NERC

NORTH AMERICAN ELECTRIC
RELIABILITY CORPORATION

Data Collection

Approaches for Probabilistic Assessments Technical Reference Document

June 2021

RELIABILITY | RESILIENCE | SECURITY

3353 Peachtree Road NE
Suite 600, North Tower
Atlanta, GA 30326
404-446-2560 | www.nerc.com

Table of Contents

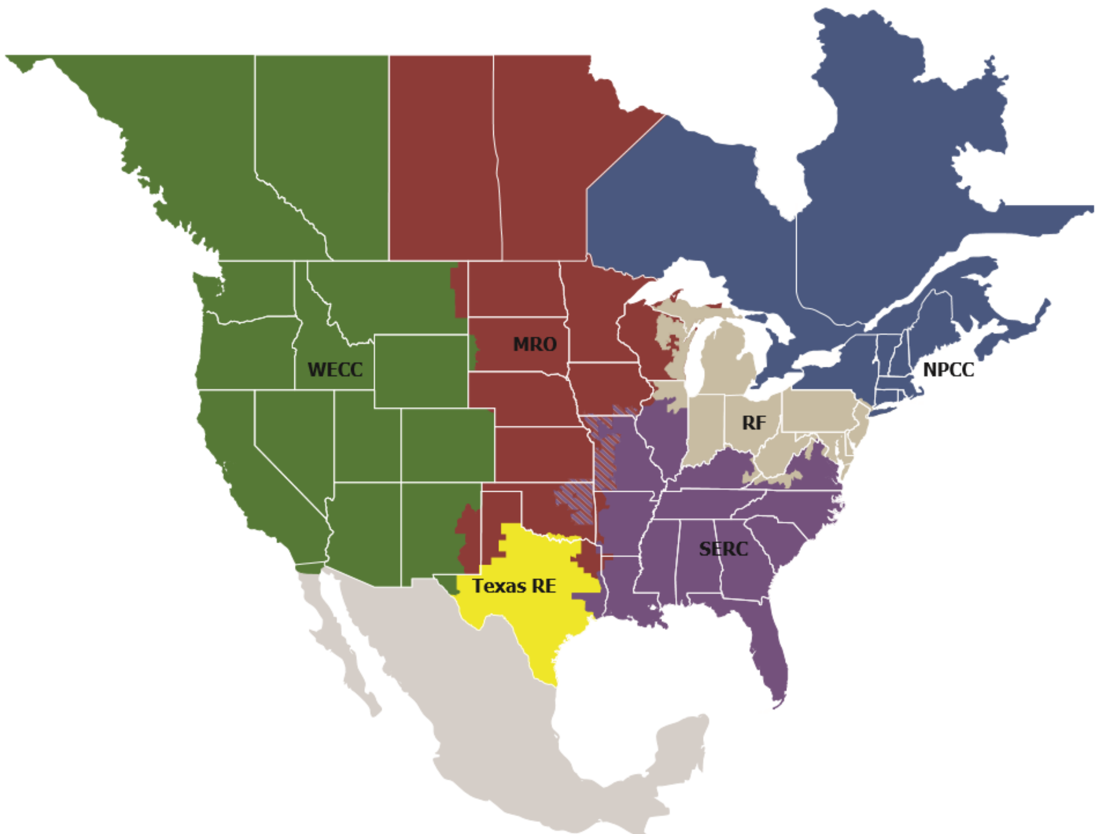
Preface

Electricity is a key component of the fabric of modern society and the Electric Reliability Organization (ERO) Enterprise serves to strengthen that fabric. The vision for the ERO Enterprise, which is comprised of the North American Electric Reliability Corporation (NERC) and the six Regional Entities (RE), is a highly reliable and secure North American bulk power system (BPS). Our mission is to assure the effective and efficient reduction of risks to the reliability and security of the grid.

Reliability | Resilience | Security

Because nearly 400 million citizens in North America are counting on us

The North American BPS is made up of six RE boundaries as shown in the map and corresponding table below. The multicolored area denotes overlap as some load-serving entities participate in one RE while associated Transmission Owners/Operators participate in another.



MRO	Midwest Reliability Organization
NPCC	Northeast Power Coordinating Council
RF	ReliabilityFirst
SERC	SERC Reliability Corporation
Texas RE	Texas Reliability Entity
WECC	WECC

Executive Summary

This document determined the general data input categories that are commonly used in loss-of-load probabilistic assessments across industry in order to build a technical reference for resource planners when performing their duties. These data categories include data considerations with a focus on parameters and collection methods for demand, thermal resources, energy-limited resources, emergency operating procedures (EOP), and transmission representations. Entities must consider procuring or obtaining enough data to accurately represent the model parameters or inputs to effectively develop and run a probabilistic reliability study.[1] An entity wishing to conduct a probabilistic study should thoroughly review these data inputs, the technical nature and aspects of the model inputs in study, and the soundness of the results with all stakeholders as a standard operating practice.

This document separates each of the major categories, (e.g., demand, generation, and transmission) in a resource adequacy study and highlights the types of data, possible sources for the data, and other qualifiers associated with the inclusion of such information in a probabilistic study.

Key Points and Possible Future Work

The Probabilistic Assessment Working Group (PAWG) identified the following key points in data collection across many different portions of a probabilistic resource adequacy study:

- Collection of weather data and any portion of the resource adequacy study related to weather should have the samples taken in the same period. If samples are not able to coincide, a cross-correlation calculation can help probabilistically align when the weather data sample was taken and when, for instance, the demand sample was taken.
- An in-depth understanding of operational characteristics of the resources represented in a study is needed to determine the requested data points in order to study the resource.
- Resource performance during ambient conditions (e.g. cold-weather or hot-weather performance) is of particular concern. Resource performance should be consistent with the assumed weather-related conditions in the case under study.
- Data collection for transmission systems in probabilistic resource adequacy assessments depends on how detailed the transmission model is represented in the study. This dependency between the quantity of data needed for the transmission elements is over and above the normal dependency that other portions of a probabilistic resource adequacy study.
- Battery energy storage systems (BESS) can be modeled similarly to other energy-limited resources, such as pumped hydro, with an emphasis on understanding the operational characteristics of the BESS.
- Planning Coordinators (PC), Transmission Planners (TP), and other modelers require access to detailed information in order to build and maintain their models for use in probabilistic studies.

The PAWG also highlighted the following objectives for possible future ERO work to be further explored and addressed as needed:

- When utilizing the Generation Availability Data System (GADS) or other historical outage reporting data, the thermal resources future outage rate may not be adequately represented by use of this historic data, especially when the facility moves to different operational characteristics. A thorough review should be done before using historic outage data when representing future risk.
- PCs, TPs, and other entities should work to gain access to data not otherwise made available that may affect the results of their resource adequacy studies or assumptions. Some entities do not have access to data sets

[1] In terms of reporting results and the metrics associated with probabilistic studies, the PAWG has published a separate document here.

to feed their models, and the need for more accurate studies may require access to data outside of those publicly available. This is paramount as resource planners are not able to perform studies without well-developed models that require a wide range of data.

- Careful understanding of data source assumptions and restrictions should be used when vetting a new or previous data source.

Introduction

Today's electricity industry is under a period of significant transition. The ERO Enterprise notes several high-level trends that have affected the North American bulk power system's (BPS) planning and operations, such as the continued retirements of traditional baseload resources accompanied with the proliferation of renewable and other forms of variable generation. These trends have highlighted an increasing need for the industry to properly model, study, and plan for the future state and reliability of the grid. The ERO Enterprise recognizes that these trends are highly variable and carry increasing uncertainties that further emphasize the need to enhance the traditional and deterministic forms of resource adequacy and reliability assessments. As was identified in the *2019 NERC Long Term Reliability Assessment* (LTRA)[2] and the *2019 NERC State of Reliability,*[3] NERC looks to enhance its resource and transmission adequacy assessments by incorporating more probabilistic approaches into its mission of a highly secure and reliable BPS. NERC continues to promote the use of more probabilistic approaches into reliability assessments by providing further insights into assessing the adequacy and reliability of the BPS.

The NERC Probabilistic Assessment Working Group (PAWG) was tasked to explore and highlight the current data collection processes across the industry that are used to produce loss-of-load probabilistic studies that assess emerging reliability risks. This document explores and identifies requirements, sources, and techniques for obtaining and modeling data for possible usage in conducting probabilistic assessments. The objective of this document is to discuss and raise awareness of probabilistic methods and techniques available to help entities conduct reliability assessments of systems with resources of increasing performance uncertainty. This document supports the PAWG's goal to promote the usage of probabilistic techniques and studies.

While NERC has historically assessed resource adequacy by using deterministic Planning Reserve Margins, the purpose of this document is to discuss data collection considerations for a probabilistic assessment. The intended audience is the industry at large with the objective of raising the collective awareness of available data collection methods. This report is written as a reference document for industry participants to understand the options available for these data sources and to highlight any benefits or considerations that these methods require.

In the spring of 2017, the PAWG conducted a survey of registered entities to better understand their assessment capabilities and identified challenges as they relate to probabilistic resource adequacy assessments. One of the recurring themes in the survey responses was the challenges with selecting and managing large sums of data in order to develop realistic inputs to probabilistic models. The *2019 NERC LTRA* key findings indicate that future probabilistic assessments should incorporate the increasing uncertainty of resources and demand while also considering the increasing amounts or sources of data. The PAWG has developed this document for entities that wish to or are engaged in conducting probabilistic assessments. The PAWG welcomes and invites subject matter expert discussion and comments on this document to further develop widespread industry participant knowledge, application, and acceptance of probabilistic studies and methodologies to assist in meeting the challenges posed to the electricity sector. This document is intended to complement ongoing industry work as there may be other groups that rest outside of NERC that are engaged in data collection discussions and probabilistic approach developments. As technical discussions and methods evolve further, the PAWG will update this document to meet industry needs.

There are numerous public and private sources of data that entities (e.g., PCs or TPs) can use to develop a probabilistic study. NERC plays a valuable role in providing some of these sources via the NERC GADS and Transmission Availability Data System (TADS); however, these are not the sole sources of data for a probabilistic study nor are they sufficient for every probabilistic reliability study. Many NERC REs and registered entities utilize different models for their probabilistic reliability studies, and this document attempts to summarize the collective approach and basic data

[2] https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_LTRA_2019.pdf
[3] https://www.nerc.com/pa/RAPA/PA/Performance%20Analysis%20DL/NERC_SOR_2019.pdf

needed to perform this work. Depending on the tools available to the entity, additional data from other sources may be required as the models available to that platform may require more information than the data source collects.

Chapter 1: Demand

Demand modeling in a probabilistic resource adequacy study is typically conducted through a combination of several inputs, including the utilization of historical data, demand forecasts, uncertainties, and assumptions specific to the system under study. Demand or load shapes can be modelled on historical monthly or hourly peak demand profiles and shapes and scaled to reflect forecasted conditions. In many cases in this chapter, the words "demand" and "load" may be used to reflect the modeling of end use customer MW draw. In the case of demand, the emphasis is on the MW amount and its time distribution while the term load can encompass other complexities outside of demand that may indirectly capture demand acting as a resource to offset the electrical system's draw at that time. Some models may not have the complexity to identify the nuances between the two terms, or some definitions may not be as clear as the above distinction. However, in terms of the data, most of the sources and procedures will not vary between demand or load, so the terms can be used in the following chapter interchangeably.

Demand Considerations

In a probabilistic resource adequacy study, accounting for specific assumptions regarding the amount and uncertainty of demand plays a significant impact on probabilistic indices results. Entities should consider the use of multiple demand-level scenarios in assessing the resource adequacy of their systems under study. An example of these demand levels could be specific forecasts, such as 50/50 or 90/10 system forecasts that represent the probabilities of exceeding explicit levels. Different techniques can also be employed by using statistical calculations, such as probability-weighted averaging. Probability-weighted averages calculate load level indices with corresponding probabilities of occurrence, representing the uncertainty in system demand due to external inputs, such as weather and economic factors. An example of this could be by using distributions of monthly peak demands versus the annual system peak demands. The selection and usage of multiple load levels can assist entities in planning against uncertainties, such as the occurrence of more extreme demand conditions or extended stressed system conditions. To gather some of these selections, a demand curve can be developed. To build demand curves, the RTO/ISO can utilize their metered data as the granular data provides an easy way to sample the demand.

Demand Curve Selection

Demand can follow many different socio-economic causes that would shift the shape of the demand curve in a multitude of ways; however, weather or climate is commonly identified as a primary driver of demand impacts. To help mitigate this, the demand curve should be constructed by considering the impact of differing weather conditions to better capture temperature sensitivity. Some of the considerations for selection can include ambient temperature for seasonal conditions, wind speed, and precipitation. Each of these meteorological markers has demonstrated impact onto the demand curve and should be considered when gathering data surrounding demand during those time periods. Specifically related to the curve construction, the peak, nadir, and ramping rates have substantial influence on the reliability impacts to the system in study.[4] Accurate characterization for those periods is important for the planning and scheduling of generation and ancillary resources during the study.

Because the resource planner desires to capture a full distribution of possible demand conditions, the demand curve selection is important when collecting a proper sample of data. These conditions include cool, average, hot, and extremely hot summers; warm, average, cold, and extremely cold winters; and low, average, and high meteorological conditions, such as irradiance or wind speed. These will emphasize some of the peaks, nadirs, and ramping rates. Accurate characterization of the identified risk depends on the samples taken and the selection of the curves those samples produce. For instance, if the demand data collected contains 25 years of curves, selecting the curves that accentuate peaks, nadirs, and ramping rates will allow the resource planner to more accurately capture the anticipated risk conditions of the peaks, nadirs, and ramping rates. In the same light, selecting all the curves will weigh all years as equally probable.

[4] Historically, the planning process typically accentuated peak conditions. As risk moves away from the on-peak periods (over a season or a day), looking at curves that accentuate other aspects of the demand curve is warranted.

Load Scenarios

Loading level directly determines the required amount of resources in the study due to the load and generation balance. In addition, the load level and composition play a significant influence on the system in study. When performing a resource adequacy study, a TP/PC must select the appropriate scenarios that either stress or relate demand to differing extreme conditions. In order to do this, planners will need to gather demand data associated with the weather conditions specified above. More specifically, this will be a distribution of load scenarios across demand curves. One example distribution is cool, average, hot, and extremely hot summers along with warm, average, cold, and extremely cold winters. To ensure a diverse number of scenarios are available in the study, combine scenarios with high, average, and low wind speeds with high, average, and low precipitation (or water flows). As many of these scenarios are study dependent, the specific study scope can assist in either paring this list down or adding to it. Additionally, sensitivities can also accentuate specific loads and can assist the planner in studying the impact on their system. For example, a load scenario that assumes very aggressive electrification of the transportation system will accentuate the usage of demand during the hours in use as well as on the days of the week that transportation is more heavily used.

Load Forecast Uncertainty Considerations

Realized load can differ from projected load for multiple reasons: First, weather cannot be exactly predicted and will cause peak load to differ from the normalized-weather forecast (as discussed in the weather-related LFU section). Second, there are uncertainties in population growth, economic growth, energy efficiency adoption rates, and other factors. Data for these topics can be regulatory-based and would vary by jurisdiction and program. These non-weather drivers of load forecast uncertainties (LFU) differ from weather related LFUs because they increase with the forward planning period while weather uncertainties will generally remain constant and be independent from the period being studied.

Non-Weather Related LFU

From the above, the uncertainties in population growth and the associated demand forecast can be addressed by a statistical approach at quantifying the uncertainty. To best illustrate this, consider the following example: For each weather-year load forecast, five non-weather LFU multipliers are applied to all load hours. **Figure 1.1** shows the uncertainty as a percentage of the 50th percentile (P50 or “50/50”) peak load forecast, indicating that the forecast uncertainty increases as one moves further into the future. Each multiplier is assigned an associated normal-curve-based probability with the sum of the probabilities totaling 100%. **Figure 1.2** shows the three-year forward LFU multipliers.[5] To calculate the weighted-average results across all load scenarios, the weather-year probability weights and the non-weather probability weights are multiplied to create joint probability weights. More details about non-weather LFU can be found in other reports in the industry.[6]

[5] While the figure shows symmetric forward LFE, these points may not be symmetric.

[6] A few relevant reports are posted on the Electric Reliability Council of Texas (ERCOT) website, which contains material listed here: http://www.ercot.com/content/wcm/lists/114801/Estimating_the_Economically_Optimal_Reserve_Margin_in_ERCOT_Revised.pdf; http://www.ercot.com/content/wcm/lists/167026/2018_12_20_ERCOT_MERM_Report_Final.pdf; http://www.ercot.com/content/wcm/lists/114801/ERCOT_Study_Process_and_Methodology_Manual_for_EORM-MERM_12-12-2017_v1.0.docx

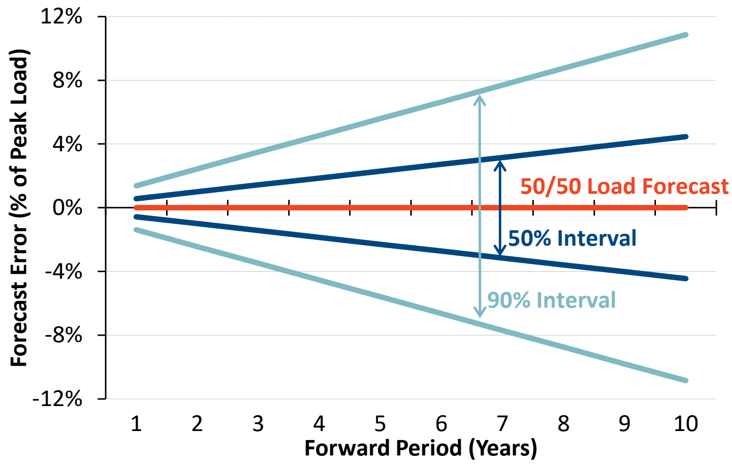


Figure 1.1: Non-Weather Forecast Uncertainty with Increasing Forward Period

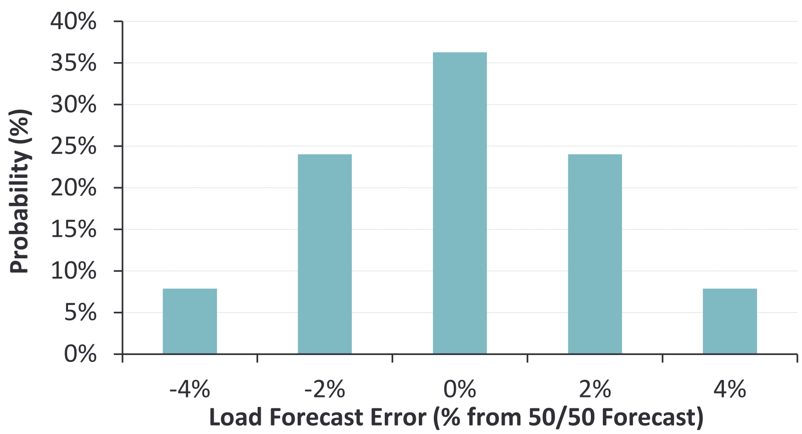


Figure 1.2: Three-Year Forward LFE with Discrete Error Points Modeled

Weather-Related LFU

While LFU methods have the ability to capture many uncertainties related to the load, weather factors are a significant driver of load and their uncertainties can be captured when undertaking a probabilistic assessment. Weather-related methods can be utilized to capture the uncertainty with respect to year-over-year differences. Typically, weather-related LFU captures the variance of conditions documented in the historic conditions. If the resource adequacy study simulates extreme conditions outside of what historic conditions can predict (e.g., sustained higher than record wind speeds), the resource planner will need to adjust or produce data that captures those conditions.[7]

Some data points to consider are ambient temperature, dew point, wind speed, and cloud cover across a variety of stations in the assessment area. These variables have been determined to relate to the variance in load, and one of the sources of data on those variables is from weather stations. To provide enough accuracy to depict the weather-related LFU, multiple years of weather are required to capture this uncertainty. The Independent Electricity System Operator (IESO) currently uses 31 weather years and runs the load model forecast on those years shifted up to seven days to account for each numeric day falling on a given day of the week. This is to allow each calendar day to fall on

[7] Note that this type of adjustment would need to be study-wide in order to have consistent study conditions for these extreme weather scenarios. This does not, however, adjust the data collection technique for weather-related LFU.

each day of the week. That is, calendar day 100 will lie on a Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday to account for differences the load has based on shifts due to being a particular day of the week. This equates to 465 distinct weather simulations,[8] and the LFU could be determined from there. Other entities, such as Argonne National Labs, have taken the information at weather stations and numerical weather prediction (NWP) data to determine the weather-related LFU.

SERC gathers this weather information from FERC Form 714 Part 3, Schedule 2.[9] This source is by no means the only resource for weather-related uncertainty as there exists data through metering at the ISO/RTO level. The ISO/RTO granular data opens up more ways to construct the LFU similar to the benefits in the **Non-Weather Related LFU** section above. The FERC data source requires that the electric utility planning area provide hourly demand levels in megawatts and the source starts at year 1993 for some portions of the database. The format changes based on the year as per **Table 1.1**.

Table 1.1: FERC Data Source Format

Reporting Year	File Format	Notes on Use
1993 to 2004	.zip files organized by reporting year and NERC regions (legacy and current). Microsoft Windows compatible programs to read spreadsheet and text files, there exists a file that needs conversion in the archive, but many programs exist to convert to Microsoft products. Each entity has a separate format for each	Ensure that the data conversion for .wk1 files can be converted to Microsoft Excel. No database viewer exists and the user must download one in order to view the data. Conversions for analysis regarding multiple entities are needed to ensure the data gathered is uniform in the study.
2005	Similar to 1993 to 2004	Individual Entity filings can be viewed through the FERC eLibrary.
2006 to present	All responding entities have the data and have the .zip archive to download. That archive contains .csv file formats	The FERC form viewer is able to visualize fully the data prior to download. This year a unified format is applied across entities.

It is suggested that the data be converted to a daily hour ending (1–24) matrix format. In order to perform that conversion, a few cleansing techniques can be utilized. Associated hourly trends and other whole filling algorithms will help to complete the database when holes or incompatible formats occur when adjusting time zones. FERC has placed a relational database viewer to assist with the collection of this data. See **Figure 1.3** for the database schema provided. Additional screening approaches to detect anomalies with the data that include outlier detection are also needed to ensure a good quality data set prior to utilization in the study.

[8] Seven days forward, seven days backward, and the day that the historic measurement was taken multiplied by the number of years. For 31 weather years, this is (7+7+1)*31 = 465.

[9] https://www.ferc.gov/docs-filing/forms/form-714/data.asp?csrt=18240670882965036364

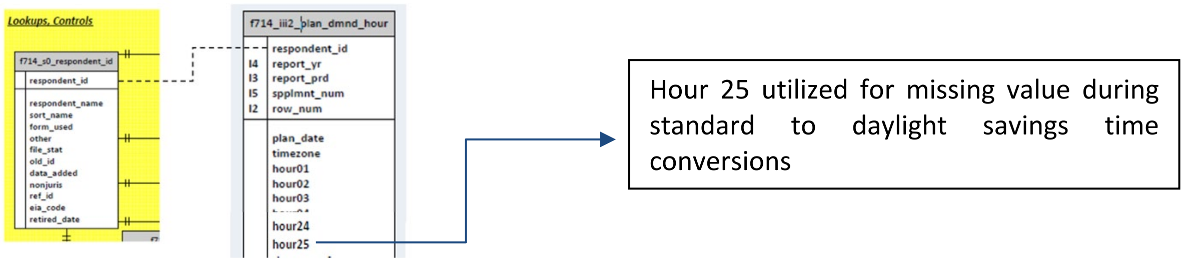


Figure 1.3: FERC Database Schema

In addition to just these hole filling requirements and other changes as required for outlier detection, additional screening approaches are needed to reconstruct the data relationships. An example of what SERC has done to additionally adjust the FERC database forms can be found in **Figure 1.4**. As shown, the additional approaches can impose a slight difference between the NERC Long Term Reliability Assessment (LTRA) data and what is filed in the FERC database. For probabilistic studies, it is best to use the data in the LTRA (i.e. post additional screening) in order to calculate the weather-related LFU.

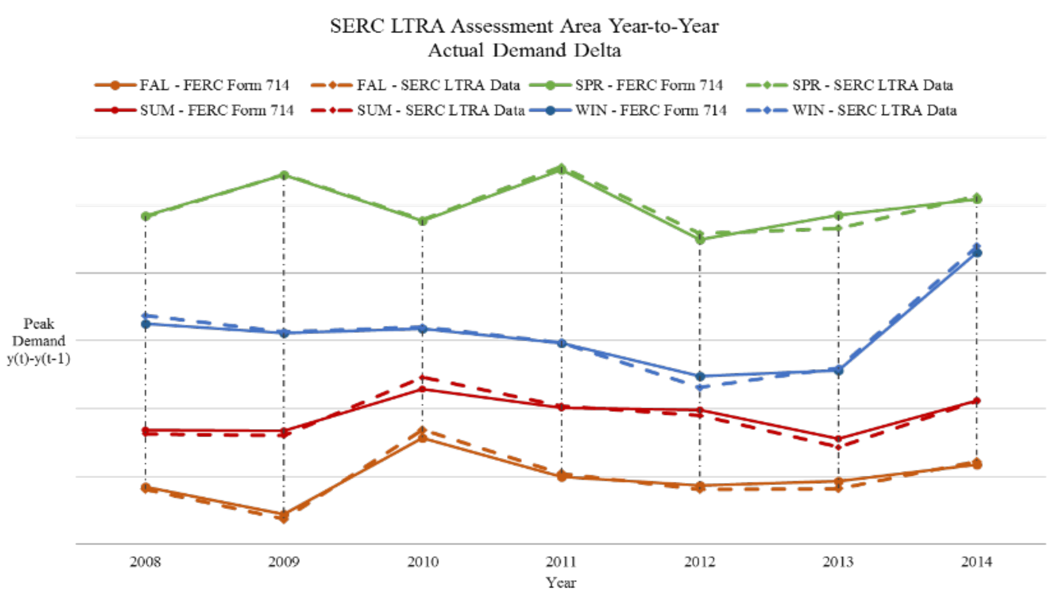


Figure 1.4: SERC Adjustment Example Utilizing FERC Databases

Complexities in Modeling Demand

While the basics of demand modeling in probabilistic studies is detailed above, multiple issues arise when allocating operational characteristics and other contractual obligations into the probabilistic study. Some of these complexities arise especially during the NERC probabilistic assessment process and are reflected in the following sections.

Modeling Multi-Area Systems

Entities should consider the correlation of peak demands with neighboring area systems in developing composite load shapes. These periods, perhaps due to heightened weather or economic conditions, represent high degrees of peak load correlations and the highest amount of coincident demands. The highest peak demands represent a conservative assumption about the ability of the system to meet demand without assistance from neighboring systems to meet peak loads. To capture this in the probabilistic study, load shapes from different assessment areas' geographic boundaries should have the same time frame as the study. Sometimes these areas change their boundaries; however, the goal is to stay consistent across the study in terms of the data quality that feeds the different areas in the study. In cases where the boundaries of the probabilistic study cross different assessment areas, data should be coordinated to capture the system coincident peak as a composite of the many areas in the study.

Demand Response

For the probabilistic incorporation of demand response, the particular mechanics of each program or structure will dictate the utilization of the demand response. Primarily of concern is the amount of load relief the demand response provides at every stage, the number of times the resource can be called in a given period, and any other limitations on the duration or amount of relief the response.[10] For areas where this is required to be registered, the above information can be found on the registration forms; however, not all areas are able to provide those registration forms. In the areas where demand response is registered, the program can define many of the parameters; however, historical usage information can solidify the amount of load relief at each demand response tier. This historical usage, however, may be affected by more parameters than just the load relief as certain connections or disconnections affect the availability of the demand response to achieve the load relief. As these are quite complex, the PAWG recommends using a data source that captures operational conditions surrounding demand response in order to capture any cross correlations or to calculate them otherwise.

For demand response that is registered, the amount of relief, number of times it can be called, and other duration limitations or restrictions are found in their registration forms to enter into their respective databases. For unregistered resources, resource planners are encouraged to use methods to predict their availability by analysis of past performance and heuristics going into the future to obtain these values. A quick overview of the data inputs for demand response appears in **Table 1.2**.

Table 1.2: Information Required for Demand Response (DR)

Information Required	Example Collection Source
Amount of load relief	Registration database that aggregates the load relief or informal survey to non-registered devices
Number of times in a given time period demand response can be called	State/Provincial level orders or similar utility contracts regarding Demand Response
Duration limitations	State/Provincial level orders or similar utility contracts regarding Demand Response
Tiers of response	State/Provincial level orders or similar utility contracts regarding Demand Response
Other restrictions	Utility specific directives, databases on controllable loads

In addition, there are market structures that contain different levels of this type of demand response. These are sometimes labeled as emergency response services but are going to vary by when they can be called and in their service responses. Supplemental collection of similar sources[11] should be utilized to capture these tiers of response.

Demand and Demand Response as a Resource

When demand response is modeled in the demand profiles themselves, adjustments to the demand profile will apply to the demand response. Conversely, when demand response is modeled as a resource, it is not included in the demand profiles and is not included in any of the alterations in the demand section. To further clarify the difference, when DR is modeled as a resource, adjustments from weather or non-weather related LFU will be only on demand

[10] For example, one of the more difficult considerations in demand response is the expected performance versus actual performance during extreme temperatures.

[11] State or other regulatory bodies as well as other internal sources may manage these sources.

rather than both the DR and the demand. This can then be directly used in the study as all of the adjustments are on demand curve. When modeling DR as a resource, these techniques need to also be applied to the demand response modeled as a resource as such LFUs will impact the key model parameters in **Table 1.2**. The data source chosen to provide the LFU should be flexible to adjust for either modeling scenario. The key point for this separation is to ensure any adjustments to demand are adjusting the operational characteristic of the demand response or the demand, rather than both. If separating the demand response as a separate resource, then the collection of data may require more data than just the amount of load relief at any given time in the simulation and may require time-of-use or other operational profiles to determine in-simulation output of the demand response when called upon.

Distributed Energy Resources

Distributed energy resources (DER) can be a multitude of differing resource types connected through the distribution system; however, many of the current installations are solar photovoltaic. Some probabilistic studies utilize simulated profiles as a load modifier in performing the load forecasts. In some areas, DER forecasts may be available, but these forecasts are generally at the state regulation level. As such, the forecasts may vary between assessment areas and could even vary internally to the assessment area if such boundaries cross state lines. Such data can be valuable to the planner when performing the probabilistic assessments, but care needs to be taken such that the DERs are not double counted in the demand portion of the study. That is, if a load modifier for simulated profiles are used, additional forecasts should not double count this modifier. See the section on **Generation Availability in BTM Generation** to see the setup for modeling DERs as explicit stochastic resources. The difference with current DER technology, however, is the high correlation to irradiance for their availability. With this high correlation, weather-related data as demonstrated above could supply another marker into the DER's availability. These types of resources use a mix of demand techniques as seen above and parameters seen in **Chapter 2**, and as such similar data sources can be expected when modeling the DER in a probabilistic study. As there is no current database or source for DER availability, a mix of operational data and weather data can be expected to model each state of the stochastic representation.

Data Validation & Cleansing

Once data are formatted across all reporting years, entities should consider performing data reviews and validations as well as post-processing work if the data are large to ensure the underlying data in question is of sound quality. These validation and cleansing methods are not just relegated to demand data and are summarized generally in **Appendix A**.

Demand Reconstruction under Boundary Changes

FERC 714 filings are housed in a central resource so an entity can import the same submitted demand data into the resource adequacy study; however, this imposes an issue where an entity's boundary changes or is under study in a different boundary. These geographic changes will require some reconstruction of the demand in each area in order to maintain the same level of demand uncertainty across the entire study region. Two options generally exist: a time-series reconstruction or a comparison of the peak demand in each area creating a ratio. The former is more time intensive but provides a greater level of accuracy for the added or reduced demand based on the geographic change. The latter option provides a quicker way to adjust the demand shape in the study but assumes that the peak ratio is valid for all times in the year. This creates a less accurate depiction of the demand change.

Demand Data Requirements

The data requirements for use in resource adequacy studies revolve generally around the time granularity of the data. An hourly representation of demand levels is required for most studies, and associated databases may or may not have such hourly representations. In such instances, hole-filling programs and other trend-based algorithms can fill the gaps associated with transferring the data into hourly format. This is crucial as some of the current metrics the PAWG has in their previous reports have the metrics in an hourly format. The reader is reminded that many databases may not have the greatest quality of data; however, such data could be sufficient for their report or study. Such databases simply require the post processing methodologies as discussed in the SERC example in **Figure 1.4**.

Collection methods

There are varieties of both sources and mechanisms for which data can be acquired and utilized for conducting probabilistic assessments. The specific data needed can vary significantly depending on the type of assessment as well as the underlying characteristics of the system under study. Aspects that potentially affect the availability of data include the status of local, state, and federal regulatory frameworks; market construct and available operational data; underlying resource mix and trends information; and/or agreements or tariffs with other registered entities. For NERC registered entities that conduct probabilistic assessments, data sources being utilized vary by jurisdiction and applicability to their respective systems. A summary table of the various types of collection sources for different types of entities is found in **Table 1.3**. It is anticipated that other data sources exist for this data, and the table is provided as a start for collecting the type of data.

Table 1.3: Data Collection Notes on Different Entities		
Entity Category	**Entity**	**Notes on Data Available**
Federal, State, or Provincial level	U.S. Energy Information Administration (EIA)	The EIA contains a lot of valuable information on various energy products, including: LNG export, generation capacity, and an hourly grid monitor. Care must be taken to gather the source of data, or to understand the assumptions associated with the reported charts, graphs, or other tools.
	National Renewable Energy Laboratory (NREL)	The data available contains maps, models, and tools used for energy analysis. Specific ones help with association of data and others are tools to feed probabilistic studies, such as weather data.
	U.S. Census Bureau for U.S. based regions and Statistics Canada for Canadian areas	The data here contain population and census data in particular areas. Additionally, it collects and publishes nationally commissioned data on such populations.
	Public Utilities Commissions	These entities can provide state, provincial, or local agency data specific to energy and resource type.
NERC Registered	Generator Owners or Generator Operators (GOs/GOPs)	Generation entities can report their outage information to the NERC GADS, and in cases where more information is required, can assist in determining their generation availability. The latter is especially true for newer plants.
	Distribution Providers (DPs)	These entities provide their distribution system to serve end-use customers. These entities are able to provide information on their served demand
	Transmission Owners	These entities are the owners of equipment for the long distance transmission of power, and may be able to provide outage information related to the equipment they own. For example: transmission lines and transformers
Operations/Market	ISO/RTO Capacity Markets	Each ISO/RTO provides an outlook on the anticipated socio-economic changes and some of them provide outputs usable in probabilistic studies

Chapter 2: Thermal Resources

A large majority of resources in the BPS are thermal resources that convert chemical energy into electrical energy by burning fuel. These resources can vary dramatically in construction, so the focus on data collection for reliability studies is on modeling the availability of the generation and the assumed operating level of that generation. In general, a two-state Markov model is the end goal for these types of resources so data collection will center on gathering enough information to fill the model. As other models exist, this section will detail the many sources of filling out any type of stochastic model.

Outage Data

Outages must be considered for all resources in conducting probabilistic assessments as outages have the ability to materially affect the availability of generators to meet the demand. NERC Registered Entities typically utilize a combination of data sources[12] to account for planned and unplanned outages along with their associated uncertainties. These typically include a combination of historical information, performance, and potential correlations to weather data. Some of the types of forms used for the information include generator availability[13] and outage rates (NERC GADS), such as the equivalent forced outage rate, FERC 714 hourly reported data, and market data. In addition, some selected entities utilize a combination of forecasted resource price data and powerflow studies or perform regression analyses for potential correlations with outside datasets.

Key Takeaway:
Building the outage rates of thermal resources requires forced, planned, maintenance, and other outage types. A single data source may not have all the types of outages.

For thermal resources, the majority of the outage data required to formulate the equivalent forced outage rate will require a data source including parameters for planned outages, maintenance frequency and length, and forced outages, including repair and failure rates.[14] The parameters associated with the planned outages include the maintenance cycle and length, usually related to the months of the year (i.e., two of the twelve months) and the length of days associated with that outage. Cases exist where the planned outages can have durations across years and such cases will require care so that the durations are related to yearly outage metrics. In addition to these planned outage inputs, the parameters associated with the forced outages include full outage mean time to repair, full outage mean time to failure, and partial outage deratings for however many derate states there are. For partial outages, the critical component is hours for MW unavailable, no matter the derate type. The sum of the MW unavailable is the critical component for a resource adequacy study. As a second objective, grouping these derates by event type can be informative for model or data validation.

The data source for the forced outage rates can be fulfilled in the NERC GADS database; however, that data does not include reported planned outages and is a calendar-reporting database where multi-year events may have differing unique identifiers. To account for those differences, supplemental information is required to bridge the gaps. In an informal poll by PAWG membership at their meetings, many of the companies contain an internal data source that accounts for the planned outage data. Some of these functions are not in the planning departments, but rather in the operational departments. When using operational tools, it is important to remember that the data may need to be altered in order to account for errors incurred while logging the planned maintenance records. Additionally, a

[12] These data sources may be quite large. For instance, ANL has over 650 million records of customer outage data sampled at about every 15 minutes.

[13] Depending on the generation model, EFOR versus EFORd will demonstrate if the plant was in demand when the outage occurred for use in determining the generation availability. The NERC Performance Analysis Subcommittee has identified that the NERC GADS does not have enough information to calculate the EFORd modeled outages using that data source only. As such, the resource planner needs additional operational data if using this in the model.

[14] These data sources are used to develop an estimate for the FOR of the unit. IEEE Std. 859-2018 describes the statistical modeling concerns surrounding the use of point estimates or averaging of results as well as the assumption of independent outages across the generation fleet.

Canadian Electrical Association (CEA) reliability database can also provide the statistics regarding thermal outages that aren't related to event based performance sources, much like the NERC GADS. In each of these sources, cleansing of the data in order to align the information submitted to the database and aligning it with the records found in operations that take on these derates. This type of cleansing may require knowledge of the model[15] in order to align the transition rates with the submitted and forced conditions.

When utilizing the NERC GADS database, a few other peculiarities exist for thermal units as the reporting for units may not be consistent across the database. For units with a high startup rate, taking startup outage out of EFOR is a more appropriate way to model the stochastic nature of the unit; then the resource planner can utilize that reduced EFOR for those units. The startup failure rate may show up as a derate or as an outage rate. An additional consideration exists for NERC GADS: the database is set up for the immediate time frame, meaning that using it as a data source for derates will only provide the reduction of MW from the current ambient conditions. For some thermal units, this is not an adequate indication of the starting point as some units are highly sensitive to ambient temperature. For these units, additional data in the form of a temperature curve assists in developing their stochastic model.

For entities that do not use the GADS data, such as the IESO, they have an internal database that takes into account all outages (submitted, forced, and approved) on a per generator basis. Other entities also maintain an internal central database for this data. Generally, those entities utilize a set of samples from historical databases and submit planned outages to forecast the generators outage data for the study. This outage data is similar to the planned outage databases discussed above. Similar conditions exist to ensure data accuracy with reporting of planned outages in this type of system as well as the forced outage data. For the IESO, the planned outages are modeled as a part of future planned outages, 10 Year Forms with projected outage schedules, and historical planned outage rates. By collecting the data in one source, IESO is able to model their thermal resources.

Perspectives on Predictive Outage Forecasting

> **Key Takeaway:**
> Predictive outage schedulers provide methods to forecast outages in future years where the planner conducts the probabilistic resource adequacy study.

Historical GADS data collected by NERC is a common and standard data source for entities modeling conventional generation.[16] Operational schedulers can also be a source of this information, and the Control Room Operations Window would be another valid data source for predictive outage forecasting. However, access to the information within this database can be challenging, and unit-specific information is not accessible to all entities.[17] An alternative way to obtain the data is by requesting it from resource entities directly. A specific example for requesting GADS data from resource entities, including the data request notice and data submission form, can be found in **Appendix B**. Since conventional generation outage trends may change over time, it is useful to predict outages in planning studies. An example of such is in ERCOT, where staff reviewed several predictive algorithms (e.g., the Prophet[18] tool developed by Facebook) to determine its usefulness in capturing changing trends. A predictive forecast approach based on Prophet[19] has been tested to forecast fleet-wide forced outages. For unit-specific outages used in probabilistic studies, the predictive approach may not be applicable. Based on the ERCOT's experiences with such data sources, the predictive approaches can help visualize the nature of the combined historical and planned outages to provide a way to more accurately collect the correct outage rates to apply to the study. To fuel a stochastic model, these predictive outage-forecasting tools should include mean time to failure, mean time to repair, mean time between failures, and other transitions between the stochastic states to be an effective data source.

---

[15] Such as the distinction between two-state and multi-state Markov models for thermal resources

[16] These databases log historic outage data to calculate their availability. There are conversations on the use or nonuse of historic data in predictive probabilistic studies found in IEEE Std. 762-2018 and IEEE Std. 859-2018.

[17] Only entities authorized to view unit specific data are allowed access to that data due to the sensitivities surrounding the data.

[18] A link to the tool can be found here

[19] Link for the Prophet tool can be found here

Data Considerations for Capacity Constraints

> **Key Takeaway:**
> Many of the capacity modifications are highly model dependent, indicating the need for varying data source requirements. Data collection should be considered on a case-by-case basis.

Outside of planned and forced generator outages, there are other factors that can also affect supply availability that must be accounted for in reliability assessments. Factors like emissions constraints, unit deratings, fuel availability, and capacity constraints all limit the availability and ability for supply-side resources to meet demand and can have wide implications for reliability, especially in extenuating weather or stressed system conditions. Additionally, some future market conditions may impact the capacity or dispatch of a unit where the operational characteristics of a thermal generating unit are subject to market influences. Some of these constraints can be found in the source documents that dictate the market rules or in the regulatory body that imposed the rules in the present or future market.

Emissions Constraints

Entities must account for the potential application of emissions if they plan to model these constraints in their resource adequacy studies. Some of these constraints are taken into account during economic dispatch of the units, while other models require explicit states modeled based upon the study conditions. Different government agencies regulate many of these constraints, so they are generally unique in each area. In general, the assumption for emissions is that the regulators will lift relevant constraints during blackouts or resource inadequate periods; however, these constraints can be adjusted by modeling the outage rates, capacity limits, and other water flow constraints in order to model the impact these policies have on specific generators. However, since the modeling varies, the amount of data required will vary as well. Resource planners are suggested to look to government agencies or emissions regulators in order to gather enough information to model the emissions constraints.

Fuel Availability Data

The NERC Electric-Gas Working Group has helped determine the interfaces and potential interdependencies that the electricity sector has with the natural gas pipelines and potential disruptions of those pipelines[20]. As it pertains to resource adequacy, the data required to model the impact of pipelines can be cumbersome and is not available in NERC GADS. The data source selected should consider mean time to failure and mean time to repair rates associated with those operating states. These general considerations are typically accounted for by using equivalent forced outage duration (EFORd) in some REs, but others do not account for this in the EFORd as that measure is typically reserved for mechanical outages. Similarly, the fuel availability statistics will need to account for the derate associated with lack of fuel. Due to these complexities, capturing this in a probabilistic study is cumbersome and will require more than usual amounts of data to perform a study. A resource planner will require access to pipeline outages and other natural gas information systems in order to model the impact on a resource adequacy study. In some very restrictive areas for fuel availability, a resource planner can consider modeling this thermal resource as an energy limited resource with considering some aspects of other energy-limited resources in **Chapter 3**. In particular, the available natural gas (in MBTU[21] per day) from a data source in these scarce periods is important to consider.

---

[20] Link to Electric-Gas Working Group report: https://www.nerc.com/comm/PC/ElectricGas%20Working%20Group%20EGWG/Fuel_Assurance_and_Fuel-Related_Reliability_Risk_Analysis_for_the_Bulk_Power_System.pdf

[21] This is a common measurement in the natural gas industry to indicate 1,000 British Thermal Units (BTUs)

Capacity Modification on Ambient Conditions

> **Key Takeaway:**
> Thermal power curves allow the resource planner to adjust the capacity based on the ambient temperature studied. Modeling ambient conditions also requires weather data close to the resource.

To capture the capacity modifications due to differing ambient temperatures, some entities send a survey to their Generator Owners with capacities at specific temperature points. These points provide a curve and that particular curve is used to set the capacity derates under the ambient conditions; the source of those ambient temperatures is the same as the Weather-Related LFU portion discussed in **Chapter 1**. The combination of these two provides a simplified method to model correlations between the weather and generator outputs for the forecasted short-term; nevertheless, the source for these model considerations stays the same: a survey to generator owners to generate a thermal curve and the weather-related LFU sources.

Other capacity modifications depending on the ambient conditions exist. Terms like "high sustainability limit," which ERCOT defined as the real time maximum sustained energy production of a resource; "dependable maximum net capacity," which is defined as the maximum power a resource can supply under specific conditions for a given time interval without exceeding thermal or other stress violations; and "seasonal capacity," which is the capacity of a resource in a given season, come into play. These terms all describe the energy restrictions on ambient conditions and constraints that would hinder the modeled generator in the reliability study from producing its nameplate value. Should this be a major concern in the study, the data source[22] chosen should be equipped to handle the desired study conditions and gather enough data on the constraint to model it stochastically. At minimum, this means determining the mean values for transitioning between the states.

Generation Availability in BTM Generation

Data sources for behind the meter generation will be highly model dependent, but there are a few considerations for these generators that typically do not report in surveys or other generator data sources. These types of resources sometimes can be found as a load modifier, but those resources can sometimes be sensitive to a market price or other dispatch signals and are thus not related to the electrical characteristics at their point of interconnection. To gather enough data on these types of resources, a case-by-case data structure will most likely be needed or a wide swath of assumptions to be made based on the available data to the resource planner. Two approaches exist for these generators: one is to net them against the load to which they are close geographically, which carries all the assumptions of demand modeling, and the other is to model these as discrete stochastic resources with a recommendation for a simple two-state Markov model that can be developed off operational data superimposed on other time-synchronized measurements to determine the resource's full capacity. If modeling via the latter method, the same data types outlined in this chapter are expected to be placed into the model, and as such similar data are to be collected. Collecting this type of data may be cumbersome for these types of generators, so heuristics developed off knowledge of these facilities can aid in determining when to collect the data to best model the resource.

[22] This may be a survey to the GO, as the IESO example above demonstrates

Chapter 3: Energy Limited Resources

Some of the common resource adequacy discussions are based on the capacity of resources and the availability of those resources to meet the level of demand in a study. In the case of energy-limited resources, such as hydro, wind, and solar, capacity related discussions are only one facet of reliability planning. This chapter focuses on the different types of energy-limited resources to describe how to collect data representative of them for use in a probabilistic study.

Hydro Units

The vast majority of hydro generating facilities are considered as energy-limited units since these facilities are dependent on the availability of water resource. The time constant for the availability of water may be longer than that of wind or solar. The effect of unit-forced unavailability is not significant on hydro generating system reliability; therefore, many resource planners incorporate this unavailability in estimates of energy limitations when conducting probabilistic analysis. Some of the input parameters for each hydro power plant are as follows:

- Installed/in-service, Planned and retirement dates
- Monthly maximum and minimum output of each plant
- Monthly available energy from each plant
- Energy distribution (available energy to hydro unit)
- Forced outage rate or EFORd

For hydro generating facilities, some entities may assume that the available water or fuel for each plant has little or no uncertainty or that the water resource is in a drought condition. This is a conservative approach to ensure that sufficient resources will be available when needed. However, if the uncertainty is to be modeled, the data needed to incorporate that into the hydro facilities model requires similar data to other weather-related energy-limited resources.

Simulated Solar Generation

In a loss-of-load probabilistic study, it is important to cover all of the weather years of data for resources highly correlated to weather data (e.g., solar photovoltaic (PV)). In order to do so, resource planners can simulate the expected behavior of the solar plant for use in their loss-of-load probabilistic studies, and many tools are available to augment or replace observed historical generation data for a particular resource or neighboring resources. One such tool is the Weather Research and Forecasting model[23] used to generate the historical atmospheric variables, such as wind speed, temperature and irradiance as well as snow, ice, or other ground cover, which in turn simulate solar power production at each location in the model. The most important data points to produce a simulated solar profile are the types of arrays, soiling, shade, snow or ice cover, and control parameters associated with tracking the solar bodies. One tool that utilizes these input parameters to then convert into ac capacity is the NREL SAM tool[24]. Other tools can produce solar profiles off generic adjustments. These tools take into account multi-order variables when producing the curves, but requires additional site-specific data that may not be available when conducting a resource adequacy study; however, it still remains an option for more specific profiles.

Key Takeaway:
Simulated profiles can be performed for both existing and planned solar PV sites. In either case, site-specific details help refine the fidelity of the profile. Some tools provide dc capacity and others ac capacity. For use in resource adequacy studies or assessments, an ac capacity will need to be calculated if the tool does not do so.

[23] Information on this model is available here
[24] Available here. See information on the PVWatts portion of the tool

To walk through the process, ERCOT computed the atmospheric values and adjusted them using surface station data and input them into a proprietary PV model to produce the hourly power output profiles. Programs mentioned above would also provide a profile, but ERCOT utilized proprietary models to accomplish the goal, yet another option available to resource planners. More details about developing hourly solar power profiles can be found in the solar profile methodology report, available on ERCOT's Resource Adequacy webpage.[25]

If utilizing site-specific information to inform profiles, data found in **Table 3.1** is useful for providing a program or vender when gathering simulated solar profiles. Some of the information is expected to be assumed as some can be site-specific and many of those parameters are not available at the time of study.

Table 3.1: Solar Profile Data Requirements

Category	Data Point	What to Gather
Static Plant Details	Installed Plant Capacity	Capacity of dc MW
	Tracking System Type	Fixed, single, or dual axis
	Tracking Origination	Azimuth, north-south, other
	Module Tilt	Horizontal, tilt to latitude, other
	Module Azimuth	Degrees off Azimuth
	Ground Cover Ratio	Ratio of array coverage by other arrays
DC to AC Conversation	DC to AC Ratio	Efficiency of dc to ac conversion in MW
Inverter Details	Inverter Capacity	either inverter make and model, or number of inverters and the inverter capacity
PV Module Details	Module Capacity	either module make and model, or number of modules per string and the module capacity

Key Takeaway:
Public resources exist to generate the simulated solar profile; however, non-public options exist for use as well.

Site-specific parameters are not required for these profiles; however, they provide a more granular approach to modeling the contributions of solar resources. In general, the solar profile is a time-series of data on the total power production (in MW) at a solar facility. Two methods exist for this. One is to gather time-series irradiance data and convert it to MW by collecting efficiency of the solar facility to convert that irradiance into MW. This conversion acts as the solar profile for a particular resource and the NREL database for United States entities contains many years of solar data for this purpose. That database can somewhat cover Canadian areas database, but meteorological data from weather stations may be able to supplement this. The other method is to take historical generation samples from another solar generation facility, gather irradiance data as above, and then merge the two in order to capture some other uncertainties not related to irradiance. Some entities use a solar forecaster to accomplish this task, but many others do this merge of data inside their own company. This latter method allows site-specific information that is not necessarily the information detailed in **Table 3.1** but captures the effects of that table.

[25] http://www.ercot.com/content/wcm/lists/114800/ERCOT_Solar_SiteScreenHrlyProfiles_Jan2017.pdf

Hydro, Wind, and Solar Data

Hydroelectric, wind, and solar resources are similar in that their production at a given point in time is governed by fuel availability. Hydroelectric resources have varying levels of control over their availability depending on the site; run-of-river generators are entirely dependent on river inflows, while generators with large reservoirs can have daily, weekly, seasonal or even annual storage. The goal of any data collection for modeling the capability of these resources is to find data that give the best representation of the capability of these resources over a period.

For all three resources, there are two basic types of data that can be collected: production data and fuel availability data. At a high level, production data captures the amount of electricity generated over a given period while fuel availability data captures the amount of primary energy that could have been converted into electricity over a period. For all three resources, the collection of production data is the same assuming that full data availability. For many embedded generators, production data may not be available. Collected data that captures the amount of primary energy that could have been converted into electricity for each resource type is outlined below.

Key Takeaway:
Energy limited resource data gathering should have the same time frame as the demand collection in the resource adequacy study.

When gathering data for these units, ensure that the same historical time frame is used for the demand sampling. If a different historical year is excluded from the sampling for data in the solar resource, the cross correlation coefficients of the hydro, wind, or solar resource with the demand will impact the end probabilistic metrics in the study. Maintaining the same historical period as the demand sampling will alleviate the concern over these cross correlations or any other dependency between the resource availability and demand. A good way to think about this is that in times of high irradiance, many air conditioning loads are likely to be active. If a TP samples irradiance outside of the same time boundaries as the load, the correlations in the shapes need to be described; they may be misrepresented in the study otherwise.

Solar Fuel Availability Data

For installed solar PV plants, the same irradiance data that created a solar profile can act as a fuel availability curve for that resource. There are various methods to collect irradiance data with some sources (detailed above). A cloud cover or satellite analysis might be necessary to determine the availability of the solar resource to contribute in the resource adequacy study. Some models require a temperature and wind speed aspect for solar availability, and any publicly available data source or nearby weather station can have those measurements. In addition to **Table 3.1**, some models require the global horizontal irradiance, diffuse horizontal irradiance, or direct normal irradiance or some combination of the three in order to calculate the output of the solar facility. Regarding those values, some weather stations are not equipped to measure all of the values.

Wind Fuel Availability Data

Key Takeaway:
If historical generation records are unavailable for the resource, geographically close profiles are adequate. This includes weather stations.

Wind fuel availability is similarly calculated as solar fuel availability. However, since wind speed is dependent upon the height of the measurement, the turbine height needs to be accounted for in the gathering of wind speed. The historical wind generation in that area is important in order to get the distribution of wind speeds and thus the generation of that facility. For operational plants, many have wind speed recorders that can be obtained to build the curve. NREL also maintains records for wind speeds between the years of 2012 and 2015; however, recent years are not recorded. The National Oceanic and Atmospheric Administration[26] can provide the wind speed for these and other years to supplement the data from NREL. If the operational plant does not record their data, nearby weather stations are also acceptable sources of data. A power curve translates this wind speed curve into a total MW output of the wind facility in order

[26] See here: https://www.noaa.gov/

to be used in the study. Other weather data may be required based on the sophistication of the wind model in the resource adequacy study.

For future-looking resource adequacy studies, the assumption of geographically close data availability is not always good. One tactic is to collect the capacity of the facility based on the projected design to assist in ascertaining the availability of the wind resource. The key parameters to procure are the design parameters and associate the parameters to an expected wind MW curve. Design factors to consider include turbine height, cut-in speed, cut-out speed, and other speed breakpoints as well as hot or cold temperature limitations and ice-loading capability of the turbine as based upon the design. As an example, WECC samples historical wind generation from their nameplate and uses that profile at a different wind generation facility in order to supply the wind speed curves. Then any design constraints are applied to that profile to gain the total MW production curve from that resource. In general, for studies that are modeling future wind facilities, a profile of wind speeds from other facilities or meteorological stations along with design parameters from the resource developer can produce the expected MW profile of the wind facility. This process is very similar to the simulated solar PV section above.

Hydroelectric Data

Similar to the wind data, representing energy-limited hydro facilities in the study could require a translation of their water supply into a total energy production. To do so, the resource planner will consider hydrologic or fluvial conditions, such as water inflow, outflow, and head of the hydroelectric resource. If using flow data, a power curve is required to translate the water flow into a time-series MW on that resource. For these types of facilities, many regulations dictate the amount of water stored or required to be flowing across the facility, so data on spilled water can supplement production data to give a better indication of the availability of the resource to produce electricity in the study. Additionally, only using production data underestimates the potential of the hydro resource. Market bid or offer data can supplement the production data to get the energy and/or the operating reserve to express the capability of the unit because the total capacity of the unit is the current capacity of the resource is the operating reserve the unit is providing added to any current power production. Since hydro facilities have many moving parts, planned and forced outages are also a concern, albeit a lesser concern. Other outages for hydroelectric facilities can also include environmental or safety outages, which have a similar lesser concern in terms of modeling in the resource adequacy study. See **Chapter 2** on Thermal resources to find databases that these facilities can report to on outages.

The end goal of data gathering for hydroelectric resources is to build a water year for the amount of water available for the plant to use in generation of electricity and to incorporate any environmental factors, operating restrictions, and generation availability that may limit production based on the sophistication of the model. Unlike other energy-limited resources, more attention can be made to the environmental factors that dictate the amount of flow out of the plant that will describe the availability of the resource. Additionally, if the hydro facility is a run-of-river facility, the inflow of the river and environmental constraints will likely dictate the availability of the plant. Some sources for the data are Environment Canada, NOAA, and other national weather databases that measure hydrological quality.

Energy Storage Systems

As of this report, two major types of energy storage exist: battery energy storage systems (BESS) and pumped hydro storage. The inputs in **Table 3.2** are important to model energy storage systems. Not all parameters are exclusive to a pumped energy storage system or a BESS, though many parameters cross over.

Table 3.2: Energy Storage System Profile Data Requirements		
Category	**Data Point**	**What to Gather**
Resource Characteristics	Maximum Generating Capacity	The maximum MW the facility can generate when discharging its energy
	Minimum Generating Capacity	The minimum MW the facility can generate when discharging its energy
	Maximum Charging Capacity	The maximum MW the facility can take on when charging its energy supply
	Minimum Charging Capacity	The minimum MW the facility can take on when charging its energy supply
	Dispatch Order	Position in the economically constrained dispatch[27]
	Storage Cycle Efficiency	Total Roundtrip efficiency on the charge or discharge cycles.
	Maximum Energy	Pumped Storage Reservoir or BESS maximum energy storage[28]
Outage and Maintenance Data	Historical Outage Data	Time series MW production and consumption for many historical years
	Maintenance Periods	Time windows where the resource is under outage for maintenance.
	Availability of the Unit	Failure and repair rates of the unit[29]
Unit Availability during Ancillary Services*	Pumping Operation	Similar to the Outage and Maintenance Data
	Normal Operation	Similar to the Outage and Maintenance Data

*This type of data may be very difficult to obtain for battery energy storage systems as they may have many different ancillary services. An operational profile may be more informative.

> **Key Takeaway:**
> Understanding the energy storage device's operational characteristics allows for adequate modeling and informs the data collection and databases required for the study.

Initial additions of energy storage systems to systems that are capacity constrained rather than energy constrained are generally capable of providing full capacity value with 4–6 hours of continuous operation relative to conventional resources. As an example, an energy storage resource can be charged during low load periods and dispatched during the few highest load hours of the day or by other dispatch patterns depending on how the resource is procured. However, when the penetration increases above 2–3% of system peak, rigorous modeling of all constraints and capabilities of energy storage systems is required. While the dispatch methodology is still the same, the frequency and duration of high load becomes more binding on the capacity value that energy storage resources can provide since they are required to serve more of the load.

[27] This is important for EOPs or other ancillary service capacities these storage systems supply. Market data may be required

[28] In pumped hydro cases, this maximum may be quite large.

[29] In BESS systems, this is highly crucial due to the construction of the battery pack. Other energy-limited resources have resilient measures in place; however, BESS construction has either an "all or none" capacity.

It is also important to note that there are numerous possible interactions of the various energy-storage-specific inputs. For example, if the dispatch order of energy storage systems is not optimized for reliability, they may need a significantly longer duration capability to provide full capacity value. In addition, if energy storage resources can be used to serve ancillary services, their reliability value can be substantial with even shorter duration capability.

Chapter 4: Emergency Operating Procedures

EOPs are control actions or tools that system operators can utilize to modify generation or loads under stressed, abnormal, or emergency system conditions. These conditions could be resource supply or reserve deficiencies or element contingencies under the course of BPS operations. EOPs should be properly accounted for and modeled into probabilistic reliability assessments to ensure that a realistic representation of system risk concerning resource adequacy are considered. These tools can be invoked or implemented to mitigate possible resource shortages or emergencies prior to the disconnection of load and the likelihoods of use and amount of relief can vary. The procedures and details of EOPs is widely dependent on a regional, area, or entity basis and typically occurs under pre-established criteria.

Parameters

Modeling EOP types of remedial actions can vary greatly by entity and data sources can vary accordingly. EOPs will generally provide a means to relieve a constraint for a specific amount of time. Some types of EOP's that could be considered for studies can include the following:

- Interruption of Transmission Service (Transmission Loading Relief)
- Load Curtailments or Interruptible Load Programs
- Operating Reserves
- Use of import agreements with neighboring systems
- Voltage Reduction
- Special Resources
- Demand Response
- Public Appeals
- Cyclic load shedding

These types of EOPs can have specific parameters that must be considered in modeling. These could include the number of times in a given time period the EOP/resource can be performed, the duration and time period between calls, and the amount of relief on subsequent calls or fatigue factors. These constraints can be seasonally adjusted as well depending on the area as seasonal temperatures may prevent an EOP from being enacted on the demand side from an undisturbed system. With regard to these procedures, state governments or programs may have the details on the limitations and can help to associate the exact parameters required to model that specific type of EOP.

Collection Methods

Due to the rigidity for some EOPs, the duration and frequency are generally fixed, indicating a lack of major data collection efforts being needed for a probabilistic study. In terms of data collection, some programs may require a customer to sign up with the utility for the program. As such, the repository that holds those records will be the source of data for the probabilistic study to determine how much load is relieved when the EOP is enacted for those programs. Relevant load relief data (in MW) for EOPs can be determined through several methods, depending on the system; however, the majority are based on collection via source documentation or by historical availability.

Key Takeaway:
Emergency operations procedures require less data gathering to model than the other topics discussed due to their fixed duration and frequency of calls.

The source documentation methods look at the establishing papers, legislation, or programs that dictate how EOPs will be called upon and use such information as data for study. For instance, some EOPs, such as voltage reduction, can be determined through the source documents of those schemes. Other EOPs' load relief data can be collected through the registration of resources and the availability requirements for these resources in an emergency. Even further, some EOPs are spelled out in the tariffs and serve as good data sources for determining the amount of available capacity for load relief. Limitations on the number of calls for these EOPs need be considered when collecting the data as well the assumptions surrounding the source documents to see if both still hold for the study in question. This type of data may not be found in the source documents and should be considered when collecting data for study.

Regarding historical availability methods, the resource planner can also actively collect data regarding how much relief occurred from historical calls to EOPs. Trends could also be reviewed from GADS or other measured data to develop reasonable assumptions for usages for a given EOP if the other methods cannot provide the data. Availability of these resources at the time of the emergency, such as the proportionality to peak loads, should be considered when developing assumptions utilizing the availability databases.

Physical Testing or Audits for Voltage Reduction

If physical test are available to the planner, the resource planner can commission a voltage reduction test and utilize those results to determine the amount of relief that the EOP can provide in the probabilistic study. These tests may require other jurisdictional approval prior to conducting the test. Other types of tests may also exist to provide the estimated capacity relief from other EOPs and entities can look to either producing their own test or coordinating with other entities to produce a test.

Chapter 5: Transmission Representation

Transmission constraints have received increased focus in probabilistic resource adequacy assessments. There are many different parameters associated with transmission lines and not all may be useful to determine the interconnected system's reliability in a probabilistic representation, depending on the study. A majority of the data sources discussed in the other chapters are representative of the desire to determine if sufficient generation is available to meet demand. Similarly, there may be a desire to determine if sufficient transmission is available to deliver that generation to meet demand.

Key Takeaway:
Data requirements depend on the types of transmission model used in the resource adequacy study. Some require additional line parameters, but others require only transfer limits

Interface Limit and Detailed Circuit Representation–Data Requirements

Typically, there are two different ways to represent transmission constraints: interface limit models and detailed circuit representations. In the interface limit model, the transmission is modeled as a "pipe" between two areas with specific constraints and properties. In the detailed circuit representation, the transmission is modeled using all transmission lines that may be seen in positive sequence load flow software into the reliability assessment realm. These types of representations can be useful depending on the type of study being done; however, their data sources may not always be the same.

Interface Limit Model

The transmission constraints between areas are modeled with interface transfer limits. Each interface is represented as a tie line with bidirectional transfer limits. Physically, each interface may consist of two or more transmission lines and the interface limits and equivalent admittances are typically determined based on thorough steady state and/or transient stability analyses. Most of the existing tools for resource adequacy assessment are able to simulate random forced outages on the interface between areas. The minimum data required for representing the interface limits depend on the purpose of assessment and the method employed for network flow analysis. **Table 5.1** shows the minimum data requirements for using the Interface Limit Model to incorporate transmission constraints in resource adequacy assessment. NERC TADS is a database that records the type of outages associated with transmission lines and provides enough information to formulate a forced outage rate for the transmission elements. Aggregation techniques will be required to associate the specific line data with how the transmission is modeled as the records in TADS may be more specific than the tie line representation. In order to find the bidirectional transfer limits, generally an available transfer capacity (ATC) study can inform on the limiting conditions and the results of that study will provide a "source to sink" capacity between areas, which is very conducive to modeling these interfaces. If adding in the dc powerflow capabilities of load flow software, the equivalent reactance between the source and sink in that ATC study will need to be determined and provided. This may not always be provided in a single ATC study, so model reduction of the powerflow data collected for Interconnection-wide base cases created under NERC MOD-032[30] can aid in finding the equivalent reactance of the interface.

Table 5.1: Minimum Data Requirements (Interface Limit)

Network Flow Method	Import/Export Limit	Equivalent Reactance	FOR
Transportation Model	Yes	No	Maybe
DC Power Flow	Yes	Yes	Maybe

[30] NERC MOD-032 can be found here

Detailed Circuit Representation

Normally, detailed transmission models are not required in resource adequacy assessment. If detailed circuits are modeled with generation facilities, the evaluation is often referred to as composite system reliability assessment and a vast number of input data are needed for such assessment. Composite system reliability assessment mainly involves the selection of possible system states for evaluation and the assessment of the consequences of these states. Two basic methodologies are used in the system-state selection in composite system reliability assessments. These are the analytical contingency enumeration approach and the Monte Carlo simulation method. The system analysis of the consequences of selected outage states is the same for both analytical and Monte Carlo simulation methods; ac or dc power flow is employed to determine if a particular state is a success or a failure in composite system reliability evaluations.

The detailed power flow data for composite system reliability assessments typically contains information on the system topology, equipment ratings, and various potential operating conditions, for example, summer/winter, peak/light load, drought/wet, or export/import scenarios. These power flow data are maintained and updated by industry regularly. Outage statistics data, such as the failure rate and average outage duration for all of the composite system facilities, are recorded and available from NERC GADS and TADS systems for generation facilities and transmission facilities. Some system specific data (e.g., remedial action schemes, fast runback of HVDC, normal operating procedures, tapped transmission lines, common mode outage information) may be needed. The general procedure and the minimum data requirements for composite generation and transmission reliability assessments are available in existing literature.[31]

31

Billinton, R., 1969. Composite system reliability evaluation. IEEE Transactions on Power Apparatus and Systems, (4), pp.276-281.

Billinton, R. and Wenyuan, L., 1991, July. Composite system reliability assessment using a Monte Carlo approach. In 1991 Third International Conference on Probabilistic Methods Applied to Electric Power Systems (pp. 53-57). IET.

Ubeda, J.R. and Allan, R.N., 1992, March. Sequential simulation applied to composite system reliability evaluation. In IEE Proceedings C (Generation, Transmission and Distribution) (Vol. 139, No. 2, pp. 81-86). IET Digital Library.

Chapter 6: Concluding Remarks

Based on the current methods for setting up a probabilistic resource adequacy assessment, the PAWG identified a few commonalities that are of particular importance. While different studies may require additional data to address the impacts of a particular risk (e.g., cyber-related attacks), this document provides different collection experiences and highlights the key points of resource adequacy studies. In particular, the PAWG identified a few common practices that should be emphasized. In general, the probabilistic studies require large quantities of data to add more complexity to the models in their assessments.[32]

The Need for Data in Probabilistic Studies

In general, a resource planner's job is to predict and determine the level of risk for future years. Planners require a set of predictive models that they develop and maintain. In order to develop and maintain their models, planners require access to a variety of different types of data that may not be available. This particular point is crucial as sometimes engineering judgement is able to fill where data is not available; however, judgement is not a substitute for high quality data sources that are representative of the equipment being modeled. This need for high quality data applies to all the different categories of data in the previous chapters and is not relegated to demand, generation, transmission, etc. Additionally, the study objective may change the modeled parameters based on the engineering judgement of the resource planner. In any two given studies, certain resources or aspects of a resource may not be a necessary modeling requirement due to the study objective. The resource planner needs to determine the model complexity required for the loss-of-load probabilistic study and use the data sources appropriately to complete the model.

Common Key Points

The PAWG identified the following key points in data collection across many different portions of a probabilistic resource adequacy study:

- Collection of weather data and any portion of the resource adequacy study related to weather should have the samples taken in the same period. If samples are not able to coincide, a cross-correlation calculation can help reorient when the weather data sample was taken and when the demand sample was taken.
- When utilizing GADS or other historical outage reporting data, the thermal resources future outage rate may not be indicative of this historic metric, especially when the facility moves to different operational characteristics.
- BESSs can be modeled similarly to other energy-limited resources, such as pumped hydro, when performing a resource adequacy assessment with an emphasis on understanding the operational characteristics of the BESS.
- Data collection for transmission systems in probabilistic resource adequacy assessments depends on how detailed of a transmission model is represented in the study. This is over and above the normal dependency that other portions of a probabilistic resource adequacy study requires.
- PCs, TPs, and other modelers require access to detailed information in order to build and maintain their models for use in probabilistic studies.

Possible Future Work

As probabilistic resource adequacy studies develop and mature, the PAWG recommends that the ERO review this data collection document. By doing so, this document can be utilized along with other probabilistic resource adequacy

[32] This assumes that no assumptions will be made regarding the effect these new facets of the model have on the availability or performance of the element in the resource adequacy study.

documents to assist entities with developing new probabilistic requirements or improving previous ones. Additionally, the PAWG issued the following recommendations:

- When utilizing GADS or other historical outage reporting databases, the thermal resources future outage rate may not be adequately represented by use of this historic data, especially when the facility moves to different operational characteristics. A thorough review should be done before using historic outage data when representing future risk.
- PCs, TPs, and other entities should work to gain access to data not otherwise made available that may affect the results of their resource adequacy studies or assumptions. Some entities do not have access to data sets to feed their models, and the need for more accurate studies may require access to data outside of those publicly available. This is paramount as resource planners are not able to perform studies without well-developed models that require a wide range of data.
- Careful understanding of data source assumptions and restrictions should be used when vetting a new or previous data source.

Appendix A: Overview of General Data Management

In general, data used for study should be complete, of high quality, and representative of the equipment under study. As with many other modeling issues, there are times when the data is not always complete, does not follow the guidelines for data submission in the database, or is not accessible without supplemental agreements. This appendix covers some of the general considerations for vetting the data for use in the probabilistic study.

Keeping Data Aligned

When the resource planner is merging many different sources of data or when dealing with large data sets, a few common procedures should be followed. Considering much of the data in probabilistic studies is based on a time series or has a time dependence (e.g., weather years), many of the processes deal with this type of alignment. Some general data alignment techniques for entities to consider are the following:

- Convert to a common time zone, including considerations for daylight savings time changes (if applicable)
- Utilize hourly trends to fill gaps in data, such as zeros and/or blank hourly values due to time zone conversions (These gaps should not be large in size, nor should they be frequent in the data source.[33])
- Detect unit outliers in minimum and maximum daily, monthly, and annual peaks for possible data errors
- Determine the per-unit relationships between hourly values and the daily peaks throughout the years in order to detect anomalies
- Conduct benchmarking to similar data sets, such as, but not limited to, entity reported actual summer and winter peak demands for use in Regional Reliability Assessments[34]

Common Sense Validation Checks

Additionally, there are a few other common sense checks when preparing the data for use in a probabilistic study. This list is provided as an example, and other checks or metrics may exist for determining how trustworthy the data source is for providing information in a resource adequacy study. Examples of such checks are found in **Figure A.1**.

[33] For example, some data sets are not usable with more than 5% total data missing or when the largest gap of data is longer than 12 hours. These values will change depending on the data. In general, a resource adequacy study can fill these gaps; however, these two metrics should be considered when vetting a data source.

[34] A common NERC approach for determining LFU uses the variance in year-over-year deltas of actual peak demand. For this reason, a good check is to compare these deltas from FERC 714 for particular entity or area with that of another data set.

Peak Data

- Correctly filled out field
- Correct mangitude
- Correct units

Energy Data

- Correctly filled out field
- Correct mangitude
- Correct units

Load Growth

- Reason provided for growth exceeding +/- 1%

Resources

- Resource in the right zone, Balancing Authority, and organization
- Unit type matches with fuel (e.g. PV unit with secondary battery incorrectly labeled as a natural gas turbine)
- No negative or zero capacities for nameplate or seasonal capacities
- Status Code, NERC Class Code, and WECC Class Code are consistent
- No duplicate resources

Reserves

- No blank or null values provided
- Data is consistent with other portions

Hourly Data (Demand, Hydro, Wind, Solar)

- Plotted data do not contain gaps
- Plotted data are consistent with request
- Plotted data demonstrate no abnormalities (e.g. jump discontinuities of large magnitude)

Transmission

- Line type should be overhead, underground, or submerged.
- Reactances are consistent with design (e.g. X/R ratio)
- Line information consistent in lines with many taps

Trasnformer

- Line type should be overhead, underground, or submerged.
- Reactances are consistent with design (e.g. X/R ratio)
- Line information consistent in lines with many taps

Figure A.1: Common Sense Checks for Data Validation

Data Retention for Future Studies

Due to the large set of data required to gather for modeling resources in a resource adequacy study, it is preferable to store the data for use in future studies. For instance, the transmission system representation, once built, does not need to request the same level of information each time the model is updated; a notification of which elements, interfaces, or other equipment that have changes suffices. Additionally, outage data does not need to always be collected for the same period. The collection effort should be focused on the data that would supplement what has historically been collected. Because of the need to store the data for long periods, a data maintainer should be used to ensure that the data are not lost, corrupted, or otherwise changed between studies. Additionally, some data can be used for different studies that further increase the value of retaining large sets of data for probabilistic reliability studies.

Appendix B: Example GADS Data Request Example Forms

This appendix serves as an example of data forms for requesting GADS data from other entities. ERCOT has graciously provided the following two forms to provide clarity on some of the information in the chapters.

GADS Data Request Notice

The following information is contained in ERCOT's GADS Data Request Notice and an example data, the form they send to other entities to request data that accompanies the notice. All content provided is to be used as an example for these requests and should be used only where appropriate.

NOTICE DATE: January 31, 2020

NOTICE TYPE: W-X013118-01 Operations

SHORT DESCRIPTION: Requested data for the Planning Reserve Study

INTENDED AUDIENCE: Resource entities

DAY AFFECTED: April 1, 2020

LONG DESCRIPTION: ERCOT is conducting a capacity planning reserve study in 2020 that is mandated by the Public Utility Commission of Texas, as well as a loss-of-load study for the North American Electric Reliability Corporation (NERC). In order to accurately model historical thermal unit availability for both studies, ERCOT is requesting that resource entities extract certain unit-specific outage data from the NERC GADS for each of their thermal Generation Resources, and provide that data as instructed in the attached data submission form. ERCOT is requesting up to two Calendar Years (2018-2019) of GADS outage event and Equivalent Forced Outage Rate (EFOR) data for units that meet the following two criteria:

A. GADS data was submitted to NERC for Calendar Year 2018. (Wind unit outage data uploaded to the NERC GADS Wind system is not to be included in the submission.)

B. The thermal unit(s) are currently expected to be in operation, or could potentially be in operation, as of January 1, 2021.

The GADS data submissions are considered Protected Information under Nodal Protocols Section 1.3.1.1(q).

ACTION REQUIRED: Please return the attached data submission form and any accompanying data files, by April 1, 2020, via email to ClientServices@ercot.com.

CONTACT: If you have any questions, please contact your ERCOT Account Manager. You may also call the general ERCOT Client Services phone number at (512) 248-3900 or contact ERCOT Client Services via email at ClientServices@ercot.com.

If you are receiving email from a public ERCOT distribution list that you no longer wish to receive, please follow this link in order to unsubscribe from this list: http://lists.ercot.com.

GADS Data Submission Form

REPORTING INSTRUCTIONS:

1. An example GADS Data submission form ERCOT is required for all units that meet reference. Please use this as an example when improving or building similar GADS data requests. An important piece of the following two criteria:
    a. GADS "Conventional" data was form is the capability to categorize the submitted for Calendar Year 2018; wind data to each utility, unit, and solar units reported do not need to be included event in your data submission.
    b. The unit(s) are currently expected to be in operation as of January 1, 2021.
2. Data submittals are due no later than April 1, 2020.
3. In the shaded cells below, enter the contact order to feed the information for the preparer of into the data submission in case ERCOT staff has questions on the submitted GADS data probabilistic model.
4. The second and third tabs, named GADS_Unit Outage Details and GADS_EFOR, respectively, specify the GADS data elements to be reported for each thermal unit.
5. Provide the requested GADS data for Calendar Years 2018 and 2019, or for the subset of these years for which GADS data is available.
6. Resource Entities may submit the GADS data in separate files (one file for each tab) as long as the field names and ordering matches the two tabs. Although Excel files are preferred, text files (such as CSV) are acceptable.
7. This file, and any separate data files, should be sent in an email as attachments. The email address for the data submission is ClientServices@ercot.com.
8. This data submission is considered Protected Information under Nodal Protocols Section 1.3.1.1(q).
9. If the data file(s) is too large to be sent using email, a secure FTP file transfer will be arranged. Please send an email to ClientServices@ercot.com requesting a file transfer link.
10. Questions on the data form or submission process should be sent to ClientServices@ercot.com or your ERCOT Account Manager.

REPORTING INSTRUCTIONS:

1. Data submission is required for all units that meet the following two criteria:
   a. GADS "Conventional" data was submitted for Calendar Year 2018; wind and solar units reported do not need to be included in your data submission.
   b. The unit(s) are currently expected to be in operation as of January 1, 2021.
2. Data submittals are due no later than April 1, 2020.
3. In the shaded cells below, enter the contact information for the preparer of the data submission in case ERCOT staff has questions on the submitted GADS data.
4. The second and third tabs, named GADS Unit Outage Details and GADS_EFOR, respectively, specify the GADS data elements to be reported for each thermal unit.
5. Provide the requested GADS data for Calendar Years 2018 and 2019, or for the subset of these years for which GADS data is available.
6. Resource Entities may submit the GADS data in separate files (one file for each tab) as long as the field names and ordering matches the two tabs. Although Excel files are preferred, text files (such as CSV) are acceptable.
7. This file, and any separate data files, should be sent in an email as attachments. The email address for the data submission is ClientServices@ercot.com.
8. This data submission is considered Protected Information under Nodal Protocols Section 1.3.1.1(q).**Respondent Contact Information:**
9. If the data file(s) is too large to be sent using email, a secure FTP file transfer will be arranged. Please send an email to ClientServices@ercot.com requesting a file transfer link. Contact Person:
10. Questions on the data form or submission process should be sent to ClientServices@ercot.com or your ERCOT Account Manager.

GADS Data Submission Table Template

The below text is a template supplied by ERCOT to fulfil the requests. This is provided for example purposes only.

Title:
Telephone Number:
Resource Entity Name:
Email address:

Utility Code	Unit Code	Unit Name	Year	Event Type	Start of Event	End of Event	Net Available Capacity	Cause Code	Event Description

Utility Code	Unit Code	Unit Name	Year	Annual-EFOR	EFOR-Jan	EFOR-Feb	EFOR-Mar	EFOR-Apr	EFOR-May	EFOR-Jun	EFOR-Jul	EFOR-Aug	EFOR-Sep	EFOR-Oct	EFOR-Nov	EFOR-Dec