Paper 2722-2018

Distribution Circuit Load Forecasting Using Advanced Metering Infrastructure Data

Prasenjit Shil, Ph.D., Ameren; Tom Anderson, SAS Institute Inc.;
Mark Konya, P.E., SAS Institute Inc.

ABSTRACT

The ability to perform very short to very long-term forecasts of distribution circuit loads at intermediate distribution circuit locations between customer meters and substation feeder buses using AMI (Advanced Metering Infrastructure) data provides significant advantages to distribution system planners and operators in a number of areas. Some of the important applications of these forecasts include anticipation of device overloads, facilitation of switching operations, and helping with DER (Distributed Energy Resources) integration into system operations. This session presents forecasting results for distribution circuits using SAS® Energy Forecasting, which uses methods such as GLM to generate forecasts. Results for different circuit locations are derived from Ameren Illinois AMI and circuit taxonomy data. Included in the presentation are details of the forecasting methodology and a discussion of applications to distribution system operations.

INTRODUCTION

Electric power transmission and distribution systems are designed to transport electric energy from generating units to end-use customers. These systems consist of components such as cables, switches, transformers, insulators, capacitors, reclosers, and other equipment designed and connected to accomplish this objective. Figure 1 provides a conceptual representation of this system and its major components. This session focuses on the distribution portion of the system, but with sufficient data the methodologies described herein can be extended to include the transmission portion of the system.

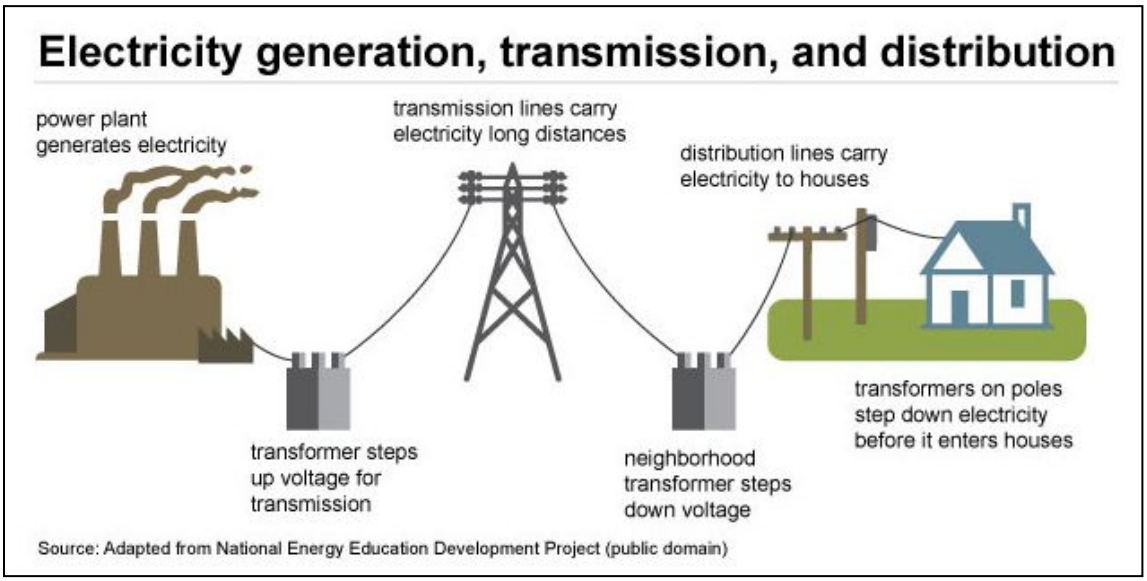


Figure 1: Electric Power Supply System

Time series forecasting of electric power distribution measures such as power (KW) and energy (KWH) is enabled by forecasting software such as SAS® Energy Forecasting, a solution which utilizes well-established mathematical techniques for predicting future values based upon the historical structure of the series. Forecasting methods such as General Linear Models (GLM), Auto-Regressive Integrated Moving Average (ARIMA) and others are utilized to generate short- to long-term forecasts, which can help distribution planning, and operations departments maintain a high level of reliability while controlling costs.

Prior to the installation of advanced metering infrastructure (including net meters), due to the unavailability of end-point use data, utilities did not have sufficient data to produce a forecast of electric power and energy at the distribution sub-circuit level.

With the installation of advanced metering systems, end-use time series data is now widely available to utilities. This data includes, but is not limited to, real and reactive power, voltage, current and various event flags which indicate, for example, loss of voltage and reverse rotation. Typically collected at 15 minute intervals, meter data can now be aggregated by time interval, transformer, device, phase, and circuit to use as input for forecasting total load at intermediate circuit nodes between meter points and substation feeder buses. Figure 2 provides a conceptual overview of a distribution circuit and the information required to generate forecasts from aggregated AMI data.

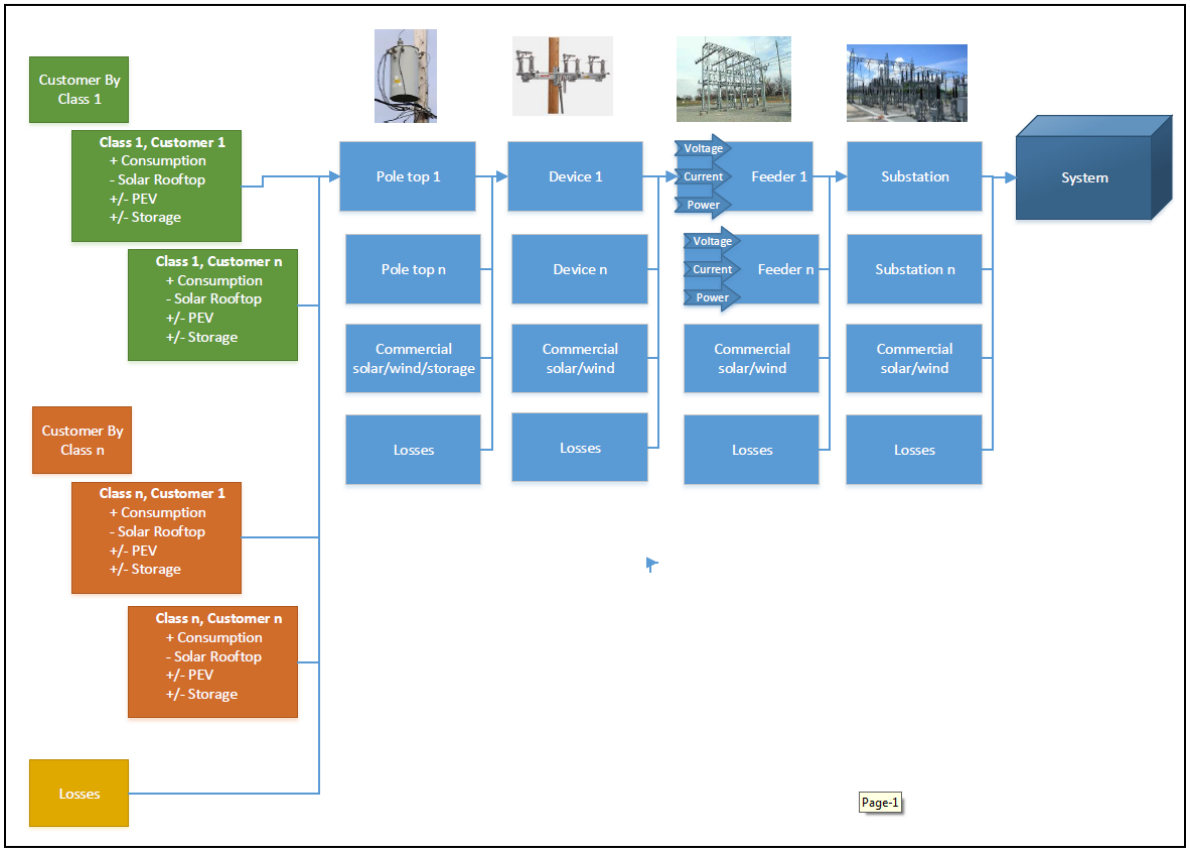


Figure 2: Overview of Information Flow

There are many uses for distribution circuit forecasts based on AMI data. One use would be to anticipate overloading of devices, substation transformers, and conductors when circuits are in their normal

configurations. By forecasting potential overloads with sufficient lead times system operators could initiate relief actions to prevent system degradation and unplanned customer interruptions. In this case, sufficient lead times can vary from very short-term (day-ahead) to very long-term (decades-ahead).

Another use would be to predict future loads of circuit sections whenever a change of circuit state is anticipated. For example, if a circuit section requires de-energization for maintenance, then customers behind the sectionalizing device could be transferred to an available alternate supply provided that the transferred customers do not overload the alternate supply during the period that maintenance is being performed on the primary supply. This overload assessment would be accomplished with very short-term load forecasts for the sectionalized circuit and alternate supply source since the two series can be added together to verify an overload would not occur.

It is important to note that SAS® Energy Forecasting enables accurate, production volume forecasting for all forecast horizons. Forecast results can also be transferred to distribution load-flow models for detailed assessment of load impact on circuit design and protection parameters. Furthermore, SAS® Energy Forecasting has the capability to reconcile hierarchal load forecasts based on AMI data with load forecasts at the feeder level based on SCADA measurements. This reconciliation can offer an alternate approach to estimating distribution system losses.

This paper illustrates application of SAS® Energy Forecasting to derive very short-term forecasts of load on selected devices and feeders in the Ameren Illinois distribution system. Based on two years of AMI data collected at hourly intervals, forecast results will provide system operators with insights into daily operations.

DATA PREPARATION AND STAGING

Prior to building the forecast models it is necessary to subject the raw data set to a process which identifies and replaces missing values and outliers. It is also necessary to add predictive variables to improve the forecast compared to a naïve model, then structure the data for analysis.

Figure 3 provides an overview of the data cleansing process flow. Each of the steps depicted are described in more detail below.

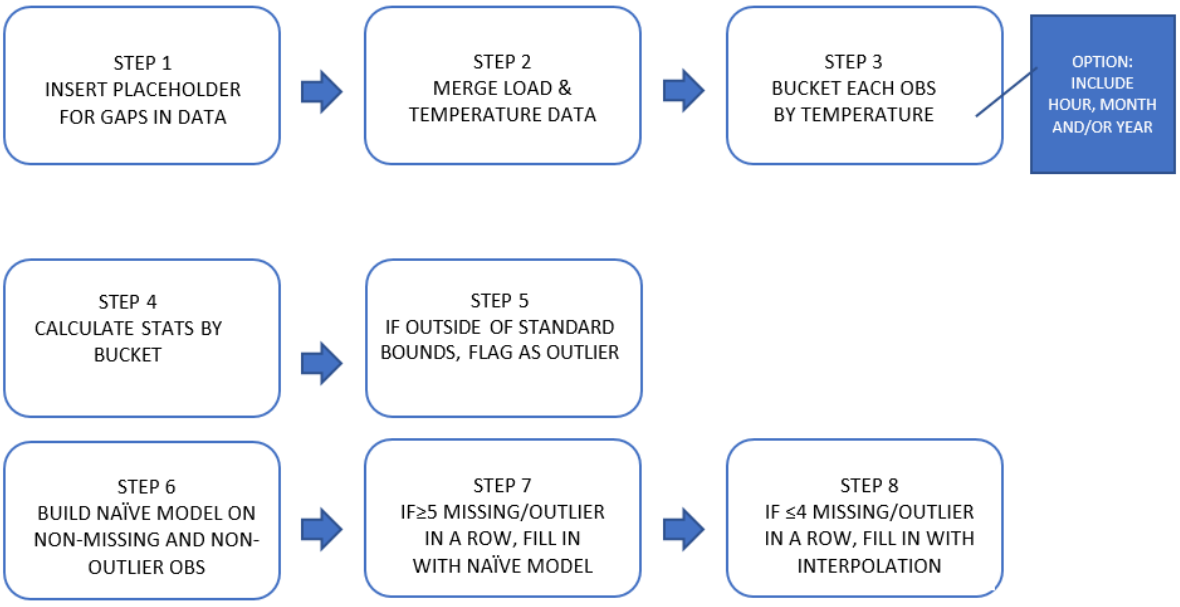


Figure 3: Data Cleansing Process Flow

DATA CLEANSING STEP 1

It is a requirement for the modeling process that the load data has a continuous time series. In step 1, SAS® uses PROC TIMESERIES to insert any missing date-time observations. The corresponding load value is left missing and this observation is marked with an indicator for a later step to use an interpolated value. This step also has the ability to insert any records at the beginning or end of the time series. This is useful if the user wants to extend the time series. The user would specify the start date and end date they would want to use. If no dates are specified, the data date range is used as is it currently exists.

DATA CLEANSING STEP 2

This step merges the load, weather, economics, and user defined variables to create one table. This table is then used for future steps for identifying outliers/missing values and for building the model to interpolate replacement values.

Weather data is especially important for identifying outliers because load will vary significantly when the temperature changes. A range of load values at 60 ^{o} F looks very different than a range of load values at 0 ^{o} F. Because of this, what an outlier is at those ranges will also vary significantly. Using weather data as a main indicator we can identify outliers properly.

Other variables such as economic indicators or 'extra' weather variables can be used as part of step 6 when building a model to interpolate replacement values for the outliers. This step also adds in the corresponding SCADA or AMI data (an option the user has). For a SCADA circuit a column is added to the table representing the AMI load data that matches the SCADA circuit. Similarly, if AMI data is available then corresponding SCADA data is added to the table.

DATA CLEANSING STEP 3

Using the table built in step 2, Step 3 "buckets" each observation based on temperature (and, optionally, month, hour, and/or year). This methodology looks at how load values which fall within a certain temperature range, as defined by a configurable 'width' value, relate to one another. A temperature value is required for this methodology.

Optionally, the analyst can also further subset the temperature bucket by month, hour, and or year. However, it is not recommended to bucket by all additional options because this could lead to very sparse data and may not give the desired results in outlier identification. An example 'bucket' would be all load values that have a temperature between 68 ^{o} F and 70 ^{o} F in August. This 'bucket' would have a width value of 1 and would be further subset by month.

DATA CLEANSING STEP 4

Based on the "buckets" created in step 3, step 4 takes the observations for each bucket and calculates descriptive statistics using PROC MEANS including mean, median, min, max, standard deviation, variance and percentiles. For cleansing purposes, the median and standard deviation are used. After generating the descriptive statistics for each bucket, the data cleansing process defines the boundaries for outliers by bucket.

Figure 4 below is an illustration of uncleansed data. The observations marked in red have been flagged as outliers by the process.

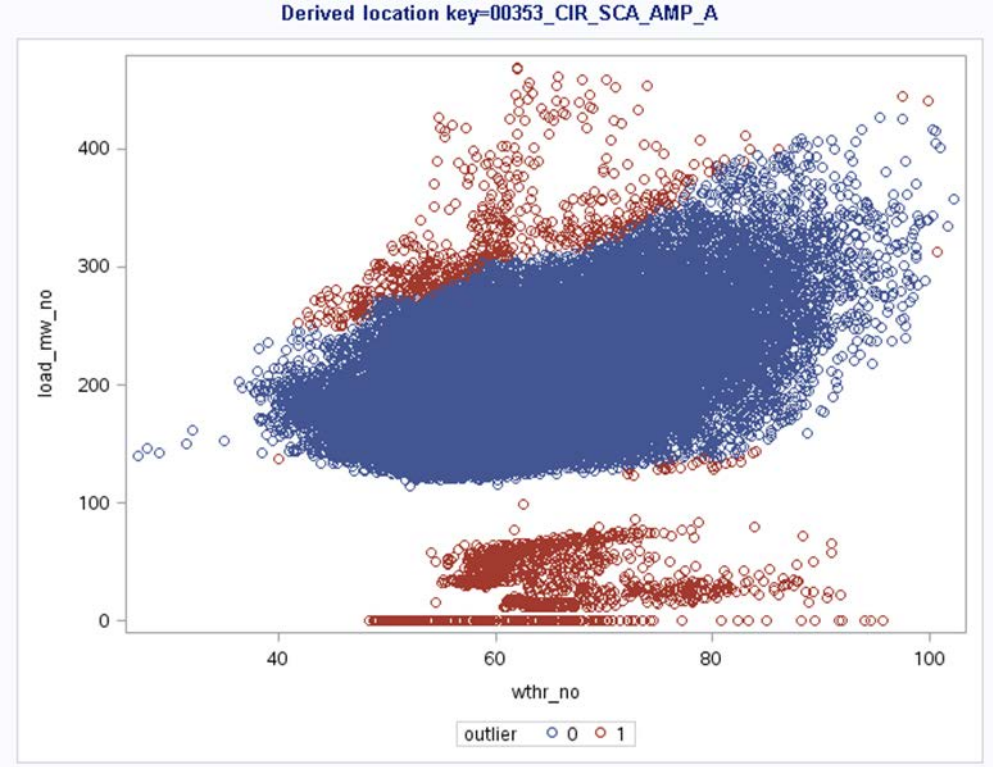


Figure 4: Uncleansed Data with Flagged Outliers

DATA CLEANSING STEP 5

After calculating the bounds for each bucket in step 4, step 5 compares each observation to the bounds of its bucket. If the load value exceeds the bounds then the value is flagged as an outlier.

DATA CLEANSING STEP 6

In this step a model is built to identify potential replacement values.

Using the SAS® procedure PROC GLM, a model is built using the data not previously flagged as missing or as an outlier. SCADA/AMI data can be used (if specified) to help drive the shape of the model. If modeling SCADA then the variable would consist of AMI data and vice versa.

Additional input variables could include, but are not limited to economic indicators, tariff class, and additional weather concepts. (The data must exist in the data set to be used). These additional variables can help improve the model fit and the accuracy of the interpolation of the load values.

Temperature is a configurable variable in that the user can decide which weather concept to use (e.g. they could use dry bulb temperature or wet bulb temperature assuming they have the data to support it).

The model estimates load based on a trend variable, the interaction of weekday and hour ending, month, interaction of month and temperature, interaction of hour and temperature, any user defined input variables (optional) and SCADA/AMI (optional).

To build and train the model, the non-outlier and non-missing data is used and the parameters are saved. Once the model is trained, then the model is applied to the 'bad' data (missing values and flagged outliers) to calculate load value interpolations. These values are added into the dataset and the process moves to steps 7 and 8.

DATA CLEANSING STEPS 7 AND 8

After Step 6 is completed, Steps 7 & 8 suggest replacement values for the flagged outlier and missing data to be used through the rest of the ETL process.

The process determines how many flagged outlier/missing data values occur consecutively over the entire dataset. (For example the dataset could contain 10 good values followed by 1 flagged values followed by 1000 good values followed by 6 flagged values)

For a series of five or more missing or outlier observations, the suggested value will be taken from the naïve model. For a series of four or fewer missing or outlier observations, the suggested value will be a linear interpolation between the two non-missing and non-outlier values at the beginning and end of the missing/outlier series. The SAS® procedure used for the linear interpolation is PROC EXPAND. The cutoff between the two methodologies is a configurable option.

The final output from the cleansing process is a data set which includes all days and hours within the historical date range. The original load value will appear in one column. Another column will show the suggested value. The outlier/missing flag variable is also retained. This table may be manually edited if the analyst desires.

After the analyst is satisfied with the cleansed data it is passed back to a data integration process which drops the original values and outlier/missing flag, keeping only the recommended value for each date/time.

Figure 5 illustrates the cleansed data. Observations in red indicate values, which have been altered.

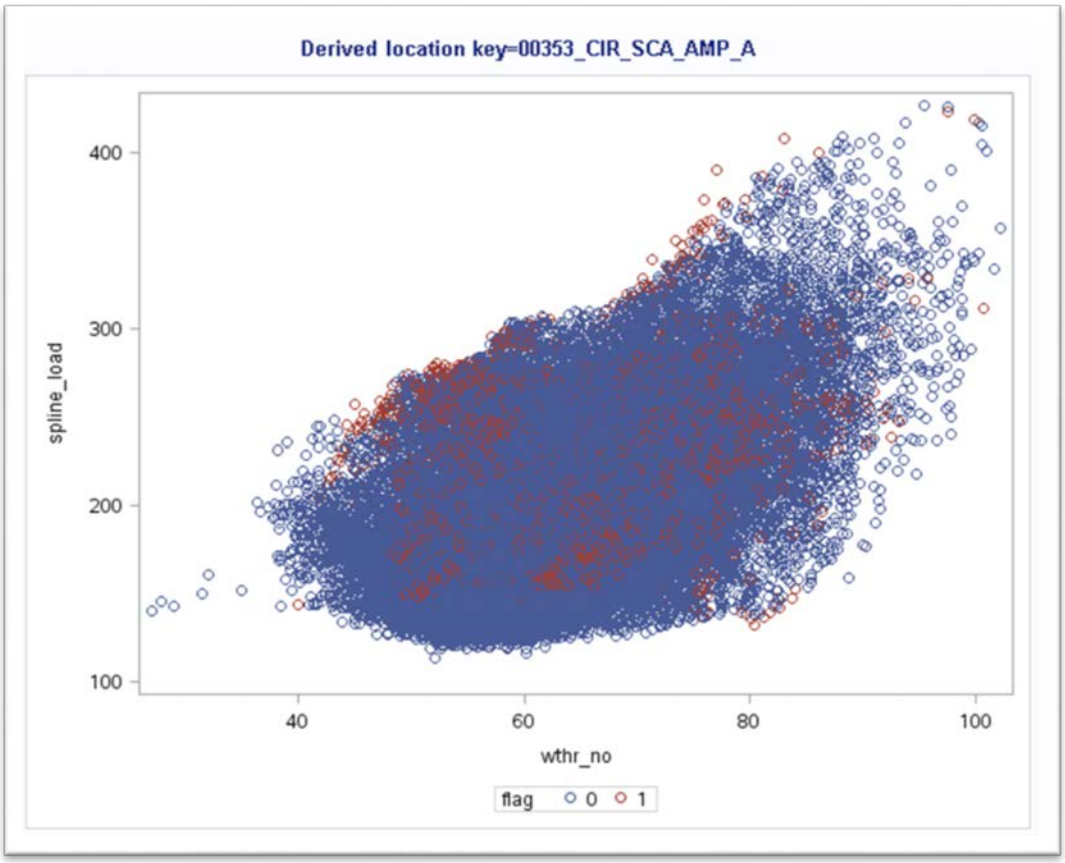


Figure 5: Cleansed Data

DATA MANAGEMENT STRATEGY

Data preparation for energy forecasting requires a landing area for raw input data for the forecasting process, a staging area for the final input data for energy forecasting, and the staging load process for transforming the data from the landing area to the staging area.

The landing data is the collection of raw data tables, extracted from operational data systems, third party vendors, and other sources, that serve as the raw input for the forecasting process. This data at this stage will generally not be suitable for forecasting due to the specific organization of the data, data quality issues, and data consistency issues.

The staging data is a collection of tables in a standardized format that serve as the final input to the forecasting process. To the extent possible, all data quality issues have been addressed in this data store: missing data filled in, outliers detected and replaced, inconsistencies across tables detected and resolved, summaries of data throughout the forecasting hierarchy generated. Any energy variables that are derived from measures in the raw data are created.

The staging load process automates the task of transforming landing data tables into staging data tables. While this is an automated process, it is derived in part from the results of the forecasting analyst's and data integration analyst's exploratory data analysis, a manual and iterative process to detect data quality issues and determine the best approaches to resolve them. Once these approaches are determined, they are incorporated into the staging load process for automatically generating staging tables from landing tables. The process is designed to handle automatic updates of data, which can be based on any of several standard approaches, depending on data volumes and other considerations.

ANALYSIS

SAS® Energy Forecasting supports the forecasting process from beginning to end, using AMI and SCADA data-driven analytical insights to assess future loading of grid assets for all forecast horizons – from very short-term day-ahead to very long-term decades ahead. With best-in-class analytics, analysts are empowered to answer the most complex load forecasting questions that arise from utilities' distribution systems.

After the Ameren Illinois AMI data was cleansed and aggregated by time stamp and phase to the feeder level, SAS® Energy Forecasting was used to create models characterizing the training data set for each phase. During this process multiple different models were automatically tested by the software and the champion model was selected based upon the mean absolute percentage error (MAPE).

The steps followed to generate the Ameren Illinois feeder load forecasts included:

1. Determining the best historic relationship between load and the factors that influence load – primarily temperature. This could be called fitting a model or creating a forecast equation. That model or equation is coupled with a forecast of temperature, and possibly other factors as an economic index or population, to create a forecast of load.
2. Splitting the history into two pieces – a longer one to create a model and a shorter recent period to test the model, often referred to as a training period and a holdout period. This lets the program select the best model and use a second stage model to tailor the model to most recent history.
3. Stepping through a sequence which creates many different models by successively adding new factors and testing that accuracy is improved. This DIAGNOSE process actually tested multiple different model combinations and tested against history to find a subset of best forecasts. Some of those models plotted here.
4. Selecting the champion model based on any one of over 20 available error measurements. For this work, mean absolute percentage error (MAPE) was used to determine the winner.

DEPLOYMENT OF RESULTS

The foundation of load forecasting work started with constructing multiple regression models where various independent variables and combinations of independent variables are tested sequentially for model improvement. Models are tested at each iteration to prevent over-fitting. Second-stage models are developed using UCM (Unobserved Components Model), ARIMAX (Autoregressive Integrated Moving Average with Explanatory Variable), exponential smoothing and neural network model. The software also allows the analyst to create forecast by combining various models to improve MAPE. For this work, we chose the default level of automation for the forecasting process. Model results and statistics are available at each step of the process and outlier files are automatically generated and can be reviewed and visualized within the application. Figure 6 shows error matrix for all the tested models (MAPEs).

In this paper, forecasting models were constructed at the system level (all five feeders combined) as well as at individual feeder level. In the table view of FCST_STAT, records are sorted by the type of error statistic that is used to judge the best or champion model. The champion model has the lowest error. The error matrix shown in the FCST_STAT contains the MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and ME (Mean Error) for annual energy, annual peak load, daily energy, daily peak load, monthly energy, monthly peak load, and hourly load. Based on Hourly MAPE comparison, GLM_ARIMAX emerged to be the best performing model for the dataset.

Model Name	Hourly ME	Hourly MAE	Hourly MAPE	Daily Peak ME	Daily Peak MAE	Daily Peak MAPE	Daily ME	Daily MAE	Daily MAPE	Monthly Peak ...	Monthly Peak ...	Monthly Peak ...
GLM_ARIMAX	-19.31135766	606.02968788	2.893571256%	93.361617403	764.5248268	3.165114349%	-405.5385108	7195.7126864	1.645211747%	43.916117622	359.77670266	1.442639977%
COMBINE	-237.2922345	721.52218976	3.509325473%	386.06413714	940.80782999	3.971553451%	-5695.013628	12127.765742	2.514035804%	807.48434359	807.48434359	3.127757418%
GLM_UCM	-108.2280987	842.35828086	4.085904891%	154.90763938	939.80112005	3.979325725%	-2272.790073	13250.040116	2.883610203%	264.4424547	432.14833823	1.699957032%
GLM_ESM	-130.6768491	866.59520841	4.214259154%	198.59945982	983.64128446	4.190589089%	-2744.213831	13723.034629	2.995879281%	264.0744536	432.14833823	1.700014190%
GLM_NN	-690.9526326	897.02451984	4.270904697%	1053.0698546	1162.6076193	4.854950159%	-14510.00528	16803.276316	3.845120162%	1169.9768353	1169.9768353	4.672658980%
GLMLFBRW	-412.8846263	862.10068503	4.294901064%	494.83893376	862.16806082	3.659934873%	-9909.231032	16605.686634	3.453956693%	567.53007575	567.53007575	2.198304423%
GLMLFBRWH	-412.8846263	862.10068503	4.294901064%	494.83893376	862.16806082	3.659934873%	-9909.231032	16605.686634	3.453956693%	567.53007575	567.53007575	2.198304423%

Figure 6: Comparing Model Stats from Various Models

Figure 7 compares output from the various forecasts. The hierarchy is shown on the left with the instances of each forecast described on the bottom left.

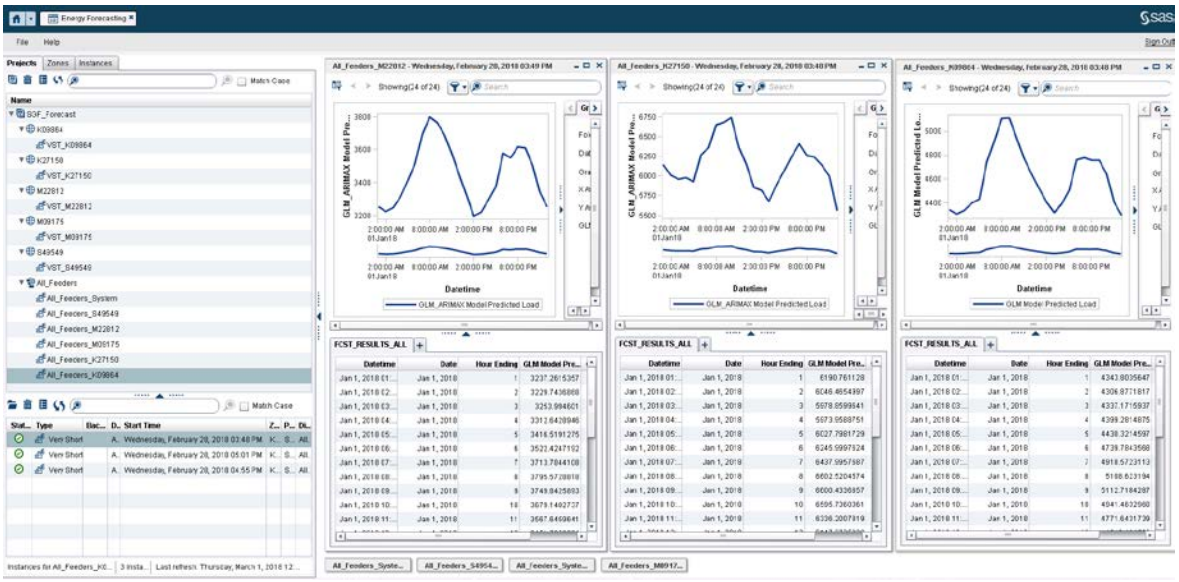


Figure 7: Comparing Forecast Results

The SAS Energy Forecasting software also provides ability to compare individual model results. Figure 8 compares GLM models and their respective MAPE values based on hourly, daily, monthly and annual periods. This is extremely beneficial in determining the best model and provide direction for improvement. In addition, SAS Energy Forecasting will identify hours (or observations) with the highest error statistics per model. This allows identification of data errors and model specification problems by reviewing error statistics, load, predicted load, and temperature for each of those hours. These observations are captured in the Results OUTLIER datasets. The APE Cutoff determines which observations are put into the outlier table. Observations with absolute percentage error greater than this threshold will be put into the table (Figure 9).

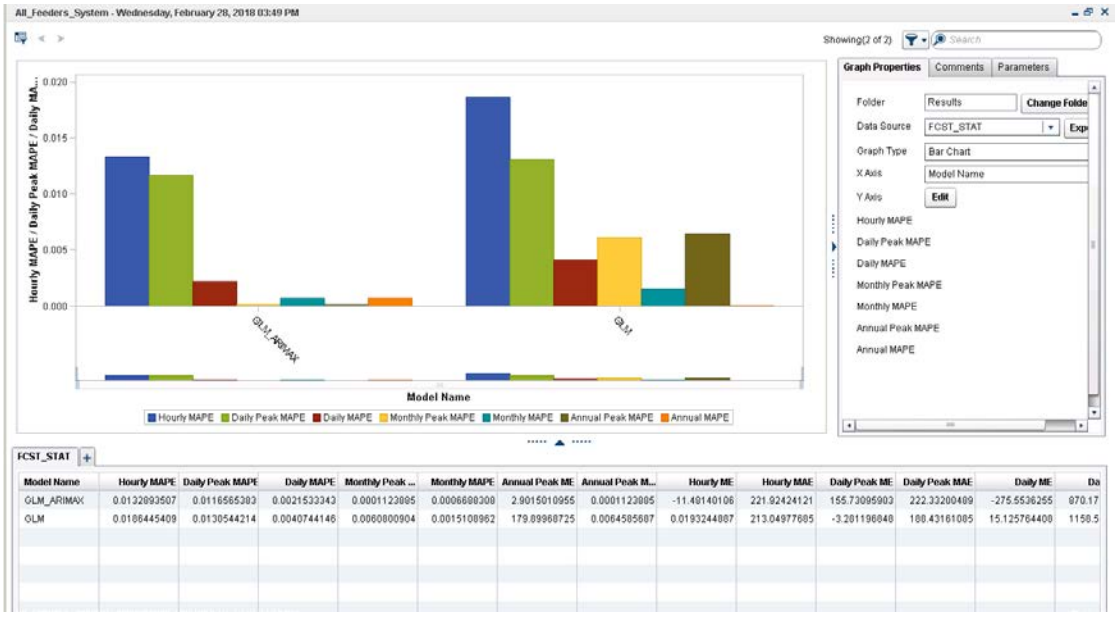


Figure 8: Comparing GLM models

The forecaster has the opportunity to view outliers in the forecast and view the specific data that may have driven the outliers so that adjustments can be made. For each model, the hours with the highest error statistics are identified. This allows identification of data errors and model specification problems by reviewing error statistics, load, predicted load, and temperature for each of those hours. The data in the following sample outlier table is based on the GLM model. It provides the actual values and the predicted load along with the error, absolute error and absolute error percentage.

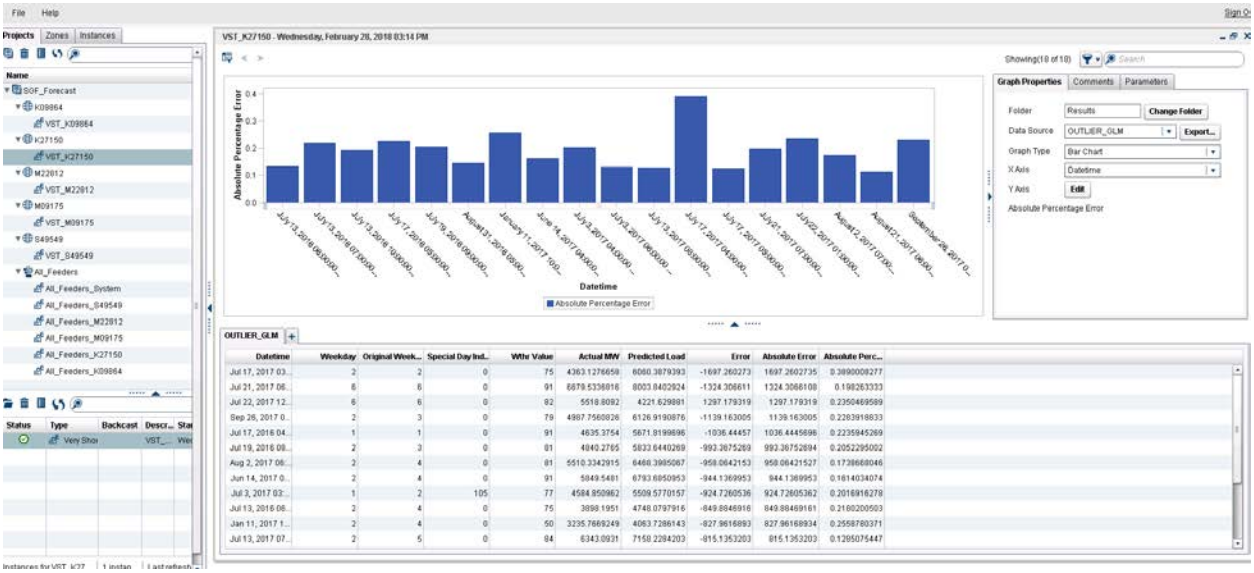


Figure 9: Outlier detection based on APE

The SAS Energy Forecasting provides the ability to enter the percentage of observations that are to be considered as outliers. The default value is 0.001, which means the generated outlier table includes the top 0.1 percent of all observations with the largest forecast absolute error or, AE (Figure 10).

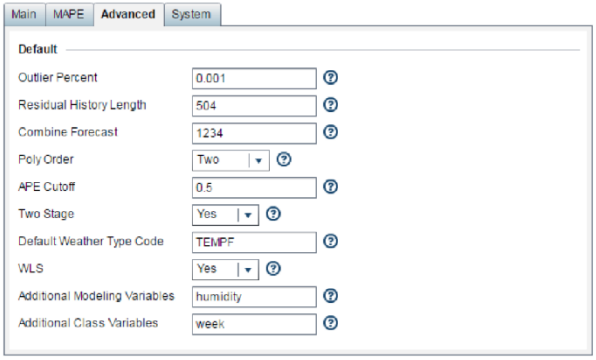
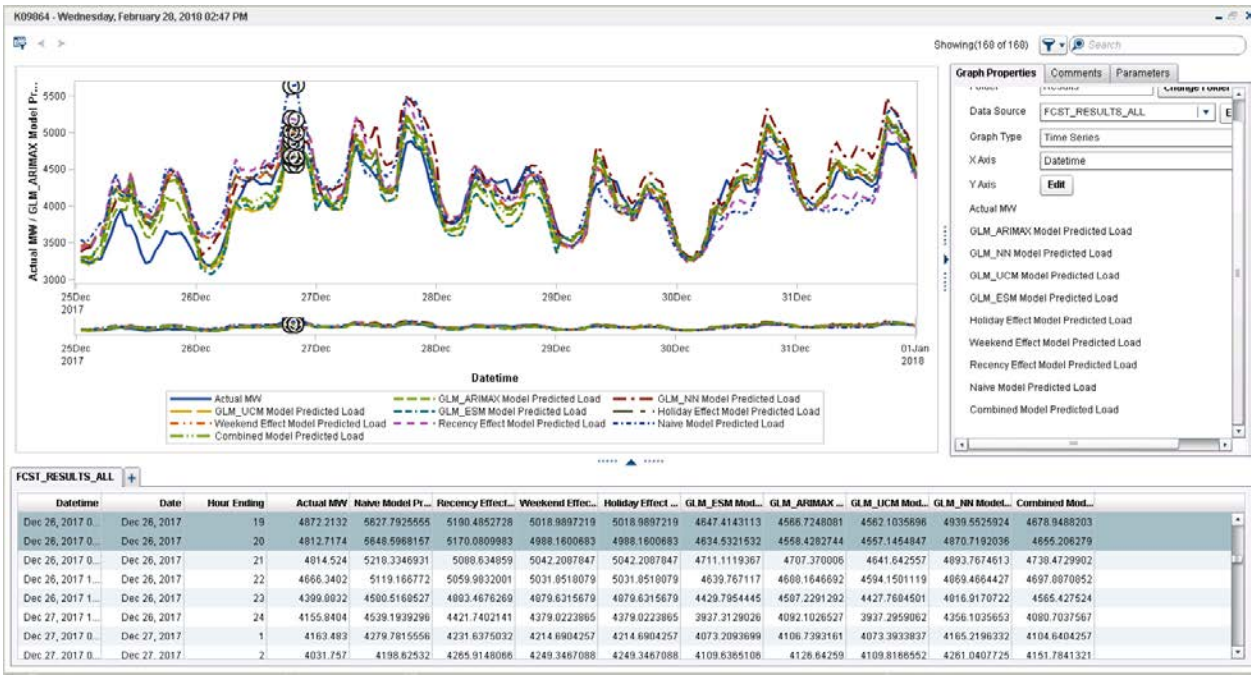


Figure 10: Advanced Parameters

The forecast results for all the models generated are provided. The winning ("champion") model is the model selected as best by diagnose. Figure 11 show the very short term forecast results from 10 models produced from the project.



Datetime	Date	Hour Ending	Actual MW	Naive Model Pr...	Recency Effect...	Weekend Effec...	Holiday Effect ...	GLM_ESM Mod...	GLM_ARIMAX ...	GLM_UCM Mod...	GLM_NN Model...	Combined Mod...
Dec 26, 2017 0...	Dec 26, 2017	19	4872.2132	5627.7925555	5190.4852728	5018.9897219	5018.9897219	4847.4143113	4566.7248081	4562.1035696	4939.5525924	4678.9488203
Dec 26, 2017 0...	Dec 26, 2017	20	4812.7174	5648.5968157	5170.0809983	4988.1600683	4988.1600683	4834.5321532	4558.4282744	4557.1454847	4870.7192036	4655.206279
Dec 26, 2017 0...	Dec 26, 2017	21	4814.524	5218.3346931	5088.634859	5042.2087847	5042.2087847	4711.1119367	4707.370006	4641.642557	4893.7674613	4738.4729902
Dec 26, 2017 1...	Dec 26, 2017	22	4666.3402	5119.166772	5059.9832001	5031.8518079	5031.8518079	4639.767117	4688.1646692	4594.1501119	4869.4664427	4697.0870852
Dec 26, 2017 1...	Dec 26, 2017	23	4399.8032	4580.5160527	4983.4676269	4979.6315679	4979.6315679	4429.7954445	4597.2291292	4427.7604501	4816.9170722	4565.427524
Dec 27, 2017 1...	Dec 26, 2017	24	4155.8404	4539.1939298	4421.7402141	4379.0223865	4379.0223865	3937.3129026	4092.1026527	3937.2959062	4356.1035653	4080.7037567
Dec 27, 2017 0...	Dec 27, 2017	1	4163.483	4279.7815556	4231.6375032	4214.6904257	4214.6904257	4073.2093699	4106.7393161	4073.3933837	4165.2196332	4104.6404257
Dec 27, 2017 0...	Dec 27, 2017	2	4031.757	4198.62532	4265.9148066	4249.3467088	4249.3467088	4109.6385106	4126.64259	4109.8186552	4261.0407725	4151.7841321

Figure 11: Comparing All Models

CONCLUSION

SAS® Energy Forecasting was applied to Ameren Illinois AMI data to generate very short-term forecasts of load on distribution circuits based on usage profiles of end-point customers. An automated process was developed to cleanse and stage the data, thereby enabling creation of accurate, high volume forecasts which can be readily visualized to help planning engineers and system operators maintain a high level of service reliability for customers while concurrently managing costs. A number of applications for these forecasts were identified, including long-term system design planning and short-term assessment of day-ahead circuit performance. These forecasts generate significant value to utilities by helping avoid customer interruptions, reducing the duration of interruptions, improving long-term planning, and offering a foundation for Integrated Distribution Planning.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the substantial contributions of others without whom this paper would not be possible. Those contributors included SAS employees Emily Forney, Sen-Hao Lai, Glenn Good, Tae Yoon Lee, Bradley Lawson, and Jennifer Whaley.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Prasenjit Shil, Ph. D.
Ameren Corporation
pshil@ameren.com
www.ameren.com

Tom Anderson,
SAS® Commercial Presales
tom.anderson@sas.com
www.sas.com

Mark J. Konya, P.E.
SAS® Global Utilities
mark.konya@sas.com
www.sas.com