

# Aggregated Residential Load Modeling Using Dynamic Bayesian Networks

Maria Vlachopoulou  
PNNL, Richland  
WA 99354, USA

Email: maria.vlachop@gmail.com

George Chin  
PNNL, Richland  
WA 99354, USA

Jason Fuller  
PNNL, Richland  
WA 99354, USA

Shuai Lu  
PNNL, Richland  
WA 99354, USA

**Abstract**—It is already obvious that the future power grid will have to address higher demand for power and energy, and to incorporate renewable resources of different energy generation patterns. Demand response (DR) schemes could successfully be used to manage and balance power supply and demand under operating conditions of the future power grid. To achieve that, more advanced tools for DR management of operations and planning are necessary that can estimate the available capacity from DR resources. In this research, a Dynamic Bayesian Network (DBN) is derived, trained, and tested that can model aggregated load of Heating, Ventilation, and Air Conditioning (HVAC) systems. DBNs can provide flexible and powerful tools for both operations and planning, due to their unique analytical capabilities. The DBN model accuracy and flexibility of use is demonstrated by testing the model under different operational scenarios.

**Keywords**—Aggregated load, Bayesian networks, demand response, load modeling

## I. INTRODUCTION

Demand response has become an essential element in the smart grid [1]. Thermostatically controlled loads, such as HVAC and electric water heaters, are the most promising candidates for demand response. These loads represent a large portion of power consumption by residential and commercial customers, coincide with system peak, and have negligible impact on customer comfort when controlled properly.

Many demand response methodologies and techniques have been proposed in the literature and implemented in smart grid demos or utility practice, including time-of-use rate schedules, direct load control, transactive control [2], real-time pricing [3], online exchange for DR capacity [4], and others. Through these various approaches, DR becomes a resource for power system operators to manage the balance of generation and load. However, one important piece of information to the operators has not been well addressed, i.e., the available capacity from DR resources as a function of time/time of day, in terms of both power and energy. Especially for residential DR resources, because of the large number of end-use devices involved, e.g., HVACs, it is usually unclear to the operator how many MW of HVAC load is online and under control, and how many MWh of energy consumption the HVAC devices would be able to reduce or increase. A model that can help estimate the availability of DR resources is needed to more effectively utilize DR resources in both operations and planning. Some previous research has been performed along this direction [5].

The focus of this paper is to build a Dynamic Bayesian Network model for the prediction of aggregated HVAC loads

online for DR resource management. The model can be used for providing information for the next operation day, operations in real time, or for planning studies of different time horizons. DBN modeling has been shown to accurately model aggregated water heater load, another thermostatically controlled load, in [6].

In this research, DBNs were chosen for modeling due to some of their unique properties that can provide value for their use as a tool for DR management of operations and planning. DBNs afford specific analytical capabilities including

- Providing an intuitive graphical model that is directly related to the physical or conceptual model
- Supporting both diagnostic (tracing roots/causes of effects) and predictive (forecasting future measurements) analysis capabilities
- Utilizing probability distributions that naturally incorporate uncertainty
- Potential to facilitate temporal analysis where time is explicitly encoded into the graphical model

Compared to other modeling methods that provide similar accuracy for aggregated load modeling, DBNs have been found to produce query results in less time, once trained. They allow the user to have control of all external variables, and to observe the influence that the change of these variables has on the queried output. Modification of the values of external variables does not require retraining of the Bayesian network. The flexibility, ease, and low computational cost that characterize DBN models renders them unique and extremely valuable in the area of DR management.

General background information on BNs and DBNs is provided in Section II. The DBN model developed for aggregated HVAC load modeling is described in detail in Section III. Section IV includes details of the DBN model application for load modeling and a set of results that illustrates the accuracy of the model under different operating scenarios. Finally, in Section V some conclusive remarks with regards to this research are presented.

## II. DYNAMIC BAYESIAN NETWORKS

A *Bayesian network* (BN) is a probabilistic graphical model, where nodes represent random variables and directed arcs/edges represent conditional dependencies. Every random variable has an associated conditional probability table which

contains the probabilities of the variable being assigned to specific values or states based on the values of parent variables, which is denoted as  $P(X_i|Pa(X_i))$ , where  $Pa(X)$  relate to the immediate parents of  $X$ . Probabilities are derived from collected data or prior knowledge. The full joint distribution is given by  $P(X_1...X_n) = \prod_i P(X_i|Pa(X_i))$ . Once a BN is constructed, the values of certain variables are set based on evidence or observations. We then compute the posterior probabilities of *query variables* given the set of *evidence variables* as knowledge. Inferencing refers to the propagation of the evidence through the BN followed by the computation of the updated probabilities of the query variables.

A *Dynamic Bayesian Network* (DBN) models the stochastic evolution of a set of variables over time. In a DBN, discrete time is introduced and conditional distributions are related to parent variable values of the previous time point. The process is typically modeled as discrete time-slices and denoted as  $P(X^{(1)}, ..., X^{(t)}) = P(X^{(1)})P(X^{(1)}|X^{(2)})...P(X^{(t)}|X^{(t-1)})$ , where  $X^{(t)}$  is state at time  $t$ . A DBN satisfies the Markov assumption that variables in  $X^{(t+1)}$  cannot depend directly on variables in  $X^{(t')}$  for  $t' < t$ . Since current events lead to future events, but not vice-versa, arcs always flow forward in one direction in a DBN. A common example of a DBN is a hidden Markov model. Fig. 1 shows a DBN where a 5-node BN is replicated across three time-slices with dependencies flowing only within a time-slice (red edges) or forward to the next time-slice (grey edges).

DBNs have been found to be useful in science and engineering applications such as in simulation meta-modeling [7], climate change modeling [8], and gene regulatory network identification [9]. The size of a DBN generally multiplies out the number of random variables by the number of time-slices that are tracked. While the size of a DBN may grow rapidly, the cognitive load to users is fairly fixed to the set of variables of a particular time-slice with the additional understanding that the variables values or states will evolve over time. Most existing DBN models are still limited in size to hundreds of variables as the number of time-slices is often small.

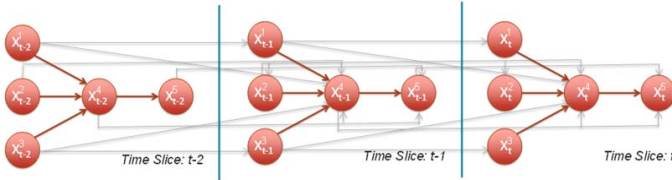


Fig. 1. 3-time-slice Dynamic Bayesian Network

### III. HVAC DYNAMIC BAYESIAN MODEL

DBNs can be applied to create models for aggregated residential load of a specific end-user appliance. For example, the aggregated load of a neighborhood of houses resulting from the use of only HVACs or only water heaters can be successfully modeled with DBNs. In this section, the application of DBNs to model aggregated HVAC load is described. The final structure of the network is shown in Fig. 2 and the node variables of the network are listed and explained in Table I.

The ability of DBNs to model changing behavior over time is an important feature for load modeling, because load is

highly correlated to time and season-dependent variables. A 2-time-slice DBN has been deemed appropriate for modeling aggregated HVAC load. The left-most time-slice of the network will be referred to as the first time-slice, and the variables in it are denoted with subscript 1 as shown in Fig. 2. Time-slice 1 captures the present time behavior. The right-most time-slice of the network will be referred to as the second time-slice, and subscript 2 is used for its variables. Time-slice 2 captures the future time-step behavior. The definition of time-step depends on the training and testing data rate and will be discussed in Section IV.

There are diverse ways of deriving the network structure, therefore for a given modeling problem there are more than one DBN structures that can produce satisfactory modeling results. In this research, the structure of the network per time-slice is based on an existing mathematical model, the Equivalent Thermal Parameter (ETP) model, described in detail in [11] and [12]. The ETP model captures the relationships between a number of physical variables, some of which are listed in Table I. By combining an existing validated mathematical model, such as the ETP model, and the DBN probabilistic modeling features, a robust and powerful DBN model is derived. This DBN can model the behavior of multiple HVACs operating under different conditions and settings.

Examining the ETP model variables and their relationships aids in determining which variables will be used as nodes of the DBN and which edges will be added between nodes to express the existence of a relationship between them. The structure of the network is explained by concentrating on the first time-slice and starting from its right-most node, the load variable, then moving leftwards until the left-most nodes of the slice are reached.

The load is driven by  $Q_h$ , or the amount of thermal energy removed or added to the house by the HVAC system, where a negative value indicates cooling and a positive value indicates heating. The ON/OFF cycle is controlled by a thermostat, maintaining the internal air temperature within a certain bound (deadband) around the thermostat setpoint,  $T_{set}$ . The electrical load,  $load$ , is a function of  $Q_h$  and the current efficiency of the unit, or the Coefficient Of Performance (COP); as COP increases, the same amount of heat energy can be produced/removed with less electrical energy. The current COP,  $actualCOP$ , is a function of the HVAC unit's base efficiency,  $baseCOP$ , and the current outdoor temperature,  $T_o$ ; when cooling, as  $T_o$  increases, the COP decreases. These connections can be traced in Fig. 2, where the edges indicate dependency. The high degree of correlation between the load variable,  $Q_h$ , and COP justifies the edges on the network connecting  $Q_h$  to  $load$  and  $actualCOP$  to  $load$ .

The second time-slice has similar structure to the first time-slice, however the second time-slice is simpler since some variables that remain static over time do not have to be repeated in it. Static variables like,  $R_{roof}$ ,  $R_{wall}$ ,  $R_{floor}$ ,  $acph$ ,  $U_A$ ,  $C_M$ ,  $F_A$ ,  $T_{set}$ , and  $baseCOP$ , do not change over time for the purpose of this research.

Determining and validating the optimal DBN structure for aggregated load modeling is not a trivial task. For this research, the performance of multiple models has been tested and the final DBN model described above and shown in Fig. 2 has an

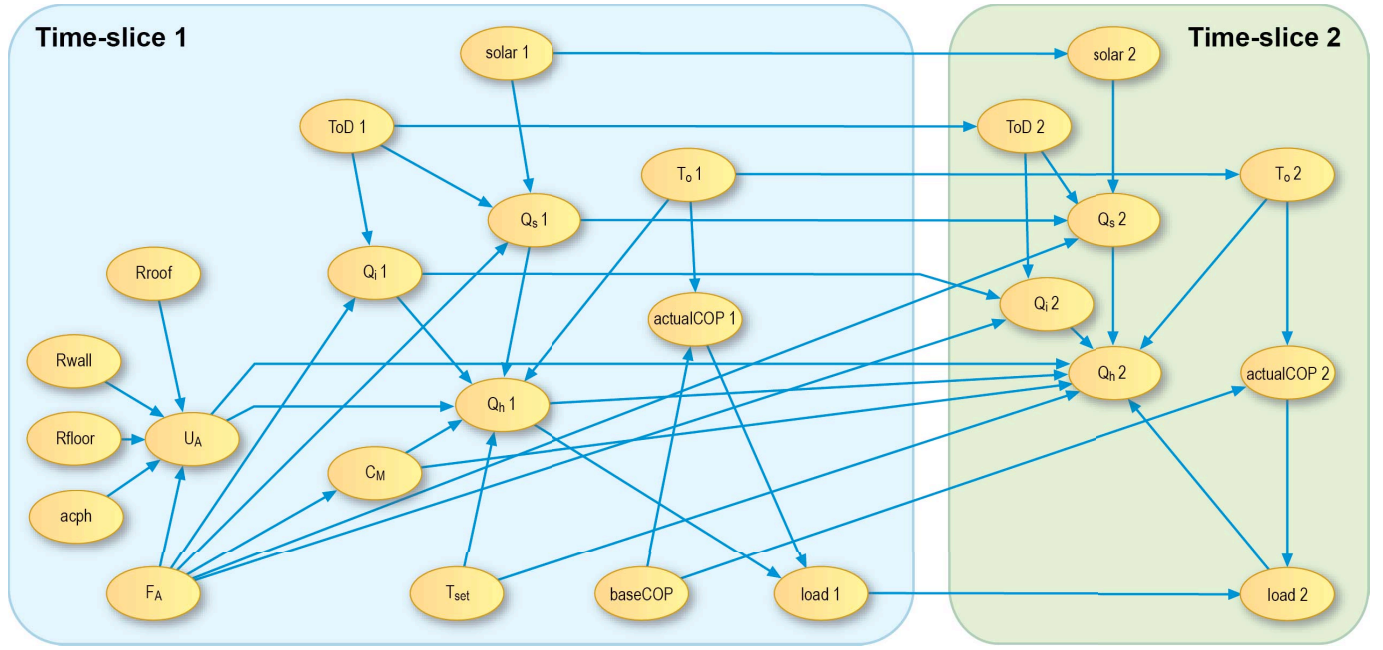


Fig. 2. HVAC Dynamic Bayesian Network

adequately good performance that will be further discussed in Section IV.

TABLE I. DBN MODEL VARIABLES

Symbol	Physical meaning
<i>Basic ETP Model Variables</i>	
$U_A$	Thermal conductivity (BTU/°F·h)
$C_M$	Amount of energy stored in house material (BTU/°F)
<i>Extended ETP Model Variables</i>	
$Q_i$	Internal thermal energy (BTU/h)
$Q_s$	solar thermal energy (BTU/h)
$Q_h$	HVAC thermal energy into/out of house (BTU/h)
<i>actualCOP</i>	Actual coefficient of performance (BTU/kWh)
<i>load</i>	Load demand (kW)
<i>External Variables</i>	
$T_o$	Outside temperature (°F)
<i>solar</i>	solar irradiation (W/s.f.)
<i>ToD</i>	Time of Day
$T_{set}$	HVAC set point
<i>baseCOP</i>	Base or rated coefficient of performance (BTU/kWh)
$F_A$	Floor area (s.f.)
$R_{roof}$	Roof R-value (°F·s.f.·h/BTU)
$R_{wall}$	Wall R-value (°F·s.f.·h/BTU)
$R_{floor}$	Floor R-value (°F·s.f.·h/BTU)
<i>acph</i>	Air changes per hour

#### IV. EXPERIMENTS

After the structure of the DBN has been determined, the network is ready for the learning and inferencing stage, which consist the fundamental stages of processing graph models. Both these stages use data, either real-world or simulated, to derive the relationship strength between network variables and to determine the value of a variable. The results presented in this section are based on Bayesian inferencing on the network.

##### A. Learning and Inference

During the learning phase, the conditional probabilities, mentioned in Section II, for each node are derived. The

conditional probability of a node expresses the likelihood of a node variable having a specific value, given the variable values of its parent nodes. The data used for learning should represent diverse operational conditions of the system that the network is aiming to model.

The data used in this research are simulated data produced by the power grid simulation tool GridLAB-D. GridLAB-D is an agent-based, open-source, power grid simulation tool developed at PNNL and the Department of Energy to simulate the complexities of the smart grid from the substation to the end-use load [13]. 1000 houses have been simulated in order to produce datasets for aggregated residential load modeling. The simulation produces results at a 5-min time-step, therefore the time-slices of the DBN model the variable value change every 5 minutes. The time step of the simulation, and therefore of the DBN model, can vary. It should be small enough to capture changes in time-varying variable behavior of the system under analysis. Seven house types are considered that capture features found in older to newer houses. House type 7 is the newest and most energy efficient since it has the highest R-values. House type 1 is the oldest and least energy efficient with the lowest R-value.

For the learning stage to account for diverse energy consumption scenarios, a mixture of houses has been simulated; that is some house are more energy efficient and larger than others. The HVAC  $T_{set}$  varied in the range  $65^{\circ}F - 70^{\circ}F$ . The base or rated COP varied in the range of 1.5 - 3.5. Data for the three summer months, June to August, are used for learning. After the learning stage is complete, inferencing can be used to query the DBN for results. The load variable of the second time-slice is the query variable that is of most interest, since this is the end-use load due to HVAC cooling in the future. Evidences, i.e. known values, have to be set on a certain number of variables prior to querying the load variable. Not all variable values need to be known as evidence in order

to query the query variables. This is one of the strengths of DBNs since many times there are missing data, particularly when setting evidences using real-world data. Evidence is set on the first time-slice only, on nodes  $R_{roof}$ ,  $R_{wall}$ ,  $R_{floor}$ ,  $acph$ ,  $F_A$ ,  $T_{set}$ ,  $baseCOP$ ,  $ToD$ ,  $T_o$ , and  $solar$ .

### B. HVAC Results

The DBN model for aggregated HVAC load provides both an accurate and flexible mechanism for modeling. In this section, these two aspects of its functionality will be demonstrated and discussed. Performance accuracy of the DBN model is essential for the model to be reliably used. Flexibility of the DBN model is equally important for the model to provide diverse and insightful information. The DBN model of Fig. 2 is tested under different scenarios of operation to validate its modeling capabilities of aggregated HVAC load. The test data are created using GridLAB-D simulation outputs for Summer 2012 (Aug 20 - Aug 26) in Yakima, WA.

The following set of figures (Fig. 3 - Fig. 5) aims in demonstrating the accuracy with which the DBN estimated the aggregated load. The comparison of the estimate is done with respect to actual load data, which represent the ground truth and which for this research are simulated data. The x-axis spans 168 points, equivalent to 7-day hourly outputs. The y-axis represents load in Watts and ranges from 0 to 5 MW. A 5 MW aggregated load is extremely rare to occur, for the purposes of the test cases of this research, so it is used as the upper bound in these figures.

A typical week-long scenario is presented in Fig. 3. COP is set to 2.5 and the house type considered is no. 4. The DBN estimate (dashed line) follows closely the actual load curve (solid line). It is observed that the load peak occurs at around 1pm to 4pm which is reasonable since during the summer months it is during those hours that the outside air temperature reaches its peak. The difference between the estimate and actual load ranges from 0 - 0.5 MW, which per house corresponds to a 0-0.5 KW. This is a small error compared the 31 KWh average house power consumption.

The base COP is one of the main driving factors affecting the quantity of HVAC load. To evaluate the DBN model performance under different base COP the model was tested for base COP within range 1.5 - 3.5. An average house of type 4 was chosen for the test. In Fig. 4 the DBN model estimate is compared to the actual load for base COP 1.5 that is a really low efficiency HVAC unit, and for base COP 3.5 that is a highly efficient HVAC unit. The low efficiency of the HVAC units used for the 1000 houses simulated justifies the higher load peaks shown in the left-hand side plot of Fig. 4 in comparison to the load peaks shown in the right-hand side plot. For the case where base COP is 1.5, the load peaks are close to 3.5 MW which are much higher than the approximately 2 MW observed in the average case represented in Fig. 3. For the case where base COP is 3.5, the load peaks are close to 1.5 MW which are lower than the average case of Fig. 3.

The house type is another important factor with regards to HVAC load. Fig. 5 shows the performance of the DBN model for both energy efficient, type 7, and non-energy efficient, type 1, houses. As expected the load peaks for house type 7 are lower, they do not exceed 2 MW, compared to the load peaks

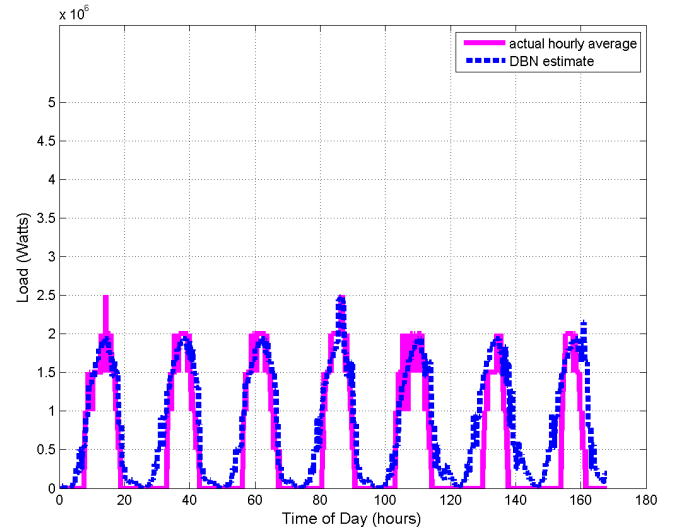


Fig. 3. Comparison of actual vs. estimate aggregated load.  $T_{set} = 70^{\circ}F$ ,  $baseCOP = 2.5$ , house type 4.

for house type 1 that are reaching 2.5 MW. In comparison to the average case scenario presented in Fig. 3, the neighborhood of inefficient houses consumes more energy than the average case, while the neighborhood of efficient houses consumes less than the average case.

The following set of figures (Fig. 7 and Fig. 6) aims in demonstrating the flexibility with which the DBN can be used to produce estimates. The DBN model, after the learning process has been complete, can be queried for different variables. So far in this discussion the load variable of the second time-slice has been the chosen variable to query, since its value is highly correlated to the estimate of the total household load. However, it is not the only variable of interest; the HVAC thermal energy,  $Q_h$ , is another variable whose value is essential for the analysis of load modeling. In Fig. 7, the query output of the DBN model of Fig. 2 is shown when  $Q_{h2}$  is used as the query node. In this case, the average case scenario is considered with base COP set to 2.5, house type set to 4, and  $T_{set}$  set to  $70^{\circ}F$ . Evidence is set on network nodes  $R_{roof}$ ,  $R_{wall}$ ,  $R_{floor}$ ,  $acph$ ,  $F_A$ ,  $T_{set}$ ,  $baseCOP$ ,  $ToD$ ,  $T_o$ , and  $solar$  of the first time slice, similar to the case when load was queried. The  $Q_h$  plot of Fig. 7 shows that the DBN can accurately estimate  $Q_h$ . Positive values of  $Q_h$  indicate heat flow caused by an HVAC unit operating on the heating cycle. Negative values of  $Q_h$  indicate heat flow caused by an HVAC unit operating on the cooling cycle. In a similar manner, other nodes can be used as query nodes, like COP and  $Q_i$ , that influence the load value. They can be queried conditioned on very detailed information of the values of other network nodes when setting evidences. The potential of such detailed analysis of the variable values demonstrates the flexibility gained by using DBN modeling for aggregated load over other methods and models.

The value of DBN modeling also stems from the fact that the query output for a variable is not just a value, but a distribution. Fig. 6 provides a graphical illustration of load marginal probability distributions. In this example, the load



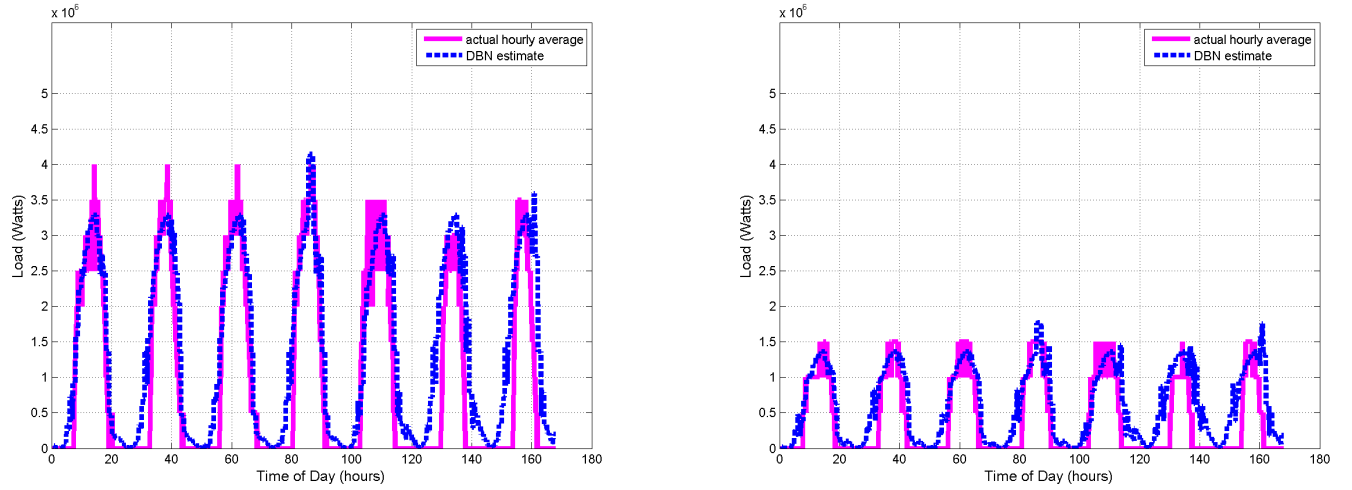


Fig. 4. Comparison of actual vs. estimate aggregated load when based COP is low at 1.5 (left-hand side plot) and when it is high at 3.5 (right-hand side plot).  $T_{set} = 70^{\circ}F$  and house type 4 for both plots.

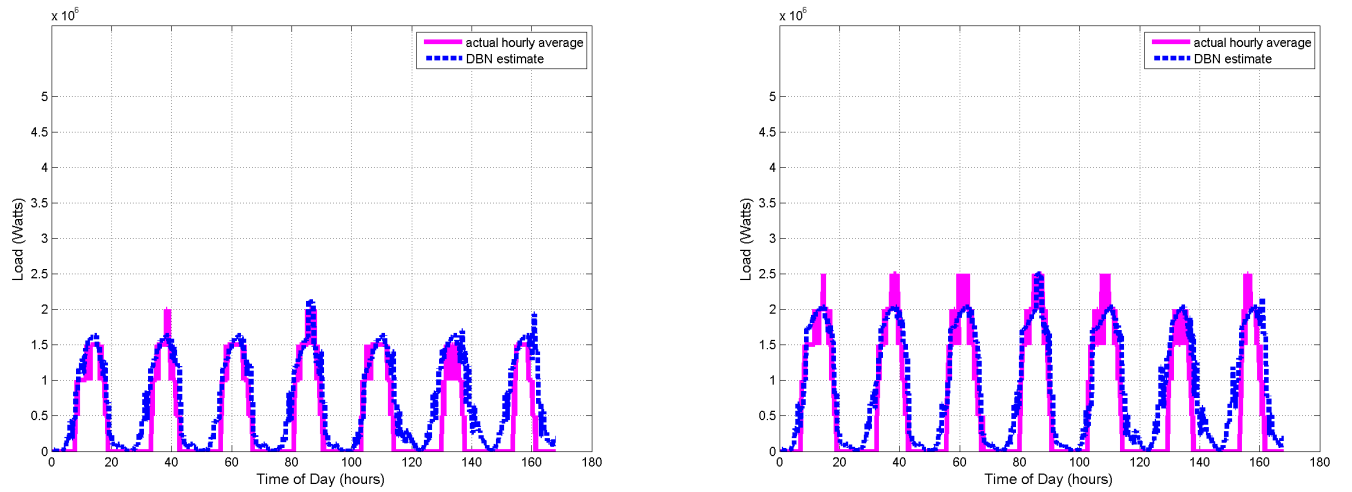


Fig. 5. Comparison of actual vs. estimate aggregated load when the house efficiency is high, i.e. house type 7 (left-hand side plot) and when it is low, i.e. house type 1 (right-hand side plot).  $T_{set} = 70^{\circ}F$  and  $baseCOP = 4$  for both plots.

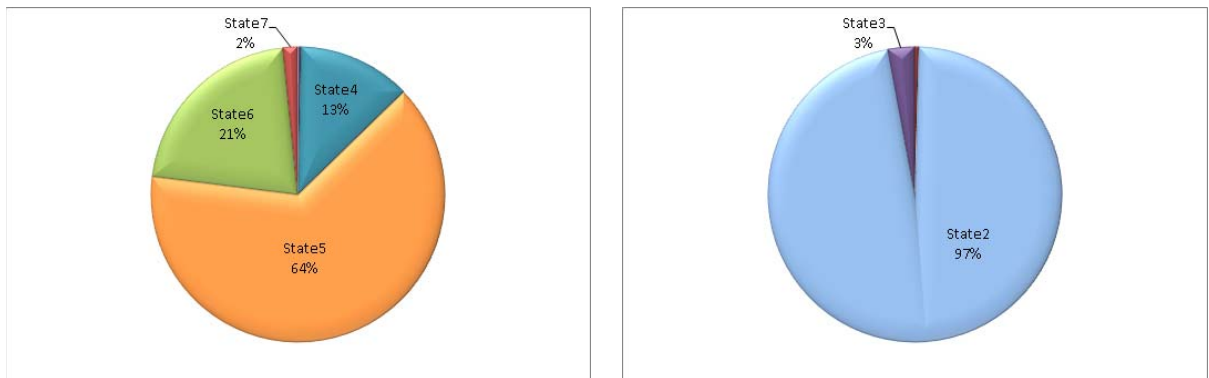


Fig. 6. Aggregated load marginal probability distribution when based COP is low at 1.5 (left-hand side plot) and when it is high at 3.5 (right-hand side plot).  $T_{set} = 70^{\circ}F$  and house type 4 for both plots.

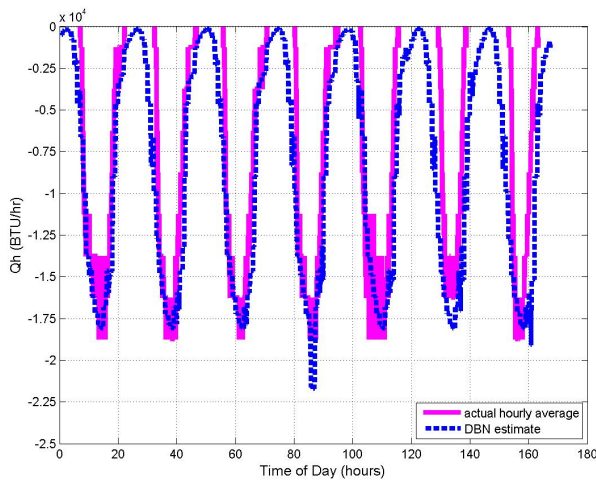


Fig. 7. Comparison of actual vs. estimate  $Q_h$ .  $T_{set} = 70^\circ F$ ,  $baseCOP = 2.5$ , house type 4.

probability distribution is compared when base COP is low i.e., 1.5, on the left-hand side plot versus when COP is high i.e., 3.5, on the right-hand side plot. In both cases,  $T_{set}$  is set to  $70^\circ F$ , the house type is 4, and ToD corresponds to 2:00pm. On the left-hand side plot the DBN assigns high marginal probabilities to discrete load states 4, 5, and 6, with state 6 being assigned the highest percentage of 64%. Consequently, given the value of the other DBN variables that were set as evidence, the load value is most probably going to be within the range of state 6. State 6 corresponds to range 3 - 3.5 MW, state 5 corresponds to range 2.5 - 3 MW, and state 4 corresponds to range 2 - 2.5 MW. A small percent is also assigned to state 7 that corresponds to range 3.5 - 4 MW. For the figures presented in this paper, when referring to the DBN estimate the weighted average of all possible load states is used. The more probable to occur states have a greater weight coefficient. The right-hand side plot of Fig. 6 shows a different load probability distribution where the DBN estimates with 97% certainty that the load value is most probably going to be within the range of state 2. State 2 corresponds to range 1 - 1.5 MW and state 1 corresponds to range 0.5 - 1 MW. This is a case where base COP of 3.5 derives a very different probability distribution compared to case of base COP 1.5. It is normal that different evidence setting produces diverse load probability distributions. The query variable probability distributions depend on the training dataset used, and different training datasets will produce different distributions. Fig. 6 demonstrates the advantage of DBN modeling in providing probabilistic analysis; something that a lot of currently used power simulators and tools lack.

Compared to the traditional way of modeling power demand using power simulator, the DBN model can produce results much faster, a few minutes in this study. This is a result of having a trained statistical model that accounts for diverse operational scenarios of a diverse house population. The power simulator model took a few hours to provide equivalent estimates of the power demand. Training of the network can be time consuming, but it only needs to be updated periodically and offline. Additionally, the DBN provides

probability distributions of the queried variables, a feature that is currently not supported by most power simulators.

## V. CONCLUSION

Demand response schemes aim in balancing power demand and supply without necessarily increasing electricity generation to meet high power demand. Instead a DR scheme could use information and data collected from the distribution system to shift load reducing load peaks. Statistical and data mining methods used for analyzing a collection of data can naturally provide great insight of data patterns and trends. This research was motivated by the fact that, to the authors' best knowledge, Dynamic Bayesian Networks have not yet been used as a tool for data analysis in the area of residential load modeling. This initial experimentation with DBNs aims in demonstrating that DBNs can be good candidates for modeling residential aggregated load. The results summarized in Fig. 3 - Fig. 5 show the accuracy with which the DBN modeled HVAC aggregated load under different scenarios of operation. Furthermore, Fig. 6 and Fig. 7 show the diverse ways in which a DBN model can be used for load analysis. Undoubtedly, there are still other aspects of DBN modeling can be investigated to further this study.

## REFERENCES

- [1] S. Lu, N. Samaan, R. Diao, M. Elizondo, C. Jin, E. Mayhorn, Y. Zhang, and H. Kirkham, "Centralized and decentralized control for demand response," in *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, 2011, pp. 1-8.
- [2] Pacific Northwest Smart Grid Demonstration Project. [Online]. Available: <http://www.pnwsmartgrid.org/transactive.asp>
- [3] P. Samadi, A. Mohsenian-Rad, R. Schober, V. W. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 415-420.
- [4] The Top Five Players in Demand Response. [Online]. Available: <http://www.greentechmedia.com/articles/read/top-5-demand-response/>
- [5] J. A. F. Moreno, A. M. García, A. G. Marín, E. G. Lázaro, and C. A. Bel, "An integrated tool for assessing the demand profile flexibility," *Power Systems, IEEE Transactions on*, vol. 19, no. 1, pp. 668-675, 2004.
- [6] M. Vlachopoulou, G. Chin, J. Fuller, S. Lu, and K. Kalsi, "Model for aggregated water heater load using dynamic bayesian networks."
- [7] J. Poropudas and K. Virtanen, "Simulation metamodeling with dynamic bayesian networks," *European Journal of Operational Research*, vol. 214, no. 3, pp. 644-655, 2011.
- [8] J.-L. Molina, D. Pulido-Velazquez, J. L. García-Aróstegui, and M. Pulido-Velázquez, "Dynamic bayesian networks as a decision support tool for assessing climate change impacts on highly stressed groundwater systems," *Journal of Hydrology*, 2012.
- [9] M. Zou and S. D. Conzen, "A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data," *Bioinformatics*, vol. 21, no. 1, pp. 71-79, 2005.
- [10] U. Lerner, R. Parr, D. Koller, and G. Biswas, "Bayesian fault detection and diagnosis in dynamic systems," in *AAAI/IAAI*, 2000, pp. 531-537.
- [11] K. P. Schneider, J. C. Fuller, and D. P. Chassin, "Multi-state load models for distribution system analysis," *Power Systems, IEEE Transactions on*, vol. 26, no. 4, pp. 2425-2433, 2011.
- [12] J. C. Fuller, N. P. Kumar, and C. A. Bonebrake, "Evaluation of Representative Smart Grid Investment Grant Project Technologies: Demand Response," Pacific Northwest National Laboratory, Tech. Rep., 01 2010.
- [13] GridLAB-D, ver. 2.2. [Online]. Available: <http://www.gridlabd.org>