# Advanced Electricity Load Forecasting combining Electricity and Transportation Network

Lalitha Madhavi K.S[1], Jose Cordova[1], Mehmet Baran Ulak[2], Michael Ohlsen[3], Eren E. Ozguven[2], Reza Arghandeh [1], and Ayberk Kocatepe[2]

[1]Department of Electrical and Computer Eng, Florida State University (FSU) email: lk14f@my.fsu.edu
[2]Department of Civil and Environmental Eng, FSU
[3]City of Tallahassee Utilities

***Abstract*—Load forecasting plays a very crucial role in many aspects of electric power systems including the economic and social benefits. Previously, there have been many studies involving load forecasting using time series approach, including weather-load relationships. In one such approach to predict load, this paper investigates through different structures that aim to relate various daily parameters. These parameters include temperature, humidity and solar radiation that comprises the weather data. Along with natural phenomenon as weather, physical aspects such as traffic flow are also considered. Based on the relationship, a prediction algorithm is applied to check if prediction error decreases when such external factors are considered. Electricity consumption data is collected from the City of Tallahassee utilities. Traffic count is provided by the Florida Department of Transportation. Moreover, the weather data is obtained from Tallahassee regional Airport weather station. This paper aims to study and establish a cause and effect relationship between the mentioned variables using different causality models and to forecast load based on the external variables. Based on the relationship, a prediction algorithm is applied to check if prediction error decreases when such external factors are considered.**

***Index Terms*—ARIMA; Causality; Forecasting; Multivariate ARIMA**

## I. Introduction

Electrical load forecasting has been an important part in the field of Power systems. There have been several studies investigating load forecasting for several years [1] - [3]. Most of these studies were conducted considering weather-load relationships, and taking into account weather parameters such as temperature and humidity [4] - [5]. So far there have been load forecasting studies based on historical electricity consumption data [1] - [3] and also based on electricity-weather relationship [4] - [5]. Apart from weather, human activities is also one of the key factors that affect load consumption. Through mobility pattern which depicts day to day urban life routine, we try to enhance load forecasting. Therefore, this paper is a novel approach for electricity forecasting considering transportation pattern. The other novelty of this paper is using the causality analysis method to illustrate the impact of traffic load on electricity consumption.

To understand the factors affecting electricity consumption, a causal study is helpful. Cause-effect analysis can help us determine a better fit for a prediction algorithm. Understanding past values of the existing datasets helps us reveal causal models.

In general, causality tests are conducted to analyze if the values of one time series $X$ is useful in the prediction of another time series $Y$. In this paper, different causality analysis such as Granger causality, Peter-Clarke's algorithm are achieved and based on the relationships achieved, electricity load forecasting is calculated by Autoregressive integrated moving average (ARIMA) model. The functional idea of Granger Causality test implies that if X and Y were two signals which had a causal relationship between them, then in prediction Y, the previous values of X is also used along with previous values of Y. In other words, the calculation of Granger Causality is based on linear prediction models [6]. To establish a causal relationship for the entire system consisting of multi-variables, we move to a more complex causal model called the Peter Clarke's causal model. The Peter-Clarke's causal algorithm is based on constraints for estimating a graph corresponding to the causal graph which represents the causal relationships among different datasets [9].

As a validation for the causal model achieved, a forecasting method is applied on our dataset. Amongst the several time series forecasting techniques, is the Auto Regressive Integrated Moving-Average (ARIMA) model. There have been many studies conduction load forecasting based on ARIMA models [10] - [11] ARIMA model belongs to the class of stochastic processes generally applied in non-stationary cases. It consists of different terms: the autoregressive (AR), the integrated (I) and the moving average (MA). Basically, a time series $y_t$ is characterized as an ARIMA(p, q, d) process if it needs to be differenced d times $(wt = \Delta^d y_t)$ to be altered as a stationary time series which can then be modeled by an ARIMA(p, q) [14].

To encapsulate different time series datasets a Multivariate model which is also known as ARIMAX model has been considered. It is a statistical algorithm that describes the relationship between an output variable Y and one or more input variables [17]. Starting from the univariate models, an ARIMAX model with three explanatory variables (Load consumption, Weather and Transportation parameters) have been considered. Next section presents the methodology providing data description and describing causal relationships. Forecasting methods are then applied to achieve load forecasting and accuracy is calculated based on original load value.

## II. Methodology

The dataset considered in this study comprises of electricity, traffic and weather data. The electricity consumption of 222 houses of a neighborhood for the year 2015 accounts for the electricity data. Traffic count has been further divided into North bound and South bound traffic based on their direction of movement. Weather data consists of Temperature, Humidity and Solar Radiation values.

### A. Data Description

The City of Tallahassee has a Meter Data Management System (MDMS) that provides the readings of customers remotely and stores it for future analysis. For this case study, the energy consumption of a neighborhood, for the year 2015 was analyzed.

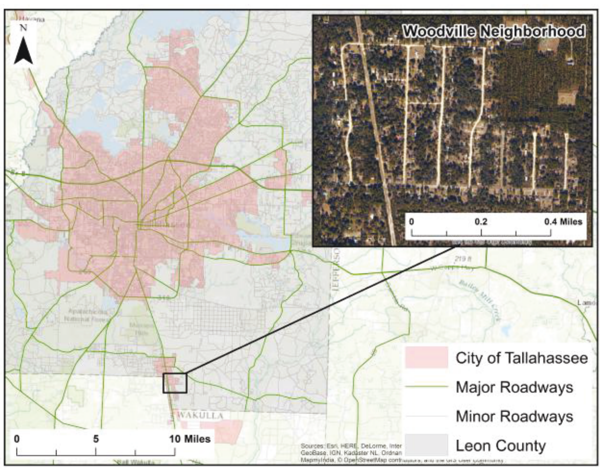


Fig. 1. City of Tallahassee Map : Woodville neighborhood highlighted

Figure 1 highlights the location, from where the data was collected and analyzed.

### B. Granger Causality

Granger Causality was first introduced by Granger in 1969 [6] . The functional idea implies that if X and Y were two signals which had a causal relationship between them, then in prediction Y, we can also use the past values of X along with past values of Y. In other words, the calculation of Granger Causality is based on linear prediction models [7].

Suppose $X^n = [X_1, X_2, ..., X_n]$ and $Y^n = [Y_1, Y_2, ...Y_n]$ are two time series. Comparing the two prediction errors $e_i$ and $e'_i$ is one of the basic step of Granger causality test:

$$Y_i = \sum_{j=1}^{L} a_j Y_{i-j} + e_i \tag{1}$$

$$Y_i = \sum_{j=1}^{L} [b_j Y_{i-j} + c_j X_{i-j} + e'_i] \tag{2}$$

In the above equation, prediction $Y_i$ based only on the past values of Y $(Y_{i-1}, ...Y_{i-L})$ consists of $e_i$ as the error, and predicting $Y_i$ based on both past values of Y, $(Y_{i-1}, ...Y_{i-L})$ and also the past values of X, $(X_{i-1}, ...X_{i-L})$ gives $e'_i$ as the error. If $e'_i$ is less than $e_i$ , that is, if by considering the previous values of X the accuracy in prediction of Y has been improved then we say that X granger Causes Y [8]. In case of strong non-linear interactions between two regions, Granger causality test may sometimes yield ambiguous result. [6]

### C. The Peter-Clarke's Algorithm

In Peter-Clarke's algorithm the procedure takes a database as an undirected graph G with a set of vertices V, and conducts a test of conditional independence *I(x,y,S)* and a significance level $0 < \alpha < 1$. This algorithm begins initially with a complete undirected graph, then shapes the graph by removing the edges with zero conditional independence relations, and again reshaping based on first order conditional independence relations, and so on. Also, the algorithm also takes an order (V) over the nodes which specifies the order to check for the independencies [12].

In the Gaussian case, conditional independencies can be inferred from partial correlations. They can be defined recursively for vertices as: $i, j \in V$ , $k \subset V\{i, j\}$ and some $h \in k$, [9].

$$\rho_{i,j|k} = \frac{\rho_{i,j|k/h} - \rho_{i,h|k/h}\rho_{j,h|k/h}}{\sqrt{(1-\rho^2_{i,h|k/h})(1-\rho^2_{j,h|k/h})}} \tag{3}$$

In the above equation, we assume that distribution P of the random vector X is multivariate normal.For $i \neq j \in 1, .., p, k \subseteq 1, ..., p$ $i, j$ that is denoted by $\rho_{i,j|k}$ the partial correlation between $X^i$ and $X^j$

Peter Clarke algorithm is based on the steps presented in the algorithm :

**Algorithm 1** Peter Clarke's Algorithm

**Input:** A dataset with V variables
and significance level $\alpha$
**Output:** Undirected graph G having edges E
**for** *each ordered pair of adjacent vertices X and Y in G* **do**
  **if** $|adj(X, G)| \setminus Y <= d$ **then**
    then **for** *each subset* $Z \subseteq adj(X, G) \setminus Y$ **do**
      Test I(X,Y| $Z$)**end**
      **if** *I(X,Y|* $Z$*)* **then**
        then Remove edge between X and Y
        Save Z as the separating set of (X,Y)
        Update G and E and Break
      **end**
    **end**
  **end**

### D. ARIMA model for Electricity Load Forecasting

This section describes the entire process that involves building the ARIMA model for electricity load forecasting. Theoretically, electricity load data is a time series data that is nonstationary and thereafter to be modeled and predicted, it needs to be stationarized by using the differencing. Based on the time profile of load data, it can be categorized into seasonal and continuous occurrences. First order continuous difference is achieved by using difference operator V:

$$VY_t = Y_t - Y_{t-1} = (1-L)Y_t \tag{4}$$

where, $Y_t$ is the $t^{th}$ time series, $Y_{t-1}$ is the $(t-1)^{th}$ time series, L is the Lag operator and V is the backward difference operator

For the original electricity load data which is periodic, to stabilize the load data, we use the periodical difference operator $V_s$ as shown below:

$$V_s Y_t = Y_t - Y_{t-s} \tag{5}$$

To use the continuous and seasonal difference together in the differencing procedure,

$$V_s{}^D V^d Y_t = (1-L^s)^D (1-L)^d Y_t \tag{6}$$

where, s specifies sample space of the data, D is the seasonal difference order and d is the continuous difference order After completion of differencing of the order, ARIMA with parameter (p,d,q) model is calculated by using the equation below:

$$\phi_p(L) V^d Y_t = \theta_q(L) a_t \tag{7}$$

where, $\phi_p(L)$ is the AR operator, $V^d$ is the dth order difference, $\theta_q(L)$ is the MA operator, p defines order of AR, q defines order of MA and $a_t$ : random white noise

Initially, the time series dataset is observed and a difference order d is selected to stationarize the data. Next, the orders of AR and MA that is p and q values are selected based on auto correlation and partial auto correlation functions. A complete ARIMA model is expressed as in Eq.(8).

$$\phi_p(L^s)\phi_p(L) V_s^D V^d Y_t = \Theta_Q(L^s)\theta_q(L) a_t \tag{8}$$

where $\phi_p(L^s)$ is the seasonal AR operator, $\Theta_Q(L^s)$ is the seasonal MA operator, $V_s^D$ is the seasonal $d^{th}$ order difference, P is the seasonal AR order, Q is seasonal MA order, s is sample space (or length of time span) and D is seasonal difference operator

With the solved coefficients in the above equation, final step involves checking residual and performing a fitting check to investigate the final model which is then followed by achieving load forecasting.

### E. ARIMAX Model for Electricity Load Forecasting

Even though univariate ARIMA model explains the relationship of time series data and fits a prediction, it does not give causal relationship measure between different independent variable and interpreted variable. It also models only for one time series data. Based on our causality models, energy consumption is known to have been influenced by external time series data such as weather and transportation. Therefore, we use the ARIMAX model that deals with different time series data related to electricity consumption such as weather and transportation, which enhances the prediction by univariate ARIMA model.

$$y_t = \sum_{i=1}^{m} \frac{\omega_i(L)}{\delta_i(L)} X_{1,t-b_i} + \frac{\theta(L)}{\phi(L)} a_t \tag{9}$$

where, $\theta(L), \phi(L), \delta(L), \omega(L)$ are defined as below, with degree *q,p,r* and *s*, respectively and $a_t$ is the Gaussian white noise series.

$$y_t = V^d Y_t \tag{10}$$

$$X_{1,t-b_i} = V^{d_t} x_{1,t-b_i} \tag{11}$$

Therefore, using lag operator, it can be represented as:

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q \tag{12}$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p \tag{13}$$

Where q and p are the order of moving average and auto regressive operator respectively.

$$\delta(L) = 1 - \delta_1 L - \delta_2 L^2 - ... - \delta_r L^r \tag{14}$$

$$\omega(L) = 1 - \omega_1 L - \omega_2 L^2 - ... - \omega_s L^s \tag{15}$$

Equation (13) and (14) represent time series of dependent variable and time series of interpreted variables with order *r* and *s* respectively.

Similar to univariate ARIMA models , fitting an ARIMAX model is followed by model identification, estimation of parameters, check for fitting and forecasting analysis. Prewhitening is applied during model identification to remove the pseudo-correlations among variables [17]. Consecutively, the model is achieved by examining the cross correlation function (CCF) among the different time series data. Furthermore, fitness of the model is checked with the help of residual analysis and fitting check. Finally, the suitable model is created that considers the weather, transportation and electricity consumption data to forecast the future electricity load trend.

## III. Evaluation and Discussion

The main aim of constructing a causal model is to understand the different factors that influence electricity consumption significantly. The three time series datasets were analyzed initially for any kind of correlations and causations. Granger causality and Peter Clark's causal models were applied to

Fig. 2. Granger causality test for year 2015

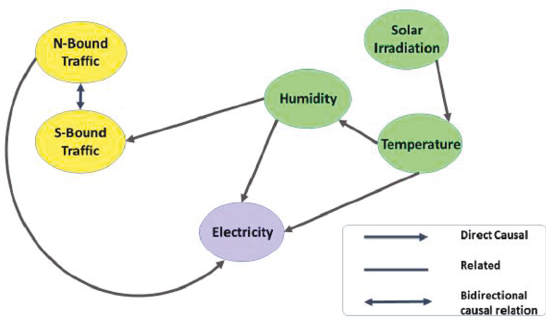


Fig. 3. Peter Causal model for year 2015

confirm a better fit. After the Casuality test, we fit the ARIMA models considering a combination of the variables.

In Figure 2, the causal map denotes the generated result from Granger Causality test. Since Granger causality best works on bivariate time series [13] , it was tested on pairwise variables As a result, it was noted that north bound traffic has a bidirectional relationship with south bound traffic. This can be explained that a high percentage of people who travel towards Northern part of the city, return after a time delay. The same stands true for people heading south. However, while dealing with more diverse data sets such as weather and electricity consumption, an ambiguous result is achieved with no direct causal relationship with electricity, though it did show some amount of interdependencies between humidity and temperature.Therefore, a more detailed causal model is considered which is based on conditional dependence. Causal model based on the Peter-Clark algorithm describes a causal network between the different parameters considered. Fig.3 represents the causal model achieved when applied on dataset spanning the entire year 2015.

According to the causal model, Humidity and Temperature have a direct causal relationship with Electricity consumption. And similarly, North bound traffic has a direct causal effect on the energy consumption.

To understand the effect of these factors on electricity consumption, load forecasting techniques are applied considering the different variables From the causal model achieved, the causal variables such as humidity, temperature into effect and perform load forecasting. It is seen that the error percentage has a significant reduction when compared to forecasting based on previous values of Energy consumption only. Therefore, traffic count is then introduced along with electricity and weather variables, to check for an improvement in the prediction of load.

To check the effectiveness of the predicted values, the mean absolute percentage error (MAPE) is calculated as expressed in the equation below:

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \frac{|Y_t^{'} - Y_t|}{Y_t} \tag{16}$$

where $n$ is the number of forecasted values, $Y_t^{'}$ is the forecasting value and $Y_t$ is the actual value.

### *A. CASE 1 : ARIMA model on Electricity consumption data*

An ARIMA model was fitted on this data to predict load for the following day based on the past values of electricity consumption alone.

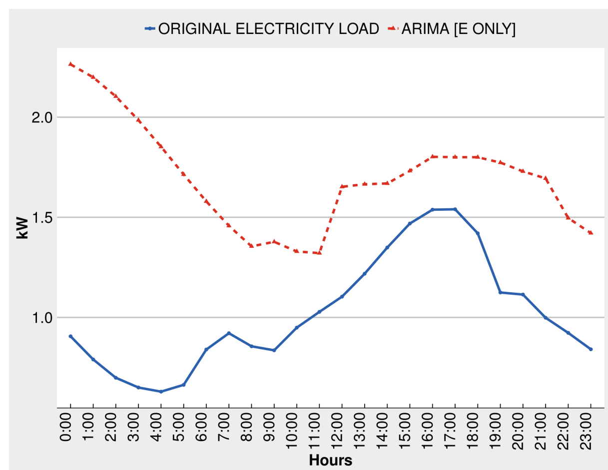


Fig. 4. Predicted data vs Original data

The above figure shows the predicted values and the original load values for 1 January 2016. The MAPE in in this case is 9.22%. To understand the effect of weather on electricity consumption, an ARIMA Transfer function model is built considering weather and traffic parameters.

### *B. CASE 2 : ARIMAX model considering Load-Weather relationship*

Based on the causal model achieved, since humidity and Temperature were seen to have a direct causal relationship with electricity consumption, an ARIMAX model was built using these values along with past values of electricity for forecasting electricity load over a period of 24 hours. Figure below shows the predicted values vs the original value of electricity consumption for this case.

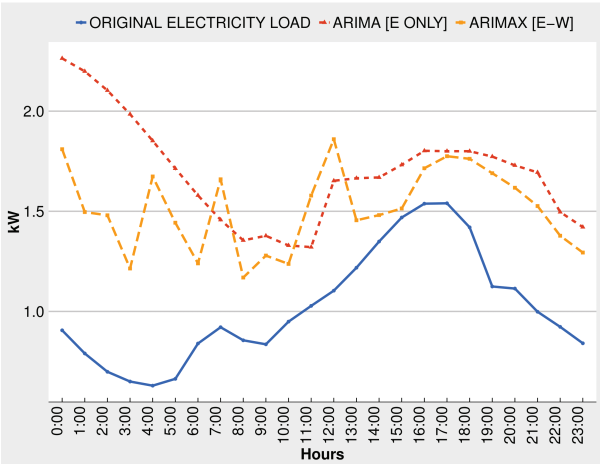


Fig. 5. ARIMA prediction based on Load Weather relationship

When forecasting electricity based on past values of electricity and weather data, the MAPE is reduced to 6.281%.

### C. CASE 3 : ARIMAX model considering Load-Weather and Traffic pattern relationship

The ARIMAX model was fitted combining the different time series data to forecast electricity load. We consider the past values of electricity, traffic and weather data to fit a model predicting electricity data output. As a result, it is observed that with an increase in number of predictor variables, there is a significant improvement in the prediction analysis. Figure below shows the prediction based on ARIMA Transfer function model. It is noted that the prediction accuracy improves and thus the predicted values show a very similar behaviors as the original load profile when compared to values predicted by univariate ARIMA model.

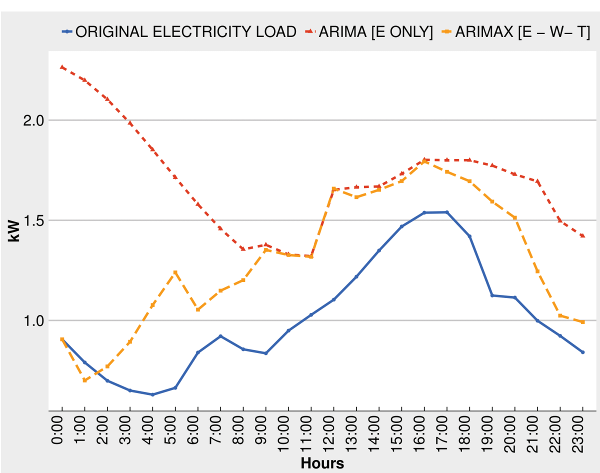


Fig. 6. Predicted data vs Original data for 1 day

In this case, where we forecast electricity based on past values of electricity, weather and transportation data as well, the MAPE has been reduced significantly to 3.38%.

TABLE I
MAPE FOR FORECASTING BASED ON ARIMA, ARIMA TRANSFER FUNCTION MODEL WHERE E IS ELECTRICITY, T IS TRAFFIC AND W IS WEATHER.

| ARIMA (E only) | ARIMAX (E-W) | ARIMAX (T,W,E) |
|---|---|---|
| 9.22 | 6.281 | 3.3 |

Table 1 lists the MAPE of the three cases considered. As seen in the table, prediction errors have significantly reduced when electricity load is forecasting based on Electricity, Weather and Traffic data.

After achieving the load forecasting for a day, the same model is applied to predict values over a period of 1 week. Fig.7 shows the forecasting for the first week of January 2016 based on both models, it is again seen that the ARIMAX model gives a better prediction when compared to ARIMA models. Therefore, considering weather and traffic parameters, our prediction error has decreased significantly.

TABLE II
MAPE FOR ARIMA MODELS(CONSIDERING JUST ELECTRICITY LOAD CONSUMPTION VALUES(E)) AND ARIMAX MODEL (CONSIDERING TRAFFIC, WEATHER AND ELECTRICITY (T,W,E)

| Days | ARIMA (E) | ARIMAX (T,W,E) |
|---|---|---|
| 1/1/2016 | 38.54 | 6.69 |
| 1/2/2016 | 20.50 | 9.11 |
| 1/3/2016 | 44.44 | 11.01 |
| 1/4/2016 | 37.13 | 8.73 |
| 1/5/2016 | 40.92 | 6.22 |
| 1/6/2016 | 38.90 | 5.63 |

Table 2 enumerates the MAPE of the predicted values over a period of one week. The original load value was compared with the ARIMA forecasted values and ARIMAX forecasted values where the effect of weather and traffic flow on electricity were included in the load forecasting. Therefore, when different factors that directly affect electricity are considered, a better load forecasting is achieved.

The above mentioned forecasting technique was applied onto different seasons of a year to check for the variations in errors. Table 3 consists of the forecasting errors of univariate ARIMA model and both the ARIMAX model, consisting of Load -Weather relationship and the Load-Weather-Traffic relationship.

## IV. CONCLUSION

In this paper, univariate ARIMA model and ARIMAX models have been built for forecasting of residential type loads. The univariate ARIMA model considers the time series data, time delay variables, auto correlation function and partial correlation functions of residuals to obtain accurate load forecasting. Whereas, the multivariate ARIMA model uses the interpreted variable of weather and transportation into regular ARIMA model to improve the prediction accuracy. It is also

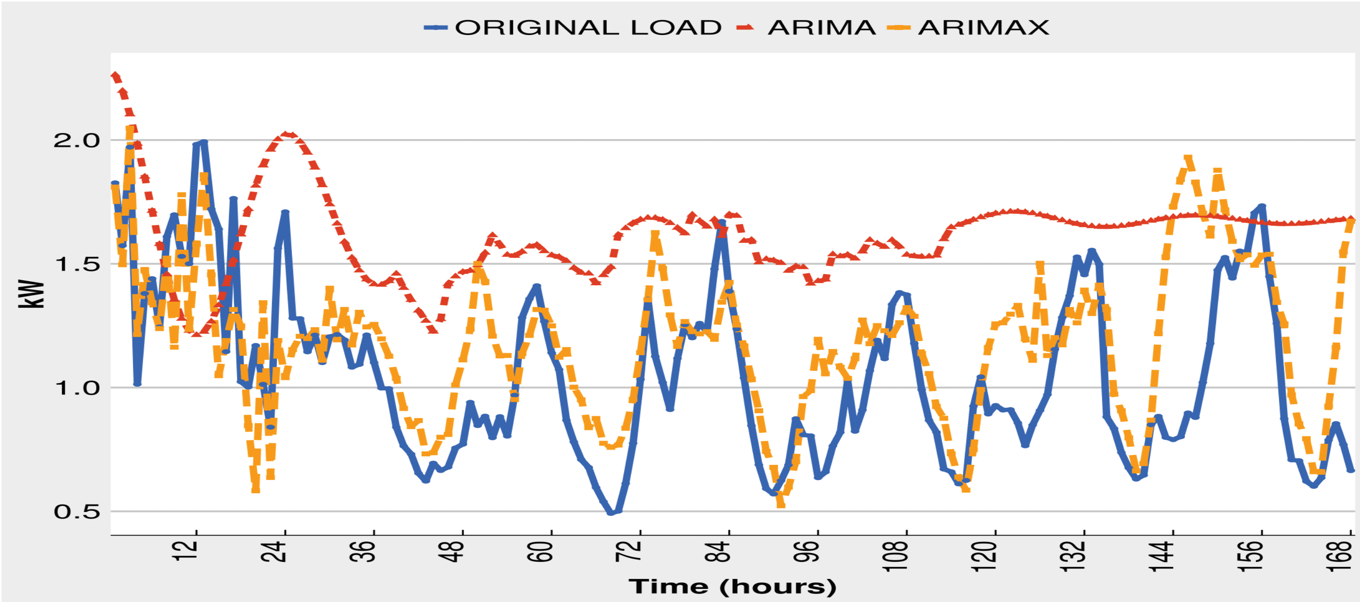


Fig. 7. Load forecasting for one week

TABLE III
MAPE FOR DIFFERENT SEASON'S FORECASTING BASED ON ARIMA, ARIMA TRANSFER FUNCTION MODEL WHERE E IS ELECTRICITY, T IS TRAFFIC AND W IS WEATHER.

| SEASONS | ARIMA (E) | ARIMAX (E-W) | ARIMAX (T,W,E) |
|---|---|---|---|
| Spring | 9.96 | 6.28 | 3.39 |
| Summer | 10.162 | 5.561 | 2.79 |
| Fall | 11.96 | 8.85 | 3.21 |

seen that, the load profile exhibited by the values predicted by ARIMAX model is very similar to the original load profile.

This study shows how different environmental and human behavioral factors can have an impact on electricity consumption and load forecasting. If with the knowledge of traffic or weather, it is possible to predict load consumption, it implies that weather forecasts and GPS real time data can come in handy whenever there is a rise or fall in electricity demand. Since accuracy of load forecasting is very crucial, we need to study more environmental factors that could affect load forecasting, thereby creating a better environment. Future work involves investigating other input parameters and integrating other types of load such as Industrial and Commercial buildings.

## ACKNOWLEDGMENT


The authors would like to thank the City of Tallahassee and Florida Department of Transportation for providing data, their support and expertise. The contents of this paper represents the authors opinion and not the official view of City of Tallahassee. This research is supported by the U.S. NSF award 1640587.


## REFERENCES


[1] D.C.Park, M.A. El-Sharwaki, R.J. Marks, L.E.Atlas and M.J.Damborg *Electric load forecasting using An Artificial Neural Network*, May 1991.

[2] Martin T. Hagan and Suzanne M. Behr *The Time series approach to Short Term load forecasting*, August 1987.

[3] George Gross and Francisco D. Galiana *Short-Term Load Forecasting*, December 1987.

[4] Jingrui Xie, Ying Chen, Tao Hong and Thomas D. Laing*Relative Humidity for Load Forecasting Models*, 2016

[5] James W. Taylor and Roberto Buizza *Using weather ensemble predictions in electricity demand forecasting*, 2002

[6] C.W.J Granger *Investing Causal Relations by Econometric Models and Cross-spectral Methods*, August 1969.

[7] Judea Pearl *The Causal Foundations of Structural Equation Modeling*, June 4, 2012

[8] Zhe Wang, Ahmed Alahmadi David C. Zhu, and Tongtong Li, *Causality Analysis of fMRI Data Based on the Directed Information Theory Framework*, 2016

[9] Markus Kalisch, Peter Buhlmann *Estimating high dimensional Directed Acyclic Graphs with the PC Algorithm*, 2007

[10] Zhe Wang, Ahmed Alahmadi David C. Zhu, and Tongtong Li, *A new hybrid correction method for shirt-term load forecasting based on ARIMA, SVR and CSA*, 2013

[11] Markus Kalisch, Peter Buhlmann *Univariate versus Multivariate Models for Short-term Electricity Load Forecasting*, 2015

[12] Peter Sprites, Clark Glymour and Richard Scheines, *Causation, Prediction and Search*, 2016

[13] Cunlu Zou and Jianfeng Feng *Granger Causality vs. dynamic Bayesian network inference: a comparative study*, 2009

[14] Abdollah Kavousi Fard and Mohammad-Reza Akbari-Zadeh *A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting*, 2014

[15] Karin Schermelleh-Engel, Martin Kerwer and Andreas G.Klein *Evaluation of model fit in nonlinear multilevel structural equation modeling*, frontiers in Psychology, March 2014.

[16] Mir Hamza Mahmood and Mark R. Bell *On Bayesian Transmit Signal Design using Information Theory*, IEEE 2014.

[17] M.Y.Cho, J.C.Hwang and C.S.Chen *Customer Short term Load forecasting by using ARIMA transfer function model*, IEEE 1995.

[18] Jiann-Fuh Chen, Wei-Ming Wang and Chao-Ming Huang *Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short term load forecasting*, Elsevier 1995