Causal Markov Elman Network for Load Forecasting in Multinetwork Systems

Lalitha Madhavi Konila Sriram, Mostafa Gilanifar, Yuxun Zhou, Eren Erman Ozguven, and Reza Arghandeh, *Senior Member, IEEE*

***Abstract*—This paper proposes a novel causality analysis approach called the causal Markov Elman network (CMEN) to characterize the interdependence among heterogeneous time series in multinetwork systems. The CMEN performance, which comprises inputs filtered by Markov property, successfully characterizes various multivariate dependencies in an urban environment. This paper also proposes a novel hypothesis of characterizing joint information between interconnected systems such as electricity and transportation networks. The proposed methodology and the hypotheses are then validated by information theory distance-based metrics. For cross validation, the CMEN is applied to the electricity load forecasting problem using actual data from Tallahassee, Florida.**

***Index Terms*—Causality, Elman network, electricity network, load forecasting, neural network, transportation network.**

I. Introduction

UNDERSTANDING the cause–effect relationships among different variables or objects has been an important step in most of the natural and social sciences over the history of human knowledge. Causality is utilized to understand the flow of information between different systems, and there have been several studies to understand the causality and its underlying factors. Lately, causality methods such as structural equation models (SEMs) have gained increased attention in the machine learning community [1], [2]. There has also been an increase in the use of deep learning algorithms in the predictive modeling area [3]. However, neural networks and deep learning algorithms have not been embraced much in the causal relationship methodologies. This paper takes a step toward developing a causation tool with recurrent neural networks (RNNs) for a better understanding of causal relationships in multiple heterogeneous time-series applications.

RNNs constitute a standard architecture designed for dealing with time-series data streams and has been applied in various disciplines such as clinical trial, predicting DNA structures [4], and machine translation [5]. To the knowledge of authors, there are very few applications of neural network structure for causality analysis. The study [6] develops a form of neural network for causal prediction known as the entangled recurrent neural network (ERNN). The ERNN approach for causal prediction propagates the backward hidden states of a bidirectional Recurrent Neural Network through an additional forward hidden layer. A simpler network structure is thus being proposed in this paper for the prediction of causal relationships, with a single direction propagation that consists of one hidden layer and one feedback layer. The simpler structure of the proposed causal Markov Elman network (CMEN) method reduces its complexity in addition to the ease of implementation on large-scale datasets. The proposed CMEN method is primarily validated comparing the outcomes with the state-of-the-art causality techniques. This comparison is evaluated based on information theory metrics called the maximum-mean discrepancy (MMD) [7] and Kullback–Leibler divergence (KLD). The achieved causal relationships are then used as a preprocessing step before load forecasting. The results of load forecasting thus serves as a cross validation of causality by utilizing the joint information between predictors that are down selected with causality.

Electricity load forecasting is an integral part of power transmission and distribution networks operation and planning. It is further categorized as short-term forecasting for 1 h to 1 week and long-term forecasting where the forecast is made over a period greater than 1 week. There have been many load forecasting studies conducted in the past, which mainly use the historical electricity consumption data as the only source for time-series forecasting [8]. There have also been some literature characterizing the dependence of electricity consumption on external factors such as weather and environment [9], [10]. However, other than external factors such as weather and environment, human

Manuscript received October 19, 2017; revised January 29, 2018, April 14, 2018, and May 24, 2018; accepted June 12, 2018. Date of publication August 15, 2018; date of current version September 28, 2018. This work is supported in part by U.S. National Science Foundation Award 1640587 and in part by U.S. DOE Award RADIANCE GMLC. *(Corresponding author: Lalitha Madhavi Konila Sriram.)*

L. M. Konila Sriram is with the Department of Electrical and Computer Engineering, Florida State University, Tallahassee, FL 32310 USA (e-mail: lk14f@my.fsu.edu).

R. Arghandeh is with the Department of Computing, Mathematics and Physics at the Western Norway University of Applied Sciences (HVL), 5020 Bergen, Norway (e-mail: r.arghandeh@fsu.edu).

M. Gilanifar is with the Department of Industrial and Manufacturing Engineering, Florida State University, Tallahassee, FL 32304 USA (e-mail: mostafa.gilanifar@gmail.com).

Y. Zhou is with the Department of Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA 94720 USA (e-mail: yxzhou@berkeley.edu).

E. E. Ozguven is with the Civil and Environmental Engineering Department, Florida A&M University-Florida State University College of Engineering, Tallahassee, FL 32310 USA (e-mail: eozguven@fsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2018.2851977

mobility is also an important factor that influences the electricity consumption at different locations and times in an urban area. Human mobility is represented by location of citizens at any given point of time. Generally, when inhabitants of a city move from one location to another, the electricity consumption at the former location decreases, and consequently, the consumption at the latter increases when that person reaches the final destination. This critical link between the electricity and mobility patterns helps us in hypothesizing the characteristic dependence between electricity consumption and transportation.

The contributions and novelties in this paper are listed as follows:

1) Proposing a new causality analysis method using Elman network with global Markov property-based filtering. This proposed methodology is named as CMEN.
2) Using CMEN to propose a novel hypothesis of characterizing joint information between interconnected systems such as electricity and transportation networks in a city.
3) Adding the proposed CMEN method as a preprocessing block to the time-series forecasting algorithms and utilizing the proposed mobility hypothesis to improve electricity forecasting accuracy.

This paper is organized as follows. Section II provides details about CMEN causality and compares it with other established causality methods. This is followed by the description in Section III. Section IV presents the overall analysis, comparison, and validation of CMEN causality model with the electricity load forecasting application. Two different information theory-based indexes have been used to validate causal relationships. In addition, five different time-series forecasting methods have been used and compared to illustrate the impact of including causal relationships into the load forecasting problem. Finally, Section V presents the summary, advantages, and limitations of the study, further mapping the future directions.

II. Methodology

In this paper, we propose the CMEN methodology for causal relationship characterization. CMEN helps us to characterize the interdependencies between urban infrastructure networks such as electricity and transportation due the interdependent nature of city's infrastructure and citizens in the context of urban mobility. The proposed electricity load forecasting framework includes the casual engine for selecting the most informative time series among the electricity, transportation, and weather datasets. By doing a causal analysis, we can select the most informative predictor for electricity load forecasting. This, in return, increases the information richness and reduces the uncertainty thereby increasing the accuracy of load forecasting.

Input variables to the CMEN causal model are historical data of the electricity consumption, weather, and traffic counts (number of vehicles on roadways). Electricity consumption of a neighborhood consisting of 222 houses has been archived and analyzed for a sample size of 1 year, which includes the electricity consumption data are collected every 30 min. Similarly, traffic count along the neighboring highway is monitored for a period of 1 year, where the data are collected every hour.

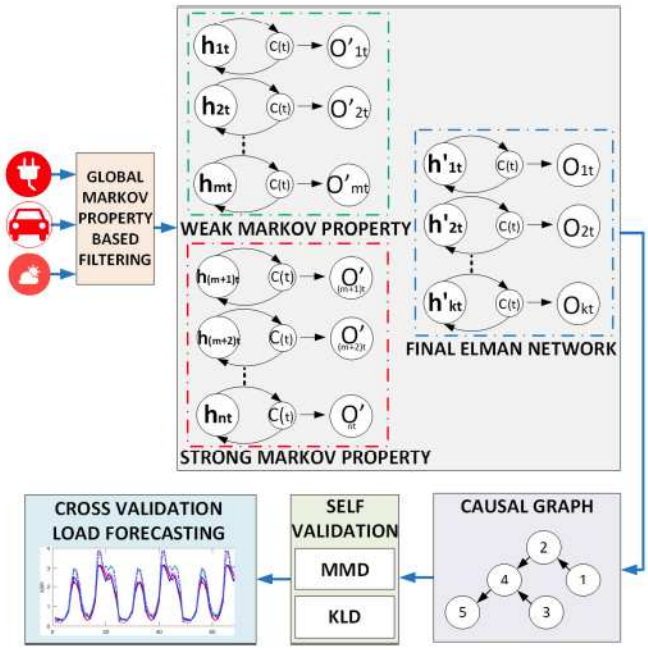

Fig. 1. Overview of the proposed CMEN structure.

A. CMEN Method

The Elman network used in CMEN belongs to the class of RNN structures. The architecture of CMEN is illustrated in Fig. 1.

The inputs to the CMEN structure are filtered based on the Global Markov property. A stochastic process is known to contain Markov property if the conditional distribution of future states of the process (conditional on both past and present states) depends only upon the present state and not on the sequence of future events.

Definition 1 (The Global Markov Property): A probability distribution P satisfies Global Markov property for an undirected graph G if for any arbitrary disjoint subset of nodes X, Y, Z such that Z separates X and Y on the graph G, the distribution satisfies $X \perp\!\!\!\perp Y | Z$.

Global Markov property consists of many conditional independence relations where it is feasible to test for a subset of those independencies that implies to others. A local Markov property designates a smaller set of conditional independence relations that implies all other conditional independence relations that is held under the Global Markov property [11].

For a better clarification of using Global Markov property for down selecting variables, a graphical model is illustrated in Fig. 2. Consider the entire system of nodes as V in this figure. As observed, the direct dependence is characterized by the edges as shown in Fig. 2(b), between nodes 1 and 2, nodes 2 and 4, nodes 3 and 4 that are the down selected group of variables as seen in Fig. 2(c). As seen in Fig. 1, the input data in this paper include historical data of electricity, transportation, and weather. The inputs are grouped into two categories based on strong and weak Markov property.

The categorized strong and weak Markov based data are fed into two Elman networks as shown in Fig. 1. Time-series vari-

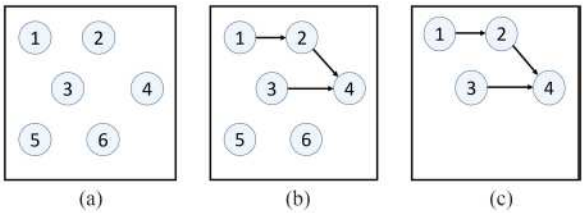

Fig. 2. Example of Directed acyclic structure representation based on Global Markov property.

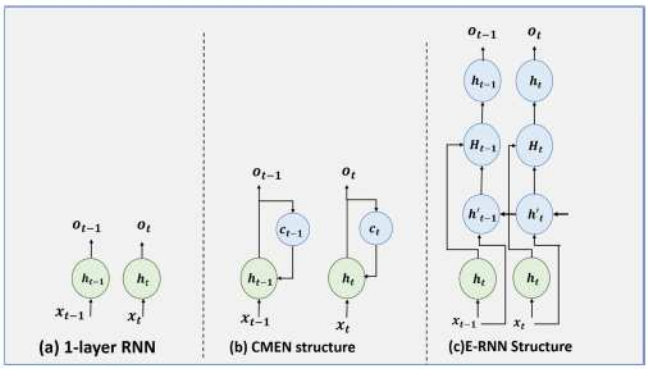

Fig. 3. Comparing CMEN with other standard neural network structures.

ables are fed into each Elman networks, which are evaluated for conditional independence to establish causal relationships in the form of edges in the output causal graph.

As observed in Fig. 1, Elman network 1 consists of m inputs selected with weak Markov property and rest of the inputs consisting the strong Markov property $m+1,\ldots,n$ are fed into Elman network 2. The outputs of these networks combined together form the input to the final Elman network that gives the causal predicted outcome.

In an Elman network, the hidden layer outputs are used as feedbacks onto itself through a context layer. Therefore, the difference between Elman network and the other two layers networks lies in the connection of the first recurrent layer. The presence of a context layer provides a delay in the connection storing values from the previous time step, which can be used in the current time step.

As seen in Fig. 3, a standard RNN consists of one forward hidden layer, with a single input and output layer. CMEN has four layers of function, the input layer, hidden layer, feedback, and the output layer. This network structure consists of one hidden layer and one output layer. The hidden layer neurons and the output layer neurons use nonlinear hyperbolic-tangent-sigmoid and pure linear activation functions, respectively. x_{t-1} and x_t are the input time series that have been selected based on strong Markov property between them, whereas $h_t, h_{t-1},\ldots$ are the forward hidden layer. In the ERNN structure, there is an additional backward hidden layer $h'_t, h'_{t-1},\ldots$, which is then added to the concatenate layer [6]. Due to the complexity of ERNN and its additional layers, it usually has longer computation time as compared with CMEN due to the simpler structure of CMEN.

The output o_t of CMEN network is shown in Fig. 3(b)

$$O_t = G + Vh_t \tag{1}$$

$$h_t = [\sigma(b + Wh_{t-1} + Wc_t + Ux_t)]. \tag{2}$$

From Figs. 1 and 3, the output layers of CMEN are given as follows:

$$O'_{t1} = R_1(n)[f_1W_1, f_2W_2,\ldots,f_mW_m] \tag{3}$$

$$O'_{t2} = R_2(n)[f_{m+1}W_{m+1}, f_{m+2}W_{m+2},\ldots,f_nW_n] \tag{4}$$

$$O_t = R_3(n)[O'_{t1}W'_1, O'_{t2}W'_2]. \tag{5}$$

In aforementioned equations, O'_{t1}, O'_{t2}, O_t represent the outputs of each Elman network inside the CMEN structure as illustrated in Fig 1. R(n) represents each Elman network structure and W_n represents the weights from the hidden layer. The following equations define the logistic and hyperbolic tangent activation functions

$$\sigma = \psi(z) = \frac{1}{1+e^{-z}} \tag{6}$$

$$\sigma = \psi(z) = \tanh(z) = \frac{2}{1+e^{-2z}} - 1. \tag{7}$$

The outputs from the CMEN structure form a directed graph (causal graph) that is created based on the conditional independence test from each Elman Network. That is, a lower value of conditional independence results in a directed edge while a higher value results in no edges between the nodes or disconnected nodes. In the associated causal graphs, there are various representations involved that emphasize the relations between the variables under consideration. Arrows (directed edges) represent the direct causal relationship, which represents the direction of information flow. Lines without arrows represent an association.

B. State-of-the-Art Causal Models

The proposed CMEN method is validated based on the comparison with other established causality methods such as Granger causality, Peter–Clark's algorithm, and SEM methods.

1) Multivariate Granger Causality Test: The spectral density matrix $S(\omega)$ forms an important section of multivariate granger causality, which can give many statistical inferences such as multiple coherence, autopower, and partial coherence. This approach is beneficial in both theoretical and practical terms. Consider a spectral density matrix $S(\omega)$ that satisfies $\int_{-\pi}^{\pi} \log|S(\omega)|d\omega > -\infty$, the spectral density matrix factorization theorem ensures that it can be decomposed into a set of unique minimum-phase functions

$$S(\omega) = \psi(\omega)\psi^*(\omega). \tag{8}$$

In aforementioned equation is the minimum-phase spectral density matrix factor that has a Fourier series expansion in nonnegative powers of $e^{i\omega}$: $\psi(\omega) = \sum_{k=0}^{\infty} R_k e^{i\omega}$, and ψ^* is its complex conjugate transpose [12].

2) The Peter–Clarke's Causality Test: The Peter–Clarke's algorithm, starts with assigning a multivariate dataset

to an undirected graph G comprising of a set of vertices V, is tested for conditional independence relations consisting of a significance level $0 < \alpha < 1$. The algorithm is initiated with an undirected graph, then modifies the structure of the graph by eliminating the edges with zero conditional independence, and altering again based on first-order conditional independence relations, iteratively

$$\rho_{\{i,j|k\}} = \frac{\left\{\rho_{\{i,j|\frac{k}{h}\}} - \rho_{\{i,h|\frac{k}{h}\}}\rho_{\{j,h|\frac{k}{h}\}}\right\}}{\left\{\sqrt{\{(1-\rho_{\{i,h|\frac{k}{h}\}^{\{2\}}})(1-\rho^2_{\{j,h|\frac{k}{h}\}}}\right\}}. \tag{9}$$

In the aforementioned equation, we assume that distribution P of the random vector X is multivariate normal. For $i \neq j \in 1, \ldots, \rho, k \subseteq 1, \ldots, \rho, i, j$ and ρ is the partial correlation between X^i and X^j [13].

3) Structural equation modeling: SEM is an approach used to quantize and evaluate the models revealing linear relationships among multivariate systems. SEM models consist of manifest variables, unobserved variables, or latent variables which can be independent (exogenous) or dependent (endogenous). Latent variables are mostly hypothetical elements which are not measurable quantities. The manifest variables, on the other hand, serve as the identifiers of the underlying elements [14].

C. CMEN Causal Model Validation

The validation of observed causal relationships is based on two well-known information theory-based indexes as follows:

1) Maximum mean discrepancy: MMD provides a score for each candidate solution, reflecting how well this candidate solution describes the observational data. The MMD computes the Reproducing Kernel Hilbert Spaces (RKHS) distance between two sample sets based on some kernel k defined on R^d where d is the dimension of the candidate time-series dataset. Generally, if $\mu_k(z)$ is a function defined on R^d with $\mu_k(z)(x) = \frac{1}{n}\sum_{i=1}^{n} k(z_i, x)$ associated with $z = (z_1, \ldots z_n)$, where z and z' are the original and generated candidate solutions, x is a function of the input time series, and n is the sample size of the dataset. The MMD measures the squared distance between the functions associated with samples z and z' as follows:

$$\mathrm{MMD}_k(z, z') = \frac{1}{n^2}\sum_{i,j}^{n} k(z_i, z_j) + \frac{1}{n'^2}\sum_{i,j}^{n'} k(z'_i, z'_j) - \frac{2}{n * n'}\sum_{i=1}^{n}\sum_{j=1}^{n'} k(z'_i, z'_j). \tag{10}$$

The sought score function, measuring the quality of the candidate solution from the MMD distance between the observational and the estimated samples, is noted L_G^{MMD}. As MMD describes the distance between the original and predicted dataset, smaller scores signify a better match [7].

TABLE I
PARAMETER TUNING FOR DIFFERENT LOAD FORECASTING METHODS USED IN THIS PAPER WITH 1 YEAR TRAINING DATASET

Forecasting Method	Parameter Tuning	References
ARIMAX	Function *auto.arima* in R autoregressive part AR(1) and factor p $\geq$ 0.05	[15],[16]
MLR	Function *auto.arima* from R beta weight = 0.710	[15],[9]
SVR	Function *tune* in R error tolerance $\epsilon = 2.3$	[17]
ANN	Package *neuralnet* from R Sigmoid activation function 4 hidden layer neurons	[18], [19]
RNN	Package *neuralnet* from R Sigmoid activation function 4 hidden layer neurons	[18],[20]
HMM	Package *depmixs4* from R Baum-Welch Algorithm based selection Hidden states (n) = 4	[21],[22] [23]
DNN	Package *neuralnet* from R 3 layers, 200 perceptrons epochs=10,learning rate=0.005	[3] [18]

2) Kullback–Leibler divergence: Measuring the disparity between two probability distributions over the same variable x is called the KLD denoted by D_{KL}. The KLD is a measure of the difference between the two probability distributions $p(x)$ and $q(x)$ with a discrete random variable x. The D_{KL} of $q(x)$ from $p(x)$ is derived as follows:

$$D_{\mathrm{KL}}(p(x)||q(x)) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)}. \tag{11}$$

D. CMEN Cross Validation Using Load Forecasting Problem

As a cross validation, a short-term load forecasting is performed by combining the various direct causal variables to predict the electricity consumption (E). The inputs to the different forecasting techniques include normalized form of the time-series datasets consisting of electricity consumption, traffic count, and weather. The training dataset has been varied from a dataset spanning 1 year, 6 months, 3 months, 1 month, and 1 week. The various techniques considered in this paper for load forecasting are autoregressive moving average with explanatory variables (ARIMAX), multilinear regression (MLR), support vector regression (SVR), feedforward neural network (ANN), recurrent neural network (RNN), and deep neural network (DNN). It is to be noted that state-of-the-art load forecasting techniques were performed using relevant packages in R environment with self-tuning capabilities. Table I provides some more details about the R packages used, the best parameters selected as the outcome of the self-tuning process. In the following results section, Table IV enumerates the error percentages that were obtained as a result of using the parameters from Table I.

The causality graphs were primarily validated using MMD and KLD. The direct causal variables were then utilized in load

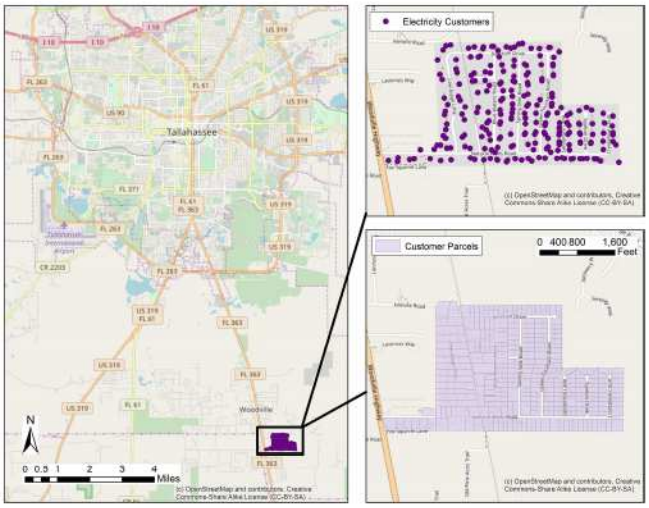

Fig. 4. Tallahassee: Southeast neighborhood highlighted.

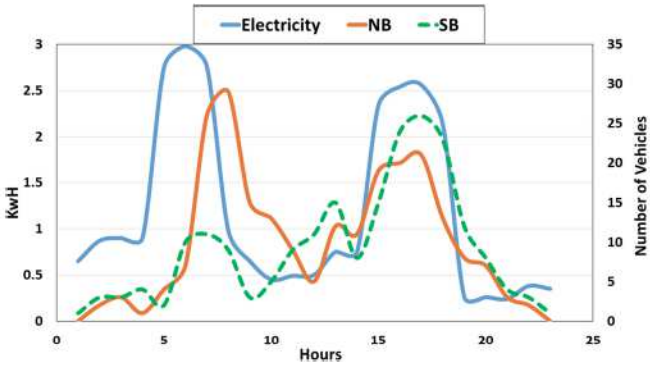

Fig. 5. Twenty-four hour curve for electricity consumption for a typical house and the traffic count on the nearby highway for the selected neighborhood as shown in Fig 4.

TABLE II
MI BETWEEN ELECTRICITY AND TRANSPORTATION

Variable 1	Variable 2	MI
Electricity	North bound Traffic	0.67
Electricity	South bound Traffic	0.26

forecasting. The load forecasting profiles are then evaluated by comparing three different error indexes MAPE, RMSE, and MAE as discussed in the Appendix.

III. Case Study

Tallahassee, the capital of Florida, is the most populated city in Leon County, hosting 180 741 population according to the U.S. census. Tallahassee area hosts six telemetered traffic monitoring sites (TTMS) that continuously collect traffic data. These data show traffic counts for every hour of each day by month and direction for any permanent TTMS location. For the case study, a TTMS site located on the Southeast Highway (2-lane, 2-way) adjacent to the neighborhood of the southeast of Tallahassee is used. This highway is the only roadway that connects this neighborhood to the downtown Tallahassee, and therefore, people use it every day for commuting purposes.

Fig. 4 shows the residential community near the selected TTMS location, where the total population is approximately 560 inhabitants. For the case study application, the 2015 electrical power consumption of approximately 4 million measurements are analyzed monthly for every customer. Our case study also consists of weather variables including dew point (D), heat index (HI), humidity (H), rain rate (R), solar irradiation (SI), temperature (T), and wind speed (W). These parameters are sampled at 30 min for year 2015.

IV. Results and Discussion

In this section, to begin the discussion regarding the proposed hypothesis of joint information between electricity and transportation networks, an example is illustrated in Fig. 5 that shows the electricity consumption for a typical house in the abovementioned neighborhood for a period of 24 h. It also shows the north and southbound traffic counts in the highway adjacent to the neighborhood. It can be observed that traffic count and electricity consumption curves exhibit similar patterns such as similar morning and evening peaks. Since the neighborhood in the case study is located in the southeast part of the city, the northbound traffic observes a peak in the morning when residents head to work and an evening peak is observed in the southbound traffic. The provided causality tests further strengthen the hypothesis of direct interdependence between electricity and traffic patterns in the neighborhood.

To validate the accuracy of the hypothesis of existence of information transfer between transportation and electricity consumption, an Information Theory-based tool, namely mutual information (MI) is adopted. MI measures the information on which variable discloses about another variable. That is, if two variables are interdependent, their MI will be greater than zero. In other words, stronger interdependence results in larger MI values.

Table II indicates the values of MI between electricity and northbound traffic as well as between electricity and southbound traffic. The positive values indicate some amount of information flow between both the datasets. Also, the higher value of MI between electricity and north bound traffic implies a higher amount of information flow between them.

The proposed CMEN causality approach is implemented on the dataset mentioned in Section III. Every dataset consists of a different sampling rate. That is, electricity consumption and weather data are sampled at every 30 min, whereas transportation datasets are achieved every 60 min. This disproportion has been corrected by using interpolation so that the interpolated transportation data values are resampled each 30 min.

As a validation for the proposed methodology, outputs of CMEN method is compared with other state-of-the-art causal methodologies. Fig. 6 illustrates the resultant causality graphs for the case study using the aforementioned causality methods. In the causal graphs, arrows signify direct causal relationship, and nodes signify different variables considered. Granger causality graph from Fig. 6(a) illustrates the direct causal relationship of electricity (E) with northbound traffic (N), heat index (HI), humidity (H), and temperature (T). On the other hand, E is observed to have some indirect association with rain

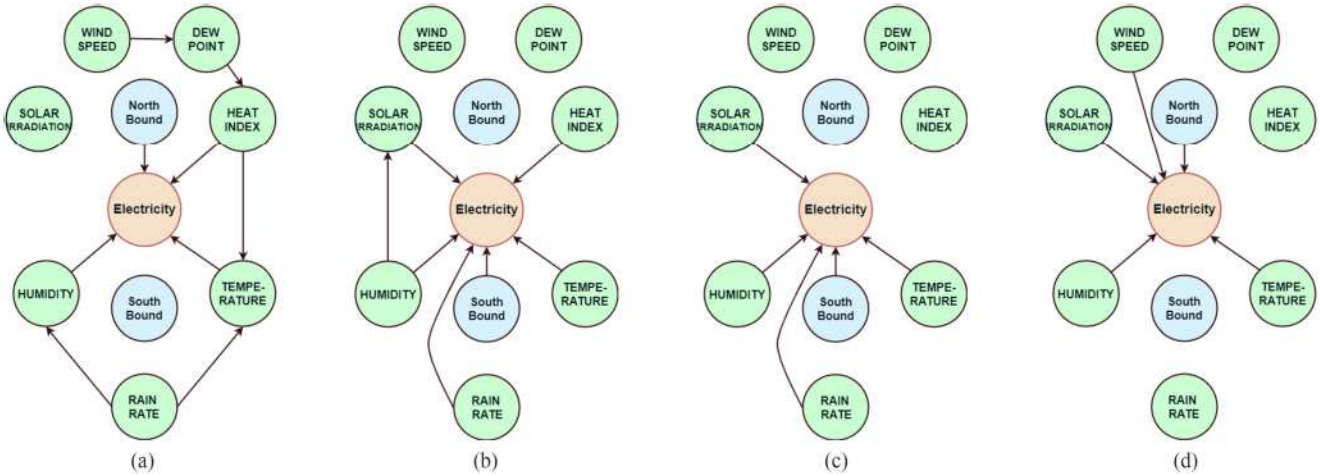

Fig. 6. Causal graphs based on a 1-year training dataset. (a) Granger Causality Test. (b) Peter–Clark's algorithm. (c) SEM. (d) CMEN.

TABLE III
MMD AND KLD FOR DIFFERENT CAUSAL MODELS

Datasets	Causal method	MMD	KLD
E-W-T	Granger	0.21	0.1792
E-W-T	PC	0.16	0.158
E-W-T	SEM	0.17	0.148
E-W-T	CMEN	0 .127	0.0798

TABLE IV
MAPE AND RMSE FOR LOAD PROFILES FORECASTED FOR 1 WEEK AHEAD BY COMBINING E-W-T DATASETS WITH A HISTORICAL TRAINING DATASET OF 1 YEAR

Method	Error Index	No Causal	Granger	PC	SEM	ERNN	CMEN
ARIMAX	MAPE	8.3	4.2	6.3	5.7	3.9	3.3
	RMSE	21.7	9.7	12.1	11.7	7.3	7.2
	MAE	12.18	8.21	10.93	9.62	7.05	6.74
MLR	MAPE	7.1	3.7	3.1	4.2	4.5	2.6
	RMSE	17.2	7.2	6.9	8.1	8.3	5.3
	MAE	15.16	8.28	11.56	10.25	8.21	6.42
SVR	MAPE	7.42	4.4	2.9	3.8	4.2	2.08
	RMSE	18	8.3	5.7	7.2	8.2	4.2
	MAE	17.8	7.97	7.24	9.6	6.47	6.13
DNN	MAPE	7.1	3.5	3.23	3.6	3.46	2.0
	RMSE	16.24	7.89	5.2	7.2	8.2	3.9
	MAE	20.54	8.02	9.3	11.2	8.09	7.01
HMM	MAPE	6.93	3.3	3.1	3.2	3.16	2.3
	RMSE	14.31	6.74	4.92	7.55	7.74	3.74
	MAE	19.65	9.81	11.62	9.64	5.23	4.11
ANN	MAPE	8.12	5.4	5.11	6.2	4.72	3.69
	RMSE	16.84	12.13	9.42	8.43	7.96	4.12
	MAE	18.32	12.11	10.61	8.46	6.11	5.2
RNN	MAPE	6.9	3.4	3.1	3.3	3.12	1.8
	RMSE	13.8	7.2	6.9	7.1	8.3	3.4
	MAE	12.61	7.96	6.9	8.12	9.16	3.98

rate R, dew point D, and wind speed W, whereas SEM implies a direct causal relationship for electricity with southbound traffic (S), T, R, H, and SI. The causal model achieved by the PC algorithm follows a similar pattern as the SEM, which illustrates that electricity is directly related to temperature, humidity, rain rate, and also the southbound traffic. CMEN implies that the electricity is impacted from temperature, humidity, rain rate, and northbound traffic flow. The CMEN is the only causal method that illustrates causal relationship of wind speed on electricity consumption. These dependencies are characterized and analyzed for the amount of information they provide in predicting electricity consumption.

To validate the causality relationships that are observed in Fig. 6, we calculate the MMD and KLD which represent the divergence of the predicted time series as explained in Section II-C. For calculating the MMD and KLD, we obtain the probability distribution function of the direct causal time-series variables represented as edges between pairwise nodes. Through the amplitude statistics, the interval $[a, b]$ (with a and b indicating the minimum and maximum of a time series $S(t) = y_t; (t = 1, \ldots, n)$ is initially divided into a finite number N_{bin} of nonoverlapping subintervals. The conventional histogram method is performed, which is established by counting the relative frequencies of the time-series variables within each subinterval. A lower value of MMD signifies a better match for the candidate solution z, which is the electricity load estimation. Similarly, a lower value of the KLD divergence indicates a better causal prediction. The validation of all four causal methods is performed by comparing the KLD and MMD values for each model. As shown in Table III, the MMD and KLD scores for CMEN are the lowest among the other causal models, therefore, CMEN provides a better causal relation between different predictors.

To understand the advantages of conducting causality analysis, we compared the load forecasted after biasing input time series based on causal models (combined electricity E, weather W, and transportation T datasets). Table IV enumerates the MAPE, RMSE, and MAE over a forecasted period of one week with 1 year training dataset.

To emphasize the effect of causal estimation as a preprocessing step for load forecasting, a noncausal-based load forecasting is performed. The result of the one week load forecasting is shown in Fig. 7. It is observed from Table IV that the MAPE, RMSE, and MAE values for load forecasting are con-

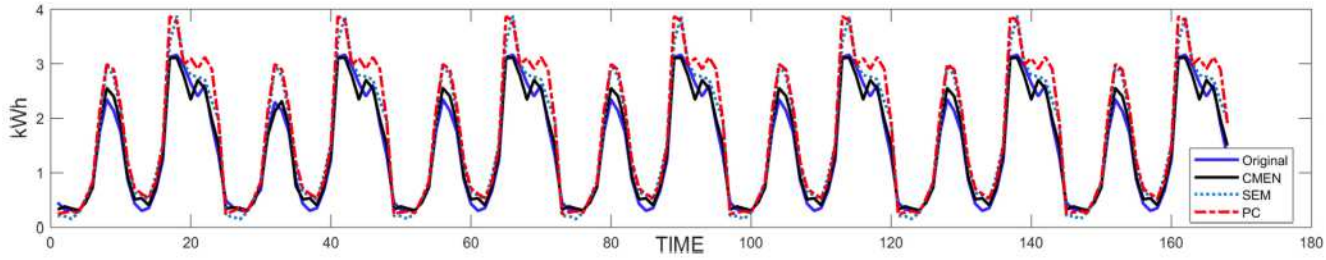

Fig. 7. One week ahead load forecasting (for an average household) performed based on CMEN causal prediction with a 1-year historical training dataset.

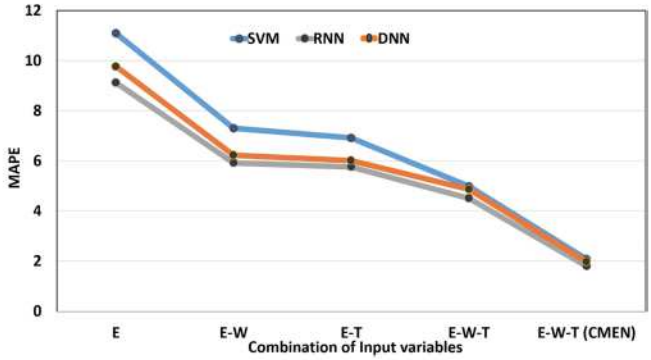

Fig. 8. MAPE for different input combination of 1 day ahead load forecasted profiles with a historical training dataset of 1 year.

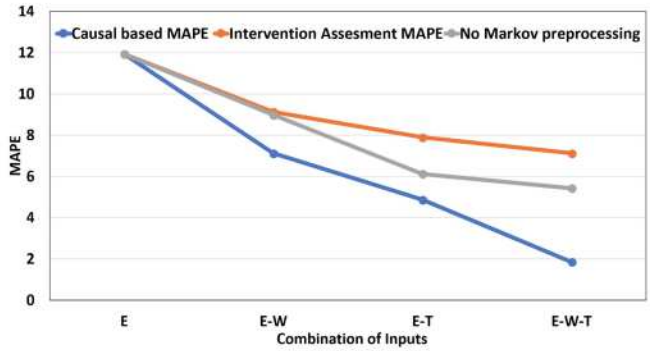

Fig. 9. Intervention assessment on a 1-day-ahead load forecasting profile for causality and Markovian based selection with a training dataset of 1 year.

siderably higher when a causality approach is not conducted. An interesting observation as seen in this table is that the CMEN outperforms the other state-of-the-art causality approaches. For example, the MAPE reduces by 73.91% when CMEN causality approach has been considered along with RNN-based load forecasting when compared with the noncausal approach. When compared with outcomes of Granger, PC algorithm, and SEM, the MAPE (forecasting error) for CMEN-based load forecasting is reduced by 47.06%, 41.94%, and 45.45%, respectively, for RNN-based load forecasting. For a more fair comparison, CMEN has also been compared with a neural network-based causal methodology called ERNN [6] that was published recently. CMEN outperforms ERNN with an error reduction of 16.66% when load forecasting has been performed.

For better validation of our proposed hypothesis regarding joint information between electricity and traffic flows, load forecasting was performed based on different combinations of input variables. The result of these different combinations is illustrated in Fig. 8.

In Fig. 8, "E" is the forecasting using the past values of electricity only. "E-W" is the forecasting based on inputs combining electricity and weather parameters. Similarly, "E-T" stands for the inputs combining electricity and transportation variables, and "E-W-T" is the combination of electricity, weather, and transportation values without considering causality relationship. Also, "E-W-T(CMEN)" signifies a combination of electricity, transportation, and weather based on CMEN causal model output. It is observed that the MAPE is higher in the case of "E only." As observed in Fig. 8, the MAPE reduces by 36.96% when load forecasting is performed by "E-W," and "E-T" reduces the MAPE further. Since "E-W-T" has more information about electricity consumption, the MAPE is reduced by 54.55%. Referring to our hypothesis in Section I, which indicates that the interdependence between electricity and transportation networks results in additional information regarding the electricity consumption and transportation patterns. The MAPE is further reduced by 73.91% when CMEN-based causality is considered.

To emphasize the introduction of causality as a preprocessor, Fig. 9 illustrates the results of using causality analysis tools such as CMEN in load forecasting, in comparison to load forecasting performed by combining the entire dataset without considering causal relationship. To validate the advantage of using global Markov property as a preprocessing step to the causal model, load forecasting was performed for the outcome of causality without Global Markov property. It is seen that, without Markov preprocessing, the MAPE is significantly higher by 42% when compared with Global Markov property-based preprocessing. An interesting observation from this figure is that the forecasting error gap between causal and noncausal approach lies in the number of variables. That is, more number of predictors implies more information. However, causality helps us in selecting the most relevant variable that consists of higher joint information factor. As seen in Fig. 9, using causality as a preprocessing block reduces the error percentage by 73.91% for the case of "E-W-T."

Another important aspect of load forecasting study is the horizon of forecast. Various combinations of training and testing sets of data are studied, and the results are depicted in Fig. 10. Three forecasting scenarios are performed including an hour ahead, a day ahead, and a week ahead forecasting. The training dataset is varied from 1 year, 3 months, 1 month, 1 week, and

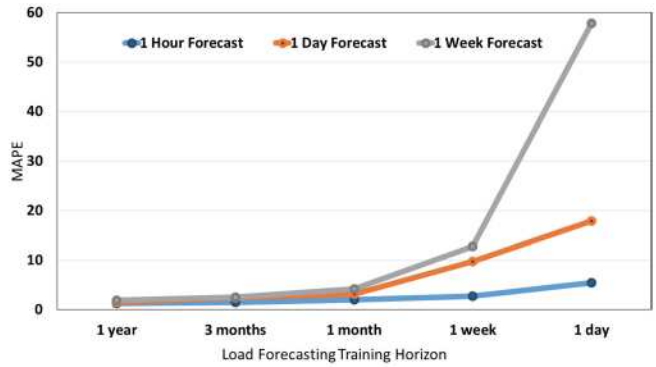

Fig. 10. MAPE for different load forecasting horizons.

1 day variables. For example, it is observed that increasing the horizon from 3 months to 1 year does not result in a considerable reduction in MAPE. Therefore, it shows that there is a saturation point in terms of the training dataset horizon. In this paper, the load forecasted for 1 week, 1 day, 1 h consisted of 3 months, 1 month, and 1 day training datasets, respectively.

Since weather plays a significant role in load forecasting, it is also important to observe the change in prediction error when load forecasting is performed in different times of the year.

V. Conclusion

This paper presented a causality-based approach in order to understand the relationship between the transportation and electricity flow in urban environments through a novel causality analysis, namely the CMEN. The CMEN was validated through the comparison with the state-of-the-art causality analysis methods. It was further evaluated using different load forecasting techniques to prove the proposed hypothesis regarding the information affinity between electricity and transportation network loads. Using a year worth of actual electricity and transportation network data in Tallahassee, Florida, the analysis showed a tremendous improvement in electricity load forecasting accuracy. For example, based on CMEN and RNN combining weather and transportation data, the MAPE for electricity load forecasting is reduced by 73.91% compared to the noncausal approach for typical load forecasting that only utilizes electricity data. Future work will be directed toward using this causality method on sparse and low rank matrices and also scaling to a larger dataset.

Appendix

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y(t)-y'(t)}{y(t)}\right| \tag{12}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left[y(t)-y'(t)\right]^2} \tag{13}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y(t)-y'(t)|}{n}. \tag{14}$$

In the above equations, $y(t)$ is the original value and $y'(t)$ is the forecasted value, whereas n is the total number of observations.

Acknowledgment

The authors would like to thank the City of Tallahassee for providing the data and expertise. The contents of this paper represent the authors' opinion and does not reflect the official view of the City of Tallahassee. Special thanks also goes to Prof. M. Zorzi and Prof. A. Chiuso from the University of Padova, Italy, for their advice and counsel about causality analysis.

References

[1] J. Pearl, "Graphs, causality, and structural equation models," *Sociol. Methods Res.*, vol. 27, no. 2, pp. 226–284, 1998.
[2] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA, USA: MIT Press, 2000.
[3] S. Ryu, J. Noh, and H. Kim, "Deep neural network based demand side short term load forecasting," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Nov. 2016, pp. 308–313.
[4] H. Zeng, M. D. Edwards, G. Liu, and D. K. Gifford, "Convolutional neural network architectures for predicting DNA–protein binding," *Bioinformatics*, vol. 32, no. 12, pp. i121–i127, 2016.
[5] J. Devlin, R. Zbib, Z. Huang, T. Lamar, R. Schwartz, and J. Makhoul, "Fast and robust neural network joint models for statistical machine translation," in *Proc. 52nd Annu. Meet. Assoc. Comput. Linguistics (Vol. 1: Long Papers)*, 2014, pp. 1370–1380.
[6] J. Yoon and M. van der Schaar, "E-RNN: Entangled recurrent neural networks for causal prediction," in *Proc. ICML Workshop Principled Approaches Deep Learn.*, 2017, pp. 1–5.
[7] A. Iyer, S. Nath, and S. Sarawagi, "Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection," in *Proc. 31st Int. Conf. Mach. Learn.*, Jun. 2014, vol. 32, pp. 530–538.
[8] P.-F. Pai and W.-C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Convers. Manage.*, vol. 46, no. 17, pp. 2669–2688, 2005.
[9] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 456–462, Jan. 2014.
[10] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *Int. J. Forecasting*, vol. 32, no. 3, pp. 585–597, 2016.
[11] J. Behrstock, V. Falgas-Ravry, M. F. Hagen, and T. Susse, "Global structural properties of random graphs," *Int. Math. Res. Notices*, vol. 2018, no. 5, pp. 1411–1441, 2018, doi: 10. 1093/imrn/rnw287.
[12] L. Barnett and A. K. Seth, "The MVGC multivariate granger causality toolbox: A new approach to Granger-causal inference," *J. Neurosci. Methods*, vol. 223, pp. 50–68, 2014.
[13] P. Spirtes and C. Glymour, "A fast algorithm for discovering sparse causal graphs," *Social Sci. Comput. Rev.*, vol. 9, no. 1, pp. 62–72, 1991.
[14] S. A. Martinez, L. A. Beebe, D. M. Thompson, T. L. Wagener, D. R. Terrell, and J. E. Campbell, "A structural equation modeling approach to understanding pathways that connect socioeconomic status and smoking," *PLOS ONE*, vol. 13, no. 2, pp. 1–17, Feb. 2018, doi: 10.1371/journal.pone.0192451.
[15] R. J. Hyndman, "Forecast: Forecasting functions for time series and linear models," R package version 8.2, 2017.
[16] K. S. L. Madhavi *et al.*, "Advanced electricity load forecasting combining electricity and transportation network," in *Proc. North Amer. Power Symp.*, Sep. 2017, pp. 1–6.
[17] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, "e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)," Tu Wien, R Package Version 1.6–8, 2017.
[18] S. Fritsch, "Training of neural networks," *R Package Version*, vol. 1, p. 33, 2016.
[19] W. Charytoniuk and M. S. Chen, "Neural network design for short-term load forecasting," in *Proc. Int. Conf. Elect. Utility Deregulation Restructuring Power Technol. (Cat. No. 00EX382)*, 2000, pp. 554–561.
[20] S. N. Yelamali, and K. Byahatti, "Electricity short term load fore-casting using Elman recurrent neural network," in *Proc. Int. Conf. Adv. Recent Technol. Commun. Comput.*, Oct. 2010, pp. 351–354.

[21] I. Visser, "Dependent mixture models—hidden Markov models of GLMs and other distributions in S4," R package version 1.3-3, 2016.
[22] D. X. Niu, B. E. Kou, and Y. Y. Zhang, "Mid-long term load forecasting using hidden Markov model," in *Proc. 3rd Int. Symp. Intell. Inf. Technol. Appl.*, vol. 3, Nov. 2009, pp. 481–483, doi: 10.1109/IITA.2009.422.
[23] M. Andersson, "Modeling electricity load curves with hidden Markov models for demand-side management status estimation," *Elect. Energy Syst.*, vol. 27, Mar. 2016, Art. no. e2265.
[24] M. Langkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, 2014.
[25] C. Esteban, O. Staeck, S. Baier, Y. Yang, and V. Tresp, "Predicting clinical events by combining static and dynamic information using recurrent neural networks," in *Proc. IEEE Int. Conf. Healthcare Inf.*, Oct. 2016, pp. 93–101.
[26] T. H. Team, "H2O: R interface for H2O," R package version 3.14.0.3, 2017.

Lalitha Madhavi Konila Sriram received the B.S. degree in electrical engineering from the Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India, in 2013. She is currently working toward the Ph.D. degree at Florida State University, Tallahassee, FL, USA, carrying out research for the Center for Advanced Power Systems.

Her research interests include Internet of Things, machine learning, and data analysis for decision support in smart cities and smart grids.

Mostafa Gilanifar received B.Sc. degree in 2011 from the Sharif University of Technology, Tehran, Iran, and the M.Sc. degree in 2014 from the Amirkabir University of Technology, Tehran, both in industrial engineering. He is currently working toward the Ph.D. degree at the Department of Industrial & Manufacturing Engineering, Florida State University (FSU), Tallahassee, FL, USA.

His research interests include data analytics for complex systems, machine learning algorithms, and statistical learning models for heterogeneous spatiotemporal measurement data.

Yuxun Zhou received the Diplome d'Ingenieur degree in applied mathematics from Ecole Centrale Paris, Châtenay-Malabry, France, and the B.S. degree from Xi'an Jiaotong University, Xi'an, China. He is currently working toward the Ph.D. degree at the Department of EECS, UC Berkeley, Berkeley, CA, USA.

His research interests include machine learning theories and algorithms for modern sensor rich, ubiquitously connected cyber-physical systems, including smart grid, power distribution networks, smart buildings, etc.

Eren Erman Ozguven received the Ph.D. degree in civil and environmental engineering from Rutgers University, New Brunswick, NJ, USA, in 2012 with concentration in emergency supply transportation operations.

He is currently an Assistant Professor with the Department of Civil and Environmental Engineering, Florida A&M University-Florida State University, Tallahassee, FL, USA. His research interests include smart cities, urban mobility, traffic safety and reliability, emergency transportation, and intelligent transportation systems.

Reza Arghandeh (S'01–M'13–SM'18) received the Ph.D. degree in electrical engineering with a specialization in power systems from Virginia Tech, Blacksburg, VA, USA, 2013.

He has been a Postdoctoral Scholar with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, 2013–2015. He is an Associate Professor in the Department of Computing, Mathematics and Physics at the Western Norway University of Applied Sciences (HVL), Bergen, Norway. He was an Assistant Professor in ECE Dept, Florida State University, USA from 2015–2018. He was a Power System Software Designer at Electrical Distribution Design, Inc., Virginia, USA, 2011–2013. His research interests include data analysis and decision support for smart grids and smart cities.