Energy 179 (2019) 214–221

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Long-term forecast of energy commodities price using machine learning

Gabriel Paes Herrera [a, b, *], Michel Constantino [b], Benjamin Miranda Tabak [c], Hemerson Pistori [b, d], Jen-Je Su [a], Athula Naranpanawa [a]

[a] Department of Accounting, Finance and Economics, Griffith University, Nathan Campus, Queensland, 4111, Australia
[b] Department of Environmental Sciences and Sustainability, Dom Bosco Catholic University, Campo Grande, MS, Brazil
[c] School of Public Policy and Government, Getulio Vargas Foundation (EPPG/FGV), Brasilia, DF, Brazil
[d] Department of Computer Science, Federal University of Mato Grosso do Sul (UFMS), Campo Grande, MS, Brazil

A R T I C L E I N F O

Article history:
Received 23 January 2019
Received in revised form
10 April 2019
Accepted 13 April 2019
Available online 28 April 2019

Keywords:
ANN
Random forests
Natural gas
Coal
Oil

A B S T R A C T

We compare the long-horizon forecast performance of traditional econometric models with machine learning methods (Neural Networks and Random Forests) for the main energy commodities in the world using monthly prices provided by the International Monetary Fund (IMF). We study the case of Oil (Brent, WTI and Dubai Fateh), Coal (AU) and Gas (US and Russia). Models accuracy are measured using RMSE and MAPE and the M-DM test is applied to evaluate whether there is a statistically significant difference between the methods. We computed thousands of tests regarding the machine learning parameters combinations as there is no method to set the optimal structure for these models. The results show that machine learning methods outperform traditional econometric methods and also that they present an additional advantage, which is the capacity to predict turning points. This study adds further evidence for the discussion on the use of machine learning algorithms for the development of more accurate forecasts to support policymakers and help the decision-making process in the international energy market.

1. Introduction

The global energy consumption in 2040 is expected to reach 739 quadrillion Btu [1]. In sectors with high energy consumption such as commercial and residential buildings and wastewater treatment, energy conservation is becoming a major concern and studies with alternatives to improve sustainability as Wang et al. [2] and Sepehri and Sarrafzadeh [3] are very important. From 1971 to 2016 the world total primary energy supply (TPES) increased by almost 2.5 times and the international energy market has become more competitive and complex than ever [4].

Price of energy sources have an intimate relation with the world economic and political environment, in history, wild fluctuation in oil prices has led to recessions and even collapsing regimes and oil shocks have been seen as one of the main dampeners of economic growth since the Second World War [5,6]. As stated by Samarasinghe [7]; time series forecasting is an important area of forecasting in which past observations are analyzed to develop a model describing the basic relationship in a sequence of event outcomes and this model is then used to forecast future events. Government agencies tend to rely on long-term forecast of oil price, for example, to develop oil reserve strategies and ensure a country‘s economic development level, social stability and political orderliness [5].

The price of some energy sources might result in inflation, economic depression and political turmoil. Therefore, one might conclude that predicting the price of such energy sources is a vital topic [8]. According to the U.S. Energy Information Administration (EIA), in 2017 the global primary energy sources were: oil (32%), coal (27%) and natural gas (22%). By 2040 natural gas is projected to surpass coal and become the second most consumed energy source in the world, with oil remaining the first and coal becoming the third [1]. As stated by Chai et al. [9]; WTI and Brent crude oil prices are a wind indicator for the international crude oil market and play a significant role in stabilizing the international financial market. The United States is the largest producer of oil and natural gas, with

* Corresponding author. Department of Accounting, Finance and Economics, Griffith University, Nathan Campus, Queensland, 4111, Australia.

E-mail addresses: gabriel.paesherrera@griffithuni.edu.au, gabrielherrera27@hotmail.com (G.P. Herrera), michel@ucdb.br (M. Constantino), benjaminm.tabak@gmail.com (B.M. Tabak), pistori@ucdb.br (H. Pistori), j.su@griffith.edu.au (J.-J. Su), a.naranpanawa@griffith.edu.au (A. Naranpanawa).

https://doi.org/10.1016/j.energy.2019.04.077
0360-5442/

Russia being the second largest producer of the latter. As for coal, the People's Republic of China is the world's biggest producer but also the biggest consumer. In matters of market, Indonesia and Australia are the largest coal exporters [4,10]. Therefore, studies aiming to forecast the price scenario of these energy commodities are fundamental to policymakers and to help in the decision-making process.

As stated by Armstrong [11]; there is no single forecasting method that dominates all other methods, each one has its pros and cons. For the last decades plenteous forecasting approaches have been proposed. Classic econometric and statistic models include Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH). However, these methods are based on strict linear assumptions and the price series must be linear or near linear. Machine learning methods can effectively distinguish random factors and capture the hidden nonlinear features which traditional econometric models are unable to do. Commonly used machine learning methods include artificial neural networks (ANN), support vector machine (SVM) and random forests (RF). Recently, hybrid models have become popular for their forecasting accuracy gains and numerous comparative studies have been conducted. Nevertheless, findings are mixed regarding which method is better.

Neural networks proved to be a superior method for forecasting wheat price in China by Zou et al. [12]; cattle price in Omaha and wheat price in Chicago by Kohzadi et al. [13]. Contrarily, Ho et al. [14] concluded that the neural network model did not perform well and was inferior to the ARIMA model in predicting compressor failures of a repairable system. The results of Jovanović et al. [15] showed that an ensemble of three different artificial neural networks is better for prediction of heating energy consumption.

According to the study conducted by Kane et al. [16]; the random forest model showed better results than the ARIMA model in predicting H5N1 outbreaks in birds in Egypt. Similarly, Wang et al. [2] proposed a better performing model based on ensemble bagging trees to predict building energy use. Zhang [17] recommended a hybrid model combining ARIMA and ANNs, the empirical results suggested that the hybrid model is able to outperform both methods used in isolation. Differently, Ding [8] recommended a method combining an ensemble empirical mode decomposition and artificial neural networks for better forecasting accuracy. The study conducted by Darbellay and Slama [18] showed that the ARIMA model is better suited for short-term forecasts and that neural networks are better suited for long-term forecasts. Safari and Davallou [19] proposed a hybrid model combining econometric methods and neural networks for forecasting monthly OPEC crude oil prices and WTI crude oil spot prices, the results show a decrease in forecast error using their proposed method.

Long-term forecasts are what really interests policymakers and affect investment decisions. For example, crude oil futures contract may have maturities as long as 7 years and it is precisely these long horizons that many policymakers focus on. Our contribution extends the discussion and add arguments to the literature by applying and evaluating three forecasting methods, i.e. a hybridization of traditional econometric models, the popular artificial neural network and the less common random forest. We used a large dataset of historical prices provided by the International Monetary Fund (IMF) for the main energy commodities in the world. In addition, our results are based on thousands of tests regarding the machine learning parameters. We believe the results will be important for future researches in the academic field and to support policymakers in the decision-making process in the international energy market.

The reminder of this paper is structured as follow. The subsequent section describes the data source, as well as the models applied in this study for time series forecasting. In the third section, we present the results and the discussion. The last section concludes and brings recommendations for further research.

2. Material and methods

The historical prices of oil, coal and natural gas are provided by the International Monetary Fund (IMF). The time series were chosen according to their importance for the international energy market and the availability of data. Prices are monthly, reported in nominal U.S. dollars, period average and not seasonally adjusted. The time series of oil and coal consists of prices from January 1980 to June 2017 (450 observations), the time series of Russian natural gas consists of prices from January 1985 to June 2017 (390 observations) and the time series of US natural gas consists of prices from January 1991 to June 2017 (318 observations). The description of each time series can be found in Table 1.

We are aware that volatility of energy prices, i.e. oil, coal and natural gas reflect internal and external pressures of several areas such as politics and economics. For example, it is well known that WTI price has a pre- and post-1973 period due to government regulations. However, the description of all episodes that influenced energy prices in history is beyond the scope of this work.

The series were divided into training and test-sets for modeling and verification purposes, where the training-set contains approximately the first 80% observations and the test-set the remaining 20% observations. All analyzes were conducted using R 3.4.4 (R [20].

2.1. Hybrid models

The hybrid model applied in this study combines five traditional econometric methods for time series forecasting: autoregressive integrated moving average (ARIMA); error, trend and seasonality (ets); seasonal and trend decomposition using Loess (stl); exponential smoothing state space model with Box-Cox transformation, ARMA errors, trend and seasonal components (tbats) and Theta model.

The conventional and simplest method of combining the five forecasts is to set equal weights to each individual model and in many cases this approach performs better and generates the best forecasting results. However, some authors state that determining the weight for each model is the key step in developing a hybrid forecast model as one individual model might make a greater contribution than the other ones [21,22].

In order to determine the weights, the algorithm used in this study applies a non-rolling time series cross-validation to minimize the root mean square error (RMSE) and set the optimal weight coefficients. In addition, we also performed the hybrid models with equal weights and compared the results to choose the best one. The analyzes were conducted using the forecastHybrid package [23] v2.1.11 in R.

2.2. ANN models

Artificial neural networks are information processing models inspired by the way the interconnected structure of a biological brain processes information. The power of ANNs lies in their ability to capture nonlinear relationships inherent in data. As stated by Samarasinghe [7]; while linear models depict a linear relationship between the current and next observations, neural networks portray a nonlinear relationship between the two that can be described as:

Table 1
Time series description and summary statistics.

Time series	Description	Mean	Std. Dev.
Oil Brent	Crude Oil (petroleum). Dated Brent, light blend 38 API, fob U.K., US$/barrel.	41.946	30.927
Oil WTI	Crude Oil (petroleum). West Texas Intermediate 40 API, Midland Texas, US$/barrel.	41.309	27.720
Oil Dubai	Crude Oil (petroleum). Dubai medium Fateh 32 API, US$/barrel.	39.737	30.246
Coal AU	Australian thermal coal. 12,000- btu/pound, FOB Newcastle/Port Kembla, US$/metric ton.	52.475	29.603
Gas US	Natural Gas spot price at the Henry Hub terminal in Louisiana, US$/Million Metric BTU.	3.875	2.260
Gas Russia	Russian Natural Gas border price in Germany, US$/Million Metric BTU.	5.097	3.510

$$y_t = f\left(y_{t-1}, y_{t-2}, \ldots, y_{t-p}\right) + \varepsilon_t$$

where $f(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ is the nonlinear function that maps a series of past observations nonlinearity to the next outcome and ε_t is the error, which is expected to be a random variable with a mean of zero and variance σ^2.

A widely used ANN structure is the feedforward multi-layer perceptron (MLP), which we have employed in this study. As described by Lasheras et al. [24]; the MLP architecture is characterized as having its neurons grouped into layers of different levels. Each of the layers is formed by a set of neurons and different kinds of layers are distinguished: the input layer, the hidden layer and the output layer. A MLP artificial neural network can have more than one hidden layer and each neuron, also often called nodes, in one layer connects with a certain weight to every other node in the following layer.

The most serious drawback of feedforward networks such as MLP is the trial and error involved in the selection of model parameters, particularly the number of hidden neurons [7]. As stated by Crone and Kourentzes [25]; there is no methodology universally accepted to guide the architecture specification of MLPs for time series prediction.

In our study we opted to use an iterative neural filter (INF) methodology that applies a feature evaluation algorithm to automatically identify the number and the period of the seasonalities that are present in a time series and capture them in the input vector of the ANN. For parameterization, data is presented to the MLP as a randomized set of input vectors of fixed length formed as a sliding, overlapping window over the time series observations. Including only lagged realizations of the dependent variable, a feedforward NN is constructed where each node receives an input signal which is the total information, or external stimulation, from other nodes, process it locally through a hyperbolic tangent (TanH) transfer function and produces a transformed output signal to other nodes or external outputs. The analyzes were conducted using the nnfor package [26] v0.9.2 in R. For more details of the specific methodology please see Kourentzes et al. [27] and Crone and [25].

2.3. Random forest

Random forest (RF) is a machine learning algorithm introduced by Breiman [28]; that can be used for classification and regression. As stated by Biau and Scornet [29]; RF operates according to the “divide and conquer” principle by applying three simple but effective steps: sample fractions of the data, grow a randomized tree predictor on each small piece and then aggregate these predictors by averaging.

Performing regression using random forest for time series forecasting is not identical to a simple regression case. As stated by Tyralis and Papacharalampous [30]; in this situation the role of the predictor variables is taken by previous values of the time series (lagged variables). Therefore, increasing the number of predictor variables inevitably results in reducing the length of the training set and, at the same time, using fewer predictor variables may reduce the information obtained by the available knowledge of the temporal dependence. The right balance between number of predictor variables and length of the training set is important to obtain the most accurate results.

According to Breiman [28]; random forests for regression are formed by growing trees depending on a random vector Θ such that the tree predictor $h(x, \Theta)$ takes on numerical values as opposed to class labels. The output values are numerical, and it is assumed that the training set is independently drawn from the distribution of the random vector Y, X. The mean-squared generalization error for any numerical predictor h(x) is:the random forest predictor is formed by taking the average over K of the trees $\{h(x, \Theta_k)\}$. The same author also highlights that random forests are an effective tool in prediction and because the Law of Large Numbers they do not overfit. The analyzes were conducted using the randomForest v4.6-14 [31], the rminer v1.4.2 [32] and the caret v6.0-79 (from Ref. [33] packages in R.

$$E_{X,Y}(Y - h(X))^2$$

2.4Performance evaluation criteria.

A forecast error represents the difference between the forecast value and the actual value. For validation purposes, it is common practice to separate the available data into two segments, one used to estimate the parameters of the model (training set), and the other used to evaluate the accuracy of the forecast made with the model (test set) as described by Hyndman and Athanasopoulos [34]; Arlot et al. [35] and Armstrong [11]. Authors normally use most of the data to train the model and a smaller sample to evaluate. For this study, we split the data into 80%–20% for training and test samples respectively following Hyndman and Athanasopoulos [34].

To evaluate the forecasting performance of the models we selected two widely used statistical loss functions, i.e. the mean absolute percentage error (MAPE) and the root mean square error (RMSE). These two indicators have been successfully adopted in similar studies in recent years, a few examples are Chai et al. [9]; Wang et al. [36]; Zhao et al. [37] and Tang et al. [38]. The definitions are presented as below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{x}_t - x_t)^2}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right|$$

where N is the number of data for testing set, and x_t and $\hat{x}_t$ are real value and estimating value at time t respectively.

Additionally, in order to statistically test the significant difference regarding the predictive accuracy amongst the forecasting models we applied the modified Diebold-Mariano (M-DM) test as proposed by Harvey et al. [39]. The main objective of the test can be comprehended by defining the original DM test [40]. The loss function is set to mean square prediction error (MSPE) and the null hypothesis is that the MSPE value of the tested model A is equal to the benchmark model B. The test can be defined as:

$$S = \frac{\overline{g}}{\left(\widehat{V}_{\overline{g}}/N\right)^{1/2}}$$

where $\overline{g} = {}^{1}\!/\!{}_{N}\sum_{t=1}^{N}[(x_t - \widehat{x}_{A,t})^2 - (x_t - \widehat{x}_{B,t})^2]$, $\widehat{V}_{\overline{g}} = \gamma_0 + 2\sum_{l=1}^{\infty}\gamma_l$, $\gamma_l = cov(g_t, g_{t+1})$ and $\widehat{x}_{A,t}$, $\widehat{x}_{B,t}$ are the predicted values for x_t calculated by the tested method A and its benchmark method B, respectively, at time t.

3. Results

Initially we plotted each time series, as well as their log return, in order to visually analyze the data and the behavior of these commodities prices along the time. Fig. 1 reveals the non-seasonality of the data in the six cases, which was also statistically confirmed using the seasonal package [41] v1.6.1 in R. This package uses the automatic procedures of X-13 ARIMA-SEATS [42] and the results confirmed no significant seasonal peaks in all six series. It is possible to observe that the prices of these energy commodities are similarly impacted by some global events such as the 2008–2009 financial crisis, while other shocks are attributed to the particularities of each market [37,43].

Subsequently, the historical prices of the six energy commodities were tested under the random walk hypothesis using wild bootstrapping of the automatic variance ratio (VR) test of Choi [44] as proposed by Kim [45]. The null hypothesis of this test is that the time series is serially uncorrelated, or in other words, follows a random walk behavior. In this study, the tests conducted with each commodity rejected the null hypothesis and thus confirmed that the observations in each time series presents significant correlation.

The study conducted by Alquist et al. [46] provides strong evidence that after 1973 changes in the nominal price of oil are predictable, contradicting the popular claim that changes in the price of oil are inherently unpredictable. In addition, Bontempi et al. [47] highlight that according to the theory of deterministic chaos the apparently random behavior may be generated by deterministic systems with only a small number of degrees of freedom, interacting nonlinearly.

After verifying that the time series does not follow a random walk behavior we proceeded with the forecasting analyzes, beginning with the hybrid models. The optimum weights for each model were set using the methodology described in item 2.1 of this study. For the Oil Brent and Gas Russia time series, the hybrid model with equal weights showed to perform better, while for the other four cases setting different weights for each individual model presented better accuracy. A full description of all methods and their respective parameters applied in this study can be found in Table 2.

Regarding the ANN models, 1400 combinations with different numbers of hidden layers and nodes were tested for each energy

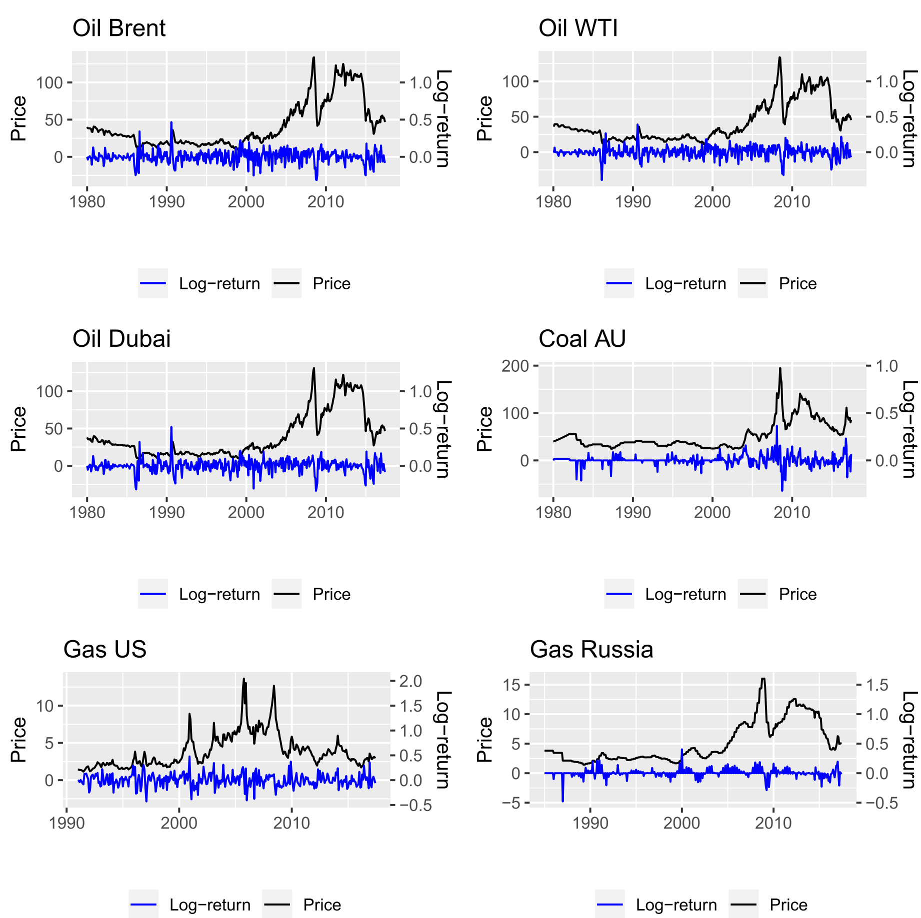

Fig. 1. Individual time series and their log return.

Table 2
Models parameters.

	Hybrid	ANN	RF
Oil Brent	weights: 0.2 - ARIMA(2,1,2) 0.2 - ets/0.2 - theta 0.2 - stl/0.2 - tbats	8 lag inputs: 1, 2, 3, 5, 6, 9, 11, 12 1st hidden layer: 15 nodes 2nd hidden layer: 10 nodes 3rd hidden layer: 5 nodes 1 output node	Lag variables: 72 Number of trees: 500 mtry: 72
Oil WTI	weights: 0.216 - ARIMA(3,1,1)(0,0,1) 0.204 - ets/0.208 - theta 0.169 - stl/0.208 - tbats	8 lag inputs: 1, 2, 3, 5, 6, 9, 11, 12 1st hidden layer: 15 nodes 1 output node	Lag variables: 18 Number of trees: 500 mtry: 18
Oil Dubai	weights: 0.187 - ARIMA(0,1,2)(0,0,1) 0.209 - ets/0.217 - theta 0.168 - stl/0.219 - tbats	7 lag inputs: 1, 2, 3, 6, 9, 10, 12 1st hidden layer: 15 nodes 1 output node	Lag variables: 36 Number of trees: 500 mtry: 36
Coal AU	weights: 0.205 - ARIMA(2,1,2) 0.170 - ets/0.246 - theta 0.166 - stl/0.213 - tbats	6 lag inputs: 1, 3, 4, 6, 7, 8 1st hidden layer: 10 nodes 1 output node	Lag variables: 36 Number of trees: 500 mtry: 36
Gas US	weights: 0.198 - ARIMA(0,1,0) 0.207 - ets/0.205 - theta 0.2 - stl/0.190 - tbats	4 lag inputs: 5, 7, 9, 12 1st hidden layer: 10 nodes 2nd hidden layer: 5 nodes 1 otput node	Lag variables: 13 Number of trees: 500 mtry: 7
Gas Russia	weights: 0.2 - ARIMA(5,1,4)(0,0,2) 0.2 - ets/0.2 - theta 0.2 - stl/0.2 - tbats	9 lag inputs: 1, 2, 3, 4, 5, 7, 9, 10, 11 1st hidden layer: 5 nodes 1 otput node	Lag variables: 16 Number of trees: 500 mtry: 9

commodity using the training sets, followed by a forecast matching the test sets for validation. The values of RMSE and MAPE of each parameter combination for all six series were computed and the best combination of hidden layers and nodes for each time series can be found in Table 2. Additionally, an illustration of each ANN architecture is presented in Fig. 2.

As we can observe, the Oil Brent time series presented the most complex ANN structure in this study, involving three hidden layers. The Gas US time series contains two hidden layers in its structure and the other four series, i.e. Oil WTI, Oil Dubai, Coal Australia and Gas Russia only one hidden layer in their structures, the latter can be considered the simplest structure of the six models analyzed. As described in section 2.2, each node in one layer process the received information and connects with a certain weight to every other node in the following layer, therefore the architecture of the Oil Brent series has 325 connections while the Gas Russia time series has 50 connections.

The third forecasting method was the random forests. We tested 1000 different combinations of number of trees and lagged variables for each energy commodity using the training sets, followed by a forecast matching the test sets for validation. We followed the same process used in the ANN analyzes and the values of RMSE and MAPE of each parameter combination for all six series were computed. In all six cases the best results were obtained with the number of trees set as 500. The number of lagged variables used for each time series models, as well as the mtry parameter are presented in Table 2. The mtry parameter is the number of variables randomly sampled for splitting at each tree node as described in Breiman [28].

Notwithstanding that earlier results in this section reveals that

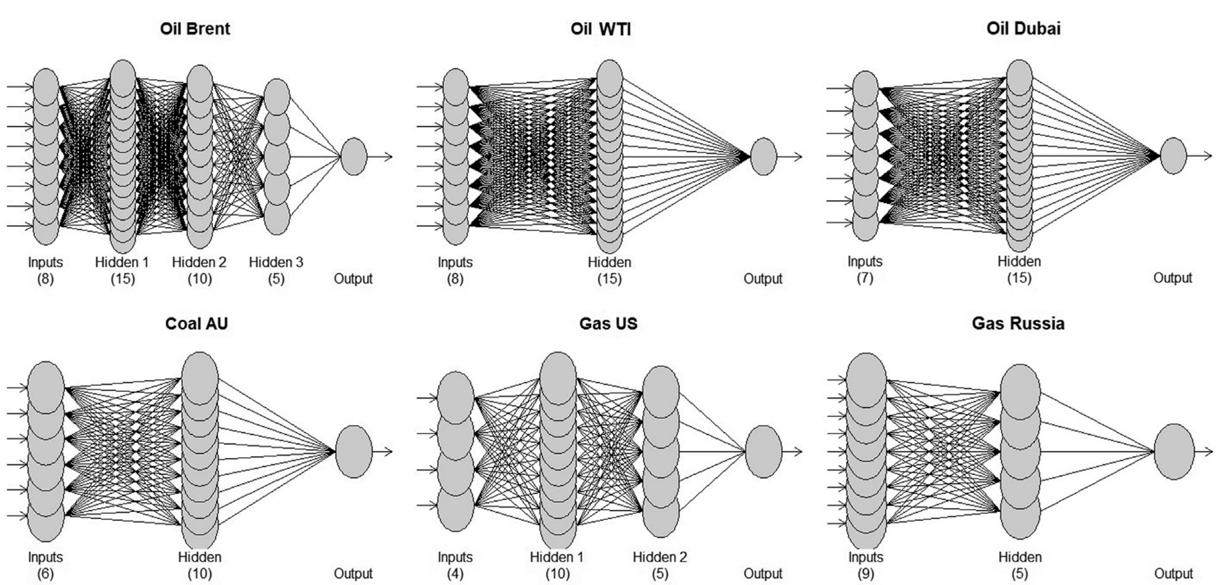

Fig. 2. Structure of each Artificial Neural Network model.

Table 3
RMSE and MAPE values.

	No-change		Hybrid		ANN		Random forests	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Oil Brent	29.7464	35.9681	29.6577	36.2480	17.3383	17.9247	7.8334	8.1483
Oil WTI	23.3258	33.4339	23.3935	33.7099	14.7208	16.5554	7.8539	8.9603
Oil Dubai	28.0263	37.4157	27.8845	38.0074	20.2872	20.4330	7.6483	8.3276
Coal AU	23.3881	21.7668	23.6744	24.1706	12.7965	11.5431	8.8422	7.3209
Gas US	0.8334	22.4672	0.8425	23.3467	0.7751	23.7169	0.4151	10.6293
Gas Russia	2.8695	37.2933	2.9458	37.4533	2.6755	32.8813	0.8491	10.0467

Table 4
M-DM test results.

	Hybrid-ANN	Hybrid-NC	Hybrid-RF	ANN-RF	NC-ANN	NC-RF
Oil Brent	10.6799***	−0.6296	12.6641***	7.0751***	10.924***	12.542***
Oil WTI	5.8676***	0.34,537	10.8199***	5.1128***	6.072***	11.162***
Oil Dubai	7.3218***	−0.7429	12.7603***	8.2356***	8.5633***	12.9537***
Coal AU	7.2169***	0.41,492	8.4155***	3.4135***	6.2233***	7.3243***
Gas US	1.1661	0.72,283	4.3657***	4.7235***	0.90,673	4.1158***
Gas Russia	6.5534***	2.4476**	12.1664***	11.8703***	2.8735***	10.7063***

***p-value<0,01; **p-value<0,05; *p-value<0,1.

the time series does not follow a random walk it is conventional to compare forecast results with the no-change (NC) forecast, considered a natural benchmark [46]. The no-change forecast is also known as the random walk model without drift, this model implies that changes in an observation value are unpredictable, so the best forecast is simply the current observation value.

The error values of the forecasts performed with the hybrid, the ANN, the random forests and the no-change models are presented in Table 3. The error values results for the six time series using the random forest method were significantly lower than those found using the hybrid, the artificial neural network and the no-change methods, demonstrating a superior performance of the RF forecast. Our results corroborate with studies conducted by Lahouar and Slama [48]; Kane et al. [16] and Dudek [49].

The same conclusion is reinforced by the M-DM test results (Table 4). All tests pairing the random forest with the other three models are statistically significant under 99% confidence level (p-value<0.01). Therefore, the null hypothesis is rejected, that is the accuracy level of the compared models are different and the positive S value shows the superiority of model b over model a. When we compare the hybrid against the no-change model the test shows no significant difference between the accuracy of both models for all time series, except for the Gas Russia series which result is only significant at a 5% confidence level. Additionally, for the Gas US time series the M-DM test reveals no significant difference between the accuracy of the hybrid, the ANN and the no-change forecasts.

Traditional econometric models, such as ARIMA, are important linear time series models that assume that the future value of a variable is the linear function of several past observations and random error. However, it cannot encompass nonlinear correlations and structures [19]. Consequently, forecasting using these models might result in low accuracy, as is the case presented in our results.

Moreover, the machine learning methods presented an additional advantage over the hybrid and the no-change methods which is the capacity of predicting turning points. As stated by Safari and Davallou [19]; machine learning methods have the ability to estimate a general level of function with a higher degree of precision and to discover nonlinear patterns in input datasets without defaults.

Fig. 3 shows all four forecasts for Oil Brent, Oil WTI, Oil Dubai, Coal Australia, Gas US and Gas Russia using the training sets and then forecasting the exact period matching the test sets. In order to better visualize the plots, we limited the X axis only for the period from 2000 to 01-01 to 2017-06-01. Some authors state that the prediction of turning points is as important as the value of the forecast itself [12,13]. Clearly the hybrid model combining traditional econometric methods and the no-change model performed poorly. Notably, the no-change forecast even showed better performance than the hybrid model when we compare the RMSE and MAPE values.

Comparing the machine learning methods, the RF is also more advantageous regarding the parameters configuration. There are many combinations of number of layers and neurons that can be used in a neural network which makes this step exhaustive. Further, depending on the number of layers and neurons it can be very computationally expensive. Since there is no formula to calculate the optimal structure for neither of these methods it is convenient that the method with the best performance is also the easier for configuring.

Finally, it is notable that the random forest method is able to achieve low error values for such long horizons. A similar work conducted by Safari and Davallou [19] testing a hybrid model of econometric methods and neural networks reported a MAPE value of 8.31 for a short forecasting period of three years, while our work presents a similar MAPE value for a forecasting period of more than seven years. Our results are even more remarkable considering that no exogenous variable is employed. This is a particular advantage since usually it is difficult to find relevant exogenous variables that match the frequency and the period of a specific time series. Additionally, even an external variable such as the oil futures prices, which is considered the best available oil forecast, does not improve forecast accuracy, as shown by Alquist et al. [46].

4. Conclusions

The analyzes performed in this study indicate that forecasting using random forests may deliver more accurate results. We used almost four decades of real data of six main energy commodities in the world and in all examples the RMSE, MAPE and M-DM test results were significantly better for the RF models when compared to the benchmark models.

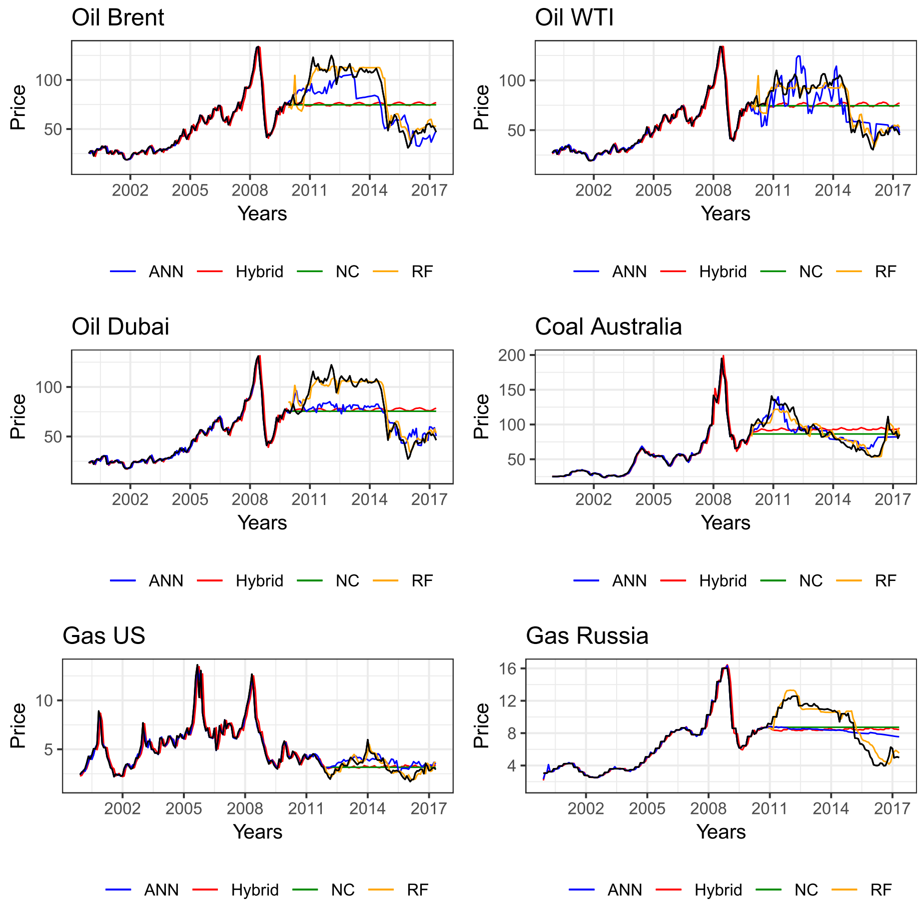

Fig. 3. Forecasts of energy commodities prices.

The results are based on thousands of tests regarding the machine learning parameters combinations as there is no method to set the optimal structure for these models. One limitation, but at the same time a distinction, of this study is that we only used the time series as input, while most of similar works employ exogenous variables. The reason for that was to try to simplify the forecasting procedure and make it easier and uncomplicated, or as one might say "do more with less". We believe we achieved this goal since all MAPE results for the RF method were inferior or equal to 10% in the long-term performed forecasts.

Considering these results, we strongly suggest researchers to include RF method as a valuable alternative for forecasting monthly prices of energy commodities. We suggest further studies to investigate the performance of RF forecast in different research areas as well as the accuracy gains of including exogenous variables. Further research could also exploit long-term relationships as some energy commodities may have time-varying long-range dependency [50,51]. Finally, this study contributes with the development of more accurate forecasting models and we believe the results will be important to support policymakers and help the decision-making process in the international energy market.

Acknowledgements

The authors would like to express their sincere appreciation to the editor and the two anonymous referees for their valuable comments and suggestions, which helped to improve the quality of this paper. Gabriel P. Herrera would like to thank Griffith University for the Postgraduate Research Scholarship. Benjamin M. Tabak and Hemerson Pistori gratefully acknowledges financial support from CNPq Foundation. Hemerson Pistori also acknowledges financial support from FUNDECT Foundation. The authors would like to thank NVIDIA Corporation for the donation of resources utilized in this research.

References

[1] EIA EIA. International energy outlook 2018. Technical report. Washington, DC: US Department of Energy; 2018.
[2] Wang Z, Wang Y, Srinivasan RS. A novel ensemble learning approach to support building energy use prediction. Energy Build 2018b;159:109–22. http://www.sciencedirect.com/science/article/pii/S0378778817335272. https://doi.org/10.1016/j.enbuild.2017.10.085.
[3] Sepehri A, Sarrafzadeh MH. Effect of nitrifiers community on fouling mitigation and nitrification efficiency in a membrane bioreactor. Chemical Engineering and Processing - Process Intensification 2018;128:10–8. http://www.sciencedirect.com/science/article/pii/S0255270117312916. https://doi.org/10.1016/j.cep.2018.04.006.
[4] IEA IEA. World energy balances: overview. Technical report. International Energy Agency (IEA); 2018b.
[5] Cheng F, Fan T, Fan D, Li S. The prediction of oil price turning points with log-periodic power law and multi-population genetic algorithm. Energy Econ 2018;72:341–55. http://www.sciencedirect.com/science/article/pii/S0140988318301312. https://doi.org/10.1016/j.eneco.2018.03.038.
[6] van de Ven DJ, Fouquet R. Historical energy price shocks and their changing effects on the economy. Energy Econ 2017;62:204–16. http://www.sciencedirect.com/science/article/pii/S0140988316303553. https://doi.org/10.1016/j.eneco.2016.12.009.
[7] Samarasinghe S. Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications; 2016.
[8] Ding Y. A novel decompose-ensemble methodology with aic-ann approach for crude oil forecasting. Energy 2018;154:328–36. http://www.sciencedirect.com/science/article/pii/S0360544218307497. https://doi.org/10.1016/j.

energy.2018.04.133.
[9] Chai J, Xing LM, Zhou XY, Zhang ZG, Li JX. Forecasting the wti crude oil price by a hybrid-refined method. Energy Econ 2018;71:114–27. http://www.sciencedirect.com/science/article/pii/S0140988318300549. https://doi.org/10.1016/j.eneco.2018.02.004.
[10] IEA IEA. Coal information: overview. Technical report. International Energy Agency (IEA); 2018a.
[11] Armstrong JS. Principles of forecasting: a handbook for researchers and practitioners, vol. 30. Springer Science & Business Media; 2001.
[12] Zou H, Xia G, Yang F, Wang H. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. Neurocomputing 2007;70:2913–23. Applications in Electrical Engineering Selected papers from the 3rd International Work-Conference on Artificial Neural Networks (IWANN 2005), https://doi.org/10.1016/j.neucom.2007.01.009. neural Network.
[13] Kohzadi N, Boyd MS, Kermanshahi B, Kaastra I. A comparison of artificial neural network and time series models for forecasting commodity prices. Neurocomputing 1996;10:169–81 [financial Applications, Part I], https://doi.org/10.1016/0925-2312(95)00020-8.
[14] Ho S, Xie M, Goh T. A comparative study of neural network and box-jenkins arima modeling in time series prediction. Comput Ind Eng 2002;42:371–5. https://doi.org/10.1016/S0360-8352(02)00036-0.
[15] Jovanović R, Sretenović AA, Živković BD. Ensemble of various neural networks for prediction of heating energy consumption. Energy Build 2015;94:189–99. http://www.sciencedirect.com/science/article/pii/S0378778815001577. https://doi.org/10.1016/j.enbuild.2015.02.052.
[16] Kane MJ, Price N, Scotch M, Rabinowitz P. Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. BMC Bioinf 2014;15:276. https://doi.org/10.1186/1471-2105-15-276. 10.1186/1471-2105-15-276.
[17] Zhang G. Time series forecasting using a hybrid arima and neural network model. Neurocomputing 2003;50:159–75. https://doi.org/10.1016/S0925-2312(01)00702-0.
[18] Darbellay GA, Slama M. Forecasting the short-term demand for electricity: do neural networks stand a better chance? Int J Forecast 2000;16:71–83. https://doi.org/10.1016/S0169-2070(99)00045-X.
[19] Safari A, Davallou M. Oil price forecasting using a hybrid model. Energy 2018;148:49–58. http://www.sciencedirect.com/science/article/pii/S0360544218300070. https://doi.org/10.1016/j.energy.2018.01.007.
[20] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org/.
[21] Wang JJ, Wang JZ, Zhang ZG, Guo SP. Stock index forecasting based on a hybrid model. Omega 2012;40:758–66. https://doi.org/10.1016/j.omega.2011.07.008. special Issue on Forecasting in Management Science.
[22] Zhu S, Wang J, Zhao W, Wang J. A seasonal hybrid procedure for electricity demand forecasting in China. Appl Energy 2011;88:3807–15. https://doi.org/10.1016/j.apenergy.2011.05.005.
[23] Shaub D, Ellis P. forecastHybrid: convenient functions for ensemble time series forecasts. https://CRAN.R-project.org/package=forecastHybrid. r package version 2.1.11; 2018.
[24] Lasheras FS, de Cos Juez FJ, Sánchez AS, Krzemień A, Fernández PR. Forecasting the comex copper spot price by means of neural networks and arima models. Resour Pol 2015;45:37–43. https://doi.org/10.1016/j.resourpol.2015.03.004.
[25] Crone SF, Kourentzes N. Feature selection for time series prediction - a combined filter and wrapper approach for neural networks. Neurocomputing 2010;73:1923–36. http://www.sciencedirect.com/science/article/pii/S0925231210000974. https://doi.org/10.1016/j.neucom.2010.01.017. subspace Learning/Selected papers from the European Symposium on Time Series Prediction.
[26] Kourentzes N. Nnfor: time series forecasting with neural networks. 2017. https://CRAN.R-project.org/package=nnfor. r package version 0.9.2.
[27] Kourentzes N, Barrow DK, Crone SF. Neural network ensemble operators for time series forecasting. Expert Syst Appl 2014;41:4235–44. http://www.sciencedirect.com/science/article/pii/S0957417413009834. https://doi.org/10.1016/j.eswa.2013.12.011.
[28] Breiman L. Random forests. Mach Learn 2001;45:5–32. https://doi.org/10.1023/A:1010933404324. 10.1023/A:1010933404324.
[29] Biau G, Scornet E. A random forest guided tour. TEST 2016;25:197–227. https://doi.org/10.1007/s11749-016-0481-7. 10.1007/s11749-016-0481-7.
[30] Tyralis H, Papacharalampous G. Variable selection in time series forecasting using random forests. Algorithms 2017;10. http://www.mdpi.com/1999-4893/10/4/114. 10.3390/a10040114.
[31] Liaw A, Wiener M. Classification and regression by randomforest. R News 2002;2:18–22. https://CRAN.R-project.org/doc/Rnews/.
[32] Cortez P. Rminer: data mining classification and regression methods. 2016. https://CRAN.R-project.org/package=rminer. r package version 1.4.2.
[33] Jed Wing MKC, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, , The R Core Team, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C, Hunt T. Caret: classification and regression training. https://CRAN.R-project.org/package=caret. r package version 6.0-79; 2018.
[34] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018.
[35] Arlot S, Celisse A, et al. A survey of cross-validation procedures for model selection. Stat Surv 2010;4:40–79.
[36] Wang M, Tian L, Zhou P. A novel approach for oil price forecasting based on data fluctuation network. Energy Econ 2018a;71:201–12. http://www.sciencedirect.com/science/article/pii/S0140988318300793. https://doi.org/10.1016/j.eneco.2018.02.021.
[37] Zhao Y, Li J, Yu L. A deep learning ensemble approach for crude oil price forecasting. Energy Econ 2017;66:9–16. http://www.sciencedirect.com/science/article/pii/S0140988317301871. https://doi.org/10.1016/j.eneco.2017.05.023.
[38] Tang L, Wu Y, Yu L. A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting. Energy 2018;157:526–38. http://www.sciencedirect.com/science/article/pii/S0360544218309800. https://doi.org/10.1016/j.energy.2018.05.146.
[39] Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. Int J Forecast 1997;13:281–91. http://www.sciencedirect.com/science/article/pii/S0169207096007194. https://doi.org/10.1016/S0169-2070(96)00719-4.
[40] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econ Stat 1995;13:253–63. https://amstat.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599. 10.1080/07350015.1995.10524599.
[41] Sax C. seasonal: R Interface to X-13-ARIMA-SEATS. 2017. https://CRAN.R-project.org/package=seasonal. r package version 1.6.1.
[42] Bureau UC. X-13ARIMA-SEATS reference manual. 2017. https://www.census.gov/ts/x13as/docX13AS.pdf. version 1.1.
[43] Kang W, Ratti RA, Vespignani JL. Oil price shocks and policy uncertainty: new evidence on the effects of us and non-us oil production. Energy Econ 2017;66:536–46. http://www.sciencedirect.com/science/article/pii/S0140988317300373. https://doi.org/10.1016/j.eneco.2017.01.027.
[44] Choi I. Testing the random walk hypothesis for real exchange rates. J Appl Econom 1999;14:293–308. https://doi.org/10.1002/(SICI)1099-1255(199905/06)14:3<293::AID-JAE503>3.0.CO;2-5.
[45] Kim JH. Automatic variance ratio test under conditional heteroskedasticity. Financ Res Lett 2009;6:179–85. https://doi.org/10.1016/j.frl.2009.04.003.
[46] Alquist R, Kilian L, Vigfusson RJ. Chapter 8 - forecasting the price of oil. In: Elliott G, Timmermann A, editors. Handbook of economic forecasting. Elsevier. Volume 2 of handbook of economic forecasting; 2013. p. 427–507. http://www.sciencedirect.com/science/article/pii/B9780444536839000086. https://doi.org/10.1016/B978-0-444-53683-9.00008-6.
[47] Bontempi G, Ben Taieb S, Le Borgne YA. Machine learning strategies for time series forecasting. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 62–77. https://doi.org/10.1007/978-3-642-36318-4_3. 10.1007/978-3-642-36318-4_3.
[48] Lahouar A, Slama JBH. Hour-ahead wind power forecast based on random forests. Renew Energy 2017;109:529–41. http://www.sciencedirect.com/science/article/pii/S0960148117302550. https://doi.org/10.1016/j.renene.2017.03.064.
[49] Dudek G. Short-term load forecasting using random forests. In: Filev D, Jabłkowski J, Kacprzyk J, Krawczak M, Popchev I, Rutkowski L, Sgurev V, Sotirova E, Szynkarczyk P, Zadrozny S, editors. Intelligent Systems'2014. Cham: Springer International Publishing; 2015. p. 821–8.
[50] Sensoy A, Hacihasanoglu E. Time-varying long range dependence in energy futures markets. Energy Econ 2014;46:318–27. http://www.sciencedirect.com/science/article/pii/S0140988314002497. https://doi.org/10.1016/j.eneco.2014.09.023.
[51] Tabak BM, Cajueiro DO. Are the crude oil markets becoming weakly efficient over time? a test for time-varying long-range dependence in prices and volatility. Energy Econ 2007;29:28–36. https://doi.org/10.1016/j.eneco.2006.06.007.