

# Investigation of forecasting methods for the hourly spot price of the Day-Ahead Electric Power Markets

Radhakrishnan Angamuthu Chinnathambi  
Department of Electrical Engineering  
University of North Dakota  
Grand forks, USA  
radhakrishnan.angamu@und.edu

Prakash Ranganathan  
Department of Electrical Engineering  
University of North Dakota  
Grand forks, USA  
prakash.ranganathan@engr.und.edu

**Abstract**— Forecasting hourly spot prices for real-time electricity usage is a challenging task. This paper investigates a series of forecasting methods to 90 and 180 days of load data collection acquired from the Iberian Electricity Market (MIBEL). This dataset was used to train and test multiple forecast models. The Mean Absolute Percentage Error (MAPE) for the proposed Hybrid combination of Auto Regressive Integrated Moving Average (ARIMA) and Generalized Linear Model (GLM) was compared against ARIMA, GLM, Random forest (RF) and Support Vector Machines (SVM) methods. The results indicate significant improvement in MAPE and correlation co-efficient values for the proposed hybrid ARIMA-GLM method.

**Keywords**—ARIMA-GLM, Hybrid model, Generalized Linear Model; Electricity Price Forecasting; Random Forest, Support Vector Machine, Iberian Market, Day-ahead Price.

## I. INTRODUCTION

The electricity markets are becoming sophisticated because of the recent changes in the trading structure for market bids on prices. These market usually include two instruments for trading: the pool and bilateral contract [1]. In the pool, both the consumers and producers submit bids which gets cleared by the market operator. These operators announce the prices for the next day. The companies might also want to use bilateral contacts for hedge against the risk of price volatility.

For both these instruments, price forecasting for the next day or next few months is vital for adjusting their bids to maximize the profit or for schedule outage, design load response and various decision making process. The market clearing prices are publicly available for all the electricity market as it is the case of the day-ahead pool of mainland Spain ([www.omel.es](http://www.omel.es)), the Californian pool ([www.caiso.com](http://www.caiso.com)), or the Australian national electricity market ([www.aemo.com.au](http://www.aemo.com.au))

Therefore, an accurate price forecast will greatly help the consumers or producers in the bidding strategies and also in the price negotiation of the bilateral contract. This paper focuses on the day-ahead price forecast of a daily electricity market using various statistical and computational intelligence models. This paper provides models to forecast

the next 24-hour market clearing prices of next day. These models provide reliable estimates of forecasts of prices in the Iberian electricity market of mainland Spain and Portugal.

The remainder of the paper is organized as follows: section 2 discusses spot electricity market representation, section 3 presents literature-review of the existing electricity price modelling approaches, section 4 presents steps involved in modelling the electricity price, section 5-10 presents techniques used in this study, section 11 details input variables considered for forecasting, section 12 presents forecasting results and discussion, and section 13 ends with conclusion.

## II. THE ELECTRICITY MARKET REPRESENTATION

The electricity spot market is a day-ahead market in which the prices for the next day is finalized before a particular market closing time. This is different from the commodity or financial markets which allows continuous trading [2]. The system operators require advanced notice to check whether the schedule falls under transmission constraints. The agents usually submit their bids for each hour of the next day in the day-ahead electricity market before a particular market closing time. We define average of the 24 hourly prices as daily spot price or the base-load price. The average for the on-peak hourly prices typically range from 8am to 8pm is called as the peak load price.

## III. LITERATURE REVIEW

An excellent state-of-the-art review on electricity price forecasting that include methods can be found in Ref. [2]. This paper discusses various modelling approaches such as concepts from multi-agent theory, reduced-form, statistical and computational intelligence. Weron in [2] discusses the strengths and weakness of the existing forecast methods and enforces the need for a robust error evaluation procedure. For smart grid applications, deployment of Auto Regressive Integrated Moving Average (ARIMA) techniques has been used for load forecasting and has some effectiveness considering the seasonality on weather, and also used in predicting the short term electricity price [3–9]. Datasets

that have used ARIMA are for Spanish, Californian and EPEX power Markets [1].

In [10], an ensemble learning method known as Random Forest (RF) has been applied to predict next day price for New York electricity market. In [11], a Support Vector Machine (SVM) method has been applied in Australian Market. This technique was used as a hybrid model along with ARMAX and Least Square [12-14].

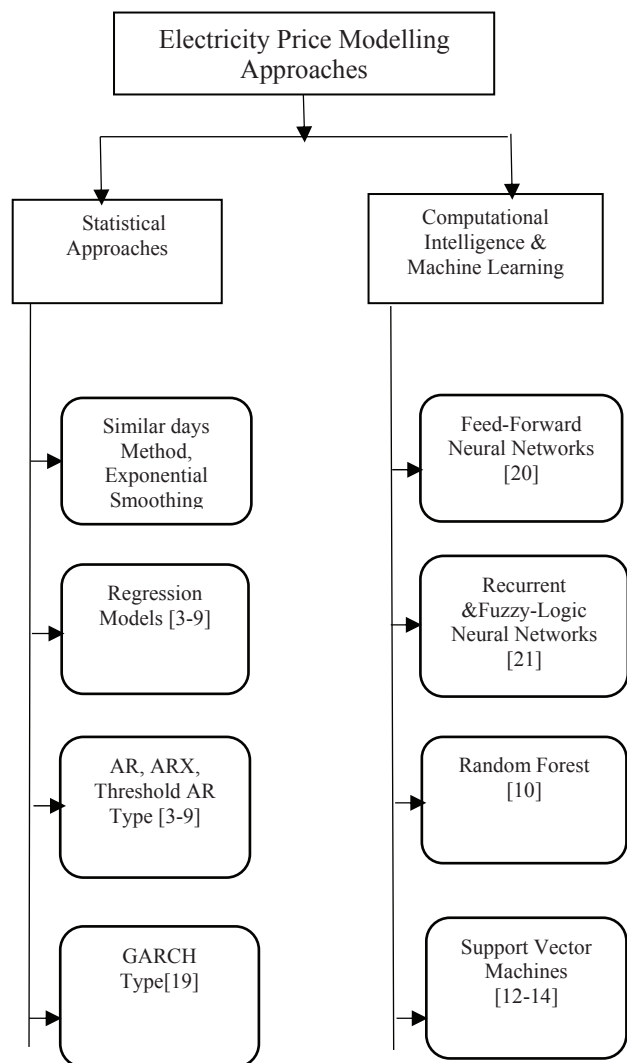


Fig.1. Taxonomy of Electricity Price Modelling Approach  
IV. MODELLING OF ELECTRICITY PRICE.

The Electricity price forecast model involves data cleansing, data preparation, and data evaluation. Complete steps involved are as follows.

*Step 1-* (Gather Load data). We collected the load consumption data through the web-link provided by the Iberian Electricity market. We had collected all the data in to a single file for easy computation of the price.

*Step 2* – Glean and order the data

We used the three & six months of dataset to build the model and evaluate how well the model generalizes to future results.

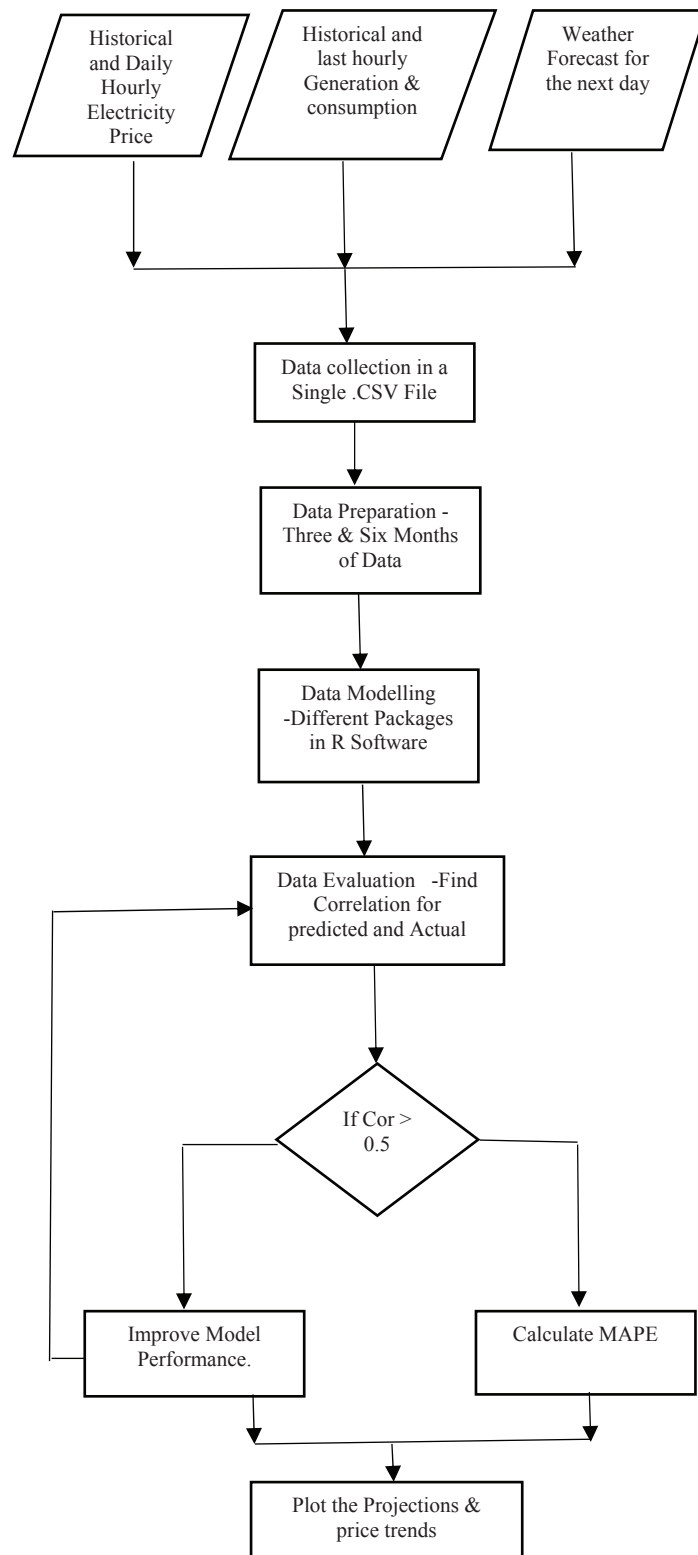


Fig.2. Flowchart showing modelling process of Electricity Price Forecasting

### Step 3 – Training a model on the data

To model the relationship between the predictor variables used in modelling and the electricity price, we used several Statistical and Machine Learning packages in Open source R software which provides a standard and easy-to-use implementation of such models.

In this study, we have used four techniques. The following R packages were used in implementing the model. Auto-arma' function in forecast package in R helps us identify the best fit ARIMA model. 'Ksvm' function in 'kernlab' package in R helps us to fit the SVM model. 'randomForest' function in randomForest package in R helps us to fit the RF model. 'glm' function under gaussian family in 'stats' package in R helps us to fit the GLM model.

### Step 4 – Evaluating model performance

We must measure the correlation between our predicted electricity price and the true value. This correlation values helps us in evaluating the model and helps us to find the direct relationship between the two variables.

Correlations close to 1 indicate strong linear relationships between actual and forecasted price. Therefore, the correlation of more than 0.9 for 14 variables for two different dataset shown in table 2 indicates a fairly strong relationship. This values show that our model is fitting well to predict the future values.

## V. AUTO REGRESSIVE INTEGRATED MOVING AVERAGE METHODOLGY (ARIMA)

ARIMA method is a stochastic process used to analyze time series data. ARIMA is a mixture of three time series components i.e. AR (Autoregressive), I (Integrated), and MA (Moving Average). A convenient notation for ARIMA model is ARIMA (p,d,q) [15] Here p, d, and q represents AR, I, and MA components. Each of these components is used to reduce the final residuals display white noise or no residuals at all.

### A. 1ST STEP OF ARIMA TO EXTRACT INFORMATION

*Integrated (I)* – subtracting the data from the previous or lagged one to extract trends to make it stationery [15]. This step is basically used to extract trend from the original time series data. Differencing is one of the most popularly used method for extraction of trends. Here, the original series is subtracted with its lagged or previous.

The residues of most time series data becomes trendless after differencing for the first time which is represented as ARIMA (0, 1, 0). If the time series data has trends still, it is further differenced to remove the trend which is denoted as ARIMA (0, 2, 0). This is called 2nd order differencing. This trend-less series is called as Stationery on mean series. This shows that the mean does not change over time.

### B. 2ND STEP OF ARIMA TO EXTRACT INFORMATION

Auto Regressive (AR) – uses the previous value influence on the current value.

After we difference the data to make it stationery, then the AR component of the ARIMA starts. As we mentioned earlier, it takes the previous value influence on the current values. This is done through obtaining a simple multiple linear regression .model with the independent or predictor variables as time lagged values. The general notation of the equation for this multiple linear regression is shown below. Here  $Y_t$  presents the price at time 't',  $\phi_1$  denotes regression co-efficient and  $e_t$  denotes error term.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots \phi_p Y_{t-p} + e_t \quad (1)$$

AR model of order 1 i.e. p=1 or ARIMA (1, 0, 0) is denoted with the following regression equation

$$Y_t = c + \phi_1 Y_{t-1} + e_t \quad (2)$$

### C. 3RD STEP OF ARIMA TO EXTRACT ERROR TERMS.

Moving Average (MA) – uses the previous value error term influence on the current value error.

After we take the auto-regression is performed, here we form the relationship between the error term of the previous and current values as shown in the equation (3). This component of the ARIMA, is formed with the simple multiple linear regression with the predictor variables as lagged error terms.

$$Y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots \theta_q e_{t-q} \quad (3)$$

## VI. ARIMA IMPLEMENTATION

### Step 1: Identification of best fit ARIMA model

Auto-arma function in R under the forecast package help us in finding the best fit for the ARIMA model. It gives the best fit by giving the value for the three components (p,d,q) which we can use it for prediction. The best fit model obtained from the Auto-arma function is based on the lowest values of the Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC).

### Step 2: Forecast using the best fit ARIMA model

The next day hourly spot prices are forecasted using the Function Forecast in R. After finding the ARIMA model from the auto.arma function, arima function is used to predict the price using the given set of variables.

## VII .SUPPORT VECTOR MACHINES (SVM)

Support Vector Machines (Support Vector network) are supervised learning model that analyses data for regression analysis in this work. SVM is assumed to be a surface that has a boundary between numerous points in a data that represents example plotted in a multidimensional space [16]. The main aim of the SVM is to create a flat boundary that leads to the equal partition of data on both sides. This boundary helps in creating SVM to model complex relationship.

The mathematics behind the SVM has been there for long time but it has become extremely popular due to the availability of these algorithm in various software. These algorithms were well supported in open source software like R that is implemented in libraries. Availability of these packages in open source software has increased the usage of these algorithm which is otherwise quite complex to implement. SVM can be used for both Classification and Prediction In this case we use these algorithm for predicting the prices.

### VIII. RANDOM FOREST METHOD

Ensemble-based method called random forests (or decision tree forests) emphasis only on ensembles of decision trees [16]. This method combines the base principles of bagging with random feature selection to add additional variation to the decision tree models. The model uses a vote to combines the trees prediction after the ensembles of trees or forest is generated.

Random forest brings both versatility and power to this machine learning approach. Because, the ensembles use the small portion of the larger dataset, it is extremely effective in handling the large dataset which might cause other methods to fail because of the dimensionality problems. Also, the error rate for the learning tasks are on par or equal to other machine learning approaches.

### IX . GENERALISED LINEAR MODEL (GLM).

In statistics, generalized linear model is a simplified generalization of the normal linear regression that allows the response variable to have an error distribution model rather than normal distribution [17].

It generalizes the simple linear regression by allowing the model to be related to the response variable through link function. This is achieved by allowing the variance of each sample to be function of its forecasted value.

In this model, each outcome Y of the dependent variable is assumed to be generated from family of probability distribution that includes the normal, binomial, Poisson, and gamma distribution. The mean depends upon on the independent variables, X through

$$E(Y) = \mu = g^{-1}(X\beta) \quad (4)$$

Where  $E(Y)$  is the expected value of Y;  $X\beta$  is the linear predictor, a linear combination of unknown parameters  $\beta$ ;  $g$  is the link function.

$$Var(Y) = V(\mu) = V(g^{-1}(X\beta)) \quad (5)$$

In this framework, the variance is typically a function, V, of the mean: It is convenient if V follows from the exponential family distribution, but it may simply be that the variance is a function of the predicted Value. The unknown parameters,  $\beta$  are typically computed with likelihood, maximum quasi-likelihood, or Bayesian techniques.

### X. HYBRID MODEL (ARIMA-GLM).

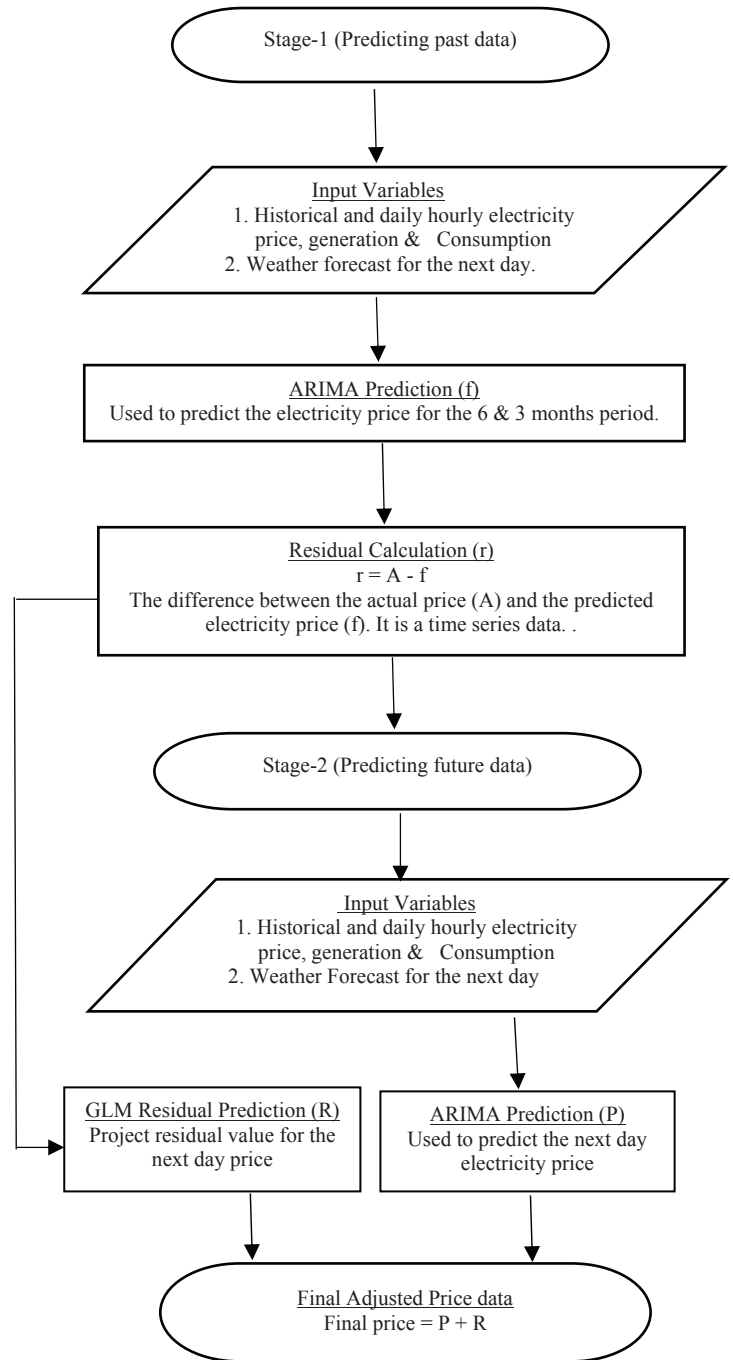


Fig.3. Flowchart showing modelling process of proposed hybrid ARIMA-GLM Model.

We propose a new two stage process to improve the forecasting estimation that mutually benefits from ARIMA and GLM method. This is a two-step process, which were executed in R programming environment. In the first step, the dataset which contains all the predictor variables is used to build an ARIMA model. For this, a pre-built software package known as 'arima' was used. This arima model is



then used to predict the price for the next day in the dataset. It should be noted here that the prediction is done for the same dataset that was used to fit the ARIMA model. The predicted values are then differenced from the actual values to create a residual dataset.

In the second step, the residuals are used as a time-series input to a GLM model. GLM is available as an R package, which can be used by calling the function 'glm'. This model is then used to predict residual values for the next day. The ARIMA model created in the first step is used here to get price forecasts for the next day as well. Finally, the forecasted price from ARIMA is added to the predicted residuals from GLM to get a final, adjusted price forecast. This data is compared against actual Iberian data set for the corresponding dates to calculate errors.

### XI EXPLANATORY (INPUT) VARIABLES FOR DAY-AHEAD PRICE FORECAST

#### A.Data Explanation

The day-ahead hourly price forecasting can be influenced by different kinds of explanatory variables:

- (a) Hourly Real Electricity Price, with correspondent date, for Iberian electricity market (MIBEL).
- (b) Actual recorded hourly power system data, mainly regional aggregated hourly power demands and power generations aggregated by generation type.
- (c) Hourly weather forecasts, including wind speed, solar irradiance and temperature.

TABLE I  
VARIABLES USED FOR FORECASTING

Variable No.	Description
1	Hourly Price D
2	Hourly Price D -6
3	Hourly Power Demand D-1
4	Hourly Power Demand D-6
5	Hourly Hydropower Generation D-1
6	Hourly Hydropower Generation D-6
7	Hourly Solar power D-1
8	Hourly Solar power D-6
9	Hourly Coal power Generation D-1
10	Hourly Coal power Generation D-6
11	Hourly Wind Power Generation D-1
12	Hourly Wind Power Generation D-6
13	Hourly Combined Cycle Power Generation D-1
14	Hourly Combined Cycle Power Generation D-6
15,16,17	Hourly Temperature , Wind speed, Radiation D+1

We tried to reduce the number of variables by taking into the significance of the variables into consideration. We were

able to get better predictions after reducing the variables. All the model performed better when there is a reduction of variables, but it is important to develop a model that improves the estimation with these 14 identified key variables. We found that last day's price 'D' and last week price 'D-6' plays an important predictor variable in determining the next day's price. The Analysis of forecast results are compared and explained in the next section.

### XII. RESULTS AND DISCUSSIONS

Several Models have been applied to predict the Electricity prices of the Iberian Markets in this study. They include ARIMA, Random Forest, Support Vector Machine, Generalized Linear Model and the Hybrid ARIMA –GLM.

The proposed model was trained and tested using the two data set (from Feb 1, 2015 to July 30, 2015) and (from May 1, 2015 to July 30, 2015) derived from the Iberian electricity market. Sets of data include hourly price, load & temperature data. For the 3 & 6 Months data under study, standard average percentage errors is computed corresponding to the 24h h of the next day. The performance was measured mainly using a Statistical measure named as MAPE.

$$MAPE_{day} = \frac{1}{24} \sum_{i=1}^{24} \frac{p^{actual} - p^{pred}}{p^{actual}} \quad (6)$$

Where  $MAPE_{day}$  is the daily error.

For the Iberian electricity market, a single day has been selected (July 31, 2015) to forecast and validate the performance of the proposed model. Figs. 4-7 shows the day-ahead price forecasting results for July-31 2015. This paper does not include any price volatility analysis and risk. The forecasting error is the main concern for power Traders, transmission planners, and utilities; a lower error indicates a better result.

Figures 4-7 provides a comparison of the different methods using MAPE for different datasets. Figure 4,5 shows the comparison graph of forecasted next day price of different methods such as ARIMA-GLM, ARIMA, Random Forest,(RF), Support vector machines(SVM) for six months and three months dataset respectively. We used different dataset to check the scalability of the model proposed.

We found that the three months dataset were giving better results due to the high relevancy of data. This is because the forecasted day (July-31 2015) has high correlation with the dataset used (from May 1, 2015 to July 30, 2015) than with the dataset (from Feb 1, 2015 to July 30, 2015).

This paper contributes to solve an important problem in price forecasting in which our test results show that the proposed hybrid algorithm could provide a considerable improvement in accuracy, efficiency and high adequacy.

Therefore, if the proposed algorithm is implemented to a forecasting system, it could improve the forecasting error.

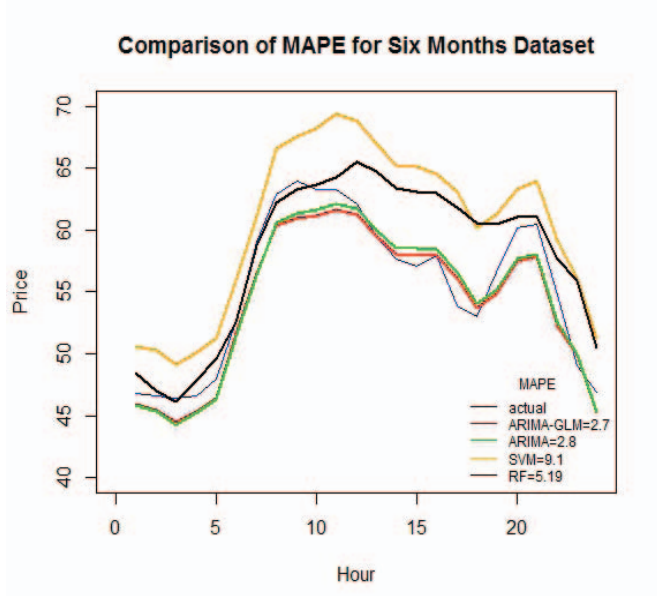


Fig.4. Comparison of MAPE for Six Months dataset (from Feb 1, 2015 to July 30, 2015) for day-ahead Iberian Electricity Market (July 31, 2015).

We used six months dataset (from Feb 1, 2015 to July 30, 2015) comprising of price, load and temperature variables to predict. As seen from the Figs. 4 and 5, the forecasted day-ahead price from the proposed hybrid model is very close to the actual value. Also, we observe that our model donot forecast the spike in prices. MAPE for the proposed hybrid model was close to 2.7% for six month dataset and was reduced to 2.4% when we used the three month of data which had high relevancy to the forecasted day price.

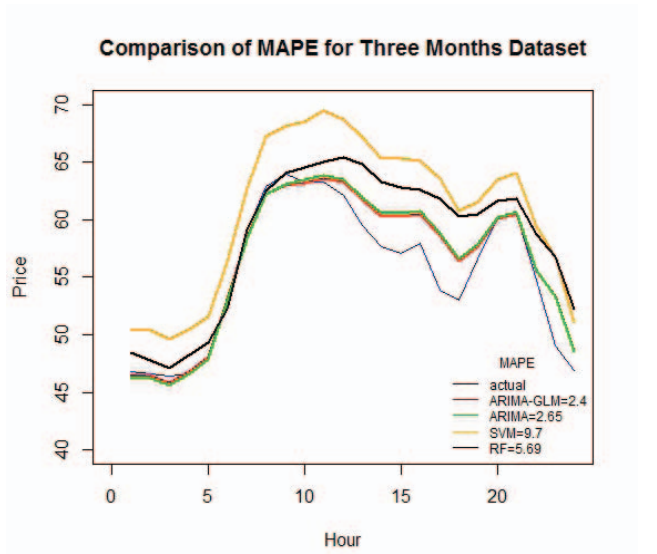


Fig.5. Comparison of MAPE for three Months (from May 1, 2015 to July 30, 2015) for day-ahead Iberian electricity Market (July 31, 2015).

Important inference form this graph is that all the methods performed better with the three months dataset except random forest and Support Vector Machines (SVM) which performed well in six months dataset. These methods uses small portion of the large dataset. This shows that these two methods are effective in handling extremely large dataset which might cause other methods to fail because of the dimensionality problems.

TABLE 2. COMPARISON OF DAY AHEAD FORECASTING PERFORMANCE OF SEVERAL MODEL

Parameter	ARIMA	RANDOM FOREST	SVM	ARIMA-GLM
MAPE-6M	2.8%	5.19%	9.1%	2.744%
MAPE-3M	2.65%	5.69%	9.7%	2.4%
Correlation-6M	97.08%	90.5%	95.8%	97.48%
Correlation-3M	96.18%	91.22%	96.06%	96.44%

Table 2 provides a numerical overview of the prediction behavior of proposed models and other existing method for the selected day where the values of MAPE, and Correlation Coefficients are compared among various methods.

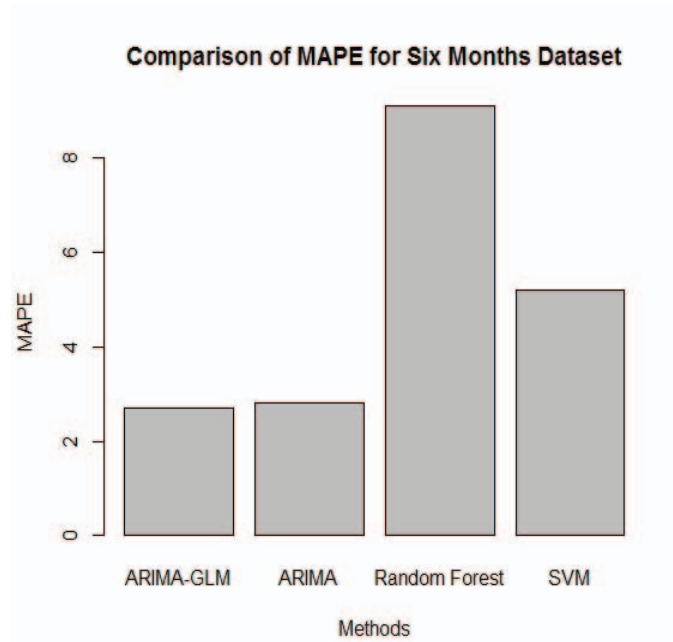


Fig.6. Comparison of MAPE for Six Months (from Feb 1, 2015 to July 30, 2015) for day-ahead Iberian Electricity Market (July 31, 2015).

Figure 6, 7 shows the simple bar chart comparison of the Mape of forecasted next day price of different methods for the two dataset used as discussed earlier. From the graph, we can see that the proposed hybrid method outperforms the other commonly used methods.

**Comparison of MAPE for Three Months Dataset**

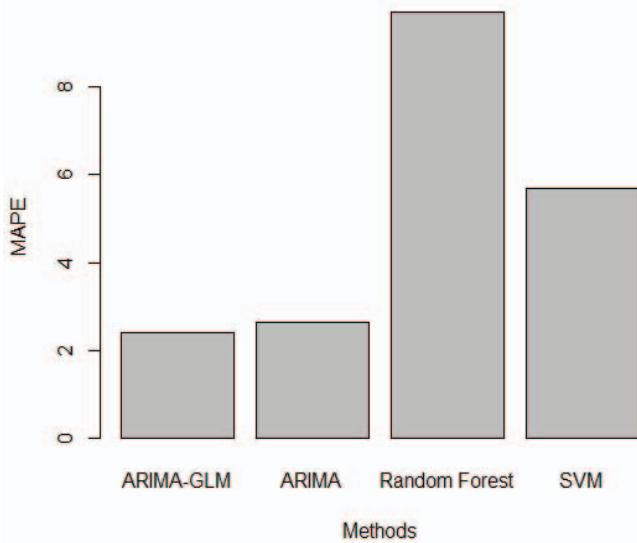


Fig.7. Comparison of MAPE for three Months (from May 1, 2015 to July 30, 2015) for day-ahead Iberian Electricity Market (July 31, 2015).

From tables 2, we can see that the proposed hybrid model gives better prediction results for both tested three months and six months of dataset. Also, we can see that the Correlation value is high for the proposed hybrid model indicating a strong relationship between the actual prices.

TABLE 3.COMPARISON OF DAY AHEAD FORECASTING PERFORMANCE OF SEVERAL MODEL INCLUDING TEMP VARIABLES (TEMP, WINDSPEED, RADIATION)

Parameter	ARIMA	RANDOM FOREST	SVM	ARIMA-GLM
MAPE-6M	3.48%	5.75%	9.92%	3.11%
MAPE-3M	3.29%	6.27%	12.02%	3.22%
Correlation-6M	94.6%	91.3%	95.8%	95.24%
Correlation-3M	95.25%	90.9%	96.06%	95.70%

Table 3 provides a numerical overview of the prediction behavior of proposed models and other existing method including temperature variables like wind speed, radiation and temperature for the selected day where the values of MAPE, and Correlation Coefficients are compared among various methods.

TABLE 4.COMPARISON OF DAY AHEAD FORECASTING PERFORMANCE OF SEVERAL MODEL FOR TWO VARIABLES (PRICE D & PRICE D-6)

Parameter	ARIMA	RANDOM FOREST	SVM	ARIMA-GLM
MAPE-6M	2.38%	6.27%	3.7%	2.39%
MAPE-3M	2.27%	6.27%	5.3%	2.26%
Correlation-6M	97.4%	90.9%	97.17%	95.09%
Correlation-3M	94.6%	90.9%	97.46%	95.09%

Table 4 provides a numerical overview of the prediction behavior of proposed models and other existing method for two variables (price D and price D-6) for the selected day where the values of MAPE, and Correlation Coefficients are compared among various methods. We were able to get better predictions after reducing the variables by taking significant variables into consideration.

### XIII. CONCLUSION

The paper investigated a collection of forecasting methods for day-ahead electricity price markets. Data from the Iberian electricity market were used to demonstrate the functioning of the proposed hybrid ARIMA-GLM Model. For the selected days and weeks, MAPE values obtained using the proposed model indicate promising preliminary results in the Iberian electricity market. Our future work includes scalability of data sets and the selection of more appropriate input variables including price volatility analysis.

### REFERENCES

- [1] Contreras, J.; Espinola, R.; Nogales, F.J.; Conejo, A.J. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* 2003, 18, 1014–1020.
- [2] Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* 2014, 30, 1030–1081.
- [3] Karakatsani, N.V.; Bunn, D.W. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *Int. J. Forecast.* 2008, 24, 764–785.
- [4] Jonsson, T.; Pinson, P.; Nielsen, H.A.; Madsen, H.; Nielsen, T.S. Forecasting electricity spot prices accounting for wind power predictions. *IEEE Trans. Sustain. Energy* 2013, 4, 210–218.
- [5] Cruz, A.; Muñoz, A.; Zamora, J.L.; Espinola, R. The effect of wind generation and weekday on Spanish electricity spot price forecasting. *Electr. Power Syst. Res.* 2011, 81, 1924–1935.
- [6] Dong, Y.; Wang, J.; Jiang, H.; Wu, J. Short-term electricity price forecast based on the improved hybrid model. *Energy Convers. Manag.* 2011, 52, 2987–2995.
- [7] Knittel, C.R.; Roberts, M.R. An empirical examination of restructured electricity prices. *Energy Econ.* 2005, 27, 791–817.
- [8] Weron, R.; Misiorek, A. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *Int. J. Forecast.* 2008, 24, 744–763.
- [9] Lira, F.; Muñoz, C.; Nuñez, F.; Cipriano, A. Short-term forecasting of electricity prices in the Colombian electricity market. *IET Gener. Transm. Distrib.* 2009, 3, 980–986.
- [10] Jjie Mei; Dawei He; Ronald Harley; Thomas Habetler; Guannan QuA random forest method for realtime price forecasting in New York electricity market
- [11] Sansom, D. C., Downs, T., & Saha, T. K. (2002). Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *Journal of Electrical and Electronics Engineering, Australia*, 22(3), 227–233.
- [12] Zhao, J. H., Dong, Z. Y., Xu, Z., & Wong, K. P. (2008). A statistical approach for interval forecasting of the electricity price. *IEEE Transactions on Power Systems*, 23(2), 267–276.
- [13] Che, J., & Wang, J. (2010). Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Conversion and Management*, 51(10), 1911–1917.

- [14] Yan, X., & Chowdhury, N. A. (2010b). Mid-term electricity market clearing price forecasting: a hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power and Energy Systems*, 53(1), 20–26.
- [15] <http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/>
- [16] B. Lantz, *Machine Learning with R*, Packt Publishing, 2013.
- [17] [https://en.wikipedia.org/wiki/Generalized\\_linear\\_model](https://en.wikipedia.org/wiki/Generalized_linear_model)
- [18] Monteiro, C.; et al. Explanatory Information Analysis for Day-Ahead Price Forecasting in the Iberian Electricity Market. *Energies* 2015, 8, 10464-10486.
- [19] Guirguis, Hany S.; Felder, Frank A., Further Advances in Forecasting Day-Ahead Electricity Prices Using Time Series Models. *KIEE International Transactions on PE*, Vol. 4-A No. 3, pp. 159~166, 2004
- [20] S. Eslahi Tatafi , Gh. A. Heydari , A. A. Gharaveisi. Short-Term Electricity Price Forecasting Using Optimal TSK Fuzzy Systems. *Journal of mathematics and computer science* 13 (2014), 238-246.
- [21] Anbazhagan, S; Kumarappan, N. Day-Ahead Deregulated Electricity Market Price Forecasting Using Recurrent Neural Network. *IEEE SYSTEMS JOURNAL*, VOL. 7, NO. 4, DECEMBER 2013