Time series forecasting (Group Project 7).
COMPGI15 / M052 Information Retrieval and Data Mining.
Group 22

Niki Sfika
Student Number 15108218
`niki.sfika.15@ucl.ac.uk`

Vladislav Ligay
Student Number 15061475
`vladislav.ligay.15@ucl.ac.uk`

Vinit Jadhav
Student Number 15091240
`vinit.jadhav.15@ucl.ac.uk`

September 28, 2016

Abstract

Energy consumption forecasting is a tricky task given the presence of complex linear as well as non-linear patterns in the energy consumption timeseries. While the orthodox method of auto-regressive integrated moving average (ARIMA) performs well to diagnose the linear aspects within the timeseries it fails to account for its non-linear aspects. On the other hand, neural network architecture accounts for the non-linear aspects of the timeseries but often ignores the linear aspects. In this research, we employ and compare the performance of the statistical method of auto-regressive integrated moving average (ARIMA), the econometric method of vector autoregression (VAR) and the machine learning method of artificial neural networks (ANN RNN LSTM) in forecasting energy consumption. Ultimately we devise and test a hybrid model based on VAR and ANN to capture both linear as well as non-linear aspects of the energy timeseries. We observe that the ANN (RNN LSTM) model outperforms all other models in terms of accuracy when the accuracy is measured using mean absolute percentage error (MAPE).

Contents

1   Introduction

The task of forecasting energy demand is critical for decision makers from the energy industry to reduce wastages and maximize profits as well as for policy makers from national governments to optimize energy regulation and utilization in their national interests. Energy demand forecasting, alternatively termed as load forecasting, involves the accurate prediction of energy consumption across various geographical zones with the quantity under consideration typically being the hourly load. Forecasting energy demand is however a tricky task. The timeseries for energy consumption are signified by the presence of linear as well as non-linear aspects within them. Experts from all around the globe have significantly contributed to address the problem of load forecasting. The Global Energy Forecasting Competition 2012 aimed at attracting participants all over the world to elicit novel ideas in the energy forecasting field. We use the load as well as temperature data provided as a part of this competition to employ and test the efficiency of three statistical models in load forecasting.

Initially we forecast the daily load based upon the orthodox technique of auto-regressive integrated moving average (ARIMA) which is heavily based upon linear analysis. The ARIMA model aims at predicting the values based upon the past values of the load timeseries. It is also important to consider if the temperature values have an effect on the hourly load. To uncover such causality we use the vector autoregressive (VAR) model in order to capture the linear interdependence between the load and temperature timeseries. While the ARIMA and VAR models work well to forecast daily load based upon the linear aspect (trend) they fail to account the sensitive non-linear aspects of the load timeseries which represents randomness induced by unaccounted emergencies and weather conditions. In order to uncover the non-linear aspects about the load timeseries we use machine learning method of artificial neural networks (ANN) based upon the recurrent neural network long short-term memory architecture. However the focus on non-linear fittings of the ANN model often hampers the forecasting accuracy of the linear basic aspect. Ultimately we intend to devise a hybrid approach harnessing the best of the linear as well as non-linear models to achieve higher forecasting accuracy.

2   Literature Review

During the past few years many approaches have been adopted for forecasting the energy consumption in various parts of the world. These approaches have been linear as well as non-linear in nature. Early research on this topic used the ARIMA linear model for energy forecasting and to investigate the dependence on external economies for energy supply [1]. By using the ARIMA model, these studies found that the energy consumption is about to increase in the future and they suggested that this particular model is robust method in predicting energy consumption in general. Another study used the ARIMA model for the energy consumption forecasting in China [2]. The study focused on China because the energy consumption in this particular country has been growing rapidly and forecasting takes prime importance in such a context to formulate effective energy policies. The results of this study also indicated that ARIMA model provides high accuracy in its prediction and its an important model that can be implemented concerning energy consumption. Brazil is another country where researchers focused on forecasting the energy consumption along with CO_2 emissions [3]. By implementing both the ARIMA model and the VAR model to predict variables from 2008 until 2013, they found that both of them have strong forecasting performance. As far as the VAR model is concerned, it is also has been used to predict energy consumption in China as we have already mentioned that it is important to find the future demand for energy in this country [4]. VAR model has also been implemented in [5] for forecasting energy consumption in Iran as this country is one of the most powerful import and export energy of the world. Again, this model predicted a significant growth in energy demand in Iran until 2015. VAR model has also been effective for predicting energy demand in Ghana, showing that a significant growth is about to happen in energy consumption [6]. Apart from ARIMA and VAR, one more model that has been used for predictions is artificial neutral networks (ANN). Particularly, ANN has been used many times by researches as a technique for prediction energy consumption and has been suggested as an extremely important and promising method for predictions [7]. In [8], it is described an ANN method for the prediction of energy consumption in Greece. The produced results were really close to the real values and they were much better than the results produced by a linear regression model and

produced also by a support vector machine model. This method can be useful further for the implementation of energy policies since the predictions tend to be much more accurate. An ANN approach was also used for the prediction of energy demand in Thailand [9]. Again they found that ANN method produced accurate results compared to other models. Although ANN was successful in producing accurate predictions, the interpretation of these results was a tricky task. On the other hand, the study also tested ARIMA model, a model that is easy to implement, and they get almost similar results to the ANN methodology and they were arguing about whether ARIMA or ANN should be used. Another work where both ARIMA and ANN were used in order to create a new hybrid model is introduced in [10]. In this paper the authors argue that neither ARIMA nor ANNs are adequate in energy consumption forecasting because ARIMA is not useful with nonlinear relationships and ANN cannot deal with both linear and non-linear patters equally well. Based on that, they created a new hybrid model that combines the advantages of the ARIMA and ANNs. They tested this model for the energy demand prediction in China and they found that this model can be more effective for increasing the energy consumption forecasting accuracy compared to the aforementioned two models.

3 Research

3.1 Research Motivation

The exposure to data mining and data analysis concepts as a part of the Information Retrieval and Data Mining module for the master's programme and application of these concepts to gain meaningful and purposeful insights supporting the decision making process has been a great motivator behind this research. Previous work experience in the energy sector, a brief understanding of energy markets and the plethora of opportunities data mining and analysis methodologies open towards building models capable of predicting energy consumption data have generated great curiosity to undertake this research. Lastly the significant contribution of the quality of this research to the module grade cannot be ignored and is a crucial motivating factor behind this research.

3.2 Research Objective

In our research we address a hierarchical load forecasting problem from the load forecasting track of this competition to forecast hourly loads for a US utility with 20 zones. The dataset provided to solve this forecasting problem includes the energy consumption data for a period of four and half years from 1/1/2004 to the 6th hour of 30/6/2008 out of which data for 8 weeks is set to be missing. Additionally, we have been given the temperature data across 11 stations for the same period without the missing weeks. The relationship of these 11 stations with 20 zones has not been provided. Our research objective is to

- Create ARIMA, VAR, ANN and a hybrid model to forecast the load data for missing weeks.

- Evaluate the forecasting accuracy of these models based upon the mean absolute percent error (MAPE).

- Discuss our results, summarize our learnings and suggest further work which can be undertaken.

3.3 Research Methodology

A flowchart providing us the overview of the research steps undertaken can be seen below. We discuss the different methods we adopted in detail proceeding the flowchart.
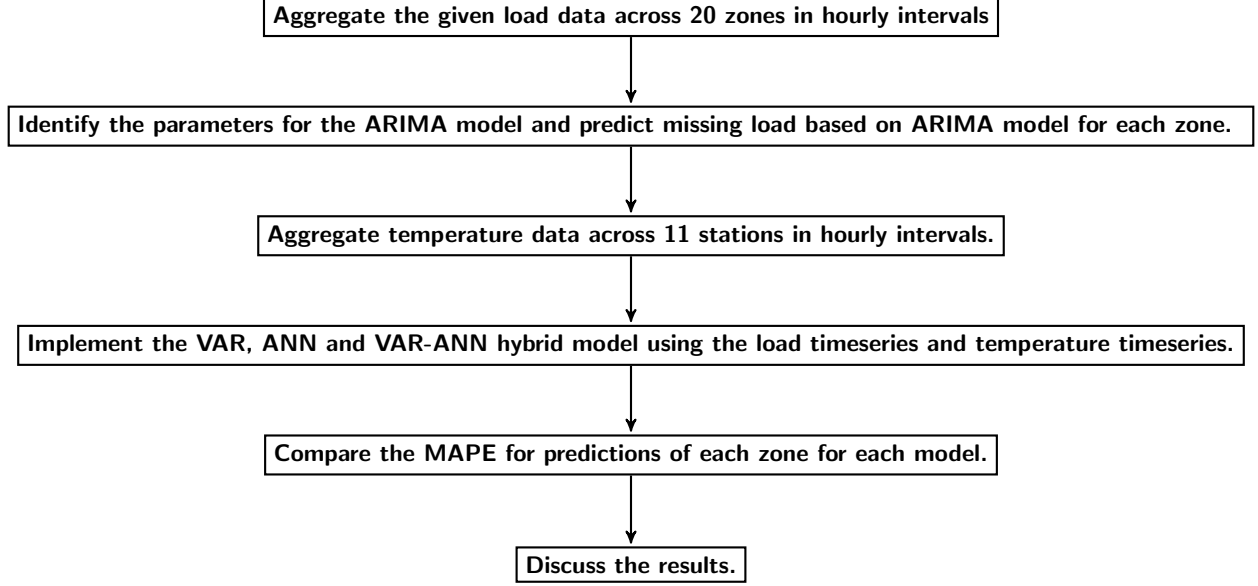
4

```
┌─────────────────────────────────────────────────────────────────────┐
│      Aggregate the given load data across 20 zones in hourly intervals │
└─────────────────────────────────────────────────────────────────────┘
                                    │
                                    ▼
┌─────────────────────────────────────────────────────────────────────┐
│ Identify the parameters for the ARIMA model and predict missing load based on ARIMA model for each zone. │
└─────────────────────────────────────────────────────────────────────┘
                                    │
                                    ▼
┌─────────────────────────────────────────────────────────────────────┐
│        Aggregate temperature data across 11 stations in hourly intervals. │
└─────────────────────────────────────────────────────────────────────┘
                                    │
                                    ▼
┌─────────────────────────────────────────────────────────────────────┐
│ Implement the VAR, ANN and VAR-ANN hybrid model using the load timeseries and temperature timeseries. │
└─────────────────────────────────────────────────────────────────────┘
                                    │
                                    ▼
┌─────────────────────────────────────────────────────────────────────┐
│        Compare the MAPE for predictions of each zone for each model. │
└─────────────────────────────────────────────────────────────────────┘
                                    │
                                    ▼
                         ┌─────────────────────┐
                         │   Discuss the results. │
                         └─────────────────────┘
```

Figure 1: *Flowchart of Research Methodology*

3.3.1   ARIMA

The ARIMA model defined as ARIMA(p,d,q) uses data at previous instances of time to fit a linear equation which is further used to forecast values at later instances of time. An important consideration here is stationarity of the data in the timeseries. If the data is non-stationary, differencing is performed on the timeseries till it achieves stationarity. The order of differencing in this case corresponds to d. The order of moving average q is determined from the decaying nature of the autocorrelation function (ACF) of the timeseries while the order of autoregression p is determined on from the decaying nature of the partial autocorrelation function (PACF).

According to the ARIMA(p,d,q) model, value of y at time t, y_t is the weighted sum of its previous values in time y_{t-1}, y_{t-2},...,y_{t-p} and error values e_t, e_{t-1}, e_{t-2},...,e_{t-q}. The equation for the same is written as,

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + e_t + b_1 e_{t-1} + b_2 e_{t-2} + b_q e_{t-q} \tag{1}$$

The coefficients $a_1, a_2, ..., a_p$ and $b_1, b_2, ..., b_q$ are estimated using the Box-Jenkins method. The ARIMA model can be thought of as a three layer filter which acts upon the given time series. In the first layer if the filter we extract the trend from the given timeseries data. This is where the differencing acts on the timeseries. Usually differencing of the first order, that is creating a timeseries of the difference between two consecutive values in the timeseries gives us a stationary (trendless) timeseries. Stationarity here is stationary around mean and variance. The stationarity of the timeseries can be validated using the KPSS and ADF tests. Differencing of further order can be taken into consideration if the first order differencing does not remove trend from the data. Differencing is an important tool to remove autocorrelation present in the timeseries as well.

Autocorrelation is a measure of the internal correlation within a time series. It is a way of measuring and explaining internal association between observations in a time series. It can be termed as the similarity between observations as a function of the time lag between them. Given measurements $y_1, y_2, ..., y_n$ at time $t_1, t_2, ..., t_n$ the lag k autocorrelation function A_k is defined as

$$A_k = \frac{\sum_{i=1}^{n-k}(y_i - \hat{y})(y_{i+k} - \hat{y})}{\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{2}$$
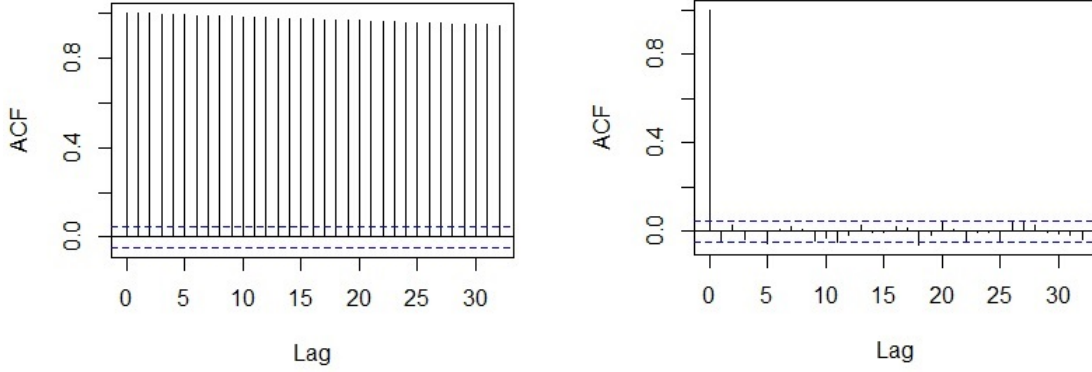
5

Figure 2: *Autocorrelation observed for load timeseries for Zone 1 (left exhibit) upto 35 days. The first difference of this timeseries reduces the autocorrelation observed to insignificant levels (right exhibit)*

Note that the autocorrelation at lag 0 is 1 as the time series being compared are the same without any lag. The autocorrelation index assigns a value of +1 to strong positive association, -1 to strong negative association and 0 to no association. The second filter extracts the influence of previous values in the timeseries on the current values. The value of p which determines how many previous values are chosen in the regression to determine the current values which can be seen in equation 1. The third filter which is dependent upon the value q takes into consideration the impact of the previous error terms on the current error terms. Once this three stage filter acts upon the given timeseries we need to ensure that the residuals left have no information remaining. This can be checked by examining the residuals for autocorrelation and partial autocorrelation. The best parameters for the model can be chosen based upon the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. Parameters for which a model corresponds to a low AIC and BIC score should be chosen.

3.3.2 VAR

Unlike ARIMA which is a univariate method taking into consideration the past values of the same timeseries, vector autoregression (VAR) is a multivariate method which takes into consideration the past values of the same timeseries as well as another timeseries to forecast future values. A VAR model n-variable model where each variable is explained by its own lagged values as well as the lagged values of the remaining n-1 variables. The equations describing a VAR model with 2 variables y_1, y_2 and lag 1 is formulated as below.

$$y_{1,t} = c_1 + (a_{11})(y_{1,t-1}) + (a_{12})(y_{2,t-1}) + e_{1,t} \tag{3}$$

$$y_{2,t} = c_2 + (a_{21})(y_{1,t-1}) + (a_{22})(y_{2,t-1}) + e_{2,t} \tag{4}$$

Similarly a VAR_p model uses p previous values of its own timeseries as well as the p previous values of the timeseries of other variables to determine its current value. p is also called the lag order of the VAR model. The stationarity of the participant timeseries in the model is an important consideration for building this model. The stationarity of the timeseries should be in terms of mean as well as variance. This can be achieved by first order of differencing or log transformation of differences or both. The choice of lag order p is an important choice which can be made based upon AIC/BIC.

The set of variables which are included in the VAR model are called endogenous variables on account of the correlation between them. Further improvisations of the VAR model such as the structural vector autoregressive models (SVAR) and the bayesian vector autoregression model (BVAR) have been developed and implemented. After the choice of p the model needs to be checked before it is implemented. This can be achieved by checking the residuals for autocorrelations. Non-autocorrelated residuals ensure the robustness of the model as they imply that all the necessary information from the endogenous variables has been

extracted. Another important consideration in this regards is checking the variables for granger causality; however this offers limited help in the multivariate context as compared to the bivariate context. This is due to the fact that a variable which is granger causal of another variable in a bivariate model may lose its importance in a multivariate model involving other variables which may drive both these variables and as such disturb the bivariate granger causality.

3.3.3 ANN (RNN LSTM)

The ARIMA and VAR models harness the linear relationships between the timeseries to predict future values. However, it is only obvious that the load timeseries is characterised by non-linear aspects representing the unaccounted emergencies due to external factors. Artificial Neural Networks (ANN) are models which help in uncovering the non-linear aspects of the timeseries. The architecture of an ANN is similar to that of the brain where neurons are substituted by nodes arranged in different layers. A simple three layer architecture of a neural network can be seen in Figure 1 where an input layer, a single hidden layer and an output layer is seen. As seen in Figure 1, the input layer of the model has multiple nodes which are the timeseries values at various lags. The equation of the model is given as,

$$y_t = f(y_{t-1}, y_{t-2}, ..., y_{t-k}) + e_t \tag{5}$$

where k is the lag and f is the transfer function of the hidden layer which can be linear, sigmoid, tan-sigmoid etc. The weights of the links and the bias values are the coefficients of the model. As seen in Figure 1 the leftmost layer in the diagram is called the input layer and the neurons are called input neurons. The rightmost layer in the network is called the output layer and the neuron called the output neurons. We can one or many input / output neurons. The middle layer having neurons which are neither input nor output is termed as the hidden layer. A neural network architecture may have one or many hidden layers. In our ANN model we used the lagged values of the load timeseries as the inputs and get the current load value as a single output. The specifications of the neural network we put to use and details about the same are explained in the testing and results section.
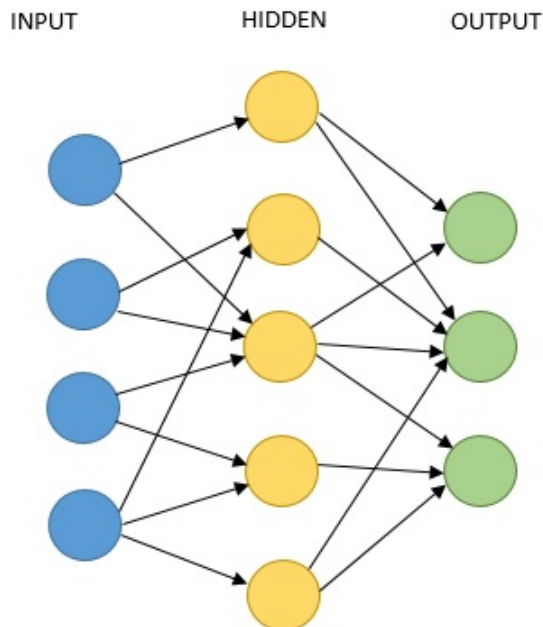

Figure 3: *A simple neural network architecture showing the input, hidden and the output layers.*

In our implementation of this model we use the Long short-term memory recurrent neural network which is a class of artificial neural networks which contain LSTM blocks in addition to the ordinary nerual network elements. The recurrent neural network is a type of artificial neural network which differs from the feedforward neural netwroks in the way that they are capable of using current inputs as well as their own one past step outputs. The LSTM is an improvistion of the RNN where the LSTM memory cell preserves the outputs from previous steps maintaining a constant error rate. The learning rate of this network as a result is hence higher.

3.3.4 VAR-ANN Hybrid

In the VAR ANN hybrid model we try to capture the multivariate linear aspect of the VAR model as well as the non-linear aspects from the ANN model. Compared to the ANN model where we supply only the lagged load values as inputs, in the hybrid model we supply the lagged temperature values in addition. The value of p from our VAR model is used as the lag value and we provide the load and temperature values upto this lag as the input.
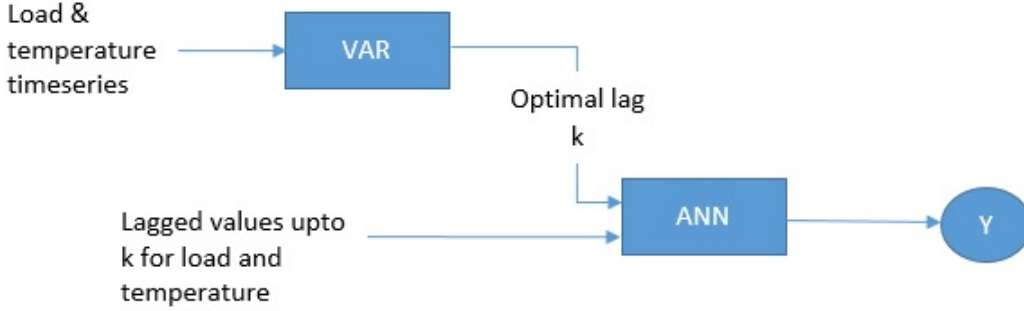

Figure 4: *Architecture of our hybrid model involving the VAR model and the ANN model. The value of k is diagnosed from the VAR model and those lagged values from the temperature timeseries are used as input to the ANN (RNN LSTM) model.*

3.3.5 Evaluation using MAPE

The mean absolute percentage error is a metric to measure the accuracy of a forecasting model. It is expressed as a percentage and can be calculated by the formulae below where A_t is the actual value and F_t is the forecasted value. We take the absolute value of the difference between the actual value and the forecasted value divided by the actual value. This is done for every point in the forecast and the sum of these absolute values is then divided by the number of points in the forecast.

$$M = \frac{1}{n} \sum_{t=1}^{n} |\frac{A_t - F_t}{A_t}| \tag{6}$$

Our motivation in the choice of MAPE as an evaluation metric lies in the simplicity of this metric. A major drawback of this metric is the presence of zero values in the actual values, however, this case does not apply in the context of our data as the energy demand is never zero in any of the zones.

4 Testing & Results

Based upon the methodology described above and the models taken into consideration we take the hourly load data available across all zones along with the temperature data across 11 stations and predict the load data for the missing weeks. The results we get for each of the four models are described below.

4.1 ARIMA Model

We consider each zone separately and provide our model the hourly load data for that zone. The choice of the model parameters that is p,d,q is initiated by taking the first differences of the timeseries values. We subsequently increase the order of differencing till we stationarity is reached. The stationarity is validated using the KPSS and the ADF tests. It should be noted that the null hypothesis for each of these tests are different. In most of the cases differencing of the first order guarantees the stationarity of the timeseries. The parameter values d and q are selected based upon the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the timeseries. We use the parameter values corresponding to the minimum AIC score for our model.

Zone	Model Specifications	MAPE
9	ARIMA(3,0,1)	0.1038282
10	ARIMA(3,1,2)	0.1314534
20	ARIMA(3,1,3)	0.1336297
19	ARIMA(2,1,1)	0.2815757
14	ARIMA(2,1,2)	0.3234994
16	ARIMA(2,1,1)	0.324753

Table 1: ARIMA model specifications for top 3 Zones and bottom 3 zones as per MAPE.

Table 1 shows the top 3 zones with the least MAPE and the corresponding model specifications which resulted in these MAPE values. The 3 zones with the maximum MAPE and the model specifications can also be seen in the table. The model specifications for models adopted across all 20 zones can be seen in Appendix Exhibit 1. A timeseries plot showing the actual load and the load predicted by the ARIMA model for the first missing week for Zone 1 can be seen in Figure 5. It is seen that the ARIMA predictions seem to converge to the mean of the load timeseries value. Although the ARIMA model successfully replicates the trend the non-linear dynamics of the timeseries are not accounted for by the ARIMA model. The robustness of the model is validated by checking the residuals of the model for autocorrelation and partial autocorrelation functions. If these values fall below insignificant levels then we have a robust model.
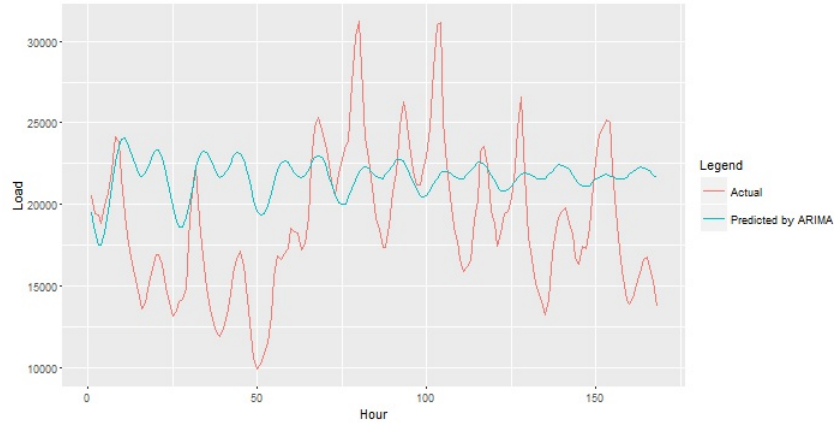

Figure 5: *Timeseries plot showing the actual load values and the predicted load values using the ARIMA model for Zone 1.*

9

4.2 VAR Model

While implementing the VAR model we take into consideration the hourly temperature timeseries across 11 stations in addition to the load timeseries of each zone. The relationship between the 11 stations and the 20 zones is not provided. This makes the choice of the temperature timeseries for each zone an interesting choice. We experiment with correlation coefficient and mutual information between the load timeseries and the temperature timeseries for 11 stations for this choice. However, we observe that considering all 11 timeseries for temperature data yields more accuracy in our VAR models. Hence we build independent VAR models for each zone prediction taking into consideration the temperature timeseries of all 11 stations.

Zone	MAPE(ARIMA)	MAPE (VAR)
10	0.1314534	0.105022319
9	0.1038282	0.106505778
3	0.13512	0.111456701
19	0.2815757	0.258008769
14	0.3234994	0.30683286
16	0.324753	0.309774759

Table 2: Top 3 zones and bottom 3 zones as per MAPE of VAR model.

Table 2 shows the top three and bottom three zones in terms of the MAPE as per the VAR model. Also seen is the MAPE as per ARIMA model for these zones. The VAR model outperforms the ARIMA model for all zones in terms of MAPE except for Zone 9 for which the MAPE increases marginally in case of the VAR model. The MAPE across all zones as per the VAR model can be seen in the summary of results. Figure 6 shows a timeseries plot of the actual load and predicted load as per the VAR model. Unlike the ARIMA model where we saw a constant repetition of a pattern the VAR model shows much more variation in the predicted values. This is due to the impact of the temperature timeseries values which are a part of the VAR model.
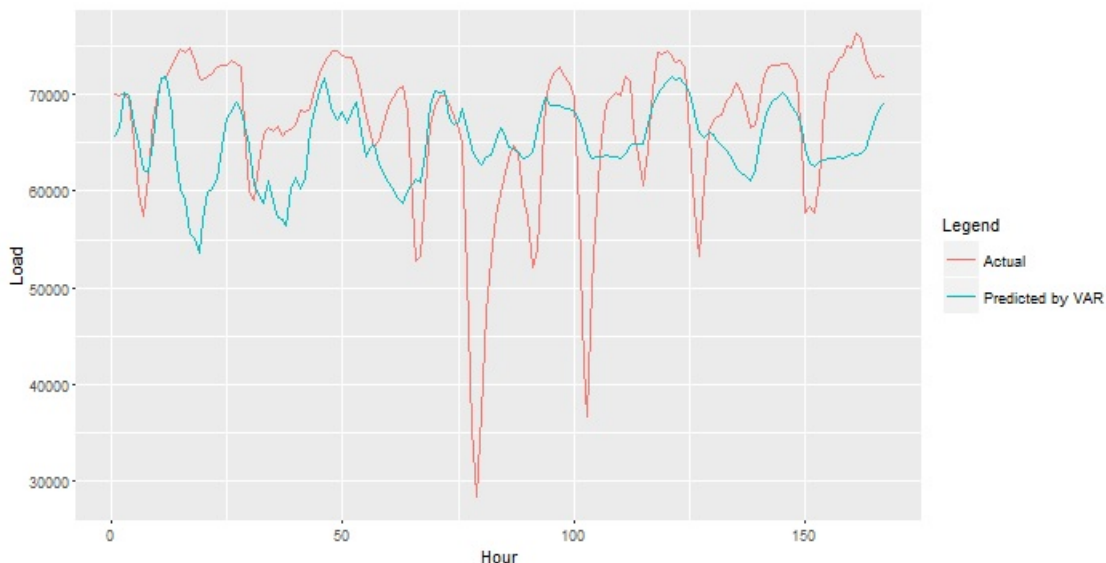

Figure 6: *Timeseries plot showing the actual load values and the predicted load values using the VAR model for Zone 9.*

4.3 ANN (RNN LSTM) Model

In the ANN model we use the Long short-term memory recurrent neural network architecture. The inputs we provide to the LSTM network are the lagged load values which are used by the network to calculate current load values. Unlike the feed forward neural networks which only allow forward call as the input goes through the hidden layer to the output layer, LSTM RNN architecture allows for backward calls within the hidden layer in addition to the capacity to avail memory blocks. The average accuracy in terms of MAPE across all zones and all missing periods of load is the highest for this model.

Zone	MAPE(ARIMA)	MAPE (VAR)	MAPE (ANN)
20	0.1336297	0.119380932	0.102215203
9	0.1038282	0.106505778	0.102335377
3	0.13512	0.111456701	0.111884055
18	0.2794444	0.257120725	0.185656416
14	0.3234994	0.30683286	0.198429407
16	0.324753	0.309774759	0.225437682

Table 3: Top 3 zones and bottom 3 zones as per MAPE of ANN (RNN LSTM) model.

Table 3 shows the top 3 and bottom 3 zones in terms of accuracy as per MAPE when using the ANN model. IT is observed that in certain cases the accuracy of the ANN model is inferior to that of the VAR model however always superior to the ARIMA model. The average MAPE across all zones and all missing periods is however the least in case of the ANN model. Figure 7 below shows the actual load values and the predicted load values for Zone 16. This is the zone where the MAPE is maximum corresponding to 0.2254 when the ANN model is put to use. The ANN model however reduces the MAPE for this zone from 0.3247 observed for the ARIMA model and 0.3097 in the VAR model. The MAPE for all zones when using the ANN model and the comparison of this model with the other models can be seen in the summary of results.
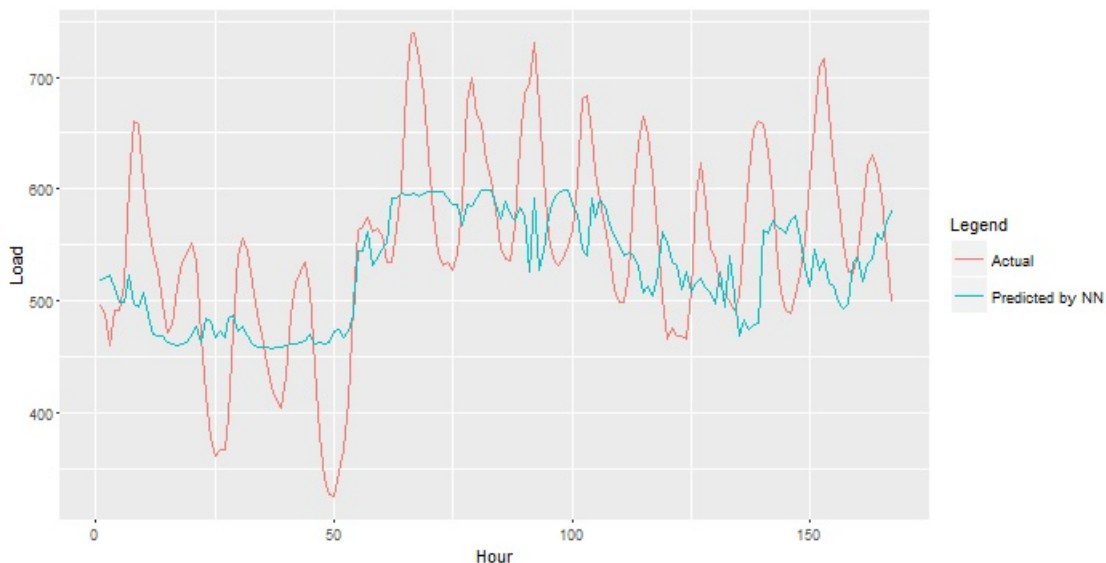

Figure 7: *Timeseries plot showing the actual load values and the predicted load values using the ANN (RNN LSTM) model for Zone 16.*

4.4 VAR-ANN Hybrid Model

We devise the VAR-ANN hybrid model to preserve the linear aspects as well as the non-linear aspects in the load timeseries. The model is similar to the ANN (RNN LSTM) model apart from the fact that the inputs provided are the lagged values of the temperature timeseries which we encountered in the VAR model. The optimal lag value i.e. the number of lags upto which the temperature values should be taken are derived from the VAR model. We setup this model on an experiential basis and compared its forecasting accuracy to the previous three models.

Zone	MAPE(ARIMA)	MAPE (VAR)	MAPE (ANN)	MAPE (HYBRID)
20	0.1336297	0.119380932	0.102215203	0.10346449
9	0.1038282	0.106505778	0.102335377	0.114828383
13	0.1607305	0.144126171	0.116205323	0.12199426
18	0.2794444	0.257120725	0.185656416	0.20564151
14	0.3234994	0.30683286	0.198429407	0.223349099
16	0.324753	0.309774759	0.225437682	0.230392548

Table 4: Top 3 zones and bottom 3 zones as per MAPE of VAR-ANN hybrid model.

Table 4 shows the top 3 and bottom 3 zones in terms of accuracy as per MAPE when using the VAR-ANN hybrid model. It is observed that the accuracy of the VAR-ANN hybrid model is slightly inferior to that of the ANN model however superior on an average to the ARIMA and VAR models. The average MAPE across all zones and all missing periods is less compared to the ARIMA and VAR models but more than of the ANN model. Figure 7 below shows the actual load values and the predicted load values for Zone 9. The prediction seems to be hypersensitive to the temperature inputs and hence the volatility in the predictions. The phenomena of overfitting the data can possibly result into this.
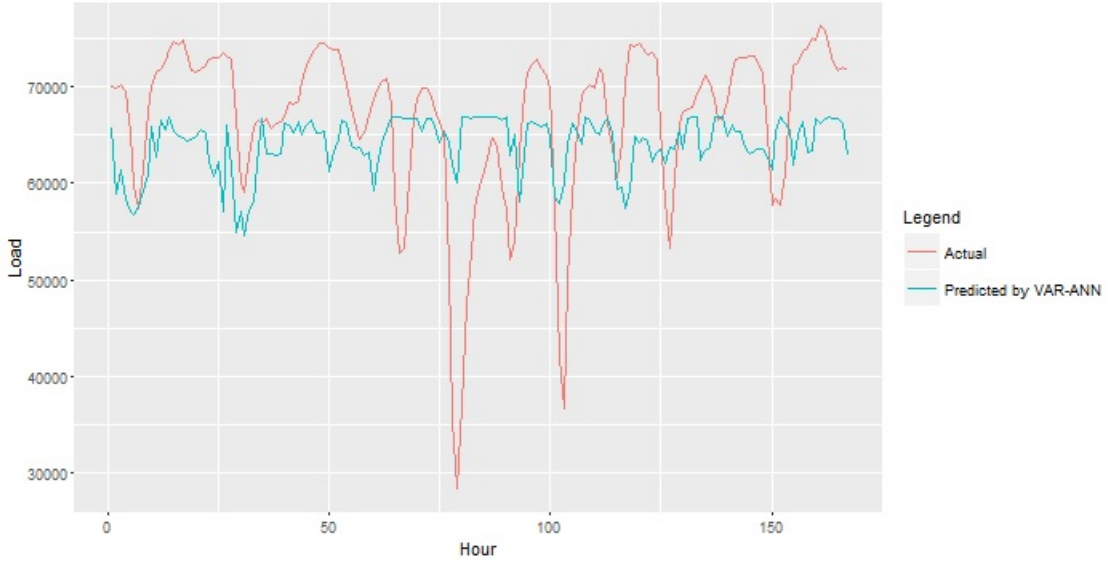

Figure 8: *Timeseries plot showing the actual load values and the predicted load values using the VAR-ANN hybrid model for Zone 9.*

4.5 Summary of Results

Table 5 summarizes the MAPE for each zone for prediction of missing load values across all the missing weeks. The average MAPE for each model across all the 20 zones is calculated as well. The ARIMA model

has the highest MAPE of 0.1978 followed by the VAR model with an average MAPE of 0.1786. The ANN (RNN LSTM) model has the lowest MAPE of 0.1454 and hence is the most accurate model from the four we compare. The hybrid model we implemented has a slightly greater MAPE value compared to the ANN (RNN LSTM) with a MAPE of 0.1561.

Zone	MAPE(ARIMA)	MAPE (VAR)	MAPE (ANN)	MAPE (HYBRID)
1	0.269913	0.227823137	0.170925291	0.192263355
2	0.1350996	0.111457003	0.118590732	0.12367616
3	0.13512	0.111456701	0.111884055	0.124860464
4	0.1491109	0.129850814	0.117069452	0.127062918
5	0.2750808	0.244593483	0.174185352	0.200273649
6	0.1353031	0.115009316	0.122159048	0.12500125
7	0.13512	0.111456701	0.113095614	0.125774908
8	0.1716779	0.142720523	0.123565008	0.135926336
9	0.1038282	0.106505778	0.102335377	0.114828383
10	0.1314534	0.105022319	0.129423948	0.127870517
11	0.222633	0.219358037	0.168950837	0.185713516
12	0.2258354	0.221606966	0.183690213	0.195080002
13	0.1607305	0.144126171	0.116205323	0.12199426
14	0.3234994	0.30683286	0.198429407	0.223349099
15	0.1659001	0.15346111	0.130114531	0.123649946
16	0.324753	0.309774759	0.225437682	0.230392548
17	0.1962608	0.177196641	0.129724685	0.140665138
18	0.2794444	0.257120725	0.185656416	0.20564151
19	0.2815757	0.258008769	0.184473629	0.194519142
20	0.1336297	0.119380932	0.102215203	0.10346449
Average	0.197798445	0.178638137	0.14540659	0.15610038

Table 5: MAPE for all zones using all models.

5 Conclusions & Future Work

We tested linear as well linear models to forecast the energy demand for a US utility with twenty zones and observed that the non-linear model of neural networks based upon the long short-term memory recurrent neural network architecture outperformed all other models in terms of accuracy when the metric of mean absolute percent error was used. When comparing the linear models the vector autoregressive model (VAR) was better in terms of accuracy to the orthodox autoregressive integrated moving average (ARIMA) model. The hybrid model was built upon the trivial approach of including the temperature timeseries lagged values in our neural network model. This model can be improvised using the neural network to process errors from the VAR model and develop predictions based upon the cumulative outputs from the linear VAR model and the non-linear ANN (RNN LSTM) model.

6 Code

The code for the assignment can be found at $https://github.com/vladislavligay/IRDM\_UCL\_2016\_GROUP\_22$. The code is developed in R. We use the packages sqldf for data preprocessing. Further packages used to derive our results include the package forecast, entropy, NNnet, vars etc.

7 References

1 Yeboah, Samuel Asuamah, Manu Ohene, and T. B. Wereko. "Forecasting aggregate and disaggregate energy consumption using arima models: a literature survey." Journal of Statistical and Econometric Methods 1.2 (2012): 71-79.

2 Miao, Junwei. "The Energy Consumption Forecasting in China Based on ARIMA Model." (2015).

3 Pao, Hsiao-Tien, and Chung-Ming Tsai. "Modeling and forecasting the CO 2 emissions, energy consumption, and economic growth in Brazil." Energy 36.5 (2011): 2450-2458.

4 Crompton, Paul, and Yanrui Wu. "Energy consumption in China: past trends and future directions." Energy economics 27.1 (2005): 195-208.

5 Yazdan, Gudarzi Farahani, Varmazyari Behzad, and Moshtaridoust Shiva. "Energy consumption in Iran: past trends and future directions." Procedia-Social and Behavioral Sciences 62 (2012): 12-17.

6 zer, Mustafa, and Charles Mensah. "A Bivariate Causality between Energy Consumption and Economic Growth in Ghana."

7 Owda, Hasan MH, et al. "Using Artificial Neural Network Techniques for Prediction of Electric Energy Consumption." arXiv preprint arXiv:1412.2186(2014).

8 Tamizharasi, G., S. Kathiresan, and K. S. Sreenivasan. "Energy forecasting using artificial neural networks." International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 3.3 (2014): 7568-7576.

9 Kandananond, Karin. "Forecasting electricity demand in Thailand with an artificial neural network approach." Energies 4.8 (2011): 1246-1257.

10 Wang, Xiping, and Ming Meng. "A Hybrid Neural Network and ARIMA Model for Energy Consumption Forcasting." Journal of computers 7.5 (2012): 1184-1190.

Appendix

Exhibit 1. ARIMA model specifications for models adopted across all zones.

Zone	Model Specifications	MAPE
1	ARIMA(4,1,2)	0.269913
2	ARIMA(2,1,2)	0.1350996
3	ARIMA(2,1,2)	0.13512
4	ARIMA(2,1,2)	0.1491109
5	ARIMA(2,1,3)	0.2750808
6	ARIMA(5,1,3)	0.1353031
7	ARIMA(2,1,2)	0.13512
8	ARIMA(2,1,3)	0.1716779
9	ARIMA(3,0,1)	0.1038282
10	ARIMA(3,1,2)	0.1314534
11	ARIMA(5,1,4)	0.222633
12	ARIMA(5,1,4)	0.2258354
13	ARIMA(2,1,4)	0.1607305
14	ARIMA(2,1,2)	0.3234994
15	ARIMA(2,1,2)	0.1659001
16	ARIMA(2,1,1)	0.324753
17	ARIMA(2,1,2)	0.1962608
18	ARIMA(2,1,3)	0.2794444
19	ARIMA(2,1,1)	0.2815757
20	ARIMA(3,1,3)	0.1336297

Table 6: Model specifications for ARIMA models used for predicting load across all zones.