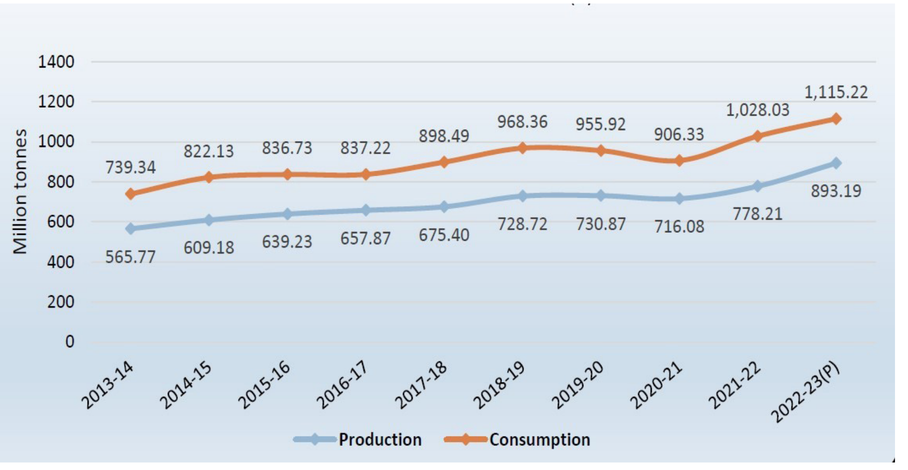
Introduction

Coal mining in India started in 1774, and at present, India is the second largest country, after China, in terms of coal production and consumption. Surpassing the historical figures, India produced 1 billion metric ton (MT) of coal and lignite in March 2024 and touches the 'historical high'. Despite its enormous production, India has to import coking coal to meet its domestic demands. Non-coking coal is imported by the coal-based power plants, captive power plants, sponge iron plants, cement industries and coal traders. Being the most significant and abundant fossil fuel in India, the coal fulfils almost 55% of the energy demand in the country. The major coal-producing states in India are; Jharkhand, Odisha, Chhattisgarh, West Bengal and Madhya Pradesh, with their combined contribution being around 90% in total (see, for example, MOSPI 2024) and the same is demonstrated in Figure 1.



[*Data source: Based on data from MOSPI (2024).]
Figure 1. *Contribution by the major states in India.*

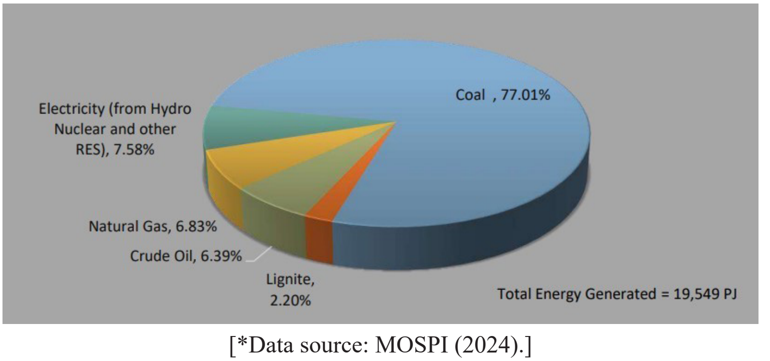
The study says that India's reliance on coal, especially in the power sector is more due to political inclusion than economic and technological reasons (see, for example, Montrone et al. 2021). Both coal production and consumption in India have been increasing over the years, and even more consumption is recorded from 2013-14 to 2022-23 (see, for example, MOSPI 2024). A comparison of the trend plots is shown in Figure 2, which is borrowed from MOSPI (2024). At present, India has sufficient stock of coal with a record production of 1 billion MT.



[*Data source: MOSPI (2024).]

Figure 2. *Trend of Production and Consumption of Coal in India from 2013-2014 to 2022-2023(P).*

Moreover, coal has the largest share in the overall energy generated in India from the different commercial sources. Figure 3 demonstrates that coal is the source of energy and it contributes maximum to the total energy generated in India during 2022-2023 (see, for example, MOSPI, 2024). Due to its vast consumption in different sectors like, steam energy, metallurgy, electricity, raw material for the industries, domestic purpose, etc., the coal is high in demand and, therefore, its production becomes a crucial aspect for various sectors in India. It helps the economic growth in an indirect way through energy generation, industrial developments, job creation, etc. Despite its needs in different sectors, the environmental concerns and health issues are some major points where the government has to think about some cleaner energy sources.



[*Data source: MOSPI (2024).]

Figure 3. *Share of total energy generated (in petajoule) from different commercial sources in India during the financial year 2022-2023(P).*

The above study from the different government organisations motivates us to analyse the coal production in India and to make a reasonable forecast based on historical time series data. With this aim, we have considered some basic time series models with their specific features of statistical model buildings. Loosely speaking, we consider the two component-based time series models, namely, Naive and Holt's linear trend models; and a correlation-based model, namely, the autoregressive integrated moving average (ARIMA) model. The component-based model incorporates the changes in the dataset due to the different components of a time series. For instance, Holt's model captures the inherent trend in the dataset and provides future estimates accordingly. On the other hand, correlation-based models incorporate the fluctuations due to the serial correlations in a data set (see, for example, Hyndman and Athanasopoulos 2018). By considering the ARIMA model, one may ensure that the future estimates are available due to the consideration of the autocorrelation feature of the real data in the past. It is, then, assumed that the correlation characteristics shall remain unchanged in the future. Now, a comparison between the two methods will tell us which feature is more relatable when we go for a future prediction of coal production in India.

The statistical literature on coal production in India is hardly available except, perhaps in the form of data accumulation, survey and report formats on the government websites. Our aim is to analyse the coal production data in India and to choose the most effective model on the basis of forecast accuracy measures. Our study also reveals the fact that which type of model is appropriate for the considered real data between the two described categories of time series models. Some of the recent works on coal data using time series approaches include Li et al. (2019), Makkhan et al. (2020), Chen et al. (2021), Parren˜o (2022), Jai Sankar et al. (2023), Mohanty and Nimaje (2023), among others. In particular, Li et al. (2019) forecasted the coal production in India by 2030 through a combined time series model after some modifications in their linear and non-linear setups. Makkhan et al. (2020) have analysed the black carbon in the regions of coal mines in India through the classical tools of statistics, such as; correlation coefficient, rank and Kandall's tau correlations and ARIMA models. Chen et al. (2021) used the descriptive and graphical tools for a statistical analysis of 'Chongqing coal mine' accidents somewhere in China. Parren˜o (2022) used the ARIMA model to forecast the coal production and consumption in the Philippines. Mohanty and Nimaje (2023) have applied the neural network and ARIMA models to forecast the fatal coal mine accidents in India, where they found the neural network model to be the best fitted model for the data under consideration. In another study, Jai Sankar et al. (2023) have modelled and analysed the coking coal production in India through the ARIMA model and concluded that coking coal production will be declining in India by, 2031.

Since the time series literature on coal production in India is quite limited, no recent study has unveiled the gross coal production in India using a time series model. This paper attempts to fill this gap and provide a systematic future prediction of coal production in India by choosing the most suitable time series model.

The rest of the paper is organised as follows. Section 2 defines the basic structures of the considered time series models. Section 3 describes the dataset and the necessary steps to proceed with the whole analysis. This particular section completes the whole analyses, for the real data set of coal production in India, under different subsections. The last section concludes the paper, which provides a smooth ending to the whole work.

Some basic time series models

Naive model

It assumes that the future value of a time series will be matched by the last observation of the series, and is defined as follows:

$$\hat{y}_{t+h} = y_t. \qquad \textit{Eq. 1}$$

In Eq. 1, $\hat{y}_{t+h}$ is the h -step ahead forecasted value of the most recent observation y_t at current time t. The forecasts obtained by the naive model are also termed as the 'random walk forecasts' (see, for example, Hyndman and Athanasopoulos (2018)).

Holt's linear trend model

Holt (1957) proposed a linear trend model that allows the forecasting of a time series with trend. Holt's model smoothens the time series data in the two phases; one for the level and the other for the trend. Eq. 2 defines the mathematical form of Holt's model as,

$$\boldsymbol{y_{t+h} = L_t + hT_t}$$

$$\boldsymbol{L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})}$$

$$\boldsymbol{T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}}, \qquad \textit{Eq. 2}$$

where L_t and T_t denote the estimates of the level and trend, at time t, respectively, of the time series data; and $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are the smoothing parameters for the level and trend, respectively. The trend of the time series plot is identified by its slope at a time (see, for example, Hyndman and Athanasopoulos 2018).

ARIMA model

This model was proposed by Box and Jenkins (1970) and termed as 'Box-Jenkins Model' in the time series literature. The ARIMA model is defined for the differenced series of data, $x_t = \Delta^d y_t$, and can be written as;

$$x_t = c + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t, \quad Eq.\ 3$$

where c, is the intercept term, and (p,d,q) denotes the combined order of the ARIMA model consisting of the order of the autoregressive (AR) components, degree of differencing and order of the moving average (MA) components in sequence. Also, in Eq. 3, e_t's are independently and identically distributed (i.i.d.) normal variates each with a common mean and variance of 0 and σ^2 respectively (see, for example, Agarwal et al. 2021).

Data description and related analysis

For our analysis, we have considered real data on the annual coal production in India approximately 1980-81 to 2021-22 (see the table in Appendix). The dataset is borrowed from 'Economic Survey 2022-23' released by the Government of India (see India 2024). A time series plot of the data is shown in Figure 4. The plot exhibits an increasing trend of coal production in India approximately 1980-81 to 2021-22, which is obviously a non-stationary pattern of the time series.

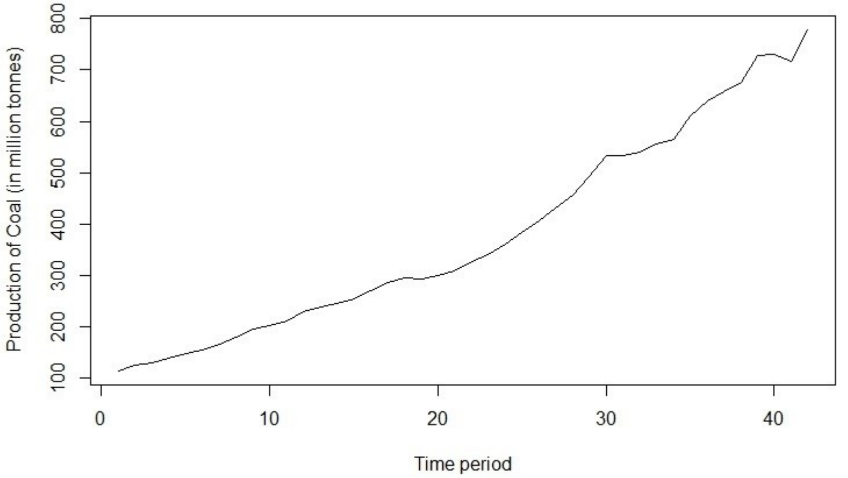


Figure 4. *Time series plot for coal production in India approximately 1980-1981 to 2021-2022.*

We, first, have proceeded with the component-based time series model prediction and classified our data into two parts; the first 36 observations as 'training-sample data' and the remaining 6 as 'test-sample data'. We have analysed the training-sample data by applying the Naive and Holt's models and then predicted for the next 6 observations corresponding to the 'test-sample' data using the R software, and the same is exhibited in Figure 5. As observed, the prediction based on naïve method is constant, and it can be considered as the base level for a further future prediction. On the contrary, Holt's linear trend method exhibits a sharply increasing pattern of the coal production in India. The retrospective predictions based on the two methods can be concluded in two different ways. First, Holt's method provides an over-fitted prediction than the naïve method. Second, none of the two methods is capable of to capture the actual fluctuation of the considered time series. An overall conclusion may be that the component-based time series models, considered here, are not good enough to obtain the future prediction of the coal production in India. We shall, therefore, be looking into some other choices of time series models like ARIMA models in the coming section.

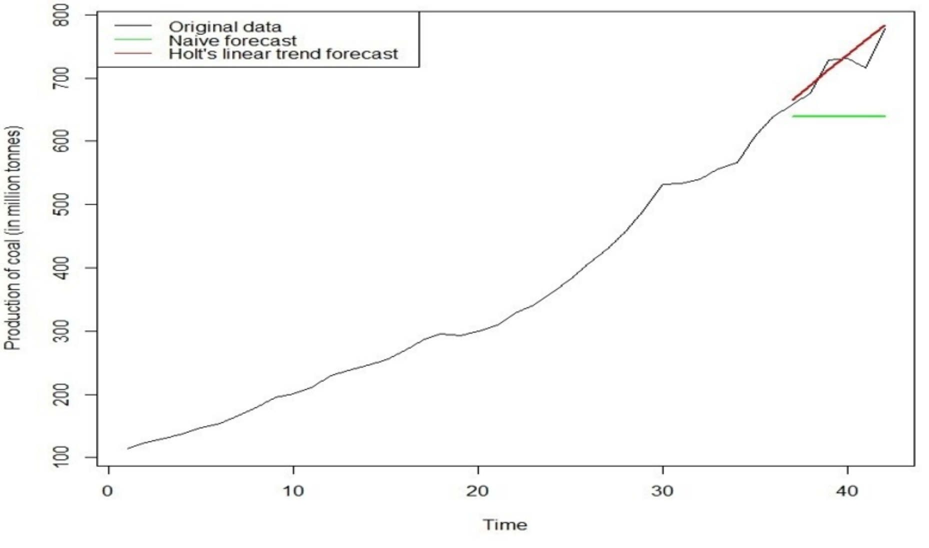


Figure 5. *A retrospective prediction based on naive and Holt's linear trend methods.*

Analysis with the ARIMA model

Analysis using the ARIMA model consisting three stages; model identification, estimation and validation (see, for example, Box et al., 2015). Model identification involves deciding the order of the ARIMA model, that is, realisation of (p, d, q). Model estimation includes the estimation of the AR and MA coefficients in the ARIMA model equation (3); and model validation consists of the authentication of the finally selected ARIMA model for the data in hand. Before we proceed with the three steps, we shall verify whether the time series is stationary or not. As mentioned before, Figure 4 depicts a non-stationary movement over time and, therefore, an appropriate transformation is needed to make it stationary (see, for example, Agarwal et al., 2021). In our case, we made the double differencing of the log-transformed data to obtain a stationary pattern (see Figure 6). We have also performed the 'stationary check' at every stage of transformation for the assurance by using the two most commonly used tests, namely, the augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The ADF test rejects a null hypothesis of the unite root in the data at the 5% level of significance; while the KPSS test does not wish to reject the null hypothesis of stationarity again at the 5% level of significance. For more details on the two tests, one may refer to Tripathi et al. (2017, 2021), among others. The outputs of the two tests are shown in Table 1, which indicates that the log-transformed double-differenced data achieved stationarity.

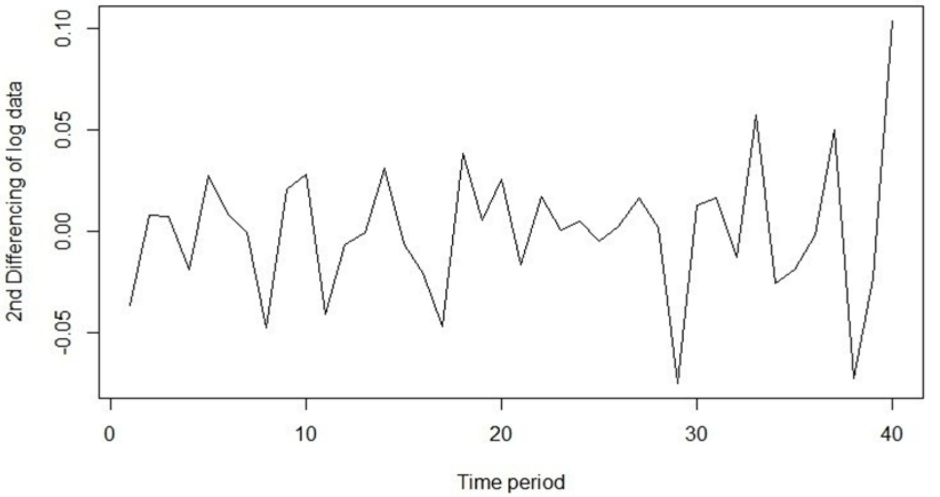


Figure 6. *Time series plot of double-differenced and log-transformed coal production data.*

Table 1. *Outputs of the ADF and KPSS tests for the different forms of time series data.*

Time series	p-value	
	ADF test	KPSS test
$z = \log(x)$	0.53	0.01
Δz	0.32	0.02
$\Delta^2 z$	0.01	0.10

To assess the order of the ARIMA model, we plotted the autocorrelation function (ACF) and partial ACF (PACF) values against different lag values in Figure 7, and observed the exact cut-off beyond the significant limits (the dotted lines). One may observe the exact cut-off at lag 2 in each of the two plots of Figure 7. Generally, the PACF cut-off decides the order of the AR process, that is, p and the ACF cut-off decides the order q in the MA process. Since our, further analysis will be based upon stationary data only, which is obtained after double differencing, the order of integration in the ARIMA model will be d = 2. Hence, our selected ARIMA model for the data is ARIMA (2, 2, 2).

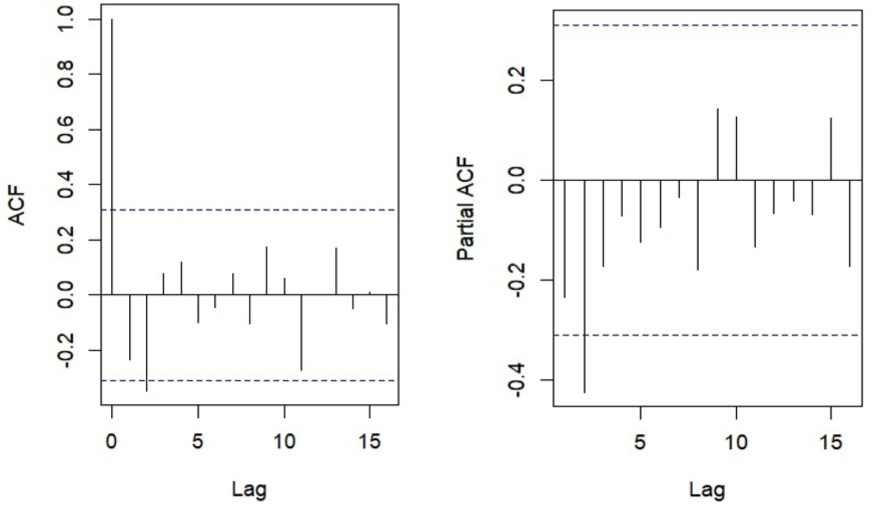


Figure 7. *ACF and PACF plots for the stationary data.*

To be more selective, we may consider some other nearby choices of ARIMA models such as ARIMA (0,2,1), ARIMA (0,2,2), ARIMA (1,2,0), ARIMA (1,2,1), ARIMA (1,2,2), ARIMA (2,2,0) and ARIMA (2,2,1). Such an approach may avoid any misleading conclusion based on a single choice of model assessment (see, for example, Tripathi et al. 2018). A final selection of the ARIMA model is done on the basis of two criteria, namely, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which are defined as;

$$\text{AIC} = -2log\hat{L} + 2k,$$
$$\text{BIC} = -2\log\hat{L} + k * log\,(T - p), \qquad Eq.\ 4$$

where k and $\hat{L}$ represent the number of parameters and the estimated likelihood of the ARIMA model respectively. Table 2 reports the AIC and BIC values of the considered ARIMA models for the coal production stationary data. As a thumb rule, the model having the least values of AIC (BIC) will be selected. Table 2 shows that the ARIMA (2,2,2) model was selected as the best fitted model for the data in hand.

Table 2. *AIC and BIC values of the considered ARIMA models.*

ARIMA models	AIC value	BIC value
ARIMA (0,2,1)	272.79	282.01
ARIMA (0,2,2)	274.79	281.74
ARIMA (1,2,0)	272.79	282.00
ARIMA (1,2,1)	275.00	281.96
ARIMA (1,2,2)	275.78	284.47
ARIMA (2,2,0)	274.79	281.75
ARIMA (2,2,1)	276.32	285.01
ARIMA (2,2,2)	**270.29**	**281.29**

To estimate the model parameters in ARIMA (2,2,2), we have maximised the likelihood function, given in Eq. 5, by using the non-linear minimisation (*nlm*) function in the R software. Eq. 5 represents the conditional likelihood function, up to proportionality, of the ARIMA (2,2,2) model for the stationary data set $x: x_1, x_2, \dots, x_T$.

$$f(\underline{x}|\Theta) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T-2}{2}} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{t=3}^{T}\left(x_t - c - \sum_{i=1}^{2}\phi_i x_{t-i} - \sum_{j=1}^{2}\theta_j e_{t-j}\right)^2\right), \quad Eq.\ 5$$

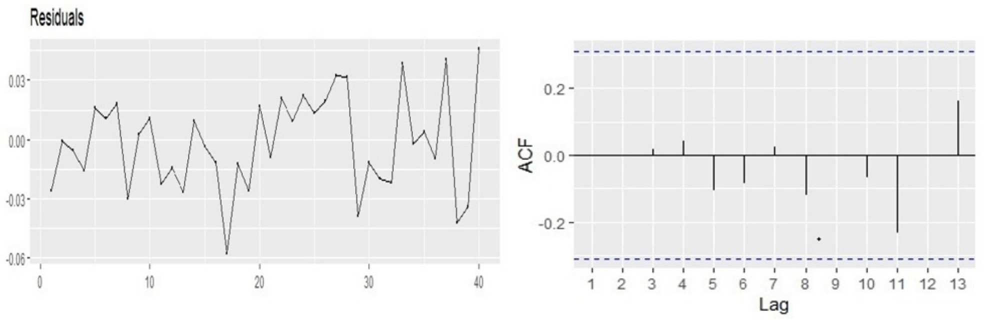
where $\Theta = (\sigma^2, c, \phi_1, \phi_2, \theta_1, \theta_2)$, a set of model parameters. The maximum likelihood estimates (MLE) of the parameters of the selected ARIMA (2,2,2) model are reported in Table 3.

Table 3. *MLE of the ARIMA (2,2,2) model*

Parameters	σ^2	c	ϕ_1	ϕ_2	θ_1	θ_2
MLE	0.80	-0.07	-1.02	-0.70	1.20	0.95

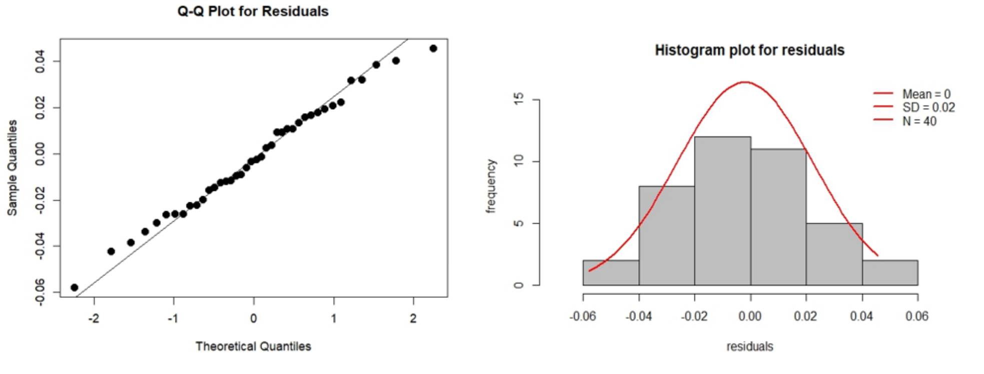
Next, to determine the appropriateness of the ARIMA (2,2,2) model, we have gone through the residual diagnostic checks. A residual is simply the difference between the actual and fitted observations of the data. Ideally, the residuals should be uncorrelated and possess the mean equal to “zero” to get utilised all the information from the data and to avoid any kind of bias in the forecasting respectively (see, for example, Hyndman and Athanasopoulos, 2018). The calculated residuals and their ACF are plotted in Figure 8-a and Figure 8-b, respectively. In the residual plot, no abrupt fluctuation is observed and the ACF plot further indicates that there is no serial correlation between the residuals over the different lags.

Another qualification in the residual diagnostic is that the residuals should follow a normal distribution with a mean of zero and a constant variance. Nothing can be more preferable than a pictorial demonstration of the calculated residuals to demonstrate this feature (see, for example, Tripathi et al., 2022). We, therefore, obtained the Q-Q plot (see Figure 9-a) between the sample and the theoretical quartiles of the residuals, and a histogram-polygon normal curve plot of the calculated residuals (see Figure 9-b). It is observed that the sample quantiles follow the straight line of the theoretical quantiles and further, the plotted histogram of the residuals are well within the normal polygon curve with the mean equal to zero and standard deviation 0.02 (see Figure 9-b). The above diagnosis of the calculated residuals strengthens our assumption that the residuals of the data, based on the ARIMA (2,2,2) model, are not correlated and follow a normal distribution with specific white noise property.



(a) (b)

Figure 8. *Residual and ACF plots of the coal production data based on the ARIMA (2,2,2) model.*



(a) (b)

Figure 9. *Q-Q plot and histogram-polygon plot of residuals based on the ARIMA (2,2,2) model.*

To conclude our work in the favour of most accurate model among the naive, Holt's model and ARIMA (2,2,2) model, we have performed the two 'forecast accuracy measure' tools, namely; root mean square error (RMSE) (see Eq. 6) and mean absolute percentage error (MAPE) (see Eq. 7) which are defined as below:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \qquad \textit{Eq. 6}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100, \qquad \textit{Eq. 7}$$

where $\widehat{y_i}$ is the i^{th} observation to be predicted from the 'test sample data' (see, for example, Tripathi and Agarwal 2023). The corresponding predicted values for the original observations can easily be obtained and they are summarised in Table 5 for each of the considered models. Table 4 summarises the results of the forecast accuracy measures for each of the considered models. This clearly indicates that the ARIMA (2,2,2) model outperforms the other two models for the given data with a better accuracy. Therefore, we recommend ARIMA models to predict the coal production in India as these models admit correlation within the data, which causes a major difference in forecasting when compared with the component-based time series models.

Table 4. *Results of the forecast accuracy measures*

Accuracy measures	Naive model	Holt's model	ARIMA (2,2,2) model
RMSE	84.94	20.18	0.02
MAPE	10.27	2.11	0.3

Based on the recommendation of 'forecast accuracy measure', we have made the future prediction of coal production in India by using the ARIMA (2,2,2) model. With the available data of 42 observations, we have predicted for next 5 years coal production in India, which is shown (by red dotted line) in Figure 10. Certainly, the Graphical pattern of the future prediction is enough to conclude that the coal production is going to be increased in the next five years in India.

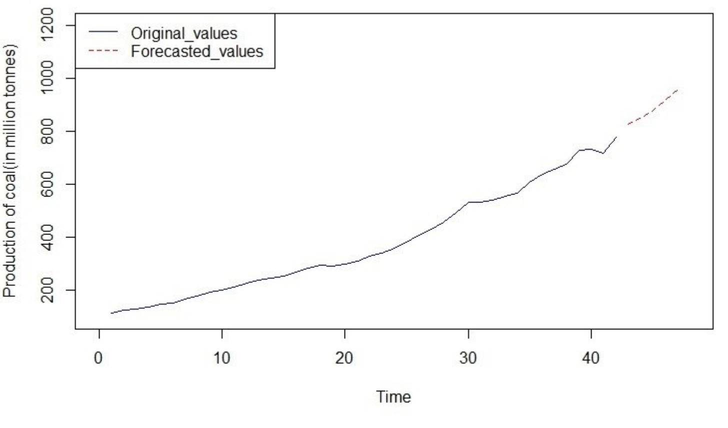


Figure 10. *Forecasts for next 5 years of coal production in India by the ARIMA (2,2,2) model.*

Conclusion and Findings

This paper has successfully analysed the coal production data of India through the time series models. The analyses tackle the component-based and correlation-based time series models for the given data. First, the forecasts were made using naive and Holt linear trend models and, then, the data is being used to analyse the ARIMA model. A suitable ARIMA model is chosen by adopting the two-fold strategy consisting of the Box-Jenkins technique and then on applying the two model selection criteria, namely, AIC and BIC. The two-fold identification of the ARIMA model suggests that ARIMA (2,2,2) is the best candidate model among others. After the successful identification steps, we performed the retrospective forecasts for the competing models (see Table 5). The forecasted values by the ARIMA (2,2,2) model are closer than those by the other component-based models, which can also be verified by their accuracy measures in Table 4. It is, therefore, decided to perform the further analysis with the selected ARIMA (2,2,2) model. After estimating the parameters of the ARIMA (2,2,2) model, the residual diagnostic checks were performed for the validity purpose of the chosen model. Finally, a 'five years' future forecast is being made by using the ARIMA (2,2,2) model graphically. It is observed that the coal in India will be in high demand and its production will increase in the years to come.

Table 5. *Retrospective six-year forecast of coal production in India based on the competing time series models.*

Year	**True Value**	**Forecast value**		
		Naive model	**Holt's model**	**ARIMA (2,2,2) model**
2016-17	657.8	639.2	664.3	657.6
2017-18	675.8	639.2	688.9	674.8
2018-19	728.7	639.2	713.1	729.1
2019-20	730.8	639.2	736.7	731.0
2020-21	716.9	639.2	759.9	715.8
2021-22	778.2	639.2	782.6	778.5

Our analysis is restricted by the use of time series models and it does not include the other factors that may affect the coal production in India, such as; infrastructure and availability of modern technology, accidental hazards during the production process, natural calamities, etc. One may consider these factors and see their effects on the coal production over the years. Also, a significant analysis can be done to observe the most significant factor and, ultimately, can tune the productivity of energy sources at a desired level whenever necessary. On the theoretical ground, the regression analysis could be another alternative to estimate the coal production with its relevant explanatory variables. Our analysis is limited, but it is not restricted for any future development of its current version. Such an analysis is going to be beneficial for the industry resource planners and government/semi-government entities to a good inventory setup and to design a reasonable path of coal consumption at the ground level.

Availability of data and materials: The data is provided in the manuscript and is open to access.

Ethics Committee Approval: As this study is based entirely on a review of existing literature, rather than primary data collection, ethics committee approval was not required. The analysis and synthesis of previously published research do not involve new interactions with human subjects or require ethical clearance.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- A.G., D.R.; Data Acquisition- A.G.; Data Analysis/ Interpretation- A.G., D.R.; Drafting Manuscript- A.G., D.R.; Critical Revision of Manuscript- P.K.T.; Final Approval and Accountability- P.K.T.; Technical or Material Support- A.G., D.R.; Supervision- P.K.T.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

References

Agarwal, M., Tripathi, P. K., & Pareek, S. (2021). Forecasting infant mortality rate of India using ARIMA model: a comparison of Bayesian and classical approaches. *Statistics & Applications*, 19(2), 101–114.

Box, G. & Jenkins, G. (1970). *Time series analysis: forecasting and control*. Holden-Day, San Francisco.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Chen, Y., Xu, J., Tan, K., & Gou, Y. (2021). Statistical analysis and regular research on Chongqing coal mine accidents. *In 2021 International Conference on Intelligent Computing, Automation and Applications (ICAA)*, pages 64–70. IEEE.

Holt, C. C. (1957). Forecasting trends and seasonals by exponentially weighted averages. *ONR Memorandum*, 52, 1957.

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts.

Jai Sankar, T., Angel Agnes Mary, I., & Nalini, K. (2023). Stochastic time series modeling for coking coal production in India. *International Journal of Scientific Development and Research*, 7(10), 443–449.

Li, S., Yang, X., & Li, R. (2019). Forecasting coal consumption in India by 2030: using linear modified linear (MGM-ARIMA) and linear modified nonlinear (BP-ARIMA) combined models. *Sustainability*, 11(3), 695.

Makkhan, S. J. S., Parmar, K. S., Kaushal, S., & Soni, K. (2020). Correlation and time-series analysis of black carbon in the coal mine regions of India: a case study. *Modeling Earth Systems and Environment*, 6, 659–669.

Mohanty, A. & Nimaje, D. S. (2023). Time series forecasting of Indian coal mines fatal accidents. *Journal of Mining Science*, 59(6), 1076–1082.

Montrone, L., Ohlendorf, N., & Chandra, R. (2021). The political economy of coal in India evidence from expert interviews. *Energy for Sustainable Development*, 61, 230–240.

Parren˜o, S. J. E. (2022). Application of time series analysis in forecasting coal production and consumption in the Philippines. *ICTACT Journal on Soft Computing*, 13(1).

Tripathi, P. K. & Agarwal, M. (2023). A bayes analysis of random walk model under different error assumptions. *Annals of Data Science*, pages 1–18.

Tripathi, P. K., Mishra, R. K., & Upadhyay, S. (2018). Bayes and classical prediction of total fertility rate of India using autoregressive integrated moving average model. *Journal of Statistics Applications & Probability*, 7(2), 233–244.

Tripathi, P. K., Ranjan, R., Pant, R., & Upadhyay, S. (2017). An approximate Bayes analysis of ARMA model for Indian GDP growth rate data. *Journal of Statistics and Management Systems*, 20(3), 399–419.

Tripathi, P. K., Sen, R., & Upadhyay, S. (2021). A Bayes algorithm for model compatibility and comparison of ARMA (p, q) models. *Statistics in Transition new series*, 22(2), 95–123.

Tripathi, P. K., Tripathi, A., Harshika, & Meenakshi (2022). Statistical insights on the health of Indian economy: a multivariate analysis. *Journal of Data, Information and Management*, 4(3), 305–327.

MOSPI (2024). Energy Statistics India 2024. https://mospi.gov.in/sites/default/files/publication_reports/Energy_Statistics_2024/Energy_Statistics_India2024.pdf

India (2024). Statistical Appendix. *Economic Survey 2023-24.* https://www.indiabudget.gov.in/economicsurvey/doc/Statistical-Appendix-in-English.pdf

Appendix

Coal production (in million tonnes) in India approximately 1980-81 to 2021-22 (see India 2024).

Year	Production	Year	Production	Year	Production
1980-81	113.9	1994-95	253.8	2008-09	457.1
1981-82	124.2	1995-96	270.1	2009-10	492.8
1982-83	130.5	1996-97	285.7	2010-11	535.7
1983-84	138.2	1997-98	295.7	2011-12	540.0
1984-85	147.4	1998-99	292.3	2012-13	556.4
1985-86	154.2	1999-00	300.0	2013-14	565.8
1986-87	165.8	2000-01	309.6	2014-15	609.2
1987-88	179.7	2001-02	309.6	2015-16	639.2
1988-89	194.6	2002-03	327.8	2016-17	657.8
1989-90	200.9	2003-04	341.3	2017-18	675.8
1990-91	211.7	2004-05	361.3	2018-19	728.7
1991-92	229.3	2005-06	382.6	2019-20	730.8
1992-93	238.3	2006-07	407.0	2020-21	716.08
1993-94	246.0	2007-08	430.8	2021-22	778.19