Paper 11360-2016

Breakthroughs at Old Dominion Electric Cooperative with Energy Load Forecasting Innovation

David Hamilton, ODEC; Emily Forney, SAS Institute Inc.; Steve Becker, SAS Institute Inc.

ABSTRACT

As the electrical grid has become more complex, utilities are revisiting their approaches, methods, and technology to accurately predict energy demands across all time horizons in a timely manner. With the advanced analytics of SAS® Energy Forecasting, ODEC provides data-driven load predictions from next hour to next year and beyond. Accurate intraday forecasts mean meeting daily peak demands saving millions of dollars at critical seasons and events. Mid-term forecasts provide a baseline to the cooperative and its members to accurately anticipate regional growth and customer needs in addition to signaling power marketers where, when and how much to hedge future energy purchases to meet weather-driven demands. Long-term forecasts create defensible numbers for large capital expenditures such as generation and transmission projects. Much of the data for determining load comes from disparate systems such as SCADA and internal billing systems combined with external market data (PJM), weather, and economic data. This data needs to be analyzed, validated, shaped, and conformed to fully leverage predictive methods. Business insights and planning metrics are achieved when flexible data integration capabilities are combined with advanced analytics and visualization. These increased computing demands at ODEC are being achieved by leveraging Amazon Web Services (AWS) for expanded business discovery and operational capacity. Flexible and scalable data and discovery environments allow ODEC analysts to efficiently develop and test models which are IO intensive. SAS visualization for the analyst is a graphic compute environment for information sharing that is memory intensive. Also, ODEC IT operations require deployment options tuned for process optimization to meet service level agreements which can be quickly evaluated, tested and promoted into production. What was once very difficult for most utilities to embrace is now achievable with new approaches, methods, and technology like never before.

INTRODUCTION

Innovations in Data Management and Forecasting

ODEC has initiated its strategy and planning to develop automation, advanced custom code and technical sequencing which will streamline operations in future years. These significant changes and improvements allow us to maximize our investment in SAS applications, exceed our business needs, and enhance stakeholder's expectations. ODEC continues to maintain and improve our current Mid-Term and Long-Term forecasting operations with SAS® Forecast Server and SAS® Forecast Studio. Our longer term plans include more automation (stored processes) of data management, data evaluation and data filtering along with increased use of data-marts to maintain our model inputs allowing us to build flexibility and accuracy.

The evolution of SAS® Energy Forecasting and SAS® Data Integration is providing us with the needed advanced applications to tackle very short term and short term forecasting in a real-time, full automation platform. The Regional Transmission Organization (RTO) project is one of several projects that should leverage the progress achieved through SAS based energy forecasting while expanding our knowledge and capability to tackle future operational challenges. ODEC concurs that the future of Data Management, Analytics, Forecasting, & Reporting solutions will gravitate to the cloud environment. ODEC looks forward with eagerness as SAS develops cloud based application platforms which opens new worlds of "Forecasting Anywhere, Anytime" realities for our business.

Data Management Innovation

Any successful energy forecasting operation begins and ends with proper data management. Databases and Data Warehouses must be properly constructed to house and feed the flow of required inputs that form the backbone of successful models used to predict future consumer behavior. At ODEC we are using SAS applications such as SAS® Enterprise Guide & SAS® Data Integration along with custom SAS code to help us manage, investigate, identify, verify and structure our source data from which our eventual Data marts are populated. Automated scheduling, retrieval, processing & validation are vital to our goals of constructing a fully software controlled environment platform feeding our diverse and comprehensive forecast modeling engines. ODEC has accepted and embraced the concepts that human intelligence can identify and set the parameters for a successful forecasting platform while allowing software applications to quickly implement them. Automated data management and the highway of connections to our forecast modeling applications creates a highly optimized process framework with which to initiate our modeling operations.

Forecast Modeling Innovation

Over the last decade, forecast modeling has dramatically changed as the "Great Recession" ravaged the historical models and practices of energy companies. The reasons for the failures point at utilities being slow to adapt to changing consumer behavior and the failure of economic forecast firms to implement a recession scenario in their forecasts. It has become apparent to forecasting practitioners' that multiple model approaches, combination models, and a fully diversified set of independent driver variables including varying scenarios must be developed and incorporated in every forecasters' toolkit. ODEC has been using a diversified set of SAS forecasting applications over the last 8 years and is implementing another SAS forecasting application SAS® Energy Forecasting this year in a fully automated data feed, model selection and, model promotion to production process platform server. ODEC continues to deploy SAS® Enterprise Guide and SAS high performance forecasting multiple model/combination model approaches to mid-term and long-term forecasts on its Forecast Server platform using SAS® Forecast Studio which allows for complete and sustained model testing, analysis, scoring, selection, and implementation. New model forms allow the forecaster to examine and adjust quickly to major changes in the economy and consumer behavior which drives energy sales and peak demand.

SHORT-TERM FORECASTING MODERNIZATION PROJECT FOR RTO

The Scope – ODEC's development for the Regional Transmission Organization (RTO) forecasting development project.

The Goal - To provide ODEC the capabilities using SAS® Energy Forecasting to support the RTO forecasting process by implementing SAS® Energy Forecasting models and automation to replace current manual methods of reporting.

The History - The repeated manual manipulation of data is time consuming and eats into the time needed by ODEC to analyze, recognize and report on the electricity usage characteristics of customers in each of ODEC's four operating zones in addition to PJM:

Allegheny Power System (APS)
Delmarva Power and Light (DPL)
Dominion Resources (DOM)
American Electric and Power (AEP)

The result of our analysis is prepared in report form and distributed at least daily and more often when needed. It is made available to ODEC Members, ODEC contractors, Large Retail Commercial, and Industrial Customers for their use in the determination of the practicality of taking advantage of Direct Load Control of Member's water heaters and air conditioning equipment and voluntary Load Management activities to minimize costs to ODEC and its members during specific periods of time.

The management of peak-time usage during specific periods of time is encouraged to minimize the cost to retail customers and the member Cooperatives when energy usage reaches a peak level and peak prices. Peak time levels are determined by the amount of energy usage for specific hours in the zone in which each Cooperative is located. Peak times are heavily impacted by local weather conditions.
In addition, for our retail customers that participate in the PJM load management program, this report offers early warning indications that PJM might declare an "emergency load management event."
Below are three fundamental costs for which load management efforts are encouraged. There may be other times when the Cooperative asks members to manage their electricity usage, such as a time when the reliability of the system is in jeopardy or during periods of very high power costs. In any case, the Cooperative will issue an advisory some time before the important times are expected to happen. Early indication of a pending peak in any zone has a high value to participants.

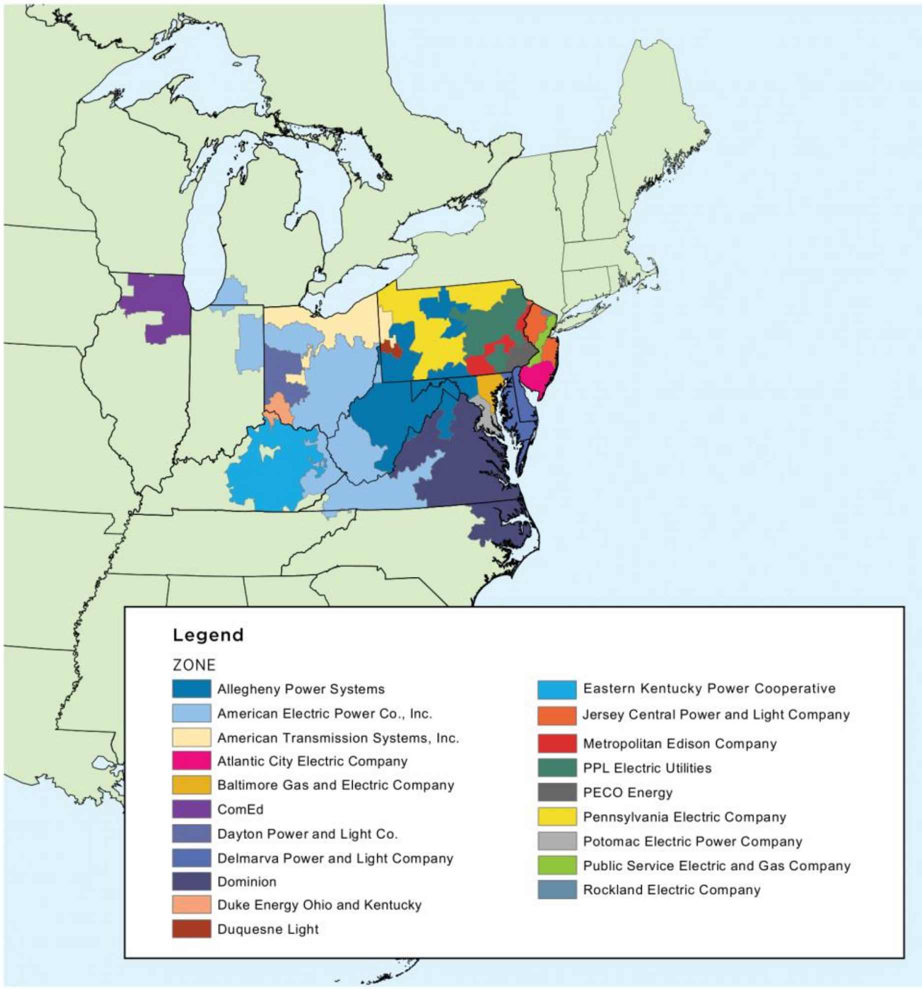
1. Network Service Peak Load (1CP) is a cost to "rent" the high voltage transmission wires over which energy flows to all customers. It is based on the single highest one-hour usage for the Cooperative over the period starting November 1st and continuing through October 31st of each year. All attempts to reduce usage/load at that time, which is identified in Panel #4, will lower the cost to all customers of the Cooperative.

2. Peak Load Allocation (5CP) is a cost based on the requirement to maintain the capability to provide energy to all customers. It is based on the five highest system peak loads occurring at peak-time usage which is during the summer period of June 1st through September 30th. When a new high-five peak is expected to occur, attempts to reduce usage/load during that hour will lower the cost to all customers of the Cooperative.

3. Monthly Billing Peak (12CP) is also a cost based on the requirement to maintain the capability to provide energy to all customers much like the 5CP cost, but it is experienced and measured each month and attempts to reduce usage/load during that hour will lower the cost to all customers of the Cooperative and in many instances individual customers who are billed demand charges for usage at the 12CP times. It is based on the highest zonal peak load occurring during the calendar month.

The report maintains current standing of 1CP & 5CP dates and hours for wholesale rates planning. Implementation of this project requires PJM Load telemetry and forecasted weather both of which would be updated in real-time for up to 10 days of forecasts for the PJM TO loads (AEP, APS, DOM, & DPL).

<u>**The Zonal Map of PJM**</u> – This zonal map of PJM and its RTO zones depicts the areas of concern and interest to ODEC forecasting.



20160112

Display 1. PJM Zonal Map

SHORT-TERM FORECASTING OUTLOOK REPORT

Understanding the Outlook Report

There are eight (8) panels containing information to help with load management decisions. The following is a detailed depiction of the report makeup and the information provided to users. This initiative requires the Outlook Report to be data-driven, self-correcting, and automated unlike current state.

This sample report, was revised on July 23rd at 08:45 and is a short-term forecast and therefore it can change frequently during any given day.

1

July '15	NPFL	MTD	MTD/NPFL	HE	Date
AEP	23,511	20,709	88.1%	17	20-Jul
DPL	4,177	3,730	89.3%	17	20-Jul
APS	8,734	8,282	94.8%	17	20-Jul
DOM	19,999	18,588	92.9%	17	20-Jul
PJM	155,544	143,931	92.5%	17	20-Jul

2

July '14	Peak	HE	Date
AEP	21,411	17	22-Jul
DPL	3,875	18	2-Jul
APS	8,084	17	23-Jul
DOM	18,692	16	2-Jul
PJM	139,386	18	1-Jul

Data Revised: JUL-23, 2015 @ 8:45

3

Today's forecast	AEP	DPL	APS	DOM	PJM
23-Jul-15	18,632	2,946	6,942	15,095	120,237
HE	17	18	18	18	17
M-T-D peak	20,709	3,730	8,282	18,588	143,931
Today's forecast ÷ M-T-D peak	90.0%	79.0%	83.8%	81.2%	83.5%
Forecast high next 7 days	20,063	3,540	7,998	16,999	136,223
Date	28-Jul	29-Jul	29-Jul	29-Jul	29-Jul
DOW	TUE	WED	WED	WED	WED
HE	17	18	17	18	17
7 day forecast ÷ M-T-D peak	96.9%	94.9%	96.6%	91.4%	94.6%
7 day forecast ÷ NPFL	85.3%	84.8%	91.6%	85.0%	87.6%
NSPL [1CP] forecast	23,511	4,621	9,024	19,999	
Forecast month	July '15	July '15	July '15	July '15	
1CP Y-T-D	24,739	4,114	9,594	21,651	
Date	2/20/15	2/20/15	2/20/15	2/20/15	
HE	8	8	8	8	
7 day forecast high ÷ 1CP Y-T-D	81.1%	86.1%	83.4%	78.5%	
1CP Y-T-D ÷ NSPL [1CP] forecast	105.2%	89.0%	106.3%	108.3%	
Likelihood of a new PJM summer peak					
PJM forecast peak for summer 2015					155,544
Max of 7 day forecast ÷			136,223		87.6%
2015 PJM 5CPs				% of summer peak	
	7/20/2015	17	143,931	92.5%	
	6/23/2015	17	137,287	88.3%	
	6/22/2015	18	134,545	86.5%	
	6/15/2015	17	134,129	86.2%	
	7/21/2015	17	131,980	84.9%	
7 day forecast ÷ Lowest 5CP					103.21%

4

DOM				
Day Max				
15,095	THU	7/23	18	81.2%
15,803	FRI	7/24	17	85.0%
15,326	SAT	7/25	17	82.4%
15,793	SUN	7/26	18	85.0%
16,284	MON	7/27	17	87.6%
16,778	TUE	7/28	17	90.3%
16,999	WED	7/29	18	91.4%

5

DPL				
Day Max				
2,946	THU	7/23	18	79.0%
3,109	FRI	7/24	18	83.3%
3,123	SAT	7/25	18	83.7%
3,242	SUN	7/26	18	86.9%
3,321	MON	7/27	18	89.0%
3,415	TUE	7/28	18	91.5%
3,540	WED	7/29	18	94.9%

6

APS				
Day Max				
6,942	THU	7/23	18	83.8%
7,072	FRI	7/24	17	85.4%
7,078	SAT	7/25	17	85.5%
6,989	SUN	7/26	18	84.4%
7,693	MON	7/27	17	92.9%
7,806	TUE	7/28	17	94.2%
7,998	WED	7/29	17	96.6%

7

AEP				
Day Max				
18,632	THU	7/23	17	90.0%
19,183	FRI	7/24	17	92.6%
18,384	SAT	7/25	18	88.8%
17,813	SUN	7/26	18	86.0%
19,950	MON	7/27	17	96.3%
20,063	TUE	7/28	17	96.9%
19,850	WED	7/29	17	95.9%

8

PJM				
Day Max				
120,237	THU	7/23	17	83.5%
124,876	FRI	7/24	17	86.8%
119,114	SAT	7/25	18	82.8%
119,311	SUN	7/26	18	82.9%
131,430	MON	7/27	17	91.3%
134,526	TUE	7/28	17	93.5%
136,223	WED	7/29	17	94.6%

Display 2. The Eight Panels of an Outlook Report

Panel #1 explains the current's month's forecasted load and actual activity month-to-date (MTD).

July '15	NPFL	MTD	MTD/NPFL	HE	Date
AEP	23,511	20,709	88.1%	17	20-Jul
DPL	4,177	3,730	89.3%	17	20-Jul
APS	8,734	8,282	94.8%	17	20-Jul
DOM	19,999	18,588	92.9%	17	20-Jul
PJM	155,544	143,931	92.5%	17	20-Jul

Display 3. Panel #1 of an Outlook Report

The MTD Peak is the highest actual hourly load amount recorded before today for the month indicated. In the example it happened for the hour ending 17 (6 pm) on July 20th.

FRCST is the highest forecasted load expected to happen sometime during the month. The forecast is a 'normal forecast' that may or may not come true depending mostly on actual weather conditions experienced during the month.

MTD/FRCST represents how close the current peak is to the forecasted amount. In this case, it is 92.5% for PJM, which tells you the peak for this month has not yet reached the forecasted amount.

Panel #2 shows the current's month's actual activity for the same month of last year.

July '14	Peak	HE	Date
AEP	21,411	17	22-Jul
DPL	3,875	18	2-Jul
APS	8,084	17	23-Jul
DOM	18,692	16	2-Jul
PJM	139,386	18	1-Jul

Display 4. Panel #2 of an Outlook Report

Panel #3 represents a 7-day forecast for the coming 7-day period. It is made up of 5 sub-panels that describe various forecasted loads.

Panel #3a explains the current day's forecasted load.

Today's forecast		AEP	DPL	APS	DOM	PJM
23-Jul-15		18,632	2,946	6,942	15,095	120,237
	HE	17	18	18	18	17
M-T-D peak		20,709	3,730	8,282	18,588	143,931
Today's forecast ÷ M-T-D peak		90.0%	79.0%	83.8%	81.2%	83.5%

Display 5. Panel #3a of an Outlook Report

Today's forecast is the highest forecasted load expected to happen for the hour ending 17. The forecast is a 'normal forecast' that may or may not come true depending mostly on weather conditions so subsequent issues of the Outlook emailed during the day might be different.

Today's forecast/M-T-D represents how close the forecasted peak is to the M-T-D amount.

Panel #3b shows the highest peak load forecasted for the next 7 days.

Forecast high next 7 days					
	20,063	3,540	7,998	16,999	136,223
Date	28-Jul	29-Jul	29-Jul	29-Jul	29-Jul
DOW	TUE	WED	WED	WED	WED
HE	17	18	17	18	17
7 day forecast ÷ M-T-D peak	96.9%	94.9%	96.6%	91.4%	94.6%
7 day forecast ÷ NPFL	85.3%	84.8%	91.6%	85.0%	87.6%

Display 6. Panel #3b of an Outlook Report

Top row is the highest forecasted peak load for the 7 day period in each zone plus PJM. Dates and times are shown.

Panel #3c represents the forecasted 1CP for the November through October period.

NSPL [1CP] forecast	23,511	4,621	9,024	19,999	
Forecast month	July '15	July '15	July '15	July '15	
1CP Y-T-D	24,739	4,114	9,594	21,651	
Date	2/20/15	2/20/15	2/20/15	2/20/15	
HE	8	8	8	8	
7 day forecast high ÷ 1CP Y-T-D	81.1%	86.1%	83.4%	78.5%	
1CP Y-T-D ÷ NSPL [1CP] forecast	105.2%	89.0%	106.3%	108.3%	

Display 7. Panel #3c of an Outlook Report

The forecasted 1CP is expected to occur in July of 2015 in all four zones. PJM does not have a 1CP determinant. The actual 1CP year-to-date (Y-T-D) is in the highest peak load recorded as of yesterday. The percentages are a metric which indicates how close the 7 day forecasted high comes to the actual 1CP and the forecasted 1CP.

Panel #3d indicates the likelihood of a new PJM summer peak occurring in the next 7 days.

Likelihood of a new PJM summer peak		
PJM forecast peak for summer 2015		155,544
Max of 7 day forecast ÷	136,223	87.6%

Display 8. Panel #3d of an Outlook Report

Panel #3e is titled 2015 PJM 5CPs. It may also be called the PJM 5CP stack or table or just plain 5CP.

2015 PJM 5CPs				% of summer peak	
	7/20/2015	17	143,931	92.5%	
	6/23/2015	17	137,287	88.3%	
	6/22/2015	18	134,545	86.5%	
	6/15/2015	17	134,129	86.2%	
	7/21/2015	17	131,980	84.9%	
7 day forecast ÷ Lowest 5CP					103.21%

Display 9. Panel #3e of an Outlook Report

PJM does not determine the 5CP times until mid-October of each year, so this "high five stack" is preliminary data representing a best estimate of the 5CP times existing at the time of this report.

Panels 4 through 8 are individual zone daily forecasted peaks plus PJM. The example is that of PJM.

PJM				
Day Max				
120,237	THU	7/23	17	83.5%
124,876	FRI	7/24	17	86.8%
119,114	SAT	7/25	18	82.8%
119,311	SUN	7/26	18	82.9%
131,430	MON	7/27	17	91.3%
134,526	TUE	7/28	17	93.5%
136,223	WED	7/29	17	94.6%

Display 10. Panels #4-8 of an Outlook Report

To begin the process of having the report process being more automated, ODEC is using SAS® Energy Forecasting to automate the forecasting process. The process we have been going through includes using SAS to explore the data, using SAS® Energy Forecasting to build and simulate different models for each time series, and then use SAS to evaluate the results.

SHORT-TERM FORECASTING MODELING

Descriptive Analytics

Before we try any modeling, we want to visualize the historical data we have to work with to gain insight into past trends and locate potential data issues. We looked at the data at the lowest time granularity and summarized levels. We looked at the relationships between all the variables – time, weather, load, and any independent variable we had for a particular time series.

Before evaluating the data for trends, we need to ensure we have a complete history of load, weather, and any user defined variables. If we had missing data, we would need to handle the missing data.

To get quick information about our data, we can use PROC TIMEID and PROC MEANS. PROC TIMEID will give us information about the time series. It will tell us how many intervals there are, the interval length, the number of observations, the start and end date and time stamps. We will check the load data, the weather data, as well as any independent variables. See Table 1 and 2 for sample output below:

Table 1 - PROC TIMEID Sample Output. Time Series: DOM

TABLE	TIMEID	START	END	NOBS	N	NMISS
Load	LOAD_DTTM	01Jan2012:01:00:00	02Apr2015:00:00:00	28488	28488	0
Weather	WTHR_DTTM	01Jan2012:01:00:00	02Apr2015:00:00:00	28488	28488	0

Table 2 - PROC TIMEID Sample Output Cont'd. Time Series DOM

TABLE	TIMEID	NINVALID	NINTCNTS	INTERVAL	MULTIPLIER	SEASONALITY
Load	LOAD_DTTM	0	1	HOUR	1	24
Weather	WTHR_DTTM	0	1	HOUR	1	24

The output from Table 1 and 2 shows us that we have a history of equal length, we aren't missing any observations, and that the interval is consistent and hourly. This is what we expect to see.

PROC MEANS allows us to get a look at quick stats on load, weather, and user defined variables. In this instance we looked at min, max, mean and median. We looked at these values to see if they made sense in context of the region where the data comes from. In Table 3, we have sample PROC MEANS output from the DOM time series. For both the load and the temperature, we see that we have reasonable minimum and maximum values.

Table 3-PROC MEANS sample output. Time Series: DOM

VARIABLE	LABEL	N	N MISS	MINIMUM	MAXIMUM	MEAN	MEDIAN
LOAD_MW_NO	Hourly Load (MW)	28488	0	4724.2	21607.59	10990.8	10612.84
WTHR_NO	Temperature (F)	28488	0	0.018773	101.137085	54.99866	55.82721

After looking at the PROC TIMEID and the PROC MEANS output, we used PROC SGSCATTER, PROC SGPANEL and PROC SGPLOT to look at a series of different plots to gain different views of the data.

To start, we looked at a general overview of the data using the compare option or PROC SGSCATTER. In Figure 1, we are looking at the DPL time series over the whole history available and we compare the weather data to the load data. Nothing is standing out as wrong.

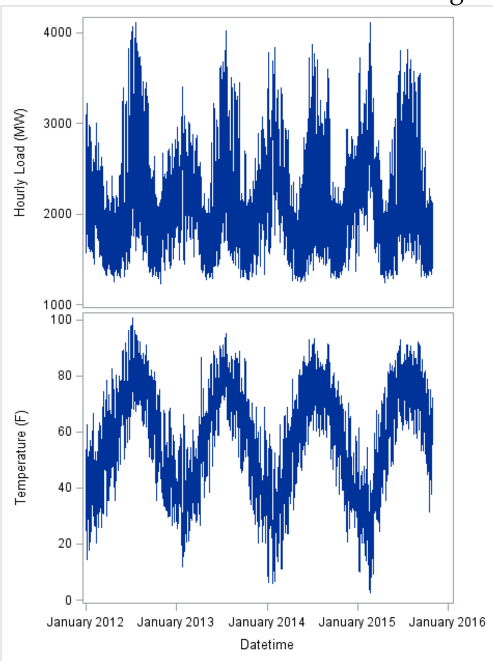


Figure 1-PROC SGSCATTER; compare option; Time Series: DPL

After looking at the data at a high level, we can zoom in on the series to see what it looks like closer up. In Figure 2, we are looking at a different time series, ODEC – AEP. This time series was used in a previous exercise. This set of data has the load, weather, and a user defined variable, PJMMW, which is the PJM load for the AEP region. In this graph, it is easy to see that there is a drop in load in the ODEC Load series which cannot be explained by either the PJM load or the temperature. It turns out this was an outage that we were previously unaware of. This plot also takes advantage of the PROC SGSCATTER with the compare option.

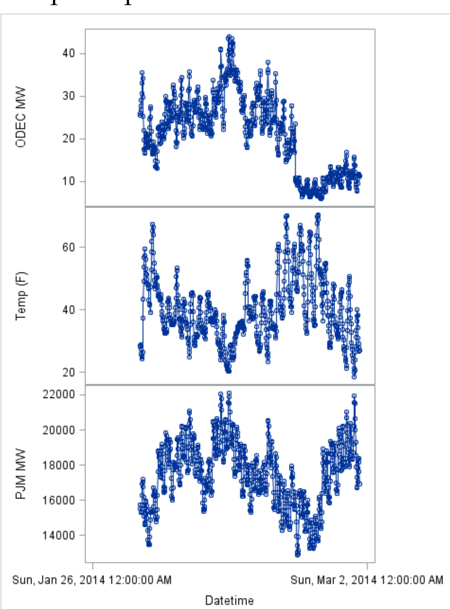


Figure 2-Monthly view of data – ODEC AEP 02/2014

This drop in load is also seen when we look at the data from a different view point. Using the PROC SGPANEL and PROC SGPLOT, we can construct two other plots that point out a problem in the data with the ODEC – AEP time series. In Figure 3, we use PROC SGPANEL to compare the temperature and the load and to split it out by year. In the graphs, we can see that there are some data problems in warmer months in 2012 and 2013 and in the colder months in 2014. This colder month time frame represents the same time period as shown in Figure 3.

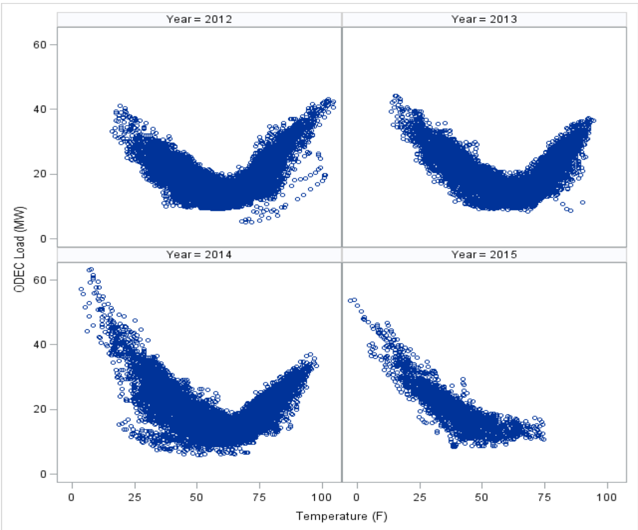


Figure 4- Load v Temperature by Year - AEP

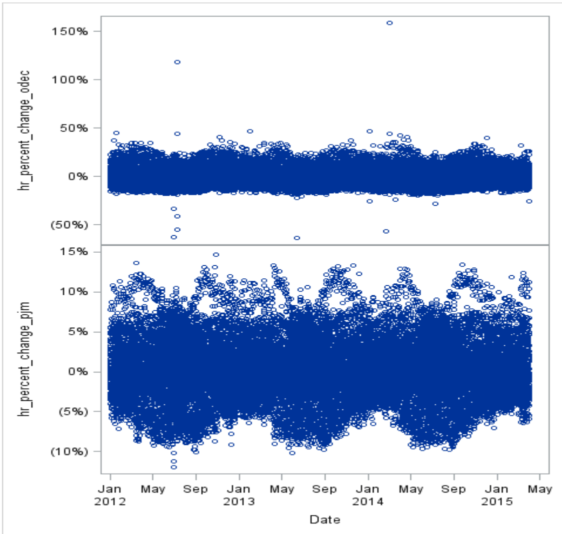


Figure 3- Percent Change in load from hour to hour; AEP

In Figure 4, we use PROC SGSCATTER with the compare option to show by what percentage the load changes from hour to hour. When there are big drops or gains, the points stand out and indicate something may have happened with the load that is outside the normal range. In the graph, we can compare the changes in the ODEC load to the changes in the PJM load. Like in Figure 3, we see that the ODEC portion of the load (graph on top) has big changes in warmer months as well as in the winter of 2014. This type of graph helps draw your eye to periods when you have a drastic change from one hour to the next.

We can continue to look at the data in different ways. In Figure 5, we use box plots to show the trends in the average hourly load and temperature day. We show this for a season, winter in Figure 5, and across years. Figure 5 shows that the average hourly daily temperature was decreasing while the average hourly load was increasing during the winter season. We can also see there were some outlier days for each season in load, but this also corresponded with low temperatures.

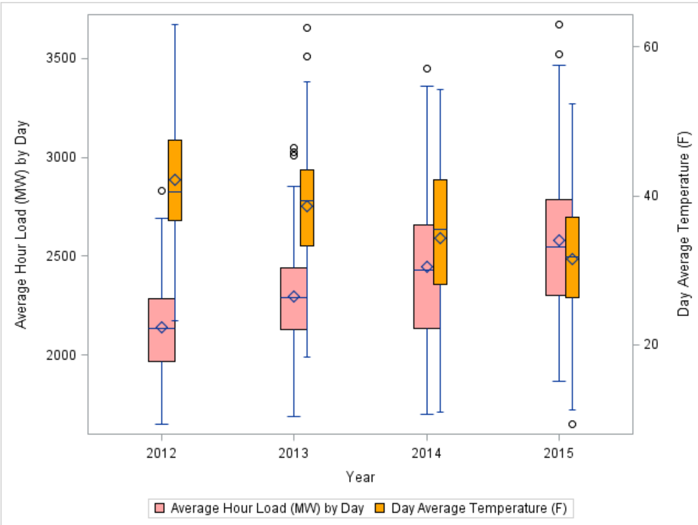


Figure 5-Winter Average Hourly Day Load and Temperature - Time Series: DPL

In addition to looking at boxplots we compare the load and temperature profiles by like days. In Figure 5, we use a PROC SGSCATTER to compare the load and temperature profiles for like days. We look at a specific month, year and day of week and see how the curves look. In Figure 5, we are looking at Saturdays in July 2013 for the DPL time series. The top graph represents load while the bottom graph represents temperature. We notice that there are two days that have a significantly higher load curve than the other two days. Looking at the bottom graph we can see this is easily explained by the temperature graphs. On the same two days, the temperature was significantly warmer. This type of graph and SAS analysis allows us to zoom in more closely on areas of data and see if strange shapes can be explained by another variable.

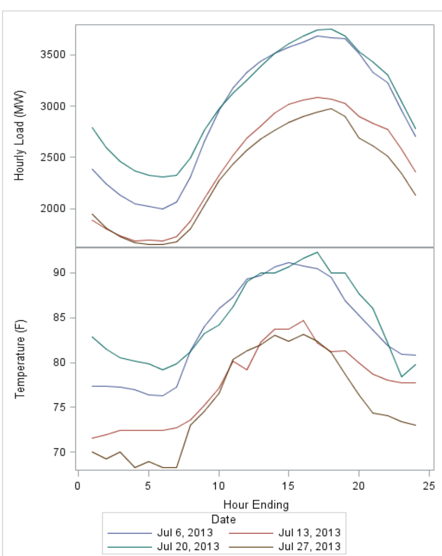


Figure 6-Like Days Saturdays July 2013; Time Series DPL

Using simple SAS procedures, we can gain greater insight into the data and identify any data issues and see relationships and trends.

Building Models
The SAS® Energy Forecasting engine

Forecasting with the SAS® Energy Forecasting engine is a two-step process. First, there is the diagnose (or model and variable selection) process which will go through an iterative process to find the champion model for the data provided. After the champion model is selected, the forecasting process can occur. This process is what generates the forecasts for the desired date range.

In the diagnose phase of forecasting with the engine, an iterative process is used to construct the champion model. The process starts with a very basic model and then iteratively adds new layers and tests the new layers to see if it added value to the forecast. The layers in the process as follows:

1. Stage 1
    a. Naïve Model – This is the benchmark. It models the load based several seasonalities and temperature combinations.
    b. Recency Effect – Takes the Naïve model and adds what we call the recency effect. This effect looks at the information from the previous hours and looks for an impact on the current hour.
    c. Weekend Effect – Takes the champion model from iteration b and adds a layer for the weekend. It looks for days of the week that are similar and will treat them similarly. E.g. in the U.S. Saturday and Sunday might look similar to each other but different from Monday – Friday.
    d. Holiday Effect – Takes the champion model from Iteration c and will account for holidays. This model will evaluate if a holiday's day of week should be treated differently than its actual day.
    e. Weighted Least Squares (optional) – Takes the champion model from above and adds an emphasis to the more recent errors using different weightings to minimize the residual sum of squares approach.

2. Stage 2 (optional) – This takes the Stage 1 model and calculates the residuals. It then fits a model to these values. This procedure can help improve model accuracy. Models considered for the residuals are Neural Network, UCM, ESM, ARIMAX, and a combination of those models.

Forecast Model Development

Before we started the RTO project, we did a proof of value exercise to establish the credibility of SAS® Energy Forecasting with ODEC data and as a way to make a decision about whether to move forward with using the SAS® Energy Forecasting engine. To establish the benchmark, we built models using multiple regression of general linear models. These models took into account seasonalities, temperatures, as well as an additional variable of a parent load.

The purpose of the models for the RTO project are to provide forecasts when the daily peak load is going to occur and the magnitude of that forecast. These models are most important in summer when the 5CP is being decided. With this in mind, we can construct our forecast models.

Using the interface we can get familiar with the kinds of parameters we want to use and we can run sample forecasts. In this phase of the model building, we experimented with two different MAPE criteria (Hourly MAPE and Daily Peak MAPE), the handling of the outlier data, and adding in additional driver variables.
Why use of hourly and Daily Peak MAPE criteria? The idea behind exploring these two MAPE criteria is that with hourly MAPE we are just trying to reduce our MAPE overall whereas with the daily peak MAPE, we are more focusing on how the models behave with respect to the peak load.

With the handling of outlier data, we explored how to identify the outliers and once identified what to do with the value whether we replace it with a modeled value or we create a user defined variable and give our outlier values a value of 1 and our non-outlier values a value of 0. Further development work is currently in process.

During this phase of development, the data we've been working with is what is known as metered load history. In our future production system we will have up to three types of load data - metered load history (billing quality data – available after preliminary load data), the preliminary load data (available at 5 am the next day), and instantaneous load data (available instantaneously). Metered load history is the highest quality data followed by the preliminary load data and then the instantaneous loads. In production, we will be working with these three loads to build our forecast.

The additional driver variable that we explored is load information for the parent loads of the time series we are interested in. For the DPL time series, this is known as Mid-Atlantic. For the three remaining time series, this is PJM. It is thought that both of these variables have very good data and so there is interest in exploring how this variable will impact the forecast and exploring the relationship between it and the RTO load.

Forecast Model Simulations

By looking over a long period of time, such as summer, we can gain understanding with how the models will behave in the long run. This can also allow us to compare several parameters over the same period of time to come to better conclusions about which models will make the most sense in the long run.

We were able to take advantage of SAS code to run these simulations in batch. Our test period looked at June – August 2015. During this time period we ran a diagnose every Saturday and ran a short term and a very short term forecast. For simulation purposes, the short term forecast was a 7 day forecast and the very short term forecast was a 24 hour forecast starting at 10 am. The idea behind this schedule was to mimic a mini production environment.

We had a simulation for several different sets of parameters. In the simulations, we investigated the use of different MAPE criteria (Daily Peak MAPE and Hourly MAPE), user defined variables (including parent loads), order of temperature polynomial (2 or 3) and model stage (1 or 2).

This past fall we began development on one of our models, DPL and the results below pertain to this time series. Updates to this model are currently in progress as we update our load data to the production environment as well as a change in our weather source.

Analyzing Results

SAS® Energy Forecasting saves the results of each run of the forecast in a separate instance folder. Within these folders we have the champion model forecast as well as the forecasts of what the champion is based on. (E.g. If the Champion Model is a two stage ESM model, then we have results for the one stage linear regression model and the two stage ESM model). Using SAS code, we can combine our forecasted values with the actual values and calculate the desired MAPE values to see how the forecast performed over the out-of-sample period.

This first set of results, as shown in Table 4, shows the overall results from Summer of how often we achieved at least one of our goals, our goals being to be within about 3% of our daily peak load and having our forecasted peak hour be within one hour of the actual peak hour. The first two columns represent models for a very short term forecast and the second two columns are the short term forecast. In the title, the HR indicates it was a model built using an hourly MAPE as selection criteria while the DP represents models built using a daily peak MAPE selection criteria. To keep this comparison fair, only the peak load from the first day of the 7 day short term forecast is evaluated.

Table 4-Results of Simulation Runs – DPL Summer 2015

	VSTF - HR	VSTF - DP	STF - HR	STF - DP
On Time or On Target	94.4%	97.7%	88.9%	80.6%
On Time	86.1%	90.7%	77.8%	68.1%
On Target	61.1%	55.8%	61.1%	47.2%
Neither	5.6%	2.3%	11.1%	19.4%

In Table 4, we see that for both Very Short Term Forecast Models we missed both criteria less than about 5% of the time and we hit either criteria at least 94% of the time. This particular set of models does not include any user defined variables. This set of results is for the DPL time series, which we began development work on this past fall.

Table 5-Results of Simulation Runs – DPL Summer 2015 with Additional Variable

	VSTF – HR	STF – HR
On Time or On Target	95%	97%
On Time	89%	90%
On Target	66%	70%
Neither	5%	3%

In Table 5, we see our summarized results for our simulation model when we added in the Mid-Atlantic information as a driver for the model. We see that we were able to add value when we compare to Table 4, especially with respect to the Short Term Forecast – Hourly Model. Further exploration is needed to determine how this variable will behave long term.

Using SAS, we can look at our results further to see how a specific forecast looks. In Figure 6, we see the 7-day forecast for DPL starting July 22, 2015. The solid line represents the predicted load while the points represent the actual load. In the graph, the actual and predicted values look pretty close. This graph was generated using PROC SGPLOT with the series and scatter options.

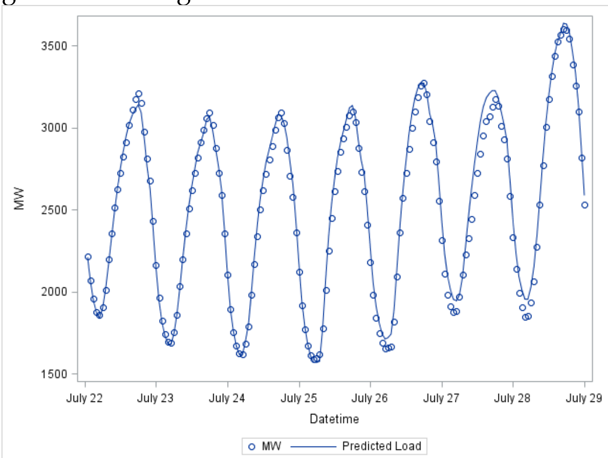


Figure 7- 7-day forecast and actuals; Time Series - DPL

We can also use SAS to look at how we are on forecasting a day leading up to a day. In Table W, We look at how the peak prediction was in the days leading up to the peak.

Table 6-Tracking Forecast Models

Date of Forecast	Date	Actual Peak Hour	Forecasted Peak Hour	Actual Peak Load (MW)	Predicted Load at Actual Peak Hour (MW)	Maximum Predicted Load on Day	APE at Peak Hour	Difference between Predictions
03SEP15	03SEP15	17	18	3534.86	3555.89	3562.90	0.60%	0.20%
02SEP15	03SEP15	17	18	3534.86	3550.80	3554.16	0.45%	0.09%
01SEP15	03SEP15	17	18	3534.86	3518.38	3525.30	0.47%	0.20%
31AUG15	03SEP15	17	18	3534.86	3518.43	3524.31	0.47%	0.17%
30AUG15	03SEP15	17	18	3534.86	3518.42	3524.75	0.47%	0.18%
29AUG15	03SEP15	17	18	3534.86	3518.77	3525.00	0.46%	0.18%
28AUG15	03SEP15	17	18	3534.86	3532.79	3536.26	0.06%	0.10%

In this table, we are looking at the peak forecast predictions for September 3, 2015. Each observation is for each 7 day forecasts that includes September 3, 2015 in its horizon. We show the hour of the actual peak as well as the hour of the predicted peak. In this case, we were within tolerance for all 7 days of the forecast. The next three columns show the Actual peak load, the prediction of the load at the actual peak hour (17) and then our prediction for what the peak load value was going to be. When comparing our two predictions, we see that the values were all less than 0.2% apart from each other. We can also see that our accuracy was very good as well and within tolerance. This particular results come from a Short Term Forecasting Model for DPL that used Hourly MAPE as a selection criteria and had the independent variable. This date, September 3, 2015, was also part of the 5CP for summer 2015.

Using SAS® Energy Forecasting and other SAS tools, we look forward to future development work to further improve the existing models as well as build new ones.

ARCHITECTURE

A number of forecasting efforts have been discussed or implied in this paper. Following is a matrix to align initiatives, technology, and their current status. Note that load forecasting is an iterative process, both in changing operational components and evolving business needs. Not listed but on the ODEC development roadmap include a 30-year long-term forecasting project.

Effort Name	Business Need	Data Mart	Forecasting Engine	Reporting	Status
Power Requirement Studies (PRS)	Mid-term forecasting models	Data Sets via Code	Forecast Studio	Enterprise Guide	Production on-site VMWare Server
PRS Data Mart	Mid-term forecasting data mart	Data Sets via DI Studio	As above	Enterprise Guide	Production on-site VMWare Server
PRS Reporting	Mid-term forecasting reporting	As above	As above	Currently in Progress	Development on-site VMWare Server
RTO Proof of Value	Intraday forecasting models	Data Sets via Code	EF 3.1 VSTLF/STLF Models	Code and Visual Analytics	Completed on AWS Instance
RTO intraday forecasting	Intraday forecasting models	As above	As above	Results Data feed Current Excel Reports	Currently in Progress

Table 7. ODEC Efforts

The SAS® Energy Forecasting solution contains a number of SAS products for both custom and out-of-the box data management, modeling, and reporting. The solution allows for development, test, and production configurations. Here is a description of the SAS® Energy Forecasting solution offering: http://www.sas.com/en_us/industry/utilities/energy-forecasting.html

Product highlights in the solution (technical and contract agreements apply):
SAS® Energy Forecasting and Workbench
SAS® Forecast Server and Studio
SAS® Enterprise Miner
SAS® Data Integration
SAS® Visual Analytics Administration and Reporting
SAS® Enterprise Guide
SAS® Studio
SAS® Add-in for Microsoft Office (Excel only)
SAS/ACCESS® Engines (2)
SAS/ACCESS® Interface to SAP HANA
SAS/GRAPH®
SAS/STAT®
SAS/ETS®
SAS/OR®
SAS® Visual Analytics Mobile Clients
SAS® Mobile BI

As mentioned above, the RTO proof of value (POV) was accomplished using SAS in AWS. In parallel the POV models were tested on VMWare and metal installations. It was noted that the metal installation was proportionally faster than the VMWare configuration – this would come to no surprise given that the hardware specifications were similar. VMWare has an overhead that is relatively consistent. VMWare is a good option for virtualization, so having lower performance should not be the only reason to not use VMWare. In our case AWS was proportionally faster than the metal installation. The reason for this was that the hardware used for the metal installation and VMWare installation was over 1-year old, while the AWS instance which was bound at the time of need was effectively newer and therefore faster hardware. Again, having faster performance should not be the only reason to use AWS.

Flexible and scalable data and discovery environments allow ODEC analysts to efficiently develop and test models which are IO intensive. SAS visualization for the analyst is a graphic compute environment for information sharing that is memory intensive. Also, ODEC IT operations require deployment options tuned for process optimization to meet service level agreements which can be quickly evaluated, tested and promoted into production. Increased computing demands at ODEC can be achieved by leveraging AWS for expanded business discovery and operational capacity.

CONCLUSION

ODEC has been able to revisit their approaches, methods, and technology to accurately predict energy demands across all time horizons in a timely manner. With the advanced analytics of SAS® Energy Forecasting, ODEC provides data-driven load predictions from next hour to next year and beyond. Accurate intraday forecasts mean meeting daily peak demands saving millions of dollars at critical seasons and events. Mid-term forecasts provide a baseline to the cooperative and its members to accurately anticipate regional growth and customer needs in addition to signaling power marketers where, when and how much to hedge future energy purchases to meet weather-driven demands. Long-term forecasts create defensible numbers for large capital expenditures such as generation and transmission projects. Much of the data for determining load comes from disparate systems such as SCADA and internal billing systems combined with external market data (PJM), weather, and economic data. This data needs to be analyzed, validated, shaped, and conformed to fully leverage predictive methods. Business insights and planning metrics are achieved when flexible data integration capabilities are combined with advanced analytics and visualization. These increased computing demands at ODEC are being achieved by leveraging Amazon Web Services (AWS) for expanded business discovery and operational capacity. What was once very difficult for most utilities to embrace is now achievable with new approaches, methods, and technology like never before.

CONTACT INFORMATION

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

David Hamilton
Old Dominion Electric Cooperative
804-968-4017
DHamilton@odec.com
http://www.odec.com/

Emily Forney
SAS Institute Inc.
919-531-0984
Emily.Forney@sas.com
http://www.sas.com/

Steve Becker (Architecture)
SAS Institute Inc.
919-531-9852
Steve.Becker@sas.com
http://www.sas.com/