The Suitability of Sarimax Time Series and LSTM Neural Networks for Predicting Electricity Consumption in Ireland

MSc Research Project
MSc Data Analytics

Edmond Connolly
Student ID: 16126556

School of Computing
National College of Ireland

Supervisor: Majid Latifi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Edmond Connolly
Student ID:	16126556
Programme:	MSc Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Majid Latifi
Submission Due Date:	16/08/2021
Project Title:	The Suitability of Sarimax Time Series and LSTM Neural Networks for Predicting Electricity Consumption in Ireland
Word Count:	5904
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Edmond Connolly
Date:	23rd September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

The Suitability of Sarimax Time Series and LSTM Neural Networks for Predicting Electricity Consumption in Ireland

Edmond Connolly
16126556

Abstract

The recent decade has focused attention more than ever before on the increasing carbon emissions and detrimental effect energy consumption has on the world we live in. Smart Grids have brought a new level of control and a new capacity to understanding this activity. A vital component of addressing energy consumption is the ability to accurately forecast it. Through this mitigating waste and anticipating the effect of future government policies may be achieved. This research will develop two predictive models and examine the capability of using meteorological data to forecast the daily electricity consumption in the Republic of Ireland for a two month period in 2020. Input features are historical energy consumption and meteorological data from the period 2014 to the end of 2020. It comprises 2558 rows of 64 columns of data. A Sarimax time series and a Long Short Term Memory (LSTM) Neural Network were employed to forecast electricity demand in the short term (60 days). The Long Short Term Memory Network achieved a higher degree of accuracy in forecasting the electricity consumption of the test period than the Sarimax model achieved.

1 Introduction

The following paper will seek to determine an accurate means of predicting energy consumption in the Republic of Ireland. At the time of writing Ireland is undergoing considerable change in regard to its energy provision. Infrastructure is being upgraded to facilitate the contribution of renewable sources of energy into the electrical grid. An accurate means of predicting energy consumption is of paramount importance in being able to meet current renewable energy obligations and commitments Ireland has entered into with our EU partners. The Climate Action Plan was entered into by the Irish government in 2019. This commits the Irish state to achieving a target of 70% of all electricity demand being met by renewable sources. (http://www.eirgridgroup.com/annual-report-2019/).

The use of energy forecasting in developing energy efficiency measures (EEMs) are a significant component in the challenge of meeting these targets. In order to develop measures to successfully counteract wasteful energy expenditure, accurate predications of future growth must be considered (Mishra et al. (2020)). The Sustainable Energy Authority of Ireland (SEAI) has been committed to developing incentive schemes to encourage energy efficiency in households over the last number of years. A number of grants have been created to achieve these goals. One such grant being the Home Energy

Saving (HES) scheme (Aravena et al. (2016)). Through measures such as this, reductions in per capita energy use can assist in the fight against ever increasing carbon emissions.

While some research into energy consumption has included embedded energy use, such as the energy consumed in produce bought by Irish citizens or been catagorised into distinct realms such as Residential,Travel and Work (Brophy et al. (2021)). This paper will focus purely on the electrical consumption of the country. This includes both household and industrial uses of electrical output.

The applications of energy forecasting have significance for policy formulation and strategic infrastructure development. The energy consumption of a country has a direct relationship with its economic growth. The relationship between economic growth and the development of the financial sector of a country and its energy consumption has been known for some time(Sadorsky (2010)). It is therefore beneficial to know how energy requirements might increase as economic development increases. Not unrelated to the economic development of a country is the fact that increased urbanisation of a country also leads to increased carbon emissions. The impact urban planning has on energy consumption and consequently on carbon dioxide emissions is yet another consideration that can be addressed adequately by accurate forecasting.

Multi-variate Statistical Forecasting in this paper will take the form of a Time Series analysis on data provided by Eirgrid. This is the company responsible for operating the Irish electrical grid. This data is from the period 2014 through 2020. A number of techniques will examine trends and seasonality and the impact this has on forecasting. An Seasonal Auto Regressive Integrated Moving Average with Exogenous Time series Analysis(SARIMAX) will be performed to predict energy demand at a given time. This is a variation on the Auto-Regressive Integrated Moving Averages(ARIMA) time series to facilitate the inclusion of a seasonal element and exogenous variables. A time series is a particularly good predictor of future behaviour as any influence of significant independent variables are included in the movements of data points between periods. The influence of these variables are contained in the values as they change between periods. In addition to this algorithm a Bi-Directional Long Short Term Memory Neural Network will also be trained. This form of Recurrent Network is a common solution to modelling sequential time series data.

The research question is as follows: *Whether historical Irish electricity consumption data combined with Irish meteorological data can accurately forecast future electricity demand in the short term?*

More accurate levels of predication can inform policy making and identify areas where wasteful energy consumption can be reduced. The following research objectives are considered:

1:*Conduct literature review of recent research in the domain.*
2:*Investigate data available and select appropriate features.*
3:*Determine the technological requirements of conducting the research and select requisite frameworks for implementation.*
4:*Implement the SARIMAX and LSTM models.*
5:*Evaluate and interpret the results with visual aids.*
6:*Critique the implementation and discuss further refinements or future work.*

After this introduction, Section 2 discusses work that has already been conducted in the domain of energy consumption and other domains where similar algorithms have been implemented. In Section 3, I propose the methodology being employed in order to model the acquired data. Section 4 will look at the models being employed and contain a description of each. Section 5 will focus on the implementation of the algorithms. The following section will examine the results of the experiment and discuss the implications of the results for the research (section 6).

2 Related Work

The application of machine learning within the electrical utility industry has inspired comprehensive research over the last decade. With the advent of smart grids for electrical power systems came the application of Artificial Intelligence. This technology has assisted in identifying faults, simulated controlling smart grids and provided health monitoring of wind generation systems (Bose (2017)).

Producing accurate Short Term Load Forecasts (STLF) can help in determining generational resource allocations, environmental and apparatus handling constraints and also identify the optimal operational state of the power grid. There are economic implications associated with the design of infrastructure and establishing working limits both of which benefit from accurate STLFs'. STLFs' are important determinants when ensuring Power Grid security, stability and reliability (Mishra et al. (2020)).

Through their research Khatoon et al. (2014) have determined that there are a number of feature types which are useful in multivariate forecasting of energy consumption. They have categorised these feature types as temporal, meteorological, economic, customer, random and other factors. This type of broad categorization of feature types that affect load prediction echoes the research of Xue and Geng (2012). This research categorized the features as being Long-Term Influence Factors, Middle-Term Influence Factors or Short-Term Influence Factors. It is the influence of Meteorological feature types that are the subject of this research. This research will try to determine their influence on the short term load demanded from the electrical grid. In the following section i discuss some of the current research in this domain and the techniques which are being employed.

2.1 Research in Time Series Forecasting and Prediction.

A number of studies have used the ARIMA method of forecasting to predict energy consumption. Akpinar and Yumusak (2016) use Arima Time Series Analysis to determine trends in the Turkish Natural Gas sector. They argue that the increase in computational complexity has allowed them to achieve a higher accuracy rate for their predictive model. To do this they employed times series decomposition, Holt Winters Exponential Smoothing and Auto Regressive Integrated Moving Averages (Arima). Their study found that these methods were successful in predicting energy consumption and had corresponding R^2 values of 0.915%,0.846% and 0.956%.

Aurna et al. (2020) performed a study on electricity demand in Kentucky using daily, weekly and monthly data from the years 2012 and to 2018 inclusive. They compared two models predictions. In their study they found that the Holt Winters model was a superior predictor than a regular Arima model. The results of this experiment have shown that the Holt Winters approach achieved an accuracy rate of 89.74%, 94.589% and 95.64% for daily weekly and monthly predictions. The equivalent rates for the Arima model were 88.51%, 89.814% and 90.85%. The metric used to determine the accuracy was a Mean Absolute Percentage Accuracy (MAPA) value.

There have also been Deep Learning approaches to energy demand prediction. Jana et al have performed research using first a Maximal Overlap Discrete Wavelet Transformation (MODWT) and then performing a Long -Short Term Memory (LSTM) network(Jana et al. (2020)). This research suggests that such methods outperform previous Time Series based methods.

Bu and Cho (2020) also used a Deep Learning model in their study. This study was developed on a residential energy consumption dataset from the University of California. It employed a Convolution Neural Network (CNN) and LSTM with Multi Headed Attention. The dataset measured energy consumption in 1, 15, 30, 45 minutes, 1 hour, 1 day and 1 week intervals. Bu et al found that the Mean Squared Error(MSE) of their predictions increased as the time interval being predicted increased. They compared the results with models built from the same dataset and achieved a lower MSE than all of them. These models included Linear Regression, Decision Trees, Random Forest, Multilayered Perceptron, Support Vector Regression and CNN-LSTM without Multi Headed Attention.

Another Deep learning approach to energy consumption prediction is studied in Khan et al's research(Khan et al. (2020)). This research used a hybrid model including a Multilayered Perceptron, Support Vector Regression and CatBoost. CatBoost is a library which optimizes deep-learning algorithms through gradient boosting. This research found favourable results and the authors state that the Root Mean Squared Logarithimic Error (RMSLE) is the lowest compared to the results of other models they implemented. The mean absolute percentage error (MAPE) recorded was 4.29%.

Rueda et al. (2019) conducted research in predicting daily energy requirement of buildings in the University of Granada. They used Symbolic Regression to predict the relationship between energy consumption for different business days of the week. Specifically they use Straight Line Programs(SLR) for the Symbol Regression. They state that their methods give greater interpretability to the modelling process than black box methods as the process selects the best predictors to use in forecasting accurate results in a way that is transparent to end users of the model such as managers or CEOs'. The SLR algorithm allows for the reduction of features used in prediction as only those features that are optimal are used.

Ilbeigi et al. (2020) also built a predictive model to gauge energy consumption in a building. Their research used a Multi-layer Perceptron/ Artificial Neural Network to model sensor data collected from a research building in Iran. This model also featured a Genetic Algorithm to assist in feature selection and optimise its performance. Their research suggests that up to 35% of the energy consumption of this building could be mitigated. Variables that were identified as of being of particular importance in the accurate prediction of energy consumption were the amount of occupants in the building and the degree of insulation the building had.

Khan and Byun (2020) used a combination of different machine learning models in research which analysed electricity consumption. The data they studied was Jeju islands electricity consumption. The models implemented were support vector regression(SVR), K-nearest neighbours and XGBoost. The performance of these models was optimized using a genetic algorithm. The research indicated that the mean absolute percentage error was 3.35% by using the ensemble model. This was a significant reduction in error when compared to each individual model acting in an independent capacity.

Alonso et al. (2020) used a Recurrent Neural Network to measure short term energy consumption loads. Their focus was on disaggregated consumption as might be found in individual buildings or households. This is as opposed to modelling substation energy consumption which would include multiple premises composed of commercial and residential buildings. They believed this approach had implications for smart metering purposes which they foresee as being of growing relevance in the near future. This was a Times Series based study which has the distinction of including exogenous variables. These variables were meteorological features such as temperature. Alonso et al compared three approaches in modelling this data. A Naïve approach which is forecasting the next 24 hours ahead using the previous 24 hours consumption, an Arima method and also the aforementioned Recurrent Neural Network. The comparative performance of these three models are as follows: The Arima method was 5% better than the Naïve approach while the RNN performed approximately 19% better than the Arima model. The metric in question is the MSE.

Chen et al. (2019) modelled energy consumption in the Chinese steel industry. They used a novel approach which made use of unlabelled data. This approach has been called 'Deep Learning embedded Semi Supervised Learning (DLeSSL). Ultimately a dataset containing unlabelled data was processed with what is referred to as a label propagation stage. This is the first of a two step process whereby the dataset which contains a small amount of labelled data and a large amount of unlabelled data, is processed to another 'enriched dataset'. The next step is 'label compensation' which involves a deep learning technique to determine the ultimate label to be applied. After this data is aggregated, Deep Learning models are applied to the new enriched dataset. This study showed that the above approach compared favourably with other methods including Linear Regression, Support Vector Machines and Decision trees. The MSE respectively is as follows: 1.254, 2.103, 1.709 and 1.946. Although it has to be stated that the authors found that for results to be favourable the ratio of unlabelled data to labelled data needs to be as high as possible.

Deng et al. (2019) used a multi step Convolution Neural Network to model electricity consumption in Ireland between 2014 and 2018. This is noteworthy because traditionally a recurrent neural network is used for time series modelling. Their results found that the CNN out performed the RNN and a number of other algorithms. CNN's are mainly used for classification purposes in fields such as artificial vision. Performing a regression with this algorithm is a novel application. A number of variations of CNNs' were used in this experiment. These included a Time Cognition Multi-Step CNN (TCMS-CNN), a Direct Multi-Step CNN(DM-MS-CNN) and a Direct Multi-Step Gated CNN (CNNDM-GCNN). The results of this research point to the TCMS-CNN as being the optimal in modelling the data having achieved an improved MAPE of 34.7%,14.2% and 19% on the LSTM, DM-MS-CNN and the GCNN respectively.

Gencer and Başçiftçi (2021) conducted research to predict vulnerabilities in the Android operating system by implementing a time series analysis. Using a LSTM neural network and ARIMA they modelled data from the American National Vulnerability Database. This data set has catalogued vulnerabilities and exposures to the operating system with identification numbers and collected the monthly instances over an eleven year period. The results of their research concluded that the ARIMA was model produced the less error but that the LSTM was a competitive alternative. The LSTM error rate was 26.8%RMSE while the ARIMA method was 18.5%RMSE.

Pacella and Papadia (2021) conducted an experiment in predicting Supply Chain Management using a LSTM and a BLSTM(Bidirectional LSTM). The BLSTM adds another layer which processes data in two directions ,forward and backwards, independently and connects to the output layer. Regular LSTMs' operate in one direction. Being able to gauge customer demand accurately facilitates efficient manufacturing activities and optimal stock control levels. This study monitored the market demand of ten products over the years 2015, 2016 and 2017. This data measured customer demand for the products in a monthly time frame. The research concluded that the BLSTM added an extra degree of accuracy in modelling the demand for all the ten products. Pacella et al also mention that LSTM without the Bidirectional component produced a competitive means of modelling the non-linear data.

The above literature review illustrates the many diverse methodologies in modelling energy consumption and time series data.The combination of different models and the optimizing of these models with additional algorithms is a feature of much of the contemporary research. A summary of significant research using Deep Learning or Arima/Sarimax models is presented in table 1. The table presents an overview of research conducted using these algorithms and a brief description of the domain and results of each study.

2.2 Comparison of Related Research

Table 1

Table 1: Summary of Literature Review

Literature			
Author	**Model/Method**	**Data**	**Performance**
Akpinar and Yumusak (2016)	Combined model/Arima/Holt-Winters Exponential Smoothing/Time Series Decomposition	Turkish Natural Gas Consumption	Improved R-squared value compared to individual methods.
Aurna et al. (2020)	Holt-Winters/Arima	Daily/Weekly/Monthly electrical consumption in Kentucky	Greater Accuracy with Holt-Winters determined by MAPA
Jana et al. (2020)	Combined MODWT - LSTM/Arima	Monthly energy consumption data USA 1973 -2018	Deep learning LSTM proved superior to Arima method
Bu and Cho (2020)	CNN/LSTM	Energy consumption 1/15/30/45 minute intervals.	Lower mse for the deep learning models than LR/DT/RF/SVR
Gencer and Başçiftçi (2021)	Arima/LSTM	Android vulnerability data	Lower mse with the Arima method than LSTM
Pacella and Papadia (2021)	BLSTM/LSTM	Supply Chain Data for predicting product demand	BLSTM produced an extra degree of accuracy compared to the LSTM
Alonso et al. (2020)	RNN/Arima/Naïve Approach	Short Term Load Forecasting Energy Consumption	RNN proved to be the best predictor with the Arima method being the second best.

3 Methodology

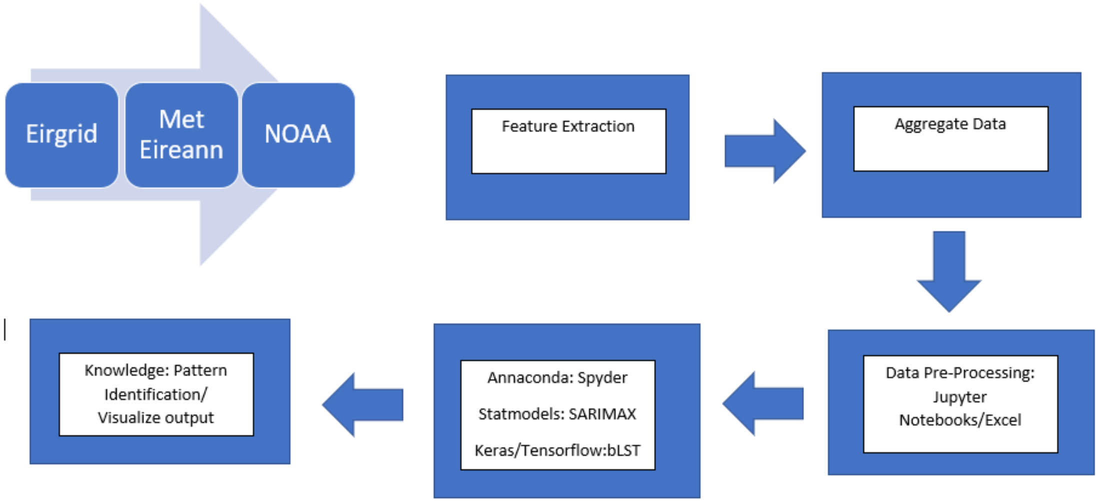

Figure 1: Process Diagram

3.1 Introduction

Figure 1 gives a high level view of the process implemented in this research. This involves the preliminary feature extraction and aggregation of the data, the pre-processing of the data, the final implementation of the models and finally the visualisation of the results of the research. The following subsections give a detailed explanation of how specific elements of these steps were implemented.

3.2 Data Sources

Data for the research was collected from three sources. Irish Energy consumption was collected form the Eirgrid website. This data ranges from 2014 to 2020. It is composed of a number of features in csv format. Data regarding energy consumed in Ireland is the only feature out of the dataset that was required. This data pertains to the whole of the republic of Ireland and is not broken down by region. It is in 15 minute intervals.

Meteorological data was gathered from the Met Eireann website for the time range previously mentioned. There are many weather stations collecting data throughout Ireland. Some stations provided more comprehensive data than others and it is these stations that were selected for this research. Weather data was collected from a number of weather stations throughout the country. In order to achieve a geographic spread of data a weather station was selected in each of Irelands Regional Authorities. These Regional Authorities were a means by which the government organised public services for the country. For the purposes of this research they are simply a means of associating weather variables with geographic areas.

Further data was collected from the National Oceanic and Atmospheric Administration. This organisation is based in the United Sates and is affiliated with the US Department of Commerce. Solar data can be collected from this site through a model which takes longitude and latitude values to calculate solar activity for a given region. From this data the daily amount of daylight that would occur in Ireland for each day throughout the data range of the research is obtained. Further features that relate to the Solar activity at Irelands' longitude and Latitude are also included.

3.3 Data Aggregation

The first step with this data was to aggregate it into a daily granularity. The preliminary processing of data used an environment created in Anaconda. Using the Pandas library the csv files from the Eirgrid website were amalgamated. Using Pandas I averaged the data into daily periods.

3.4 Data Pre-Processing

After creating a Pandas data frame from the data, the date field was set to be the index for the subsequent algorithms. As only the total energy demand of the Republic of Ireland was required, the remaining columns were removed so as not to have an extra overhead when running the Time Series.

The data was then examined for missing values. The only missing values that were uncovered related to some of the weather station data. Given that there were very few missing values, imputing the missing values was deemed sufficient to address this issue. The knn algorithm was selected as an appropriate method to achieve this imputation. The knn imputation was conducted using the Pandas library.

Using Microsoft Excel the corresponding days of the week were matched with the date range and then encoded for inputs to the algorithms. To achieve the encoding a Periodic Cyclic transformation was applied to the data(Mahajan et al. (2021)). The purpose of this encoding is that it preserves a relationship between the variables in its final encoded format. For example, in their encoded form Sunday would share a close proximity to Monday but be several degrees further from Thursday. This is something that cannot be achieved through one-hot encoding or by simply labelling the days of the week 1 through 7. The process of encoding days of the week in this manner creates two variables for each day of the week. The formula and output for this process is illustrated in figure 2.

```
data = [['Sunday', 0], ['monday', 1], ['tuesday', 2],['wednesday', 3],['thursday',4],['friday',5],['saturday',6]]
df_day = pd.DataFrame(data, columns=['day_name','day_number'])
```

```
df_day['sin_time']= np.sin(2*np.pi*df_day.day_number/7)
df_day['cos_time']= np.cos(2*np.pi*df_day.day_number/7)
df_day
```

	day_name	day_number	sin_time	cos_time
0	Sunday	0	0.000000	1.000000
1	monday	1	0.781831	0.623490
2	tuesday	2	0.974928	-0.222521
3	wednesday	3	0.433884	-0.900969
4	thursday	4	-0.433884	-0.900969
5	friday	5	-0.974928	-0.222521
6	saturday	6	-0.781831	0.623490

Figure 2: Encoding for Days of the Week

The features in this dataset were then analysed in a correlation matrix to examine the relationship each had with the total consumption demand for electricity in Ireland. The features with the strongest correlation to the dependant variable were then selected as predictors for the finalised dataset. An example of the correlation matrix can be seen in figure 3. The input features used in the final dataset are the features which showed the strongest correlation with the output variable. The correlation values were output to a csv and features with a correlation above 50% were used for the final dataset. The features with lower correlations were removed from the dataset. The reduced dataset has 64 features including the dependent variable. The same dataset was used for both the Sarimax and LSTM algorithms.

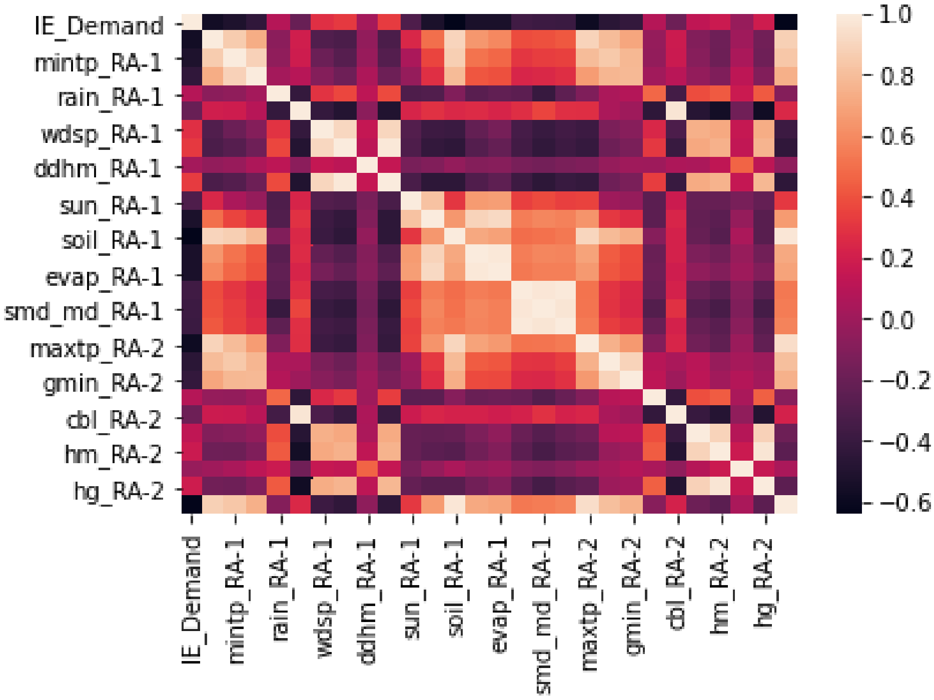

Figure 3: Correlation Matrix

3.5 Modelling

Once the finalised dataset was amalgamated and the missing values imputed it was loaded into an Anaconda python environment. The python packages used to perform the modelling were Statsmodels and Keras for the Sarimax and LSTM modelling respectively.

The first step when modelling the data for the Sarimax time series was to perform a Seasonal Decomposition. With this analysis the data could be examined for seasonal trends. A seasonal component can be seen to exist when certain patterns in data appear periodically. As a result of the seasonal decomposition a seasonal component was determined to exist. For this reason the selection of the Sarimax model was deemed to be the most appropriate for the data. Further examination of the data included Augmented Dickey-Fuller test for Stationarity, Autocorrelation Function to test for correlations with past values and a Partial Autocorrelation Function to determine correlations with past averages of residuals.

Once this was complete a training set up to the 2496 row of data was created. The remaining 60 rows of the data pertained to the testing set and corresponded with the months of November and December 2020. The Sarimax method from Pythons' Statsmodels package was then used to train and test the model.

The implementation of the Bi-Directional LSTM was conducted with the Keras library. Keras is a high-level Application Programming Interface (API) for the Tensorflow framework. It provides access to this deep learning framework in a more user friendly manner. Keras is supported by Tensorflow and has an almost official position as its API of choice.

Training and test datasets were compiled from the original data. The data was scaled using RobustScaler from the python library scikit-learn.

To evaluate the models performance the results were visualised. The actual values for electricity consumption were plotted alongside the models forecast results for comparison. The metrics used to evaluate the models were as follows: Mean Squared Error(1), Root Mean Squared Error(2), R Squared(3) and Mean Absolute error(4).

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{1}$$

$$\text{RMSErrors} \ = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}} \tag{2}$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{3}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{4}$$

4 Design Specification

Preliminary aggregation of the data was conducted in Microsoft Excel and pythons Pandas library in Jupyter Notebooks. In these environments feature extraction, missing value imputation and variable encoding were performed. Once the finalised datasets were completed the models were implemented in an Anaconda environment. The keras library was used to train the bLSTM Recurrent Neural Network on the tensorflow framework. For the SARIMAX time series the python package Statsmodels Sarimax was used.

4.1 Recurrent Neural Network

A Recurrent Neural Network is a specific kind of Neural Network optimized for use with sequential data. They have a temporal dimension which allows time and sequences to be processed in a manner which influences its predictive capability. This is analogous to memory in a human brain. A RNN is composed of at least three layers. An input layer, a hidden layer and an output layer. Each layer is composed of a number of perceptrons/nodes which take inputs, applies a weight, and then produces an output or outputs.A Recurrent Neural Network processes data in a loop as opposed to feedforward Neural networks which process one through a network. This feedback/temporal loop ultimately means a layers outputs also become inputs as data is processed. This looping occurs in the hidden layer of an RNN. It not only produces an output but also feeds back into itself. The neurons/nodes connect back to themselves through time. In effect they remember what was in that neuron previously. This process further adjusts the weights associated with the outputs and is illustrated in figure 4.

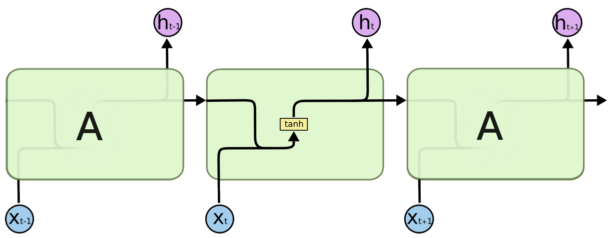

Figure 4: Recurrent Neural Network

A problem associated when trying to find the global minimum of the cost function can occur. The cost function can be said to compare the error from the output layer with the correct output. As data is processed through the neural network to the output layer an error is calculated and propagated back through the neural network to update the weights. This is known as the recurrent weight in an RNN.

As the weight propagates back through the network it can become smaller. As the weight is applied to itself when it passes through the hidden layers in declines. The lower this value is, the harder it is for the network to update the weights. What this means is that if there are numerous hidden layers, the layers near the output layer may be trained optimally while the earlier layers nearer the input layers may not be. The problem is that these earlier layers are inputs to the later layers. This creates a cycle that has a negative influence on the training of a model. Ultimately it will mean that the whole network is not trained in a correct manner. This is known as the vanishing gradient problem. Solutions to this problem can include selecting appropriate initial weights to counteract the gradient becoming too small. Another alternative solution is called a Long Short-Term Memory Network (LSTM).

4.2 LSTM

LSTM neural networks are one of the most commonly used Recurrent Neural Networks for sequential data (Sherstinsky (2020)). They address the issue of vanishing gradients through a means of *Uniform Credit Assignment* (Arras et al. (2019)). This is a process whereby a cells inputs are each assigned similar weights. Each cell receives an input vector. Within the neuron there is a input gate, a forget gate and an output gate (figure 5). The combined effect of these gates is to determine the weights that are significant and associate a value close to one for each. Through this process the weight of neurons in layers further away form the target neuron do not get reduced due to a vanishing gradient. In an LSTM the output of every layer is fed as an input to the next layer (Hewamalage et al. (2021)). This stacking approach is fundamental to achieving the temporal dimension to the LSTM. An LSTM inputs data from prior state time periods to be input simultaneously to a neuron along with the current state time period. Processing in this manner provides a *context* to the target neuron.

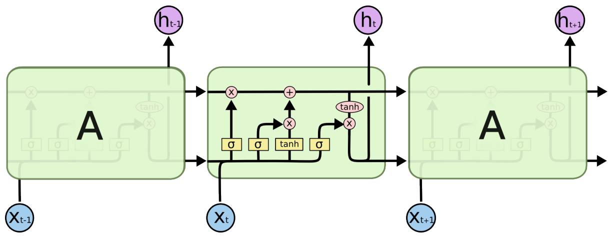

Figure 5: LSTM

In addition to the uni-directional LSTM as just described there are also Bi-directional LSTMs'. The only difference being that a bi-directional LSTM also includes time periods in the future of the target neuron while being trained. Input sequences are processed both in a forward and reverse manner at each time step before producing a single output. This means that twice the amount of weights are calculated at each time-step. Benchmarking research by Breuel (2015) suggests that Bi-Directional LSTMs' perform better on time sequenced data than regular LSTMs at all network sizes. It is this type of LSTM being employed in this research.

4.3 SARIMAX

SARIMAX is a variant of the Auto Regressive Integrated Moving Averages(ARIMA) model. The Autoregressive(AR) component relies on past period values and past periods only to predict current values (5). It is a linear model, where current period values are a sum of past outcomes multiplied by a numeric factor. AR uses the past value multiplied by a coefficient. The Moving-Averages(MA) component uses the residual multiplied by a coefficient instead (6). Also as previously mentioned the absolute value of the coefficient should be less than one. Otherwise the values will exponentially increase through a compounding effect similar to the exploding gradient problem in Recurrent Neural Networks. The more past values or lags the more complicated the model becomes and the more coefficients we have to determine and the more likely that some of them would not be significant. These lags can be thought of as the past data superimposed over the current data. Usually more data means a better prediction. But if the coefficients are close to zero/not significant they would have no effect on the predicted values. So they should not be included. We use the The ACF and PACF are used to help determine the optimal number of lags. The more lags that are used the more likely overfitting will occur. The integration component(I) is used for data that is non-stationary. By adding an integration step stationarity is introduced to the data. The SARIMAX model adds a seasonal component(S) and exogenous variables(X). In the Sarimax model the exogenous variables are added to the model in a multiple linear regression (Arunraj et al. (2016)). The seasonal component is added to the model in order to incorporate the presence of seasonal trends in the forecasting process. Exogenous' variables facilitate the use of outside variables in modelling an outcome (7). As regards the terminology used for the Sarimax models it is customary to use p for the ar component, d for the integration component and q for the ma component. Capitalised letters refer to these components for the exogenous variables. The order is referred to as (p,d,q)(P,D,Q,S) for inclusion in a Sarimax model.

$$\mathrm{Y_t = C + \phi_1 Y_{t-1} + \varepsilon_t} \tag{5}$$

$$\mathrm{Y_t = c + \beta_1 \varepsilon_{t-1} + \varepsilon_t} \tag{6}$$

$$\begin{gathered} Y_t = \beta_0 + \beta_2 X_{1,r} + \beta_2 X_{2,r} + \cdots + \beta_k X_{k,r} + \\ \left(\frac{\theta_q(B)\Theta_Q\left(B^S\right)}{\phi_p(B)\Phi_P\left(B^S\right)(1-B)^d\left(1-k^S\right)^D} \varepsilon_t \right) \end{gathered} \tag{7}$$

Following the implementation of the above models the results were plotted and the evaluation metrics calculated.

5 Implementation

The implementation of both the LSTM and SARIMAX models used the same finalised data set consisting of 64 features. This data set had missing values imputed and days of the week encoded. Unused features had been removed. Aside from scaling of the data for the LSTM there was no more processing need before the modelling took place. The data set was comprised of 2558 rows spanning data between the first of January 2014 through the thirty first of December 2020. This data was in a daily granularity.

5.1 The Implementation of the Sarimax Model

The preliminary steps in implementing this model involved reading a csv into the python environment in Spyder programming editor. Once this had been loaded the date column was set to datetime using the Pandas library and set as the index of the dataset. Auto correlation Function and Partial Auto-Correlation Function graphs were plotted to see if values correlated with past values and past period averages. After this the data was tested for stationarity using the Augmented Dickey-Fuller test. Following this the dataset was tested for seasonality using statsmodels STL package.

SARIMAX	
Parameter	Value
p	1
d	1
q	2
P	1
D	0
Q	2
S	7

Table 2: Sarimax Configuration

A training set of the first 2496 rows was then created in order to train the SARIMAX model. The remaining 60 rows were the testing set. The dates corresponding to the testing set were the first of November 2020 through the thirty first of December 2020. The parameters for the SARIMAX model were arrived at largely by trial and error. The ACF and PACF graphs were of some assistance in providing guidance for the ar and ma components value. The Augmented Dickey-Fuller test gave a value outside the critical value for stationarity with a 95% level of significance. For this reason the integration step was set to *1*. The seasonal component of *7* was selected as it is the earliest period where repetition occurs. Namely weekly. The model was implemented with parameters as in table 2.

Subsequent to the training of the model, a visualisation of the following two months forecast was plotted along with the actual values for the time period concerned.

5.2 The Implementation of the Bi-Directional Long Short Term Memory Neural Network

The initial loading of the data into the Anaconda environment followed the same process as above. As the LSTM algorithm can be prone to over-fitting a validation process was incorporated into the modelling. The training set also was also of a different size to the previous set to allow prior time periods to incorporated into the testing set. For this reason the training set included the the first 2466 rows of the dataset. Ten percent of the training dataset was allocated for validation purposes. The testing set was the remainder of the dataset and corresponded with the time period as above. The data was scaled using robustScaler. Each data set was scaled independently. This included all features apart form the Encoded Days values. The data was then reshaped into a 3 dimensional numpy array of time steps, batch size and features. It is this format that is required for Tensorflow LSTMs'. This was done separately for each of the train and test sets.

The LSTM was configured with two hidden layers of 128 neurons in addition to the input layers 128 neurons. The batch size was the default of 32 and the dropout was 0.2. It was optimized by the 'adam' algorithm. The activation function used was *tanh*, and the recurrent activation function was *sigmoid* which are the default parameters for the algorithm in the keras framework. The remaining parameters for the BLSTM are as illustrated in table 3.

A graph was plotted to measure the loss function of training and validation sets. This

bLSTM	
Parameter	Value
Input Layer	128
Hidden Layer 1	128
Hidden Layer 2	128
Dense Layer	1
Optimizer	adam
Batch Size	32
Epochs	35
Dropout	0.2
Loss	MSE

Table 3: LSTM Configuration

provided a visual reference to when the loss of each converged. This was the most important guidance in determining the optimal amount epochs.

The forecast from the model was again plotted along with the actual data from the relevant period after the test values had been rescaled.

6 Evaluation

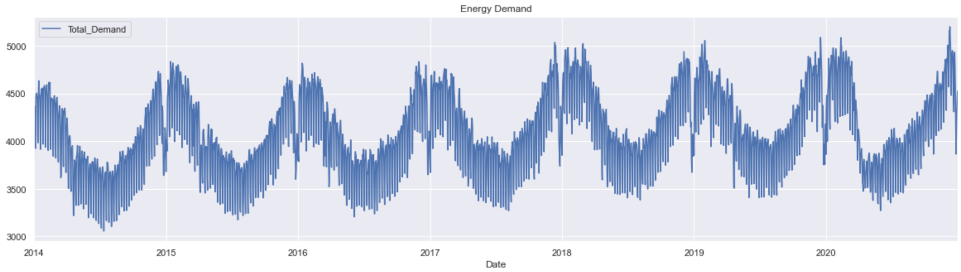

Figure 6: Eirgrid Data

6.1 SARIMAX

The Sarimax model for the sixty days of the test data achieved an R^2 value of 24% meaning only a modest amount of the variance of the feature space was explained by this model. The Mean Absolute Percentage Error was 3.58%. The RMSE value was 194MW and the Mean Absolute Error was 128MW. Running the same experiment and shortening the testing set to a 30 day period resulted in a much improved R^2 value 74%. Improvements in the other metrics were also observed with RMSE being 82.67MW and MAE being 66.34MW. The MAPE reduced considerably to 1.85%

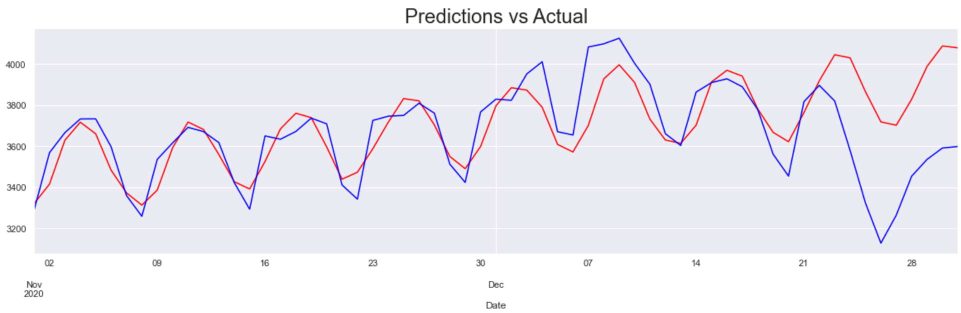

Figure 7: Sarimax

SARIMAX Results	
Parameter	Value
RMSE	194.13 MW
MAPE	3.58%
R^2	24%
MAE	128.28MW

Table 4: Sarimax Results

6.2 bLSTM

The Bi-Directional LSTM was trained for 35 epochs and achieved a MAPE of 1.85% and R^2 of 81%. The RMSE and MAE were 95.33 and 0.15MWs' respectively.

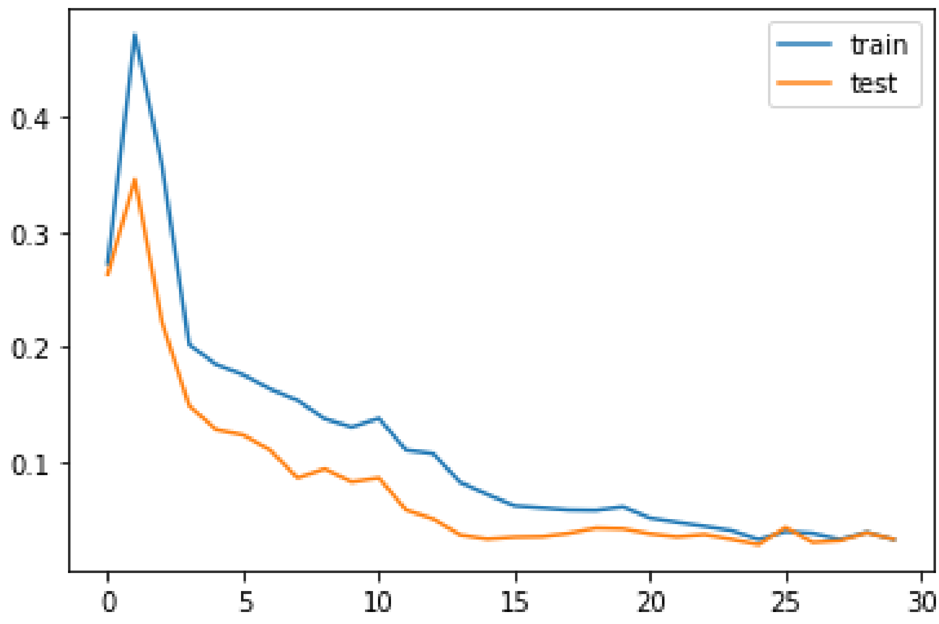

Figure 8: Validation Loss

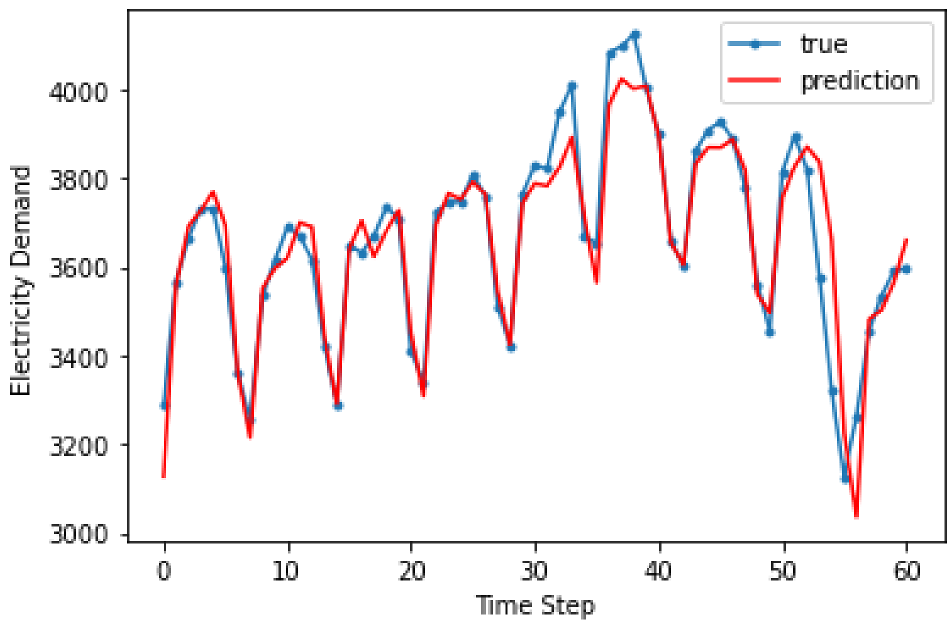

Figure 9: Actual Vs Predicted Electricity Demand

LSTM Results	
Parameter	Value
RMSE	95.33MW
MAPE	1.85 %
R^2	81%
MAE	0.15MW

Table 5: LSTM Results

6.3 Discussion

The use of a Bi-Directional LSTM for modelling a Time Series of electricity consumption has a number of advantages. First and foremost is its capacity to add a temporal dimension to its predictive abilities. This allows the algorithm to determine at what a point in a trend the current state is. At each point in the training of the model prior and future values corresponding to the output variable are present. Also this algorithm is specifically designed for sequenced data such as this research seeks to model. It has the capacity to map non-linear space and as such is able to determine relationships between variables that may not be captured by alternative methods such as the Sarimax model.

However the extra computational overhead associated with the Bi-Directional LSTM does mean that it is slower to train. It also has an extra degree of complexity involved in its development when compared to regular Neural Networks. For example the Keras/Tensorflow LSTM requires data to be re-shaped into a three dimensional numpy array before it can be used as an input.

Both models achieved a degree of success in addressing the research question. Prediction of electricity consumption over a 60 day time period has been achieved with results that have an acceptable level of error.

The Sarimax model achieved a significantly better forecast when the time period was reduced. It may be that the coefficients for further predictions are just not produced by this model. A feature reduction technique might be worthy of investigation. A Principal Component Analysis may be able to reduce the input features and therefore lead to less noise in the dataset. Perhaps this could improve the predictive capacity of the independent variables.

The Sarimax models main disadvantage in modelling electricity consumption and meteorological values is that the algorithm is linear in nature. The relationships between the dependent and independent variables may not have be best determined in this manner.

The success of the LSTM in forecasting the 60 days of the testing set shows the suitability of this algorithm for modelling electricity consumption. Perhaps there is not so much of a linear relationship between the variables in this research and the Neural Network could determine this while the Sarimax model could not. The work of Akpinar and Yumusak (2016) recorded a significant improvement of the R^2 value when a combined model was developed. Perhaps combining another optimizing algorithm with these models might produce an even better R^2 value.

7 Conclusion and Future Work

The models evaluated in this paper both achieved a degree of success in addressing the research objectives. Both models have successfully managed short-term forecasts for the electricity demand on the Irish electrical grid.

The inclusion of features relating to calendar events such as holidays might be beneficial in similar work. A distinct change in the pattern of data occurs at the Christmas holiday season. This is only the most visible occurrence of such a trend. Other holidays or calendar events may occur at a lower level throughout a year and representing these events may facilitate greater modelling accuracy. Incorporating economic indicators and energy pricing features would also be worthy of further research.

References

Akpinar, M. and Yumusak, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series methods, **9**(9): 727.

Alonso, A. M., Nogales, F. J. and Ruiz, C. (2020). A single scalable LSTM model for short-term forecasting of massive electricity time series, **13**(20): 5328.

Aravena, C., Riquelme, A. and Denny, E. (2016). Money, comfort or environment? priorities and determinants of energy efficiency investments in irish households, **39**(2): 159–186.

Arras, L., Arjona-Medina, J., Widrich, M., Montavon, G., Gillhofer, M., Müller, K.-R., Hochreiter, S. and Samek, W. (2019). *Explaining and Interpreting LSTMs*, Springer International Publishing, Cham, pp. 211–238.

Arunraj, N. S., Ahrens, D. and Fernandes, M. (2016). Application of sarimax model to forecast daily sales in food retail industry, **7**(2).

Aurna, N. F., Rubel, T. M., Siddiqui, T. A., Karim, T., Saika, S., Arifeen, M., Mahbub, T. N., Reza, S. M. S. and Kabir, H. (2020). Time series analysis of electric energy consumption using autoregressive integrated moving average model and holt winters model, **19**(3): 11.

Bose, B. K. (2017). Artificial intelligence techniques in smart grid and renewable energy systems—some example applications, *Proceedings of the IEEE* **105**(11): 2262–2273.

Breuel, T. M. (2015). Benchmarking of LSTM networks, *CoRR* **abs/1508.02774**.

Brophy, B., McCormack, E., Smith, W. and Timoney, D. (2021). Interactive, bottom-up model to audit the true per-person energy consumption (direct and embedded) of irish citizens per annum, **7**: 1025–1045.

Bu, S.-J. and Cho, S.-B. (2020). Time series forecasting with multi-headed attention-based deep learning for residential energy consumption, **13**(18): 4722.

Chen, C., Liu, Y., Kumar, M., Qin, J. and Ren, Y. (2019). Energy consumption modelling using deep learning embedded semi-supervised learning, **135**: 757–765.

Deng, Z., Wang, B., Xu, Y., Xu, T., Liu, C. and Zhu, Z. (2019). Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting, **7**: 88058–88071.

Gencer, K. and Başçiftçi, F. (2021). Time series forecast modeling of vulnerabilities in the android operating system using ARIMA and deep learning methods, **30**: 100515.

Hewamalage, H., Bergmeir, C. and Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions, **37**(1): 388–427.

Ilbeigi, M., Ghomeishi, M. and Dehghanbanadaki, A. (2020). Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm, **61**: 102325.

Jana, R. K., Ghosh, I. and Sanyal, M. K. (2020). A granular deep learning approach for predicting energy consumption, **89**: 106091.

Khan, P. W. and Byun, Y.-C. (2020). Genetic algorithm based optimized feature engineering and hybrid machine learning for effective energy consumption prediction, **8**: 196274–196286.

Khan, P. W., Byun, Y.-C., Lee, S.-J., Kang, D.-H., Kang, J.-Y. and Park, H.-S. (2020). Machine learning-based approach to predict energy consumption of renewable and non-renewable power sources, **13**(18): 4870.

Khatoon, S., Ibraheem, Singh, A. K. and Priti (2014). Effects of various factors on electric load forecasting: An overview, pp. 1–5.

Mahajan, T., Singh, G. and Bruns, G. (2021). An experimental assessment of treatments for cyclical data, p. 6.

Mishra, M., Nayak, J., Naik, B. and Abraham, A. (2020). Deep learning in electrical utility industry: A comprehensive review of a decade of research, **96**: 104000.

Pacella, M. and Papadia, G. (2021). Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management, **99**: 604–609.

Rueda, R., Cuéllar, M., Pegalajar, M. and Delgado, M. (2019). Straight line programs for energy consumption modelling, **80**: 310–328.

Sadorsky, P. (2010). The impact of financial development on energy consumption in emerging economies, **38**(5): 2528–2535.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, **404**: 132306.

Xue, B. and Geng, J. (2012). Dynamic transverse correction method of middle and long term energy forecasting based on statistic of forecasting errors, pp. 253–256.