Itron
20 Park Plaza, 4th Floor
Boston, MA 02116
(617) 423-7660

2022 PJM MODEL REVIEW

FINAL REPORT

Prepared For:

PJM

Prepared By:

Eric Fox, Senior Forecast Consultant
Michael Russo, Forecast Consultant
Dr. Frank A. Monforte, Senior Forecast Consultant
Dr. Stuart McMenamin, Managing Director

TABLE OF CONTENTS

LIST OF FIGURES

1 OVERVIEW

Itron, Inc. (Itron) was contracted by PJM to review the current forecast models and develop a set of recommendations for transitioning to an hourly forecasting framework. This report discusses the forecast, recommendations, and addresses issues raised by the PJM Market Participants.

Over the last three months, Itron has worked closely with PJM forecast staff in reviewing current forecast approach, identifying forecast issues, and developing new models for evaluation and ultimately a set of recommendations for the next phase of model improvements. The core task is to transition from daily peak and energy models to hourly models where the need for hourly long-term models is driven by the penetration of new technologies that are either currently or are expected to reshape system hourly loads.

Each year, PJM updates long-term energy and demand forecast for 22 Planning Zones (EDC) that are mapped into 6 Load Deliverability Areas (LDA). PJM faces a unique and complex forecasting problem as the PJM service area stretches across a large section of the country that includes utilities from Chicago to Philadelphia. Planning Zones are spread across different time zones, include different mixes of residential, commercial, and industrial customers, and at any one point in time, significant differences in weather conditions. All this must be pulled together to not only generate load forecasts for individual zones, but for the entire PJM system.

The current approach entails estimating a set of daily energy, own-peak, and coincident peak models for each zone and simulating with historical zonal daily weather data. Simulation results are combined across zones with the system forecast based on the 50% (expected case) outcome. Model-based zone peaks and coincident peaks are then exogenously adjusted for solar demand, electric vehicles, battery storage, and zone-specific load impact programs and large changes in industrial loads including data centers. The problem is a forecast based on daily demand models implicitly assumes the timing of the peak is constant over time. This is a relative safe assumption when factors that potentially reshape loads are small. But this is no longer the case; the projected impact of electric vehicles, increase in solar load, battery storage, and other programs designed to address climate change will significantly reshape loads impacting both the level and timing of system peak demand. The only means to address the issue is to transition to hourly load modeling framework that still addresses the weather, time zones, and customer diversity across the PJM Zones.

The report is organized into the following sections:

1. Overview
2. Review of the Current PJM Model
3. Developing the Long-Term Model Drivers
4. Hourly Model Analysis and Recommendations
5. Reshaping Demand – Modeling Technology Impacts
6. Other Issues
    - Measuring Forecast Accuracy
    - Weather Normalization
    - Impact of State and Utility Energy Efficiency Programs
    - Capturing Temperature Trends
7. Recommendations

2 REVIEW OF THE CURRENT PJM MODEL

PJM's current Long-term Forecast System (LTFS) is presented in Figure 2-1. The current PJM modeling process includes three sets of daily models for each zone. These are daily energy, daily noncoincident zone peak (NCP), and daily CP (zone load at the time of the daily PJM peak). For each zone, CP models are estimated for daily loads coincident with the Locational Deliverability Area (LDA) peak and coincident with the overall PJM system peak. The LTFS gathers the data required to estimate and forecast the load zone models, estimates the models, and then develops long-term forecasts using a multi-year weather simulation approach. The LTFS is a SAS™ based software solution that can be summarized into four major sections. For a detailed description of the current LTFS see **2022 Load Forecast Supplement**, PJM Resource Adequacy Planning Department, January 2022 (https://www.pjm.com/-/media/planning/res-adeq/load-forecast/load-forecast-supplement.ashx)

Data Input Section. The data inputs that feed the LTFS come in a variety of electronic formats. This section prepares the raw input data to be ingested by the Data Transformation Section of the LTFS. Data inputs include:

- Economic history and forecasts,
- End-use saturation and efficiency trends (SAE),
- Sector (residential, commercial, and industrial) sales (MWh) by EDC,
- Hourly Loads by EDC,
- Weather data by weather station, and
- Solar PV, EV Charging, Battery Storage, and Demand Response peak demand impacts.

Data Transformation Section. This section converts the raw input into daily coincident and non-coincident model variables. Key modules within this section are:

Sector Models. This series of three modules – Residential, Commercial, Industrial – are used to combine historical and forecasted economic data, end-use appliance saturation and efficiency trends, and annual sector sales to construct Space Heating, Space Cooling, and Other Non-Weather sensitive load indices that are used to drive long-term coincident and non-coincident peak forecasts by Electric Distribution Company (EDC). Figure 2-2 shows a map of the EDCs that make up the PJM control region. The Space Heating index is combined with daily Heating Degree Days to form an estimate of daily space heating load (MWh). The Space Cooling index is combined with daily Cooling Degree Days to form an estimate of historical and forecasted daily space cooling load (MWh).

Historical Load & Weather Data. Hourly historical load & weather data by EDC and weather station are input into the LTFS. The non-weather sensitive portion of the load is extracted from these data and is used to scale the Other Non-Weather sensitive load index to daily MWh for both the historical and forecast period.

Reconstituted Coincident and Non-Coincident Peak Loads. To account for load loss due to embedded solar PV generation, historical estimates of aggregate embedded solar PV generation for

each hour is added to the historical coincident and non-coincident peak loads that are computed directly from the historical measured load data with estimates of demand response added back. The coincident and non-coincident peak forecasts are then adjusted downward in the forecast period by subtracting off forecasted embedded solar generation values for hour ending 17:00.

Model and Forecast Engines Section. This section is used to specify and estimate the Daily Energy, and Coincident and Non-Coincident Peak Models by EDC. The second module generates the long-term daily energy, and coincident and non-coincident peak forecasts by EDC and Weather Simulation trace.

Daily Energy, and Coincident and Non-Coincident Peak Models. The Space Heating and Space Cooling indices are combined with historical Heating Degree and Cooling Day variables to form MWh estimates of daily space heating and cooling. The daily values are included along with the estimated Other Non-Weather sensitive load data as explanatory variables in a Daily Coincident Peak and Daily Non-Coincident Peak models. A separate set of Coincident and Non-Coincident Peak models is estimated for each EDC using the EDC specific daily Heating MWh, Cooling MWh, and Other MWh values.

Weather Simulations. To account for the geographical and temporal diversity of the large PJM operating footprint weather simulations are used to generate a distribution of daily coincident and non-coincident peak day forecasts for each EDC. The weather simulations use historical hourly weather data by weather station to form forecasts of daily Heating Degree Days and Cooling Degree days by EDC. The setup of the weather simulation traces is depicted in Figure 2-3. Each year of historical weather data is used to construct 13 separate load forecasts for each year in the forecast horizon. The 2022 vintage of the LTFS utilizes 27 years of historical weather data that drives 351 (computed as 13 x 27) daily load forecasts by EDC and year in the forecast horizon. Using the 351 forecast simulations, the LTFS constructed Cumulative Frequency Distributions of Coincident and Non-Coincident Peak Loads by Month/Season and EDC. The Cumulative Frequency Distributions are used to select load bands under design conditions (e.g., 50%, 10%, 90%). Advantages of the weather simulation framework are:

- Actual and Consistent Weather Patterns Across EDCs. Actual weather data for all weather stations are rotated together ensuring consistent and realistic weather patterns are used to drive the load forecasts.
- Consistency Across Weather Concepts. All weather concepts are rotated together ensuring consistent and realistic movements in temperatures, humidity, wind speed, wind direction, cloud cover, and solar irradiance. This is critical to align HVAC and solar PV generation interactions.

Forecast Adjustments. At the end of each weather simulation embedded Solar PV, EV Charging, and other technology impacts are applied to the coincident and non-coincident peak forecasts. The last step is to make adjustments for expected distribution system capacity projects that are not captured by the base case load forecast process.

Forecast Summary Report Module. This section is used to create formal and ad hoc summaries of the long-term forecasts.

FIGURE 2-1: LTFS 2022 MODEL FRAMEWORK

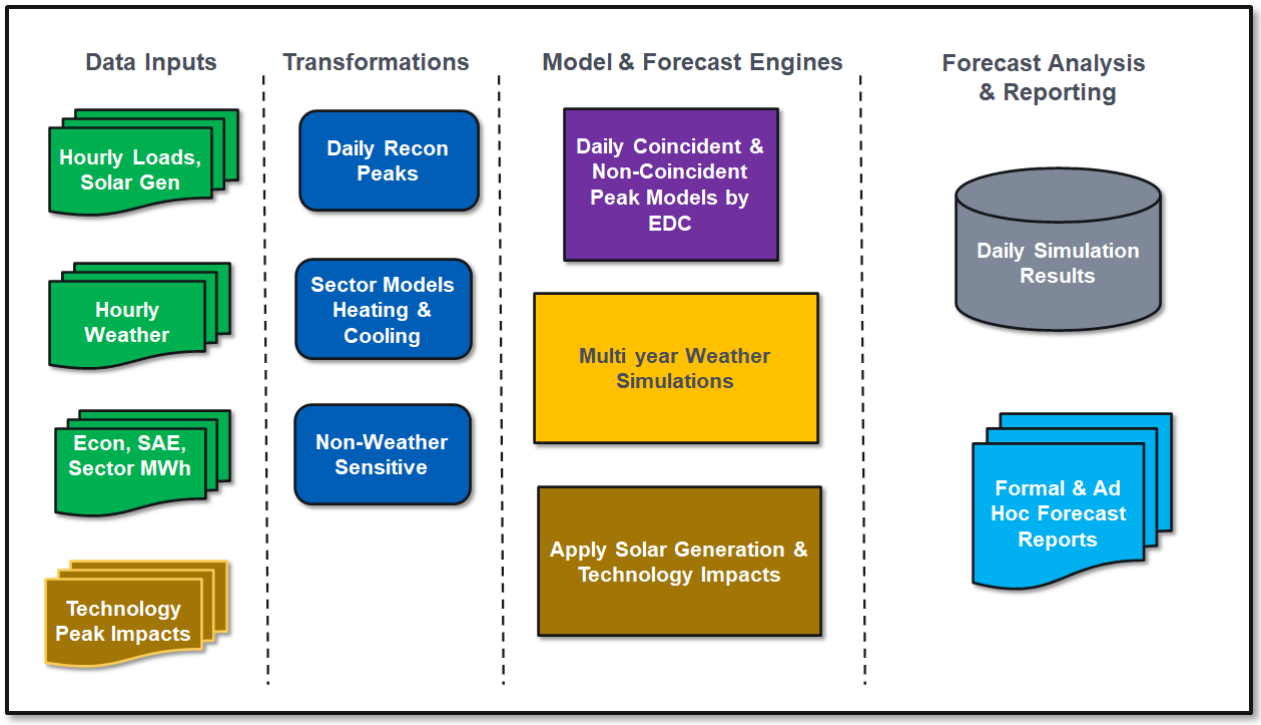


FIGURE 2-2: 2022 MAP OF PJM PARTICIPATING EDCS

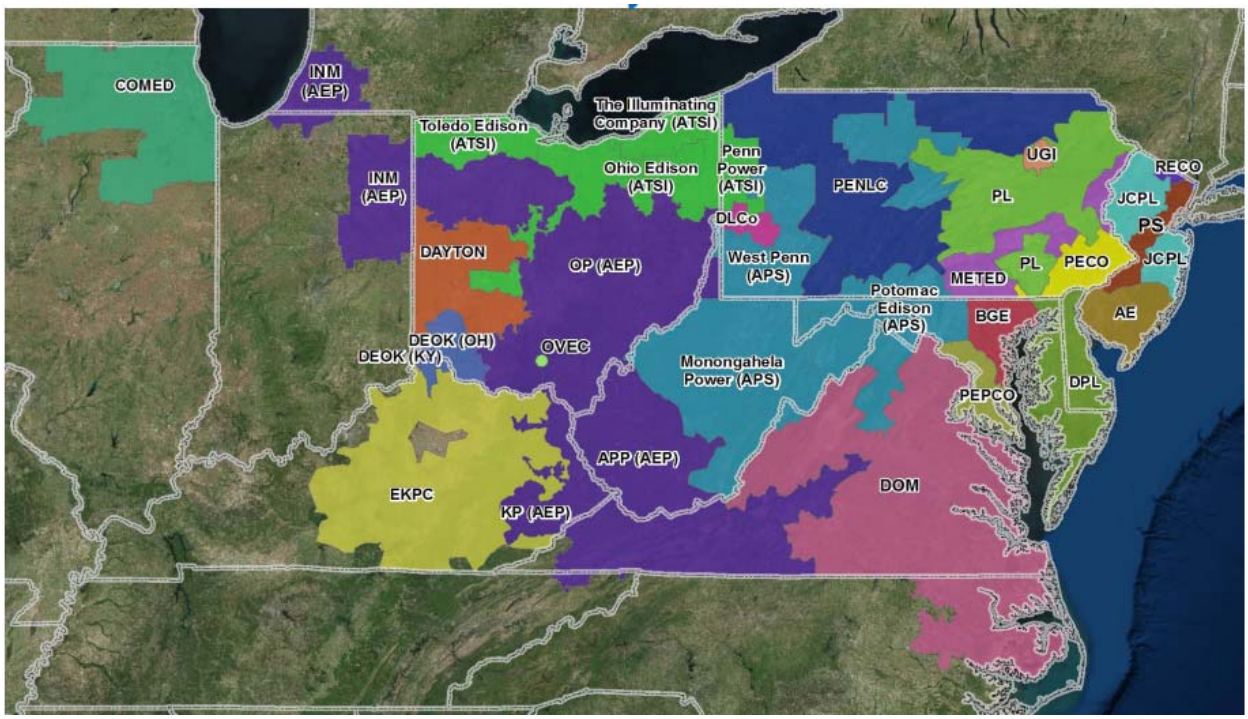


FIGURE 2-3: CONFIGURING WEATHER SIMULATION FORECAST TRACES

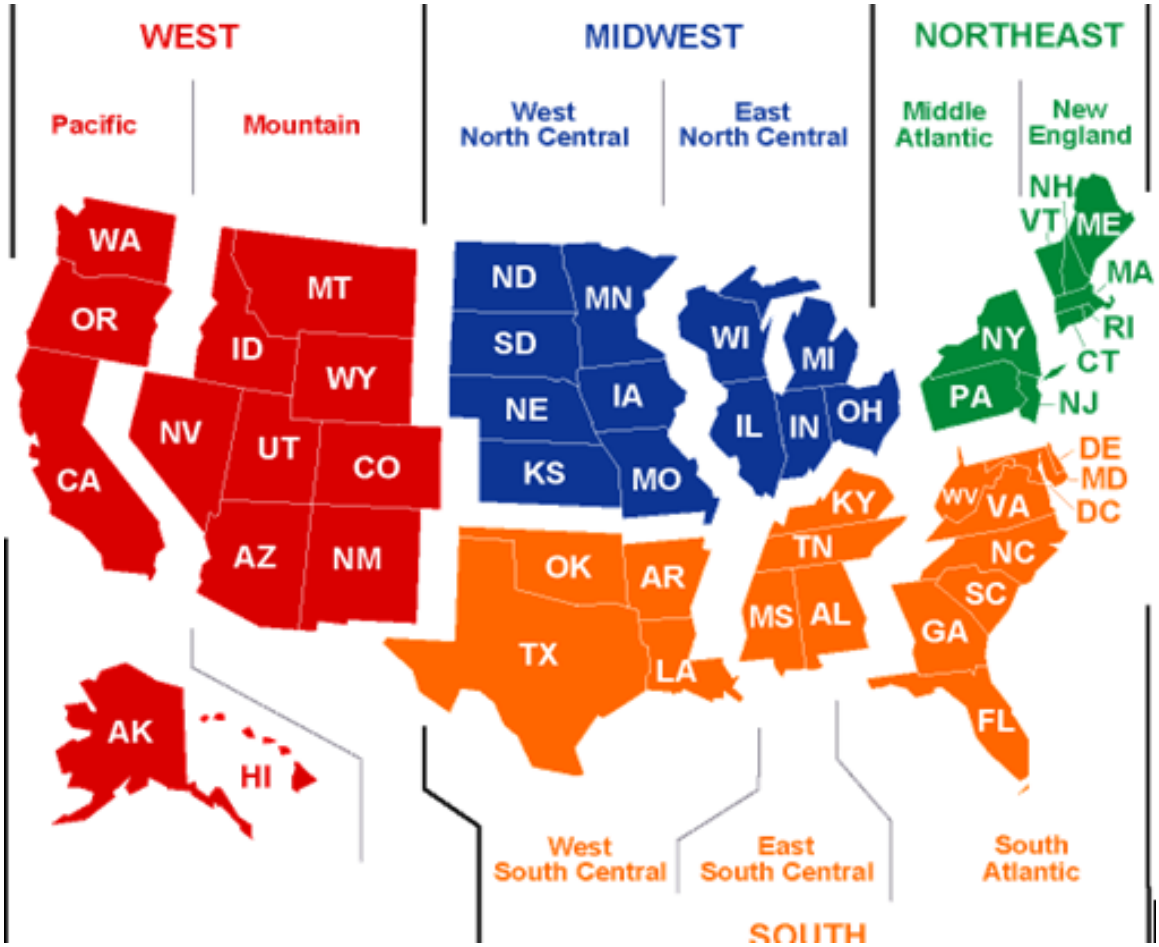
	Weather Scenarios												
		Rotate Forward						Rotate Backward					
Date	A1995	B1995	C1995	D1995	E1995	F1995	G1995	H1995	I1995	J1995	K1995	L1995	M1995
1-Jan	1/1/1995	1/2/1995	1/3/1995	1/4/1995	1/5/1995	1/6/1995	1/7/1995	12/31/1995	12/30/1995	12/29/1995	12/28/1995	12/27/1995	12/26/1995
2-Jan	1/2/1995	1/3/1995	1/4/1995	1/5/1995	1/6/1995	1/7/1995	1/8/1995	1/1/1995	12/31/1995	12/30/1995	12/29/1995	12/28/1995	12/27/1995
3-Jan	1/3/1995	1/4/1995	1/5/1995	1/6/1995	1/7/1995	1/8/1995	1/9/1995	1/2/1995	1/1/1995	12/31/1995	12/30/1995	12/29/1995	12/28/1995
4-Jan	1/4/1995	1/5/1995	1/6/1995	1/7/1995	1/8/1995	1/9/1995	1/10/1995	1/3/1995	1/2/1995	1/1/1995	12/31/1995	12/30/1995	12/29/1995
5-Jan	1/5/1995	1/6/1995	1/7/1995	1/8/1995	1/9/1995	1/10/1995	1/11/1995	1/4/1995	1/3/1995	1/2/1995	1/1/1995	12/31/1995	12/30/1995
6-Jan	1/6/1995	1/7/1995	1/8/1995	1/9/1995	1/10/1995	1/11/1995	1/12/1995	1/5/1995	1/4/1995	1/3/1995	1/2/1995	1/1/1995	12/31/1995
7-Jan	1/7/1995	1/8/1995	1/9/1995	1/10/1995	1/11/1995	1/12/1995	1/13/1995	1/6/1995	1/5/1995	1/4/1995	1/3/1995	1/2/1995	1/1/1995
-	-	-	-	-	-	-	-	-	-	-	-	-	-
25-Dec	12/25/1995	12/26/1995	12/27/1995	12/28/1995	12/29/1995	12/30/1995	12/31/1995	12/24/1995	12/23/1995	12/22/1995	12/21/1995	12/20/1995	12/19/1995
26-Dec	12/26/1995	12/27/1995	12/28/1995	12/29/1995	12/30/1995	12/31/1995	1/1/1995	12/25/1995	12/24/1995	12/23/1995	12/22/1995	12/21/1995	12/20/1995
27-Dec	12/27/1995	12/28/1995	12/29/1995	12/30/1995	12/31/1995	1/1/1995	1/2/1995	12/26/1995	12/25/1995	12/24/1995	12/23/1995	12/22/1995	12/21/1995
28-Dec	12/28/1995	12/29/1995	12/30/1995	12/31/1995	1/1/1995	1/2/1995	1/3/1995	12/27/1995	12/26/1995	12/25/1995	12/24/1995	12/23/1995	12/22/1995
29-Dec	12/29/1995	12/30/1995	12/31/1995	1/1/1995	1/2/1995	1/3/1995	1/4/1995	12/28/1995	12/27/1995	12/26/1995	12/25/1995	12/24/1995	12/23/1995
30-Dec	12/30/1995	12/31/1995	1/1/1995	1/2/1995	1/3/1995	1/4/1995	1/5/1995	12/29/1995	12/28/1995	12/27/1995	12/26/1995	12/25/1995	12/24/1995
31-Dec	12/31/1995	1/1/1995	1/2/1995	1/3/1995	1/4/1995	1/5/1995	1/6/1995	12/30/1995	12/29/1995	12/28/1995	12/27/1995	12/26/1995	12/25/1995

3 LONG-TERM MODEL DRIVERS

PJM's long-term energy and demand forecast is driven by zone-level population, economic activity as captured by regional output and employment and end-use energy trends. In general, sales growth associated with new customers and business activity is offset by improvements in energy efficiency. The 2022 PJM baseline forecast (before forecast adjustments) is basically flat when end-use energy intensities are combined with population and economic projections; decline in use per customer is roughly equal to customer growth. Long-term demand is largely driven by expected electric vehicle and data center load growth.

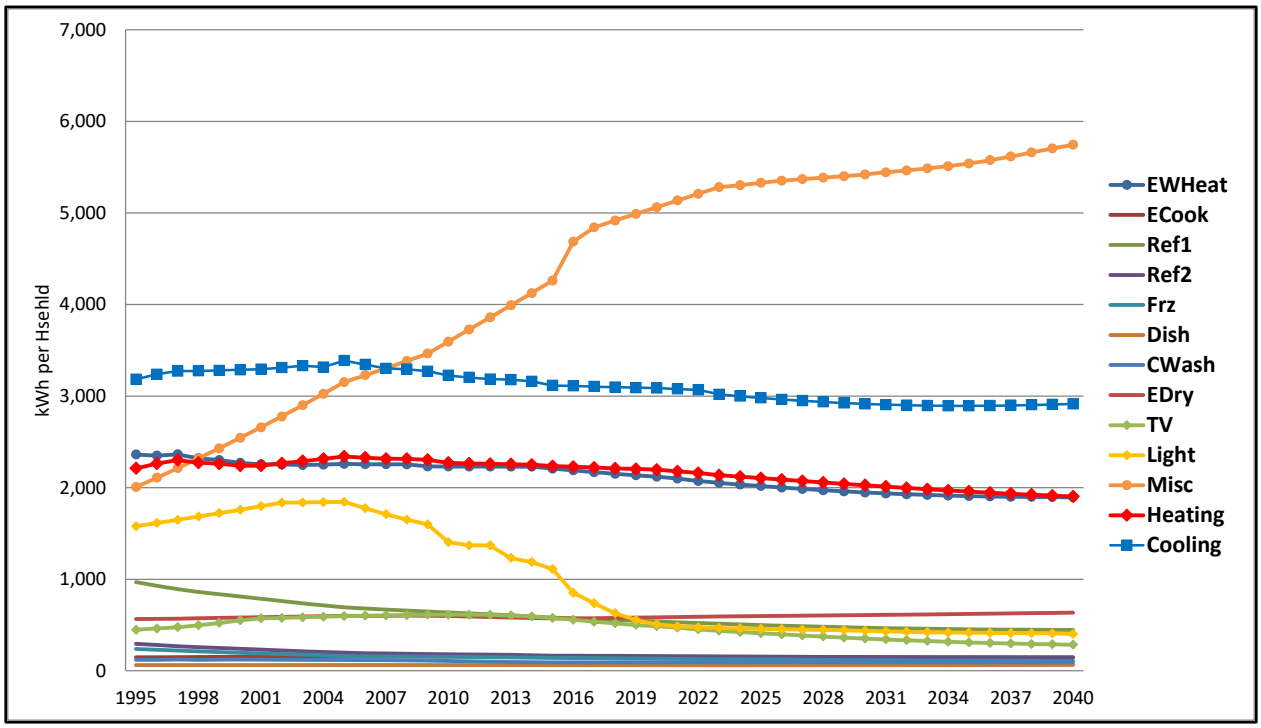
The PJM energy trend is consistent with other regions where long-term population and economic output are relatively slow. What has driven the decline in customer usage are end-use efficiency standards, utility efficiency programs that provide rebates for adoption of more efficient technology options such as lighting, and just the natural replacement of old appliances and commercial systems with new, more efficient equipment and thermal shell improvements. The current PJM model captures these energy efficiency impacts through incorporation of end-use saturation and efficiency projections. End-use saturation and average stock efficiency are derived from The Energy Information Administration (EIA) Annual Energy Outlook (AEO). The AEO forecast is based on a set of end-use models for the residential, commercial, and industrial sector. The AEO generates detail end-use information including number of households, number of end-use units, total consumption, and for many end-uses stock efficiency. Forecasts are generated for 9 census divisions as depicted in Figure 3-1.

FIGURE 3-1: FORECAST CENSUS DIVISIONS



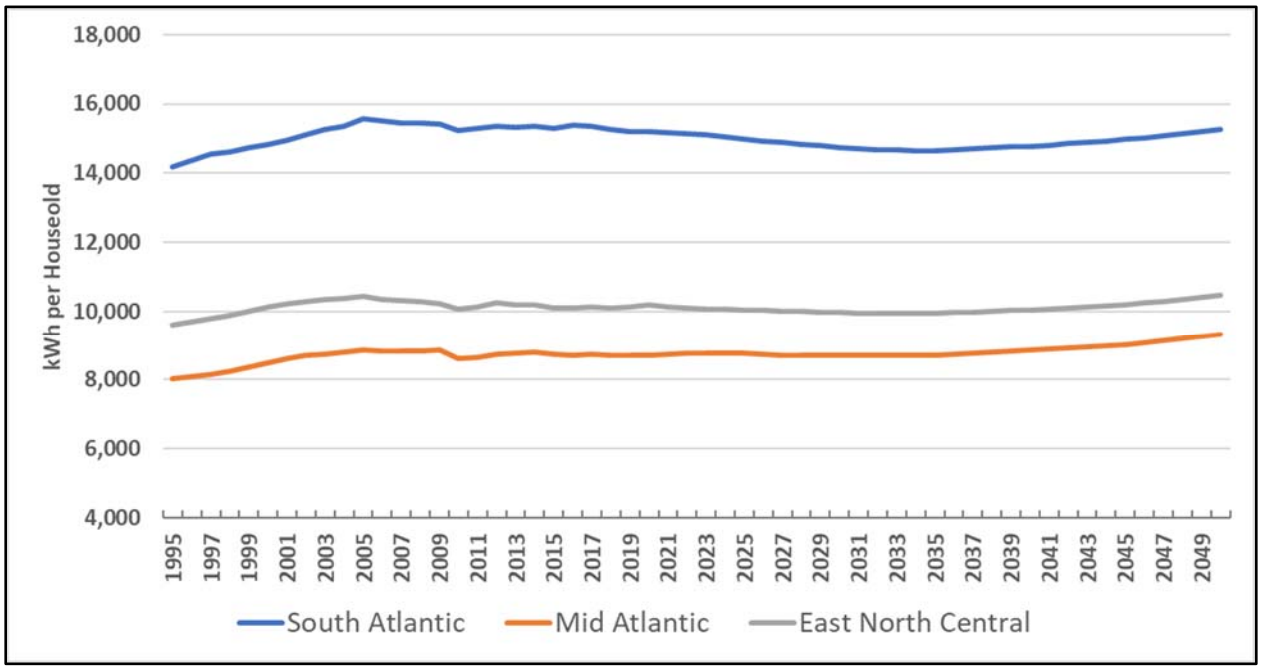
Each year, Itron processes the AEO forecast database and develops and maintains historical and projected end-use consumption, saturation, and efficiencies. Saturation and efficiency are used to construct end-use intensities. In the residential sector, intensities are on a kWh per household basis, and in the commercial sector on a kWh per square foot basis. At the national level, industrial intensities are calculated on a kWh per employee basis or kWh per $ GDP. The PJM models use intensities from the Mid-Atlantic, South Atlantic, East South Central, and East North Central Census Divisions; the specific set of intensities used in the zone model construction is determined by the zone's location. Figure 3-2 shows residential end-use intensities for the South Atlantic Census Division.

FIGURE 3-2: SOUTH ATLANTIC END-USE INTENSITIES



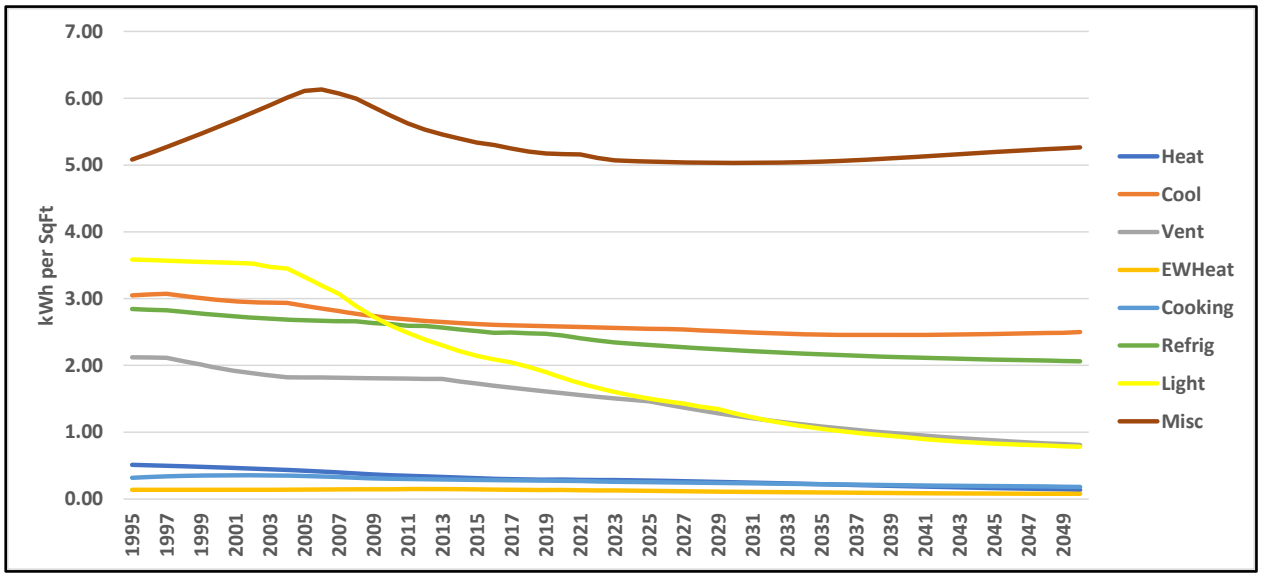
Like all Divisions, the miscellaneous category is the only end-use showing meaningful growth. Miscellaneous includes end-uses not specifically defined such as plug loads, pool pumps, spas, and wine coolers to name a few. In aggregate total intensity declines into through 2035; afterwards starts increasing as in the reference case, there are no new standards after that point in time. Figure 3-3 shows total residential intensity trend.

FIGURE 3-3: TOTAL RESIDENTIAL INTENSITIES



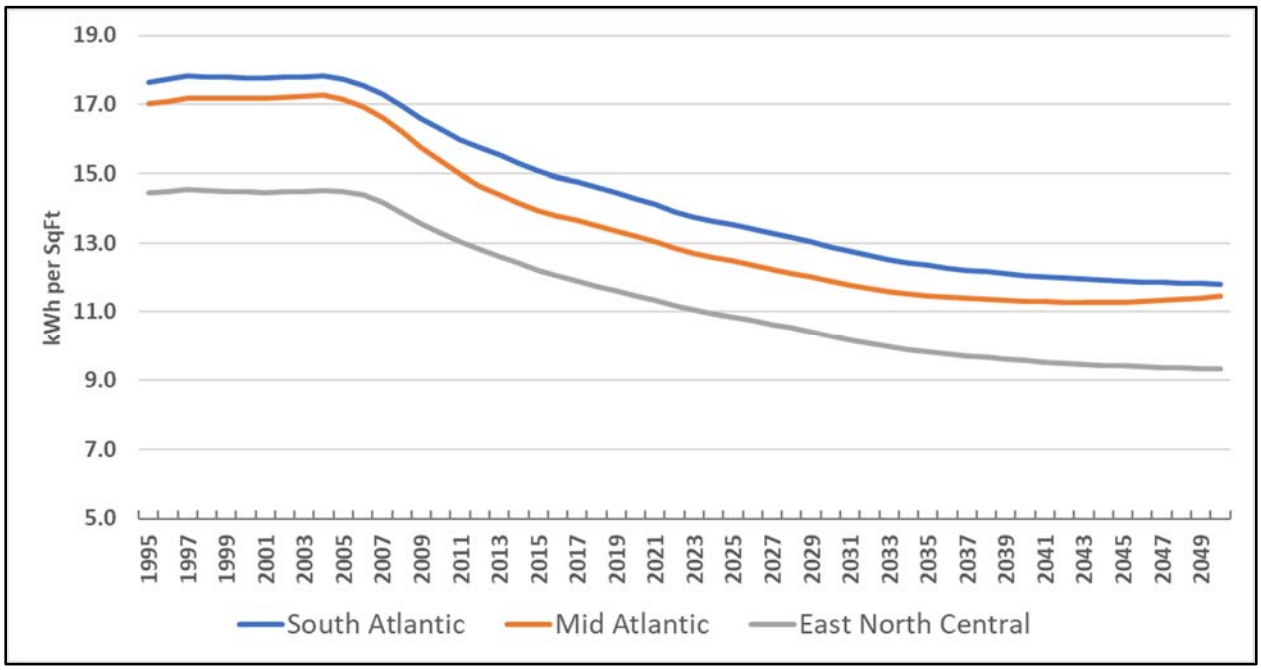
End-use intensities are also declining in the commercial sector. Figure 3-4 shows commercial end-use intensities.

FIGURE 3-4: SOUTH ATLANTIC COMMERCIAL END-USE INTENSITIES



While decline in end-use intensities are slowing across several end-uses, both lighting and ventilation are still expected to see strong overall efficiency improvements. Total commercial intensity has been declining since 2007 and is expected to continue to decline through the forecast horizon. Figure 3-5 shows total commercial intensities.

FIGURE 3-5: TOTAL COMMERCIAL INTENSITIES



TRANSITION FROM ANNUAL TO MONTHLY SECTOR MODELS

Heating Index (*HeatIdx)*, Cooling Index (*CoolIdx*), and Base Use Index (*BaseIdx*) are incorporated in the Zonal daily energy, peak, and coincident peak models. These indices are derived by combining population, economic growth, with end-use intensities allowing the indices to capture customer growth, economic activity, and current and expected gains in end-use efficiency. The construction of the indices is theoretically strong and effectively capture the impact end-use efficiency gains have had on regional sales. Resulting forecasts are consistent with EIA's long-term outlook.

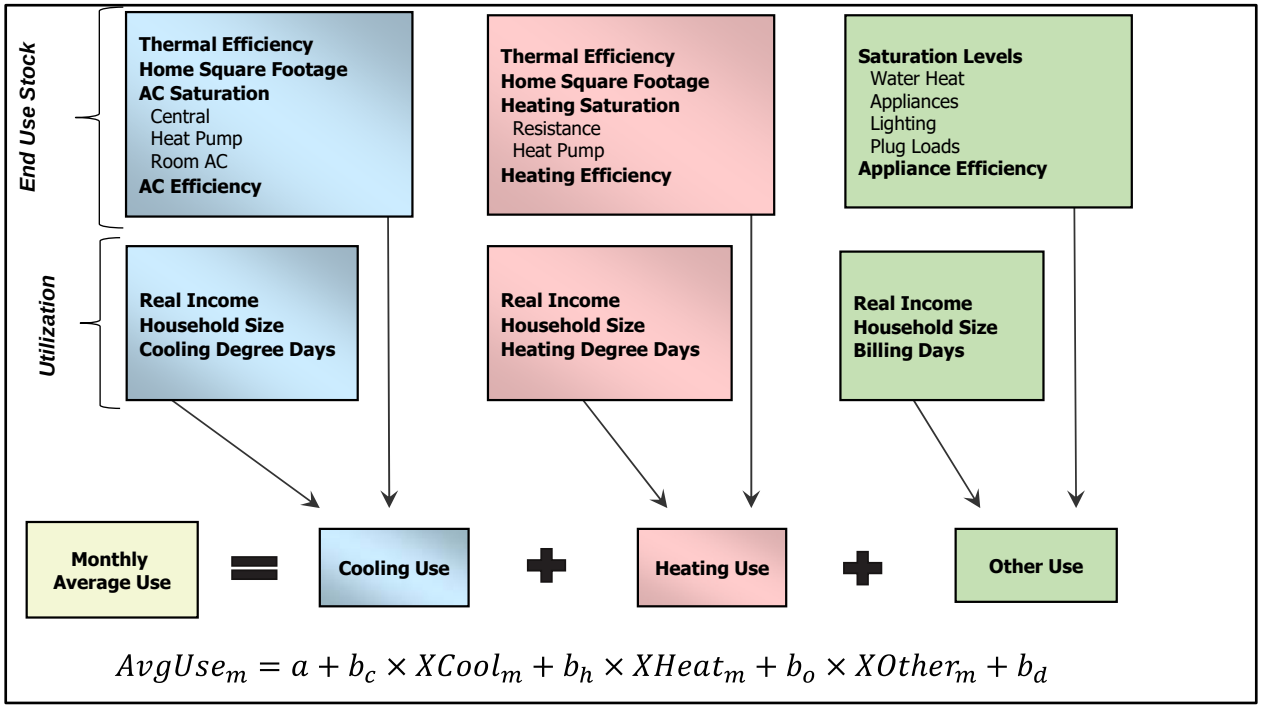
In the current PJM models, quarterly end-use indices are derived from annual customer class sales models (residential, commercial, and industrial). Annual models are estimated for each zone and reflect regional population and economic growth. While we believe the current approach adequately captures efficiency improvements, estimated indices can be strengthened and calculations simplified by deriving the indices from monthly rate class sales models rather than through annual models and calculation of quarterly indices. Monthly zone-level sales and customer can be derived from the *EIA 861 and 861m* reports. EIA 861 provides annual utility sales that includes both utility and retail served energy and customers. EIA 861m provides utility-supplied energy (which excludes retail delivered sales) that is used to shape total

annual sales to months. Monthly model provides more observations and greater variation in historical monthly use due to both weather and economic activity. This in turn results in stronger model coefficients both in terms of their impact (elasticity) and statistical strength. The use of monthly model also addresses Market Participants' concern as fewer years are required to estimate strong model coefficients used in constructing the model indices. Market Participants' issue was that using older annual data (which is necessary to have enough data points for modeling) may give too much weight to past years that are not reflective of current end-use characteristics. PJM sought to address this concern by shortening the estimation period to ten years but found they could not estimate reasonable models with just ten observations. Moving to monthly models addresses this issue; ten years of historical data gives 120 observations vs. 10 observations in an annual model. The shorter historical period will result in indices that are more reflective of current period end-use mix and economic activity impacts.

ESTIMATE RATE CLASS MODELS

The proposed approach is to construct end-use indices from monthly Statistically Adjusted End-Use (SAE) models for the residential and commercial sectors and more generalized model specification for the industrial sector. In the residential sector, the model is estimated for monthly average use and in the commercial sector for monthly sales. The SAE model is designed to capture economic growth as well as structural changes reflected in end-use saturation and efficiency trends and building shell improvements. Figure 3-6 shows the residential SAE average use model.

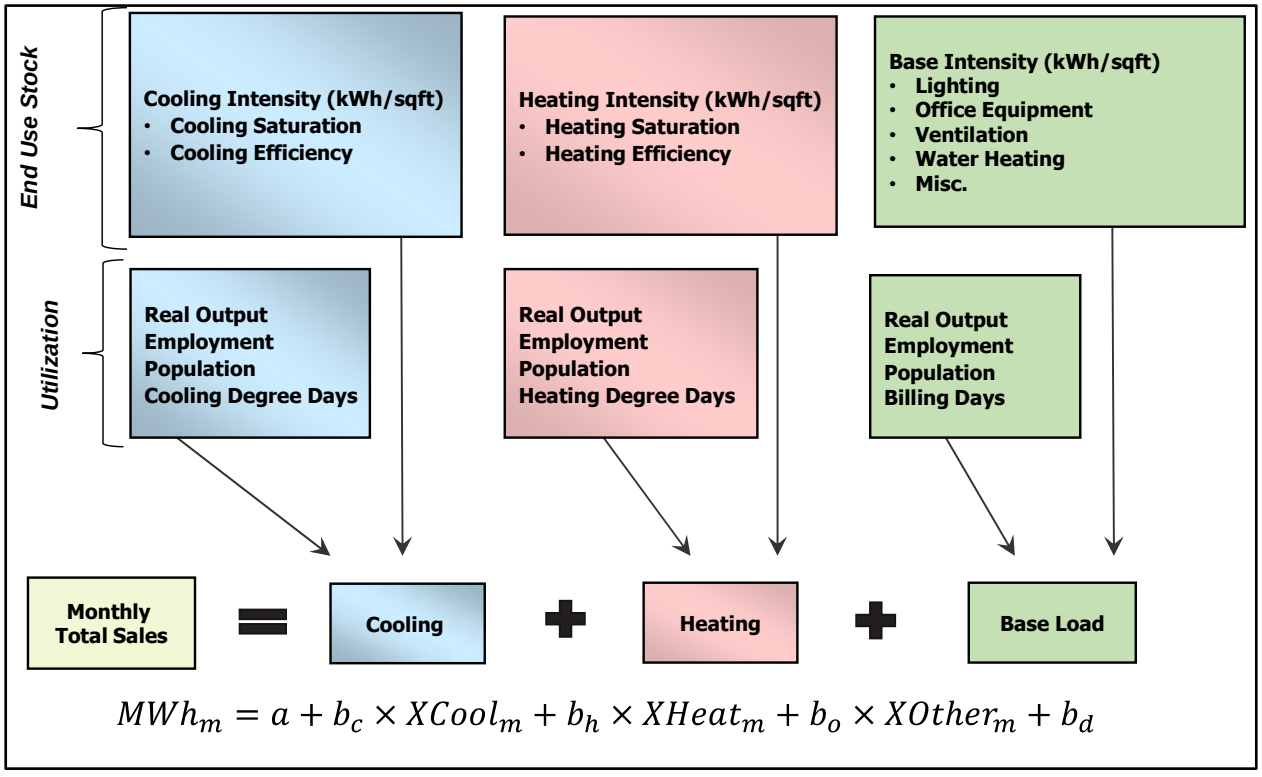
FIGURE 3-6: RESIDENTIAL SAE AVERAGE USE MODEL



The model variables *XCool*, *XHeat*, and *XOther* are initial estimates of average monthly cooling, heating, and non-weather sensitive use. The model variables are constructed by combining long-term annual saturation and efficiency trends (the end-use stock variables) with monthly weather, household income, and household size (the monthly utilization variables). The coefficients – b_c , b_h, and b_o are estimated using linear regression that relates average residential use to the constructed end-use model variable. The estimated model coefficients statistically adjust the end- use energy estimates to customer use and are used as describe below to estimate historical and projected heating, cooling, and base use loads. As parameters are estimated using an average use model, total residential heating, cooling, and baseloads are derived by multiplying per customer end-use sales by number of residential customers. Residential customer projections are based on a monthly linear regression model that relates customers to number of households.

A similar model can be estimated for the commercial sector where cooling, heating, and base-use variables are derived by combining commercial end-use energy intensity (measured in kWh per square foot) with regional economic drivers that capture both economic growth and short-term business activity. Figure 3-7 shows the commercial SAE model.

FIGURE 3-7: COMMERCIAL SAE SALES



In general, commercial models are estimated at total sales level. The estimated model coefficients calibrate commercial end-use energy estimates to actual commercial sales.

The monthly SAE specification was evaluated and presented to PJM staff using DPL Zone rate class monthly sales data, economic data, and EIA Census Division residential and commercial end-use intensities. Recent data developed by the National Renewable Energy Laboratory (NREL) can potentially be used to help calibrate the census-level saturation estimates to Zones. NREL has constructed detail residential (ResStock) and commercial (ComStock) end-use data bases derived from thousands of residential and commercial building simulations. 2018 end-use saturation and consumption can be calculated for each Public Use Microdata Area (PUMA). PUMA disaggregate states into regions with no fewer than 100,000 people. Figure 3-8 shows residential actual and predicted average use and Figure 3-9 shows DPL commercial sales model.

FIGURE 3-8: DPL RESIDENTIAL SAE MODEL

Variable	Coefficient	StdErr	T-Stat	P-Value
mStructResVar.XHeat	0.611	0.041	14.962	0.00%
mStructResVar.lag_XHeat	0.385	0.040	9.596	0.00%
mStructResVar.XCool	0.579	0.029	20.275	0.00%
mStructResVar.lag_XCool	0.271	0.028	9.580	0.00%
mStructResVar.XOther	0.694	0.016	42.501	0.00%

Model Statistics	
Iterations	1
Adjusted Observations	132
Deg. of Freedom for Error	127
R-Squared	0.899
Adjusted R-Squared	0.896
AIC	8.507
BIC	8.616
F-Statistic	#NA
Prob (F-Statistic)	#NA
Log-Likelihood	-743.74
Model Sum of Squares	5,411,707.85
Sum of Squared Errors	605,458.10
Mean Squared Error	4,767.39
Std. Error of Regression	69.05
Mean Abs. Dev. (MAD)	50.32
Mean Abs. % Err. (MAPE)	5.02%
Durbin-Watson Statistic	1.739
Durbin-H Statistic	#NA
Ljung-Box Statistic	21.44
Prob (Ljung-Box)	0.6124

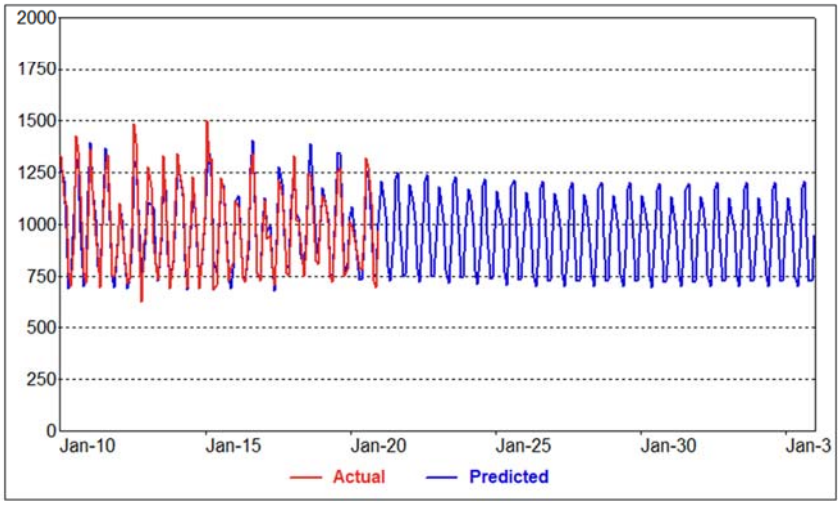
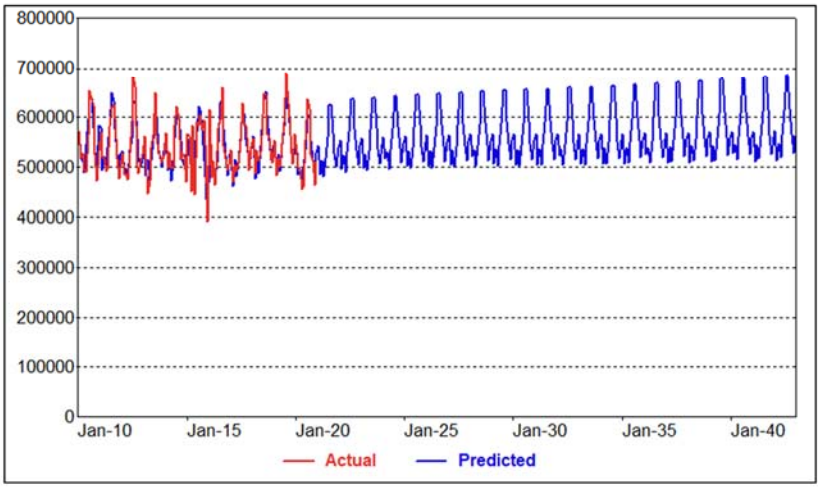


FIGURE 3-9: DPL COMMERCIAL SAE MODEL

Variable	Coefficient	StdErr	T-Stat	P-Value
mStructComVar.XHeat	603682.895	79802.942	7.565	0.00%
mStructComVar.XCool	169975.308	16510.469	10.295	0.00%
mStructComVar.lag_XCool	44790.922	15989.044	2.801	0.59%
mStructComVar.XOther	37659.345	441.829	85.235	0.00%
mBin.Yr2018Plus	11113.086	5608.705	1.981	4.97%
mBin.Dec2015	-107557.755	29095.540	-3.697	0.03%

Model Statistics	
Iterations	1
Adjusted Observations	132
Deg. of Freedom for Error	126
R-Squared	0.747
Adjusted R-Squared	0.737
AIC	20.58
BIC	20.71
F-Statistic	#NA
Prob (F-Statistic)	#NA
Log-Likelihood	-1,539.42
Model Sum of Squares	308,160,817,274.05
Sum of Squared Errors	104,181,634,515.42
Mean Squared Error	826,838,369.17
Std. Error of Regression	28,754.80
Mean Abs. Dev. (MAD)	19,750.20
Mean Abs. % Err. (MAPE)	3.67%
Durbin-Watson Statistic	2.350
Durbin-H Statistic	#NA
Ljung-Box Statistic	30.64
Prob (Ljung-Box)	0.1643
Skewness	0.077
Kurtosis	5.269
Jarque-Bera	28.440
Prob (Jarque-Bera)	0.0000



The models generate statistically significant end-use coefficients that can be used to generate strong estimates of cooling, heating, and base-use energy requirements. Another strength is that given the structure, the SAE model generally results in consistent estimated parameters when there are just small changes in model parameters as new data is added and older data dropped.

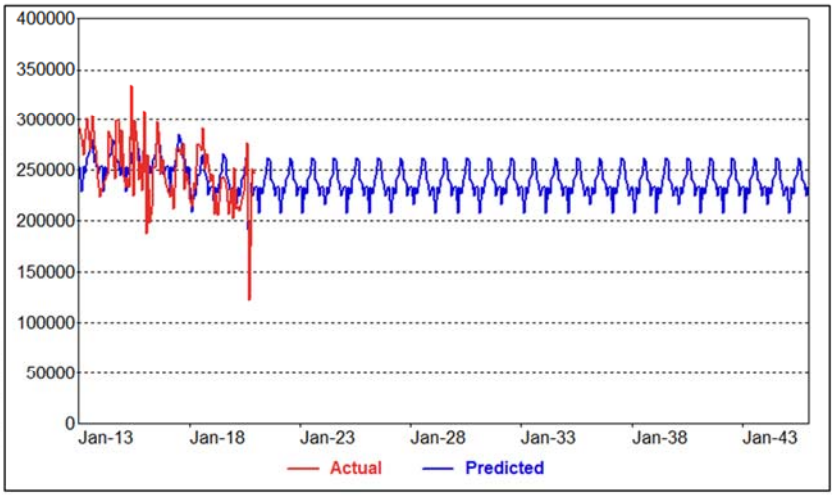
Similar structured models can be estimated for the industrial sector. To that end, PJM constructed industrial models that leverage the EIA industrial sector energy forecasts. In the industrial sector, EIA generates energy per dollar output projections by industrial sector. Using a combination of state-level output estimates and regional sector employment data, PJM constructs annual industrial indices that reflect Zonal industrial sector real output, productivity, and efficiency improvements. Itron has developed similar inputs for industrial modeling and believe that the PJM current industrial model process is theoretically strong. While structured models have worked in some regions, often they don't as in many service areas monthly industrial sales are not "well-behaved"; in these cases, PJM will need to use reasonable judgment in constructing monthly sales models and may require exogenous adjustments based on expected activity from large industrial customers. Industrial sales are often dominated by a few large customers; there can be significant month-to-month sales variation as a result of large swings in individual customer usage and large billing adjustments.

This is illustrated in Figure 3-10 where we could not estimate a reasonable DPL industrial sales model. Not only are there large monthly sales swings that are not weather-related, but there is a drop in 2018 sales that could be the result of one or more large customer closing down part or all of their operations.

FIGURE 3-10: DPL INDUSTRIAL SALES MODEL

Variable	Coefficient	StdErr	T-Stat	P-Value
mWthrCal.Days	8182.112	123.024	66.508	0.00%
mWthrCal.CDD60	65.106	14.350	4.537	0.00%
mBin.Yr2018Plus	-20698.365	5168.287	-4.005	0.01%
mBin.Sept20	-119005.506	24678.138	-4.822	0.00%

Model Statistics	
Iterations	1
Adjusted Observations	96
Deg. of Freedom for Error	92
R-Squared	0.431
Adjusted R-Squared	0.413
AIC	20.24
BIC	20.34
F-Statistic	#NA
Prob (F-Statistic)	#NA
Log-Likelihood	-1,103.56
Model Sum of Squares	41,138,783,394.60
Sum of Squared Errors	54,275,336,134.38
Mean Squared Error	589,949,305.81
Std. Error of Regression	24,288.87
Mean Abs. Dev. (MAD)	18,817.72
Mean Abs. % Err. (MAPE)	7.55%
Durbin-Watson Statistic	1.913
Durbin-H Statistic	#NA
Ljung-Box Statistic	35.85
Prob (Ljung-Box)	0.0567
Skewness	0.182
Kurtosis	3.263
Jarque-Bera	0.807
Prob (Jarque-Bera)	0.6679



Without any strong identifiable model drivers, the forecast is held constant at current levels of activity. Generally, the industrial sales forecast would then be adjusted for expected large customer load additions or load losses. While limited in terms of information, model coefficients can be still be used to isolate industrial cooling and base-use contribution to system energy requirements.

SYSTEM END-USE MODEL INPUT CONSTRUCTION

The next step is to isolate long-term system heating, cooling, and other-use energy requirements for both the historical and forecasted periods. End-use energy requirements are estimated by multiplying the end-use model coefficients with the constructed XHeat and XCool for normal weather and the XOther model variables. Normal or expected weather conditions are used in the XHeat and XCool index calculations to avoid double counting weather impacts. Weather impacts are captured in the daily or hourly zonal load weather simulations. Figure 3-11 shows the derived residential heating, cooling, and base-use loads,

Figure 3-12 the commercial and industrial monthly energy requirements, and Figure 3-13 total system cooling, heating, and baseload requirements.

FIGURE 3-11: DPL RESIDENTIAL END-USE ENERGY REQUIREMENTS (MWH)

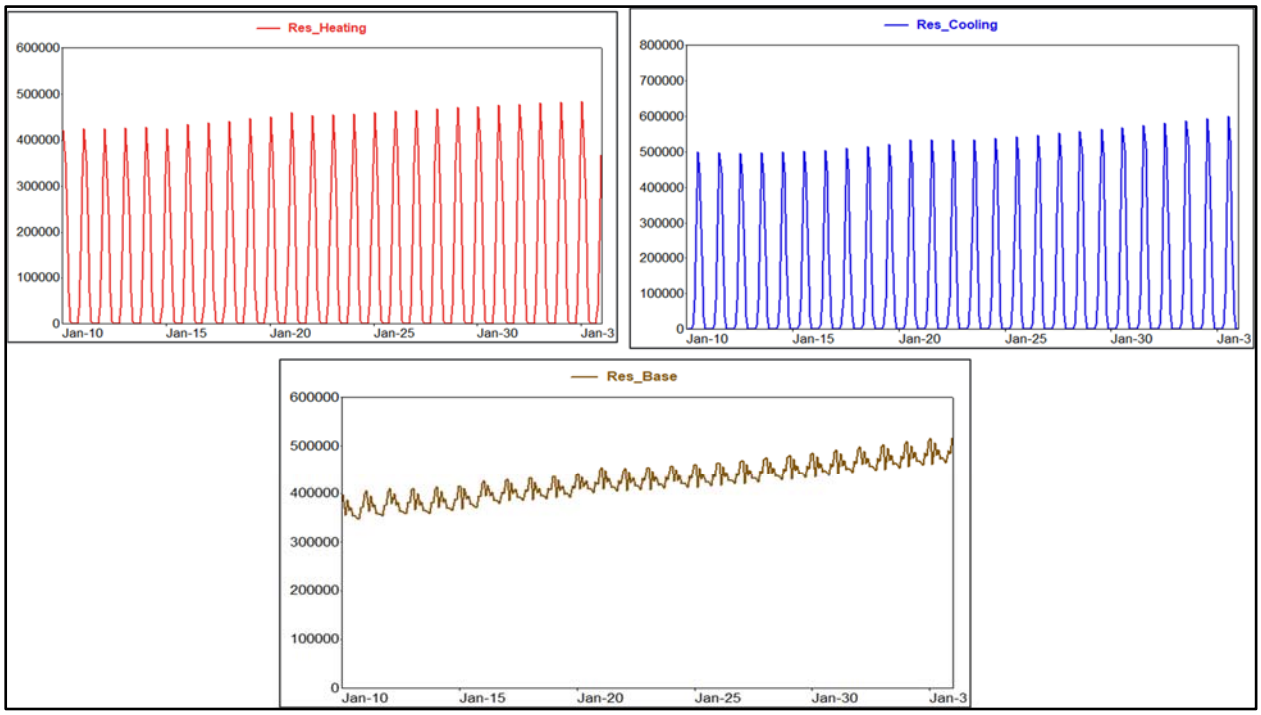


FIGURE 3-12: DPL C&I END-USE ENERGY REQUIREMENTS (MWH)

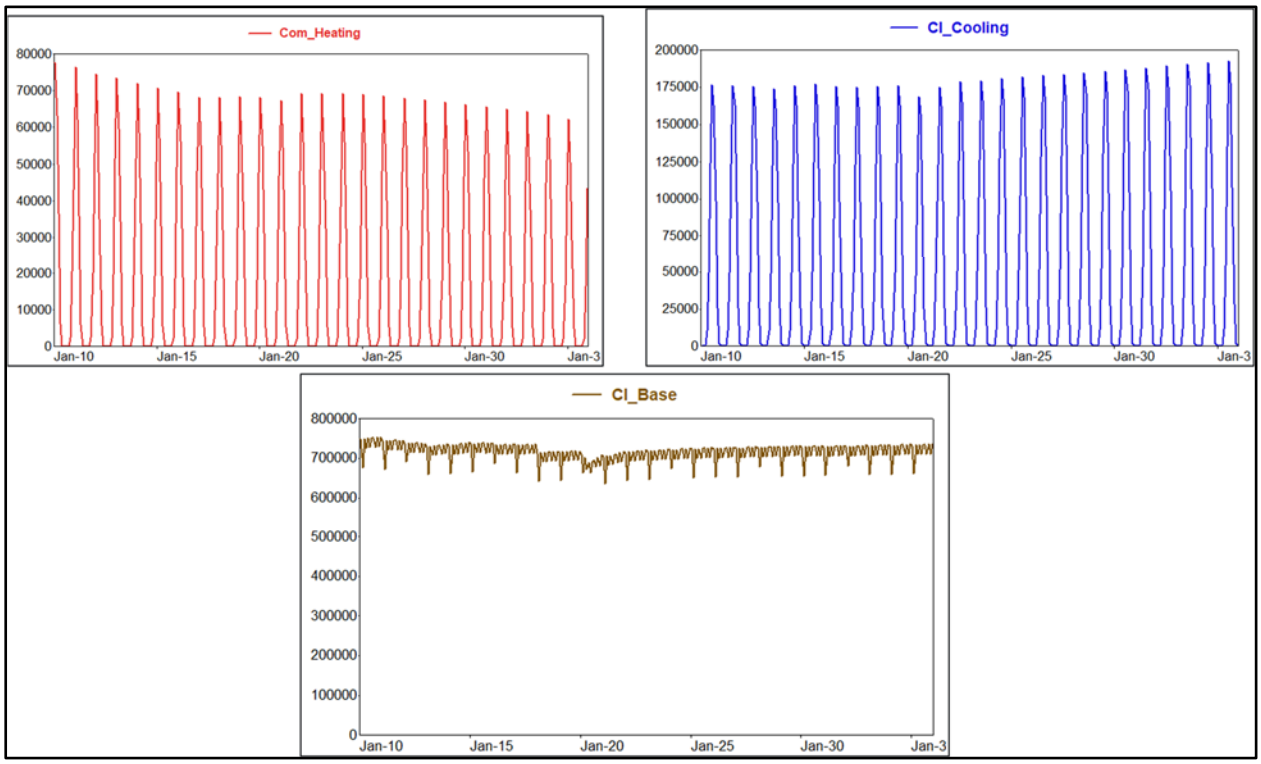
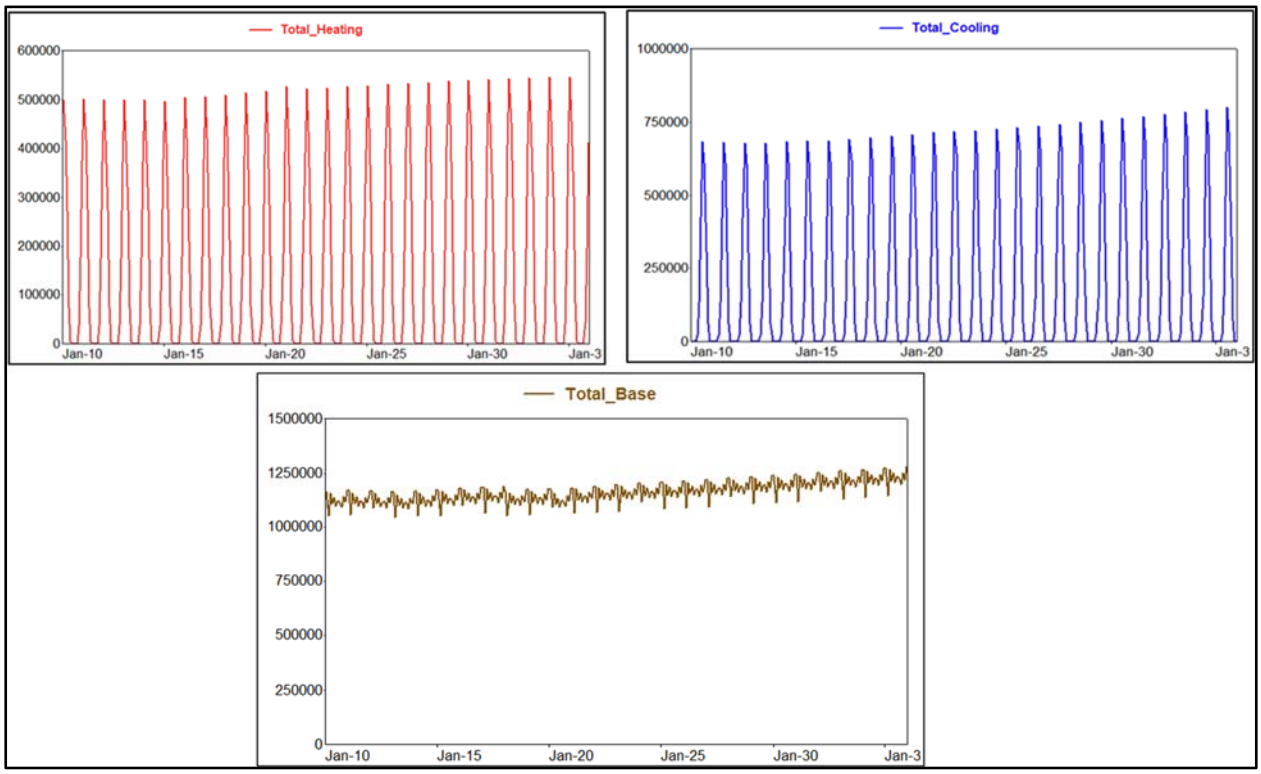
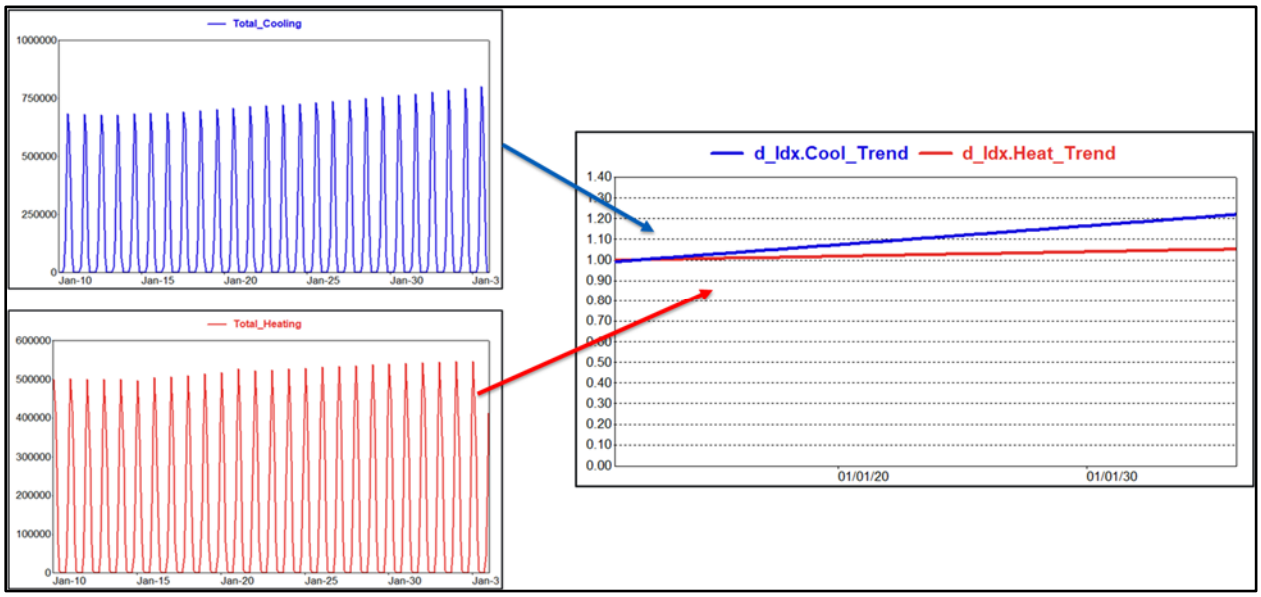


FIGURE 3-13: DPL TOTAL END-USE ENERGY REQUIREMENTS (MWH)



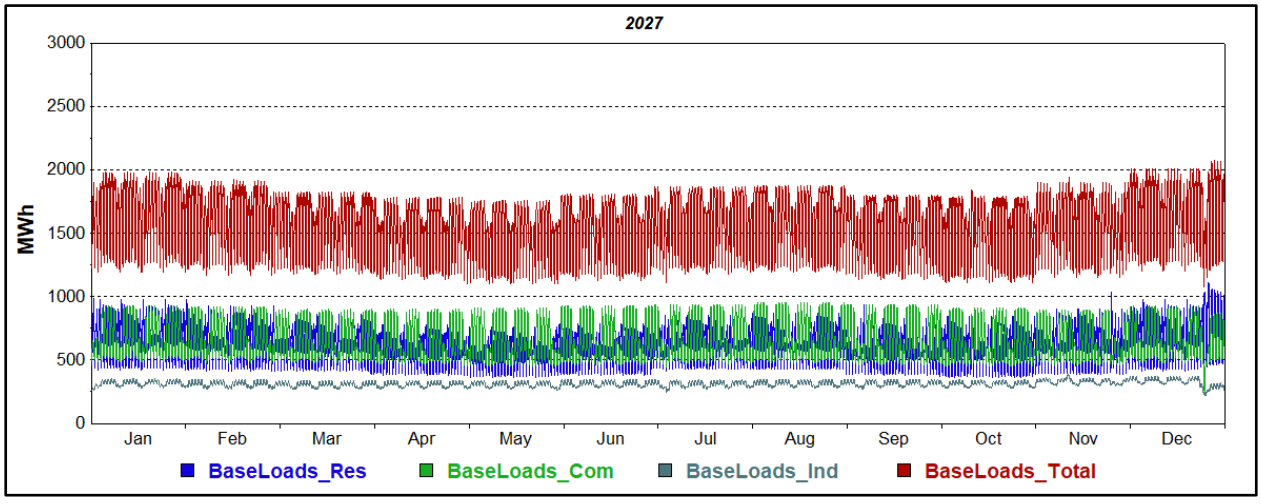
Calculated heating, cooling, and base-use energy requirements are converted to daily indices that are then incorporated into the Zonal hourly load models. *HeatIdx* is combined with the hourly model heating-related weather variables, *CoolIdx* with cooling weather variables, and *BaseIdx* with seasonal and daily binary variables. Figure 3-14 shows the conversion of monthly end-use energy requirements to *HeatIdx* and *CoolIdx*.

FIGURE 3-14: DPL DAILY INPUTS FOR HOURLY ZONAL MODELS



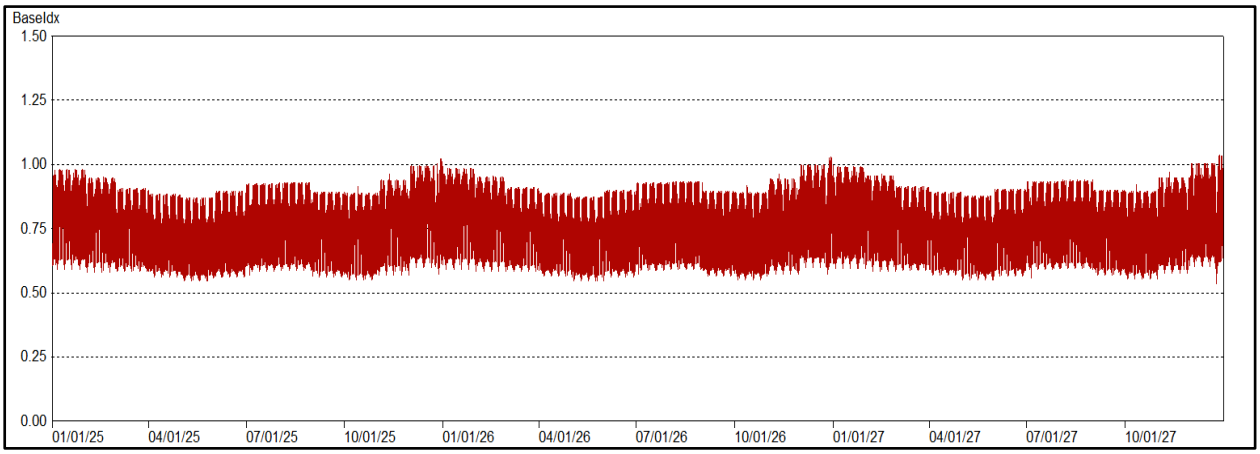
Base-use energy requirements are also converted to a daily index and interacted with hourly model binaries designed to capture non-weather sensitive load including day of the week, holiday, and seasonal daily variables. Alternatively, AMI data or end-use profiles available from NREL could be used to convert customer class base-use energy requirements to hourly load estimates. Figure 3-15 shows estimated DPL hourly base-use energy requirements by class and total for 2027.

FIGURE 3-15: DPL HOURLY BASE-USE ENERGY REQUIREMENTS (2027)



Total base-use load requirements are indexed and used in estimating the hourly Zonal models. Figure 3-16 shows the hourly BaseIdx variable for 2025 through 2027.

FIGURE 3-16: DPL BASEIDX 2025 - 2027



The integration of the end-use indices into the hourly models is discussed in the Hourly Model Analysis and Recommendations section.

4 HOURLY MODEL ANALYSIS AND RECOMMENDATIONS

The current PJM modeling process includes three sets of daily models for each zone. These are daily energy, daily noncoincident zone peak (NCP), and daily CP (zone load at the time of the daily PJM peak). For each zone, CP models are estimated for daily loads coincident with the Locational Deliverability Area (LDA) peak and coincident with the overall PJM system peak.

The application of this approach is complicated by the saturation of new technologies (PV and EV in particular) that modify the timing of system energy requirements. In particular, the penetration of behind the meter solar systems is changing both the timing of zone peaks and the coincidence factors across zones. The best way to understand the impact of these changes is through an hourly modeling process. This section shows how the hourly modeling process would work, provides an example of the level of accuracy that can be achieved, and provides recommendations about the construction of hourly weather variables.

The examples are developed using hourly load and weather data for the DPL (Delmarva) zone. As part of this project, Itron did not evaluate model performance for any other PJM zones. However, in the past we have implemented similar models and performed similar analysis for a wide variety of utilities, including many of those in the PJM footprint. Based on this experience, we believe that the conclusions presented below will generalize well to the other PJM zones.

The general conclusions of this section are:

- PJM has the data to develop accurate and robust hourly models
- Hourly models are well suited to the weather simulation approach
- Hourly models can accurately model system zone peaks (NCP) and the coincident load (CP) values.
- The conclusions include specific recommendations for construction of interactive weather variables.

EXAMPLE OF HOURLY LOAD DATA

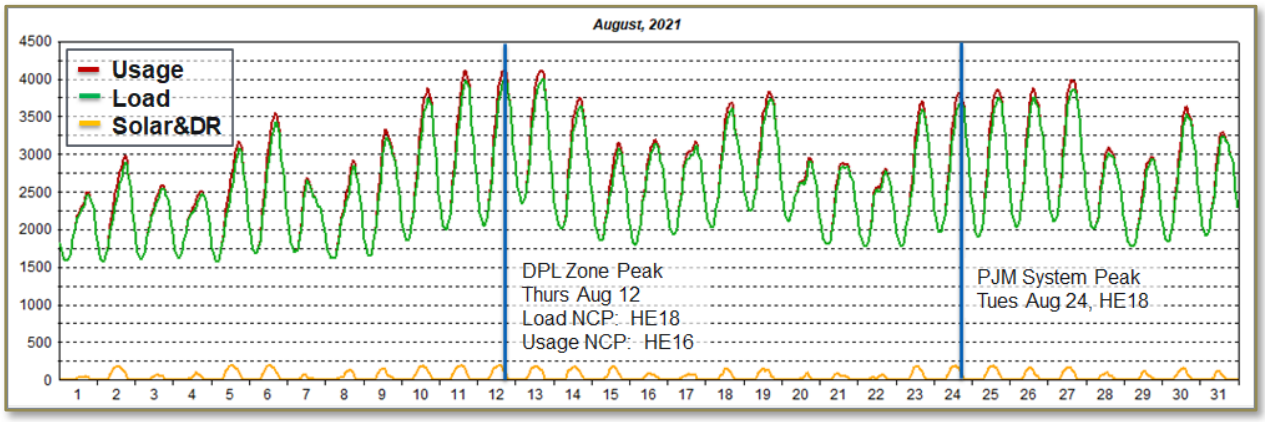
Figure 4-1 shows an example of hourly load data for the DPL (Delmarva) zone. The chart shows hourly data for one month, August 2021. The top line (Red) is labeled Usage and shows the estimate of hourly energy usage (customer end-use consumption). The Green line is labeled Load and shows the level of zone resources dispatched to meet customer energy requirements. The bottom line (Orange) shows the estimated level of behind-the-meter solar generation.

The utility load is computed as the measured hourly zone load plus estimated hourly impacts of DR programs. (Note, there were no estimated DR impacts in the DPL zone in the month displayed). The orange line (Solar) is estimated and is based on hourly solar irradiation data and installed capacity of behind-the-meter solar panels. The red Usage line is calculated as the sum of utility load (Green) and

solar Generation (Orange). This is an estimate of total customer energy usage, including energy supplied by BTM solar systems and energy supplied by the zone resources.

In the chart, the two vertical blue lines represent the day and time of the DPL zone peak (on August 12) and the PJM system peak (on August 24). On the zone peak (NCP) day, customer usage peaked at HE17 (4 pm to 5 pm), but the peak utility peak load occurred in the following hour (HE18) as BTM solar production ramped downward. It was a clear day, so this is just part of the normal ramp-down process for solar generation and the resulting ramp-up process for utility loads in the late afternoon. As a point of reference, peak solar generation on this day was estimated to be 197 MW.

FIGURE 4-1: DPL HOURLY USAGE, LOAD, AND BTM SOLAR – AUGUST 2021



DPL Data	Usage Peak 8/12/HE17	NCP 8/12 HE18	CP 8/24 HE18	Coincidence Factor
Usage	4,106.9	4,100.3	3,753.6	91.4%
Solar_DR	132.9	94.3	83.6	
DPL Load	3,974.0	4,006.0	3,670.0	91.6%

The right-hand blue line on August 24th represents the PJM System peak load. Although not shown in the grid above, customer usage on this day peaked in HE16, load peaked in HE17, and the PJM peak is at HE18. Customer usage is estimated to have declined by about 73 MW over this three-hour window, but solar generation values ramped down by about the same amount (from 156 MWh to 125 MWh to 84 MWh, respectively), and as a result, net DPL load was relatively flat from the time of the customer peak to the time of the PJM peak.

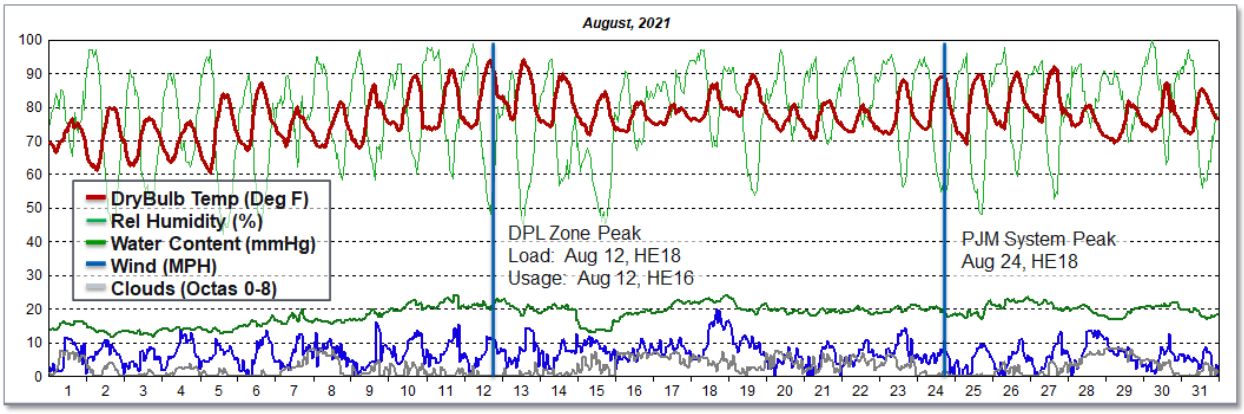
The grid below the chart pulls together the data for the hour of usage peak, the hour of DPL peak load (NCP) and the DPL load at the time of the PJM system peak (CP). The coincidence factor for usage (CP usage/Peak usage) is estimated to be 91.4%. The coincidence factor for DPL load (CP load /NCP load) is 91.6%. This is well below the typical coincidence values for the DPL area but is within the boundaries of coincidence results generated by the weather simulation approach with rotation.

EXAMPLE OF HOURLY WEATHER DATA

Figure 4-2 shows an example of weather for the DPL (Delmarva) zone. The chart shows hourly values for the same month as shown above, August of 2021. The weather data values are computed by averaging hourly data from the two weather stations, Wilmington, Newcastle (ILG) in the north and Wallops Island (WAL) in the south; based on associated population, ILG has a 70% weight and WAL a 30% weight. The red line shows data for hourly drybulb temperature. The light green line shows data for relative humidity. The dark green line shows hourly data for the moisture content of the air in millimeters of mercury (mmHg). The blue and gray lines at the bottom are for Wind speed in MPH and Cloud Cover in Octas (ranging from 0 to 8).

As with the hourly load chart, the hours for the DPL zone peak and the PJM system peak are marked with vertical blue lines. The hourly weather data for these two peak hours is displayed below the chart.

FIGURE 4-2: HOURLY WEATHER DATA FOR DPL ZONE – AUGUST 2021



Summer Peak Hour Weather Variables	DPL 2021 NCP 8/12 HE18	PJM 2021 Peak 8/24 HE18
AvgDB Temperature (Deg F)	92.7	88.8
Relative Humidity (%)	54.3	53.3
Moisture Content (mmHg)	21.3	19.1
Temp Hum Index (Deg F)	95.4	91.1
Wind Speed (MPH)	11.4	8.4
Cloud Cover (Octas)	0.12	1.6

The peak hour data show the main difference between the two days, which is that the DPL zone is four degrees warmer on the NCP day than on the peak day. The THI variable that is used in the model is 4.3 degrees higher. Estimated equations indicate that 4 additional degrees of temperature in HE18 and the preceding hours adds about 290 MWh to the load in HE18. This explains why the DPL peak on August 12 is so far above the DPL CP value on August 24.

While we are on the topic of weather, a brief comment about the THI variable and relative humidity. The normal daily cycles are clear in the weather chart. Each day, the temperature increases to a mid-day high and then decreases at night to lower values. As this cycle occurs, the amount of water that the air can hold changes significantly, since warmer air can hold much more water than colder air. Relative humidity represents ratio of the amount of moisture in the air to the amount of moisture the air can hold. The amount of moisture in the air (the numerator) remains relatively constant over most days, but as the temperature drops at night the amount of water the air can hold (the denominator) goes down, and therefore relative humidity goes up. This continues until the air temperature hits the dew point temperature, at which point relative humidity approaches 100%.

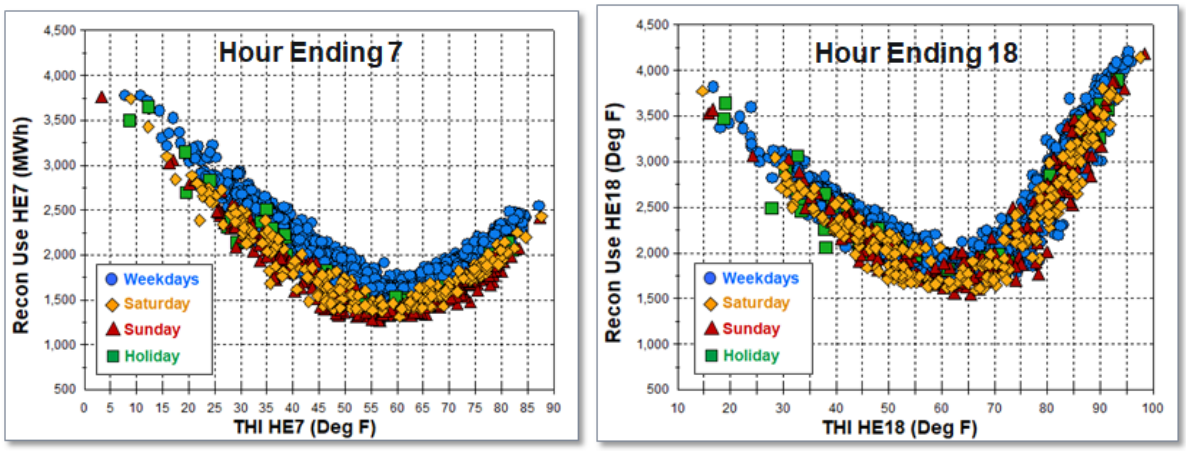
The dark green line shows a measurement that is proportional to the numerator in the relative humidity equation. It measures the absolute amount of moisture in the air in each hour. Unlike relative humidity, this water content measurement does not change much during a day. In the work below, we discuss the THI formula based on Relative Humidity. It may, however, make sense to research alternatives that utilize a more stable measurement like moisture content or dewpoint temperature.

HOURLY MODELS

This section provides an analysis of alternative approaches to constructing hourly models. Itron has studied a wide range of modeling methods including a variety of regression approaches (least squares, ridge, lasso, elastic net, quantile, support vector), decision tree regressors, and neural network models. While neural network models typically have the best estimation period accuracy, they are more complex than regressions, and can be problematic when explanatory variables move far outside their historical range, as is likely the case in a long-term forecast. We have found that least squares regression (usually called ordinary least squares or OLS) works as well as any of the alternatives in terms of out-of-sample accuracy and provide a robust approach for long-run forecasting of daily and hourly energy usage. Based on these past findings, we restrict our attention here to OLS methods.

The typical approach for long-run forecasting is to specify 24 independent structural models, with one model for each hour. The idea is illustrated in Figure 4-3 that shows scatter plots of loads at HE07 (the hour from 6 am to 7 am) and HE18 (the hour from 5 pm to 6 pm). In these charts, the height of each point represents the estimated hourly usage for the selected hour on a specific day. The position on the X axis is determined by the corresponding hourly temperature (THI in this case) on each day. The points are color coded by day of the week. Each chart shows data for mid-2017 to mid-2021, providing about 1,500 observations for each hour.

FIGURE 4-3: SCATTER PLOTS FOR HE07 AND HE18 – 2017 TO 2021



While temperature and day-of-week variables clearly explain much of the variation in loads in an hour, a variety of other factors are also important. This includes zone growth in terms of customers and economic activity, other weather variables (prior hour temperatures, humidity, wind speed, and cloud cover), seasonal factors (hours of daylight, sun angles, ground water temperatures), changes in end-use saturation and efficiency, and exogenous shocks, such as phases of the Covid pandemic.

Proper construction of explanatory variables that capture the important nonlinearities and interactions among these factors is the key to a strong and robust explanatory model. In what follows, most of the focus is on weather variable definitions and interactions since this will be central to the effectiveness of the weather simulation approach.

TRANSFORMING HOURLY WEATHER

The first analysis concerns the granularity of weather variables that are used in the hourly modeling. Data used in the analysis ranged from mid-2017 through early 2022, giving a total slightly over 1,700 observations for each hour. The hourly Y variable in this analysis is utility load (in later sections, the final model evaluations were performed using hourly usage).

Four sets of models were constructed and compared in terms of in-sample accuracy. All four sets have a common set of calendar-based variables, including the following:

- Monthly binary variables
- Monthly trend variables to capture within-month trends
- Holiday variables for individual holidays
- Long weather lags (10 day and 28-day CD and HD)
- Covid variables Phase 1 to Phase 4 to capture net covid impacts
- Time trend variable

The four sets of models differ in the way hourly weather is transformed into explanatory factors. The approaches are:

1. **Daily Average**. With this approach, hourly values are averaged for each day and the daily average variables are used to construct explanatory factors.
2. **Time-of-Day Average**. With this approach, hourly values are averaged within four time of day blocks (night, morning, afternoon, evening). The TOD averages are then used to construct explanatory factors.
3. **Rolling Hourly with Splines.** With this approach weather variables are computed for each hour as the centered moving average of the prior hour, current hour, and next hour values. The rolling hourly variables are used to construct heating degree and cooling degree variables using 5 degree temperature ranges. For example, THI degrees are computed using base values of 60, 65, 70, 75, 80 and 85. These variables are then used to construct a single heating degree spline and cooling degree spline based on temperature derivatives computed using hourly neural network models. The spline variables are then interacted with other calendar and weather variables to capture slope shifts and nonlinear interaction effects.
4. **Rolling Hourly with Two-Part Degree Days**. With this approach, two cooling degree variables and two heating degree variables are computed for each hour. For example, for HE18, cooling degrees are computed using base temperatures of 70 and 80 and heating degrees are computed using base temperatures of 60 and 50. The first heating degree variable (HD1) and the first cooling degree variable (TD1) are then interacted with other calendar and weather variables to capture slope shifts and nonlinear interaction effects.

To evaluate the relative power of these approaches, a full set of hourly models is estimated using each approach. The mean absolute percent error (MAPE) values from these models are used for the comparison. For each hour, this statistic is computed as the average of the absolute values of the estimated model errors for that hour.

Daily Average Models. For the Daily Average model, daily weather values are computed by averaging the hourly values across the hours in each day. Steps in the process are:

Hourly drybulb temperatures and hourly humidity values are used to compute hourly THI values.

Hourly values are then averaged for each day to get a daily average values (AvgDB, AvgTHI, AvgWind, AvgClouds).

- Two cooling degree variables are computed from AvgTHI using base temperatures of 60 and 70 (AvgTD1 and AvgTD2).
- Two heating degree variables are computed from AvgDB using base temperatures of 60 and 50 (AvgHD1 and AvgHD2)
- AvgTD1 and AvgHD1 are lagged one day and are also interacted with calendar and other weather variables:
    - TD1 and HD1 lagged one day
    - Cold Clouds (Clouds interacted with HD base 50)

- Warm Clouds (Clouds interacted with CD base 70)
- Cold Wind (Wind interacted with HD base 50)
- Warm Wind (Wind interacted with CD base 70)
- SpringTHI and FallTHI (Seasonal binaries interacted with TD1)
- SpringHD and Fall HD (Seasonal binaries interacted with HD1)
- WkEndTHI, WkEndHD (Weekend binary interacted with TD1 and HD1)
- TrendTD1 and TrendHD1 (Trend variable interacted with TD1 and HD1)

Time of Day (TOD) Models. For the Time-of-Day models, hourly weather values are averaged for the four quarters of the day: Night (HE01 to HE06), Morning (HE07 to HE12), Afternoon (HE13 to HE18) and Evening (HE19 to HE24).

- Hourly drybulb temperatures and hourly relative humidity values are used to compute hourly THI values.
- Hourly values are then averaged for each TOD block to get an average variable values (AvgDB), AvgTHI, AvgWind, and AvgClouds for each block.
- Two cooling degree variables are computed for each block using base temperatures that vary across blocks (60 and 70 for Night, 65 and 75 for Morning, Afternoon and Evening)
- Two heating degree variables are computed for each block using base temperatures that vary across blocks (55 and 45 for Night and Evening, 60 and 50 for Morning and Afternoon).
- Lag and interaction variables similar to those in the Daily Average model are then constructed for each TOD block.

For each hour, the TD1 and HD1 variables are included for the block the hour falls in as well as for the prior block and the following block. This surrounding block values are included to improve performance in the hours near the block boundaries.

Rolling Hourly Models. For the Rolling hourly models, weather variables are computed for each hour. Moving from one hour to the next, all calculations shift one hour to the right, including current hour variables and lagged hour variables. Steps in the process are:

For each weather variable (temperature, humidity, clouds, and wind), the hourly value is computed as a centered moving average of the values for the prior hour, the current hour, and the next hour. In test estimations, this three-hour average (MAPE = 2.29%) worked a bit better than a two-hour average (MAPE = 2.32%), which worked a bit better than the current hour reading alone (MAPE = 2.36%).

Starting with the 3-hour average for each hour, drybulb temperatures and relative humidity values are used to compute THI values for each hour.

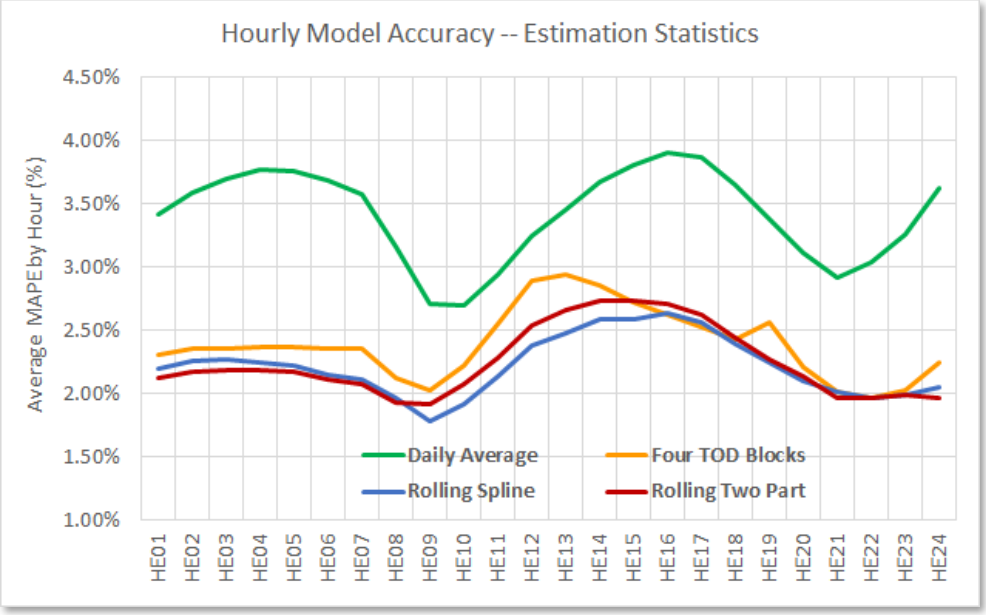
Rolling Spline Models. Starting with the 3-hour average drybulb values for heating and THI values for cooling, a full set of HD and TD variables are computed with 5 degree buckets. Base temperatures for HD are 65, 60, 55, 50, 45, 40, and 30. Base temperatures for TD are 55, 60, 65, 70, 75, 80, 85, 90.

HD and TD values in successive buckets are combined into a single HDSpline variable and a single TDSpline variable.

Rolling Two-Part Models. Starting with the 3-hour average drybulb values for heating and THI values for cooling, HD values are computed for each hour with two base temperatures (HD1 and HD2) and TD values are computed for each hour with two base temperatures (TD1 and TD2). The base temperature values are all at 5-degree points, and vary by hour, as shown in Figure 4-4.

Lag and interaction variables similar to those in the Daily Average model are then constructed for each hour of the day, using the 3-hour average weather variables.

FIGURE 4-4: ESTIMATION HOURLY MAPE STATISTICS – COMPARISON OF METHODS



Hour	DailyAvg Wthr	Four TOD Blocks	Rolling Spline	Rolling Two Part
HR00	3.42%	2.31%	2.20%	2.12%
HR01	3.59%	2.36%	2.26%	2.17%
HR02	3.69%	2.36%	2.27%	2.19%
HR03	3.77%	2.37%	2.24%	2.18%
HR04	3.76%	2.37%	2.22%	2.17%
HR05	3.68%	2.35%	2.15%	2.11%
HR06	3.57%	2.35%	2.11%	2.08%
HR07	3.16%	2.12%	1.96%	1.93%
HR08	2.71%	2.03%	1.78%	1.91%
HR09	2.69%	2.22%	1.92%	2.08%
HR10	2.94%	2.55%	2.14%	2.28%
HR11	3.24%	2.89%	2.38%	2.54%
HR12	3.45%	2.94%	2.48%	2.66%
HR13	3.67%	2.86%	2.59%	2.73%
HR14	3.81%	2.72%	2.59%	2.73%
HR15	3.90%	2.62%	2.64%	2.71%
HR16	3.86%	2.53%	2.56%	2.62%
HR17	3.65%	2.43%	2.39%	2.44%
HR18	3.38%	2.56%	2.25%	2.27%
HR19	3.11%	2.21%	2.10%	2.14%
HR20	2.92%	2.01%	2.01%	1.96%
HR21	3.04%	1.96%	1.97%	1.96%
HR22	3.26%	2.02%	1.99%	1.99%
HR23	3.62%	2.24%	2.05%	1.97%
Avg	3.41%	2.39%	2.22%	2.25%
#Coef	78	75	79-81	82

Two-part TD & HD base temperatures

Hour	TD1	TD2	HD1	HD2
HE01	65	75	55	45
HE02	65	75	55	45
HE03	65	75	55	45
HE04	65	75	55	45
HE05	65	75	55	45
HE06	65	75	55	45
HE07	65	75	55	45
HE08	65	75	55	45
HE09	65	75	55	45
HE10	65	75	55	45
HE11	70	80	60	50
HE12	70	80	60	50
HE13	70	80	60	50
HE14	70	80	60	50
HE15	70	80	60	50
HE16	70	80	60	50
HE17	70	80	60	50
HE18	70	80	60	50
HE19	70	80	60	50
HE20	65	75	55	45
HE21	65	75	55	45
HE22	65	75	55	45
HE23	65	75	55	45
HE24	65	75	55	45

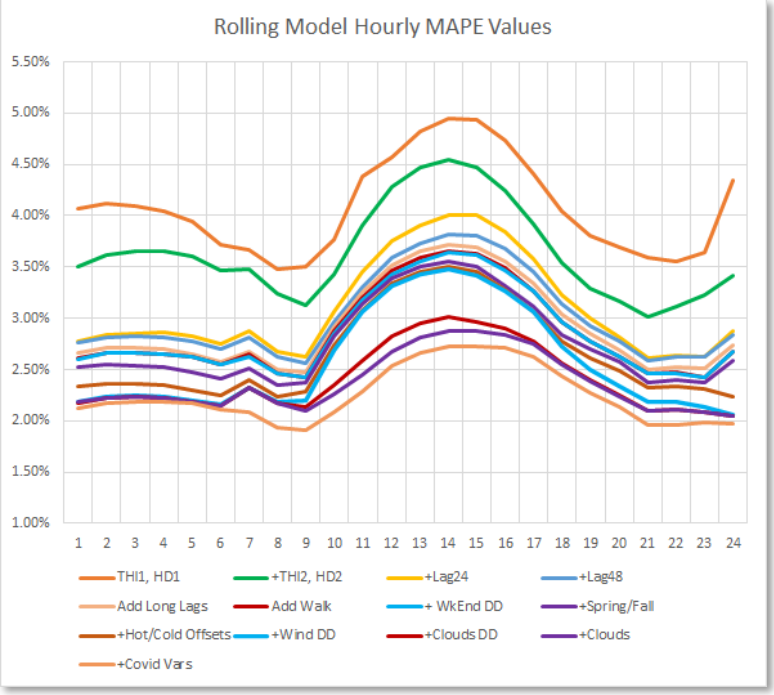
Recommendation: As shown in Figure 4-4 and the associated table, the method with TOD blocks performs much better than the Daily Average method. And the Rolling methods perform a bit better than the TOD-block method. As a result, we recommend using the rolling hourly approach. The complex rolling spline approach works slightly better than the two-part approach. Despite this slight accuracy advantage, we recommend the two-part approach because it is much simpler, it is easier to implement, and it is much easier to understand.

ROLLING TWO-PART MODEL: EXPLANATORY FACTOR CASCADE

Figure 4-5 shows a cascade of statistics for the rolling hourly Two-Part model as groups of explanatory factors are added sequentially. The chart shows the MAPE value for each hourly model and the associated table shows the average of these values across hours.

The first model which includes only the calendar-based factors and does not include any weather variables is not shown in the graph. The second model adds TD1 and HD1 variables for low powered degrees. The third model adds the TD2 and HD2 variables for higher powered degrees. Then Lag HD and TD values are added. This is followed by calendar and weather interaction variables. The last step, which adds a first order autoregressive error term is not shown in the graph.

FIGURE 4-5: ROLLING TWO-PART MODEL, EXPLANATORY FACTOR CASCADE STATISTICS



Rolling Model Components	Average MAPE
Month, DayType, Hols	10.31%
Add THI1, HD1	4.08%
Add THI2, HD2	3.64%
Add Lag24 DD	3.10%
Add Lag48 DD	3.02%
Add Long Lags	2.92%
Add Month Trends	2.87%
Add WkEnd Shift	2.86%
Add Spring/Fall Shifts	2.77%
Add Hot/Cold Lag Offsets	2.66%
Add Wind * HD,CD	2.58%
Add Clouds * HD,CD	2.41%
Add Clouds	2.37%
Add Covid Vars	2.25%
Add AR1	2.04%

This cascade reveals the accuracy gain that results as each variable group that is added. The chart shows hourly pattern of these accuracy impacts. For example, the large gap toward the bottom (from light blue to red) reflects the influence of adding the cloud interaction variables. These have the greatest impact in the day-time hours (between HE9 and HE18), reducing the mid-day MAPE values from 3.5% to 3.0%.

TEMPERATURE-HUMIDITY INDEX CONSTRUCTION

There are a variety of THI-type indexes that have been used over the years. There are formulas that use drybulb and wetbulb temperatures, drybulb and dewpoint temperatures, and drybulb and humidity. The index currently used by PJM falls into the third category and is shown on the left-hand side of Figure 4-6. This formula has some nice properties, in that humidity interacts with temperature. As a result, the higher the temperature, the more important an extra percent of humidity becomes.

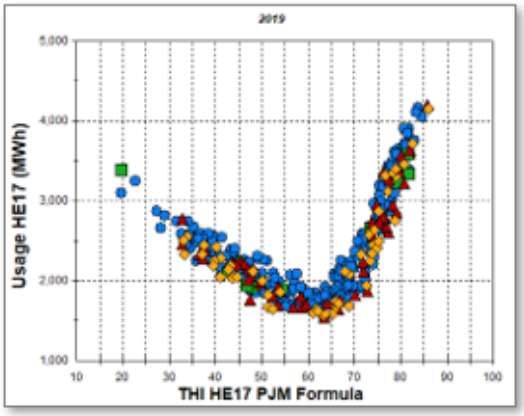
Although it is not obvious, this formula has the unfortunate effect of compressing the temperature variable rather than creating a higher "feels-like" type of variable. This is seen in the four cases that are shown in Figure 4-6. All cases have a drybulb temperature of 90 degrees. In the first case, humidity is at 40%, which is about as low as humidity gets in the DPL zone in summer months. At 40% humidity, the PJM formula gives a THI value of 79.4. As humidity levels rise above 40%, THI values increase and hit their highest value at 100%, at which point the THI value is the same as the drybulb temperature.

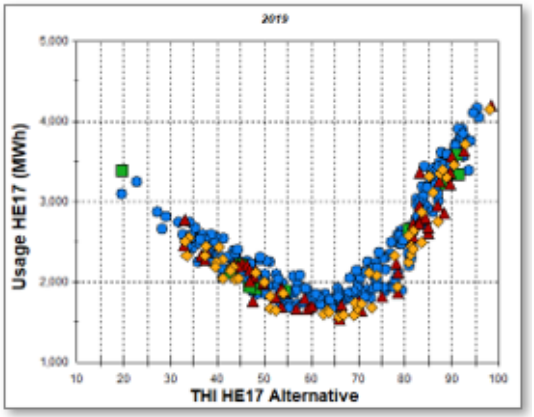
The alternative formula works differently. At humidity levels rise above 40%, the THI temperature goes above 90 degrees. At the extreme of 90 degrees and 100% humidity, the THI reached a level over 100 degrees.

FIGURE 4-6: THI FORMULAS AND HE17 SCATTER PLOTS

	Current PJM Formula T - .55 * (1 - Hum/100) * CD58			
	Case 1	Case 2	Case 3	Case 4
T (°F)	90.0	90.0	90.0	90.0
Hum (%)	40.0	65.0	90.0	100.0
THI	79.4	83.8	88.2	90.0

	Recommended Formula T + .55 * (Hum/100 - .4) * CD58			
	Case 1	Case 2	Case 3	Case 4
T (°F)	90.0	90.0	90.0	90.0
Hum (%)	40.0	65.0	90.0	100.0
THI	90.0	94.4	98.8	100.6





The compression effect is clear in the graphs, which show scatter plots for HE17 loads against the HE17 THI value. The scatter plots show about 4 years of data from mid-2017 to mid-2021. The cold sides of the two graphs (temperatures below 58 degrees) are identical because the humidity component only impacts warmer temperatures. In the left-hand chart, which uses the current PJM formula, THI cooling impacts appear to begin a bit before 70 and the highest THI value is at 85, giving a 15-degree range. With the alternative formula, the highest values are at about 98, giving close to a 30-degree range. This makes it easier to see the difference between low powered degrees (70 to 80 using the alternative definition) and the high-powered degrees (degrees above 80).

By choosing different base temperatures for the two approaches, the two formulas can be made to work about the same in terms of accuracy. But the alternative definition makes it easier to visualize the use of multi-part TD variables on the hot side. Also, we believe that the alternative definition is easier to understand, because it is consistent with our intuition that higher levels of humidity make it feel hotter than the drybulb air temperature. As a result, the following recommendation is based more on convenience and consistency with common sense than performance improvement.

Recommendation: Use the alternative THI formula shown above. As time allows, optimize the parameters in this formula (the .55, the -.4, and the 58-degree base for the temperature interaction), noting that the best values of these parameters may vary across zones. Also, as time allows research alternatives using the same type of interactive specification with the more stable measures of humidity, like dewpoint temperature and moisture content.

WIND/TEMPERATURE INTERACTIONS

PJM currently uses a Wind-Adjusted Temperature (WWP) variable. It is defined as follows:

- WWP = T - .5 * Max (Wind – 10, 0)

This variable is used to compute heating degree values, which come into play in the winter for forecasting energy and system peaks. This variable is the only place that wind appears in the modeling. This implies:

- Wind only matters when it is cold
- Wind does not matter until it exceeds 10 MPH
- Beyond 10 MPH, each additional MPH is like a 0.5-degree temperature decline

In terms of the effect of wind on building HVAC systems, there are at least two effects. First, when it is still, there is a thin layer of air surrounding the building that acts like an insulating layer and that is heated when the sun is shining on the building surface.

When it is cold and cloudy, this layer is warmer than the surrounding air. When the wind blows, this layer of insulation is removed, which increases the heating load.

When it is warm this layer is much hotter than the surrounding air on the roof and sunny side of the building. When the wind blows, it removes this superheated layer, which reduces the cooling load.

These effects probably start to be felt by building systems well below 10 MPH.

Second, wind impacts infiltration levels. High levels of wind result in higher rates of air turnover in buildings that are not perfectly sealed. When humidity levels are high, this has the further effect of replacing dry cooler air inside with warmer humid are from outside.

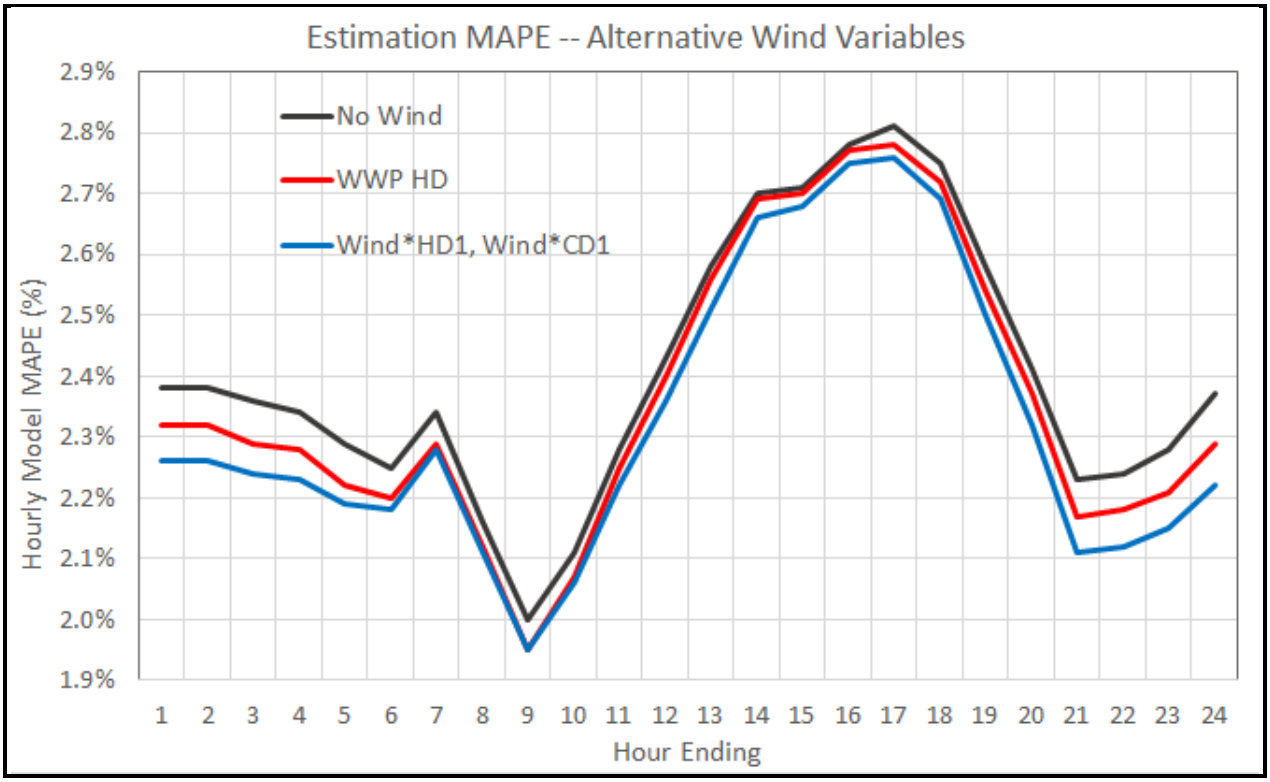
To evaluate this, we estimated models that interact wind speeds with HD and TD variables in each hour. The two variables used are:

- ColdWind = WindSpeed ×HD1
- HotWind = WindSpeed × TD1

On the heating side, the estimated coefficients are negative and significant throughout the day, implying that more wind raises the heating load in all hours when it is cold. The slopes are weaker during the mid-day hours, possibly reflecting the influence of sunlight on building surface temperatures during the day. On the cooling side, coefficients are positive and significant in the night hours and negative and significant in the mid-day hours. During the day, the wind blows the superheated air on sunny building surfaces away, lowering the cooling load. At night, wind causes higher infiltration of warm and possibly humid air, increasing the cooling load.

Figure 4-7 shows MAPE statistics without Wind (black line), with the current wind variable (red line) and the alternative hot side and cold side wind variables defined above.

FIGURE 4-7: IMPACT OF WIND VARIABLES ON MODEL ACCURACY



Inclusion of the cold-side wind variables improves model performance in all hours, with the biggest impacts in the nighttime hours. Use of the alternative variables on both the cold and hot sides result in a further accuracy improvement, with the biggest impacts in the afternoon and nighttime hours.

Recommendation: Use the alternative wind variables on both the cold and hot side. As time allows, optimize the parameters in this formula (the base temperatures in the TD and HD variables that interact with wind speed).

CLOUD COVER/TEMPERATURE INTERACTIONS

Cloud cover variables are not currently included in the PJM models. These are powerful variables, especially in the mid-day and afternoon hours. Cloud cover has three main impacts on utility loads:

On a cold day, clouds interrupt the warming effect of solar radiation on building surfaces. This will result in higher electricity use by electric heating equipment and circulating fans for central heating systems.

On a hot day, clouds interrupt the warming effect of solar radiation on building surfaces. This will result in lower electricity use by electric cooling equipment.

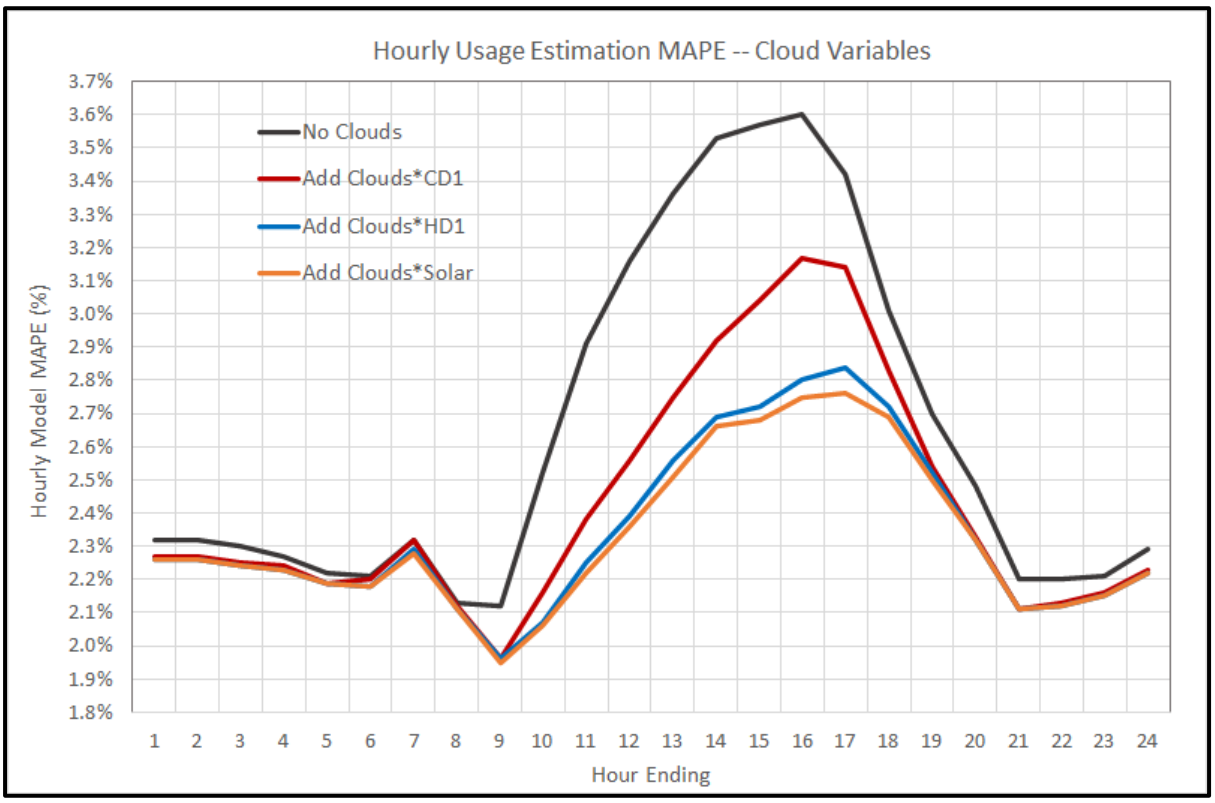
During day-time hours, clouds reduce generation levels for behind the meter solar systems, which will increase the proportion of usage supplied by utility generation sources.

In the early work on this project, we estimated models using utility load, and the third effect was large (as big as 200 MW) and significant. This is shown in the explanatory factor cascade presented above. However, as we will see below, when the explanatory variable is usage (end-use consumption) rather than utility load, the third effect is likely to be small.

Clouds have a further effect of moderating the decline in air temperature at night. But this effect is captured by the measured temperature variable.

Figure 4-8 shows MAPE statistics with usage (reconstituted load) as the variable to be explained. Hourly accuracy statistics are shown for models without Clouds (black line), models with clouds interacted with warm temperatures (red line), and models with clouds interacted with both warm and cold temperatures (blue line).

FIGURE 4-8: IMPACT OF CLOUD VARIABLES ON MODEL ACCURACY



As Figure 4-8 shows, the impact of clouds on model accuracy occurs mostly in the day-time hours, with the biggest impacts about mid-day. The improvement from the black line (average MAPE = 2.64%) to the red line (average MAPE = 2.43%) shows the impacts of clouds on warm days. The further improvement from the red line to the blue line (average MAPE = 2.34%) shows the further impact of clouds on cold days. The final step to the orange line (average MAPE = 2.33%) shows that clouds interacted with clear-sky solar irradiance have a minimal further effect, as expected.

Recommendation: Use the cloud variables interacting with cooling and heating degrees to account for the impact of clouds on hourly energy usage. As time allows, optimize the parameters in this formula (the base temperatures in the TD and HD variables that interact with cloud cover).

INTERACTING HOURLY MODEL VARIABLES WITH SAE EXPLANATORY FACTORS

The following figure shows an example of an estimated hourly model that is built to include SAE variable indexes. Three daily variables are included in the model.

Cooling Index: Accounts for changes in saturation, efficiency, and utilization of cooling equipment (room air, central air, and heat pumps). This index is interacted with the cooling degree variables

(TD1, TD2, LagTD) and the associated cooling degree interaction variables (Wind, Clouds, Weekend, Spring, Fall, etc.)

Heating Index: Accounts for changes in saturation, efficiency, and utilization of heating equipment (resistance heating and heat pumps). This index is interacted with the heating degree variables (HD1, HD2, LagHD) and the associated heating degree interaction variables (Wind, Clouds, Weekend, Spring, Fall, etc.)

Base (Other) Index: Accounts for changes in saturation, efficiency, and utilization of non_HVAC equipment (water heating, cooking, refrigeration, clothes washing and drying, dishwashers, office equipment, lighting, and miscellaneous). This index is interacted with the Constant, Monthly binary variables, day-of-week binary variables, monthly trends, holidays, and Covid phases.

Figure 4-9 provides as an example, the HE18 model (note: the rolling hourly explanatory factors in the model are labeled as hour beginning 17). The variables that are interacted with the Base Index are highlighted in green. The variables that are interacted with the Heating Index are highlighted in red. The variables that are interacted with the Cooling Index are highlighted in blue.

The variables that have clear backgrounds are not interacted with the SAE index variables. Specifically, the Trend, TrendHD, and TrendCD terms are included to allow the model to pick up any trends in the data that are not explained by the SAE variables.

FIGURE 4-9: EXAMPLE OF ESTIMATED MODEL (HE18)

Variable	Coefficient	StdErr	T-Stat
DayTypes.Intercept_XOther	689.463	20.493	33.643
MonthVars.Jan	-5.379	10.614	-0.507
MonthVars.Feb	-29.739	8.883	-3.348
MonthVars.Mar	-77.148	8.422	-9.161
MonthVars.MarDST	-4.426	2.792	-1.585
MonthVars.Apr	-103.989	10.321	-10.075
MonthVars.May	-88.767	10.386	-8.547
MonthVars.Jun	-67.343	11.844	-5.686
MonthVars.Jul	-16.406	14.683	-1.117
MonthVars.Aug	-21.176	16.810	-1.260
MonthVars.Sep	-53.893	15.671	-3.439
MonthVars.Oct	-81.147	12.993	-6.245
MonthVars.Nov	-32.847	12.326	-2.665
MonthVars.NovDST	-16.495	3.604	-4.577
MonthVars.JanWalk	-1.844	1.361	-1.355
MonthVars.FebWalk	-3.773	1.100	-3.429
MonthVars.MarWalk	-0.718	1.583	-0.454
MonthVars.AprWalk	0.915	1.050	0.872
MonthVars.MayWalk	4.216	1.070	3.940
MonthVars.JunWalk	5.269	1.122	4.695
MonthVars.JulWalk	-0.608	1.049	-0.580
MonthVars.AugWalk	-2.870	0.855	-3.357
MonthVars.SepWalk	-4.592	1.119	-4.104
MonthVars.OctWalk	0.475	1.103	0.431
MonthVars.NovWalk	3.100	1.454	2.132
MonthVars.DecWalk	0.283	1.311	0.216
DayTypes.Monday	39.083	3.532	11.066
DayTypes.Tuesday	34.006	3.565	9.540
DayTypes.Wednesday	33.286	3.572	9.320
DayTypes.Thursday	33.936	3.558	9.537
DayTypes.Friday	25.642	3.552	7.219
DayTypes.Saturday	-12.776	2.736	-4.669

Interacts with XOther

Variable	Coefficient	StdErr	T-Stat
Calendar.MLK	37.283	50.244	0.742
Calendar.PresDay	-19.129	47.730	-0.401
Calendar.GoodFri	-8.112	54.883	-0.148
Calendar.MemDay	-114.256	49.459	-2.310
Calendar.July4th	-134.103	47.159	-2.844
Calendar.LaborDay	-39.744	50.150	-0.793
Calendar.Thanks	-378.458	49.853	-7.591
Calendar.FriAThanks	-127.570	49.667	-2.569
DayTypes.WkBeforeXMas	-7.743	15.597	-0.496
Calendar.XMasEve	-145.364	81.256	-1.789
Calendar.XMasDay	-359.132	50.854	-7.062

Variable	Coefficient	StdErr	T-Stat
Calendar.NYEve	-40.271	66.697	-0.604
Calendar.NYDay	-183.177	46.315	-3.955
DayTypes.WkAfterNewYear	20.083	10.327	1.945
DayTypes.WkDayBeforeHol	22.668	26.707	0.849
DayTypes.WkDayAfterHol	-10.909	17.254	-0.632
DayTypes.Phase1	-78.117	14.309	-5.459
DayTypes.Phase2	90.195	11.040	8.170
DayTypes.Phase3	93.289	12.771	7.305
DayTypes.Phase4	31.587	11.900	2.654
DayTypes.Trend2015	-23.962	4.451	-5.384

Heating Vars **Interacts with XHeat**

Variable	Coefficient	StdErr	T-Stat
HD1.HD1_17	12.198	4.552	2.680
HD2.HD2_17	21.169	2.171	9.751
Lag6.Lag6HD_17	7.148	2.182	3.276
Lag24.Lag24HD_17	4.650	0.729	6.381
Lag24HC.Lag24CD_HD17	-1.440	7.269	-0.198
WkEndDD.WkEndHD17	0.475	0.628	0.757
SeasHD.SpringHD17	-0.543	0.636	-0.854
SeasHD.FallHD17	-0.977	0.700	-1.397
ColdWind.WindHD17	18.443	2.617	7.047
ColdClouds.CloudHD17	36.500	4.051	9.011
Daily.MA10_HDD	1.248	0.841	1.484
Daily.MA28_HDD	0.289	1.299	0.222
TrendDD.Trend_HD17	-0.670	0.314	-2.137

Cooling Vars **Interacts with XCool**

Variable	Coefficient	StdErr	T-Stat
CD1.TD1_17	26.219	4.517	5.804
CD2.TD2_17	8.522	1.778	4.792
Lag6.Lag6CD_17	32.113	2.969	10.817
Lag24.Lag24CD_17	6.928	1.307	5.299
Lag24HC.Lag24HD_CD17	-25.499	9.861	-2.586
WkEndDD.WkEndCD17	-0.237	0.870	-0.272
SeasCD.SpringCD17	-19.997	1.493	-13.391
SeasCD.FallCD17	-15.404	2.052	-7.509
HotWind.WindCD17	-2.640	6.026	-0.438
HotClouds.CloudCD17	-137.085	10.435	-13.138
Daily.MA10_CDD	5.254	1.645	3.195
Daily.MA28_CDD	1.096	3.090	0.355
TrendDD.Trend_CD17	1.298	0.330	3.930

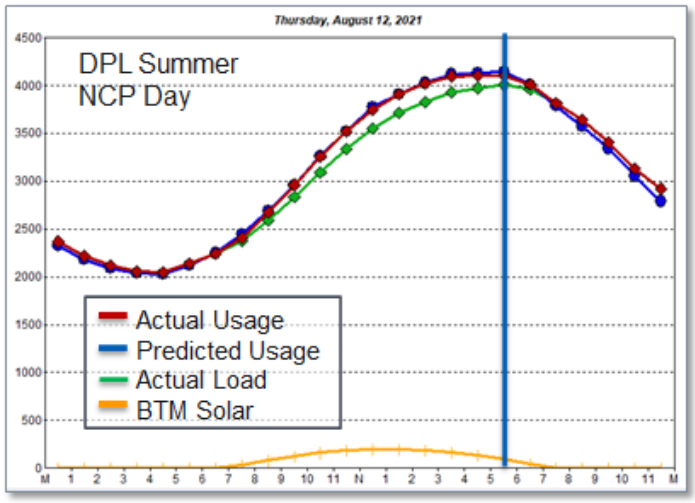
PEAK DAY PREDICTED VALUES AND POST PROCESSING

To evaluate the power of the estimated models on peak days, the following figures focus in on the actual 2021 peak days for DPL and PJM. There are three figures, and each figure shows hourly data for the following four series:

- Actual Usage: This is the estimate of actual usage (reconstituted load), which is measured utility load adjusted for estimated DR effects plus estimated behind-the-meter solar generation.
- Predicted Usage: This is the predicted value from the hourly usage models with explanatory factors interacted with SAE index variables.
- Actual Load: This is the measured DPL zone load, representing the fraction of total usage that is supplied by zone generation resources.
- Solar Generation: This is the estimate of hourly energy supplied by behind-the-meter solar generation.

Figure 4-10 shows the DPL zone peak day, August 12, 2021. Estimated usage on this day is relatively flat between HE16 and HE18. The estimated BTM solar generation, however, is ramping down from 167 MWh to 133 MWh to 94 MWh over these three hours. The hourly models correctly predict that the usage peak occurs on HE17 and that the zone peak load occurs on HE18. In the NCP hour, the model over predicts the zone load by about 46 MWh (1.2%).

FIGURE 4-10: HOURLY MODEL RESULTS – SUMMER 2021 NCP

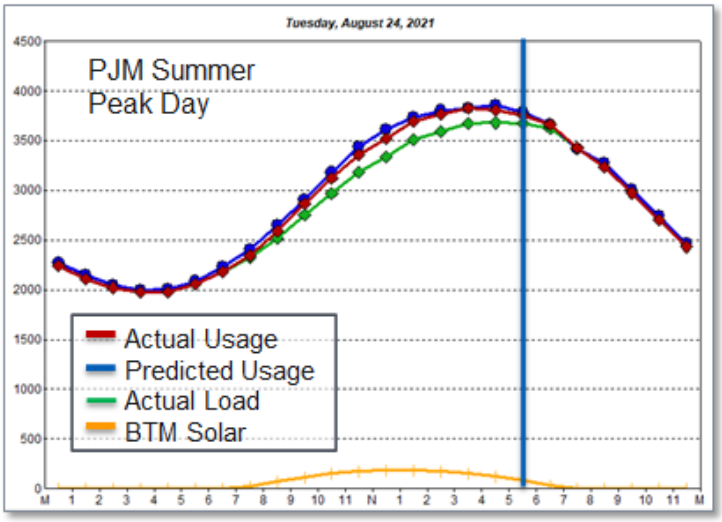


DPL Summer NCP Hour Data	DPL NCP 8/12 HE18
Actual Usage	4,100.3
Predicted Usage	4,146.7
Usage Error	-46.4
Solar_DR	94.3
Actual Load	4,006.0
Predicted Load	4,052.4
Load Error	-46.4

Figure 4-11 shows the DPL zone usage and loads on the PJM peak day, August 24, 2021. On this day, the DPL zone usage peaked in HE16, the measured zone load peaked in HE17, and the PJM peak occurred in HE18. The CP statistics provided to the right of the chart are for HE18, the hour of the

annual PJM peak. In the CP hour, the hourly model overpredicts the DPL zone load by about 27 MWh (.7%).

FIGURE 4-11: HOURLY MODEL RESULTS – SUMMER 2021 CP



DPL Summer CP Hour Data	DPL CP 8/24 HE18
Actual Usage	3,753.6
Predicted Usage	3,781.0
Usage Error	-27.5
Solar_DR	83.6
Actual Load	3,670.0
Predicted Load	3,697.5
Load Error	-27.5

There is a significant difference between the DPL loads on the DPL peak day (Aug 12) and the PJM peak day (Aug 24). Weather statistics for the peak hours on these days are shown in Figure 4-12. The main factor is the difference in average drybulb temperature on these days (93 degrees on the DPL peak day and 89 degrees on the day of the PJM peak day). This difference of nearly 4 degrees persisted throughout the afternoon and evening hours on these days.

FIGURE 4-12: NCP AND CP HOUR WEATHER

Summer Peak Hour Weather Variables	DPL 2021 NCP 8/12 HE18	PJM 2021 Peak 8/24 HE18
AvgDB Temperature (Deg F)	92.7	88.8
Relative Humidity (%)	54.3	53.3
Moisture Content (mmHg)	21.3	19.1
Temp Hum Index (Deg F)	95.4	91.1
Wind Speed (MPH)	11.4	8.4
Cloud Cover (Octas)	0.12	1.6

The difference in loads and the implied coincidence factors are summarized in Figure 4-13. The table focuses on the measured loads. The coincidence factor is the ratio of the CP value to the NCP value. For the summer peak, the actual coincidence factor is 91.6% and the predicted coincidence factor is 91.2%. This are both well below the typical coincidence factor for DPL and are toward the bottom of the range generated by the multi-year weather simulations.

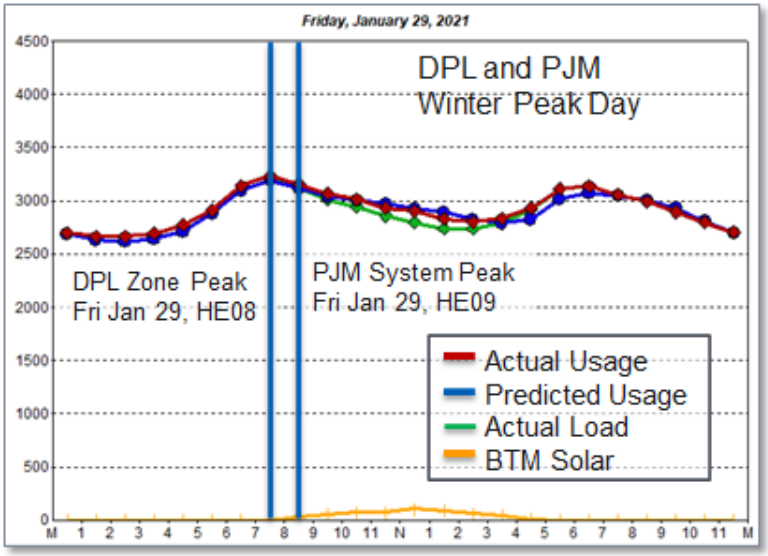
FIGURE 4-13: STATISTICS FOR 2021 SUMMER CP COINCIDENCE FACTOR

DPL Summer Peak Hour Results	Date/Time of NCP	NCP	CP 8/24 HE18	Coincidence CP/NCP
Actual Load	8/12 HE18	4,006.0	3,670.0	91.6%
Predicted Load	8/12 HE18	4,052.4	3,697.5	91.2%
Error	0	-46.4	-27.5	0.4%
Error %		-1.2%	-0.7%	

Figure 4-14 shows the DPL and PJM winter peak day (January 29, 2021). Both peaks occurred on the same day, but at different hours (HE08 and HE09, respectively). The hourly model underpredicted the load in these hours by 39 MWh (1.2%) for the NCP and 43 MWh (1.4%) for the CP. Behind-the-meter solar generation ramped up between the two hours from 2 MWh to 36 MWh.

In terms of coincidence, the hourly model correctly predicts the DPL peak to occur in the hour before the PJM Peak. The predicted winter peak coincidence factor (96.5% is within .1% of the actual measured coincidence factor (96.6%).

FIGURE 4-14: HOURLY MODEL RESULTS – WINTER 2021 NCP AND CP



	DPL NCP 1/29 HE08
Actual Usage	3,232.8
Predicted Usage	3,193.4
Usage Error	39.4
Solar_DR	1.8
Actual Load	3,231.0
Predicted Load	3,191.6
Load Error	39.4

	DPL CP 1/29 HE09
Actual Usage	3,158.0
Predicted Usage	3,115.0
Usage Error	43.0
Solar_DR	36.0
Actual Load	3,122.0
Predicted Load	3,079.0
Load Error	43.0

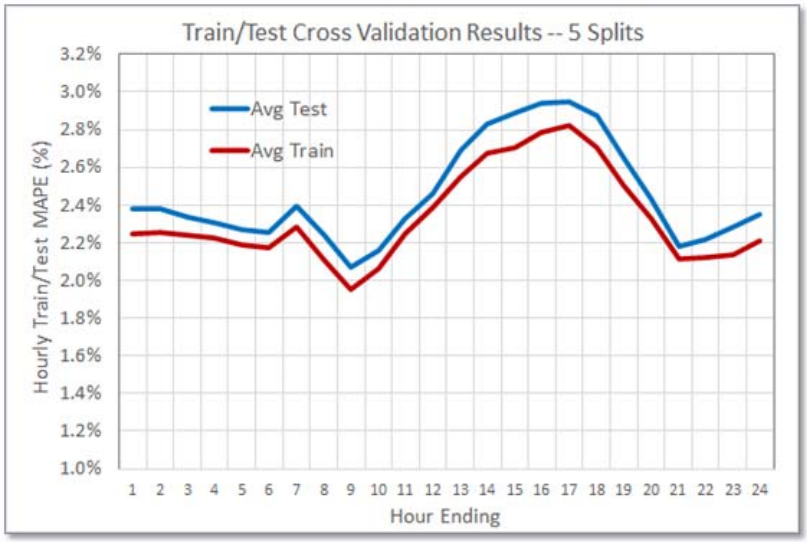
DPL Winter Peak Hour Results	Date/Time of NCP	NCP	CP 1/29 HE09	Coincidence CP/NCP
Actual Load	1/29 HE08	3,231.0	3,122.0	96.6%
Predicted Load	1/29 HE08	3,191.6	3,079.0	96.5%
Error	0	39.4	43.0	0.2%
Error %		1.2%	1.4%	

The accuracy results observed for the 2021 peak days are better than expected given that the average MAPE values for the hourly models are in the 2% to 3% range, depending on the hour. So, it is probably unrealistic to expect results this good for all years. However, the results do show that the hourly modeling approach is capable of explaining NCP, CP, and coincidence factor outcomes. By extension, simulations with alternative actual weather patterns should realistically represent the range of coincidence results that would occur under varied conditions. As part of this project, we did not estimate models for PJM zones other than DPL, nor did we analyze model accuracy for other zones in terms of peak loads and coincidence factors.

MODEL TESTING

The final step with the hourly models is out-of-sample testing. This was performed by running 5 test cases, with each case withholding about 20% of the data. Models were estimated using the training subset of the data in each case and accuracy statistics were computed using the test subset of the data in each of the cases. Figure 4-15 present the train and test statistics averaged across the 5 test cases.

FIGURE 4-15: HOURLY MODEL RESULTS – WINTER 2021 NCP AND CP



	Average	Test1	Test2	Test3	Test4	Test5
All Data Estimation	2.34%					
Average Train	2.34%	2.35%	2.35%	2.33%	2.31%	2.34%
Average Test	2.45%	2.39%	2.48%	2.46%	2.52%	2.41%

Hour	Average Train	Average Test
HR00	2.25%	2.38%
HR01	2.26%	2.38%
HR02	2.24%	2.34%
HR03	2.23%	2.30%
HR04	2.19%	2.27%
HR05	2.18%	2.26%
HR06	2.28%	2.39%
HR07	2.11%	2.24%
HR08	1.95%	2.07%
HR09	2.07%	2.16%
HR10	2.25%	2.33%
HR11	2.39%	2.46%
HR12	2.55%	2.69%
HR13	2.67%	2.83%
HR14	2.70%	2.89%
HR15	2.78%	2.94%
HR16	2.83%	2.95%
HR17	2.71%	2.88%
HR18	2.51%	2.66%
HR19	2.33%	2.43%
HR20	2.11%	2.18%
HR21	2.12%	2.22%
HR22	2.13%	2.28%
HR23	2.21%	2.35%
Average	2.34%	2.45%

As the statistics show, the accuracy for the test cases is close to the estimation (training) accuracy. Averaged across all tests and hours, the test MAPE is about .11% larger than the training MAPE. This indicates that the hourly models are robust (generalize well out of sample). This result (accuracy loss of about .10%) is what we generally expect to see in a model that is well specified in terms of nonlinearities and important interactions and that is not over specified.

5 RESHAPING DEMAND: MODELING TECHNOLOGY IMPACTS

EXTENDED WEATHER SIMULATION FRAMEWORK

The current LTFS forecast process utilizes a multi-year weather simulation framework to construct Cumulative Distribution Functions (CDF) daily energy, daily noncoincident zone peak (NCP), and daily CP (zone load at the time of the daily PJM peak). For each zone, CP models are estimated for daily loads coincident with the Locational Deliverability Area (LDA) peak and coincident with the overall PJM system peak.

The application of the CP and NCP modelling is complicated by the saturation of new technologies (PV and EV in particular) that modify the timing of system energy requirements. In particular, the penetration of behind the meter solar systems is changing both the timing of zone peaks and the coincidence factors across zones. The best way to understand the impact of these changes is by simulating the impact of new technologies on hourly loads. Summarized below are the high-level steps of the recommended Extended Weather Simulation Framework.

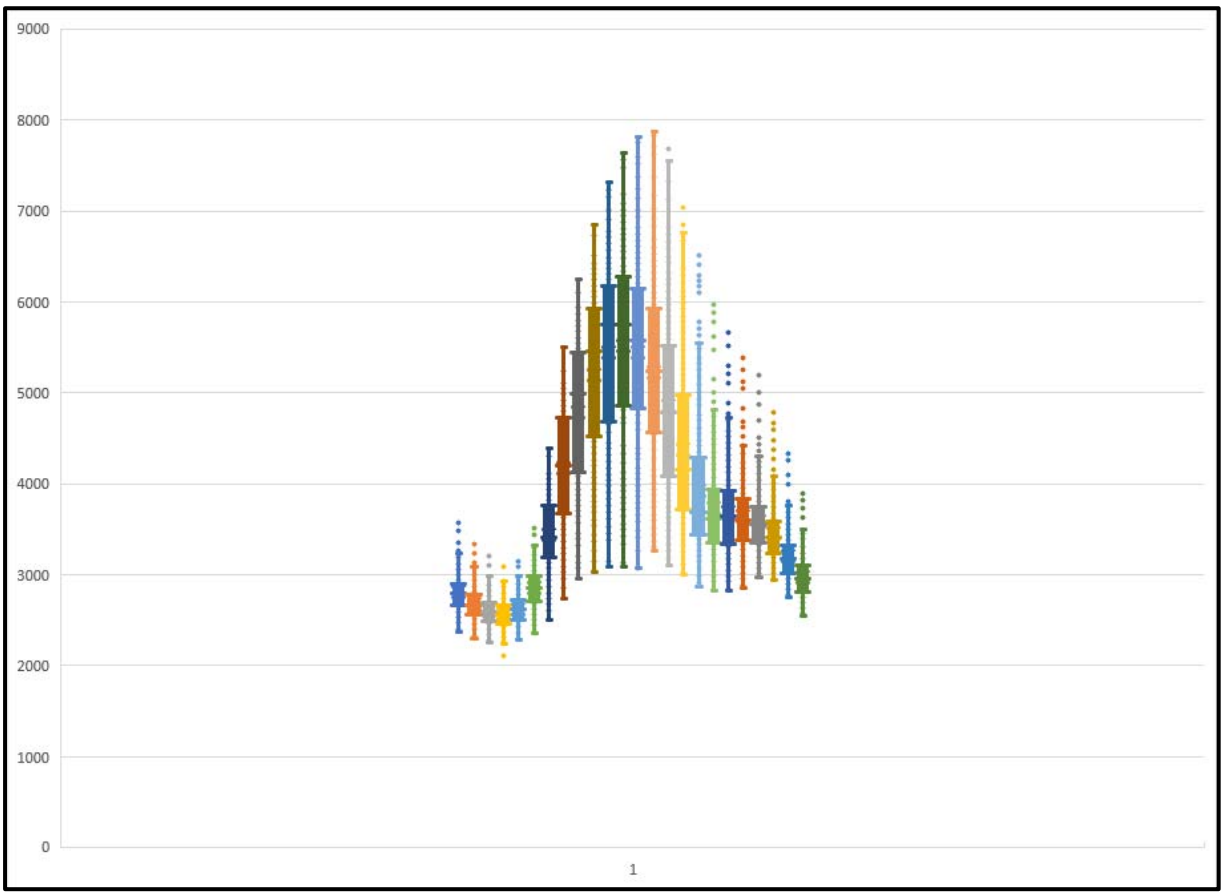
Step 1. Create Weather Forecast Simulation Traces by EDC and Forecast Year. Under this step the existing Multi-year Weather Simulation framework is extended to include hourly Solar Generation Capacity factors that are consistent with the corresponding hourly temperature, wind speed, wind direction, humidity, precipitation, and cloud cover or Global Horizontal Irradiance. The Solar Generation Capacity factors when multiplied by forecasts of embedded solar PV capacity provides an estimate of hourly embedded solar PV generation. These solar PV generation values reduce baseline forecasts of hourly reconstituted loads derived from the hourly load forecast models. The result is a forecast of PJM generation requirements net of distributed solar PV generation.

In a similar fashion, any new technologies that are weather sensitive (e.g., electric heat pumps) should have utilization or capacity factors that are consistent with the hourly weather data.

Step 2. For each EDC, Weather Trace and Forecast Year Simulate EDC-Level Hourly Reconstituted Loads using the Hourly Load models and rotated weather data. For each Weather Trace and Forecast Year Construct PJM Total & PJM Region Hourly Loads by summing across EDCs by Forecast Year, Weather Trace and EDC.

Step 2.a. For PJM Total, PJM Region and for each EDC and Forecast Year, construct hourly Probability Density Functions (PDF) and Cumulative Distribution Functions (CDF) of Reconstituted loads. An example of the PDF for a single EDC and forecast year is presented in Figure 5-1. In this figure each hourly vertical slice represents the PDF of 365 reconstituted loads in that hour.

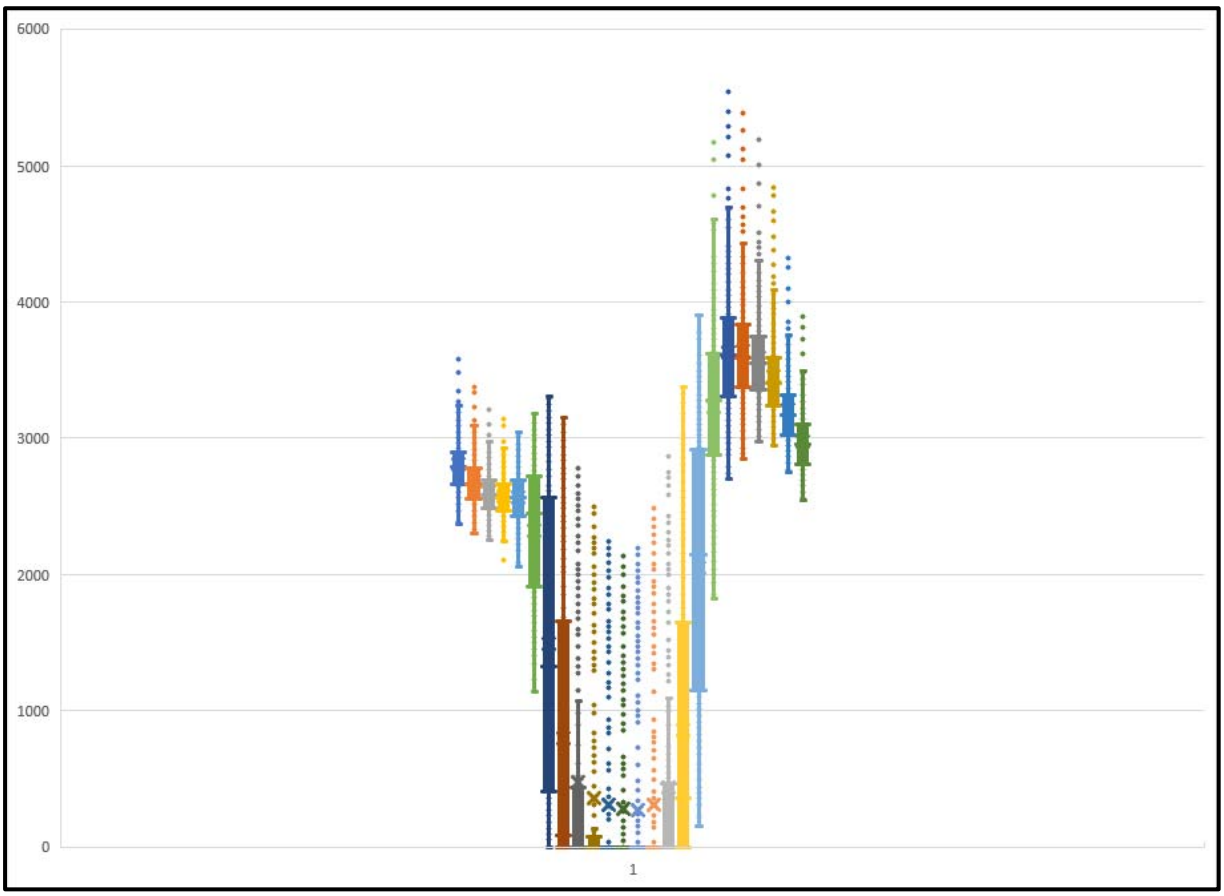
FIGURE 5-1: EXAMPLE PDF OF RECONSTITUTED LOADS FOR A SINGLE EDC AND FORECAST YEAR



Step 3. For each EDC, subtract off simulated EDC-Level embedded hourly solar PV generation Weather Trace and Forecast Year. For each Weather Trace and Forecast Year Construct PJM Total & PJM Region Hourly Loads by summing across EDCs by Forecast Year, Weather Trace and EDC.

Step 3.a. For PJM Total, PJM Region and for each EDC and Forecast Year, construct hourly PDF and CDF of Net Loads (i.e., Reconstituted loads less embedded solar PV generation). An example of the PDF for a single EDC and forecast year is presented in Figure 5-2. In this figure each hourly vertical slice represents the PDF of 365 net loads in that hour.

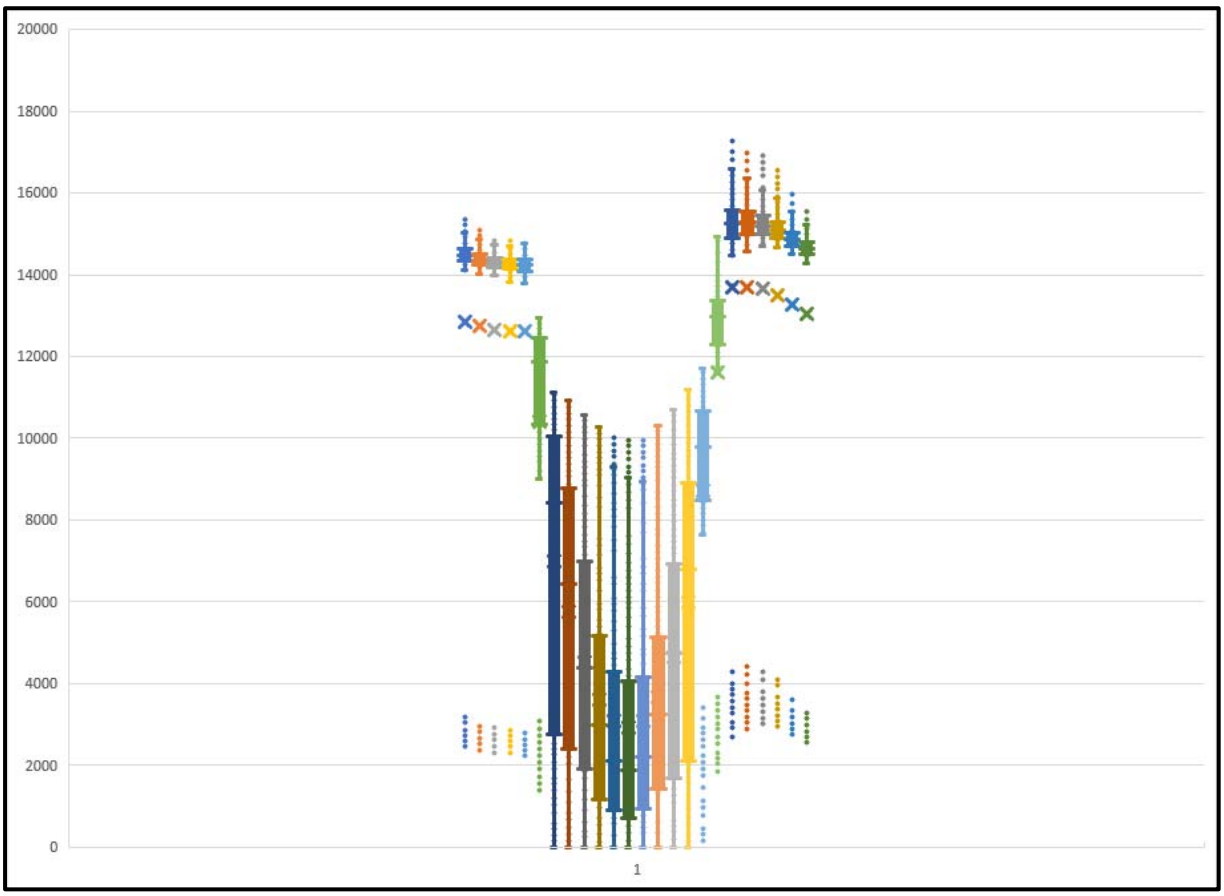
FIGURE 5-2: EXAMPLE SOLAR PV GENERATION ON NET LOADS FOR A SINGLE EDC



Step 4 – Step N. Here N represents the total number of technology impacts that are simulated. For each EDC, Layer in the hourly load impact of the remaining Weather Sensitive and Non-Weather Sensitive Technology Impact Shapes by Weather Trace and Forecast Year. For each Weather Trace and Forecast Year Construct PJM Total & PJM Region Hourly Loads by summing across EDCs by Forecast Year, Weather Trace and EDC.

Step 4.a. – Step N.a. At each step, form PDF and CDF values of the resulting hourly loads by PJM Total, PJM Region, and each EDC and Weather Forecast Year. An example of the PDF for a single EDC and forecast year is presented in Figure 5-3. In this figure each hourly vertical slice represents the PDF of 365 net loads in that hour after residential, commercial, and fleet EV Charging is added in.

FIGURE 5-3: EXAMPLE OF EV CHARGING IMPACT ON NET LOADS FOR A SINGLE EDC



6 MODELING ISSUES

The Market Participants identified several issues to be addressed as part of the study. These issues include:

- Measuring Forecast Accuracy,
- Weather Normalization,
- Capturing Impacts of Energy Efficiency Programs,
- Accounting for Temperature Trends, and
- Large Load Adjustments.

MEASURING FORECAST ACCURACY

Forecast errors for an hourly SAE model can be decomposed into three broad categories:

- Actual weather deviations from "normal"
- Economic growth deviations from forecast
- Other, including SAE saturation and efficiency forecast errors, specification errors, measurement errors, and statistical model errors.

Actual data for the first category (weather) are available on a near real time basis. Actual data for economic growth are available with varying lags and revision cycles. Updated estimates for the SAE inputs are produced annually, but the underlying research that leads to major updates take several years to develop and process, and as a result, we place these in the Other category.

The PJM zone and system forecasts are developed using the weather simulation approach, and the final forecast values for energy, NCP, and CP variables are derived by post processing these data. The best way to judge the accuracy of the underlying model that is used in the weather simulations is to substitute in the actual hourly weather that occurred and, if available, the actual economic variable values. Then compare the model predicted values to the actual values.

For energy and NCP values, this comparison is relatively straightforward. First run the hourly models with actual weather and economic data that are available. Then subtract out the estimated BTM hourly solar generation. Then compute monthly energy and NCP values for each zone and compare these results with the actual measured load values.

For the CP values, more post processing is required to compute the predicted time of the system peak and the predicted hourly zone load values at the CP times. In the end, the idea is to compare what the model would have predicts given actual weather and economics and the outcomes that actually occurred under these conditions. These comparisons are very much like the estimation error comparisons that were made toward the end of the Hourly Models section of this report, where we looked at NCP and CP errors for 2021.

WEATHER NORMALIZATION

The current LTFS forecast process utilizes a multi-year weather simulation framework to construct Cumulative Distribution Functions (CDF) daily energy, daily noncoincident zone peak (NCP), and daily CP (zone load at the time of the daily PJM peak). When this process is run over historical years the midpoint of the CDFs for each concept is taken as the *weather normalize* value.

In general, the power industry utilizes one of two broad approaches for constructing weather normalized loads where each load zone is treated independently of other load zones. In principle, the approaches are designed to provide load patterns under average or typical weather conditions.

Approach 1: Construct Normal Weather. Under this approach, a *normal* weather pattern is first constructed using historical weather data for the load zone under study. The constructed series of normal weather is then pushed through the load zone forecast model to generate a backcast of loads under normal weather conditions. There are several schemes for constructing a normal weather pattern.

- Average weather by calendar day. Under this approach, multiple years of weather are grouped by calendar date (e.g., January 1st, January 2nd, …, December 31st). The daily or hourly averages of each weather concept are then computed. The resulting time series of average data is deemed normal weather.
- An alternative to daily averages is to sort each month of historical weather data from the coldest day to the hottest day. That is, all January days across a year are sorted from the coldest day to the hottest day. The same ranking is developed for each year of available historical weather data. Once the data has been ranked, average values of the ranked data are taken. For the coldest day in January, the algorithm computes the average of the coldest day. The second rank of January data is then used to compute the average of the second coldest day. This continues until all ranked days are averaged resulting in 365 days of average data. The last step is to select a weather pattern from the historical data to allocate the ranked data. For example, if 1987 is selected as the weather pattern year, then the coldest January average is mapped to the coldest day in January 1987. The second coldest January average data is mapped to the second coldest day in January 1987. The process continues until every day in 1987 is assigned data. The resulting Rank & Average time series is then pushed through the load zone forecast model to generate a backcast of loads under normal weather conditions.
- The average approaches described above resulting in average values for all weather concepts. With the proliferation of distributed energy resources like solar PV the resulting weather normalized loads (net of solar PV generation) becomes harder to interpret. The National Renewable Energy Laboratory (NREL) has developed Typical Meteorological Year (version TMY3) weather data that is designed to support building simulations under realistic weather conditions. (See https://nsrdb.nrel.gov/data-sets/tmy). This includes solar radiation data that is consistent with temperature, humidity, wind speed, and wind direction. These data are available for most major weather stations spanning the US. The TMY3 data is constructed by selecting for each month the year that has been deemed the most typical. The actual hourly weather data for the selected year and month is then assigned as the TMY data. For example, the TMY3 data may

use for January the data from January 2010, for February the data from February 2015,…, for December the data from December 2008.

All three schemes work well when the load pattern for a single zone is considered in isolation of other load zones because the construction of normal weather is performed one weather station at a time without consideration of the weather conditions for other weather stations. Thus, it is plausible that normal weather could lead to differences in the timing of and contribution of each load zone to the PJM System and Regional Peaks that are outside historical values. If this approach is used, then of the three schemes the TMY3 data will generate the most realistic net load pattern under *normal* weather conditions.

Approach 2: Multi-year Weather Simulations. Under this approach, multiple years of observed weather data are combined with a load forecast model to backcast loads under alternative weather patterns. The resulting load backcasts are then summarized to determine *weather-normalized* loads. The LTFS constructs for each load zone CDFs of the monthly and seasonal backcast CP and NCP values. The median or 50% value from these CDFs are deemed the weather normalized values.

For multiple zones that are geographically dispersed, construction of **normal weather** patterns that yields a consistent set of "normal" loads across zones is difficult if not impossible. For multiple zones, weather simulations using historical weather patterns yield realistic distributions of loads across load zones. Weather simulations do not ensure that the weather that leads to the 50% or median value CP and NCP for a given load zone may not correspond to the same weather trace that leads to the 50% or median CP and NCP values for all other load zones. As a result, care must be taken when comparing the weather normal CP and NCP across zones.

The introduction of Distributed Energy Resources like solar PV adds complexity and begs the questions: What are "Normal" Loads? The added load volatility associated with distributed solar PV is driving a growing interest on the part of system operators and planners in quantifying hourly load uncertainty. The weather simulation process that PJM employs is well suited to quantifying hourly load uncertainty. We recommend shifting the focus away from developing "normalized" CP and NCP values toward using the weather simulation process to quantify historical and future hourly load uncertainty. The hourly load simulations can also be used to construct distributions of hourly ramp rates that are expected to evolve dramatically with deep penetration of solar PV generation and EV charging.

CAPTURING THE IMPACTS OF ENERGY EFFICIENCY PROGRAMS

Impacts of State and Utility energy efficiency (EE) programs are captured in the model end-use intensities along with new standards and natural occurring efficiency improvements as old appliances are replace with new appliance. There are two components to the end-use intensities – saturations, the share of homes or commercial floorspace that own the end-use and efficiency – the kWh or work output per kWh input. Most primary residential and commercial end-use intensities other than miscellaneous are declining or are flat as increase in end-use stock efficiency is generally increasing faster than saturation.

End-use intensities, a direct input into the SAE models, are derived from the EIA Annual Energy Outlook – saturation and efficiency projections; in the commercial sector, efficiency impacts are embedded in the intensity projections. End-use stock efficiency is derived from an end-use choice model that moves the average stock efficiency based on stock turnover, new purchases, and relative life-cycle costs of competing technology options. In this model, standards work to limit the least efficient options over time while declining costs in more efficient technologies result in greater adoption of the higher efficient technology options.

EIA captures the EE impacts in a couple of ways. First, in the last few years, EIA has made an effort to directly account for state and utility efficiency programs by mapping regional EE program expenditures to end-uses and "rebating" (lowering the cost) of the high-efficient technology options. As a result of the lower cost, more of the high-efficient technology option is adopted. Second, the underlying information on new technologies including number of units sold and associated efficiency information are updated on an on-going basis; this information is derived from annual appliance shipments data. The impact of programs that encourage adoption of more efficient technology such as the Energy Star program and utility incentive programs are partly reflected in the shipments data that in turn are used in calibrating the AEO end-use models.

The SAE estimation process itself also contributes to capturing EE program impacts as the sales and customer data used in estimating the models incorporates past EE program impacts.

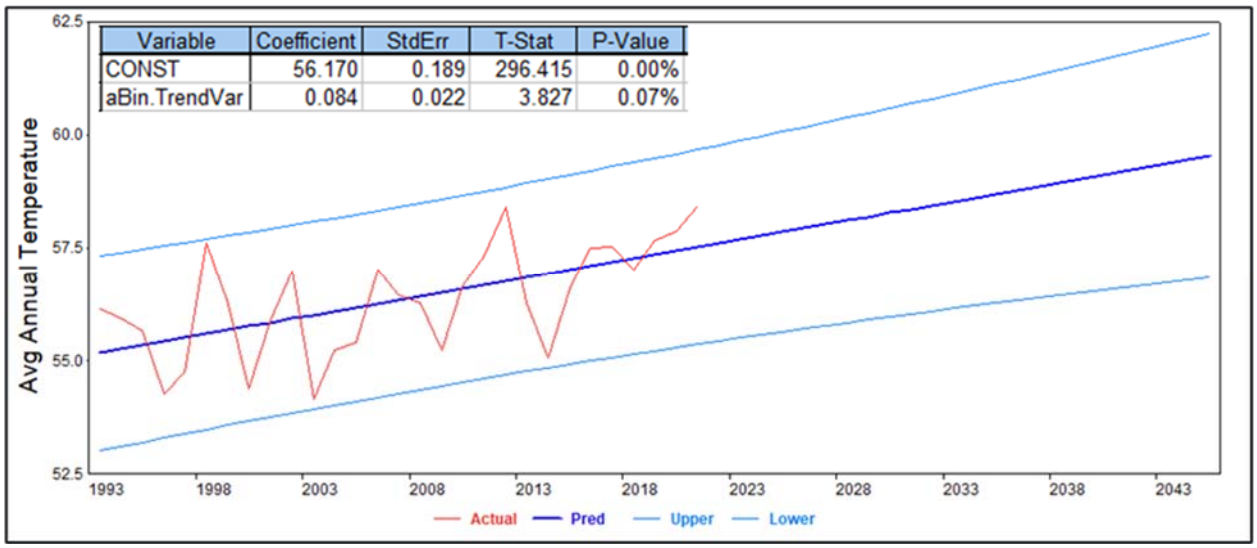
ACCOUNTING FOR TEMPERATURE TRENDS

In electric load forecasts we generally assume that current and future weather conditions will most likely look like the average of the past years. Typically, normal or expected weather is derived as an average of temperatures or degree-days for a twenty to thirty-year period; some utilities will use periods as short as ten-years. Rather than define normal weather conditions, PJM simulates with historical actual temperatures and then calculates normal as the 50% probability outcome. The outcome is approximately the same as if the average temperature profile for the PJM system could be determined.

In recent temperature trend studies, we have found that depending on location, average temperature has been increasing from 0.4 degrees to over 1.0 degree per decade. Our work has been consistent with other temperature trend studies and climate model projections. Giving increasing temperature an average based on the last 20 years would be more representative of 2011 (the mid-year) than 2022.

Figure 6-1 shows the DPL average temperature trend.

FIGURE 6-1: DPL AVERAGE TEMPERATURE TREND



Since 1993, average temperature has been increasing .08 degrees per year or 0.8 degrees per decade. This translates into increasing CDD and decreasing HDD as illustrated in Figure 6-2 and Figure 6-3.

FIGURE 6-2: DPL TRENDED CDD

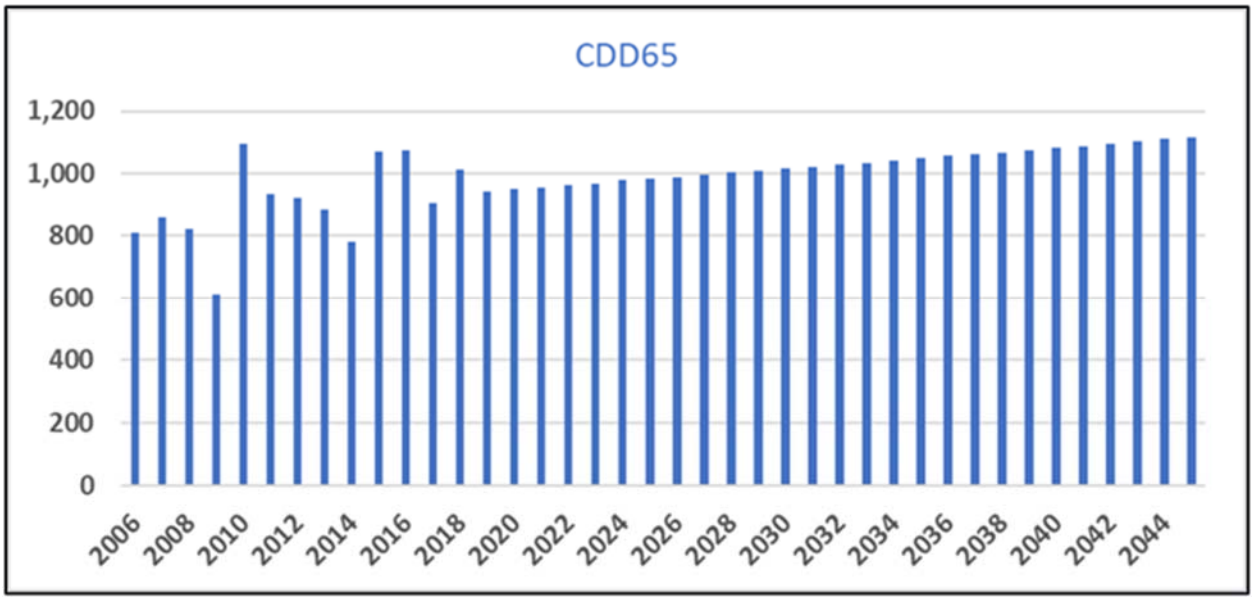
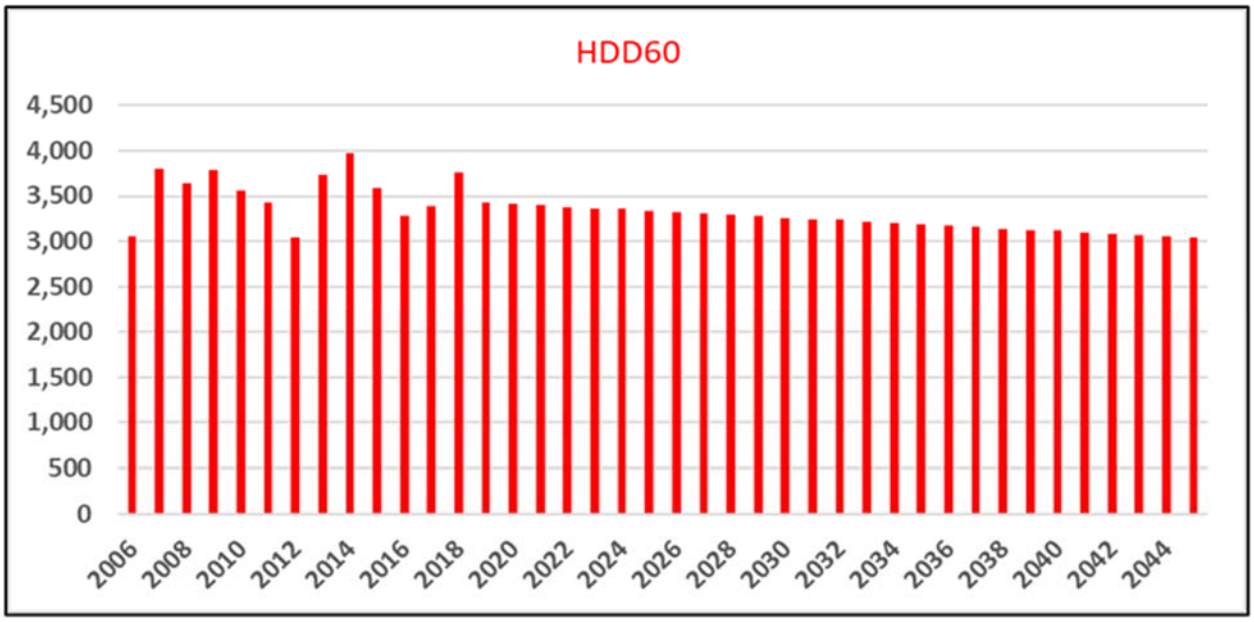


FIGURE 6-3: DPL TRENDED HDD



The weather simulation process and model complexity make it difficult to directly model the impact of increasing temperatures. The approach we recommend is less complex; that is to capture the impact of increasing temperatures with trended normal HDD and CDD in the constructed heating and cooling model indices.

LARGE LOAD ADJUSTMENTS

In the simulation framework, the impact of new technologies such as solar and EVs depends on the load/weather simulation outcome. The earlier section on reshaping loads describes the recommended approach for carrying these technology load adjustments through the forecast horizon. The load forecast may also need to be adjusted for large load additions or losses from a specific customer or group of customers; this would generally be in the industrial sector. If industrial loads adjustments are not too extreme, impacts of load loss or gain can be made to the industrial baseline energy projection that is part of the base-use model index.

One of the largest adjustments has been for data centers in Dominion Zone. Data centers load growth has been enormous – data center loads for just Dominon alone has increased from 1,250 MW at the beginning of 2020 to over 2,500 MW in June of this year. Demand is expected to continue to increase at this pace through 2026. PJM excludes the DOM Zone data center loads from the DOM Zone models and adjust the load afterwards for expected data center load growth. Given the size and growth of data center loads this is the correct approach. PJM relies on Dominion for the first five years of the load projection; this makes sense since Dominon works directly with their data center customers as Dominion plans and builds capacity for future needs.

While the DOM Zone should continue to experience strong data center load growth past the next five years, the rate of growth is likely to slow. Factors that will impact demand growth are transmission and other physical constraints, and demand for more localized data center capacity for services that require faster response time such as 5-G networks, autonomous vehicles, internet of things, and automated manufacturing processes. The long-term forecast will need to be based on historical capacity trends, continuous discussions with utility personnel that are directly involved with data center activities and transmission construction, trade and real estate publication reviews, and organizations that support regional data center development.

7 RECOMMENDATIONS

Throughout the project, we explored several modeling options, discussed issues with the current modeling approach, and ultimately through our work and meetings with PJM staff and Market Participants have developed a set of recommendations that build on the current PJM forecast model. Our recommendations are outlined below:

1. **Replace Annual/Quarterly End-Use Indices with Monthly/Daily Indices**

   Heating, cooling, and base-use load indices can be derived from monthly class SAE models. The SAE models are well documented, used by many utilities for long-term sales and energy forecasting, and are relatively robust in the sense that adding new data and dropping old data does not generally result in significant changes in the model parameters. Indices based on monthly (versus annual models) provide significantly more observations and as a result require fewer years of historical data; resulting in estimated model parameters that will be more representative of the current and forecast periods. Monthly models will also result in stronger heating and cooling coefficients because there is generally more weather variation in monthly data series than in an annual data series.

2. **Continue with Weather Simulation Approach**

   Given the diversity of weather across PJM zones, it is nearly impossible to define a normal daily or hourly weather pattern for the entire system. The current method of developing load distributions from zonal weather simulations represents the best approach for estimating expected long-term demand. Twenty-years of historical weather data with 7 rotations within in each year provides a strong basis for simulating the distribution of load outcomes.

3. **Replace Daily Models (Energy, Zone peak, and Coincident peak) with Hourly Load Models**

   The need to capture the impact of solar, EV, and other technologies that are reshaping demand requires an hourly modeling framework. Replacing the set of zonal daily models with the hourly model described in the report will meet this need. PJM should utilize the hourly rolling weather approach with two-part heating degree and cooling degree variables. PJM should interact these weather variables and other hourly model variables with heating, cooling, and base-use indices developed from the SAE models.

4. **Adjust Loads for Solar and New Technologies Through the Simulation Process**

   To correctly account for solar, EVs and other load adjustments, the hourly projections for these technologies should be constructed to be consistent with the weather simulation process. Each load simulation can then be adjusted appropriately to reflect the impact of solar and other weather-sensitive technology adjustments for each simulation. The load impact of EVs and other non-weather sensitive technologies will also need to be adjusted within the simulation process, as the impact of EVs and other technologies on load depends on the net of solar simulation outcome. The adjusted hourly load simulations can then be post-process to derive

zonal adjusted peak and energy and coincident peaks from the aggregation of the net zonal hourly load forecasts.

5. Capture Increasing Temperature Trends

Long-term temperature trends should be evaluated for each of the planning zones with results used to adjust cooling and heating indices that are inputs in the hourly load models. We expect to see increasing temperatures across the PJM service area that will contribute to an increase in cooling requirements and a decrease in space heating loads. Zone-level temperature trends can be used to construct trended HDD and CDD that are in turn incorporated into the heating and cooling model indices.