



Conference on Systems Engineering Research (CSER 2014)

Eds.: Azad M. Madni, University of Southern California; Barry Boehm, University of Southern California;
Michael Sievers, Jet Propulsion Laboratory; Marilee Wheaton, The Aerospace Corporation
Redondo Beach, CA, March 21-22, 2014

Residential Power Load Forecasting

Patrick Day, Michael Fabian, Don Noble, George Ruwisch, Ryan Spencer,
Jeff Stevenson, Rajesh Thoppay

Georgia Institute of Technology, North Avenue, Atlanta, GA 30332, USA

Abstract

The prepaid electric power metering market is being driven in large part by advancements in and the adoption of Smart Grid technology. Advanced smart meters facilitate the deployment of prepaid systems with smart prepaid meters. A successful program hinges on the ability to accurately predict the amount of energy consumed on a daily basis for each end user. This method of forecasting is called Residential Power Load Forecasting (RPLF). This paper describes the systems engineering (SE) processes and tools that were used to develop a recommended load prediction model for the project sponsor, SmartGridCIS. The basic concept is that power is treated similar to a prepaid telephone in a “pay as you go” fashion. Modeling techniques explored in the analysis of alternatives (AoA) include Fuzzy Logic, Time Series Moving Average, and Artificial Neural Networks (ANN). SE tools such as prioritization and Pugh matrices were used to choose the best-fit model, which ended up being the ANN. Cognitive systems engineering was used in conjunction with the task analysis. Requirements were developed using the commercial tool IBM Rational DOORS®.

© 2014 The Authors. Published by Elsevier B.V.
Selection and peer-review under responsibility of the University of Southern California.

Keywords: Energy Load Forecasting; Short-Term Forecasting; Long-Term Forecasting, Smart Grid

1. Introduction

The exponential growth of the Smart Grid and the rapid deployment of software-enabled meters provide utility companies and consumers alike new opportunities in the way that energy delivered and consumed. The global growth of prepaid electricity programs has been steady and gradual in recent years, but now, thanks to the Smart Grid, the prepaid metering market is now poised to take off on a larger scale. The prepaid energy market is similar to the pre-paid cellular phone concept. Energy consumers are able to pay for energy usage in advance, making an account deposit and having the daily energy usage cost debited from the account. This system provides utilities a way to service customers with poor credit rating or who have a delinquent payment history because Smart Meters allow real-time monitoring of individual customer usage. In order to provide the end user with sufficient notice and accurate account balance status, an accurate individual-customer forecasting method is needed. Residential Power Load Forecasting (RPLF) is a method to predict power usage for individual consumers based on both historical energy use and weather data.

The purpose of this paper is to show how Systems Engineering can be used to determine the most appropriate forecasting model for RPLF. The remainder of the paper is outlined as follows:

- Section 2 - background information on energy forecasting & cognitive systems engineering
- Section 3 - literature review of individual load forecasting
- Sections 4&5 - discussion of design and architecture considerations respectively
- Section 6 - outline of the proposed forecasting model
- Section 7 - outline of future work
- Section 8 - conclusion

1.1. System characterization

Figure 1 depicts an overview of the Prepaid Smart Meter System of Systems (SoS), which describes information flow between the operational elements of the system in order to set a frame of reference for this work.

The customer’s home residence is equipped with a specially designed prepaid smart meter rather than a standard Radio Frequency (RF) communications enabled smart meter or electromechanical meter. The process that the customer goes through to add more available energy to their account involves a swipe of a prepaid reloadable chip card on the card slot of the prepaid smart meter. The balance of the chip card is monitored and added to by using various customer interaction methods, such as smartphone/tablet applications, SMS, web applications, email, kiosks, etc. These interaction methods inform the customer how long their prepayment will provide available energy at their meter, and add energy (via adding credit) to their account as needed.

The smart meter installed at the customer location is monitored and controlled by a connection to the utility company’s Advanced Metering Infrastructure (AMI). Examples of the type of energy consumption statistics that are collected from customer smart meters include kilowatt usage per various time intervals, smart appliance usage details, and so forth. These data are fed into the Meter Data Management System (MDMS) at the utility company, where it is stored and analyzed. Not shown in the diagram is the input of weather data to the utility company third-party sources, used as a factor in the calculations of energy forecasts.

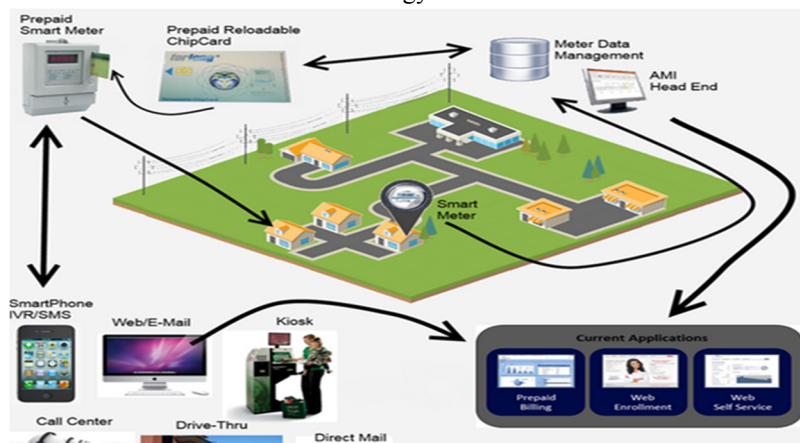


Fig. 1. System Characterization

2. Background

2.1. Energy forecasting

Load forecasting is a method used to calculate the expected energy requirements of a system at some future time. Forecasters rely on historical consumption data to determine how much power a customer (or group of customers) may require. The data are analyzed using models that vary in sophistication from linear regression to trained neural networks. Model inputs also vary and can include day of the week, holiday calendars, weather conditions and forecasts, geographical differences, demographic information, and so on. Traditionally, load forecasting has been used for classes of users resulting in a sometimes-wide variance between the mean value for a class and any particular user. Forecasting at the individual level is intricately more difficult due to the variability of human behavior. Another layer of complexity stems from the fact that large numbers of customers, the amount of historical data required, and the level of computation moves the forecasting problem into the realm of big data and high-performance computing. Traditionally, load forecasting has been used to make decisions about daily generation scheduling, planning for grid expansion, and accurately determining prices for large industrial customers. Accurate forecasting is important in prepaid energy because of the overhead involved in continuously reading a large number of meters installed on a communications network.

2.2. Cognition systems engineering

Cognitive systems engineering is a specialty discipline of systems development that addresses the design of socio-technical systems. A sociotechnical system is one in which humans provide essential functionality related to decision making, planning, collaboration, and management. Drawing on contemporary insights from cognitive, social, and organizational psychology, cognitive systems engineers seek to design systems that are effective and robust. The focus is to make cognitive work easier to perform and thus more reliable by integrating technical functions with the human cognitive processes they need to support.

Cognitive systems engineers assist with the design of human interfaces, communication systems, training systems, teams, and management systems. They employ principles and methods that bear on the design of procedures, processes, training and technology. Examples of systems that can benefit are military command and control, civil air traffic control, transportation, communication, process control, power generation, power distribution, healthcare, and large scale project infrastructure.

The need for a systematic and comprehensive approach to cognitive issues in the design of sociotechnical systems has emerged over the past twenty years as computer-based technologies have pushed the nature of operational work in a direction in which cognitive challenges dominate. Issues such as decision-making in complex and dynamic information environments, distributed collaboration, and management of extensively networked systems have, in many cases, transformed the nature of work. Cognitive systems engineers identify the cognitive states, the cognitive processes, and the cognitive strategies used by skilled practitioners to perform this work and subsequently develop design solutions for such things as decision and planning tools that support expert human cognition¹.

3. Literature Review

Load Forecasting is not a new concept; in fact, it has been the subject of research for a great many years both in academia and the power industry. The methods of approach may differ, but the underlying intent remains the same: to predict the load or power consumption for an asset or group of assets based on some historical data. Current work in forecasting related to Volt-Var Control (VVC) has faced a similar problem of predicting end user load in order to discover the efficacy of various VVC schemes to improve power quality while reducing overall generation needs.

In literature, a number of techniques have been produced some using artificial intelligence²⁻⁶ and others implementing modified forms of regression. ‘Individual Load Forecasting’ in particular appears to be an especially difficult breed of forecasting, due to the random nature of human behavior. A hand full of papers seems to address this problem directly. In one such paper, the author normalized temperatures by determining the median temperature

for weekdays and weekends over the data collection period. The median temperature was used in order to reduce bias due to unexpected temperature shifts. After further normalization and real power corrected data, this information was coupled to traditional models in order to evaluate the performance of the VVC scheme with the predicted load stemming from traditional ZIP models. The output of the combined models achieved 4% MOE for daily load predictions⁷.

Another paper focused on short-term predictions with weather compensation using a combination of Artificial Neural Networks (ANN) and regression. An ANN model was developed after analyzing the temperature and consumption data and identifying statistically significant clusters of hourly and daily temperature profiles. It was identified that these clusters represented similarities in load use, day of the week, and temperature. After this analysis, it was identified that the inputs to the ANN model would consist of day of week, hour of day, temperature cluster for the day, hourly temperature, and VVC scheme status⁸.

In most instances authors were able to leverage this historical data to predict power use with relatively accurate results based on the input data. In addition large external data set were necessary in order to train the models. A key observation in most cases was that some sort of intermediate model was used to drive a secondary model to achieve marginally acceptable accuracy.

4. Forecasting Methods

4.1 Fuzzy Logic

Fuzzy Logic (FL) provides a simple way to arrive at a definite conclusion that is based upon vague, ambiguous, imprecise, noisy, or missing input information. It mimics how a person would make decisions, only much faster. It is used extensively in control systems. Linguistic variables are used to represent an FL system's operating parameters. A good analogy is when we quickly adjust the hot and cold in the shower to get comfortable⁹. FL uses an “If X AND Y THEN Z” approach (rather than solving mathematically). This makes it extremely useful in controlling nonlinear systems that would be difficult or impossible to model mathematically. It can handle large amounts of data and is inherently robust. There are several challenges in using FL; defining the rule-based models quickly become complex when too many inputs and outputs are chosen for a single implementation¹⁰.

4.2 Time series moving average

Time Series Moving Average is a model used to analyze a set of data points by creating a series of averages of different subsets of the full data set. A moving average is not a single number, but it is a set of numbers, each of which is the average of the corresponding subset of a larger set of data points. A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. The threshold between short-term and long-term depends on the application, and the parameters of the moving average will be set accordingly. When used with non-time series data, a moving average simply acts as a generic smoothing operation without any specific connection to time.

4.1.3 Artificial neural networks

Artificial neural networks (ANNs) are mathematical models based upon the functioning of the human brain, and are composed of three different layers of input, hidden and output layers each of which are composed of a certain number of neurons.

ANNs can approximate the best function to a set of data, which is especially important when the functions are complex. Moreover, ANNs are non-linear by nature which means that they can not only correctly estimate non-linear functions, but also extract non-linear elements from the data. ANNs with one or more hidden layers can separate the space in different areas and build different functions for each of them. This means that ANNs have the capacity to build non-linear piecewise models. ANN is considered capable of identifying and treating abrupt changes in a time series pattern.

5. Architecture Considerations

5.1. Architecture goals

Overall, the development of a historical model is critical towards the data analysis functions of the system. Once a baseline model of energy usage is created, the new data input into the system from real time collection may modify the baseline as a result of changing trends in data usage. While the task analysis does not outline the performance of a function or process (external to data processing), the importance of the Input/Output (I/O) analysis is paramount towards an understanding of what the system must do to perform the required actions. The goal of the system is to create a meaningful and accurate model from which to predict future energy usage while correlating this information against the current balance of a customers' prepaid account.

5.2. Task analysis

A critical aspect of an effort of this type, which is essentially a data analysis problem, is the analysis of, "what needs to be done," within the given system. In this case, the ability to forecast energy usage leveraging meter data, both in real time and historical, is central to the analysis of the problem domain and the resultant system. I/O analysis is key towards the successful development of a predictive model. The following analysis must be completed in order to determine the most effective means towards the development of the proposed system:

- Identify the inputs into the system
- Identify the outputs from the system
- Identify the constraints affecting action

5.3. System inputs

For the proposed system, the inputs are consistent between both the historical "training" data, used to create a usage model for each customer (or type of customer) and the predictive data input into the model. The incoming data includes both usage and weather data. Usage data are collected from the various AMI data feeds, in either single aggregate reads per day or interval reads, depending on the type of AMI system employed by the customers' utility company. In addition, weather feeds, from the National Oceanic and Atmospheric Administration (NOAA) or National Weather Service (NWS) are collected to train the historical model or predict future usage. The incoming weather data is considered the most influential factor in daily energy usage.

5.4. System outputs

System outputs include the customer profile, which is used to feed the predictive model for iterative calculations over time which determines the inputs to the predictive model. Also, an output of the system is the future usage profile for the individual customer, which will then assist in determining the effect on the remaining financial balance to which usage is applied.

Ultimately, the output of the full analysis includes a measurement of how long the customer's financial payment will allow the continued use of energy until further funds are needed. This aspect, however, is outside the scope of this development as the goal of current research is to develop the predictive model.

5.5. System constraints

A critical step in model development is the analysis of the constraints resulting from differing data feeds. As identified in the literature review, an increase in data fidelity affects the accuracy of the predictive model. The less granular the data, the less accurate the prediction becomes. This can be identified in both the usage and weather data.

The usage data, as discussed above and further in Section 6, may arrive in a single aggregate value, or in an incremental read of approximately ninety-six individual reads per day. In order to normalize this data for

calculation, several iterations of data processing must be undertaken in order to provide a consistent data feed for the training and predictive models. In this case, a distribution function is applied to the aggregate read in order to synch data between the interval and single aggregate reads.

Weather data also must be pre-processed prior to use within the data models. The weather data, as provided by NOAA, does not necessarily match the read interval provided from the AMI data feed. For the purposes of correlating usage with the current weather data, additional normalization processing must be completed to match usage with the ambient temperature and humidity (for example) at the time the read was taken.

Within the reviewed literature, two approaches towards normalizing the weather data to usage were used. In one case, the median temperature was used to normalize real power usage for model training, while in the other, statistically significant clusters of temperature profiles were used as model inputs.

Based on the constraint analysis, the decision to implement the hybrid model with preprocessing of input data, was realized as a necessary component towards the successful implementation of a predictive end-user model. Ultimately, more than one model is needed in order to complete the objective of the system as a whole; those models including those used to normalize and develop intermediate preprocessing and a predictive model to generate a future usage model for each customer. Further discussion of data preprocessing and the end result is provided in Section 6.

6. Forecasting Model

The first chosen design for implementing the hybrid model approach for both long and short-term forecasts used an absolute customer history response to temperature along with a smaller, more recent, frame of reference. The absolute history is the pairing of a customer's complete usage with the corresponding temperature data so that probabilities (%) of usage per temperature degree could be created. This method allows for a generalization of customer usage determination and serves as an estimate to base future forecasts upon. The short-term history uses a timeframe on the order of weeks to create a similar profile but is not as computationally expensive.

The benefit of this approach is that customer kilowatt (kW) usage and their temperature response is comprehensive while not greatly impacted by short-term or recent behavior and is thusly more stable. However, the statistical weight of this response curve must account for short-term observations on some level. This approach uses the long-term response curve to give a general estimate of forecasted usage in kW along according to their likelihood and extending slightly beyond the highest probabilities. The short-term response curve is then applied upon the previously generated long-term probability curve thereby producing a narrowed usage forecast tolerant of possible recent deviations in consumption.

The smart meter is polled at a constant interval of every fifteen-minute and records the current customer load in kW. The weather data used within the model and retrieved from the NOAA is not consistently sampled throughout the day. These two sets of data, consisting of an unequal number of samples, are therefore difficult to compare when attempting to establish a pattern of consumption. The error is introduced when extrapolating this weather data against that of the energy usage data.

A possible solution to circumvent this problem would be by extrapolating in a piecewise manner to produce the past weather data. For example, this extrapolation function could be split into two chunks on a day and night basis to minimize major 'jumps' in the temperature function, thereby reducing error resulting from estimation of data points. However, the piecewise function would only need to be done once per customer because the model allows for data to be added to their distribution once it has been defined. When the model is ran for the first time, it calculates each of the 365 days in a year. From that point, the model would only consider 365+1, 365+2, (new *n*+1) etc., when making calculations for future predictions. However, this is only one approach of any number of possible solutions. A future systems engineering team could perform an analysis of alternatives on the most appropriate way to extrapolate the weather data across the energy usage data, given these irregular datasets.

In terms of computation, the short-term forecast is not computationally expensive and can be run several times a day with the same, relatively small set of data. The long-term forecast, which incorporates a much larger dataset, will be very computationally expensive considering the number of customers requiring a forecast. If the level of computing power needed to perform these calculations is not available on-site, it may become necessary to

incorporate a cloud computing solution such as Amazon Web Services (AWS), to better meet the computing demand.

Figure 2, shown below, is a SysML internal block diagram view (IBD) of the hybrid model concept. A data collection module handles the processing and insertion of both weather and energy usage data feeds into the

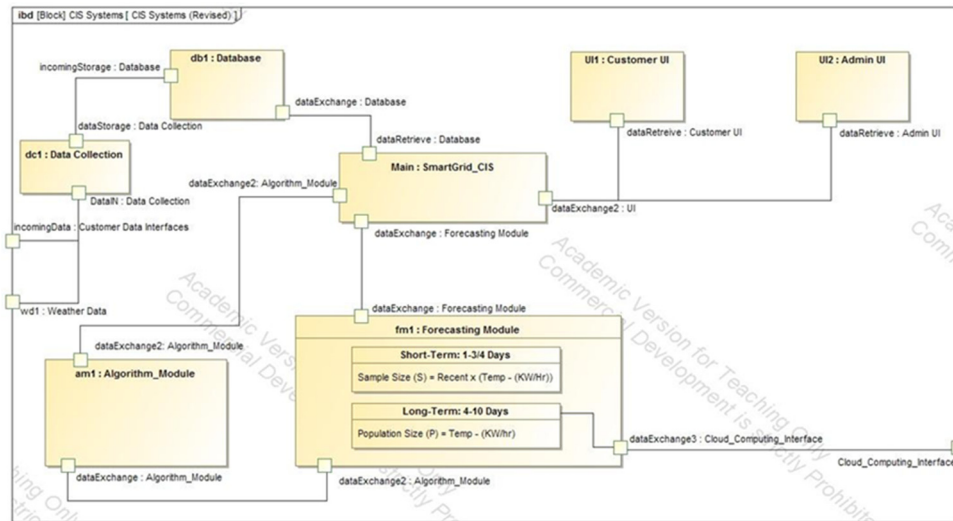


Fig. 2. SysML internal block diagram (IBD) of CIS System

database. The database stores these records by property address, with the main purpose of RPLF, but can also provide larger datasets such as entire cities, counties, or even states, for larger energy use predictions such as determining if a particular area can benefit from a power grid expansion. The SmartGridCIS module is the main method, and ties into the sponsor's existing software that handles customer billing and other functions.

The algorithm module is responsible for handling the irregularity of weather and energy usage data feeds, executes production of the distribution curve, then combines each feed into a single distribution per customer. These distribution curves are stored for use by the forecasting module to perform needed calculations for a RPLF. The forecasting module would include logic to automatically switch between a short-term and long-term approach as necessary, such as in the event of an extreme weather event. Finally, completed forecasts are then stored in the database and made available to the existing administrative and customer user interfaces developed by SmartGridCIS.

Also shown in the forecasting module is a recommended interface to a cloud computing solution, such as AWS, which will become necessary when computing long-term energy forecasts for 10,000+ customers. However, there are many other public, private, and hybrid public/private cloud solutions that should be considered for this project dependent upon the customer, end-user, or marketing of the software suite. If necessary, a future systems engineering team could perform another AoA to determine the most appropriate cloud computing solution for this project.

Future Work

Based on the research finding and the models discussion, to achieve accurate Short Term Load Forecasting (STLF) it is imperative to use a new strategy that involves combination and/or Hybrid Model (CHM). One of the key factors that drive CHM is weather. Weather variables such as temperature, cloud cover, and humidity impacts heating/cooling load whereas visibility and precipitation impacts lighten load. Further research is necessary to determine an optimal model that provides the most accurate prediction based on the location specific weather profiles. This model output could be fed into another model that combines other factors.

Additional follow up work for this project would include the verification and validation of the model and its performance. To verify that the model met the requirements established at the project initiation, several tests should be established and conducted against the model. These test cases would begin with tasks such as simple code inspection to ensure the appropriate variables are included as a step towards algorithm verification. There would also be tests to verify that the model would properly accept all of the data inputs identified in the requirements.

The model's performance would also need to be verified. This would be a measure of the accuracy of the model as called out in the different accuracy bands. In this case, predictions made by the model would be compared to the historical readings for the customers. This would be performed for all prediction and accuracy intervals. These data should also be used to determine the statistical deviations of the model to give a second performance characteristic of the accuracy of the model (i.e. is the model both accurate and consistent). This validation ideally would be incorporated as a performance metric for the model and coded into the model for continual evaluation.

Another area of research for this project would be the consideration of a second iteration of systems engineering processes which defined the derived requirements, the AoA, and the selection of the model. As previously mentioned, additional AoA should be done for determining the most appropriate algorithm for extrapolating the weather data across the energy usage data given the irregular data feeds, and also for selecting the best high performance computing solution. Given that a great deal of additional experience and knowledge would have been gained from the project, confirming the decisions made during the project would add to the confidence of the model's appropriateness and performance.

8. Conclusion

Accurate short-term load forecasting (STLF) plays a critical role in energy systems because it is the essential part of power system planning and operation, and it is also fundamental in many applications. Considering that an individual forecasting model usually cannot work very well for STLF, a possible combination/hybrid model was discussed. Based on the research conducted, weather conditions and data format play a pivotal role in determining the model combination or selection. It is suggested to determine the weather model based on the analysis of historical data for a given location and the format of the incoming data. Other factors such as time of day, season, and historical usage together with system expert inputs will lead to determination of the appropriate combination/hybrid model selection. Accuracy could be validated using historical data and then tuned based upon the current data.

References

1. Lintern, G. *Cognitive Systems Engineering Brief*. Available from: www.cognitivesystemsdesign.net.
2. Dudu, S.M.K.a.S.V., *Short-Term Load Prediction with a Special Emphasis on Weather Compensation using a Novel Committee of Wavelet Recurrent Neural Networks and Regression Methods*. IEEE, 2010.
3. Joao C. Mourdo, A.n.E.R., *Application of Computation Intelligence Techniques for Energy Load and Price Forecast in some States of USA*. IEEE, 2007.
4. K. Y. Lee, Y.T.C.a.J.H.p., *Short-term load forecasting using an artificial neural network*. IEEE Trans. Power Syst., 1992. 7(1): p. 124-132.
5. A. G. Bakritzis, V.P., S.J. Klartzis, M. c. Alexiadis and A.H. Aissis, *A neural network short-term load forecasting model for the Greek power system*. IEEE Trans. Power Syst., 1996. 11(2): p. 858-863.
6. A. D. Papalexopoulos, S.H.a.T.M.P., *An Implementation of a Neural Network Based Load Forecasting Model for the EMS*. IEEE Trans. Power Syst., 1994. 9(4): p. 1956-1962.
7. K. P. Schneider, T.F.W., *Volt-VAR Optimization on American Electric Power Feeders in Northeast Columbus*. IEEE PES, 2012. 1(8): p. 7-10.
8. B. Milosevic, A.V., and K. Mannar, *Substation Day-ahead Automated Volt/VAR Optimization Scheme*. IEEE PES, 2012. 1(5): p. 22-26.
9. *Encoder: Newsletter of the Seattle Robotics Society*. Available from: <http://www.seattlerobotics.org/encoder/mar98/fuz/flindex.html>
10. Padak, A., *Developing a Software to Determine Themicrocontroller Specification for Fuzzy Logic Control Applications*, in *Department of Electrical and Electronic Engineering*2006, University of Çukurova Institute of Natural and Applied Science. p. 73.