

Poster Abstract: Time series forecasting of European Energy Market

HUN RIM and JURAJ KARDOS*, Università della Svizzera italiana, Switzerland

Due to climate change, the EU is transitioning towards renewable energy, aiming to increase its share in gross energy consumption to 42.5% by 2030 from the current 23%. However, the intermittent and seasonal nature of renewable energy sources presents challenges in predicting their production capacity. In this dynamic and ever-evolving landscape of the energy market, the ability to accurately forecast the evolution of renewable energy's stake in the market is a critical component in the decision making process of policy makers, and market participants alike. This project aims to explore and evaluate the performance of well-established forecasting methods in anticipating the trends of individual renewable energy components, ultimately contributing to the fostering of a balanced, sustainable, and reliable energy market in the EU.

The primary focus of our framework is to assess auto-regressive forecasting methods, which leverage chronological data intervals to capture temporal dependencies from market history, thus ensuring accurate forecasting. However, instead of evaluating the model purely reliant on auto-regression, we examine more intricate models which extend to incorporate moving-average, exploit seasonality of time series data, or those utilising the correlation with exogenous variables. Furthermore, To get a better understanding of not only accuracy, but precision of the models, short-term forecasts are made iteratively to provide a bigger pool of model performance samples. All model's performance in forecasting the individual energy components are evaluated by measuring and visualising the Root Mean Square Percentage Error between forecast result and the test data.

Additional Key Words and Phrases: Time series, Forecasting, Energy Market, Predictive analytics, ARIMA

ACM Reference Format:

Hun Rim and Juraj Kardos. 2024. Poster Abstract: Time series forecasting of European Energy Market . 1, 1 (April 2024), 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Like a staple good, we have steady demand for energy in our daily lives unlike luxury goods. When visualized, monthly energy demand reveals distinct patterns, with cyclical peaks occurring during winter and lows in summer, notably in July and August. These observations highlight Europe's heightened energy consumption in colder seasons and highlight the data's **seasonality**, suggesting the possibility of translating these patterns into mathematical models for forecasting future values.

Authors' address: Hun Rim, rimh@usi.ch; Juraj Kardos, xxx@usi.ch, Università della Svizzera italiana, Via Giuseppe Buffi 13, Lugano, Ticino, Switzerland, 6900.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Association for Computing Machinery.

XXXX-XXXX/2024/4-ART \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1.1 ARIMA

ARIMA model is one of the most widely used statistical forecasting model, and it is the center of our evaluation. It is a generalized version of AutoRegressive Moving-Average (ARMA) model and difference between the two models is their ability to handle non-stationary data. Stationarity indicates constancy of statistical property of a dataset such as variance, or mean over time. When data is stationary, it guarantees the stability of underlying patterns in the dataset, and eradicates the concern of the pattern being a result of random fluctuation. ARIMA model is capable of adjusting the data to provide this stability unlike ARMA.

Autoregression is a technique used in time series analysis that assumes temporal trend between the values in the dataset, and uncover a function of order p . In the auto-regression model, the variable regresses against itself. Put simply, the current value is calculated using past values, by estimating the influence of these past values ($\phi_1, \phi_2, \dots, \phi_p$) on the present value. However, even if we assume the general trend is captured perfectly, there are inconsistent white-noise in the data which are not replicable. Hence, ARIMA model improves the accuracy by replicating these small bumps in the data through factoring in the anomalies in the recent data. All these factors are combined to formularize ARIMA model as the following:

$$\hat{L}_N^{(d)} = C + \sum_{i=1}^p \psi_i L_{N-i}^{(d)} + \sum_{j=1}^q \varphi_j \epsilon_{N-j} + \epsilon_N$$

1.2 Seasonality

As mentioned before, there is a clear seasonality in EU energy market and it would be preferable to incorporate this. There is a ARIMA's variation called S(easonal)ARIMA which extends to look into same interval in previous seasons. When the duration of the season is given, the SARIMA model breaks the data down into seasons and study recurring patterns in those segments. Hence, the estimation of the current value in SARIMA model is carried out as the following:

$$SARIMA = \hat{L}_N^{(d)} + \sum_{i=1}^P \Phi_i L_{N-im}^D + \sum_{j=1}^Q \Theta_j \epsilon_{N-jm}$$

1.3 Deep-Learning

Along with predictive statistics, deep learning is frequently used in forecasting. Especially, deep-learning using Recurrent-Neural-Network (RNN) is very popular with time series forecasting. As the name suggests, RNN has a property of repeating the process of aggregation and activation unlike normal neural networks. This regression like property of RNN makes it very effective when building a forecasting model, and it would be interesting to evaluate it's performance in comparison to conventional statistical methods.

Hence, RNN forecasting method is also in progress of evaluation.

2 METHOD & MATERIAL

2.1 Procedure

- (p, d, q) order calculation.

There are 2 common methods in calculating the (p, d, q) values. First method is calculating the d value through differencing and *Augmented Dickey-Fuller test*, then finding the order of autoregression (p) through studying temporal correlation using ACF and PACF. Another method is implementing a grid search algorithm which finds the minimum Akaike Information Criterion (AIC) value by trying different combinations of (p, d, q) within a given range. Naturally, first method is more efficient whereas the second method is computationally inefficient but simple to implement. However, as computational inefficiency problem is becoming negligent due to the significant enhancement of computational power, we decided to utilize *auto_arima()* function of **Python**. Which is a built-in grid search algorithm from *pmdarima* library.

- Train Test Split

To train and measure the performance of the forecasting model, we split our data into two parts, training set and test set. We train the model using 60% of our data and assign the remaining 40% as test data set to compare it with the forecast.

- Training & Forecasting

For training of the model, we use Python functions from *statsmodels* library. The calculated order and seasonal order are passed to the trainer functions along with the train data set. Each model has respective trainer function provided from the *statsmodels* library (*ARIMA()*, *SARIMAX()*) and are configured accordingly with required training data, order of auto-regression, differencing, and moving-average. Finally, the model is fitted to the passed training data.

Then the trained model is used to make a forecast using the model class's *Predict()* function from the starting month to the ending month of our designation. Which in our case has an interval of 3 months.

- Rolling Horizon Approach

The forecast relies on historical values for estimation, which limits the model's ability to capture incoming trends, resulting in decrease of accuracy as the forecast window increases. To avoid this issue, forecasts are made at quarterly intervals. This approach is particularly necessary because the ARIMA model cannot incorporate new forecast errors into its formula, creating a need for iterative recalculation when the training set is updated. Therefore, we adopt the rolling horizon approach, where after each forecast, the training dataset is expanded by one month, and the model is re-trained before forecasting the future values of subsequent three months. By implementing the rolling horizon approach, the model

becomes capable of incorporating real-time updates, mimics continuous learning, and avoids forecasting biases.

- Analysis

Measuring the performance of forecasting method requires specific metric as our interest is in the relative error between predicted and actual values. Hence, we are using Root-Mean-Squared-Percentage-Error (RMSPE) as the metric for the performance measurement. Lower values in RMSPE indicate better performance. RMSPE measures the average percentage deviation between predicted and actual values, providing a standardized measure of prediction accuracy that is independent of the scale of the data. The data used for forecasting is not normalized and individual energy components have radically different scale. Hence, the RMSPE value provides better insight in evaluating forecasting models. RMSPE value can be derived as the following:

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \hat{Y}_i}{Y_i} \right)^2} \times 100$$

- Exogenous Variable

The EU power sector, just like any other sector, experiences volatility due to exogenous factors such as the climate or the economy. Hence, the framework evaluates SARIMAX (SARIMA + eXogenous variables) which is capable of accounting for fluctuations due to the exogenous factors. In this model, the exogenous variable which has influence on the power sector is collected as indicators. Every data point in time series dataset needs corresponding exogenous value, and it is passed to SARIMAX function. When forecasting the market movements, the model can reflect on the previous values with similar indicator behaviors from the exogenous variable list to account for the external factors, providing a more accurate forecast.

2.2 Data

The dataset consists of monthly energy components contributing to the energy balance on the supply and demand side as time series data. The data are published by the national authorities of European countries (usually the transmission system operators or the statistical office) and are collected on the level of the high-voltage transmission grid. The focus of this study is put on the key energy components or the components with the high level of variability, particularly electricity demand, and renewable energy generation such as wind, PV or hydro. Electricity demand is crucial for understanding how much power the system needs to meet consumer needs. Demand can vary significantly based on seasonal factors (heating/cooling), economic activity, and population growth. The increasing penetration of renewables significantly impacts supply dynamics. Understanding the variability of wind and solar generation, as well as the dispatchability (controllable generation) of hydro, is critical for managing grid integration challenges and optimizing

renewable energy utilization. Data from the January of 2016 onwards is chosen as this period coincides with a significant ramp-up of renewable energy deployment across Europe. Analyzing data from this period allows us to track the evolving energy mix and its impact on power market dynamics.

3 RESULT & DISCUSSION

3.1 Result Table

Date	Mean	SARIMA	ARIMA	S-Order	A-Order
2023-03-01	35980	8.20	93.98	(1,0,0)X(2,0,2,12)	(5,0,0)
2023-04-01	32624	4.73	2.83	(1,0,1)X(5,0,0,12)	(5,0,0)
2023-05-01	31530	1.96	82.48	(3,0,0)X(0,0,2,12)	(5,0,0)
2023-06-01	31202	9.86	48.31	(1,0,0)X(2,0,2,12)	(5,0,0)
2023-07-01	31312	13.93	1.08	(1,0,0)X(2,0,1,12)	(5,0,0)
2023-08-01	31660	11.80	2.83	(1,0,1)X(2,0,1,12)	(5,0,0)

Fig. 1. Forecast result of power demand in France

The Figure (1) is a showcase of the framework evaluating the performance of forecast models using rolling horizon approach through RMSPE metrics, where SARIMA and ARIMA column contains the RMSPE of the forecast against the test value. The table indicates SARIMA model's RMSPE is evenly distributed between approximately 2 to 14, and ARIMA model's RMSPE ranges from approximately 1 to 94 with higher concentration on the lower end (below 3). As mentioned before, lower RMSPE indicates higher performance of the model. Which means SARIMA model in general performs better than ARIMA model, and is more stable in terms of both accuracy and precision. Furthermore, SARIMA-Order (S-Order) seems to be constantly updated, whereas ARIMA-Order (A-order) remains constant. This could be interpreted as SARIMA model better reflects the change in trend and seasonality with the consideration of *seasonality* factor, in comparison to the ARIMA model.

3.2 Visualization

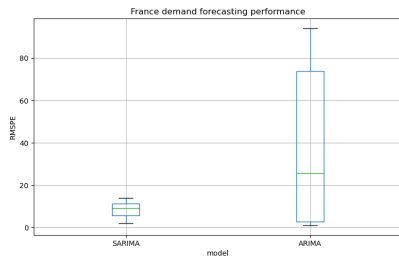


Fig. 2. France energy demand forecast performance visualization

Figure (2) above is the visualization of the SARIMA and ARIMA's performance through box-plotting the RMSPE values. This visualization is particularly useful when we are viewing larger scale of performance data from different energy sources and different

countries as we don't have to individually read all tables. It displays how accurate the majority of the forecast's are while denoting the spread of the RMSPE values, and the anomalies in the data that are far off from general performance. Hence, enabling simple visual analysis of the result while making the anomalies easier to spot. The box-plots indicate SARIMA model was approximately 15% off at worst case, and almost completely accurate in the best cases with median of approximately below 10%. Whereas ARIMA model has similar result to SARIMA model in case of best case scenario, while its median and worst case is far worse than that of the SARIMA. It is quite obvious that forecast error of this scale is not originating from residual factors, and it is more likely to be due to the trend.

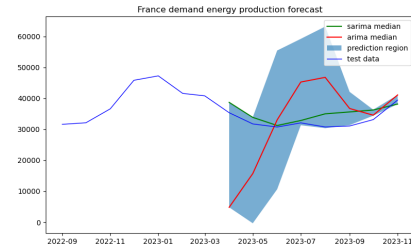


Fig. 3. France energy demand forecast

Figure (3) displays the forecast made from SARIMA and ARIMA model against the actual values. The shaded region is calculated from finding the maximum and the minimum of a merged forecast results. The green function is generated from the median value of the SARIMA forecasts, and the red function is the function generated from the median value of the ARIMA forecasts. As shown on demand forecast, ARIMA model performs quite poorly in most of the time excluding the last 2 months while SARIMA model performs consistently well over time. However, this is quite a rare case.

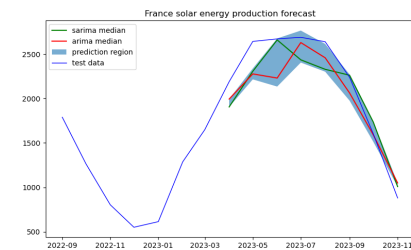


Fig. 4. France solar energy production forecast

Figure (4) visualizes the forecasts of solar energy production in France against the real values. From the visual analysis, both models perform quite well with ARIMA model showcasing slightly better performance in some cases, and most of the forecasting results exhibit similar pattern.

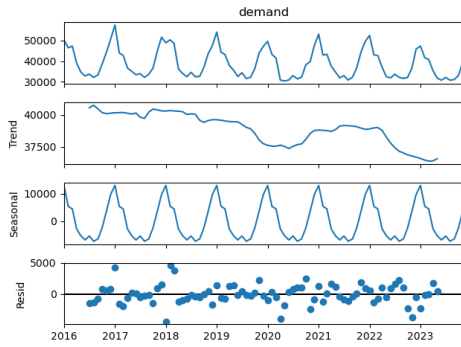


Fig. 5. Seasonal decomposition of energy demand in France

3.3 Discussion

In most cases, both ARIMA, and SARIMA model showcased quite accurate and precise forecasts with the median RMSPE being below 20%, excluding few exceptions of ARIMA forecasts. Naturally, one would wonder why figure (3) would exhibit such an abnormal behavior. If we model the trend of those abnormal individual energy components, we can understand why. As it can be seen from figure (5), the trend is inconsistent, especially at the end where our forecast is made, and as auto-regression is heavily effected by the trend, it is understandable ARIMA performs poorly while the SARIMA model avoids this issue through incorporating seasonality.

4 CONCLUSION

In conclusion, this paper has conducted an investigation into the forecasting performance of well-known methods such as ARIMA, SARIMA and more, providing insights into their strengths and limitations in forecasting demand, and production capacity of renewable energies. Through thorough analysis, we have evaluated the effectiveness of these traditional forecasting approaches in evolving European energy market.

Our findings demonstrate that while traditional methods such as ARIMA and SARIMA have shown promise in modeling and forecasting renewable energy production, they may encounter challenges in accurately capturing rapidly shifting trend, seasonality, and residual factors of evolving renewable energy sources. Hence, depending on the properties of the energy components, certain models will outperform the others. These challenges underscore the need for tailor-made forecasting methodologies that can better adapt to the complex dynamics of renewable energy production while using the traditional models as the base.

Moving forward, future research endeavors should focus on exploring alternative forecasting techniques, such as hybrid models, that better incorporate factors such as weather or economic indicators, technological advancements, and policy changes. By embracing the diversity of forecasting methodologies and utilising the advancements in data science, we can enhance the accuracy and reliability

of renewable energy production forecasts, thus facilitating the transition towards a sustainable and resilient energy future in Europe.

REFERENCES