Time series analysis and forecasting of China's energy production during Covid-19: statistical models vs machine learning models

Zekai Lu

Guangzhou University <https://orcid.org/0000-0001-8015-3071>

Nian Liu

Guangzhou University

Ying Xie

Guangzhou University

Junhui Xu (✉ williamlord@163.com)

Guangzhou University

Research Article

Keywords: Machine Learning, Time series, COVID-19, Energy production forecast, Multistep forecasting

Posted Date: November 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1074872/v3>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Time-series analysis and forecasting of China's energy production during COVID-19: statistical models vs. machine learning models

Abstract

COVID-19 is a huge catastrophe of global proportions, and this catastrophe has had far-reaching effects on energy production worldwide. In this paper, we build traditional statistical models and machine learning models to forecast energy production series in the post-pandemic period based on Chinese energy production data and COVID-19 Chinese epidemic data from 2018 to 2021. The experimental results showed that the optimal models in this study outperformed the baseline models on each series, with MAPE values less than 10. Further studies found that the LightGBM, NNAT and LSTM machine learning models worked better in unstable energy series, while the ARIMA statistical model still had an advantage in stable energy time series. Overall, the machine learning models outperformed the traditional models during COVID-19 in terms of prediction. Our findings provide an important reference for energy research in public health emergencies, as well as a theoretical basis for factories to adjust their production plans and governments to adjust their energy decisions during COVID-19.

Keywords: Machine Learning; Time series; COVID-19; Energy production forecast; Multistep forecasting

1. Introduction

The COVID-19 pandemic that broke out in early 2020 is a huge public health crisis for all humanity. According to publicly available data from Johns Hopkins University(E. Dong et al. 2020), as of September 2021, 230 million people have been infected worldwide, resulting in more than 4.8 million deaths. Due to the highly contagious nature of COVID-19, the vast majority of countries around the world have had to adjust their economic and social policies to meet the challenges posed by the pandemic, with half of the world's population already experiencing a full or partial lockdown according to statistics("Context" n.d.).

Not only has the COVID-19 pandemic caused deaths, factory shutdowns and school closures(Hassan, Muneeza, et al. 2021), it has also had a profound impact on energy production (including the production of raw materials and the production of electricity). The virus was a black swan event for the energy production sector, which was hit by a series of lockdowns, shutdowns and reduced demand. In the labour-intensive industries of raw coal processing and thermal power generation, for example, companies in these sectors struggled with work stoppages, reduced orders and transport disruptions. Energy shortages pose a huge challenge to industrial recovery and social stability, especially in today's highly interconnected society, where the crisis caused by energy shortages can

easily be transmitted to all systems of society, leading to economic panic and social panic.

The impact on China, the world's largest energy producer, was clearly more pronounced, with its energy production suffering an unprecedented crisis during the pandemic. Beginning with Wuhan, the first city to be closed to the public, China imposed an almost draconian lockdown ("Wuhan lockdown 'unprecedented', shows commitment to contain virus" 2020). The unanticipated lockdown measures brought Chinese energy production to a standstill ("Electricity first, overcoming the difficulties together, promoting Covid-19 prevention and control and serving economic and social development in an integrated manner" n.d.). In particular, the raw coal production and processing industries, which make up the largest part of China's energy system, were forced to shut down their coal mining sites due to the pandemic, and railway capacity was greatly reduced due to strict traffic controls, which further transmitted to the downstream supply of thermal power plants ("Effect of Covid-19 on the power industry" n.d.). At the beginning of the pandemic, when northern China was in the midst of winter heating, the fall in energy supply caused energy prices in China to soar, with the effects continuing even into the autumn of 2021 ("Unprecedented' power cuts in China hits homes, factories" n.d.).

Interpreting and predicting trends in Chinese energy production is an important guide to the recovery of Chinese industry. There are also lessons to be learned from energy production around the world. Effective forecasting of changes in energy production in the midst of the pandemic is extremely important and will help guide factories to adjust production in a timely manner and governments to adjust economic policies at a macro level.

Traditionally, we have often used parametric statistical models such as ETS and ARIMA for energy production forecasting. However, the impact of COVID-19 has made forecasting by statistical models difficult. COVID-19 undermines the requirement for time-series stability in traditional parametric models (Ledolter 1989), and traditional parametric models have difficulty incorporating the effects of COVID-19 into forecasting models. The machine learning model used in this paper can better solve this dilemma (M. Z. Abedin et al. 2021).

Machine learning time-series forecasting models have developed rapidly in recent years, from shallow machine learning models such as decision trees and SVMs, which emerged after 2000 (Kotsiantis et al. 2007), to deep learning models such as CNNs and LSTMs, which have emerged since 2015 (Lara-Benítez et al. 2021). Machine learning has achieved good results in time-series prediction of economic and social problems. Since time series are a kind of sequence and are similar to image data and text data, which machine learning is good at processing, it is logical to migrate machine learning models to time series (G. Dong and Pei 2007). There has been tremendous progress in migrating machine learning methods to time series, including the SVM and LSTM we mentioned earlier, which have achieved excellent results in various studies. In this study, the machine learning model was able to effectively overcome the requirement for temporal stability and incorporate the effects of the epidemic into the model using covariates.

Reviewing past studies, we find that there are still relatively few studies on energy production forecasting during the COVID-19 pandemic. The only few articles that have been published have used traditional ETS and ARIMA models. For example, D'Alessandro and Chapman used the ETS model to forecast and discuss the supply and demand of electricity generation in Japan during the pandemic (D'Alessandro et al. 2021), and Werth et al. discussed in detail the impact of the epidemic on the production of various energy sources in Europe without further forecasting the time series (Werth et al. 2021).

In addition, there are very few predictions for Chinese energy in the same period, with most of the studies on Chinese energy forecasting predating the pandemic. For example, Rout et al. and Zou et al. studied long-term energy production changes in China and long-term natural gas supply changes in China (Rout et al. 2011; Zou et al. 2018). Moreover, most papers tend to use two completely different sets of modelling frameworks, namely, the traditional statistical framework and the machine learning framework, and rarely combine the two for analysis. Except for the studies by Ahmed et al. and Wang et al., which have compared the performance of models under different frameworks, they study pre-COVID-19 energy data and do not take into account the impact of the pandemic (R. A. Ahmed and Shabri 2014; J. Wang and Hu 2015). Finally, there is a problem in the current study; most of the papers on energy forecasting use a single time series, e.g., only thermal power generation data or wind power generation data. There is no consideration of the effects of different time series that provide a comprehensive analysis of the whole energy system. Naturally, these studies do not use covariates to assist in forecasting. In general, the articles on covariate-assisted forecasting are mostly theoretical and financial articles, with few articles on energy forecasting (M. Z. Abedin et al. 2019; Zhang et al. 2020).

Overall, this paper achieves innovation in the following areas. First, this paper is the pioneer in using both traditional statistical models and machine learning models for energy forecasting, filling the gap of lack of comparison and recommendation of the two forecasting frameworks in past energy forecasting. Second, we have developed high-performance forecasting models for different energy series, which provide a reference for subsequent research and theoretical support for industry and government decision-making. In addition, the data used in the study are innovative in that we use macro energy data for China during the epidemic and incorporate the impact of the epidemic into the model using machine learning models. Furthermore, the paper analyses and forecasts 10 different energy series using a variety of different types of energy production data (including raw materials and electricity generation) to assess the impact of the epidemic on the overall energy system in China. Finally, we have developed an efficient energy time-series modelling workflow with a high degree of flexibility and portability.

This paper is structured into 7 main sections, each of which is introduced as follows:

In the first section, the *Introduction*, we provide a brief discussion of the current context of the study and the problems with existing research and briefly describe the highlights and innovations of our study.

In Section 2, *Related work*, we present some of the research relevant to this study in two sections, namely the ‘Time-series forecasting model’ and ‘Time-series analysis in COVID-19’.

In Section 3, *Data and data processing*, we first use a data crawler to obtain data on energy production (10 sets of time series including energy raw materials and electricity generation) and on the Chinese pandemic. These two major types of time series are also discussed in detail, and data preprocessing and covariate matrix construction are carried out.

In Section 4, *Methodology and modelling process*, we used the time-series leave-out method to divide the training and test sets by 4:1 for each energy production time series. Finally, we present four performance metrics (RMSE, MAE, MPE and MAPE), which are used as criteria for judging the performance of our models.

In Section 5, *Empirical analysis result*, we analyse and discuss the performance of various models. In the machine learning models, we also include energy time series and pandemic data with the top 3 absolute values of correlation coefficients in each prediction time series to improve the accuracy of the predictions and use t tests to determine whether their performance is improved relative to the original models.

In Section 6, the *Discussion*, we derive the optimal model over the different series and assess the model robustness using 5-fold time-series cross-validation and the bootstrap method. This is followed by further elaboration and discussion of some of the findings from this study.

In the final section, *Conclusions and perspectives*, we summarise the findings and highlights of this study and point out the weaknesses of the study and directions for future research.

2. Related work

Since the industrial revolution, how to predict energy production effectively has been a hot topic for scientists (Clarke and Henderson 2002). Effective prediction of energy production can be of great importance to industrial production. In this section, we review some of the work relevant to this paper in 2 parts. These are the “Time-series forecasting model” and “Time-series analysis in COVID-19”.

2.1 Time-series forecasting model

In this section, we review the two main types of time-series models that are currently in vogue: statistical models and machine learning models.

2.1.1 Statistical models

Statistical models are the most common class of models used in energy forecasting. They include linear regression, SMA, ETS, ARIMA, GARCH and so on (Suganthi and Samuel 2012).

These models are inherently solid in statistical theory and highly interpretable. Therefore, they have been widely used by researchers. For example, in a review of energy forecasting published by Esteves et al. in 2015, 59.3% of energy forecasting models were statistical models, and the most cited model type in energy forecasting articles was statistical models at 67.3%(Esteves et al. 2015). Another review, published by Gargiuloin and Gallachóir in 2013, also illustrates this issue(Gargiulo and Gallachóir 2013). It is easy to see that statistical models are still the most common type of model used in energy forecasting. They are widely used by statisticians and mathematicians with many variants, including the well-known ARIMAX model and VAR model(Toda 1991; Williams 2001).

However, these models often have several shortcomings(Chatfield and Xing 2019). First, such models have a greater need for the stability of the time series, i.e., the time series must not fluctuate significantly. Second, such models often contain some strong assumptions before forecasting; for example, ARIMA models require time-series data to be univariate only. Finally, traditional models have difficulty considering correlations between time series and can only perform univariate autoregressions. Given the suddenness of COVID-19, which can cause time series to fluctuate dramatically(Buehler and Pucher 2021), statistical models tend to be less effective than the machine learning models we will mention below.

2.1.2 Machine learning models

In recent years, with the rise of machine learning research, many of the machine learning models originally used in image classification and natural language processing have been transferred to time-series prediction. This is related to the fact that time series are sequential data (similar to images and text) and that time series are autocorrelated(G. Dong and Pei 2007). Machine learning models fall into two main categories: shallow machine learning models and deep learning models. The former type was mainly used before 2015, including KNN algorithms, random forest algorithms and boosting algorithms(N. K. Ahmed et al. 2010). The latter has emerged since 2015 and includes some novel models, such as MLP, LSTM, GRU and DEEPAR(M. Z. Abedin et al. 2018). The latter models are much more complex than the former models, often using GPUs to train a large network with thousands of parameters.

In regard to using machine learning models in the field of energy, in addition to the study of Tu et al. mentioned in the introduction section, other studies exist in the literature. Liu et al. used a grey neural network model to forecast the supply of the Spanish electricity energy market(X. Liu et al. 2016), Liu et al. used a GRU artificial network model to forecast the supply and demand of major energy sources in China(B. Liu et al. 2017), and finally, Ghoddusi et al. and Kaytez et al. used various machine learning models, including neural networks and SVM algorithms, to forecast the energy supply to make forecasts(Ghoddusi et al. 2019; Kaytez et al. 2015).

The advantage of these nonparametric machine learning models over statistical models is that they can handle unstable series better and do not have as many statistical assumptions as traditional models. In addition, they can use multiple covariates to assist

in forecasting, such as MLP and LSTM models that can use past and future covariates to assist in forecasting.

For traditional time-series models such as ARIMA, covariates can only be added disguised by error-corrected regression, i.e., dynamic regression models, also known as ARIMAX; however, these models have been shown to have limited improvement in model performance for energy forecasting, and in some cases are even weaker than ordinary ARIMA models(Dehghani et al. 2017; Migon and Alves 2013).

In addition, the LSTM neural network model can achieve multiple outputs, which effectively reduces the error accumulation in multistep forecasting of traditional time-series forecasting models. However, the black-box nature of machine learning models often makes the results difficult to interpret(Xu et al. 2021), which may cause decision-makers to distrust their prediction results. Recently, however, some studies have made greater progress in machine learning interpretability. For example, Lundberg et al. used a method known as the *Shapley additive explanations* to interpret the results of machine learning models(Lundberg et al. 2020). These advances allow complex machine learning models to be better grasped.

2.2 Time-series analysis in COVID-19

2.2.1 Research in the early stage of COVID-19

The sudden outbreak of COVID-19 took people by surprise, and early forecasting studies of COVID-19 focused on how to effectively predict the course of the pandemic (i.e., the number of infections and deaths) in the hope of relieving the enormous pressure on the public health system. Two main types of forecasting algorithms were used in the choice of forecasting models. These are epidemiological models and time-series models(Rahimi et al. 2021).

Epidemic models are widely used as a nonparametric simulation model in the COVID-19 pandemic, and the basic idea of an epidemic model is to predict the multifaceted impact of a virus on a complex system(Weiss 2013). Dil et al. used a simple SIR model to predict the number of infections in Pakistan, concluding that Pakistan was likely to experience a surge in cases due to its weak medical infrastructure(Dil et al. 2020). A variant of the SIR model, the SEIR model, was used by Zhao et al. and Reno et al. in their projections for Switzerland and Northern Italy, respectively(Reno et al. 2020; C. Zhao et al. 2020). Their conclusions are similar to those of Dil et al. in that governments must take rapid public health measures to prevent large-scale spread of the pandemic.

Time-series models, the main algorithm used in this study, also have a wide range of applications in COVID-19 forecasting. Similar to the models used in this paper, there are two main categories of time-series forecasting models in these studies, namely, traditional statistical models and machine learning models. In terms of statistical models, Ahmar and Del Val used an ARIMA model to predict the number of confirmed cases in Spain and to analyse the relationship with stock index fluctuations(Ahmar and Del Val 2020), Maleki et al. used a variant of the simple moving average method to predict the number of infections and cures worldwide(M. Maleki et al. 2020), and Chowdhury and

Abedin combined the VAR model and GARCH model to predict the US stock index during COVID-19(Chowdhury and Abedin 2020). For machine models, Chimmula and Zhang used an LSTM model to predict the Canadian pandemic(Chimmula and Zhang 2020), James Fong and Herrera Viedma used a modified SVM model to predict the number of infections worldwide(James Fong and Herrera Viedma 2020), and Moftakhar et al. and Tamang et al. used an MLP model to predict the number of infections and deaths(Moftakhar et al. 2020; Tamang et al. 2020).

2.2.2 Energy forecasts in COVID-19

From the available studies, energy forecasts for the COVID-19 period are still relatively scarce, and these studies have mainly used two different sets of models with different frameworks. Among the traditional models, D'Alessandro and Chapman used the ETS model to forecast the demand and supply of electricity generation during the Japanese pandemic(D'Alessandro et al. 2021), and Narayan used a linear regression model to forecast global oil production during COVID-19(Narayan 2020). Additionally, Alhajeri et al. used a linear regression model modified by a genetic algorithm to predict oil production in Kuwait during this period(Alhajeri et al. 2020). In terms of machine learning models, Tu et al. used an LSTM model to predict the electricity consumption of buildings during COVID-19(Tu et al. 2020). Wang et al. also illustrate the advantages of machine learning techniques in COVID-19 energy prediction, although they did not perform prediction experiments(B. Wang et al. 2020).

It is easy to notice that most of the energy forecasts during the COVID-19 pandemic focused on only one energy type and used a single forecasting model. Many studies have focused on the impact of the pandemic on energy markets, with little mention of forecasting. The impact of the epidemic on different energy sources varies considerably; for example, the impact on thermal power generation is significant between epidemics, while the impact on clean energy sources such as hydroelectricity is relatively small(Agdas and Barooah 2020; Kuzemko et al. 2020).

There is also no universal model for forecasting different types of energy production. For example, Yang et al. showed that LSTM models outperform traditional statistical models for thermal power generation(Yang et al. 2020), while H. Liu et al. showed that traditional time-series models outperform MLP neural network models for wind power generation(H. Liu et al. 2010). In this paper, various models have been developed to predict and discuss different species. We also refer to Agdas and Barooah's Practice, who used epidemic data as a covariate in energy forecasting for California(Agdas and Barooah 2020). We incorporated both epidemic data and highly correlated energy data into the model for multivariate forecasting by following their idea.

3. Data and data processing

In this section, we first describe the data acquisition process and perform descriptive statistics on the data. We then preprocessed the data and performed correlation analysis. The flowchart is shown in the following diagram in [fig. 1](#).

Figure 1: Data workflow

3.1 Data collection

The focus of this paper is on the time series of Chinese energy production in the COVID-19 pandemic. We use two main types of time-series data, namely, Chinese energy data and Chinese pandemic data. Due to the large amount of data to be captured and the tedious task of manually obtaining these data, we first developed a crawler using Python’s Scrapy framework to crawl the relevant data(Myers and McGuffee 2015).

For Chinese energy data, we use publicly available energy data from the National Bureau of Statistics of China (NBS) as our data source¹. We crawl two main types of energy time-series data, namely, different types of energy raw material output and different types of electricity generation. All time series are in months and span from October 2018 to July 2021.

The types of energy raw materials are crude oil production (10,000 tons), crude oil processing output (10,000 tons), raw coal output (10,000 tons), natural gas output (100 million m³) and gas production output (100 million m³). The types of electricity generated are thermal power generation (100 million kWh), solar power generation (100 million kWh), nuclear power generation (100 million kWh), hydropower generation (100 million kWh) and wind power generation (100 million kWh). We then stored the captured data in a SQLite database, which is a lightweight database that facilitates data persistence in our research(Bhosale et al. 2015).

For the Chinese pandemic data, we used publicly available data from Johns Hopkins University as the data source² and collected six types of time-series data since the pandemic outbreak through Scrapy, including three types of incremental data, new infections, new deaths and new vaccinations, and three types of cumulative data,

¹ China National Bureau of Statistics Macro Energy Data. <http://www.stats.gov.cn/>

² COVID-19 Open Data, Johns Hopkins University.
<https://github.com/CSSEGISandData/COVID-19>

cumulative infections, cumulative deaths and cumulative vaccinations. All data are for the period January 2020 to September 2021. Again, after completing the crawl, we stored the data in a SQLite database for further analysis.

After completing the collection of energy data and epidemiological data, we next carried out descriptive analysis of these two main categories of data.

3.2 The energy data

The following figures, [fig. 2](#) and [fig. 4](#), are time-series plots of energy raw materials and electricity generation, respectively, while the other two figures, [fig. 3](#) and [fig. 5](#), are bar charts for both. We have marked the time points of the COVID-19 outbreak with red lines ³.

First, we analyse the energy raw materials in [fig. 2](#) and [fig. 3](#). It is easy to note that COVID-19 had a significant impact on crude oil processing and raw coal production. Both of them suffered severe declines in production as a result of COVID-19, due in large part to the lockdown and stoppage measures taken by China during the pandemic, which greatly affected these two labour-intensive raw material industries (“Wuhan lockdown ‘unprecedented’, shows commitment to contain virus” 2020).

The impact of other categories lagged behind, such as gas and natural gas, both of which experienced a significant drop in production by mid-2020. Finally, crude oil was apparently less affected by the pandemic, and production did not fluctuate much. Currently, raw coal is still the most important source of energy in China (Yuan 2018), while crude oil processing is of great importance for the country’s transport industry (Y. Zhao et al. 2007), and the fall in production of both has had a negative impact on China’s economic recovery.

The spike in raw coal prices and gasoline prices in China’s energy market since mid-2020 has put the country’s industrial recovery under great pressure.

³ We took the time point for activation of the WHO outbreak crisis management team, which occurred on 1 January 2020

Figure 2: Time series of various raw material production processes

Figure 3: Bar charts of various raw material production processes

A similar phenomenon can be observed with electricity generation. Comparing [fig. 4](#) and [fig. 5](#), we see a significant decline in both nuclear and thermal power generation. In particular, thermal power generation has exhibited a significant downward trend since the outbreak. It was not until mid-2020 that the previous output was gradually restored. On the one hand, it is related to the fact that in China, the correlation between thermal power generation and coal production is relatively high(Lam and Shiu 2004). On the other hand, China's thermal power generation manpower needs are high(L. Liu et al. 2014), and the lockdown caused by COVID-19 forced plants to shut down("COVID-19 lockdown in China" 2021). Similarly, nuclear energy was at risk of shutting down its output facilities and thus experienced a significant downward trend during the early period of COVID-19.

The other three major energy sources were hydro, solar and wind. The impact of the epidemic was relatively small because the winter and spring seasons, when the epidemic was most severe, are not peak production periods. Additionally, their manpower needs are low.

Thermal power generation is the most important means of power generation in China, and the decline in power generation has had an extremely negative impact on China's industrial recovery and will continue until late 2021 ("Unprecedented' power cuts in China hits homes, factories" n.d.).

Figure 4: Time series of various electricity generation processes

Figure 5: Bar charts of various electricity generation processes

With the above descriptive statistics, we find a clear difference in the degree of impact of the pandemic on different energy time series. Both raw coal production and gasoline production, important indicators for measuring China's economic recovery, experienced severe declines, and both reached their lowest points in the observed time

period (October 2018 to July 2021). Thermal power generation has also been affected by the decline in raw coal production and the lockdown, and its output is considerably lower.

Time series with different fluctuations pose some difficulties for us to model; however, traditional statistical models often have the requirement of series stability. It is obvious that the pandemic has made certain series more volatile. This makes it necessary to use nonparametric models such as machine learning for modelling such series.

3.3 The COVID-19 data

In this section, we present a descriptive analysis of Chinese pandemic data. In [fig. 6](#), we show a time-series plot of the various types of pandemic data for China, and we indicate the time when city closures started in China with a red line ⁴. The top half of the figure shows the various types of incremental data, and the bottom half shows the various types of cumulative data. We can see that the new cases and new deaths in China occurred mainly between January 2020 and March 2020. After that, the number of new cases declined rapidly, and there were no more large-scale outbreaks in China, except for occasional localised epidemic fluctuations. The overall accumulation curve thus shows a slow upward trend. The number of new deaths in China flattened rapidly after the first wave of the national epidemic. Since February 2021, when the vaccine became available, the number of vaccinations in China has been rising rapidly. It was only in mid-2021 that the number of new vaccinations began to slowly decrease.

China's epidemic prevention and control has clearly been more successful. The first wave of a nationwide outbreak was quelled in just 3 months. The time series in the figure shows that the number of new cases and new deaths declined rapidly after China began to lock down the city, suggesting that the closure measures had clearly contained the spread of the epidemic to some extent. Additionally, after the advent of the vaccine, the number of vaccinations in China rose dramatically, reaching 2.07 billion doses in China by July 2020. The rapid lockdown measures and high vaccination rates made China one of the first few countries in the world to complete a full resumption.

Effective preparedness measures and early work resumption allowed China's energy production and power generation to recover quickly. This has allowed China to recover much more than other developing countries (Hassan, Chowdhury, et al. 2021). Combined with the analysis of China's energy time series in the previous section, we can understand that the most severe period of the epidemic in China was from January 2020 to early April 2020, which was also a time of dramatic fluctuations in the production and generation of each energy source. During this period, China took measures to lock down cities and shut down production, and energy production came to a standstill, while the demand for energy from various industries was greatly reduced.

⁴ We took the time point of the lockdown in Wuhan, which occurred on 23 January 2020.

Figure 6: Time series of the COVID-19 pandemic (China)

3.4 Data standardisation

Machine learning models require us to normalise the data. We use the Z score method to normalise the data distribution to a mean of 0 and a standard deviation of 1 (Cheadle et al. 2003). The formula for the standard fraction method is as follows:

$$z = \frac{x - \tilde{x}}{S} \quad (1)$$

In eq. 1, z denotes the standardised observation, $\tilde{x}$ denotes the mean of the variable, and S denotes the standard deviation of the variable.

We use the adjusted data as new variables to overwrite the original variables. After completing the prediction, we revert the data using the inverse normalisation formula in eq. 2

$$x = z \times S + \tilde{x} \quad (2)$$

3.5 Correlation analysis and data matrix construction

We further analyse the correlation between the epidemic data and the energy data. We used the Pearson correlation coefficient to perform correlation analysis on various time series (Benesty et al. 2009). The two variables to be measured are X and Y , and the Pearson correlation coefficient is set to r , which gives us the formula,

$$r = \frac{\sum_{i=1}^n (X_i - \tilde{X})(Y_i - \tilde{Y})}{\sqrt{\sum_{i=1}^n (X_i - \tilde{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \tilde{Y})^2}} \quad (3)$$

In eq. 3, X_i denotes the value of the i th observation in variable X , and Y_i denotes the value of the i th observation in variable Y . $\tilde{X}$ and $\tilde{Y}$ denote the mean of the two variables.

In fig. 7, we show a heatmap of the correlations for each time series, where darker colours represent higher correlations between variables, with red colours representing positive correlations and grey colours representing negative correlations. The information in the figure allows us to visually show that many types of energy data have a negative correlation with new cases and new deaths (coloured grey-blue). In contrast, there was a positive correlation (coloured red) for the number of new vaccinations, cumulative cases, cumulative deaths and cumulative vaccinations. There were also differences in the strength of the correlation between the time series. For example, there was a strong positive correlation between cumulative cases and cumulative deaths; however, although it was also positive with the cumulative vaccination, the strength was much less.

We can also find some correlations between the various energy time series, such as a strong positive correlation between natural gas generation and wind power generation, a strong positive correlation between crude oil production and gas generation, and a strong negative correlation between hydroelectric generation and natural gas generation.

Figure 7: Heatmap of different sequences

If cross-sectional/panel data are employed, we can add correlated variables as covariates to the model to improve the performance of the model. However, for time-series statistical models, only autoregressive forecasts can be made using univariate variables, and correlations between time series are often ignored. However, these often contain important information in energy forecasts, as we discussed earlier, and changes in pandemics can affect energy production to some extent. Energy sources also interact with each other. However, multivariate series are difficult for traditional models to handle.

Machine learning models break this limitation by adding other relevant time-series data as covariates to the model to assist us in making predictions. In this study, we decided to use pandemic time series and highly correlated energy time series as covariates to improve the accuracy of our predictions. We construct the covariate matrix for the machine learning time-series model. We added the epidemic data of six species as covariates to each model. For the choice of the energy time series, we select the three energy time series with the highest absolute correlation with the predicted time series as covariates. A figure of the input data matrix is shown in [fig. 8](#).

Date	Predictor	Covid19 covariate	...	Covid19 covariate	Energy covariate	...	Energy covariate
1-Jan	Num	Num	...	Num	Num	...	Num
1-Feb	Num	Num	...	Num	Num	...	Num
1-Mar	Num	Num	...	Num	Num	...	Num

Figure 8: Time-series input data matrix

With the above accomplished, we can start building our model.

The modelling workflow for this study is shown in [fig. 9](#) below.

Figure 9: Modelling workflow

We first divided the training and test sets by the time-series hold-out method, after which we trained the model on the training set.

MAE, MPE and MAPE, to test the model performance on a test set. Finally, to verify the robustness of the results, we used time-series cross-validation for the optimal model. A bootstrap method was also used to estimate the confidence intervals for each performance indicator after the cross-validation results.

4.1 Time-series segmentation

The method of dividing time series is different from the traditional classification or regression tasks in machine learning. In particular, an approach that first shuffles the dataset and then divides it is inappropriate. Since there is autocorrelation in the time series, we must sort the dataset in chronological order and then divide it.

The method of dividing the dataset into a training set and a test set on a proportional basis is known as the time-series hold-out method, as shown in Figure [fig. 10](#). In contrast, [fig. 11](#) starts with a small part of the dataset, makes predictions on a later part of the data, and then tests the accuracy of the predictions. The process is then repeated until the entire dataset is exhausted. We define the number of cycles as K , and this method is known as the time-series K -fold cross-validation method.

The time series in this paper spans only 3 years, while the time-series units are represented by months. To make the best use of the data, we first divided the dataset in a 4:1 ratio using the time-series hold-out method, i.e., 80% of the data were used as the training set and 20% for the test set, as shown in [fig. 10](#). For neural networks, such as NNET and LSTM, training also requires a validation set for iterative optimisation, and since our data size is relatively small, we use the test set as a validation set as well.

We then test the robustness of the results, which we validate using the method shown in [fig. 11](#), i.e., the time-series 5-fold cross-validation method. The bootstrap method is also used to obtain confidence intervals for the metrics to ensure the robustness of the test results.

Figure 10: Time-series hold-out method

Figure 11: Time-series 5-fold cross-validation

In this study, we used ARIMA and ETS as traditional statistical models. A seasonal naive model is used as the baseline model. Due to the multistep time-series forecasting in this study, i.e., forecasting multiple consecutive time points, we used the recursive strategy method as a multistep forecasting strategy. The recursive strategy method is one of the most commonly used multistep forecasting strategies and is widely used in energy forecasting (Hadri et al. 2021; Taieb and Hyndman 2012).

4.2.1 Seasonal naive model (baseline)

We use the seasonal naive model as the baseline model. In the time-series prediction, we simply set all predictions to the value of the last observation that arrived to obtain the naive model.

Based on the naive model, each forecast is set to the last observation in each season, where the season refers to a cyclic time period that repeats regularly over time, not a real-world season, to obtain the seasonal naive model, which is defined as follows:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \quad (4)$$

In eq. 4, m is the seasonal period. In addition, k is the integer part of $(h - 1)/m$.

We find that the seasonal naive model is seasonal in nature and can be predicted according to a certain time pattern. Therefore, it is more suitable as a baseline model than the naive model.

4.2.2 ETS model

Since its introduction in the late 1950s, the *exponential smoothing model* (ETS) model has been one of the most effective and classical parametric time-series forecasting models (Holt 2004; Winters 1960). The forecasts generated using the ETS model are weighted averages of past observations, and the weights decay exponentially as the

distance of past observations from the forecast increases. In other words, the closer the observation is, the higher the corresponding weight is, and vice versa.

The ETS model is determined by three main coefficients: error, trend and seasonality. The interoperations between the three main parameters make up the different ETS models. In addition, eq. 5 is the specific form of the ETS model,

$$ETS(X_1, X_2, X_3) \quad (5)$$

In eq. 5, the first term X_1 is the error term, which indicates the additive or multiplicative error of the model, and is denoted as A if it is additive and M if it is multiplicative. The second term, X_2 , is the trend term of the model. If this term is N , it means that there is no trend in the model; if there is an additive trend, it is A ; and if there is a decaying trend, also known as additive damping, it is A_d . The last term is seasonality, again denoted by N if there is no seasonality, A if there is additive seasonality, and M if there is multiplicative seasonality. By combining the above components, we can obtain different ETS models. For example, eq. 6 represents a multiplicative Holt-Winters method model with additive error:

$$ETS(A, A, N) \quad (6)$$

To address the issue of how to select the appropriate ETS model parameters, we use the corrected *Akaike's information criterion* (AIC_c) to select the best model parameters (Bozdogan 1987). AIC_c is an improvement of *Akaike's information criterion* (AIC) and is better adapted to the case of small samples. AIC is defined as:

$$AIC = -2\log(L) + 2k \quad (7)$$

In eq. 7, L is the likelihood function of the model, and k is the sum of the number of parameters that have been estimated and the initial state. AIC_c adds a correction term, as shown in the following equation, eq. 8,

$$AIC_c = AIC + \frac{k(k+1)}{T-k-1} \quad (8)$$

The *Bayesian information criterion* (BIC) or *Akaike's information criterion* (AIC) can also be chosen for the judging criteria. In this study, we calculate the AIC_c for different combinations by continuously combining various parameters. The model with the lowest AIC_c value is chosen as the optimal ETS model.

4.2.3 ARIMA model

The *autoregressive integrated moving average model* (ARIMA) is another well-known univariate time-series forecasting model. Unlike the ETS model, which uses trends and seasonality in the data to build a model, the ARIMA model aims to portray the autoregression of the data (Nelson 1998; Saboia 1977).

The ARIMA model is first obtained by differencing the nonstationary series to obtain the stationary series, which is then combined with an autoregressive model and a moving average model to make forecasts. The ARIMA model equation is as follows.

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (9)$$

In eq. 9, y'_t is the sequence of differences. The right-hand side of the formula includes the delayed value of y_t and the error in the delayed value ε_t . For ease of representation, we will refer to this model as

$$ARIMA(p, d, q) \quad (10)$$

In eq. 10, p is the autoregressive model order, d is the difference order, and finally, q is the moving average model order. The components of eq. 9 can then be interpreted as the following eq. 11:

$$\underbrace{(1 - \phi_1 B - \dots - \phi_p B^p)}_{AR(p)} \underbrace{(1 - B)^d}_{d \text{ differences}} y_t = \underbrace{c + (1 + \theta_1 B + \dots + \theta_q B^q)}_{MA(q)} \varepsilon_t \quad (11)$$

With different combinations of p , d and q , we can obtain different ARIMA models. For the seasonality that often occurs in time series, we can introduce a seasonal term for eq. 10, i.e., eq. 12, where m denotes the number of observations per year. P , D , and Q have a similar meaning to the lower-case form explained above; however, they belong to the parameters associated with the back m of the seasonal period.

$$ARIMA(p, d, q)(P, D, Q)_m \quad (12)$$

Similar to the ETS model, we select the best model parameters by AIC_c . The lowest combination of AIC_c is obtained by continuously iterating through the system with various combinations of parameters.

4.3 Machine learning forecasting models

In this study, we used three machine learning models for prediction, namely, the LightGBM, NNAR and LSTM models. The former is a shallow machine learning model, and the latter two are deep learning neural network models. We also use the recursive strategy method as our temporal multistep prediction strategy.

4.3.1 LightGBM

H

LightGBM is a decision tree-based gradient boosting algorithm framework proposed by Microsoft(Ke et al. 2017). As an improved version of gradient boosting, it features high accuracy, high training efficiency, support for parallelism and GPUs, small memory requirements, and the ability to handle large-scale data. Its API interface is simple, efficient, and easy to use.

In machine learning, multiple weak learners are combined to form a new learner to improve the efficiency of a model. This method is called ensemble learning(Bauer and Kohavi 1999). Ensemble learning methods can be divided into bagging algorithms and boosting algorithms, and boosting algorithms can be divided into AdaBoost and gradient boosting. The gradient boosting algorithm is represented by XGBoost and the LightGBM, which are used in this paper. The core idea of gradient boosting is to use the

negative gradient of the loss function to approximate the value of the current model $f(x) = f_{j-1}(x)$ to replace the residuals.

Suppose the training sample is i , the number of iterations is j , and the loss function is L ; then, the negative gradient r_{ij} can be expressed as [eq. 13](#)

$$r_{ij} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{j-1}(x)} \quad (13)$$

The base learner $h_j(x)$ is used to determine the negative gradient value r of the loss function, and the optimal gradient value γ_j for minimising the loss function by the following formula is expressed as [eq. 14](#):

$$\gamma_j = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, f_{j-1}(x_i) + \gamma h_j(x_i)) \quad (14)$$

After that, we update the model to [eq. 15](#),

$$f_j(x) = f_{j-1}(x) + \gamma_j h_j(x) \quad (15)$$

As indicated by the above equation, gradient boosting generates a base learner in each iteration. Each round of learners is added to the previous round of learners. After the model has converged, we obtain the theoretically strongest learner.

LightGBM, as an improved lightweight gradient boosting algorithm, can effectively control the complexity of the model and reduce the training space and time, thus realising the lightweight nature of the algorithm. We follow the format in [fig. 8](#) to feed the model with lagged values of predicted and covariate timeseries and obtain the next predicted values by the recursive strategy method. To select the optimal parameters for each energy time series, we use the grid search method(Liashchynskyi and Liashchynskyi 2019) and use *RMSE* as the judging criterion, where the lower the *RMSE* is, the better the parameter combination.

4.3.2 NNAR

NNAR, or the neural network autoregressive model(A. Maleki et al. 2018), is a variant of a neural network model in time series. The NNAR model used in this study is a feedforward neural network (also known as MLP) with a single hidden layer and lagged inputs(Pinkus 1999), which has achieved excellent results in financial series forecasting(M. Z. Abedin et al. 2019). The structure of the model is shown in [fig. 12](#), where the black nodes are the neural network nodes and the blue nodes are the residuals of the neural network. The MLP consists of a combination of input, hidden and output layers, where each layer node receives the input from the previous layer, the output of each layer node is the input of the next layer, and the nodes perform a weighted linear combination of the inputs.

Figure 12: MLP model structure

In the hidden layer, we use an activation function to map the input of the neuron to the output, and [eq. 16](#) is the sigmoid activation function. The activation function tends to reduce the effect of extreme input values, thus making the neural network robust to outliers.

$$\sigma = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1} \quad (16)$$

We use a nonlinear function to output the results, as shown in the following function [eq. 17](#),

$$z_j = b_j + \sum_{i=1}^n w_{i,j} x_i \quad (17)$$

The weights $w_{i,j}$ and the deviations b_j are learned from training on the data and are updated after each training round according to the loss function. We usually need to limit the weight size to prevent it from being too large or too small, and this parameter is called the learning rate. It is computed by an optimiser, such as the well-known Adam optimiser (Kingma and Ba 2014).

MLP takes the lagged values of the time series as input and feeds them into the model with k hidden layer nodes according to the specified number of lags p . After completing the training, the updated weights are fed backwards to the model, and then, the training is cycled until the model reaches convergence. For data with seasonality, we can add the last observation of the same season as input, and we set the season length to m and feed the number of lags in the season to the model as P . This gives us the NNAR of the form in [eq. 18](#),

$$\text{NNAR}(p, P, k)_m \quad (18)$$

For the input of covariate time series, we also input these time series into the model together after lagging. We use AIC_c to choose the best model among different parameter combinations.

4.3.3 LSTM

The LSTM, *long short-term memory network model*, is a variant of a recurrent neural network model (Kong et al. 2017). The LSTM model can establish a temporal association between the previous time point $t - 1$ and the current time point t . In a sense, LSTM should be the most suitable model for predicting energy production because it has the ability to infer the logic of intrinsic energy production.

LSTM calculates weights by a back propagation learning algorithm (Kong et al. 2017). It consists of cells called memory blocks, each of which contains an “input gate”, an “output gate”, and a “forget gate”, as shown in fig. 13, which illustrates an LSTM model cell. Next, we describe the components of fig. 13.

Figure 13: LSTM network structure

The equation for the “input gate” I_t is eq. 19. The “input gate” I_t transmits the output H at moment $t - 1$ and the input X at moment t through the σ function (eq. 16).

$$I_t = \sigma(w_i \cdot [H_{t-1}, X_t]) + b_i) \quad (19)$$

After that, the hyperbolic tangent function $\tanh$ is applied to the inputs and outputs of the previous step, thus creating a new value $\tilde{C}_t$ to be used as the internal state. The update of the internal state is performed by the following equation eq. 20:

$$\tilde{C}_t = \tanh(w_c \cdot [H_{t-1}, X_t] + b_c) \quad (20)$$

The “forget gate” F_t is also computed using the σ function, which takes the previous output H_t and the input X_t to perform the computation. It is defined as shown in eq. 21:

$$F_t = \sigma(w_f \cdot [H_{t-1}, X_t] + b_f) \quad (21)$$

Last, the “output gate” O_t described by eq. 22 is changed based on the state C_t (eq. 23). The C_t state, in turn, is continuously updated by multiplying the hyperbolic tangent function $\tanh$ with the output of the function σ ,

$$O_t = \sigma(w_o \cdot [H_{t-1}, X_t] + b_o) \quad (22)$$

$$C_t = F_t \times C_{t-1} + I_t \times \tilde{C}_t \quad (23)$$

The w_i , w_c , w_f , and w_o parameters and b_i , b_c , b_f , and b_o parameters mentioned in the above equations represent the weights and deviations of different levels in the LSTM memory block, respectively (Kong et al. 2017). They are adjusted iteratively with the BPTT learning algorithm until convergence (Guo 2013).

In this study, the training data and covariate data follow the form in fig. 8 that we mentioned earlier. They are fed into the model for training.

4.4 Evaluation metrics

We have taken as references the studies from Abedin et al. and Tay et al. (M. Abedin et al. 2020; Tay and Cao 2001). Four metrics commonly used in time-series model evaluation were chosen to assess the performance of our model, namely, RMSE, MAE, MPE and MAPE. To better explain the equations below, we denote y_i as the true value and $\hat{y}_i$ as the estimated value. Finally, n denotes the number of times the model predicts.

4.4.1 RMSE

The *root mean squared error* (RMSE) is a commonly used measure for regression analysis and is the standard deviation of the residuals. A smaller RMSE means that our model is performing better. In the following, eq. 24 is the RMSE formula.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (24)$$

4.4.2 MAE

The *mean absolute error* (MAE) calculates the average size of the residuals without taking into account the positive or negative nature of the residuals; the MAE takes on values ranging from 0 to positive infinity, and again, a smaller MAE indicates a better performing model. In general, MAE values are much smaller than RMSE values because MAE is linear, whereas RMSE shows quadratic term growth. The MAE formula is expressed as eq. 25,

$$MAE = \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{n} \quad (25)$$

4.4.3 MPE

The *mean percentage error* (MPE) is the average of the percentage errors between the predicted and observed values of the model. Since the formula uses the actual value of the residuals, the positive and negative prediction errors cancel each other out. When MPE is equal to 0, it means that the model has perfect prediction performance. However, when the true value of y_i is 0, this metric will not be available. A smaller value of MPE indicates a better predictive model performance. The MPE formula is expressed as [eq. 26](#),

$$\text{MPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \quad (26)$$

4.4.4 MAPE

Mean absolute percentage error (MAPE) is an improved indicator of MPE. One advantage of MAPE is that it is more explanatory, with a MAPE value of n representing an average deviation of $n\%$ from the true outcome. As with MPE, this metric will not be available when the true value of y_i is 0. A smaller value of MAPE indicates better performance of the prediction model. We can see that the MAPE formula in [eq. 27](#) has an additional absolute value sign compared to [eq. 26](#).

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (27)$$

4.5 Machine overview

For the traditional statistical models in this study, we used the Fable package in R as a modelling tool (O'Hara-Wild et al. 2021). For the machine learning models, we used Scikit-learn (Pedregosa et al. 2011) and Keras (Chollet and others 2015) in Python as modelling tools. All models were run on a desktop computer with a GTX-2070 GPU.

5. Empirical analysis result

5.1 Statistical model results

5.1.1 Baseline model results

We first test our data on the baseline model, the seasonal naive model. First, the baseline model is implemented on the energy raw material dataset, as shown in [tbl. 1](#). We plot the performance of the baseline model on the energy raw materials in [fig. 14](#). We use the red line as the dividing line between the training and test sets, and the series after the red line are the predicted series. Two confidence intervals are also given: an 80% confidence interval and a 95% confidence interval.

The baseline model performs better for crude oil output and natural gas output. In the natural gas output, the RMSE and MAE values are the lowest of all series, while in

the crude oil output, the MPE and MAPE values are the lowest of all series, implying an average deviation of 2.68% from the true results.

The inconsistency between the RMSE and MAPE best results above is because RMSE is an absolute measure, while MAPE is a relative measure. As MAPE is a relative metric that allows for better comparison of model performance across different series, we will focus on using the MAPE metric to characterise model performance next.

However, the baseline model does not perform well on the other series. Looking at the time series of both variables, it is easy to see that there is a large deviation in the forecasts.

Table 1: Performance of the baseline model on energy raw materials⁵

Model	Variable	RMSE	MAE	MPE	MAPE
Baseline	COO	45.64	44.63	2.15	2.68
	COPO	586.13	498.66	8.28	8.52
	GO	83.88	75.21	3.95	5.68
	NGO	15.26	14.63	8.52	8.52
	RCO	3478.44	2206.73	5.76	7.09

⁵ For convenience we use abbreviations of energy raw material types to represent the chronological names, i.e., COO for crude oil output (10,000 tons), COPO for crude oil processing output (10,000 tons), GO for gas output (100 million m³), NGO for natural gas output (100 million m³), and finally, RCO for raw coal output (10,000 tons).

Figure 14: Baseline model predictions on energy raw material data

We then proceeded to test our baseline model on the generation dataset with the following results in [tbl. 2](#) and [fig. 15](#): The performance of the baseline model is clearly weaker in power generation compared to that of energy raw materials. With the exception of the hydropower generation series, all the series have MAPE values greater than 10, which means that the predictions differ on average by more than 10% from the true results.

Table 2: Performance of the baseline model on electricity generation⁶

Model	Variable	RMSE	MAE	MPE	MAPE
Baseline	HG	64.35	57.69	0.21	6.88
	NPG	42.53	40.46	12.37	12.37
	SPG	30.68	28.39	19.60	19.60
	TPG	695.75	654.71	13.64	13.64
	WPG	113.24	105.30	23.54	23.54

⁶ Similarly, for ease of presentation, we use abbreviations for the types of generation to represent the names of the time series. i.e., HG for hydropower generation (100 million kwh), NPG for nuclear power generation (100 million kwh), SPG for solar power generation (100 million kwh), TPG for thermal power generation (100 million kwh), and finally, WPG for wind power generation (100 million kwh).

Figure 15: Baseline model predictions on electricity generation

From the above analysis, it is clear that the baseline model, the seasonal naive model, is far from adequate for effective energy forecasting. In particular, there is a large deviation in the forecasting of electricity generation, but as a baseline model, it is nevertheless successful in identifying the seasonality of some series.

5.1.2 ETS model results

We tested the dataset on the ETS model. The results on the energy raw material time series are shown in [tbl. 3](#) and [fig. 16](#). Compared to the baseline model, the ETS model shows a more significant improvement in performance. All series outperform the baseline model, except for the natural gas output series, which does not perform as well as the baseline model. The best performing gas output series has a MAPE value of 2.20, which is related to the low volatility of the gas output series. However, looking at the time-series plot in [fig. 16](#), it is easy to see that the ETS model's prediction is almost a straight line and that the model does not well address the seasonality problem that exists in the time series.

Table 3: Performance of the ETS model on energy raw materials

Model	Variable	RMSE	MAE	MPE	MAPE
$ETS(M, N, N)$	COO	73.74	70.47	2.97	4.23
$ETS(M, N, N)$	COPO	173.65	142.47	-1.72	2.46
$ETS(A, N, N)$	GO	34.09	29.17	-0.52	2.20
$ETS(M, N, N)$	NGO	17.36	15.64	-9.34	9.34

$ETS(M, N, N)$	RCO	1553.64	1395.53	-3.69	4.42
----------------	-----	---------	---------	-------	------

Figure 16: ETS model predictions on energy raw material data

The next series are the results for the generation series, as shown in [tbl. 4](#) and [fig. 17](#). The ETS model does not perform as well in the generation series as it does in the energy raw material series. The best result is in the thermal power generation sequence, but the MAPE value also reaches 5.04.

The worst performance is in the hydropower generation and solar power generation series, with MAPE values of 9.55 and 27.45, respectively, both weaker than the baseline model. However, the model effectively captures the seasonal trends in the hydropower generation series.

Table 4: Performance of the ETS model on electricity generation

Model	Variable	RMSE	MAE	MPE	MAPE
$ETS(M, N, M)$	HG	96.34	84.07	-9.55	9.55
$ETS(A, A, N)$	NPG	26.80	22.69	-1.48	7.00
$ETS(A, N, N)$	SPG	44.63	41.43	27.46	27.46
$ETS(M, N, N)$	TPG	336.17	251.06	4.72	5.05
$ETS(A, N, N)$	WPG	59.93	54.41	8.31	11.75

Figure 17: ETS model predictions on electricity generation

Overall, the ETS model outperforms the baseline model in general; however, it still performs weaker than the baseline model in some series. Next, we use the ARIMA model, which is considered to be the most effective traditional statistical model.

5.1.3 ARIMA model results

The ARIMA model has been the most popular and the best performing time-series model until the advent of machine learning models. To test its performance on the energy dataset, we combined ARIMA models with different parameters for testing. The performance on the energy raw material dataset is shown in [tbl. 5](#) and [fig. 18](#).

Compared to the ETS model, which sometimes even performs worse than the baseline model, the ARIMA model outperforms the benchmark model across the board for the energy raw materials dataset. The best performing series is the crude oil output series with a MAPE value of 1.66, which shows that the model has successfully identified the pattern of the crude oil output series. The worst performing series is the raw coal output series with a MAPE value of 6.86, where the model incorrectly identifies seasonality, allowing forecasts to start from lower values.

However, we still find that the prediction of the crude oil processing output series is almost a straight line. There is clearly room for improvement, and we will be exploring ways to improve this with machine learning.

Table 5: Performance of the ARIMA model on energy raw materials

Model	Variable	RMSE	MAE	MPE	MAPE
$ARIMA(0,1,1)(0,1,0)_{12}$	COO	30.03	27.25	0.68	1.66
$ARIMA(3,1,0)$	COPO	180.60	147.54	-1.92	2.55
$ARIMA(3,1,1)(0,1,0)_{12}$	GO	79.11	77.10	-5.49	5.39
$ARIMA(3,1,1)(0,1,0)_{12}$	NGO	5.70	4.74	-2.03	2.76
$ARIMA(2,1,2)(0,1,0)_{12}$	RCO	3076.77	2339.00	2.50	6.86

Figure 18: ARIMA model predictions on electricity generation

We continue to test the ARIMA model on the generation dataset with the results shown in [tbl. 6](#) and [fig. 19](#). Compared to the baseline model, the ARIMA model continues to outperform across the board. The ARIMA model performs best for hydropower generation with a MAPE of 5.70 and worst for wind power generation with a MAPE of 13.41. [Fig. 19](#) also shows that the ARIMA model essentially identifies serial monthly fluctuations and seasonal fluctuations.

Table 6: Performance of the ARIMA model on electricity generation

Model	Variable	RMSE	MAE	MPE	MAPE
$ARIMA(2,0,1)(0,1,0)_{12}$	HG	66.39	50.18	-3.03	5.70
$ARIMA(1,1,3)(0,1,0)_{12}$	NPG	24.36	21.74	6.38	6.70
$ARIMA(1,1,2)(0,1,0)_{12}$	SPG	14.78	14.15	6.20	9.73
$ARIMA(3,1,1)(0,1,0)_{12}$	TPG	563.96	527.30	10.94	10.94
$ARIMA(4,1,0)(0,1,0)_{12}$	WPG	72.31	60.54	13.11	13.41

Figure 19: ARIMA model predictions on electricity generation

In general, the ARIMA model is indeed the best performing of the traditional statistical models. It outperforms the baseline model across the board in both broad categories of series, and there are only a few groups of series that do not perform as well as the ETS model. From [fig. 18](#) and [fig. 19](#), we also find that the ARIMA model is largely successful in identifying the volatility patterns of these series. However, we also observe that in some series, the confidence intervals for the ARIMA predictions are too large, and the predictions are not very accurate. Additionally, due to the relatively small size of our dataset, some erroneous patterns were identified, such as wind power generation.

These problems are a drawback of traditional statistical models, which have difficulty making effective predictions on small datasets and tend to overfit. This means that they are excellent for in-sample predictions but not as good for out-of-sample data. Therefore, we will use machine learning models for testing in the following sections.

5.2 Machine learning model results

Before training the model, we needed to construct the covariate matrix following the method of constructing covariates mentioned previously. We added all the epidemic time-series data as covariates to each model and selected the three energy time series with the highest absolute values of correlation coefficients with the predicted time series as covariates. The results are presented in [tbl. 7](#) below, and we added the covariates as shown in the table.

To compare the performance of the model with the addition of covariates, we added the results of the MAPE subtraction between the model with and without covariates to the tables. If the result is negative, the model outperforms the original model with the addition of covariates; otherwise, the result is positive.

In addition, we performed a 5-fold cross-validation test between the covariate model and the original model to ensure the robustness of the results. A t test was also conducted on both models to check whether the model after the inclusion of covariates truly outperformed the model without the inclusion of covariates. We set the null hypothesis that there is no significant difference between the MAPE values of the model with the addition of covariates and the MAPE values of the original model, and the alternative hypothesis is that the MAPE values of the model with the addition of covariates are significantly smaller than the MAPE values of the original model, i.e., that the model outperforms the original model.

Table 7: Matrix of covariates

Predictor	Covariate1	Covariate2	Covariate3	Special Covariate
COO	SPG	NPG	GO	Covid-19 Covariate
COPO	GO	NPG	SPG	
GO	COPO	NPG	SPG	
NGO	WPG	TPG	SPG	
RCO	NPG	COPO	GO	
HG	WPG	NGO	NPG	
NPG	COPO	TPG	RCO	
SPG	GO	NPG	COPO	
TPG	NPH	COO	NGO	
WPG	NGO	SPG	HG	

5.2.1 LightGBM model results

The first model to be analysed is the LightGBM model. To set the parameters, we use the grid search method to obtain the optimal parameters (Liashchynskiy and Liashchynskiy 2019). Our choice of metric for evaluating the optimal parameters is MAPE. The model exhausts the parameters within the specified grid to obtain the combination with the lowest MAPE value.

The results⁷ are shown in [tbl. 8](#) and [fig. 20](#). [Tbl. 8](#) has two additional columns relative to the previous table, MAPE's Δ column is the difference between the MAPE values of the original model and the covariate model, and the P value is the result of the t

⁷ Due to the large number of LightGBM parameters and the wide range of adjustable parameters, not all configurations of optimal parameter combinations are listed in this study.

test of the MAPE values of the original model and the covariate model. These have already been described in the previous section.

We can see that LightGBM performs significantly better than the original model without covariates, except for the raw coal output time series, which fails the t test.

We can also see that LightGBM outperforms the baseline model across the board, with a significant improvement in performance. The LightGBM model with the addition of covariates also outperforms the ARIMA model, which is the best statistical model in our test, except for crude oil output and crude oil output. The best performing crude oil processing output has a MAPE value of 1.79, and the worst performing natural gas output has a MAPE value of 4.00.

It is interesting to note that in contrast to ARIMA's identification of raw coal output, where the initial forecast is set very low due to seasonality, the results are not as good as they could be. LightGBM clearly breaks this barrier by making predictions based on actual conditions and achieving a MAPE of 2.98.

Table 8: Performance of the LightGBM model on energy raw materials

Model	Variable	RMSE	MAE	MPE	MAPE	MAPE's Δ	P value
LightGBM	COO	55.67	50.26	1.71	3.04	-0.34	<0.05
	COPO	132.24	107.16	1.34	1.79	-1.02	<0.05
	GO	32.95	27.66	-0.08	2.08	-0.61	<0.05
	NGO	8.20	6.93	-0.15	4.00	-0.19	<0.05
	RCO	1187.58	954.16	-0.38	2.98	0.09	0.07

Figure 20: LightGBM model predictions on energy raw materials

Let us continue to analyse the performance of the LightGBM model on the electricity generation series, with the results shown in [tbl. 9](#) and [fig. 21](#). We can see that both the hydropower generation and solar power generation series fail the t test and do not outperform the original model. Although the performance of the model improves compared to the baseline model, only the thermal power generation and wind power generation series outperform the ARIMA model significantly, while the other models fail to outperform the ARIMA model. This confirms our previous discussion that there is no uniformity in the optimal model for different series.

Table 9: Performance of the LightGBM model on electricity generation

Model	Variable	RMSE	MAE	MPE	MAPE	MAPE's Δ	P value
LightGBM	HG	112.04	84.74	1.38	8.99	1.36	0.32
	NPG	35.93	28.52	8.15	8.15	-1.72	<0.05
	SPG	15.94	14.95	10.46	10.46	-0.11	0.06
	TPG	312.98	274.11	-0.64	5.78	-2.31	<0.05
	WPG	59.74	55.75	4.27	12.60	-3.24	<0.05

Figure 21: LightGBM model predictions on electricity generation

5.2.2 NNAR model results

The next model is the NNAR model, where we use the minimum AIC_c search method for optimal parameter selection. First, the energy raw material predictions are shown in [tbl. 10](#) and [fig. 22](#).

The performance of the NNAR model improved significantly with the addition of covariates. All series passed the t test successfully, except for the crude oil output series, which did not pass the t test. The most significant improvement in performance is in the crude oil processing output series, where the MAPE value decreases by 7.94.

However, the NNAR model did not perform well against the other models; although it outperformed the baseline model, only the gas output series had a lower MAPE value compared to the ARIMA model, and it also lagged behind the previous machine learning model, LightGBM.

Table 10: Performance of the NNAR model on energy raw materials

Model	Variable	RMSE	MAE	MPE	MAPE	MAPE's Δ	P value
$NNAR(1,1,3)_{12}$	COO	52.01	48.95	2.48	2.93	1.10	0.27
$NNAR(1,2,2)_{12}$	COPO	428.05	389.11	6.54	6.54	-7.94	<0.05
$NNAR(1,1,4)_{12}$	GO	55.19	49.67	-3.75	3.75	-3.23	<0.05
$NNAR(1,1,2)_{12}$	NGO	6.99	6.48	-1.69	3.82	-5.93	<0.05
$NNAR(2,3,4)_{12}$	RCO	2768.75	2205.38	0.79	7.03	-4.96	<0.05

Figure 22: NNAR model predictions on energy raw materials

We then analysed the performance of the models on power generation, and the results are shown in [tbl. 11](#) and [fig. 23](#).

The performance of this set of models shines through. Compared to the original model, all series pass the t test, except for thermal power generation. The performance of the thermal power generation and wind power generation series improves significantly compared to the ARIMA model, with the best performance thus far on both series, especially on the thermal power generation series where the model has a MAPE value of only 2.81.

Table 11: Performance of the NNAR model on electricity generation

Model	Variable	RMSE	MAE	MPE	MAPE	MAPE's Δ	P value
$NNAR(4,3,3)_{12}$	HG	115.13	97.49	-10.04	13.49	4.08	0.82
$NNAR(1,2,3)_{12}$	NPG	30.10	24.92	4.66	7.18	-5.81	<0.05
$NNAR(2,3,3)_{12}$	SPG	11.31	9.74	-3.54	6.83	-8.55	<0.05
$NNAR(1,1,3)_{12}$	TPG	152.72	136.01	-0.02	2.81	-12.75	<0.05
$NNAR(2,2,3)_{12}$	WPG	46.65	39.20	7.71	8.60	-10.37	<0.05

Figure 23: NNAR model predictions on electricity generation

5.2.3 LSTM model results

Let us analyse the last machine learning model, the LSTM. For the setup of the LSTM model, we start by setting the initial parameters in terms of epochs to 300, batch to 6, feed length to 4 and the optimiser to the Adam optimiser. Other parameters, such as hidden dim, number of neurons, learning rate, number of loop layers and dropout rate, were automatically tuned using the Keras built-in autoreference module. We use MAPE as an evaluation metric for the optimal parameters.

We first focus on the performance of the model on energy raw materials, as shown in [tbl. 12](#) and [fig. 24](#). First, the performance of the LSTM model improves significantly with the addition of covariates, except for the natural gas output series. The MAPE value is reduced by approximately 6.21 after adding the covariates to the crude oil processing output series.

In comparison with other models, however, the LSTM model outperforms the baseline model for only 3 sets of series, and this number drops to 1 when compared to the ARIMA model. Finally, the LSTM loses across the board when compared to the LightGBM model, which performs well for energy raw materials. These results seem to indicate that the LSTM model is not suitable for forecasting energy series with large time horizons.

Table 12: Performance of the LSTM model on energy raw materials

Model	Variable	RMSE	MAE	MPE	MAPE	MAPE's Δ	P value
LSTM	COO	91.75	72.28	-0.20	4.38	-3.19	<0.05
	COPO	269.00	226.14	-0.64	3.84	-6.21	<0.05
	GO	43.72	32.57	-0.88	2.45	-0.63	<0.05
	NGO	16.28	14.74	-3.32	8.58	0.23	0.12
	RCO	3284.57	2577.29	-1.51	7.98	-5.54	<0.05

Figure 24: LSTM model predictions on energy raw materials

We then look at the model performance on each series of generation groups, and the results are shown in [tbl. 13](#) and [fig. 25](#). Although the performance of all of the models with the inclusion of covariates passed the t test, the performance is not satisfactory when compared with other models. Although the performance is improved compared to the baseline model, it still does not stand out from the ARIMA and NNAR models that perform well in terms of power generation. Except for the wind power generation series, LSTM has the best performance with a MAPE value of 8.41.

Table 13: Performance of the LSTM model on electricity generation

Model	Variable	RMSE	MAE	MPE	MAPE	MAPE's Δ	P value
LSTM	HG	91.40	83.11	11.00	11.00	-2.32	<0.05
	NPG	53.09	38.36	-2.40	11.96	-4.24	<0.05
	SPG	14.87	12.59	0.07	9.06	-7.23	<0.05
	TPG	885.67	662.28	-3.74	13.89	-10.32	<0.05
	WPG	54.68	35.40	-5.81	8.41	-3.33	<0.05

Figure 25: LSTM model predictions on electricity generation

6. Discussion

After the detailed analysis of each series in the previous section, we selected the model with the best performance on each series. Since the magnitudes of the individual series are not uniform, we choose MAPE, a relative measure, as the criterion for judging the optimal model; the lower the MAPE value is, the better the model performance.

To check the robustness of the results, we first validate our model results using the 5-fold time-series cross-validation method (fig. 11). We set the initial series length to 10 and then add 6 new observations each time to form the cross-check dataset and obtain the MAPE values for each fold. We then used the bootstrap method to obtain the 95% confidence interval for the MAPE. The results are shown below in tbl. 14.

Table 14: Optimal model performance for each time series

Type	Variable	Model	MAPE	MAPE CI
Raw output	COO	$ARIMA(0,1,1)(0,1,0)_{12}$	1.66	[1.06,2.33]
	COPO	LightGBM	1.79	[1.38,1.85]
	GO	LightGBM	2.08	[1.97,2.64]
	NGO	$ARIMA(3,1,1)(0,1,0)_{12}$	2.76	[1.78,3.59]

Power generation	RCO	LightGBM	2.98	[2.33,3.28]
	HG	$ARIMA(2,0,1)(0,1,0)_{12}$	5.70	[4.52,10.8]
	NPG	$ARIMA(1,1,3)(0,1,0)_{12}$	6.70	[3.07,7.48]
	SPG	$NNAR(2,3,3)_{12}$	6.83	[7.05,7.05]
	TPG	$NNAR(1,1,3)_{12}$	2.81	[2.15,4.94]
	WPG	LSTM	8.41	[7.55,9.06]

In terms of optimal model selection, our study indicates that the series with the lowest MAPE value is the crude oil output series predicted by $ARIMA(0,1,1)(0,1,0)_{12}$, with a value of 1.66. The highest MAPE value is the wind power generation series predicted by the LSTM model, which is only approximately 8.41. All of the optimal models have MAPE values below 10 and outperform the benchmark model. The ability to predict the production of various energy sources with reasonable accuracy (“Forecasting FAQs” 2012; Kolassa and Schütz 2007) can provide useful insights for industrial production and government policy.

Our study notes that there is no model that performs optimally on all series and that each series has its own applicable model. In terms of the classification of models, only the ARIMA model among the traditional statistical models can be used as the optimal model in certain time series, accounting for 40% of the optimal models. Machine learning models, on the other hand, account for 60%. The time-series plots (fig. 2 and fig. 4) of the individual series indicate that the statistical model is dominated by stable series with significant seasonality, such as crude oil output and hydropower generation. The machine learning model, on the other hand, dominates the more volatile and less stable series, such as crude oil processing output, raw coal output and thermal power generation.

In addition, our study also found that the machine learning model can apparently better handle time series that were heavily affected by the COVID-19 epidemic, such as crude oil processing output and raw coal output. We can clearly see in the graph that COVID-19 (with the red line as the impact time point) impacted these energy series, causing a very significant drop in production. The machine learning model LightGBM handles this problem well, with MAPE values of 1.79 and 2.98 for these two series, compared to 2.54 and 6.86, respectively, for the best statistical model in the study, ARIMA.

The most obvious difference is in the thermal power generation series, where the combination of fig. 4 and fig. 5 shows that there was a period of low production in thermal power generation after the outbreak of COVID-19, which confused the otherwise seasonal time series and made it difficult for the statistical model to predict. The traditional ARIMA model has a MAPE of 10.94, while the machine learning NNAR model has a MAPE of 2.81. For the purposes of this paper, the machine learning model is better at predicting energy production timescales that are more affected by COVID-19. Traditional models are better at predicting less affected time series, such as hydropower generation and nuclear power generation.

Referring to the discussion in the *Data and data processing* section, we know that the time series that are significantly affected by COVID-19 are some of the more influential energy sources in China's energy system. Raw coal is by far the most important raw energy material in China, while thermal power generation is still the main form of power generation in China today. These series also have the common characteristic that their manpower requirements are large and vulnerable to lockdown policies; therefore, the impact of COVID-19 is also significant. The top performers on these series are the machine learning models (LightGBM and NNAR). The time series that are less affected by COVID-19, on the other hand, are many of the clean energy sources, which have less human demand and are also much more affected by seasonality and still represent a smaller proportion of China's energy system. The statistical model still has a greater advantage in these series.

From the previous analysis, we can also see that the performance of the machine learning models improved significantly with the inclusion of covariates. Out of the 30 models with covariates, 24 models passed the t test. In particular, 9 out of 10 LSTM models with covariates passed the t test. This is related to the multiple output mechanism of the LSTM model itself, which has been noted in the literature to be better at incorporating covariates (Sagheer and Kotb 2019), which is also confirmed by the research in this paper. In addition, other machine learning models incorporating covariates have shown significant performance improvements, such as the optimal thermal power generation prediction sequence in the study. The MAPE value of the NNAR model was reduced by 12.76 after the inclusion of covariates.

From these findings, our study further infers the following. When faced with energy production time series with few observations, the addition of other correlation series to the model can be effective in improving the accuracy of the predictions. These correlation series can be divided into homogeneous energy production and shock event correlation time series (the COVID-19 correlation series in this paper). The combination of the above two time series as covariates can improve the performance of the model to some extent.

7. Conclusions and perspectives

Forecasting energy production in China during COVID-19 is an exceptionally complex task. Energy production time series have few observations, time-series stability is generally low, and the impact of COVID-19 varies between time series. This poses new challenges for effective forecasting, which in turn has important implications for plant production and national policy.

This study reveals some patterns in energy forecasting for mainland China by analysing and forecasting the time series of two major types of energy production in China. The important findings in this paper are as follows:

1. This paper provides a comprehensive analysis and projection of two major groups and ten different energy production timescales for China. We present optimal models for each energy source by combining traditional statistical modelling frameworks and the latest machine learning modelling frameworks. All of these

- models outperformed the benchmark model, and all had MAPE values less than 10. These findings have implications for guiding industrial production and government energy policy at a macro level.
2. In this study, we used two main types of temporal models: traditional statistical models and machine learning models. We find that machine learning models are better able to handle energy data significantly affected by COVID-19, while statistical models still have their own advantages in stable time series. Overall, machine learning models outperform traditional models in predicting COVID-19. Nevertheless, our study notes that there are large differences between energy production processes, indicating that there is no universal optimal model. Some findings from this paper can be used to help us choose a model.
 3. We also innovate by incorporating data energy production time series and epidemic time series into the model and discuss the effect of covariates on the performance of each set of time-series models. We found that with the use of epidemic data and energy data as covariates, we were able to effectively improve the accuracy of the predictions in general. This advantage is difficult to achieve with traditional statistical models.
 4. We discuss the results to categorise two different categories of timings, those that are strongly influenced by COVID-19 and those that are less influenced. We found that the energy sources greatly affected by COVID-19 still occupy an important position in China's energy system and are often traditional energy sources with high labour input and high pollution, such as the coal processing industry and thermal power generation industry. The clean energy sources that are less affected by COVID-19 are more likely to be marginal in China's energy system, such as hydropower and wind power, and their strong resistance to the epidemic deserves more attention. In addition, the energy sources that are greatly affected by COVID-19 are more capable of being processed by machine learning models. Moreover, in regard to the energy sources with less impact, the traditional ARIMA model still has a greater advantage.
 5. Finally, this paper provides some plausible explanations for the global energy shortage crisis that currently exists. We have found that the effects of COVID-19 are long lasting and that the lockdown and shutdown measures taken at the outbreak sites still have an impact on some types of energy production. In addition, the high correlation between raw materials and electricity generation makes it possible for shutdowns in raw material production to be transmitted to electricity generation, resulting in shortages, such as the strong link found in this paper between coal output and thermal power generation.

In this paper, we have established a reproducible, scientific and efficient modelling workflow. This workflow standardises the time-series modelling process and can be drawn upon for subsequent forecasting studies and for forecasting models in industry.

We offer the following recommendations based on this study. First, many of the energy production time series during COVID-19 are not stable, and this is where a machine learning model may give better forecasts; however, we cannot ignore the excellent performance of traditional statistical models for some of the stable series. Second, there is no single, optimal model for predicting each energy time series during

COVID-19, and different models should be used depending on the characteristics of the time series.

In addition, COVID-19, as an important shock event, poses new challenges for accurate forecasting. The use of methods such as covariates to incorporate COVID-19 into the model should be considered when forecasting. Finally, effective forecasting and assessment of energy production is important for national economies and national life, and there is a need for plants and government statistics offices to use better models to forecast energy production time series.

However, this study also has the following problems. The large span of time series used in this paper and the small number of observations in each time series tend to overfit the predictions. Classical statistical models and innovative machine learning models are used in the study; however, some models that are also considered to be good predictors are not discussed, such as random forests and grey models. Finally, we do not collect more energy-related time series as covariates to assist in forecasting.

Further research could be conducted in future studies in the following areas. More observations and more segmented time series of energy production could be used. In addition, more of the currently popular time-series forecasting models for forecasting could be explored, such as the latest transformer model. Finally, more relevant time series could be collected to improve the accuracy of the forecasts, e.g., macroeconomic data for China during the pandemic.

References

- Abedin, M., Chi, G., Uddin, M., Satu, M., Khan, Md. I., & Hájek, P. (2020). Tax Default Prediction Using Feature Transformation-Based Machine Learning. *IEEE Access*, 1–1. <https://doi.org/10/gm3v8m>
- Abedin, M. Z., Chi, G., Colombage, S., & Moula, F.-E. (2018). *Credit Default Prediction Using a Support Vector Machine and a Probabilistic Neural Network* (SSRN Scholarly Paper No. ID 3175776). Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3175776>. Accessed 18 October 2021
- Abedin, M. Z., Guotai, C., Moula, F.-E., Azad, A. S., & Khan, M. S. U. (2019). Topological applications of multilayer perceptrons and support vector machines in financial decision support systems. *International Journal of Finance & Economics*, 24(1), 474–507.

- Abedin, M. Z., Hassan, M. K., Khan, I., & Julio, I. F. (2021). Feature Transformation for Corporate Tax Default Prediction: Application of Machine Learning Approaches. *Asia-Pacific Journal of Operational Research*, 2140017. <https://doi.org/10/gm9vsg>
- Agdas, D., & Barooah, P. (2020). Impact of the COVID-19 Pandemic on the U.S. Electricity Demand and Supply: An Early View From Data. *IEEE Access*, 8, 151523–151534. Presented at the IEEE Access. <https://doi.org/10/ghbqgn>
- Ahmar, A. S., & Del Val, E. B. (2020). SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. *Science of The Total Environment*, 729, 138883. <https://doi.org/10/ggvq6v>
- Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6), 594–621. <https://doi.org/10/bgrb7r>
- Ahmed, R. A., & Shabri, A. B. (2014). Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. *American Journal of Applied Sciences*, 11(3), 425. <https://doi.org/10/gm4mrm>
- Alhajeri, H. M., Almutairi, A., Alenezi, A., & Alshammari, F. (2020). Energy Demand in the State of Kuwait During the Covid-19 Pandemic: Technical, Economic, and Environmental Perspectives. *Energies*, 13(17), 4370. <https://doi.org/10/gm5ph7>
- Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning*, 36(1), 105–139. <https://doi.org/10/b334sz>

- Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1–4). Springer.
- Bhosale, S. T., Patil, M. T., & Patil, M. P. (2015). Sqlite: Light database system. *Int. J. Comput. Sci. Mob. Comput*, 44(4), 882–885.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370.
<https://doi.org/10/dxtfsf>
- Buehler, R., & Pucher, J. (2021). *COVID-19 Impacts on Cycling, 2019–2020*. Taylor & Francis.
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: an introduction with R*. Chapman and hall/CRC.
- Cheadle, C., Vawter, M. P., Freed, W. J., & Becker, K. G. (2003). Analysis of microarray data using Z score transformation. *The Journal of molecular diagnostics*, 5(2), 73–81. <https://doi.org/10/c7cfpf>
- Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864. <https://doi.org/10/ggvzsx>
- Chollet, F. & others. (2015). Keras. <https://keras.io>
- Chowdhury, E. K., & Abedin, M. Z. (2020). COVID-19 Effects on the US Stock Index Returns: An Event Study Approach. *SSRN Electronic Journal*.
<https://doi.org/10/gm5jfx>
- Clarke, B., & Henderson, L. D. (2002). *From energy to information: Representation in science and technology, art, and literature*. Stanford University Press.

Context: A world in lockdown – Global Energy Review 2020 – Analysis. (n.d.). *IEA*.

<https://www.iea.org/reports/global-energy-review-2020/context-a-world-in-lockdown>. Accessed 11 October 2021

COVID-19 lockdown in China. (2021, September 26). In *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=COVID-19_lockdown_in_China&oldid=1046587794. Accessed 5 October 2021

D'Alessandro, K., Chapman, A., & Dargusch, P. (2021). The Impact of COVID-19 on Electricity Demand and Generation in Japan. *Sustainability*, 13(19), 10836.

<https://doi.org/10/gm4ks4>

Dehghani, M., Saghafian, B., Rivaz, F., & Khodadadi, A. (2017). Evaluation of dynamic regression and artificial neural networks models for real-time hydrological drought forecasting. *Arabian Journal of Geosciences*, 10(12), 1–13.

<https://doi.org/10/gbn9dq>

Dil, S., Dil, N., & Maken, Z. H. (2020). COVID-19 trends and forecast in the Eastern Mediterranean Region with a Particular Focus on Pakistan. *Cureus*, 12(6).

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533–534.

<https://doi.org/10/ggnsjk>

Dong, G., & Pei, J. (2007). *Sequence data mining* (Vol. 33). Springer Science & Business Media.

Effect of Covid-19 on the power industry. (n.d.). *Joint credit*.

<https://baijiahao.baidu.com/s?id=1659406607679272936&wfr=spider&for=pc>. Accessed 11 October 2021

Electricity first, overcoming the difficulties together, promoting Covid-19 prevention and control and serving economic and social development in an integrated manner.

(n.d.). *State Grid Corporation of China*.

<http://www.sasac.gov.cn/n4470048/n13461446/n14398052/n16460319/n16460370/c16465329/content.html>. Accessed 11 October 2021

Esteves, G. R. T., Bastos, B. Q., Cyrino, F. L., Calili, R. F., & Souza, R. C. (2015). Long Term Electricity Forecast: A Systematic Review. *Procedia Computer Science*, 55, 549–558. <https://doi.org/10/gm5pwp>

Forecasting FAQs. (2012). In *The Business Forecasting Deal* (pp. 193–246). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119199885.app1>

Gargiulo, M., & Gallachóir, B. Ó. (2013). Long-term energy models: Principles, characteristics, focus, and limitations. *WIREs Energy and Environment*, 2(2), 158–177. <https://doi.org/10/ggfzwm>

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, 81, 709–727. <https://doi.org/10/gg49kg>

Guo, J. (2013). Backpropagation through time. *Unpubl. ms., Harbin Institute of Technology*, 40, 1–6.

Hadri, S., Najib, M., Bakhouya, M., Fakhri, Y., & El Arroussi, M. (2021). Performance Evaluation of Forecasting Strategies for Electricity Consumption in Buildings. *Energies*, 14(18), 5831. <https://doi.org/10/gm6g93>

- Hassan, M. K., Chowdhury, M. I. H., Balli, F., & Hasan, R. (2021). A note on COVID-19 instigated maximum drawdown in Islamic markets versus conventional counterparts. *Finance Research Letters*, 102426. <https://doi.org/10/gm9vtw>
- Hassan, M. K., Muneeza, A., & Sarea, A. M. (2021). *COVID-19 and Islamic Social Finance*. Routledge.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5–10. <https://doi.org/10/fg269b>
- James Fong, S., & Herrera Viedma, E. (2020). Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak.
- Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*, 67, 431–438. <https://doi.org/10/gm47d9>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 3146–3154.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kolassa, S., & Schütz, W. (2007). Advantages of the MAD/MEAN ratio over the MAPE. *Foresight: The International Journal of Applied Forecasting*, (6), 40–43.

- Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1), 841–851. <https://doi.org/10/gfb7pv>
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3–24.
- Kuzemko, C., Bradshaw, M., Bridge, G., Goldthau, A., Jewell, J., Overland, I., et al. (2020). Covid-19 and the politics of sustainable energy transitions. *Energy Research & Social Science*, 68, 101685. <https://doi.org/10/ghpw3z>
- Lam, P.-L., & Shiu, A. (2004). Efficiency and productivity of China's thermal power generation. *Review of Industrial Organization*, 24(1), 73–93. <https://doi.org/10/bjqbz>
- Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems*. <https://doi.org/10.1142/S0129065721300011>
- Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting*, 5(2), 231–240.
- Liashchynskyi, P., & Liashchynskyi, P. (2019). Grid search, random search, genetic algorithm: a big comparison for NAS. *arXiv preprint arXiv:1912.06059*.
- Liu, B., Fu, C., Bielefield, A., & Liu, Y. Q. (2017). Forecasting of Chinese primary energy consumption in 2021 with GRU artificial neural

- networkliuForecastingChinesePrimary2017. *Energies*, 10(10), 1453.
<https://doi.org/10/gck38h>
- Liu, H., Tian, H.-Q., Chen, C., & Li, Y. (2010). A hybrid statistical method to predict wind speed and wind power. *Renewable Energy*, 35(8), 1857–1861.
<https://doi.org/10/fq42cj>
- Liu, L., Zong, H., Zhao, E., Chen, C., & Wang, J. (2014). Can China realize its carbon emission reduction goal in 2020: From the perspective of thermal power development. *Applied Energy*, 124, 199–212. <https://doi.org/10/f5368t>
- Liu, X., Moreno, B., & García, A. S. (2016). A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors. *Energy*, 115, 1042–1054. <https://doi.org/10/f9fmfg>
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10/ggjt4>
- Maleki, A., Nasser, S., Aminabad, M. S., & Hadi, M. (2018). Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics. *KSCE Journal of Civil Engineering*, 22(9), 3233–3245. <https://doi.org/10/gm5qvg>
- Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K.-H. (2020). Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel medicine and infectious disease*, 37, 101742. <https://doi.org/10/ggxfsd>
- Migon, H. S., & Alves, L. C. (2013). Multivariate dynamic regression: modeling and forecasting for intraday electricity load. *Applied Stochastic Models in Business and Industry*, 29(6), 579–598. <https://doi.org/10/gm3pqx>

- Moftakhar, L., Mozhgan, S., & Safe, M. S. (2020). Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iranian Journal of Public Health*, 49(Supple 1), 92–100.
- Myers, D., & McGuffee, J. W. (2015). Choosing scrapy. *Journal of Computing Sciences in Colleges*, 31(1), 83–89.
- Narayan, P. K. (2020). Oil price news and COVID-19—Is there any connection? *Energy Research Letters*, 1(1), 13176.
- Nelson, B. K. (1998). Time series analysis using autoregressive integrated moving average (ARIMA) models. *Academic emergency medicine*, 5(7), 739–744.
<https://doi.org/10/c6fzfs>
- O'Hara-Wild, M., Hyndman, R., & Wang, E. (2021). *fable: Forecasting models for tidy time series* (manual). <https://CRAN.R-project.org/package=fable>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. *Acta numerica*, 8, 143–195. <https://doi.org/10/bznh56>
- Rahimi, I., Chen, F., & Gandomi, A. H. (2021). A review on COVID-19 forecasting models. *Neural Computing and Applications*. <https://doi.org/10/gmgtg7>
- Reno, C., Lenzi, J., Navarra, A., Barelli, E., Gori, D., Lanza, A., et al. (2020). Forecasting COVID-19-associated hospitalizations under different levels of social distancing in Lombardy and Emilia-Romagna, Northern Italy: results from an extended SEIR

- compartmental model. *Journal of clinical medicine*, 9(5), 1492.
<https://doi.org/10/ggxfwt>
- Rout, U. K., Voß, A., Singh, A., Fahl, U., Blesl, M., & Gallachóir, B. P. Ó. (2011).
Energy and emissions forecast of China over a long-time horizon. *Energy*, 36(1),
1–11. <https://doi.org/10/btgtwq>
- Saboia, J. L. M. (1977). Autoregressive integrated moving average (ARIMA) models for
birth forecasting. *Journal of the American Statistical Association*, 72(358), 264–
270. <https://doi.org/10/gm3rb6>
- Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using
deep LSTM recurrent networks. *Neurocomputing*, 323, 203–213.
<https://doi.org/10/gjs26h>
- Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting—A
review. *Renewable and sustainable energy reviews*, 16(2), 1223–1240.
<https://doi.org/10/dk6csp>
- Taieb, S. B., & Hyndman, R. J. (2012). Recursive and direct multi-step forecasting: the
best of both worlds, 36.
- Tamang, S. K., Singh, P. D., & Datta, B. (2020). Forecasting of Covid-19 cases based on
prediction using artificial neural network curve fitting technique. *Global Journal
of Environmental Science and Management*, 6(Special Issue (Covid-19)), 53–64.
- Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time
series forecasting. *omega*, 29(4), 309–317. <https://doi.org/10/d97863>
- Toda, H. (1991). *Vector autoregression and causality*. Yale University.

Tu, H., Lukic, S., Dubey, A., & Karsai, G. (2020). An LSTM-Based Online Prediction Method for Building Electric Load During COVID-19. In *Annual Conference of the PHM Society* (Vol. 12, pp. 8–8).

‘Unprecedented’ power cuts in China hits homes, factories. (n.d.).

<https://www.aljazeera.com/news/2021/9/28/chinas-northeast-suffers-power-crunch>. Accessed 5 October 2021

Wang, B., Yang, Z., Xuan, J., & Jiao, K. (2020). Crises and opportunities in terms of energy and AI technologies during the COVID-19 pandemic. *Energy and AI*, 1, 100013. <https://doi.org/10/ghrhsb>

Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis—Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, 93, 41–56. <https://doi.org/10/f758bh>

Weiss, H. H. (2013). The SIR model and the foundations of public health. *Materials mathematics*, 0001–0017.

Werth, A., Gravino, P., & Prevedello, G. (2021). Impact analysis of COVID-19 responses on energy grid dynamics in Europe. *Applied Energy*, 281, 116045. <https://doi.org/10/gjqp4n>

Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling. *Transportation Research Record*, 1776(1), 194–200. <https://doi.org/10/d4cm84>

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages.

Management science, 6(3), 324–342. <https://doi.org/10/ckvdfv>

Wuhan lockdown “unprecedented”, shows commitment to contain virus: WHO

representative in China. (2020, January 23). *Reuters*.

<https://www.reuters.com/article/us-china-health-who-idUSKBN1ZM1G9>.

Accessed 5 October 2021

Xu, J., Lu, Z., & Xie, Y. (2021). Loan default prediction of Chinese P2P market: a

machine learning methodology. *Scientific Reports*, 11, 18759.

<https://doi.org/10/gmwnfz>

Yang, G., Wang, Y., & Li, X. (2020). Prediction of the NO_x emissions from thermal

power plant using long-short term memory neural network. *Energy*, 192, 116597.

<https://doi.org/10/gm5pqm>

Yuan, J. (2018). The future of coal in China. *Resources, Conservation and Recycling*,

129, 290–292. <https://doi.org/10/gmz6zr>

Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). Comparison of machine learning algorithms

for predicting crime hotspots. *IEEE Access*, 8, 181302–181310.

<https://doi.org/10/gm4mtf>

Zhao, C., Tepekule, B., Criscuolo, N. G., Wendel-Garcia, P. D., Hilty, M. P., Fumeaux,

T., & Van Boeckel, T. P. (2020). icumonitoring. ch: a platform for short-term

forecasting of intensive care unit occupancy during the COVID-19 epidemic in

Switzerland. *Swiss medical weekly*, 150, w20277.

Zhao, Y., Hao, L.-S., & Wan, L. (2007). Research on the spatial structure of crude oil flow and the characteristics of its flow field in China. *Energy Policy*, 35(10), 5035–5050. <https://doi.org/10/dmwc33>

Zou, C., Zhao, Q., Chen, J., Li, J., Yang, Z., Sun, Q., et al. (2018). Natural gas in China: Development trend and strategic forecast. *Natural Gas Industry B*, 5(4), 380–390. <https://doi.org/10/ghw4ct>