Model for Aggregated Water Heater Load Using Dynamic Bayesian Networks

M. Vlachopoulou[1], G. Chin[1], J. C. Fuller[1], S. Lu[1], and K. Kalsi[1]
[1] Pacific Northwest National Laboratory, Richland, WA 99354 USA

Abstract - *The transition to the new generation power grid, or "smart grid", requires novel ways of using and analyzing data collected from the grid infrastructure. Fundamental functionalities like demand response (DR), that the smart grid needs, rely heavily on the ability of the energy providers and distributors to forecast the load behavior of appliances under different DR strategies. This paper presents a new model of aggregated water heater load, based on dynamic Bayesian networks (DBNs). The model has been validated against simulated data from an open source distribution simulation software (GridLAB-D). The results presented in this paper demonstrate that the DBN model accurately tracks the load profile curves of aggregated water heaters under different testing scenarios.*

Keywords: Dynamic Bayesian network, water heater, demand response, smart grid

1 Introduction

New advances in power and energy technologies have recently accentuated the need for revision of the current power grid operation to ensure reliability and performance. The next generation power grid, known as the "smart grid", provides a new framework that includes the new technology deployments and addresses the issues of system state uncertainty and deregulation [1]. Demand response (DR), distributed generation (DG) and distributed energy storage (DES) are basic strategies applied during the smart grid operation. They formulate a new power grid paradigm that incorporates distributed architecture instead of the traditional centralized one, as well as dynamic response to real time changes of the power grid state. The organizations and corporations involved in the power generation, transmission and distribution will be required to have the necessary analysis and planning tools for a successful transition to smart grid operation.

The demand response feature is enabled by allowing devices and appliances to modulate their operation in response to an event causing a change of state of the voltage and frequency of the grid, energy prices, or a number of other factors. The Federal Energy Regulatory Commission (FERC) in [2] specifies different types of DR programs including dynamic pricing without enabling technology, dynamic pricing with enabling technology, direct load control and interruptive tariffs. A simple overview of DR strategies can be found in [3]. Multiple utility companies have expressed interest in assessing the impact that DR can have to their operations. Results of DR studies using empirical data are presented in [4] and [5]. The dynamic behaviour of the smart grid stems from the new perception of the grid as a network with real-time communication of its components. The DR strategies support the required network response flexibility, however it is essential that their application preserves and enhances the grid reliability. Reliability studies under DR operation [6] and [7] have used the DC Optimal Power Flow (OPF) model to access the impact of DR programs. A general impact of DR, DES and the penetration of renewable energy resources to the smart grid reliability is analysed in [8].

There is a prominent need for analysis tools that can be used by utilities and other power grid management participants, to forecast the DR effect on the power grid operation. Load forecasting is an important component of this analysis, where load is the power sink of an appliance or device. There are different types of load forecasting, depending on the time horizon of the forecast. There have been certain critical factors determined that affect such forecasts [9]. The forecasting methodologies range from statistical methods, like regression and time series analysis, to artificial intelligence and data mining methods, like neural networks, fuzzy logic and support vector machines. Efforts have been recently made for analytical modelling of aggregated loads [10]-[12].

This paper presents a novel approach of forecasting aggregated end-use water heater load in residential areas. This approach entails a Dynamic Bayesian Network (DBN) for modelling of the aggregated load behaviour. The developed model successfully and accurately emulates the behaviour of the aggregated water heater load due to two factors. First, the DBN structure enables modelling of the dynamic physical system behaviour. Second, end-use load information data have been used for training of the network. Currently the DBN has been trained and tested using simulated data produced by simulation software (GridLAB-D). This DBN model can provide the basis for an accurate and flexible tool that is deployed for the analysis of DR strategies. It is easier and faster to use compared to

GridLAB-D and also provides a larger degree of flexibility since it can be retrained using different data sets.

This paper is organized as follows. In Section 2, an overview of the DBN principles of operation and application examples are given. Section 2 also includes a brief description of the software which was used to generate the training and testing data for the DBN analysed in this paper. In Section 3, a detailed description of the DBN model, as applied to the problem of water heater load aggregation, is given. The results illustrating the performance of the DBN model can be found in Section 4. Finally, Section 5 includes the conclusions of this research and future work.

2 Dynamic Bayesian network principles

Bayesian networks are a widely used machine learning methodology with diverse areas of application like medical diagnosis, sensor modeling and reliability analysis [13]. Studies of dynamic Bayesian networks are relatively more recent and aim in modeling a constantly changing system. In this section, a description of the basic structure and principles of operation of DBNs is presented. Additionally, the simulation software that provided the training and testing data is described.

2.1 Dynamic Bayesian networks

A Bayesian network (BN) or belief network is a probabilistic graphical model. In a BN, nodes represent random variables and directed arcs 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. These probabilities are commonly derived from collected data or prior knowledge. Once a BN has been constructed, the values of certain variables can be set based on evidence or observations. The posterior probabilities of the query variables can then be computed given the set of evidence variables as knowledge. Inferencing refers to the propagation of the evidence through the network followed by computation of the updated probabilities of the query variables.

For temporal analysis, a dynamic Bayesian network may be used to model 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. Since current events cause future events, but not vice-versa, directed arcs always flow forward in one direction in a DBN. For many applications, the graphical representation of a DBN often takes the form of a first-order Markov or hidden Markov model. DBNs have been used in a variety of applications in areas such as speech recognition [14], distributed sensor networks [15], and computational biology [16].

In developing a BN or DBN model, domain expertise is invaluable in a number of modeling steps. First and foremost, the structure of the BN in terms of the variables and conditional dependencies rely heavily on expert input. The structure of a BN should resemble the logical or physical topology of the system or process that it is modeling. BN structure learning algorithms including score-and-search-based and constraint-based methods are also available to automatically generate BN structures from training data, but it has been found that such algorithms are most effective in verifying a manually-constructed BN rather than constructing a BN from scratch. Expert input is also important in defining variable states, as they should represent the specific conditions of logical or physical entities in the BN. With respect to the conditional probability tables, we have found that BN parameter learning algorithms such as maximum likelihood estimation and expectation-maximization are mostly effective in learning probabilities. After parameter learning, however, we typically have experts verify that the learned probabilities appear reasonable.

2.2 GridLAB-D simulation

The DBN training and testing data were produced by the simulation software, GridLAB-D. GridLAB-D is an agent-based, open-source, power grid simulation tool developed at Pacific Northwest National Laboratory (PNNL) for the Department of Energy (DOE) to simulate the complexities of the smart grid from the substation to the end-use load [17].This allows users to develop models to simulate the behaviors of individual end-use loads and their interactions with the power system, including voltage effects, weather dependencies, control functions, consumer demand and a number of other inputs which affect the behavior of the end-use loads.

To simulate the behavior of a water heater, a multi-state load model is available [18], [19]. This model uses multiple states and state transition rules to describe the power demand at any given time in the simulation [20]. The physical processes within the water heater, such as thermostat set point, water temperature, consumer hot water usage, and thermodynamic heat flow equations are described by state models. These are combined to create a simulation which can simulate the power demand of thousands of individual water heater "agents", each with individualized characteristics and parameters. While this is highly advantageous for studying the effects of a thermostat setback or direct load control program on consumers, as the drop in water temperature can be tracked on an individual level, the simulations can be time- and labor-intensive.

3 Aggregated model

This section discusses how dynamic Bayesian networks, discussed in Section 2.1, can be used to model the aggregated behavior of water heaters with regards to power consumption. First the dynamic Bayesian network model is presented, followed by a demonstration of its use for aggregated water heater end-use load forecasting.

3.1 Dynamic Bayesian network model

The structure of the dynamic Bayesian network model developed for this research is based on expert opinion. The expert opinion was used to define the relationships between a set of variables that influence the load energy consumption due to water heater operation in residential areas. These variables represent time, weather and appliance specific factors. It has been indicated by the GridLAB-D simulation that these factors are the most influential for the water heater load consumption and similar facts regarding most influential factors were pointed out in [9]. Specifically, the variables used to build the DBN are time of day (ToD), season, outside air temperature, solar radiation, water heater efficiency, water heater temperature set point, hot water usage and load consumption. The dynamic behavior of the DBN is established by using two time slices in the network structure as shown in Fig. 1. The data are extracted at 5-minute intervals from GridLAB-D, therefore the two-time slice BDN has the ability to capture the dynamic behavior of the simulation at a minimum temporal resolution of 5 minutes.

The variable relationships of the DBN model described above can be easily explained and make intuitive sense. First, regarding the time factors, it has been observed that during specific times of the day the water heater load power demand is greater. For example, the average person tends to use the shower in the morning hours, resulting in higher water usage and therefore higher water heater energy consumption at that time compared to the consumption at noon. The time of day also naturally relates to the variations observed in the outside air temperature and solar radiation. Seasonality also has an effect on the power demand since during the winter months, for example, if a water heater is located in unconditioned garage, then the lower air temperature leads to greater thermal energy loss across the insulation jacket, resulting in greater energy consumption. Seasonality also naturally relates to the variations observed in the outside air temperature and solar radiation. Finally, the appliance related variables, like water heater efficiency and thermostat set point have an intuitive relation to the load power demand. Lower efficiency water heaters will lose more heat into the ambient air, and over time consume more energy to maintain the temperature of the water as compared to a more efficient water heater. Also, the higher the thermostat set point, the more energy is consumed to maintain the temperature of the water due to the greater temperature gradient across the insulation jacket.

The network variables that relate to the first time slice are denoted with the numeric 1 of the node name, while the ones that relate to the second time slice with the numeric 2 of the node name, as shown in Fig. 1. The water heater efficiency and thermostat set point variables only appear in the first time slice, since they remain constant over time, during steady state operation. The time of day variable, belonging to the first time slice, has been observed to have an impact on the variables of the second time slice. This relationship is modeled by adding the appropriate arcs on the network as shown in Fig. 1.

3.2 Model usage example

Determining and validating a network structure that models aggregated water heater load behavior with adequate accuracy is not a trivial task. Multiple training and testing scenarios have been considered for the evaluation of the DBN model. However, once the model has been established, it provides the user with a very flexible tool for analysis and planning.

An example of its use is demonstrated in Fig. 1. The nodes of the trained network that are circled with a solid contour are the nodes to which evidences are set. The node circled with a dotted contour is the query node. In this example, the DBN querying process is used to determine the load demand, given the time of day, season, outside temperature, solar radiation, thermostat set point, efficiency and hot water usage. The querying node for the load is at the second time slice, while the evidence nodes are at the first time slice. This captures the notion that the distribution of a variable at a present time can be queried based on the values or distributions of variables at a time in the past.

Another example of using this DBN is presented in Fig. 2, in which case information related to the hot water usage needs to be derived. The nodes where evidence is set (circled with a solid contour) are time of day, season, outside temperature, solar radiation, thermostat set point, efficiency and load of the first time slice. The querying node is the hot water usage in the second time slice. In this example, similar to the aforementioned example, the evidences can be set in the form of a distribution if there is not enough information about the actual value. Also, the querying result provides the user with a distribution which is useful in accounting for forecasting errors. These two examples demonstrate the flexibility of usage that the trained DBN provides.

4 Results

In this section a description of the training and testing data is presented. The testing results of the trained DBN are compared to the simulated data under different scenarios of the water heater operation. Two different forecasting methods, soft and hard forecasting are considered based on the resulting probability distribution of the query node.

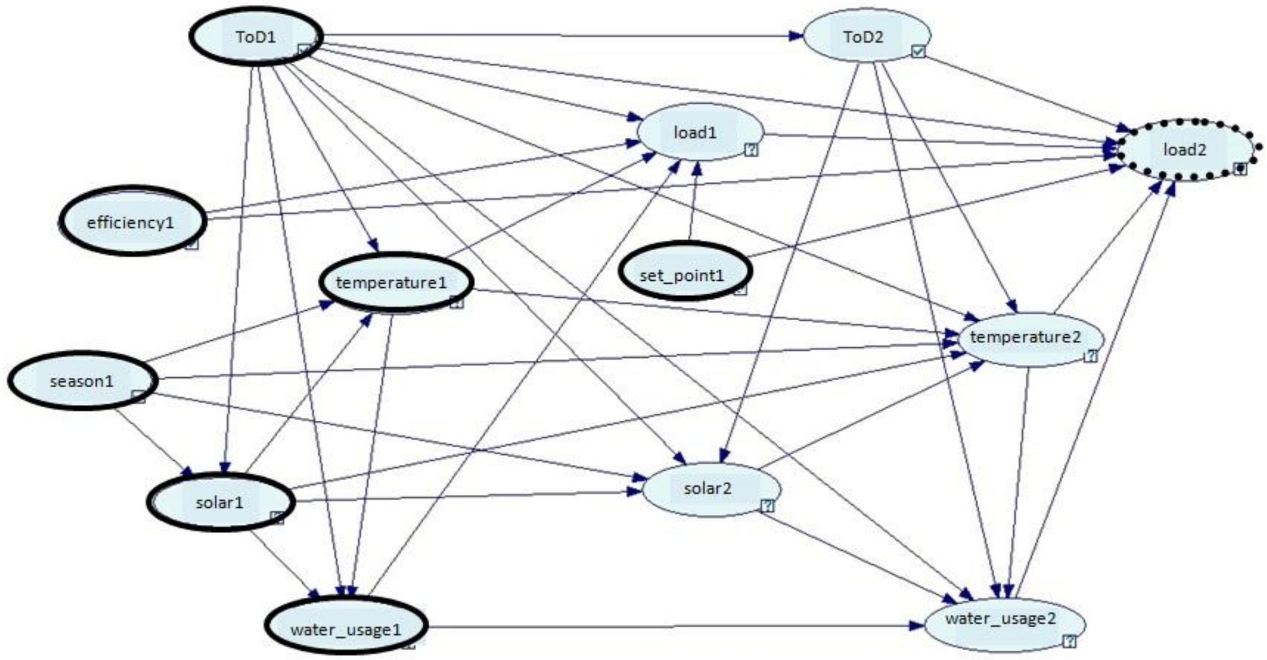

Fig. 1. Two-time slice Dynamic Bayesian Network model of aggregated water heater load. Example 1, querying load node circled with a dotted contour, using evidence information of nodes circled with solid contours.

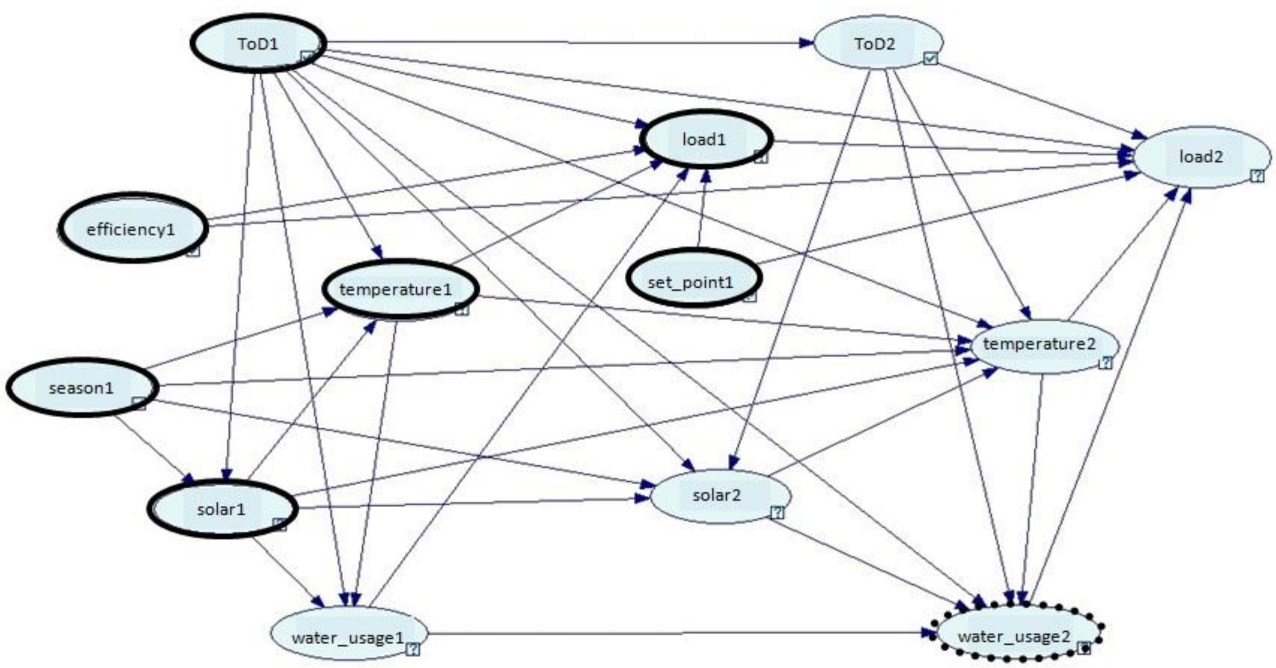

Fig. 2. Two-time slice Dynamic Bayesian Network model of aggregated water heater load. Example 2, querying water usage node circled with a dotted contour, using evidence information of nodes circled with solid contours.

4.1 Simulation data: DBN training and testing

The GridLAB-D simulation environment was used to produce the training and testing data. The simulation was of a residential neighborhood of 1000 houses. Each house had a Heating, Ventilation and Air Conditioning (HVAC) system simulated such that the inside house temperature was maintained at a reasonable level. Other characteristics, such as end-use load usage, cooling and heating set points, and thermal insulation were randomly varied across the population of homes to create a distribution of home parameters and characteristics representative of "real" building stock.

The training data were produced by running the simulation for the winter season, from December to March approximately, excluding a week in February whose data would be used for the testing dataset. This training range of data has been empirically proven to provide adequate training for the DBN. The DBN was trained using different training scenarios with the thermostat set point, efficiency and hot water usage varying between the scenarios. The water heater set point range considered was 110 to 135 ^{o}F. The water heater efficiency was set to low, medium or high, used to represent the relative amount of insulation around the thermal jacket of the water heater. Schedules (ToD) for the hot water usage were created from End-Use Load and Consumer Assessment Program (ELCAP) residential load data, while incorporating Energy Information Administration (EIA) website data on average hot water consumption in the U.S. [21],[22]. The hot water usage was set to either low or high, affecting the relative magnitude of the water flows, and was used to represent residences with low-flow rate fixtures versus older, high-flow fixtures. The simulation used a typical meteorological year (TMY) weather file that provides the outside air temperature and solar radiation information [23]. The load demand ranged from 0 to 1400 kW, approximately, between different scenarios and Time of Day.

It is a well-known fact that the discretization of the variables has a big impact on the accuracy of the querying results of the DBN [24]. The discretization method was decided based on expert opinion and experimentation with the DBN. The expert's opinion helped identify the variables with the highest sensitivity and those variables were discretized at a finer resolution. For example, the time of day is a variable with high sensitivity so it was discretized at an hourly basis. The load power demand is also an influential variable so it was uniformly discretized every 100kW. The hot water usage variable discretization is coarse, since it was only set to high or low, even if it is a highly sensitive variable. The reason for this discrepancy is that real world data do not usually contain hot water usage information with high accuracy. Future implementation of this work will involve training and testing with real world data. It is therefore appropriate to keep in consideration the realistic availability of data for a smooth transition to the real world application.

4.2 Results

The DBN was tested using GridLAB-D simulated data for a week in February that were not included in the training set. The testing of the DBN accuracy was performed by querying the load variable of the second time slice, similarly to the first example presented in Section 3.2. This example is a good indication of how a utility company would use this tool to do end-use load forecasting of an aggregated water heater load. The query results over the time period of a day, in comparison to the actual simulated data, are plotted in Fig. 3-6. The results presented in Fig. 3 correspond to a high hot water usage case and low efficiency, while the results presented in Fig. 4 correspond to low hot water usage and high efficiency. In both cases the water heater set point is set to 115 ^{o}F. Equivalently, the results in Fig. 5 and Fig. 6 present the same comparison of high water usage/low efficiency versus low water usage/high efficiency, but with the water heater set point at 130 ^{o}F. The GridLAB-D hourly average load demand data are compared to a hard and soft load demand forecasting. The hard forecasted data are obtained by selecting the load variable value that was assigned the highest probability of occurrence by the querying process. The soft forecasted data are obtained by evaluating an average of the possible load demand values weighted by their assigned probabilities.

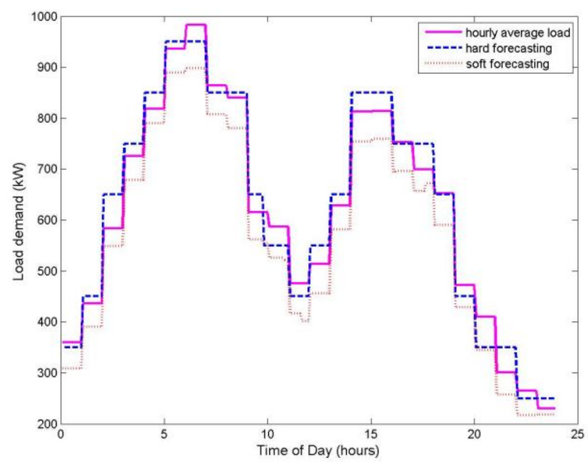

Fig. 3. Simulated vs. forecasted hourly averaged daily load profile curves. Simulation parameters: winter season, 115 ^{o}F set point, low efficiency, high water usage.

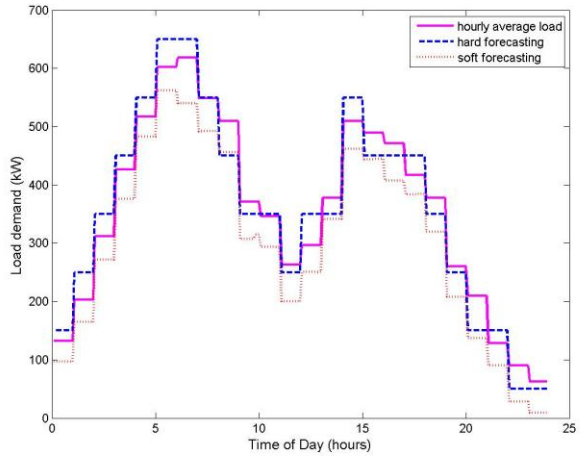

Fig. 4. Simulated vs. forecasted hourly averaged daily load profile curves. Simulation parameters: winter season, 115 ^{o}F set point, high efficiency, low water usage.

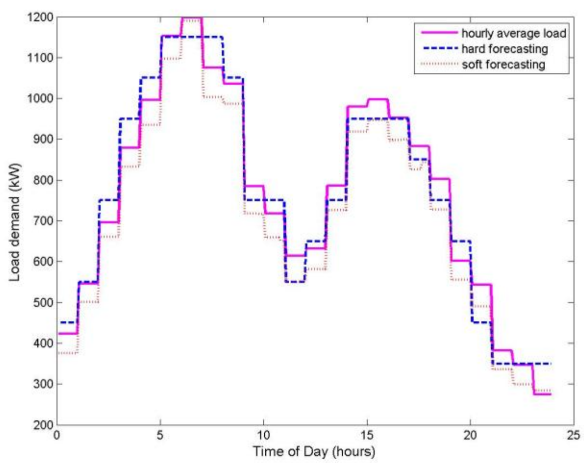

Fig. 5. Simulated vs. forecasted hourly averaged daily load profile curves. Simulation parameters: winter season, 130 ^{o}F set point, low efficiency, high water usage.

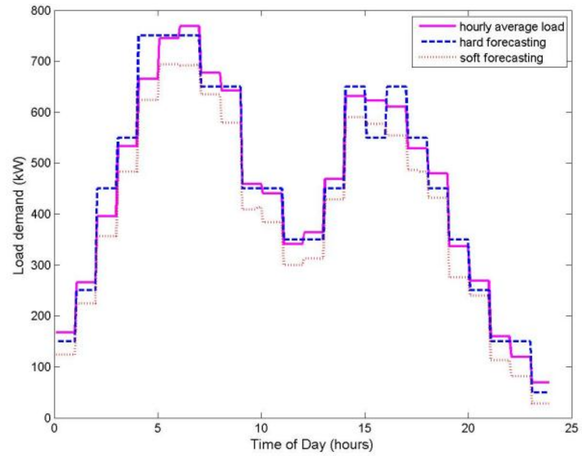

Fig. 6. Simulated vs. forecasted hourly averaged daily load profile curves. Simulation parameters: winter season, 130 ^{o}F set point, high efficiency, low water usage.

In Fig. 3 the load demand is much higher at any time in the day compared to Fig. 4, since the simulated testing data are that of a neighborhood having low efficiency water heaters and high hot water usage. Equivalently, the same statement can be made when comparing the load demand profile presented in Fig. 5 versus Fig. 6. As demonstrated by the results, the load demand additionally depends on the water heater set point. A higher set point results in higher load energy consumption as shown by comparing Fig. 3 and Fig. 5.

Both hard and soft forecasting accurately track the GridLAB-D load profile curve. It can be observed that the soft forecasting tracks the load curve variations slightly better than the hard forecasting. The average forecasting error is approximately in the order of 50kW. These results demonstrate that the DBN has been trained adequately for load forecasting of different simulated test scenarios.

5 Conclusions and future work

It has been shown that DBNs can be successfully used to model the aggregated water heater load demand, tracking the simulated data load profile curve closely. Therefore, this research provides a first indication that DBNs constitute a powerful modeling tool as applied in the area of power engineering and the smart grid. It provides the flexibility needed for energy consumption analysis and could potentially be used for the assessment of the impact of DR programs on the grid operation. It can be used to ingest a high volume of data for training under different scenarios of operation without having to modify its structure. The DBN modeling approach’s main advantage over other machine learning and data mining methodologies is that it models the physical relationship between the actual system variables. The authors are now working on applying the model presented in this paper to real world data. It is expected the new results will demonstrate the value of applying DBNs for load forecasting even further.

6 References

[1] H. Farhangi, “The path of the smart grid,” IEEE PES Magazine, vol. 8, no. 1, pp. 18-28, Jan-Feb 2010.

[2] “A National Assessment of Demand Response Potential”, Staff Report, Federal Energy Regulatory Commission (FERC), June 2009. [Online]. Available: http://www.ferc.gov/legal/staff-reports/06-09-demand-response.pdf

[3] M. H. Albadi and E. F. El-Saadany, “Demand Response in Electricity Markets: An Overview,” IEEE PES General Meeting, 2007.

[4] P. Cappers, C. Goldman, and D. Kathan, “Demand Response in U.S. Electricity Markets: Empirical Evidence,” Energy, vol. 35, pp. 1526-1535, 2010.

[5] A. Faruqui and S. Sergici. (2010, February). Household Response to Dynamic Pricing of Electricity-A Survey of the Empirical Evidence. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1134132

[6] L. Goel, Q. Wu, and P. Wang, “Reliability Enhancement of A Deregulated Power System Considering Demand Response,” IEEE PES General Meeting, 2006.

[7] R. Azami and A. F. Fard, “Impact of Demand Response Programs on System and Nodal Reliability of a Deregulated Power,” IEEE Int. Conf. of Sustainable Energy Technologies, 2008.

[8] K. Moslehi and R. Kumar, “A reliability Perspective of the Smart Grid,” IEEE Trans. Smart Grid, vol. 1, pp. 57-64, June 2010.

[9] R. Weron, Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. John Wiley & Sons, 2007.

[10] W. Zhang, K. Kalsi, J. Fuller, M. Elizondo, and D. Chassin, “Aggregate Model for Heterogeneous Thermostatically Controlled Loads with Demand Response,” to appear in proceedings of IEEE PES General Meeting, San Diego, CA, July 2012.

[11] K. Kalsi, M. Elizondo, J. Fuller, S. Lu, and D. Chassin, “Development and Validation of Aggregated Models for Thermostatic Controlled Loads with Demand Response”, Hawaii International Conference on System Sciences, Maui, Hawaii, January 2012.

[12] K. Kalsi, F. Chassin, and D. Chassin, “Aggregated Modeling of Thermostatic Loads in Demand Response: A Systems and Control Perspective”, IEEE Conference on Decision and Control and European Control Conference, Orlando, Florida, December, 2011.

[13] O. Pourret, P. Naim, and B. Marcot, Bayesian Networks, A Practical Guide to Applications, Wiley, 2008.

[14] G. Zwieg and S. Russell, “Speech Recognition with Dynamic Bayesian Networks”, Proc. of AAAI-98, Madison, WI, July 1998.

[15] G. Chin Jr., S. Choudhury, L. Kangas, S. McFarlane, and A. Marquez, “Fault Detection in Distributed Climate Sensor Networks using Dynamic Bayesian Networks”, Proc. of 6th IEEE International Conference on e-Science, Brisbane, Australia, December 2010.

[16] D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.

[17] “GridLAB-D, ver. 2.2”. October 2011. [Online]. Available: http://www.gridlabd.org

[18] K. P. Schneider, J. C. Fuller, and D. P. Chassin, “Multi-State Load Models for Distribution System Analysis,” IEEE Trans. Power Systems, vol. 26, no. 4, pp. 2425-2433, Nov. 2011.

[19] Z. T. Taylor, K. Gowri, and S. Katipamula, “GridLAB-D Technical Support Document: Residential End-Use Module Version 1.0,” PNNL-17694, Pacific Northwest National Laboratory, Richland, WA, 2008.

[20] J. C. Laurent and R. P. Malhame, “A physically-based computer model of aggregate electric water heating loads,” IEEE Trans. Power Systems, vol. 9, no. 3, pp. 1209-1217, Aug. 1994.

[21] R. G. Pratt, C. C. Conner, E. E. Richman, K. G. Ritland, W. F. Sandusky, and M. E. Taylor, “Description of Electric Energy Use in Single Family Residences in the Pacific Northwest,” DOE/BP 13795 21, Bonneville Power Administration, Portland, OR, 1989.

[22] “U.S. Energy Information Administration”. September 2011. [Online]. Available: http://www.eia.gov.

[23] “National Solar Radiation Data Base”. September 2011. [Online]. Available: http://rredc.nrel.gov/solar/old_data/nsrdb/1961-1990/tmy2/

[24] N. Friedman and M. Goldszmidt, “Discretizing Continuous Attributes While Learning Bayesian Networks”, Proc. 13th International Conference on Machine Learning, vol. 159, no. 12, pp. 157-165.