Technical Indicators for Hourly Energy Market Trading

Catherine McHugh, Sonya Coleman, Dermot Kerr
Intelligent Systems Research Centre (ISRC)
Ulster University
Northern Ireland, UK
Email: {mchugh-c24; sa.coleman; d.kerr}@ulster.ac.uk

***Abstract*—Financial trading often combines machine learning and technical indicators to accurately predict future market prices. Energy data and financial data have similar features; therefore, this research derives eight electricity price technical indicators to help control spending and reduce trading costs for the Integrated Single Electricity Market in Ireland. The proposed technical indicators were derived from electricity price data, collected on an hourly basis from February until November 2019, and used to train three regression machine learning algorithms (Random Forest, Gradient Boosting, and Extreme Gradient Boosting). The results for each of the regression algorithms were first compared using one model for all trading periods. The Random Forest algorithm was then trained with the same technical indicators for each of the 24 hours periods individually to see if an hourly approach enhanced model performance. The proposed technical indicators accurately predict electricity prices and overall accuracy was greatly improved using separate hourly forecasting models.**

Keywords—Hourly Forecasting, Machine Learning, Technical Indicators, Energy Market

I. Introduction

Energy data display volatile characteristics that make forecasting in the energy market difficult [1]. Time series models analyse patterns by observing and training with previous prices to predict future values. Price prediction machine learning algorithms are an increasingly popular tool to tackle volatility and reduce trading costs by creating optimal price models [2]. Price fluctuations arise when supply and demand are imbalanced, but price forecasting can optimise unit purchasing especially when short-term forecasting as the relationship is stronger between actual and predicted values [3]. This research centres on day-ahead electricity price prediction to examine energy market trends, with the overall aim of building an innovative system that assists electricity suppliers in future planning to reduce purchasing costs and hence enables consistent pricing for customers.

In the financial trading market, simple machine learning algorithms have been quite effective when used for prediction [4]. Technical indicators originate from historical financial data and are often used as inputs to train machine learning forecasting models. Generally technical indicators are mostly considered to aid investors in whether it is best to buy or sell in the trading market [5]. This approach could be applied to the day-ahead energy market by developing specific technical indicators that follow electricity price trends and including these derived indicators as inputs in prediction models to forecast future electricity prices. Technical indicators have only recently been applied in the energy market for day-ahead forecasting, therefore existing literature in this area is limited [6]. Energy forecasting models that apply fundamental indicators as inputs (load, weather variables, generation, etc.) have found that same hour input data have robust correlation [7]. This conclusion was also stated when separate 24-hour time-series models were applied to the Spanish electricity price market data, noting that separate hour models were more homogeneous in observing trends than a model that considers all hours [8].

This paper develops eight new technical indicators specifically for the energy market, building on the idea from [6] of calculating the indicators for each hour separately. First we examine our technical indicators by including them as inputs and using the actual electricity price as output in machine learning regression models. The forecasting performance for day-ahead predictions is evaluated for all techniques and all 24-hours are included in the training models. The model performance accuracies are then compared with the performance of models that are trained only on raw price data as input, denoted as persistence models, to determine if including technical indicators as inputs improves model performance. This research then examines 1-hourly models for each of the 24 hours to determine if technical indicators do follow hourly patterns and are therefore better at matching market trends when split by hour.

This paper is organised as follows: Section II outlines each of the proposed technical indicators for energy market prediction and describes how each is calculated in terms of electricity price. Section III discusses the three regression algorithm modelling techniques. The results are presented and discussed in Section IV highlighting first, the accuracy of each regression 24-hour model, and then displaying the results for each of the individual 24 1-hour Random Forest models. Section V concludes the paper by summarising the key findings of this research.

II. Technical Indicators

Technical analysis is common in stock market trading to capture trends and information on price movement from indicators built using raw stock price [9]. The core indicators for price prediction are: (i) trend, (ii) oscillator, and (iii) momentum [5]. The most complex part of technical analysis is deciding on parameter optimization and the sliding window size is a key feature as it relates to the corresponding number of historical records required for the calculation of each indicator [10]. A new advancement for the island of Ireland is the Integrated Single Electricity Market (ISEM) allowing energy traders greater control. This development has led to the need for novel technical price indicators to aid in forecasting decisions when to buy or sell in the ISEM.

Our research presents eight innovative energy trading technical indicators originating from, but not the same as, the common financial technical indicators. The requirement for the ISEM is day-ahead, therefore this work only focusses on technical price indicators which improve day-ahead prediction accuracy. The individual calculations for each of the indicators used in both the all hours model and separate hourly models are listed below:

1. Percentage Price Change Moving Average (PPCMA): A trend indicator in time-series that, for the energy market, we calculate price change as the difference between the current price (Hour *n*) and the price from the same time period the day before (Hour *n* Lag 24), all divided by the price at Hour *n* Lag 24. In the all hours model the moving average percentage price change was calculated for is a rolling 24-hour window. For the hourly models a pool (*i*) ranging from a rolling 1-hour window to a 100-hour window was calculated:

$$PPCMA_i = \sum_i [PPC] \tag{1}$$

where

$$PPC = \frac{Price_{Hour\,n} - Price_{Hour\,n\,Lag\,24}}{Price_{Hour\,n\,Lag\,24}} * 100 \tag{2}$$

2. Moving Average Deviation (MAD): A trend indicator that utilises the PPCMA indicator to calculate the deviation rate of the current electricity price from PPCMA. For the hourly models a pool (*i*) ranges from 1 to 100:

$$MAD_i = \frac{Price_{Hour\,n} - PPCMA_i}{PPCMA_i} \tag{3}$$

3. Percentage Range (PR): An oscillator indicator that finds a relationship between current electricity price and the highest/lowest prices over a 24-hour window for the all hours model. For the hourly models a pool (*i*) ranges from 1 to 100 to calculate the highest and lowest prices. This indicator oscillates between 0 and 100, with a value above 80 determined to indicate energy units are oversold and a value below 20 indicating that energy units are overbought:

$$PR_i = \left[\frac{HighestPrice_i - Price_{Hour\,n}}{HighestPrice_i - LowestPrice_i}\right] * 100 \tag{4}$$

4. Average True Range (ATR): A trend indicator measuring price volatility. Over a 24-hour window there are three different values calculated for the all hours model: highest price over the 24-hour period minus lowest price over the 24-hour period; highest price over the 24-hour period minus starting electricity price; and lowest price over the 24-hour period minus starting electricity price. The maximum value from these three values is selected for each trading hour and averaged over a rolling 24-hour window. For the hourly models a pool (*i*) ranging from a rolling 1-hour window to a 100-hour window were used in the calculations:

$$ATR_i = \sum_i MAX\ [\, A_i, B_i, C_i \,] \tag{5}$$

$$A_i = HighestPrice_i - LowestPrice_i \tag{6}$$

$$B_i = |\, HighestPrice_i - Price_{Hour\,n} \,| \tag{7}$$

$$C_i = |\, LowestPrice_i - Price_{Hour\,n} \,| \tag{8}$$

5. Relative Strength Index (RSI): An oscillator indicator that compares recent price gains to recent price losses. This indicator oscillates between 0 and 100, with a value over 70 determined to indicate that energy units are

overvalued and a value below 30 indicating that energy units are undervalued. For the all hours model, Price Up is the average of the previous 24 hours when price difference increased, and Price Down is the average of the previous 24 hours when price difference decreased. For the hourly models, Price Up and Price Down are calculated from the average of the previous *i* hours with *i* ranging from 0 to 100:

$$RSI_i = 100 - \left[\frac{100}{D_i}\right] \tag{9}$$

where

$$D_i = \left(1 - \frac{\sum_i Price\ Up\left[Price_n\ -\ Price_{Hour\ n\ Lag\ i}\right]}{\sum_i Price\ Down\left[Price_n\ -\ Price_{Hour\ n\ Lag\ i}\right]}\right) \tag{10}$$

6. Average Directional Movement Index (ADX): A trend indicator measuring the strength of the trend, grouping the two directional movement indexes depending whether price change, calculated as current electricity price minus previous 24-hour price (all hours model)/previous i-hour price (hourly models), is grouped as a Price Up (positive) change or Price Down (negative) change. The two indexes are combined and smoothed with a moving average:

$$ADX_i = \left[\frac{\sum_i DX\ Up(a_i)\ -\ \sum_i DX\ Down(b_i)}{\sum_i DX\ Up(a_i) +\ \sum_i DX\ Down(b_i)}\right] * 100 \tag{11}$$

where

$$a_i = \frac{\sum_i Price\ Up[Price_n\ -\ Price_{Hour\ n\ Lag\ i}]}{ATR_i} \tag{12}$$

$$b_i = \frac{\sum_i Price\ Down[Price_n\ -\ Price_{Hour\ n\ Lag\ i}]}{ATR_i} \tag{13}$$

7. Moving Average Convergence/Divergence (MACD): An oscillator indicator that considers the strength, direction, and duration of the trend as well as price momentum through moving averages of previous price values with rolling window sizes of 12 and 24 for the all hours model and rolling window sizes of 7 and 14 for the hourly models:

$$MACD = \sum_7 Price\ MA_{Hour\ n\ Lag\ 7} - \sum_{14} Price\ MA_{Hour\ n\ Lag\ 14} \tag{14}$$

8. Price Momentum (PMOM): A momentum indicator that measures the power of the market by observing the current electricity price with the previous trading value (1 hour before) for the all hours model. For the hourly models a pool (*i*) ranges from 1 to 100 to calculate:

$$PMOM_i = Price_{Hour\ n} - Price_{Hour\ n\ Lag\ i} \tag{15}$$

III. Modelling Methodology

Three machine learning algorithms were trained with the eight technical indicators, implemented through SkLearn. A Random Forest regression algorithm is an efficient ensemble technique with many benefits: straightforward tuning, robust to outliers, and expandable for data fitting [11]. During training of a Random Forest there are multiple trees split at nodes therefore no single tree perceives the complete training dataset [12]. There is transparency with the algorithm as a tuning parameter decides when to split the input data for classifying [13]. After the Random Forest is built a prediction value is outputted, which is the average of each individual regression tree's prediction [14]. In this research a Random Forest regression algorithm was implemented with 1000 trees.

Sequential learning, used in boosting algorithms, combines weak learner models to create one strong learner model [15]. In a Gradient Boosting regression algorithm a prediction model is built from weak learners through optimizing a loss function [16]. Another boosting regression algorithm is the Extreme Gradient Boosting (XGBoost) that works through ensemble sequential learning with weighted predictors [16]. This is an advanced machine learning algorithm due to speed and the ability to train large data [17]. In this research an XGBoost algorithm was implemented with 1000 trees, the fraction of column to be a random tree sample was set to 0.6, the fraction of observations to be random tree subsample was set to 0.8, the maximum depth of tree was set to 4, with a learning rate of 0.05.

The hourly models contain different technical indicator parameters depending on the hour (0-23). As a pool of indicators ranging from 1 to 100 were created for this part of the research, selecting the optimal indicators for each hour is the first step. Hyperparameters *n* and *s* represent the lag factor and span respectively. This approach of finding an optimal *n* and *s* was taken from the work presented in [6]. In our research *n* was utilized in the creation of five of our novel technical indicators (PR, ATR, RSI, ADX, and PMOM) and *s* was utilized in the creation of two of our novel technical indicators (PPCMA and MAD). To find the optimal *n* and *s* for each hour, the model which provided the lowest Root Mean Square Error (RMSE) during the testing set was selected. This is calculated in terms of electricity price as the difference between actual and predicted price values. Optimization was implemented on SkLearn by creating a list with all possible combinations for *n* and *s* and ranking the RMSE for each combination in order from lowest to highest RMSE.

Table I displays the optimal *n* and *s* values for each of the 24 1-hour models. The corresponding technical indicators are chosen based on optimal *n* and *s*. For instance, the optimal technical indicators for Hour 0 are PPCMA45, MAD45, PR48, ATR48, RSI48, ADX48, MACD, and PMOM48. If *n* or *s* reached the maximum value of 100 the indicator pool range was increased to 150 to ensure the optimal model was selected.

TABLE I. OPTIMAL N AND S FOR HOURLY MODELS

Hour	Optimal n	Optimal s
0	48	45
1	100	42
2	61	74
3	59	5
4	59	76
5	74	75
6	100	99
7	99	75
8	91	93
9	2	97
10	87	82
11	75	76
12	78	77
13	41	7
14	95	97
15	83	71
16	87	82
17	106	80
18	87	94
19	102	99
20	102	106
21	56	38
22	81	81
23	55	55

IV. RESULTS

For the experiments we use hourly ISEM electricity price data ranging from 1st February 2019 until 01st December 2019 retrieved from SEMOpx website [18]. The technical indicators derived from the raw price data, outlined in Section II, were first calculated using all data and a sliding window of 24 hours. Additionally, the indicators were calculated for each hour separately and this time a sliding window pool from 1 to 100 was utilised for each indicator to find the optimal *n* and *s*. The calculated technical indicators were applied as training inputs for the three machine learning models.

The data for the 24-hour machine learning models were split 85% for training (04th February - 16 October 2019) and 15% for testing (17th October – 30th November 2019). The input data were the eight technical indicators and the output data was the actual electricity price for the same timeframe, time *T*. To set the work of technical indicators in context a persistence model with raw electricity price data as input was observed as a baseline at time *T* and output price data at time *T*+24. Table II presents comparative results for both persistence and technical indicator 24-hour models for Random Forest, Gradient Boosting, and XGBoost.

TABLE II. SUMMARY RESULTS FOR 24-HOUR MODELS

Algorithm	Persistence Models		Technical Indicators	
	Testing Accuracy	Testing RMSE	Testing Accuracy	Testing RMSE
Gradient Boosting	75.44%	13.18	86.90%	6.66
Random Forest	73.21%	16.30	91.57%	6.77
XGBoost	75.02%	14.39	89.70%	5.34

Testing model accuracy percentage is computed as model error subtracted from 100. The model error metric chosen for the calculation is the Root Mean Squared Log Error (RMSLE) as it is robust to outliers, only observes relative error, and gives a larger penalty for underestimating [19]. From Table I the testing accuracy ranged between 73% and 76% for the persistence models and ranged between 86% and 92% for the technical indicator models highlighting that using technical indicators as inputs improves model performance. RMSE evaluates the model performance of the test set, the closer the value is to zero the better the prediction. During model testing XGBoost provided the lowest RMSE value of 5.34. However, the model accuracy was the highest at 91.57% using the Random Forest. Figure 1 illustrates the actual electricity price values plotted against the predicted price values for the Random Forest testing phase. The figure exhibits a good-fit for the majority of predicted values especially at the beginning of the testing period, but the fit is less accurate during the last few hours of the testing period.

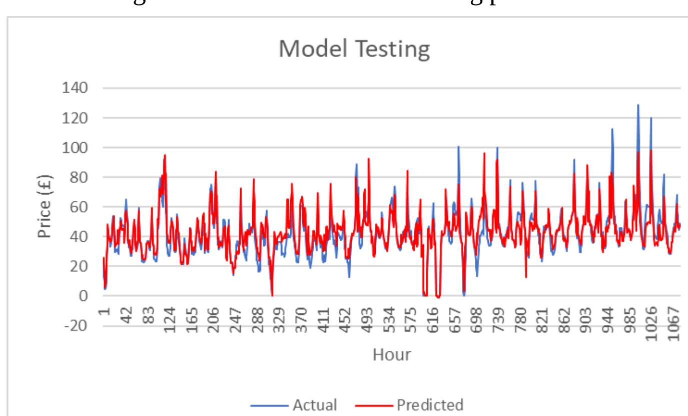

Fig. 1. Random Forest 24-Hour Model Testing

The next stage of this work was to split the data by hour before calculating separate hourly technical indicators to determine if models are better trained as 24 separate 1-hour prediction models. As well as hourly technical indicators there was also a pool of indicators with varying *n* and *s* for each hour in order to select the optimal hourly models. The technical indicators were calculated from data ranging between February and December 2019 and used as model inputs. From the previous work it was clear that no matter which machine learning algorithm was used, the use of technical indicators improved prediction performance. Therefore, we only create hourly models using the Random Forest algorithm as it presented the highest model accuracy in Table II. The Random Forest hourly models used 85% of the dataset for training (11th May - 31 October 2019) and 15% for testing (01st November – 01st December 2019).

Table III presents the optimal model testing accuracy and RMSE values for each hour separately.

TABLE III. SUMMARY RESULTS FOR RANDOM FOREST HOURLY OPTIMAL MODELS

Hour	Testing Accuracy	Testing RMSE
0	88.16%	1.55
1	98.17%	0.98
2	90.26%	0.81
3	84.78%	0.77
4	86.28%	0.87
5	89.33%	0.75
6	91.16%	3.16
7	98.74%	1.28
8	99.32%	0.74
9	98.92%	1.19
10	98.46%	1.99
11	98.64%	1.69
12	98.82%	1.47
13	98.02%	2.38
14	97.89%	2.42
15	98.50%	1.74
16	98.04%	3.24
17	94.62%	11.86
18	97.65%	4.6
19	98.58%	1.79
20	98.70%	1.44
21	98.32%	1.95
22	98.32%	1.59
23	86.87%	1.75

Overall the testing accuracy ranged from 84% to over 99% and most of the testing RMSE values were below 3. This is a significant improvement from the testing results shown in Table II. Hour 8 had the most promising results with a testing accuracy of 99.32% and a RMSE value of 0.74. The visual output of the actual and predicted electricity prices for this hour are displayed in Figure 2 illustrating an excellent fit. As this plot is for a 1-hour model (Hour 8) the x-axis (Figure 2) is the same hour in each day in the testing period, whereas in the all hours technical indicator model the x-axis (Figure 1) is every hour in the testing period.

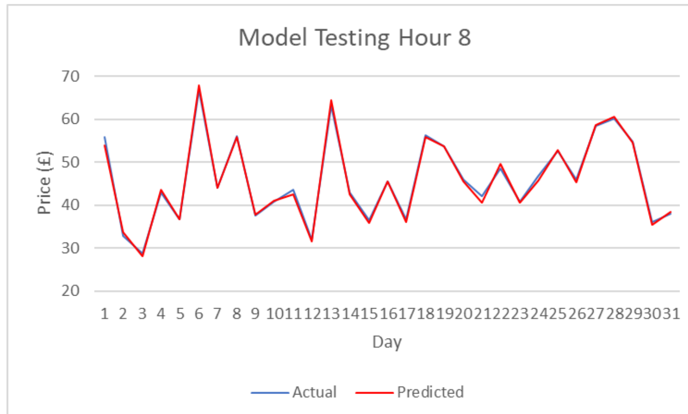

Fig. 2. Random Forest Hour 8 Model Testing

V. CONCLUSION

Eight technical indicators (PPCMA, MAD, PR, ATR, RSI, ADX, MACD, and Price Momentum) were specifically derived from raw data for energy trading and tested on three machine learning regression models (Random Forest, Gradient Boosting, and XGBoost) to forecast electricity prices. The technical indicators were first calculated using an all hours approach and then they were re-calculated when split by hour to find an optimal hourly electricity price forecasting model.

In both approaches the model data were split 85% for training and 15% for testing. In the 24-hour model approach the results were compared with a baseline persistence model which was tested with raw price data only. The three algorithms accuracy ranged between 73% and 76% for the persistence models and ranged between 86% and 92% for the technical indicator models. These results confirmed that including technical indicators as model inputs improved overall performance.

In the next experiment stage 24 separate 1-hour prediction models were generated using the Random Forest algorithm. Random Forest was selected here as in previous results it resulted in the highest model accuracy. Optimal *n* and *s* were required for each of the 24 1-hour Random Forest models split by hour and chosen through running each indicator pool combination and selecting the hyperparameters which result in the lowest RMSE. The testing accuracy ranged between 84% to over 99% for each of the 24 1-hour models and the majority had a RMSE value below 3. These promising results indicate that having individual hour models are more homogeneous and beneficial for energy trading.

To conclude, energy traders should consider technical indicators in price prediction models, especially individual models that have been optimised for each hour of the day, to capture market trends and enable accurate predictions, thus reducing purchasing costs. Further work will consider adding other energy related factors such as wind generation to the optimal models to determine if model accuracy can be further improved.

ACKNOWLEDGMENT

This research was supported by DfE CAST scholarship in collaboration with Click Energy.

REFERENCES

[1] H. Mosbah and M. El-Hawary, "Hourly Electricity Price Forecasting for the Next Month Using Multilayer Neural Network," in *Canadian Journal of Electrical and Computer Engineering*, 2016, vol. 39, no. 4, pp. 283–291.

[2] G. Gao, K. Lo, and F. Fan, "Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market," *Energy Power Eng.*, vol. 09, no. 04, pp. 120–126, 2017.

[3] N. Pandey and K. G. Upadhyay, "Different price forecasting techniques and their application in deregulated electricity market : A comprehensive study," in *International Conference on Emerging Trends in Electrical , Electronics and Sustainable Energy Systems (ICETEESES)*, 2016, pp. 1–4.

[4] E. A. Gerlein, M. McGinnity, A. Belatreche, and S. Coleman, "Evaluating machine learning classification for financial trading: An empirical approach," in *Expert Systems with Applications*, 2016, vol. 54, pp. 193–207.

[5] M. T. Ã and S. Tokuoka, "Adaptive use of technical indicators for the prediction of intra-day stock prices," in *Statistical Mechanics and its Applications*, 2007, vol. 383, pp. 125–133.

[6] S. Demir, K. Mincev, K. Kok, and N. G. Paterakis, "Introducing technical indicators to electricity price forecasting: A feature engineering study for linear, ensemble, and deep machine learning models," *Appl. Sci.*, vol. 10, no. 1, 2020.

[7] P. Li, F. Arci, J. Reilly, K. Curran, and A. Belatreche, "Using Artificial Neural Networks to predict short-term wholesale prices on the Irish Single Electricity Market," in *2016 27th Irish Signals and Systems Conference (ISSC)*, 2016, pp. 1–10.

[8] C. García-Martos, J. Rodríguez, and M. J. Sánchez, "Mixed models for short-run forecasting of electricity prices: Application for the Spanish market," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 544–552, 2007.

[9] J. Diego, J. Ignacio, J. Francisco, and L. Jose, "Multiobjective Optimization of Technical Market Indicators," in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference*, 2009, pp. 1999–2004.

[10] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, "Forecasting price movements using technical indicators: Investigating the impact of varying input window length," in *Neurocomputing*, 2017, vol. 264, pp. 71–88.

[11] J. Mei, D. He, R. Harley, T. Habetler, and G. Qu, "A random forest method for real-time price forecasting in New York electricity market," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2014-October, no. October, 2014.

[12] L. Khaidem, S. Saha, and S. R. Dey, "Predicting the direction of stock market prices using random forest," vol. 00, no. 00, pp. 1–20, 2016.

[13] K. Mulrennan, J. Donovan, D. Tormey, and R. Macpherson, "A data science approach to modelling a manufacturing facility's electrical energy profile from plant production data," *Proc. - 2018 IEEE 5th Int. Conf. Data Sci. Adv. Anal. DSAA 2018*, pp. 387–391, 2018.

[14] J. Pórtoles, C. González, and J. M. Moguerza, "Electricity Price Forecasting with Dynamic Trees: A Benchmark Against the Random Forest Approach," *Energies*, vol. 11, no. 6, 2018.

[15] R. Gandhi, "Boosting Algorithms: AdaBoost, Gradient Boosting and XGBoost," 2018. [Online]. Available: https://hackernoon.com/boosting-algorithms-adaboost-gradient-boosting-and-xgboost-f74991cad38c. [Accessed: 06-Jan-2020].

[16] S. Dey, Y. Kumar, S. Saha, and S. Basak, "Forecasting to Classification : Predicting the direction of stock market price using Xtreme Gradient Boosting Forecasting to Classification : Predicting the direction of stock market price using Xtreme Gradient Boosting," no. October, pp. 1–10, 2016.

[17] M. Pathak, "Using XGBoost in Python," 2019. [Online]. Available: https://www.datacamp.com/community/tutorials/xgboost-in-python. [Accessed: 06-Jan-2020].

[18] SEMOpx, "Day-Ahead Market data," 2019. [Online]. Available: https://www.semopx.com/market-data/market-results/. [Accessed: 10-Dec-2019].

[19] S. Saxena, "What's the Difference Between RMSE and RMSLE?," *Analytics Vidhya*, 2019. [Online]. Available: https://medium.com/analytics-vidhya/root-mean-square-log-error-rmse-vs-rmlse-935c6cc1802a. [Accessed: 08-Jul-2020].