This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2702751, IEEE Transactions on Smart Grid

Variable Selection Methods for Probabilistic Load Forecasting: Empirical Evidence from Seven States of the United States

Jingrui Xie and Tao Hong

***Abstract*-- Variable selection is the process of selecting a subset of relevant variables for use in model construction. It is a critical step in forecasting but has not yet played a major role in the load forecasting literature. In probabilistic load forecasting, many methodologies to date rely on the variable selection mechanisms inherited from the point load forecasting literature. Consequently, the variables of an underlying model for probabilistic load forecasting are selected by minimizing a point error measure. On the other hand, a holistic and seemingly more accurate method would be to select variables using probabilistic error measures. Nevertheless, this holistic approach by nature requires more computational efforts than its counterpart. As the computing technologies are being greatly enhanced over time, a fundamental research question arises: *can we significantly improve the forecast skill by taking the holistic yet computationally intensive variable selection method?* This paper tackles the variable selection problem in probabilistic load forecasting by proposing a holistic method (HoM) and comparing it with a heuristic method (HeM). HoM uses a probabilistic error measure to select the variables to construct the underlying model for probabilistic forecasting, which is consistent with the error measure used for the final probabilistic forecast evaluation. HeM takes a shortcut by relying on a point error measure for variable selection. The evidence from the empirical study covering seven states of the United States suggests that 1) the two methods indeed return different variable sets for the underlying models, and 2) HoM slightly outperforms but does not dominate HeM w.r.t. the skill of probabilistic load forecasts. Nevertheless, the conclusion might vary on other datasets. Other empirical studies of the same nature would be encouraged as part of the future work.**

***Index Terms*-- Load forecasting; mean absolute error; mean absolute percentage error; pinball loss; probabilistic forecasting; quantile score; variable selection.**

I. Introduction

VARIABLE selection, a.k.a. feature selection or attribute selection, is the process of selecting a subset of relevant variables for use in model construction. Few papers in the load forecasting literature have focused on the variable selection problem, i.e., revealing and discussing the comparison among various models constructed from subsets of relevant variables. Most, if not all, of these variable-selection-related papers were on point load forecasting. Hong proposed a variable selection algorithm for linear regression models, when the same algorithm was also shown to be applicable to the selection of other models, such as fuzzy regression models and artificial neural networks [1]. Part of Hong's algorithm was later extended by releasing the computational constraint to allow evaluation of a large amount of lagged and moving average temperature variables [2]. Xie et al. applied stepwise regression to select variables during the nPower electric load forecasting challenge 2015 [3]. Ziel and Liu used the least absolute shrinkage and selection operator (LASSO) estimator for automatic variable selection [4]. Xie et al. discussed how to select relative humidity variables for load forecasting [5]. Several papers summarizing the methodologies implemented for Global Energy Forecasting Competition 2012 (GEFCom2012) also briefly introduced the variable selection method. For example, Ben Taieb and Hyndman took a component-wise gradient boosting method to select temperature, lagged demand, and calendar variables to use in the non-parametric additive load forecasting models [6]. Lloyd combined the forecasts from the gradient boosting machine, the Gaussian process, and the benchmark model from the competition. A standard R implementation of the gradient boosting machine with a customized shrinkage factor and tree size was used by Lloyd to learn and select from a number of unknown regression functions [7]. Nedellec et al. used the random forest method to average many simple tree models with each of them including a subset of the relevant variables [8].

Probabilistic load forecasting produces forecasts in the form of quantiles, intervals or density functions, which provide more comprehensive information about the future than what point forecasts can do. Probabilistic load forecasts can be generated from three different methods or a combination of any two or three of them [9]: 1) simulating input scenarios [10]–[15], 2) applying probabilistic models such as quantile regression [16]–[18], and 3) converting point forecasts into probabilistic forecasts via residual simulation [13], [19] or forecast combination [18]. Both the first and third methods require the development of underlying point forecasting models. These models are typically constructed by evaluating their point forecast performance using point error measures such as Mean Absolute Percentage Error (MAPE) [11]–[13],

Jingrui Xie is with the Forecasting R&D of SAS Institute Inc at Cary, NC, USA (e-mail:jingrui.rain.xie@gmail.com). Tao Hong is with University of North Carolina at Charlotte, Charlotte, NC, USA, and Dongbei University of Finance and Economics, Dalian, China (e-mail: hongtao01@gmail.com).

[19], Root Mean Square Error [19], Akaike's Information Criterion [10], or Generalized Cross-validation Scores [14]. Intuitively, selecting variables for probabilistic load forecasting based on a point error measure would be more computationally efficient but less accurate than its counterpart, using a probabilistic error measure for variable selection. This raises a fundamental research question for probabilistic load forecasting: *can we significantly improve the forecast skill by taking the holistic yet computationally intensive variable selection method?*

To answer the aforementioned question, this paper presents the first and formal investigation on variable selection for probabilistic load forecasting. We propose a holistic method (HoM) for variable selection and an analysis framework to compare it with a heuristic method (HeM) in the context of probabilistic load forecasting. The empirical study involves a large number of realistic models with similar point forecast accuracy. The relevant variables of interest are lagged and moving average temperature variables, as discussed in the previous studies on recency effect [2], [11]. For the reproducibility purpose, two public available datasets are being used in this paper: 1) the data from the probabilistic load forecasting track of GEFCom2014 [20] and 2) the zonal level data of ISO New England (ISO-NE) [21]. These two datasets covers seven states of the United States.

The rest of this paper is organized as follows: Section II provides the background information, including a description of the case study data, an introduction to the relevant variables and comparative models, and the method to generate probabilistic load forecasts. Section III illustrates the variable selection process using HeM and HoM, respectively. Section IV presents the variable selection and forecasting results from both datasets. Section V comments on the computational intensity of the two variable selection methods and some practical considerations. The paper is then concluded in Section VI with a brief discussion of future research.

II. Background

A. Data

In this paper, the GEFCom2014 data will be used as the primary case study to demonstrate the variable selection process. The ISO-NE data will be used as a secondary case study to further confirm that the findings are not specific to only one dataset. For reproducibility purpose, we did not perform any specific data cleansing to the original datasets. For the ISO-NE data, we took the average of the two adjacent readings to replace the zero entry at the beginning of the daylight saving time and divided the entry by two at the end of the daylight saving time in each year.

GEFCom2014 established a benchmarking data pool for research and development in energy forecasting [20]. The probabilistic load forecasting track includes 5 years of hourly load data (2007-2011) and 10 years of hourly temperature data (2001-2010). A summary statistics of the load and temperature data are provided in Table I. We will follow the settings of GEFCom2014 to provide the one-month ahead forecast for each of the 12 months of 2011 on a rolling basis with the forecast origin rolling forward one month each time [20]. The variables are re-selected and updated for each month based on the performance of the model in the same month of year 2010 (i.e. validation year). For example, we first forecast the load of January 2010 using the data from January 2007 to December 2009 and select a subset of relevant variables based on the model performance in January 2010. The selected subset of variables will be used to construct the model for January 2011. We then roll the forecast origin forward to forecast February 2010 using the data from February 2007 to January 2010. This process is repeated until the variables are selected to construct the forecasting model for each of the 12 months of 2011.

ISO-NE is an independent regional transmission organization, serving the six states in the northeast of U.S. including Connecticut (CT), most of Maine (ME), Massachusetts, New Hampshire (NH), Rhode Island (RI), and Vermont (VT). Massachusetts is further dissected into three load zones, namely NEMASSBOST, SEMASS and WC MASS. Each of the other five states forms its own load zone. We use the zonal level hourly load and temperature data that ISO-NE publishes on its website for the secondary case study [21]. For each zone, the history data can be traced back to the year 2003. Similar to the design of the GEFCom2014 case study, for each zone under ISO-NE, we select the model for each of the 12 months of the test year (i.e. year 2015) based on their performance during the same month in the validation year (i.e. year 2014). The variable selection process is also conducted on a rolling basis with three years of history prior to the validation month as the training data. Table I presents the summary statistics of the load (2011-2015) and temperature data (2004-2014) used for the eight load zones of ISO-NE.

Table I
Summary Statistics of Case Study Data

	Load (MW)			**Temperature (°F)**		
	Min	**Mean**	**Max**	**Min**	**Mean**	**Max**
GEFCom2014	64.4	150	318	13	61	98
CT	1372	3527	7219	-8	52	102
ME	795	1312	2135	-14	48	100
NEMASSBOST	1821	2925	5658	-7	52	101
NH	521	1329	2433	-22	47	100
RI	365	936	1967	-6	52	101
SEMASS	875	1716	3645	-6	52	101
VT	407	656	985	-20	47	97
WCMASS	739	1989	3650	-12	49	96

B. Forecasting Models

Multiple linear regression (MLR) models have been widely used for load forecasting [1]–[3], [10]–[15], [22]. This paper evaluates a series of MLR models for probabilistic load forecasting using HeM and HoM. Nevertheless, the framework presented in this paper does not constrain itself to regression only. Similar as the evidence shown in [1], [5], [13], and [15], models using other forecasting techniques, such as neural networks, can be evaluated following the same framework.

In psychology, recency effect is used to describe the fact that people tend to remember the most recent item. It was

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2702751, IEEE Transactions on Smart Grid

adopted by Hong to describe how recent temperatures affect the electricity demand [1], [2]. The lagged temperatures up to the most recent three hours and the simple and weighted moving average temperatures of the preceding 24 hours were evaluated in [1]. Wang et al. then released the computational constraints to evaluate a larger number of lagged and 24-hour moving average temperatures [2]. In this paper, we apply two variable selection methods, HeM and HoM, to the same set of lagged and moving average temperatures as in [2]. Following [2], we let

$$f(T_t) = \alpha_1 T_t + \alpha_2 T_t^2 + \alpha_3 T_t^3 + \alpha_4 T_t \times M_t + \alpha_5 T^2 M_t + \alpha_6 T_t^3 M_t + \alpha_7 T_t H_t + \alpha_8 T_t^2 H_t + \alpha_9 T_t^3 H_t \tag{1}$$

where M_t, and H_t are class variables for month and hour; T_t is the temperature. Then the group of regression models can be defined as

$$g(h,d) = \beta_0 + \beta_1 Trend_t + \beta_2 H_t \times W_t + \sum_h f(T_{t-h}) + \sum_d f(\tilde{T}_{t,d}) \tag{2}$$

where $Trend_t$ is a chronological trend; W_t is the class variable for day of week; T_{t-h} is the lagged temperature of the previous h^{th} hour, $\tilde{T}_{t,d} = \frac{1}{24}\sum_{h=24d-23}^{24d} T_{t-h}$ represents the 24-hour moving average temperate of the previous d^{th} day. We let h ranges from 0 to 48, and d ranges from 0 to 7. By varying the values of h and d, we can create 392 different models to be evaluated for each forecasted month.

The reason for using recency effect to construct candidate models for evaluation in this study is three-fold: 1) we need a large number of candidate models for the comparison purpose, while modeling recency effect helps generate hundreds of models; 2) the difference among the candidate models w.r.t. point forecast accuracy could be rather small, so that we can study the skill of probabilistic forecasts from similarly accurate underlying models; and 3) these models are currently being used in the real world by power companies, so our empirical study is meaningful to the state of the practice.

C. Probabilistic Load Forecasting

In this study, the probabilistic load forecasts are generated by feeding multiple temperature scenarios into the selected deterministic point forecast model. The temperature scenarios are generated using a fixed-date method [15], which was introduced in [11] and then implemented in [12], [13], [17]. Specifically, using k years of historical temperature, we can create k temperature scenarios for the forecasted year by matching the temperature profile date by date to the forecasted year. Each temperature scenario can then be used to derive a scenario-based point load forecast. Assigning equal probability to these temperature scenarios and the resulting scenario-based point load forecasts, we can obtain the probabilistic load forecast. The value of k is chosen for each case study to exhaust all available historical temperature series. In the GEFCom2014 case study, we let k = 9 for forecasting the validation year (i.e. 2010) and k = 10 for forecasting the test year (i.e. 2011). In the ISO-NE case study, we let k = 10 for forecasting the validation year (i.e. 2014) and k = 11 for forecasting the test year (i.e. 2015).

III. Variable Selection

Two variable selection methods, e.g., HeM and HoM, as illustrated in Fig. 1 will be implemented to select the optimal *h-d* pair of lagged temperature and moving average temperature combination.

A. Heuristic Method (HeM)

HeM selects the relevant variables based on the point forecast error of the validation period. Two point forecast error measures are tested in this paper: Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE).

MAPE is a widely used error measure in point load forecasting, which can be specified as:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \tag{3}$$

where y_t and $\hat{y}_t$ are the actual and predicted load at time t, respectively, and N is the number of observations.

MAE is also frequently used for evaluating point load forecasts, which can be specified as:

$$MAE = \frac{1}{N}\sum_{t=1}^{N} | y_t - \hat{y}_t | \tag{4}$$

Because MAE is scale-dependent, it is aligned with the scale-dependent skill score that we will use later for probabilistic forecast evaluation.

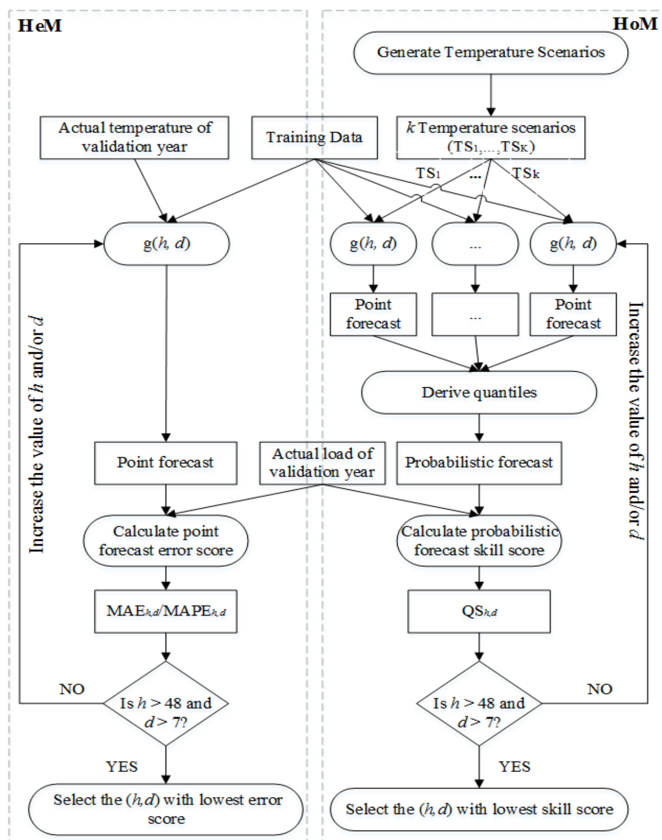

Figure 1. Flow chart of the variable selection process

For a given *h-d* pair, HeM feeds the training data and the actual temperature information of the validation period into the model g(*h*,*d*) to generate hourly point forecast for the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2702751, IEEE Transactions on Smart Grid

validation period. The hourly MAPE of the validation period is then calculated. This process is repeated 392 times until the MAPE values of all the 392 models with different *h*-*d* pair of lagged and moving average temperature combinations are generated. The set of lagged and moving average temperature variables resulting in the lowest point forecast error score (i.e. MAPE or MAE) are selected to construct the underlying model.

B. Holistic Method (HoM)

Alternatively, HoM selects the relevant variables based on the skill of the probabilistic forecast in the validation period. The most commonly used attributes for probabilistic forecast evaluation are reliability, sharpness, and resolution [9]. Quantile Score (QS) was brought to the probabilistic load forecasting literature in GEFCom2014 [20] as a comprehensive measure to evaluate the sharpness and resolution of the probabilistic load forecast. The pinball loss for each quantile and each hour is calculated as described in (4),

$$Pinball(\hat{y}_{t,q}, y_t, q) = \begin{cases} (1-\frac{q}{100})(\hat{y}_{t,q} - y_t) & y_t < \hat{y}_{t,q} \\ \frac{q}{100}(y_t - \hat{y}_{t,q}) & y_t \geq \hat{y}_{t,q} \end{cases} \quad (4)$$

where $\hat{y}_{t,q}$ is the q^{th} quantile (q = 1, 2, …, 99) of the predicted load. The mean of the pinball losses across all the quantiles and all the forecasted hours is then calculated to be the QS of the probabilistic forecast. A lower QS indicates a better forecast. In this paper, we follow the setup of GEFCom2014 by using QS as the evaluation criteria for the probabilistic load forecast of the test period.

HoM first generates *k* temperature scenarios from the *k* years of historical temperature series using the fixed-date method [15]. For a given *h*-*d* pair, it then feeds the training data and one of the *k* temperature scenarios to the model g(*h*,*d*) to generate an hourly point forecast for the validation period. This process is repeated *k* times to generate *k* point forecasts of the validation period with each coming from one temperature scenario. The empirical distribution of the *k* point forecasts is then used to derive the 1st to the 99th quantiles to serve as the probabilistic forecast, which is used to calculate the QS. This process is repeated 392 times for the 392 different combinations of the *h*-*d* pairs. The set of lagged and moving average temperature variables resulting in the lowest QS are selected to construct the underlying model.

IV. Results

A. Variable Selection Results

The two variable selection processes in Fig. 1 are first applied to the GEFCom2014 data. Table II presents the *h*-*d* pairs that yield the lowest MAPE values, MAE values, or QS for each of the 12 months of 2010. The selected *h*-*d* pair for each month will be used to generate probabilistic load forecasts for the same month of the test year (i.e. 2011).

Table II
Optimal *h* and *d* combinations selected for each month

	HeM						HoM		
Month	***h***	***d***	**MAPE(%)**	***H***	***d***	**MAE**	***h***	***d***	**QS**
1	6	2	4.02	6	2	7.25	48	1	18.20
2	5	0	4.44	4	2	8.31	26	0	11.64
3	9	5	6.23	9	7	7.56	9	6	7.45
4	24	3	8.06	48	3	7.98	1	7	4.91
5	5	0	5.64	5	0	7.39	29	6	6.18
6	9	7	3.71	9	7	6.59	0	0	11.30
7	5	0	4.55	38	4	8.38	0	1	9.70
8	9	2	3.73	9	2	7.12	0	1	6.50
9	16	6	5.11	16	4	8.00	43	2	6.01
10	17	7	5.24	17	7	6.28	10	7	4.24
11	12	4	4.77	11	4	6.20	42	7	5.73
12	2	2	5.50	2	2	10.62	48	4	17.68

Fig. 2, Fig.3, and Fig. 4 show the heat maps corresponding to the average MAPE values, the average MAE values, and the average QS for a given *h*-*d* pair across the 12 months of the validation data, respectively. We only present *h* up to 24 for Fig. 2 and Fig. 3 to avoid verbose presentation. A cooler background color indicates a smaller value. In each figure, the lowest value is highlighted in bold. The optimal *h*-*d* pairs selected within the testing range are quite close for the HeM with two different error measures. The MAPE value reaches the minimum when h = 9 and d = 4. The MAE value reaches the minimum when h = 9 and d = 3. On the other hand, HoM leads to a very different *h*-*d* pair. The QS reaches the minimum when h = 29 and d = 6. This is largely due to the different features between the point and probabilistic error measures. Point forecast error measures such as MAPE and MAE evaluate the forecast on the expected value or mean, while quantile score evaluates the forecasts at many different quantiles. Because HeM and HoM take fundamentally different approaches to variable selection, the underlying models are expected to be different. Nevertheless, when the point forecast errors of the candidate models are significantly different (e.g., the top left corner of the heat map where $h < 4$ and $d < 1$), the two variable selection methods agree with each other w.r.t. the ranking of the candidate models.

B. Forecasting Results

The *h*-*d* pair selected for each month using HeM (or HoM) as listed in Table II is then used to forecast the same month of the test year (i.e. 2011) of the GEFCom2014 data. Fig. 5 shows the QS for each month of year 2011 and the average of them. Comparing HoM with MAPE-based HeM, HoM wins 8 of the 12 months. Comparing HoM with MAE-based HeM, HoM wins 7 of the 12 months. In other words, HoM slightly outperforms but does not dominate HeM w.r.t. the skill of probabilistic load forecasts. On average, HoM, MAPE-based HeM and MAE-based HeM lead to the QS of 8.23, 8.24 and 8.22, respectively.

Fig. 6, Fig. 7, and Fig. 8 present the quantile forecasts of the annual peak week of year 2011 from applying HeM and HoM, respectively. The blue dots are the actual load. The dash lines are the 1st to the 99th percentile of the probabilistic forecasts except the medians, which are represented by the red solid lines in the figures. The forecasts from applying these two variable selection methods are quite close to each other with the QS being 13.96 and 13.47 from the MAPE-based HeM and MAE-based HeM, respectively, and 13.86 from the HoM. Again, the difference is subtle.

h \ d	0	1	2	3	4	5	6	7
0	6.481	5.673	5.588	5.570	5.573	5.575	5.589	5.581
1	6.011	5.511	5.415	5.396	5.394	5.398	5.410	5.401
2	5.818	5.455	5.365	5.345	5.342	5.345	5.357	5.347
3	5.690	5.414	5.331	5.311	5.307	5.310	5.323	5.314
4	5.592	5.377	5.304	5.283	5.280	5.282	5.293	5.287
5	5.513	5.349	5.283	5.262	5.258	5.260	5.272	5.267
6	5.459	5.332	5.273	5.252	5.250	5.251	5.265	5.261
7	5.416	5.319	5.266	5.245	5.243	5.246	5.258	5.257
8	5.380	5.306	5.257	5.237	5.235	5.239	5.252	5.253
9	5.355	5.300	5.252	5.234	**5.232**	5.238	5.249	5.251
10	5.340	5.301	5.256	5.236	5.235	5.241	5.256	5.259
11	5.331	5.305	5.258	5.238	5.239	5.246	5.261	5.265
12	5.341	5.325	5.278	5.256	5.256	5.265	5.279	5.281
13	5.347	5.341	5.291	5.268	5.269	5.278	5.292	5.293
14	5.352	5.351	5.300	5.277	5.277	5.288	5.300	5.301
15	5.359	5.357	5.308	5.284	5.285	5.296	5.304	5.306
16	5.365	5.362	5.315	5.289	5.291	5.302	5.308	5.310
17	5.374	5.362	5.314	5.286	5.287	5.298	5.303	5.306
18	5.377	5.362	5.318	5.289	5.290	5.301	5.306	5.309
19	5.379	5.356	5.317	5.288	5.290	5.301	5.308	5.312
20	5.385	5.353	5.316	5.288	5.289	5.300	5.309	5.315
21	5.397	5.358	5.320	5.292	5.293	5.304	5.315	5.321
22	5.413	5.364	5.327	5.300	5.300	5.312	5.323	5.327
23	5.427	5.377	5.338	5.314	5.315	5.328	5.337	5.341
24	5.439	5.381	5.346	5.323	5.325	5.338	5.348	5.352

Figure 2. Average MAPE (%) values of applying different lagged and moving average temperature combination for the validation year

h \ d	0	1	2	3	4	5	6	7
0	9.728	8.499	8.384	8.367	8.389	8.396	8.418	8.415
1	9.038	8.249	8.115	8.092	8.107	8.116	8.136	8.130
2	8.737	8.155	8.031	8.005	8.017	8.024	8.043	8.035
3	8.543	8.092	7.979	7.953	7.962	7.969	7.991	7.984
4	8.394	8.041	7.942	7.916	7.924	7.931	7.949	7.946
5	8.273	8.005	7.919	7.891	7.898	7.904	7.924	7.922
6	8.191	7.988	7.908	7.880	7.890	7.896	7.917	7.918
7	8.125	7.975	7.905	7.877	7.885	7.893	7.913	7.919
8	8.069	7.961	7.898	7.869	7.876	7.888	7.908	7.917
9	8.026	7.956	7.894	**7.868**	7.876	7.890	7.908	7.919
10	8.004	7.964	7.904	7.877	7.885	7.902	7.924	7.937
11	7.993	7.973	7.912	7.885	7.896	7.915	7.937	7.950
12	8.007	8.002	7.942	7.913	7.922	7.944	7.966	7.976
13	8.022	8.027	7.963	7.933	7.942	7.966	7.986	7.995
14	8.034	8.043	7.977	7.947	7.955	7.981	7.997	8.007
15	8.045	8.050	7.987	7.956	7.965	7.991	8.003	8.013
16	8.056	8.058	7.997	7.962	7.972	7.998	8.007	8.016
17	8.072	8.059	7.996	7.959	7.967	7.993	7.999	8.010
18	8.080	8.063	8.006	7.968	7.975	8.001	8.007	8.018
19	8.082	8.057	8.007	7.970	7.977	8.003	8.011	8.024
20	8.095	8.055	8.010	7.973	7.978	8.003	8.014	8.030
21	8.120	8.064	8.021	7.984	7.987	8.013	8.026	8.040
22	8.151	8.075	8.037	8.000	8.003	8.028	8.042	8.054
23	8.178	8.097	8.058	8.024	8.029	8.056	8.069	8.080
24	8.201	8.107	8.074	8.043	8.047	8.076	8.089	8.102

Figure 3. Average MAE values of applying different lagged and moving average temperature combination for the validation year

h \ d	0	1	2	3	4	5	6	7
0	9.338	9.266	9.258	9.253	9.251	9.250	9.244	9.244
1	9.327	9.275	9.264	9.261	9.258	9.258	9.251	9.252
2	9.318	9.276	9.264	9.262	9.260	9.259	9.252	9.253
3	9.312	9.278	9.266	9.264	9.262	9.262	9.255	9.255
4	9.304	9.277	9.265	9.264	9.262	9.261	9.255	9.255
5	9.296	9.277	9.265	9.264	9.262	9.261	9.255	9.255
6	9.289	9.277	9.265	9.264	9.262	9.261	9.255	9.255
7	9.284	9.277	9.266	9.265	9.262	9.262	9.255	9.256
8	9.280	9.278	9.266	9.266	9.263	9.263	9.257	9.257
9	9.276	9.277	9.264	9.264	9.261	9.261	9.255	9.255
10	9.274	9.275	9.262	9.261	9.258	9.258	9.252	9.252
11	9.272	9.272	9.259	9.258	9.255	9.254	9.248	9.248
12	9.270	9.269	9.256	9.254	9.251	9.250	9.245	9.245
13	9.265	9.265	9.252	9.250	9.247	9.246	9.240	9.240
14	9.260	9.260	9.249	9.246	9.243	9.242	9.237	9.237
15	9.256	9.258	9.247	9.245	9.241	9.240	9.235	9.235
16	9.253	9.255	9.245	9.242	9.239	9.238	9.233	9.232
17	9.250	9.252	9.244	9.241	9.237	9.236	9.231	9.230
18	9.248	9.249	9.243	9.240	9.236	9.235	9.230	9.229
19	9.247	9.249	9.243	9.240	9.237	9.235	9.230	9.230
20	9.245	9.248	9.243	9.240	9.236	9.235	9.230	9.229
21	9.244	9.247	9.242	9.239	9.235	9.234	9.229	9.229
22	9.243	9.244	9.239	9.236	9.233	9.231	9.226	9.226
23	9.242	9.241	9.235	9.233	9.230	9.228	9.223	9.223
24	9.240	9.239	9.234	9.231	9.228	9.226	9.222	9.221
25	9.237	9.236	9.232	9.229	9.226	9.224	9.220	9.220
26	9.236	9.236	9.231	9.228	9.225	9.224	9.220	9.219
27	9.236	9.235	9.230	9.227	9.224	9.223	9.219	9.219
28	9.235	9.235	9.229	9.227	9.224	9.223	9.219	9.219
29	9.233	9.234	9.229	9.226	9.224	9.223	**9.2189**	9.219
30	9.233	9.233	9.229	9.226	9.224	9.223	9.219	9.219
31	9.232	9.232	9.228	9.226	9.224	9.223	9.219	9.219
32	9.230	9.231	9.229	9.226	9.224	9.224	9.220	9.219
33	9.231	9.231	9.230	9.227	9.225	9.225	9.221	9.220
34	9.231	9.232	9.231	9.229	9.227	9.226	9.222	9.222
35	9.232	9.232	9.232	9.230	9.228	9.228	9.224	9.223
36	9.232	9.233	9.233	9.231	9.229	9.229	9.224	9.224
37	9.232	9.233	9.233	9.231	9.230	9.230	9.225	9.225
38	9.231	9.233	9.234	9.232	9.231	9.230	9.226	9.226
39	9.231	9.232	9.233	9.232	9.231	9.231	9.226	9.226
40	9.230	9.232	9.233	9.232	9.231	9.231	9.226	9.226
41	9.228	9.230	9.233	9.232	9.231	9.231	9.226	9.226
42	9.228	9.230	9.233	9.234	9.233	9.233	9.227	9.227
43	9.228	9.230	9.234	9.235	9.234	9.234	9.229	9.229
44	9.228	9.230	9.235	9.236	9.236	9.236	9.230	9.230
45	9.229	9.231	9.236	9.237	9.236	9.236	9.231	9.231
46	9.229	9.231	9.237	9.238	9.237	9.237	9.231	9.231
47	9.231	9.232	9.238	9.238	9.238	9.238	9.232	9.233
48	9.231	9.232	9.237	9.240	9.239	9.239	9.234	9.234

Figure 4. Average quantile scores of applying different lagged and moving average temperature combination for the validation year

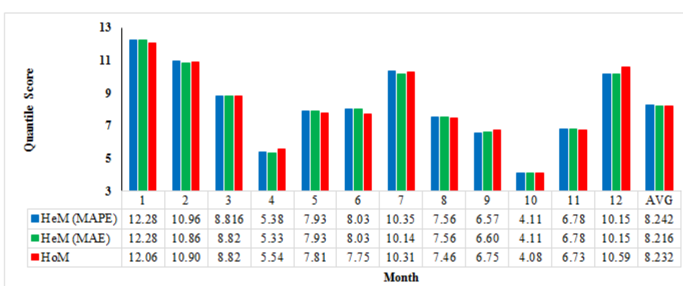

Figure 5. Quantile scores of the test year (2011)

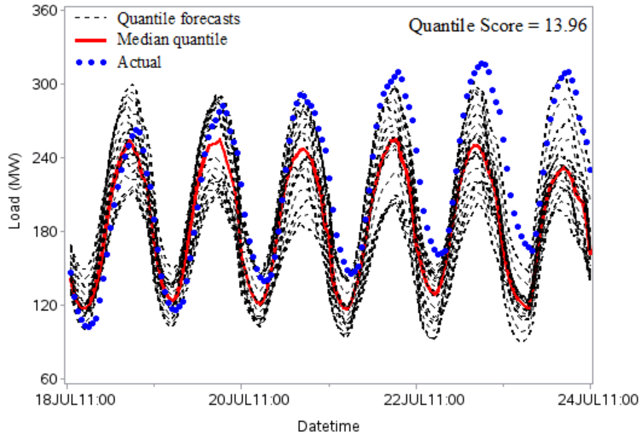

Figure 6. Actual and quantile forecasts of annual peak week in 2011 (MAPE-based HeM)

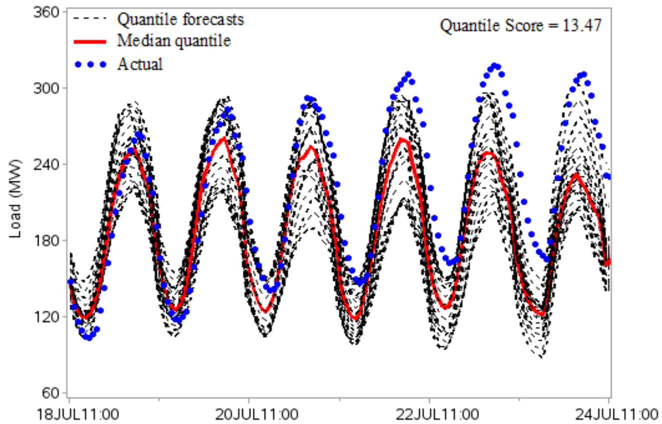

Figure 7. Actual and quantile forecasts of annual peak week in 2011 (MAE-based HeM)

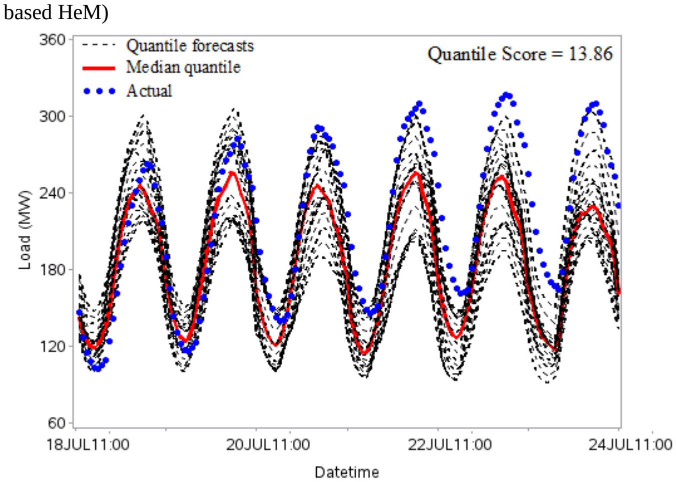

Figure 8. Actual and quantile forecasts of annual peak week in 2011 (HoM)

C. *ISO-NE Case Study*

To avoid drawing conclusions specifically for one dataset, we conduct the same variable selection experiment for each of the eight load zones under ISO-NE. Table III lists the average of the optimal *h* and *d* values selected for each of the eight load zones. In all cases, the two methods return different underlying models, which is also consistent with the earlier findings. Fig. 9 shows the average QS across the 12 months in 2015 (i.e. test year) from using the optimal *h-d* pair selected for each month using HeM and HoM, respectively. On average, the HoM wins 5 of the 8 zones. The differences in QS from implementing these two methods are subtle for all of the eight zones. Again, HoM does not dominate HeM w.r.t. the skill of probabilistic load forecasts.

Table III
The average values of optimal *h* and *d* combinations selected for each zone

	HeM						HoM		
Zone	***h***	***d***	**MAPE(%)**	***h***	***d***	**MAE**	***h***	***d***	**QS**
CT	5	5	3.16	7	5	111.59	33	4	84.73
ME	18	4	3.04	22	4	21.32	33	4	21.32
NEMASSBOST	13	5	2.84	15	5	84.25	18	5	65.64
NH	20	4	2.89	21	4	37.62	33	4	28.05
RI	22	5	3.07	22	5	28.59	21	4	21.71
SEMASS	22	4	3.19	22	4	55.04	21	4	41.47
VT	15	3	3.08	15	3	20.12	30	4	11.45
WCMASS	19	4	2.82	19	5	55.58	24	5	44.39

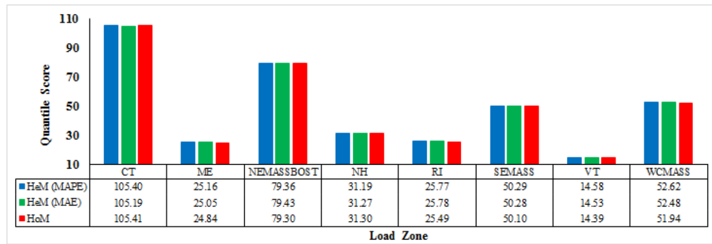

Figure 9: Average quantile scores of the test year (2015) for the load zones

V. Discussions

A. *Computational Intensity*

The technological advancement in the most recent decade has significantly released the constraints of using computationally intensive techniques for load forecasting [2], [4]. Nevertheless, variable selection may still cost a significant amount of time if many candidate variables are being considered. In our case studies, for each of the 12 months in a given load zone, each variable selection method evaluates 392 models. Furthermore, for a given model g(*h*,*d*), HoM needs to execute the model *k* times due to the use of *k* different temperature scenarios. This means the HoM actually processes 392*k* models for a given *h-d* pair.

All the data-processing and modeling tasks are implemented using SAS® 9.4 software on a 64-bit AIX 7.1 server (16 CPUs and 128GB RAM). If we define running through the g(*h*,*d*) model for all possible *h-d* pairs as one evaluation iteration. It takes us between two to three hours to finish one iteration of the evaluation. HoM on average takes about 10 minutes longer than what HeM does for each iteration. For the ISO-NE case study, we process 96 (12 months × 8 zones) iterations. This means that HoM in total takes 16 hours longer (10×12×8/60) than HeM to select the variables for each of the 12 months and all the eight zones. The extra processing time comes from the extra step to generate temperature scenarios and the probabilistic forecast. If the underlying model is more computationally intensive, HoM may require even longer extra processing time.

We can also implement the variable selection methods by leveraging the parallel processing technologies. In other words, we can evaluate multiple zones and/or periods at the same time to save processing time. In the parallel processing setup, HoM requires more storage spaces and is more I/O intensive than the HeM. For example, on the same server mentioned above, we are able to process all the eight zones for the HeM simultaneously while we can only process two zones for the HoM at the same time due to the limitations on storage spaces and I/O constraints.

B. Practical Considerations

As mentioned earlier, the candidate models in this study are having similar point forecasting accuracy when $h>4$ and $d>1$. Although the HeM and HoM returned different variable sets for forecasting, we can still observe similar probabilistic load forecasting performance. This suggests that probabilistic load forecasting performance is insensitive to the underlying point forecasting model when the underlying models are similarly accurate. Practically, when the candidate models are similar on point forecasting performance, it is unlikely to gain much improvement from taking the more computational-intensive holistic method.

To be consistent with GEFCom2014, the empirical studies in this paper were conducted for one-month ahead load forecasting, which is typically considered as a medium term load forecasting problem. Nevertheless, the variable selection strategies investigated in this paper can also be applied to load forecasting problems under other forecast horizon(s). Similarly, the proposed analysis framework can be applied to the load data of different resolution, or to other modeling techniques. As the first formal study in the area of variable selection for probabilistic load forecasting, this paper opens the door to further research about the performance of various variable selection strategies applied to other forecast horizons, data resolutions and modeling techniques.

In this paper, a large number of candidate variables have been evaluated. On one hand, not every load forecasting system evaluates such a large number of variables. The forecaster defines the set of candidate variables based on factors such as data availability, the relevance of variables, and the resource available for conducting the forecasting process. On the other hand, identifying the proper model for the forecasting process may not be conducted in real time. In practice, medium and long term load forecasting typically uses the models identified offline. Even for short term load forecasting, identifying the models once per day would be sufficient. In other words, spending a few hours for model selection would not be a concern for many real-world applications.

Both HoM and HeM in this study are based on computationally exhaustive method for a comprehensive demonstration. Some early stopping rule can be further investigated to save the processing time. Nevertheless, it is also worth pointing out that the traditional variable selection algorithms, such as stepwise regression and LASSO, are different from HoM: (1) these traditional variable selection methods still evaluate the point forecasting performance, so they belong to the HeM category; (2) many of them evaluate the performance based on in-sample-fit instead of out-of-sample tests [23].

VI. Conclusions

In this paper, we tackled the variable selection problem in probabilistic load forecasting, which is a rarely touched area in the load forecasting literature. We proposed a holistic method (a.k.a. HoM) to select a subset of relevant variables for constructing the underlying model for probabilistic load forecasting. We also propose an analysis framework to compare HoM with a heuristic method (a.k.a. HeM), which selects variables by minimizing point forecast errors. Through an empirical study using two publicly available datasets covering seven states of the United States, we found that the two variable selection methods may agree with each other when the candidate models are significantly different w.r.t. the point forecast accuracy. In reality, when the variables are leading to the models with similar point forecast accuracy, these two variable selection methods indeed return quite different underlying models for probabilistic load forecasting. HoM slightly outperforms but does not dominate HeM w.r.t. the skill of probabilistic load forecasts. The price paid to achieve this minor skill improvement includes processing time and I/O throughput. In short, it is unlikely for us to significantly improve the forecast skill by taking the holistic yet computationally intensive variable selection method.

This paper is the first formal study on the variable selection problem of probabilistic load forecasting. The empirical study reveals the pros and cons of two variable selection methods. Although this empirical research covers seven states of the United States, the aforementioned conclusion may or may not be applicable to other states or other countries. Other empirical studies on variable selection following the same comparison framework would be encouraged as part of the future work. In addition, the findings from this paper may also spark other interesting questions for future research. Recall that some recent progress was made on combining point forecasts from sister models [17], [22]. While HoM and HeM are giving probabilistic load forecasts with similar skill, we can ask another research question: *will the combination of probabilistic forecasts outperform each individual?* This paper certainly paves the way towards future research on combining probabilistic load forecasts.

VII. Reference

[1] T. Hong, "Short term electric load forecasting," Ph.D. Dissertation, Graduate Program of Operation Research and Dept. of Electrical and Computer Engineering, North Carolina State University, 2010.

[2] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: a big data approach," *Int. J. Forecast.*, vol. 32, no. 3, pp. 585–597, 2016.

[3] J. Xie, B. Liu, X. Lyu, T. Hong, and D. Basterfield, "Combining load forecasts from independent experts," in *47th North American Power Symposium (NAPS2015), October 4-6,* 2015.

[4] F. Ziel and B. Liu, "Lasso estimation for GEFCom2014 probabilistic electric load forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1029–1037, 2016.

[5] J. Xie, Y. Chen, T. Hong, and T. D. Laing, "Relative humidity for load forecasting models," *IEEE Trans. Smart Grid*, in press, 2016.[Oneline]. avaliable:10.1109/TSG.2016.2547964
[6] S. Ben Taieb and R. J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," *Int. J. Forecast.*, vol. 30, no. 2, pp. 382–394, 2014.
[7] J. R. Lloyd, "GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes," *Int. J. Forecast.*, vol. 30, no. 2, pp. 369–374, 2014.
[8] R. Nedellec, J. Cugliari, and Y. Goude, "GEFCom2012: Electric load forecasting and backcasting with semi-parametric models," *Int. J. Forecast.*, vol. 30, no. 2, pp. 375–381, 014.
[9] T. Hong and S. Fan, "Probabilistic electric load forecasting: a tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1–32, 2016.
[10] R. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1142–1153, 2010.
[11] 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, 2014.
[12] J. Xie and T. Hong, "GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1012–1016, 2015.
[13] J. Xie, T. Hong, T. Laing, and C. Kang, "On normality assumption in residual simulation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1046-1053, 2017.
[14] V. Dordonnat, A. Pichavant, and A. Pierrot, "GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1005–1011, 2016.
[15] J. Xie and T. Hong, "Temperature scenario generation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. in press, 2016.
[16] S. Bentaieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2448–2455, 2016.
[17] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 730-737, 2017.
[18] S. Haben and G. Giasemidis, "A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1017–1022, 2016.
[19] P. E. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1166–1172, 2005.
[20] T. Hong, Pierre Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *Int. J. Forecast.*, vol. 32, no. 3, pp. 896–913, 2016.
[21] ISO New England, "ISO New England Zonal Information," 2016. [Online]. Available: https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info. [Accessed: 26-Oct-2016].
[22] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40–49, 2016.
[23] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *Int. J. Forecast.*, vol. 16, no. 4, pp. 437–450, 2000.

Jingrui Xie received her bachelor degree in Finance from Sun Yat-sen University, Guangzhou, China, her master degree in Economics from Duke University, Durham, NC, USA, her master degree in Engineering Management, and her Ph.D. degree in Infrastructure and Environmental Systems from University of North Carolina at Charlotte, Charlotte, NC, USA in 2009, 2011, 2015, and 2016, respectively. Dr. Xie is a Research Statistician Developer at SAS Institute Inc. where she leads the energy forecasting research. She was the primary statistician developer for the SAS Energy Forecasting Solution and has provided consulting services on load forecasting to several utilities for system operations, planning and energy trading.

Tao Hong received the B.Eng. degree in automation from Tsinghua University, Beijing, China, in 2005, and the Ph.D. degree in operations research and electrical engineering from North Carolina State University, Raleigh, NC, USA, in 2010. He is the Director of the Big Data Energy Analytics Laboratory, NCEMC Faculty Fellow of Energy Analytics and Assistant Professor of Systems Engineering and Engineering Management, University of North Carolina at Charlotte, Charlotte, NC. Dr. Hong is the Founding Chair of the IEEE Working Group on Energy Forecasting, and General Chair of the Global Energy Forecasting Competition. He is an editor of IEEE Transactions on Smart Grid and associate editor of International Journal of Forecasting. His primary research interest is in advanced analytics and its applications to the energy, supply chain, healthcare and finance industries.