Interactive Visualization Of Electricity Load Prediction for Demand Side Energy Management

Chao Chen, Yaoyi Guan, Xiguang Liu, Guolin Yao, Ye Zhang
cchen711,yguan35,xliu765,gyao30,yzhang3412@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

1 INTRODUCTION - MOTIVATION

United States' consumption of electricity increased almost monotonically since 1960. However, electricity cannot be stored at a large scale. Hence, the power generation must balance with the load at any time. However, the load can change dramatically across hours. For instance, the load at 5pm is generally higher than the load at 11pm. The power utilities have to run high-cost power plants to cover the peak load, which induces higher electricity prices, higher electricity production cost, higher emissions and lower efficiency. Energy demand management, also known as demand-side management (DSM), is the modification of consumer demand for energy through various methods such as financial incentives and behavioral change through education. The goal of demand-side management is to encourage the consumers to use less energy during peak hours, or to use energy during off-peak period such as nighttime and weekends. Demand management could reduce the need for investments in networks and power plants for meeting peak demands, reduce emissions from low efficiency power plants, and improve reliability of power system. In peak load hours, the electricity load in U.S. is around 9,000,000 MW. Based on a study from Stanford University, every MW reduction of peak demand hour in U.S. translates to a system-wide savings of about $130 million.

However, the ability of electricity consumers to altering demand is relatively low. In recent decade, energy demand management has been promoted by power companies but remarkable improvement on demand side was not observed. One of the root causes is that the electricity consumers do not have necessary information to support managing the utilization. Probably part of the consumers might be able to adjust the utilization based on the temporal characteristics of demand. However, the true demand curve for electricity is complex and complicated since the consumption of electricity could be significantly influenced by weather and other factors, like hours of a day, day of a week, holidays and so no. Utility companies and energy trading companies predict loads for daily operation. In contrast, there has not been any product or services in the market providing accurate and reliable real-time information of power load to facilitate consumers reducing their electricity bill through participate demand side management.

2 PROBLEM DEFINITION

In order to prompt demand side energy management, it is necessary to provide electricity consumers accurate information about power load and motivate consumers utilizing the information in managing electricity utilization. The current project proposed a unique tool visualizing the weather forecast data and predicting power load in the next 12 hours in the area of interest, which is provided by a machine learning model. It will provide intuitive information to help all common consumers participate demand side energy management much easier than before, which will reduce their electricity bills, reduce emissions from power plants and improve reliability of power supply.

3 LITERATURE SURVEY

Weather conditions, such as temperature, are the major predictors of electricity load forecasting. For current project, both electricity load and weather information will be visualized. Hence, the literature survey includes two parts, data visualization and electricity load forecasting. Extensive research has been performed over the subject of data visualization, including the visualization of weather data. Traditional visualization method are surface-based. In recent years, multiple techniques have been developed to overcome the complexity of weather data. A novel visualization approach utilizing multi-dimension data to present meteorological data, hydrometer particles and the prediction from optical model for multiple hydrometer fields was proposed [16]. Similarly, natural textures as extra or main hues to represent the dimensions of weather data in a controllable manner was proposed to represent multi-dimensional weather data [17]. The interaction between weather data and other factors could be visualized in an interactive way [11]. It inspires the authors to integrate the weather data and power load. Integrated interactive models focusing on interactions flows and multivariate patterns calculated the strength of connection between locations and grouped spatial units into contiguous regions [7]. Significant information should be highlighted in the visualization [1], including the temporal-spatial correlation between variables [15][5]. Beside the correlation between time, space and weather, time-series information of weather data will be one of the key function in the proposed tool [14]. Short-term load forecasting is of great interest in electricity industries and academical institutes. The power load could be predicted by multiple candidate algorithms [18][4]. Similar day algorithm [13] is one of the candidate options, which has been applied to model development [6][2]. Machine learning techniques will be definitely another option [12]. Neural network model [3] and deep learning model [8] have been tested. In the research of short-term forecasting of power load, it was observed that the power load level could be influenced by multiple major factors [9]. As a result, variable selection would be one of the most significant step in power load prediction[10]. The aforementioned 18 papers provide various methods for visualization of weather data or forecast of power load. None of them provide interactive visualization of weather data and power load. The tool developed in current project will provide interactive visualization of weather data and power load and quantify the dynamics between them.

4 PROPOSED METHOD

4.1 Innovations

Energy demand management has big potential to bring huge benefits to customers, electricity companies, electricity operators and the entire society through minimizing power generation cost, reducing emissions and improve reliability. However, demand-side management is extremely challenging since the overwhelming majority of the electricity consumers do not have ability and necessary information to manage the utilization. **The proposed tool, as far as we know, is the first one designed to try to fill the gap between the supply side and demand side induced by the information asymmetry. It would be better than any state-of-the-art DSM system because of the key innovations listed below:**

Innovation 1: In order to prompt demand-side energy management, it is necessary to help millions of electricity customers to determine the hours with lower electricity demand, which indicates lower unit price. Although short-term load forecasting is a common practice for utility companies and energy trading organization, there has not been a product or service providing similar information to common electricity customers. The proposed tool is user friendly. It will provide local load prediction to promote demand-side management.

Innovation 2: Consumers are used to monitor the weather forecast on a regular basis. The proposed tool will incorporate the prediction of power load into the visualization of weather forecast, which will be frequently monitored by the consumers. As a result, it could provide users intuitive, visualized information and motivate the consumers to manage electricity utilization in a data-driven way without changing the consumers' behavior patterns.

Innovation 3: This tool can easily incorporated with Smart Home system in the near future. Through this tool, the Smart Home System can automatically setup or recommend how to use electricity in a house. For instance, it can recommend the time for dish washing, EV recharging, optimal setting points of air condition, and so on.

Note that the innovations of current project focuses on the application of existing theory/methods to provide unique services to both the utility companies and electricity consumers for the sake of improving the demand-side management. The metrics of measuring the success of the tool include forecast accuracy, the population of tool users and how much loads in peak hours are shifted to off-peak hours. The success of the proposed tool will significantly improve the performance energy demand management, which will lead to increasing reliability power plant operation, decreasing fossil energy consumption and less emissions. The main risks of the tool is the limitation on data sources. The data utilized in current research are free. Charge might be imposed on the data. It will induce negative impact on the marketing.

The current project is comprised of two parts. The first part is data visualization, which present the information of weather together with the predicted power load in a region of interest. The second part is development of a model predicting the power load in specific region based on geographic, population, and weather information using machine learning methods. Details are presented in the next two sections.

4.2 Data collection/organization

The data for the current project are mainly collected from open sources. Details are listed below.

Map data: The zipcode map for New York State is not readily available. It is clipped and processed based on New York State Contour shapefile and US zipcode shapefile using ArcMap 10.4. The New York State contour map is downloaded from https://cugir.library.cornell.edu/catalog/cugir-007865. The US zipcode map is from https://www.census.gov/cgi-bin/geo/shapefiles/index.php by selecting year 2019 and layer type of "ZIP Code Tabulation Areas". Shape files are eventually converted to topojson through an online conversion tool available at https://mapshaper.org/.

Geographic information: New York weather station address was manually downloaded. The population by zip code in New York States was retrieved from United States Census Bureau. There are 2169 distinct zip code in New York State in total.

Historical weather data: The historical weather data contributed by 41 New York State weather stations in a time window from 2015-1-1 to 2020-3-1 was ordered and obtained from https://www.ncdc.noaa.gov/cdo-web/datatools/lcd. There were has 41 csv files, one for each weather station. There were 3,595,660 rows of data in total. Since the raw data files were organized by multiple report types, Python was utilized to filter out one uniform type that contains 1,889,757 rows of data. Data cleaning was performed on OpenRefine to remove observations with missing values. The cleaned data contains as many as 1,482,463 rows.

Power load data: The New York State ISO zone load data in a time window from 2015-1-1 to 2020-3-1 was comprised of 1,733 csv files and were manually downloaded from https://www.nyiso.com/custom-reports?report=rt_actual_load. The data was concatenated and contained half a million records.

AWS MySQL database was created as data warehouse. All the aforementioned data was combined and loaded into the database, which contains 1,474,489 rows of data in total. The data would be queried for data visualization and model development.

Inputs for model: The model inputs requires real-time information. The forecast of weather will be retrieved from the API: http://dataservice.accuweather.com/forecasts/v1/hourly/12hour/. The holiday information will be obtained via API: https://calendarific.com/api/v2/holidays. Note that the inputs for model will be plug into the model to predict the load.

4.3 User interface

Data visualization was implemented by D3 embedded in React. React is modern front-end framework compatible with all the mainstream browsers. More importantly, a list of mobile-first designs stem from this framework. One of them, which was used in this particular project is Material UI framework. This single page app consists of three compartments, namely the top navigation bar, left side drawer, and main display area. The top nav bar is mainly to direct users to view relative links, such as our GitHub repository page (https://github.com/yzhang250/weather-electricity-load-browser) and contact information. The drawer features two options, which are historical data viewer and the prediction of power load. Once one of the options is selected, corresponding data visualization will be displayed in the main display area.

The historical data viewer enable users quickly checking past weather data and electricity load data in an interactive manner. The viewer has two clickable maps, which are showing temperature choropleth map on the left and an electricity load choropleth map on the right, and an information card above them. By clicking on different areas on the map, the corresponding weather information, such as temperature, wind speed, humidity, will be instantly shown in the information card.

The prediction section shows the temperature and predicted power load in next 12 hours of selected region on the left panel. The user interface receives clicked zip code region from the historical map, by doing which users are also allowed to pinpoint the geographic location of the selected zip code region. Once the submit button under the prediction section is submitted, the interface would present a line graph with horizontal axis showing next 12 hours, left y axis representing the predicted load in megawatts hour (MWH), and right y axis displaying the scraped weather data forecast scraped from https://developer.accuweather.com/. A tooltip of

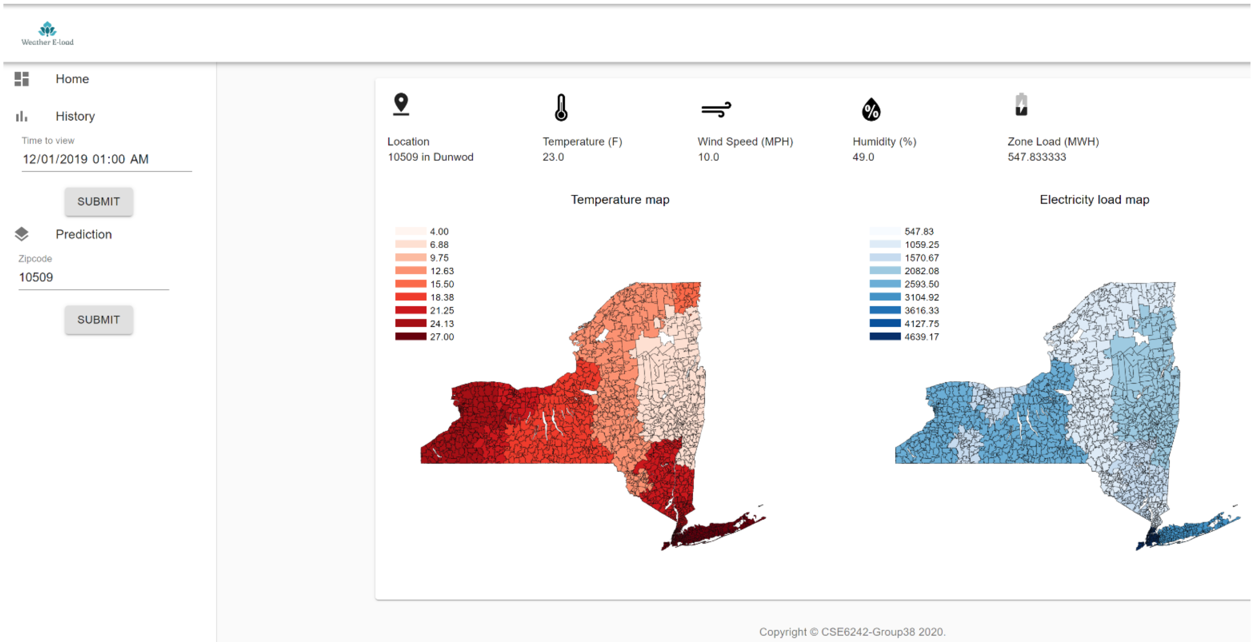

Figure 1: Snapshot of data visualization: historical information of weather and power load in New York State

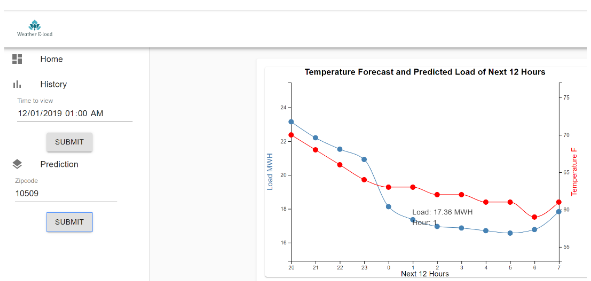

Figure 2: Snapshot of data visualization: prediction of power load and temperature forecast in New York State

selected hour and temperature/load zone is shown once a dot is clicked. The line graph aims to 1) show the correlations between the temperature (the red line) and the load (the blue line), and 2) assist users in managing electricity utility.

4.4 Algorithm for load prediction

A model need to be developed to predict the short-term loads in the region of interest in New York State. Short-term load forecast is a hot topic in electricity industries. A lot of algorithm have been tested, including linear regression, time series analysis, support vector machine, random forest, Gaussian process regression, similar day method, neural network and gradient boosting machine (GBM) and so on. The challenge is that since the load patterns of different areas are quite different, so there is no on all-around algorithm, which is definite better than others for short-term loading forecasting. Another challenge is that it is very difficult to further improve the forecasting accuracy. In current project, all the algorithms listed above were preliminarily tested to train a model based on a subset of the historical data. This subset of data include the information of one region, the New York City, in a one-year time window. Based on the model accuracy, stability, tuning requirement, overfitting risk, capability to deal with both numerical and categorical inputs, it was found that random forest and GBM, two tree-based algorithm are more suitable for the load prediction. As a result, random forest and GBM were more extensively tested using historical data over the entire New York State in a three-year time window. **In the process of model training, the dataset was split into two parts, 70% for training, 30% for test.**

Another important part in model development is feature selection. Weather conditions, especially temperature, are major factors influencing electricity utilization.

Other factors, such as hours of day, day of week also influence the pattern of the load in one area. The features listed above are widely selected in the models developed in electricity industries. However, the research in present project revealed the limitation and weakness of the common practice in industries. **Two main innovations of feature selection was proposed in the current research.**

First, among numbers of weather indicators, the common practice in industrial filed for load forecasting just take temperature as predictor. In current research, more weather indicators were explored. After incorporating humidity, wind speed and wind direction into consideration, it was found that the combination of temperature and wind speed can reduce forecasting error by 0.2% , comparing to using temperature as the single weather indicator.

Second, the weather in one region is forecasted by multiple weather stations in one state. The common practice in industry selects the data from several weather stations near big cities in the state and utilizes the simple average temperature as one of the predictors. In current project, it was found that the population-weighted average of the temperatures reported by 39 weather stations in New York State reduced the load forecasting error by 1.2%, comparing to using simple average.

The selected features for model training and the features screened out were listed in Table 1 and 2, respectively. Note that the feature selection was sophisticated despite of the limited number of features. Details of the observation from the data and model will be depicted in next section. In general, the result of feature selection is consistent with intuition and facilitates the model interpretation.

Categorical Features	Numerical Features
hours of a day	temperature
day of week	wind speed
month	population by zipcode
holiday	population of NY
zipcode	

Table 1: Features selected for model training.

Categorical Features	Numerical Features
cloud	wet bulb temperature
wind direction	feel-like temperature
	humidity
	relative humidity

Table 2: Features dropped from model training.

5 EXPERIMENTS & EVALUATION

Two models using random forest and GBM were implemented. An interface for data visualization and information delivery was designed for experiment and evaluation. The implementation of the models were performed on the platform of Microsoft Azure. API of trained model is https://ussouthcentral.services.azureml.net/workspaces/d53edbbac2e848cf8a74c034f195ea53/services/9cf6d1493a08452b9374fbd8ff9895cf/execute?api-version=2.0&details=true.

The experiment was designed to answer the following questions: 1. which algorithm will provide better prediction, random forest or GBM; 2. does the proposed model outperform the benchmark; 3. how many models are needed to provide load prediction in different regions; 4. what is the best setting for hyperparameters; 5. how accurate the final model is; 6. how to perform model maintenance.

In the proposed product, the number of models need to be determined. There were two options. The first type is one overall model using zip code as one of the input parameters, so it can predict the electricity load for all the regions based on zip code. The second type is training multiple models, one for each region. The accuracy of the second option was just slightly better (0.4%) than the single overall model. However, the time consumption in model training for the second option was much higher. The future updating and maintenance of the second option will require much more efforts. Hence, the single overall model was selected to predict the loads for different regions.

The other part of the experiments were parameter tuning for the machine learning model. Grid search was employed to determine the optimized hyperparameters, which are listed as following. Number of trees: 500; Number of variables randomly sampled as candidates at each split: 3; sample size: 80%; node size: 5. Visualization of model fitting results is presented in the appendix.

Evaluation approach and results are described as below. In the current project, the final model was selected from random forest and GBM after the preliminary screen of algorithms. The evaluation of the model performance will focuses on the diagnostics of

Algorithm	RF	GBM
MAE	51.5	57.1
RMSE	100.1	101.5
RSE	0.44%	0.46%
CoD	0.9972	0.9972

Table 3: Model performance over training data, random forest vs. GBM

Algorithm	RF	GBM
MAE	60.6	67.9
RMSE	115.8	117.5
RSE	0.53%	0.54%
CoD	0.9948	0.9946

Table 4: Model performance over test data, random forest vs. GBM

the model, especially the goodness of fit, and the accuracy of the prediction. The goodness of fit was evaluated based on coefficient of determination (CoD), which can be compared against the benchmark value of the industry. The accuracy of the prediction was evaluated in mean absolute error (MAE), root mean square error (RMSE), relative squared error (RSE). The metrics of model performance for training data and test data are listed in Table 3 and 4, respectively, to compare the random forest and GBM model.

As shown in Table 3 and 4, the test error is slightly higher than test error as expected. Through the intensive algorithm screening and novel feature selection, the short-term power load prediction presents high accuracy. The CoD is about 2% higher than the benchmark value of electricity industry. In addition, the random forest model outperforms the GBM. As a result, random forest model was selected to provide power load prediction for the proposed product.

The observations in the experiments of visualization and model implementation are listed below. First, the influence of temperature is very obvious. When temperature is lower than 60 F, the power consumption is reverse proportional to temperature. However, when temperature is higher than 60 F, it is proportional to temperature. The influence of temperature higher than 60 F is more significant. Second, power load shows strong periodic pattern. In large time scale, say a year, it was dominated by seasonality (electricity utilization in winter is higher than that in summer). In median time scale, say a week, it was influenced by workday and weekends. In small time scale, say a day, electricity load surges after work and before bed time. The periodic pattern is characterized by features related to time. Third, the geographic information will incorporate the influence of load from different types of consumers and the number of consumers. For example, the load pattern in business district is distinct from that in residential area. Forth, population-intensive region will intuitively demands higher load during night. The geographic and population influence was characterized by population and zip code. The utilization of air conditioner or heaters is sensitive to weather. This was characterized by temperature and wind speed.

Model maintenance plan: The electricity load pattern will change over time. As a result, model performance would be monitored. The characteristic stability of the input features together with the metrics listed in Table 3 would be monitored. Usually, the load forecasting model should be updated with new historical data at regularly basis, usually updating the model one time each year should be enough. If any performance deterioration was observed, model refit or re-development may be necessary.

6 CONCLUSION AND DISCUSSION

In current project, a unique product was initiated, implemented and tested. **Each team member contribute evenly to the project.** The proposed product could provide users intuitive, accurate and real-time information of weather forecast and prediction of power load in the next 12 hours in the region of interest, which has not been provided by any existing products or energy management systems. The delivered information could facilitate users, especially the electricity consumers, to improve the demand-side management and will result in reduction in power load during peak hour, suppressing the cost for both electricity consumers and producers, reduction in fossil energy consumption and emission. The current product will be compatible to existing network of smart home systems. It can be utilized to combine with the information of electricity price to automate the utilization of electricity facilities.

A APPENDIX: VISUALIZATION OF RANDOM FOREST MODEL OUTPUT

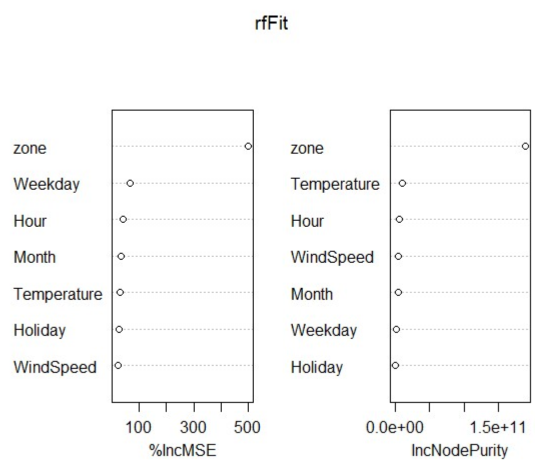

Figure 3: Information of selected features in random forest model

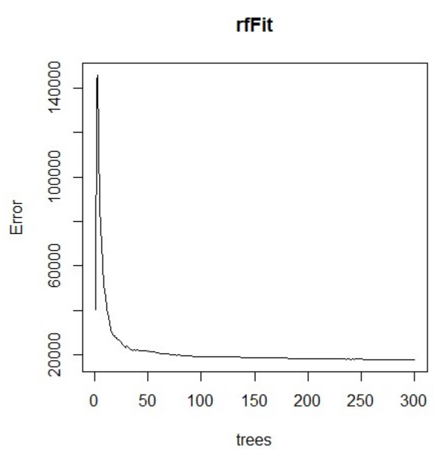

Figure 4: Model error and number of trees to grow

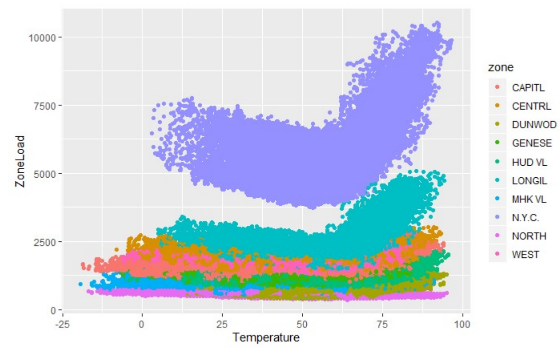

Figure 5: Power load as a function of temperature by regions in New York State

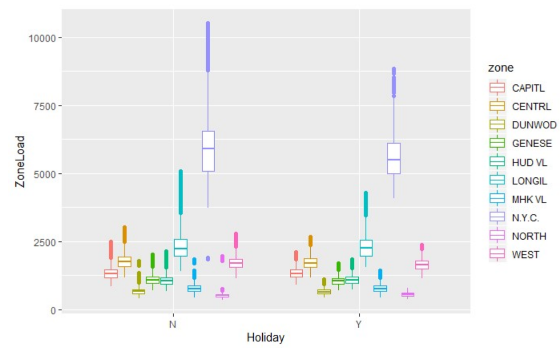

Figure 6: Power load as a function of holidays by regions in New York State

REFERENCES

[1] F. P. Boscoe, C. McLaughlin, M. J. Schymura, and C. L. Kielb. 2003. Visualization of the spatial scan statistic using nested circles. *Health & Place* 9, 3 (2003), 273–277. https://doi.org/10.1016/S1353-8292(02)00060-6

[2] Y. Chen, P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, M. A. Coolbeth, P. B. Friedland, and S. J. Rourke. 2010. Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks. *IEEE Transactions on Power Systems* 25, 1 (2010), 322–330. https://doi.org/10.1109/TPWRS.2009.2030426

[3] T.W.S. Chow and C. T. Leung. 1996. Neural network based short-term load forecasting using weather compensation. *IEEE Transactions on Power Systems* 11, 4 (1996), 1736–1742. https://doi.org/10.1109/59.544636

[4] V. Dehalwar, A. Kalam, M. L. Kalhe, and A. Zayegh. 2016. Electricity load forecasting for Urban area using weather forecast information. In *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*. IEEE, Shanghai, China. https://doi.org/10.1109/ICPRE.2016.7871231

[5] A. M. Dokter, P. Desmet, J. H. Spaaks, S. van Hoey, L. Veen, L. Verlinden, and W. Bouten. 2019. bioRad: Biological analysis and viualization of weather radar data. *Ecography* 42, 5 (2019), 852–860. https://doi.org/10.1111/ecog.04028

[6] Yonghan Feng, Dinakar Grade, S. M. Ryan, J. Watson, R. J. Wets, and D. L. Woodruff. 2013. A new approximation method for generating day-ahead load scenarios. In *2013 IEEE Power Energy Society General Meeting*. IEEE. https://doi.org/10.1109/PESMG.2013.6672564

[7] Diansheng Guo. 2009. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on*

Visualization and Computer Graphics 15, 6 (2009), 1041–1048. https://doi.org/10.1109/TVCG.2009.143

[8] Z. Guo, K Zhou, X. Zhang, and S. Yang. 2018. A deep learning model for short-term power load and probability density forecasting. *Energy* 160 (2018), 1186–1200. https://doi.org/10.1016/j.energy.2018.07.090

[9] Tao Hong. 2010. *Short Term Electric Load Forecasting*. Ph.D. Dissertation. North Carolina State University. http://www.lib.ncsu.edu/resolver/1840.16/6457

[10] Sen-Hao Lai. 2018. *When One Size No Longer Fits All - Electric Load Forecasting with a Geographic Hierarchy*. SAS White Paper. https://www.sas.com/en/whitepapers/when-one-size-no-longer-fits-all-106226.html

[11] P. Lundblad, O. Eurenius, and T. Heldring. 2009. Interactive Visualization of Weather and Ship Data. In *2009 13th International Conference Information Visualisation*. IEEE. https://doi.org/10.1109/IV.2009.38

[12] V. M.Krasnopolsky and M. S. Fox-Rabinovitz. 2006. Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Networks* 19, 2 (2006), 122–134. https://doi.org/10.1016/j.neunet.2006.01.002

[13] F. A. Monforte. 2016. *A Practitioner's Guide to Short-term Load Forecast Modeling*. Itron.

[14] H. T. Nguyen, T. V. Tran, P. V. Tran, and H. Dang. 2013. Multivariate cube for visualization of weather data. In *2013 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 311–316. https://doi.org/10.1109/ICCAIS.2013.6720574

[15] V. Pena-Araya, E. Pietriga, and A. Bezerianos. 2019. A Comparison of Visualizations for Identifying Correlation over Space and Time. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 375–385. https://doi.org/10.1109/TVCG.2019.2934807

[16] K. Riley, D. Ebert, C. Hansen, and J. Levit. 2003. Visually accurate multi-field weather visualization. In *IEEE Visualization, 2003. VIS 2003*. IEEE. https://doi.org/10.1109/VISUAL.2003.1250383

[17] Ying Tang, Huamin Qu, Yingcai Wu, and Hong Zhou. 2006. Natural Textures for Weather Data Visualization. In *Tenth International Conference on Information Visualisation (IV'06)*. IEEE. https://doi.org/10.1109/IV.2006.77

[18] Rafal Weron. 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30, 4 (2014), 1030–1081. https://doi.org/10.1016/j.ijforecast.2014.08.008