Probabilistic Reliability Modeling Inputs and Assumptions

RESOURCE ADEQUACY PROCEEDING R.14-10-010
CALIFORNIA PUBLIC UTILITIES COMMISSION – ENERGY DIVISION

July 15, 2015

Contents

I. INTRODUCTION 4

II. SERVM SOFTWARE 5

III. COORDINATION OF KEY DATA SOURCES AND KEY DEFINITIONS 5

A. Key Data Sources 5

B. Key Definitions and Reliability Metrics 7
1. Desired Reliability Level – One Event in Ten Years 7
2. Definition of Outage Event – Reserves or no Reserves 8
3. General Order of Studies in ELCC modeling 8

IV. WEATHER DATA AND REGIONS 9

A. Weather Data and Sources 10

B. Region Designations 10

C. Regional Weather Assumptions 12

V. LOAD INPUTS 12

A. Synthetic Load Shapes 12
1. Data Gathering 12
2. Processing 14
3. Neural Network Modeling 15

B. Economic and Demographic Forecasting Uncertainty 17

VI. RESOURCE INPUTS AND USE LIMITATIONS 18

A. Generic Resource Information 18

B. Disaggregating Aggregate Units Into Child Units 19

C. Thermal Resources – Types T, F, and N 20
1. Heat Rates 20
2. Ramp Rates 21
3. ELCC compared to "Perfect Capacity" 23
4. Natural Gas Price Forecasts 24

5. Variable Operating and Maintenance Cost 24
6. Imports and Direct Sales 25

D. Renewable Resources – Type R 26
1. Wind and Solar Generation Profiles 26

E. Demand Response – Type C 34

F. Energy Storage Resources - Type P 38
1. Current Inputs 38
2. Potential Future Expansions/Changes to Inputs 38

G. Hydropower Resources – Type H 39
1. Regional Aggregation of Energy 40
2. Run of River (ROR) Hydro Resources 40
3. Scheduled Hydro Resources 41
4. Emergency Hydro Resources 41

H. Variability and Uncertainty in Load, Wind ,and Solar Profiles 42
1. Data Gathering 43
2. Data Processing 43
3. Creation of Forecast Uncertainty Distributions 45
4. Use of Distributions in SERVM 46
5. Forecast Error Distributions Entered into SERVM 47

VII. TRANSMISSION INPUTS 50

VIII. SYSTEM INPUTS 52

A. System Periods 52

B. Ancillary Services 52

C. Operating Reserve Demand Curves (Scarcity Pricing) 52

IX. NEXT STEPS 53

I. Introduction

This document describes some of the inputs and assumptions recommended by Energy Division (ED) Staff for use in probabilistic reliability modeling, as part of the Resource Adequacy (RA) rulemaking (R.14-10-010). In compliance with Senate Bill (SB) X1 2 and in accordance with the RA proceeding Scoping Memo, Staff is developing a probabilistic reliability model in order to calculate the Effective Load Carrying Capability (ELCC) and Qualifying Capacity (QC) of wind and solar resources.[1]

A resource's QC is the number of megawatts (MW) eligible to be counted towards meeting a load serving entity's (LSE's) System and Local RA requirements, subject to deliverability constraints. ELCC is a percentage that expresses how well a resource is able to meet reliability conditions and reduce expected reliability problems or outage events (considering availability and use limitations). ELCC can be thought of as a ratio of the MW quantity of resource to the MW quantity of "perfect capacity" that provides equivalent reliability benefit. Reliability benefit is measured by first calculating the LOLE of the entire system with the resource, then calculating the LOLE of the entire system with the "Perfect Capacity" instead. The amount in MW of "Perfect Capacity" added to make up for the loss of the resource will determine the ELCC of the resource.

This document is meant to highlight the inputs and assumptions that go into the overall study, but also summarizes the order of studies performed to establish LOLE and ELCC.

Results are included in a separate document.

This paper includes the following key components:

- Review of SERVM (Strategic Energy Risk Valuation Model), software which is being used by ED Staff in the LOLE and ELCC analysis
- Key data sources and key definitions for modeling
- Gathering and use of weather data for development of load shapes using neural net modeling
- Data related to conventional (fossil fuel) generators
- Natural gas price forecasts
- Sources of and use of weather data and weather region definitions to create hourly profiles for wind production, and use of PVWatts data for solar generators
- Development of data inputs and hourly profiles for hydro generators
- Data for demand response and storage resources
- Development of uncertainty and forecast error parameters for flexibility modeling

[1] SBX 1 2 details can be found at http://www.leginfo.ca.gov/pub/11-12/bill/sen/sb_0001-0050/sbx1_2_bill_20110412_chaptered.pdf.

II. SERVM software

ED reliability modeling is being conducted with a modified version of the Strategic Energy Risk Valuation Model (SERVM) software developed by Astrape Consulting.[2] SERVM calculates numerous reliability and cost metrics for a given study year in light of expected weather, overall economic growth, and unit performance. For each of these factors, variability and forecasting uncertainties are also taken into account.

As with all probabilistic models, SERVM attempts to simulate the study year many thousands of times over, with each simulation reflecting a slightly different set of weather, economic, and unit performance conditions. Iteration conditions are selected probabilistically, based on how likely they are to occur. In SERVM, a given future study year is modeled based on historical weather; both load and generation profiles are simulated based on that historical weather. For each of approximately thirty possible weather years, six to eight points of load forecast error can be simulated, creating roughly 200 to 260 scenarios. Each of these scenarios is in turn run with a hundred or more unit outage draws, creating thousands of iterations for the simulation.

The results provide a comprehensive distribution of reliability costs, expected unserved energy, and other reliability metrics. Expected values and confidence intervals can then be calculated based on these distributions. ED staff plans to use these metrics in determining the effective load carrying capability (ELCC) of wind and solar resources, which indicates the contribution of these resources towards system reliability. This document addresses selected inputs and assumptions used in SERVM. Description of the ELCC calculation will be provided in a future paper.

While this document focuses on the ELCC work that is being done for the 2016 RA compliance year as part of the RA proceeding, SERVM can be useful in analyzing a number of issues facing the Commission. To that end, RA staff is coordinating with the 2014 Long Term Procurement Planning (LTPP) staff among others to ensure that any future, longer-term modeling will be consistent with the 2014 LTPP assumptions.[3]

III. Coordination of Key Data Sources and Key Definitions

A. Key Data Sources

In developing probabilistic modeling for use in the RA proceeding, ED staff is coordinating with other state agencies and organizations across the Western United States to ensure consistency of modeling and data assumptions. Specifically, ED staff is working to harmonize as much as is reasonable with key

---

[2] More information from the developer can be found at http://www.astrape.com/index.php?file=products.

[3] The most recent proposed LTPP Planning Assumptions and Scenarios can be found at http://www.cpuc.ca.gov/PUC/energy/Procurement/LTPP/ltpp_history.htm.

data sources from the California Independent System Operator (CAISO) and the Western Electric Coordinating Council (WECC).

ED staff proposes to source extensive information from the CAISO MasterFile. In order to participate in the CAISO energy market, generators must submit a wide array of information into the MasterFile database. The MasterFile is used by the CAISO in order to optimize dispatch in light of an array of unit-specific characteristics such as start-up costs and start-up time, ramp rate, heat rate, and forbidden operating ranges. Generators participating in CAISO markets maintain their information in the MasterFile in order to ensure cost effective dispatch of their plants. A number of the data fields in the MasterFile are confidential, and are accessible to ED staff via a subpoena executed annually. However, definitions of all the fields in the MasterFile are public and are posted on the CAISO website.[4]

In addition to the CAISO, the WECC also compiles a base case dataset for the WECC and its members to use as a common basis for their modeling. Each Balancing Authority may have unique access to accurate and confidential data for generators and other market participants within its footprint, but since the WECC is so interconnected, it is difficult to accurately model reliability and economic conditions in one Balancing Authority without attention to generators and loads in the surrounding Balancing Authorities. To facilitate consistent modeling by all Balancing Authorities in WECC, every two years WECC produces a Common Case dataset containing generic information for all load and supply data across WECC. Produced by a subcommittee of WECC members called the Transmission Expansion Planning Policy Committee (TEPPC), this dataset is generated for both the immediate next year and for a year ten years into the future. TEPPC has produced datasets for 2014 and 2024. The Common Case dataset is publicly available, and can be downloaded from the WECC website.[5]

ED staff proposes to source most modeling inputs related to loads and generators from the two main sources discussed above: the CAISO MasterFile and the WECC TEPPC Common Case dataset. Each dataset has advantages and disadvantages. For generators that supply information to the CAISO MasterFile, there is a larger range of information available to ED for modeling purposes. For example, unit-specific outage information and heat rate curves can be derived from the CAISO MasterFile. However, because this dataset is confidential, it brings with it the challenge of how to make as much of the input data as possible accessible to stakeholders.

The WECC TEPPC Common Case dataset, on the other hand, uses public data. However, because those data are public, they must be generic or aggregated, and thus are not unit-specific or sufficiently

[4] MasterFile field definitions can be downloaded from http://www.caiso.com/Documents/GRDTandIRDTDefinitions.xls. CAISO MasterFile data are confidential, and not able to be posted; however, it may be possible to aggregate portions of these data for stakeholder review.

[5] WECC TEPPC 2024 Common Case datasets are available for download here: https://www.wecc.biz/Reliability/Forms/AllItems.aspx WECC staff is currently updating and finalizing the latest version (v.1.5) of the 2024 Common Case. ED Staff performed analysis with version V.1.3 as modeling was going on, but has not yet updated to v.1.5. Staff intends to update to V.1.5 before further modeling is undertaken.

differentiated. For this reason, it is common for particular jurisdictions or balancing authorities within the WECC to substitute their own confidential, in-house data for the TEPPC Common Case inputs related to their own specific balancing authority. ED staff is evaluating the advantages and disadvantages of this approach. For near term modeling (such as for determining wind and solar ELCC values for the 2017 RA Compliance Year), ED staff proposes to use the TEPPC 2024 Common Case for regions external to CAISO only. For CAISO regions, ED staff proposes to use generator-specific information gained via subpoena from the CAISO MasterFile.

In addition to the datasets mentioned above, other information will be sourced from internal ED data, the Integrated Energy Policy Report (IEPR) produced by the California Energy Commission (CEC), the National Oceanic and Atmospheric Administration (NOAA), the National Renewable Energy Laboratory (NREL), and data specifically gathered from the utilities. These data and their use in SERVM will be described in further detail in the sections that follow.

B. Key Definitions and Reliability Metrics

Staff made assumptions about a number of key definitions to enable the ELCC model to progress. Staff made an assumption about what reliability standard at which to calibrate the CAISO system, the definition of an outage event, and performed a convergence analysis to evaluate the optimal number of iterations to run for each case.

In support of the ELCC analysis, staff performed three studies. First, staff studied the overall system, and calibrated the CAISO reliability area (an aggregate of SCE, SDGE, PGE_Bay and PGE_Valley) to a LOLE level of 0.1. Second, staff performed an ELCC analysis to gauge the ELCC of solar facilities in general in the CAISO area. Third, staff performed an ELCC analysis of wind generators generally in the CAISO region. Each study consisted of 165 cases (33 weather years times five load forecast error levels) with 120 iterations each case. The order of studies reflected the task to be completed. Other projects require greater or fewer studies, with perhaps greater or fewer iterations to achieve desired convergence.

1. Desired Reliability Level – One Event in Ten Years

Traditional LOLE studies focus on one peak hour of the day. Other hours are not modeled. Before the development of today's computers, LOLE studies focused on 1 peak hour of the peak weekdays, and could not model the rest. Now, computers are able to model each hour of each day, meaning instead of 261 individual hours per year, current models are able to simulate 8760 individual hours per year. Thus traditional metrics need to be rethought and potentially redefined in order to be applicable to current computing ability.

Traditionally LOLE studies have sought to calibrate the electric system to ensure meeting the one event in ten years LOLE, or a LOLE of 0.1 outage events per year. When that was modeled across just the peak hour of the day on weekdays, that meant as long as there were no measured outage events over the 261 hour data points, then there was no loss of load in the entire year. Circumstances during offpeak, or reliability stress occurring as conditions change during a day, are ignored. Applying the one in ten standard to models that include all hours of the year mean the "one event in ten" standard will be more conservative and more unclear. Would the application of the "one event in ten" standard mean that any hour of any day when there was an outage, regardless of the length of the outage or would the metric be translated into 24 hours in ten years, or LOLE of 2.4 hours per year? This is a complicated question with ramifications for the meaning of LOLE. It is not always obvious that different LOLE studies are comparable.

ED staff conducted ELCC modeling by first calibrating the overall CAISO system (not each individual service area within CAISO nor all of California including non-CAISO areas) to the LOLE of 0.1, specifying that any one outage event during the day would count against the LOLE target equally regardless of how long the outage lasted. ED staff did track and monitor other types of metrics, namely expected unserved energy (EUE) and Loss of Load Hours (LOLH) but the ELCC results ED staff calculated relied on first calibrating the CAISO reserve sharing group as a whole to the 0.1 LOLE standard.

2. Definition of Outage Event – Reserves or no Reserves

LOLE studies measure events where there is curtailment of firm load due to lack of generation resources. The point at which firm load is curtailed in a complicated electric system such as the CAISO is rarely when all reserves are expended; generally some amount of operating reserves are required at all times. Firm load can be curtailed even if there are some (depleted) reserves remaining, since the reserves are there to prevent even larger losses of firm load. Thus ED staff believes it is inappropriate to measure loss of load only when load is higher than resources; rather, it may be more appropriate to measure loss of load at the point where load plus a certain portion of reserves is greater than available resources.

In conducting this ELCC study, ED staff measured loss of load at the point where minimum operating reserves plus load are higher than the available resources. In other words, ED staff assumed firm load is curtailed at 103% of load, or at the point where it became impossible to maintain firm load and 3% operating reserves.

3. General Order of Studies in ELCC modeling

A sequence of studies is performed to establish the Effective Load Carrying Capability (ELCC) of a particular resource or set of resources within a larger electric system. The calibration and sequence of these studies is critical. The ELCC of resources reflect their ability to decrease expected loss of load when added to a system that is calibrated to a certain reliability level.

ED staff performed numerous studies to calibrate reliability within the overall CAISO system at the one event in ten year (0.1 LOLE) metric. The system is calibrated to 0.1 LOLE by adjusting the amount of capacity included in the CAISO generation fleet. In addition to the recent retirements of the Coolwater and Morro Bay facilities, ED staff further removed facilities that were projected to retire between now and the middle of 2016 or shortly thereafter, including Pittsburg unit 6 and the Cabrillo II peakers in San Diego.

Once the CAISO system has been calibrated to a LOLE of 0.1, then it is possible to begin the ELCC studies. ED staff began by removing all solar facilities in CAISO, including solar thermal, PV, and performed studies to gauge the average ELCC of all solar PV facilities within CAISO, then the same sequence of studies to assess the average ELCC of all wind facilities within CAISO. All solar facilities were added and removed as a group, without distinction to location or technology. All wind facilities were added or removed together as a group as well. This was to determine the portfolio ELCC of wind and solar generators; the portfolio ELCC of each type of generator across CAISO would later be used to allocate a diversity benefit to a specific technology class or a specific locational group of facilities when the ELCC of solar or wind facilities are calculated in more granular way.

Portfolio ELCC will be greater than the ELCC for a specific locational or technology grouping of facilities, since a smaller group will benefit from the contribution of the larger group.

In summary average annual portfolio ELCC for all solar or wind facilities in CAISO can be calculated following these steps:

1. Create the capacity portfolio brings CAISO area as a whole to the LOLE of 0.1 given loads and resources which are expected to exist in the 2016 study year.
2. Perform a study according to standard specifications and save all required output reports.
3. Remove all facilities under study in CAISO, but not those outside of CAISO.
4. Perform a study according to standard specifications and save all required output reports.
5. Make an estimate of the amount of perfect capacity needed to replace the removed facilities and add it in proportion to where the facilities were removed.
6. Perform a study according to standard specifications and save all required output reports. If the LOLE of the new system is not yet equal to 0.1, repeat steps 3 and 4 until LOLE equals 0.1 either by adding a little more (if LOLE is greater than 0.1) or take a little away (if LOLE is less than 0.1).
7. One LOLE equals 0.1, find ELCC by calculating a ratio of nameplate MW removed to "Perfect Capacity" nameplate MW added, and the result is the average ELCC of the CAISO portfolio of all the studied facilities. The resulting annual ELCC value will be a percentage less than 1.
8. Energy Division staff is currently unable to create "monthly" ELCC values, as the methodology is unestablished.

IV. Weather Data and Regions

Weather is an integral input into probabilistic reliability modeling. It is used both in the development of synthetic load shapes, which are highly correlated to temperature and humidity, and in the development of generation profiles for weather-sensitive resources such as wind and solar. In order to balance the need to model the wide range of weather across the state at any given time and the need to keep modeling times feasible, a set of representative weather stations are selected and grouped to create regions that are modeled as homogeneous areas. This section details the weather data utilized, the sources for this data, the regions modeled, and the process by which these regions were created.

A. Weather Data and Sources

Weather data is gathered by a variety of agencies and institutions, but the most commonly used data source is the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center. Consistent with common practice, ED staff utilized the NOAA Integrated Surface Data - Lite (ISD-Lite, DS3505) dataset for the 30 year period from 1981 to 2010.[6] Data are hourly and include:

- Air temperature (degrees Celsius * 10)
- Dew point temperature (degrees Celsius * 10)
- Wind direction (angular degrees)
- Wind speed (meters per second * 10)
- Total cloud cover (coded, see format documentation)
- One-hour accumulated liquid precipitation (millimeters)
- Six-hour accumulated liquid precipitation (millimeters)
- Solar irradiance (direct and diffuse) and angle

Weather station metadata (latitude, longitude, and elevation) are sourced from the U.S. Automated Surface Observing System (ASOS) Station Listing.[7] Solar irradiance data is supplemented by the National Solar Radiation Database (NSRDB), which also covers the period from 1981 to 2010.[8] To account for differences in wind capacity factors at different heights, NREL wind resource data may also be incorporated into the analysis.[9]

B. Region Designations

SERVM models eight distinct regions within California and ten outside of California. These regions are utilized throughout SERVM to associate groups of generation facilities with common weather, load, weather-related generation profiles, transmission constraints, and utility service territories. The regions modeled are listed in Table 1, below. The regions below do not correspond to Local Areas, and are not granular enough for transmission planning. Thus, this study is not currently intended to be a probablistic version of the Local Capacity Technical Study. In the future, more granularities could be achieved by splitting the regions into smaller areas; however, it is unlikely the model will ever be able to conduct a power flow assessment. That is not currently the purpose of ED's efforts at this time.

Table 1. Regions Modeled in SERVM

California Regions	Regions external to California

[6] ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/

[7] ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/isd-lite-technical-document.txt

[8] http://www.ncdc.noaa.gov/cdo-web/datasets

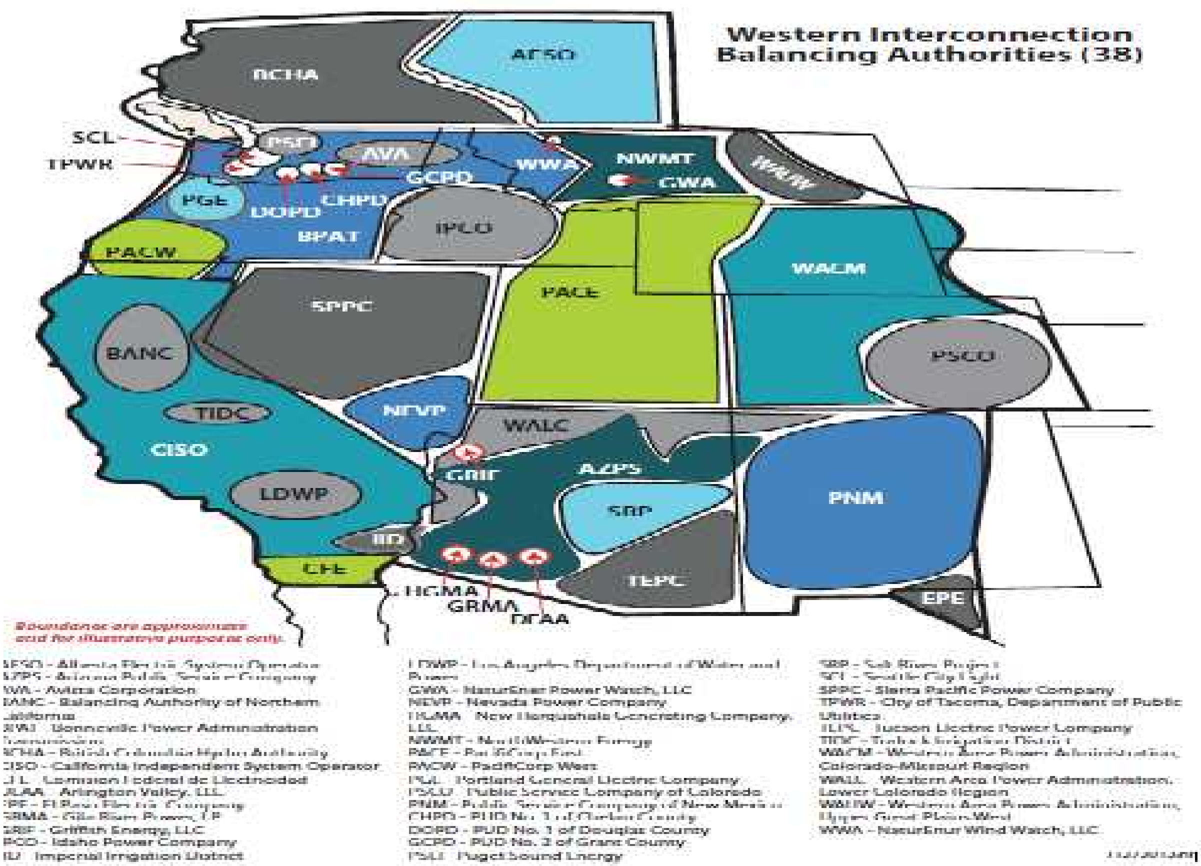
[9] http://www.windpoweringamerica.gov/windmaps/resource_potential.asp (http://wind.nrel.gov/Web_nrel/)

IID (Imperial Irrigation District) Service Territory	Arizona
LADWP Balancing Authority Area (BAA)	Canada
PG&E Bay Area *(Greater Bay Area LCR Area)*	Colorado
PG&E Valley *(Other PG&E Local Capacity Areas)*	Mexico
SCE TAC Area	Montana
SDG&E Service Territory	Nevada
Balancing Authority of Northern California (aka SMUD)	New Mexico
TID (Turlock Irrigation District) BAA	Pacific Northwest
	Utah

ED staff delineated regions in SERVM to correspond to both the TEPPC 2024 Common Case, and the CAISO PLEXOS dataset.

Figure 1 illustrates the boundaries of each balancing authority; some balancing authorities span more then one state. In the future, ED staff proposes to delineate areas by the utility balancing authority they represent, not state of location. This delineation will be both more accurate and correspond to TEPPC Common Case aggregating. For example, Colorado, Wyoming, and Utah will be represented instead by PSCO, Pacificorp East, and WACM respectively. This will entail recreating load shapes and hydro shapes correspondingly.

Figure 1. Balancing authorities in WECC



Source: WECC website, downloaded Nov. 23, 2013 from
http://www.wecc.biz/library/WECC%20Documents/Publications/WECC_BA_Map.pdf

C. Regional Weather Assumptions

For each region, one to three representative weather stations are selected as the basis for weather-dependent modeling. If multiple stations are used to represent a region, their data for a given year is weighted and combined to create a single blended weather dataset for that region and year. However, while data is uniformly sourced from the NOAA and NREL products mentioned previously, the weather stations and weightings selected are not necessarily uniform across the model. Rather, while the weather data stations and weightings utilized for creating synthetic load shapes are selected to reflect areas of greatest load in each region, the data stations and weightings utilized for forecasting wind and solar generation (if necessary) will be selected to reflect the areas of greatest installed capacity for each technology type. This difference will allow SERVM to more closely approximate different types of weather dependencies across various aspects of modeling. The specific weather stations and weightings utilized for the creation of synthetic load shapes and generation profiles will be covered in the relevant section.

V. Load Inputs

A. Synthetic Load Shapes

The objective of synthetic load shapes is to enable a more accurate approximation of whether available resources can meet an unknown future demand. Synthetic load shapes provide variability around a forecasted annual hourly load shape due to weather. When combined with distributions of load growth uncertainty, synthetic load shapes capture the variability that can be expected due to weather, economic, and demographic variation. Given how important an input load shape is in determining the final reliability indices, efforts to model load variability realistically will pay dividends in better results. For purposes of SERVM modeling, synthetic load shapes are constructed from historical hourly loads and temperature history, using a neural network model.

Creation of synthetic load shapes requires data gathering, processing of the data, and assembly of the final load shapes that are to be modeled in SERVM. Finally, some notes on neural network modeling conclude this section.

1. Data Gathering

First, historical hourly loads by utility area are downloaded from the publically posted FERC Form 714. Five years of historical hourly load data from California's utilities and surrounding states are downloaded and organized into columns of a spreadsheet. It is important to ensure that any possible impacts of demand response events are added back to the historical load data to create unmitigated gross load shapes before they are collected into the spreadsheet. That way, when demand response resources are later modeled to meet load, there is no possibility of double counting.

In addition to the load data from the FERC Form 714, historical weather data are required. Thirty years of weather data is downloaded from NOAA for selected weather stations, and imported into a spreadsheet. NOAA data includes hourly temperature, humidity, wind speed, and cloud cover, as discussed in the weather data sources section, but only the hourly temperature field is used for synthetic load shapes. The other data is useful for generation modeling purposes, however.

One to three weather stations are selected as indicative of load patterns in each of the geographic regions modeled in SERVM, as discussed in the weather and region modeling section above. Weather stations and weights in California are those used for weather normalization in the CEC demand forecast process. Summer weights are based on estimated saturation of air conditioners, while winter weights are based on the distribution of population, as shown in Table 2, below.[10] This weights urban areas in California more heavily than rural areas in California. However, because only incomplete data are available for certain weather stations, in some cases weather stations farther from load centers with better data are used in lieu of weather stations located closer to load centers with poor data. The training process verified that correlation factors were adequate, despite sometimes using more distant weather stations. For areas outside of California, simple averages were used.

Table 2. Weather Stations and Weightings for 2015

Region	Station 1	Weight	Station 2	Weight	Station 3	Weight	Station 4	Weight
IID	Imperial	1						
LADWP[11]	Long Beach	.42	Burbank	.58				
PG&E Bay	San Jose	.55	SFO	.45				
PG&E Valley	Sacramento	.35	Fresno	.65				
SCE	Fresno	.09	Long Beach	.49	Burbank	.23	Riverside	.19
SDG&E	San Diego	1						
SMUD	Sacramento	1						
TID	Fresno	1						
Arizona	Tucson	.33	Phoenix	.33	Las Vegas	.33		
Canada	Calgary	.25	Vancouver	.25	Victoria	.25	Edmonton	.25
Colorado	Cheyenne	.33	Denver	.33	Colorado Springs	.33		
Montana	Missoula	1						
Nevada	Reno	.5	Elko	.5				
New Mexico	Albuquerque	.5	Santa Fe	.5				
Pacific Northwest	Portland	.33	Spokane	.33	Seattle	.33		
Utah	Salt Lake City	.5	Boise	.5				

Source: Figures based on NOAA data and input from CEC staff

[10] This is discussed in the "Revised Short-Term (2011-2012) Peak Demand Forecast Committee Final Report". California Energy Commission, Electricity Supply Analysis Division. CE C-200-2011-002-CTF

[11] Weather stations for LAX and Downtown LA were not used because their data quality was poor.

2. Processing

a) Load Data Processing

Once the load data is downloaded and organized, the load data is all normalized to the same year, by choosing one year as the "normal" year, adjusting years before that year upwards to account for overall load growth, and adjusting years after the normal year downwards to strip out the effects of load growth. This is done by adjusting each hour of the yearly history upwards or downwards by a scaling factor that is equal to average peak load growth between the study year and the year being adjusted; this raises or lowers a yearly load profile without altering its shape. The model allows for scaling each area differently or for dynamically scaling loads, but that is not proposed at this time.

In addition to scaling to the proper year, the five years of historical hourly load shapes have been adjusted to "add back" the historical demand response events that occurred during those historical years. Were this not done, then any study would have been an inaccurate portrait of historical load conditions. Actual hourly demand response impacts (taken from utility reports of historical demand response events) are added back into historical load figures to represent historical loads as if the demand response events had not occurred. Thus, when demand response events are modeled for the study year in SERVM, there is no double counting of demand response impacts (triggering modeled events on top of or in addition to historical events).

Once the loads are scaled, it is important to spot-check the resulting loads to ensure that they remain plausible as possible variations of the chosen "normal" year. No particular year's shape is weighted as more likely than any other at this point; that may be done during the SERVM reliability modeling phase, however.

After scaling to the same reference year, day of week differences (due to regular workweek cycles) are removed from the shapes. All load data are scaled such that all days have the same shape as the Wednesday shape from the source data. Weekend days have significantly lower load than Wednesdays, and are scaled up accordingly. Other weekdays tend to have shapes that are slightly lower than Wednesdays, and are also scaled accordingly. This provides a more consistent training set for the neural network model to develop relationships. Based on these data, the neural network model will predict loads for Wednesdays, given temperatures across the entire year. These shapes are then adjusted back to the appropriate day of week at the end of the load development process. Holidays are not accounted for at this time, but could be in the future.

b) Weather Data Processing

The five years of weather data are organized into a spreadsheet and processed for entry into a neural network model. In creating synthetic load shapes, historical temperature at the weather station is chosen as the predictor of load levels. With hourly temperature as one column, another eight columns of data are created and fit to the temperature data. The other columns include hour of day, 8 and 24 and 48 hour previous aggregate temperatures, current cooling degree hours, current heating degree hours, and 5% and 50% exponentially weighted aggregate temperatures. These data are paired with the hourly load in the five scaled load shapes processed previously.

The data are then separated into months. For modeling purposes, each historical month is expanded to include the 15 days before and after the month. Five years of January (December 15 through February 15) are modeled together, five years of February (January 15 through March 15) are modeled together, and so on. This results in approximately 300 days of data for each month (five years of data, each with 60 days included per month), 270 of which will be used to train the neural network model, as described below. Thirty days of data will be excluded so that the results of the model can be checked by "predicting" load for those historical days and seeing how well the predictor model performs relative to historical load data.

c) Scaling load shapes upwards or downwards to model a particular year:

Once load shapes are generated and entered into the model, it is necessary to scale the load shapes to the individual study year being modeled. A table of peak loads per region provide the basis on which SERVM scales the load shapes up or down to represent differences between the peak load in the load shapes (load shapes were all normalized to the same peak load MW value) and the particular study year under consideration. Staff generated a table of region peak load values for each study year and loaded those values into SERVM.

The peak load in the 33 predicted load shapes is compared to the peak load in the table appropriate to the individual study year and the load shapes for each region are scaled so that the peak load in the study year matches the peak load in the individual load shape for that region. Each hour of the load shape is scaled up or down to the same ratio as the peak of the load shape is scaled up or down. Regions can vary individually between study years, and grow or reduce at different rates throughout the study horizon.

For the initial ELCC modeling presented in the RA proceeding, staff studied 2016 RA year. Region forecasts for the eight areas internal to California are all based on the same forecast used for the LTPP proceeding, which is the IEPR 2013 forecast for the mid load – mid AAEE forecast, for the 1 in 2 weather year (Form 1.5b) and linked here.[12]

3. Neural Network Modeling

The training data are then entered into a proprietary neural network model developed by Ward Systems called the NeuroShell predictor model.[13] The neural net model "trains" a network file to identify the

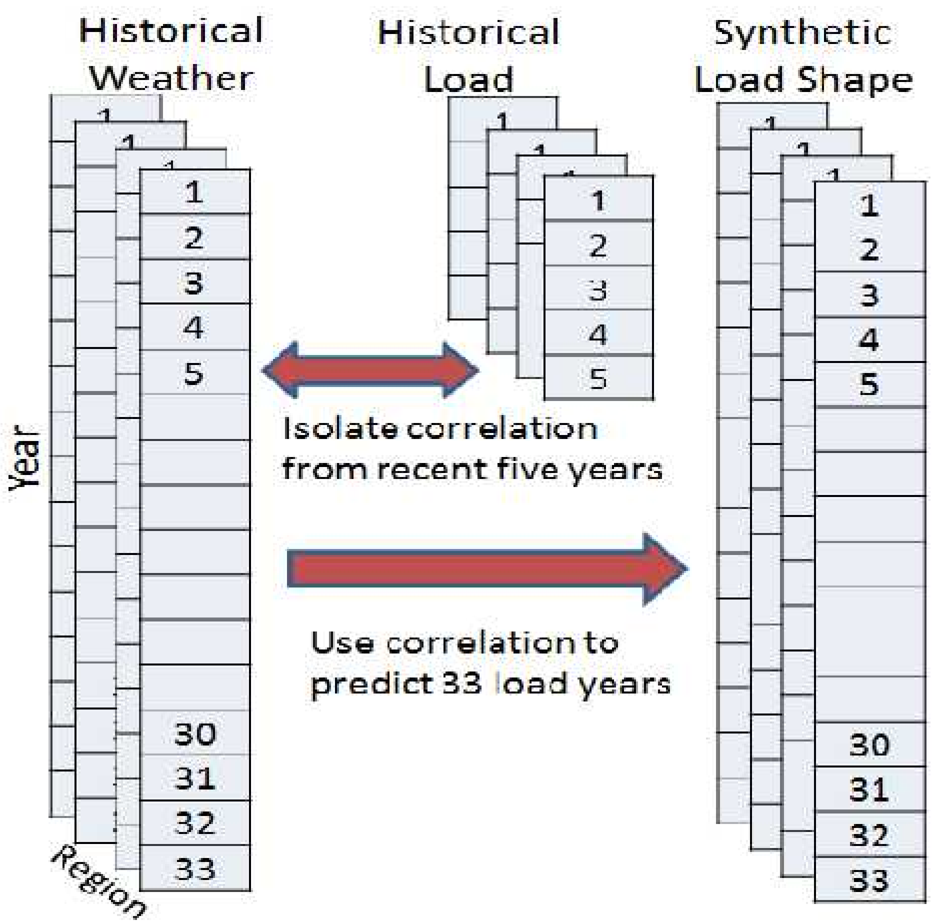
[12] http://www.energy.ca.gov/2013_energypolicy/documents/demand-forecast_CMF/LSE_and_Balancing_Authority_Forecasts/

[13] More information regarding Ward Systems NeuroShell Predictor model is available at http://www.wardsystems.com/predictor.asp.

underlying relationships between the hourly load levels and the other eight columns of data (much like a dynamic iterative regression model) that develops predictive relationships between the nine columns of data (the variables mentioned previously, such as temperature and 24 hour aggregate temperatures), and producing an algorithm that predicts relationships between temperature and load, as well as year. Day of week differences are also reintroduced, as discussed previously. The model is then able to produce a load forecast from a temperature level for any given date. The data excluded from the predictor training are then modeled with the predictor tool to see how well the tool performs relative to actual historical load and temperature.

Once staff is confident of the predictor tool's ability to generate a plausible hourly load shape from the five years of temperature and load files, the thirty three years of historical temperature data are input into the neural network predictor to generate a synthetic load shape for each historical weather year. These load shapes are created for a specific study year to be modeled for reliability and incorporate that year's load growth, days of the week, etc. The thirty three load years are then entered into the SERVM database for reliability modeling. Figure 2 illustrates the process.

Figure 2 Process of Creating Synthetic Load Shapes



a) Notes on Neural Network Models

Neural network models are good at interpolation (finding relationships with historical data sets) but struggle with extrapolation from data that was not in the historical record. For example, if temperature is used to predict load, it is important to use the resulting "predictor" relationships to predict load only for temperatures that are in the historical record. Generation of load shapes for temperatures higher or lower than in those present in the historical record could yield predictions that are close to the output at lowest or highest measured historical temperature values, which is unlikely to match the actual load seen at such extreme temperatures. As a result, SERVM models load at the most extreme temperatures separately, using a more simple regression that focuses narrowly on the impact of each degree change.

Additionally, the predictor performs better when relationships are very consistent. During winter, the relationship between weather and load is less predictable and more volatile. As a result, model performance is not as good in winter as it is in summer, when high heat has a clearer impact on loads.

Nevertheless, the NeuroShell predictor model is statistically significant, with R squared values exceeding 90%. Additional calibration using data from California loads and temperatures could improve that figure.

B. Economic and Demographic Forecasting Uncertainty

Load uncertainty is driven not only by year-to-year volatility in weather patterns, but also by long-term uncertainty in economic and demographic growth forecasts, as well as in energy efficiency impacts and other demand side effects. Unanticipated economic growth or downturns can result in peak loads that are substantially higher or lower than forecast.

SERVM accounts for this potential error by incorporating a "load forecast multiplier" into each model run. A range of load multipliers can be entered into the model, along with the probability of selecting each value. Collectively, they represent the distribution of load forecasting error due to non-weather causes (economics, demographics, etc.). At the beginning of each case, a particular weather year and its corresponding load shapes are selected. A load forecast multiplier is selected independently, and all hourly load values are adjusted upwards or downwards by that same value. For example, if a load forecast multiplier of 0.95 is selected (simulating an unexpected economic downturn), then a region with a peak load of 1000 MW in the given weather year would be adjusted to have a peak load of 950 MW. A new weather year and a new load forecast multiplier would be selected for the next case. Number of weather years multiplied by number of load forecast multipliers equals the number of total cases that are run as part of a study.

The load forecast multipliers used in Energy Division modeling are based on analysis of near term forecasting that was available via the internet from the OECD Journal.[14] Staff evaluated projections of 1 year ahead and 2 year ahead GDP growth, noting the magnitudes of GDP uncertainty and their

[14] Link here: http://www.keepeek.com/Digital-Asset-Management/oecd/economics/an-evaluation-of-the-growth-and-unemployment-forecasts-in-the-ecb-survey-of-professional-forecasters_jbcma-2010-5km33sg210kk#page9

probabilities. These figures were entered as a basis for the load forecast uncertainty variables in SERVM. The values are summarized in Table 3.

Table 3 Economic/Demographic Forecast Error

Magnitude of forecast error (Percentage)	Probability of error occurring (percentage)
2.5% error	6.68% probability
1.5% error	24.17% probability
0% error	38.29% probability
-1.5% error	24.17% probability
-2.5% error	6.68% probability

Source: OECD Journal

VI. Resource Inputs and Use Limitations

A. Generic Resource Information

There are a number of inputs that are common to all supply side resources (including demand response, intermittent renewables, thermal facilities, and storage) in order to identify and characterize their capabilities for the model. For example, the model requires each resource to be identified with a unique ID number, a region in which the resource is located, and the first and last year of expected service. Additionally, there are numerous input fields that are specific to particular unit types. The following table summarizes the resource categories in the SERVM database.

Table 4. Resource types modeled in SERVM

Resource Type	Description of Category
(T)hermal	Combustion turbine
(F)ossil	Fossil steam generators
(N)uclear	Nuclear generators
(R)enewable	Renewable generators whose output is dependent on weather patterns – non-dispatchable and not economically triggered
(C)urtailable	Demand response with constraints such as hours per day or month
(P)umped Storage (used to model all storage facilities)	Storage resources that can either consume or generate electricity; available energy and round-trip efficiency are essential modeling inputs for this resource type
(H)ydropower	Hydropower facilities that are not pumped storage; they are modeled as one of three subtypes – emergency, scheduled, or run of river

Proposed data sources for generic facility inputs are summarized in Table 5, below. The table does not list specific variable names in SERVM, but instead gives a less specialized narrative name. These data fields are common to all types of resources. For some data fields, it is easy to process existing data into SERVM data formats, but data reconciliation is difficult. For example, some plants with more than one unit are modeled as a single combined unit in one source dataset, but as two separate units in another dataset. Combined cycle plant configurations are often challenging, and judgment calls are needed. ED staff will evaluate all judgment calls with other parties to ensure the accuracy and reasonableness of decisions. It is also important to note that these values can vary by month and by year – meaning a generator can have a heat rate, ramp rate, maximum capacity, or any other variable that changes across different months and different years in the model.

Table 5. Generic data inputs common to most resource types (T, F, N, R, C, P, and H)

Variable	Applicable Gen Types	Sources/Comments
Resource name	All	CAISO MasterFile for resources located in CAISO; TEPPC 2024 Common Case dataset for resources outside of CAISO (including resources in LADWP or SMUD territories)
In service and retirement dates	All	CAISO MasterFile for resources located in CAISO; TEPPC 2024 Common Case dataset for resources outside of CAISO (including resources in LADWP or SMUD territories)
Region location	All	CAISO MasterFile for resources located in CAISO; TEPPC 2024 Common Case dataset for resources outside of CAISO (including resources in LADWP or SMUD territories)
Minimum and maximum MW production level (P_{min} and P_{max})	All	CAISO MasterFile for resources located in CAISO; TEPPC 2024 Common Case dataset for resources outside of CAISO (including resources in LADWP or SMUD territories). Values can be month-specific.
Fuel type (i.e., natural gas, biogas, nuclear, etc.)	T, F, N, R	CAISO MasterFile for resources located in CAISO; TEPPC 2024 Common Case dataset for resources outside of CAISO (including resources in LADWP or SMUD territories). Price curves for natural gas are discussed in the thermal resources section, below.

Each type of resource has some inputs that are unique to it. The following sections give more detail regarding specific resource types in SERVM and ED's proposed data sources to populate the database for modeling.

B. Disaggregating Aggregate Units Into Child Units

Staff generated unit inputs from ISO MasterFile data and the TEPPC 2024 Common Case as specified earlier in this paper. Staff did some amount of disaggregation on the two data sources, however, when it was apparent that between databases a combination of units were listed as one aggregated unit. Staff believed that in the case of peakers and combustion turbines, the model would produce more accurate results when aggregated units were modeled individually. This presented the challenge of generating unit inputs for individual units broken off of aggregates. Table 6 summarizes how individual unit inputs were generated.

Table 6 Generation of Inputs for Child Units from Aggregate Units

Input Field	Disaggregation Process
Inservice date	List same inservice date for each child unit as the aggregate unit – in effect all child units came online at same time and will retire at same time
capmax	Assume capmax of aggregate unit is total of all child units and divide capmax equally among child units unless there is a reason to do otherwise
capmin	Assume capmin is the capmin of one child unit and use that value for all child units, assuming each child unit has the same capmin
Minimum on time and minimum down time	Assume value is equal for all child units
Fuel type	Assume all child units consume same fuel as aggregate – use same value for all child units
Ramp rate	Assume ramp rate is total of all ramp rates of all child units, and divide equally among child units
Start up time	Assume start up time is the same for all child units, and use the value for the aggregate unit as the value for all child units
Start up costs	Assume value is equal for all child units

C. Thermal Resources – Types T, F, and N

The following discussion covers several types of information that are specific to thermal resources and are not common across other types of generators. They include heat rate, ramp rate, and forced and planned outage information. Because ED staff intends to conduct its reliability modeling utilizing a blend of both aggregate heat rate and ramp rate data from the TEPPC Common Case (consistent with CAISO and SCE analysis) and unit-specific heat rate and ramp rate values generated based on the CAISO MasterFile, there are some inputs that can be posted publicly and some that cannot. The difference in analytical results, and whether the differences are significant, will inform the amount of effort to put into further unit-specific analysis.

1. Heat Rates

Heat rates of dispatchable generators often vary over the operating range of the generator. It is important to characterize a generator's heat rate profile over the operating range of the plant so that SERVM can properly compute the marginal cost of dispatching the generator. SERVM can model generators with a single heat rate (usually a MW weighted average heat rate), or SERVM can create a curve based on a quadratic equation that can vary the marginal heat rate over the range of the generator's operation.

SERVM takes as inputs three values, each being the three coefficients on the quadratic equation.

The TEPPC Common Case 2024 dataset includes heat rate values for individual generators that are aggregated averages and neither unit specific nor variable across the unit dispatch range, and these simplified, constant heat rates can be input into and modeled by SERVM. However, there are tradeoffs to evaluate, balancing simplicity and transparency with accuracy and precision. For example, an individual generator would be undifferentiated from other generators in the same “class”, and thus it would be impossible to accurately project the actual dispatch of the facility in economic dispatch; as a result, the generator might be dispatched unrealistically throughout its operating range. In a sense, all similar power plants and all points in the operating range of those plants are treated equally under these modeling assumptions.

While transparency and simplicity are important, there are very important reasons to distinguish between generators. It is important for policymakers to appreciate distinctions for procurement oversight by discouraging procurement of worse functioning or less economical plants, and for operations in projecting what actual revenues and costs individual generators will encounter. It is also important for system modeling to ensure that dispatch results are realistic.

In light of these benefits to incorporating more accurate heat rates, ED staff proposes to use CAISO heat rate information for generators included in the CAISO in the MasterFile. ED staff can use MasterFile segment information to create heat rate curve coefficients for input into SERVM that represent the best fit across the whole range of heat rate segments included for a given unit. For those generators not listed in the CAISO MasterFile, ED staff proposes to use the TEPPC 2024 Common Case, which is the most current and complete source of information for all balancing authorities within WECC. The 2024 dataset also includes incremental development in between 2012 and 2024.

2. Ramp Rates

Like heat rates, ramp rates also vary across the operational range of a facility. Facilities also may have differing ramp rate ranges in the up direction versus the down direction. Thus SERVM allows for the entry of a set of ramp rate segments for each facility, both in the upwards and downwards direction. ED staff proposes to source ramp rates from the TEPPC 2024 Common Case set for generators outside of CAISO and to use the MasterFile ramp rate segments to create the portfolio of ramp rates both upwards and downwards to represent each generator for generators inside of CAISO. The TEPPC data represent what is likely a class average ramp rate; this makes the data less accurate, but enables it to be public. MasterFile data, meanwhile, is unit-specific and offers a glimpse as to the variations of ramp rates across the operating regions of the generators; however, it is confidential. Generator Forced Outage and Planned Maintenance Inputs

To model generators properly, some data regarding the chances of outages on those generators are needed. SERVM makes use of outage data by modeling generators with a distribution of time to fail, time to repair, and partial outage states. Table 7 lists the variables in SERVM that relate to forced or maintenance outages on units. The table does not list specific variable names in SERVM, but instead gives a less specialized narrative name.

Table 7. Inputs related to forced and planned outage hours and statistics for SERVM

Variable description	Comments	Sources/Comments
Availability	Percentage factor (1- percent of time unit is unavailable)	At this time, ED staff will source all of these inputs from GADS data, using class averages.
Time to fail	User can specify a distribution hourly values for how long a resource will run before it fails. SERVM draws a value from this distribution to draw outages on resources - user can specify either high values (making generators more reliable) or low values (making generators less reliable).	
Time to repair	Given in hours, this variable is how long a resource is out when it is on outage. Users can specify a number of hours for planned and forced outages separately.	
Partial outage derate	User can specify partial outage states	
Maintenance periods	Unit specific variable users can use to specify more than one maintenance period for each year	
Start up probability	Users can specify what the probability is for resources to fail upon startup	

Since 2010, generator owners operating in North America have been required to electronically submit outage data that describes each event that occurs at their generator to the North American Electric Reliability Council (NERC) in a standard format. Before that, the data submission was voluntary and non-electronic. Generator Availability Data Systems or (GADS) data is commonly used for purposes of modeling generator outages in other states. This data is available to CPUC staff via license from NERC. GADS data is reported to NERC by individual generators. Thus unit specific data is available, although unit specific data would be confidential. For the time being, ED staff has generated class averages for these variables, using the following categories to differentiate generators:

- Steam Turbines in California
- All Steam Turbines including those in California
- Combustion turbines within California
- All Combustion Turbines including those in California
- Combined Cycle plants within California
- All Combined Cycle plants including those in California
- All cogeneration facilities including those in California (there were insufficient facilities to generate averages solely for California plants)

Our use of GADS data is in contrast with the modeling that the CAISO completed in support of the Commission's Long Term Procurement Plan (LTPP) during 2012; for that modeling, the CAISO generated outage statistics based on its internal outage logging system. The CAISO uses data it gathers from generators via the Scheduling and Logging Interface for California (SLIC) database to generate class average summary statistics. The SLIC system however is due to be retired in December 2014, and the new Outage Management System (OMS) will replace it.[15] While having the advantage of being public, class average values fail to meaningfully differentiate between generators that in reality perform quite differently.

As the CPUC works to replace the San Onofre Nuclear Generating Station (SONGS) and other units that use once-through cooling (OTC), ED staff believes there is a particularly significant need to accurately differentiate between individual generators (some of which are scheduled to come into compliance with OTC requirements) in order to measure how reliability will be affected by forthcoming retirements and retrofits. Moreover, as the generating fleet moves from fossil-based resources that largely operate in baseload orientation to fewer fossil generators seeking to balance an ever increasing ratio of energy generated by intermittent resources, differentiating between generators with regards outage rates is important to gauge the reliability effects of this transition. This level of granularity is needed to accurately assess how much reliability and flexibility is served by those generators that retire (even differentiating between individual OTC generators) and how the new generators recently brought online and those in planning provide more, less, or equivalent reliability and flexibility.

3. ELCC compared to "Perfect Capacity"

ELCC is calculated by measuring the reliability of the system (in LOLE), and achieving a LOLE of 1 event in ten years. Then, the target generator is removed, a substitute is added in, and LOLE is recalculated. We calibrate the LOLE results so that we add in the right amount of substitute capacity to achieve the same LOLE as the system with the target generator included, then calculate a ratio of target generator MW to substitute generator MW. This ratio (in percent) is referred to as the ELCC of the target generator.

It is important to specify exactly what the substitute capacity is in terms of performance, outage rate, and other characteristics. It is feasible to compare the target generator (in our case, wind or solar facilities) against a generic peaker plant. One could choose an existing plant to compare against, or one could compare against "perfect capacity". A perfect generator is one with operational and performance characteristics that ensure optimal ability of that generator to contribute to reliability. In essence, a "perfect" generator contributes reliability to the system equivalent to the size of the generator – there is no derate for performance. It is an impossible standard of course, since no generator operates perfectly, without any equipment failures or with no time to start up. No generators are "perfect" and it is just a theoretical modeling convention, but comparison against "perfect capacity" allows all generators to be rated against each other. Even new peaker plants will not have an ELCC of 100%.

ED staff entered generic "peaker" generators into the database and added the generators in proportion to the reliability contribution of the wind or solar facilities removed. Table 8 lists the characteristics of the perfect capacity.

[15] The CAISO OMS project page is linked here: http://www.caiso.com/informed/Pages/StakeholderProcesses/OutageManagementSystemProject.aspx

Table 8 Resource Characteristics of Perfect Capacity

Variable description	Description	Value of Variable
Capmax	Maximum generation level	204.2 or 102
CapMin	Minimum capacity level (PMin)	1 MW
Availability	Percentage factor (1- percent of time unit is unavailable)	1 (indicating perfect availability)
Time to fail	User can specify a distribution hourly values for how long a resource will run before it fails. SERVM draws a value from this distribution to draw outages on resources - user can specify either high values (making generators more reliable) or low values (making generators less reliable).	90000 (never fail)
Time to repair	Given in hours, this variable is how long a resource is out when it is on outage. Users can specify a number of hours for planned and forced outages separately.	0 (Repairs instantly)
Startminutes	How long in minutes for the plant to start up	2 minutes
Maintenance periods	Unit specific variable users can use to specify more than one maintenance period for each year	None
Start up probability	Users can specify what the probability is for resources to fail upon startup	0 (Never fails on startup)

4. Natural Gas Price Forecasts

The natural gas price forecasts utilized by SERVM were developed by the California Energy Commission (CEC), consistent with the 2013 Integrated Energy Policy Report (IEPR) provides the burnertip prices utilized in the model, including the natural gas hub and transportation prices for 2016, in nominal dollars. ED staff has used the CEC NAMGas report to ensure consistency across agencies; however, staff recognizes that the 2015 natural gas price forecasts are above current future price strips[16].

ED staff used the CEC NAMGas data to create both annual fuel price projections for each hub, but also fuel handling inputs (the “csthnd” variable in SERVM). Each individual unit was linked to a particular fuel price curve as well as given a fuel handling variable. These values are in addition to other economic variables that SERVM uses to economically simulate operation of a particular unit. In addition to fuel price and fuel handling charge, a unit would also have cost variables for startup cost and variable operations and maintenance (“strtup” and “cstvar” variables in SERVM respectively).

5. Variable Operating and Maintenance Cost

Fuel prices and variable operating and maintenance costs make up the cost to a particular generator of generating electricity. Variable operating and maintenance costs are expressed in $/MWh and factor

[16] http://www.cpuc.ca.gov/NR/rdonlyres/BBD1123F-1900-4396-BE3B-105B726C65DD/0/FPC.xls

into dispatch costs. SERVM uses the Variable O&M as a means of dispatching facilities in economic order, and creating overall production costs. Facilities with higher or lower Variable O&M are less likely to be dispatched all else being equal than those with lower costs.

The actual variable O&M costs of each facility are both confidential and difficult to arrive at. Analysis of each individual contract would determine the cost values for each particular facility, and this value is likely impossible to publish. It is important to note that this value, though generally reflective of technical specifications of generating equipment, is also influenced by subjective contracting realities, such as labor costs. Staff located a suitable proxy in the CAISO MasterFile, with the costs used by the CAISO to developed default energy bids. Staff elected to use values from the CAISO GRDT data template for resource modeling, posted to the CAISO website. The data values are included in Table 9 below.

Table 9 Variable Operations and Maintenance Costs

Type of fuel	Variable O&M ($/MWh)
Solar or Pumping Storage Hydro	$0
Nuclear	$1.00
Coal or Wind	$2.00
Other hydro	$2.50
Combined Cycle or Steam Turbines	$2.80
Geothermal	$3.00
Biogas	$4.00
Gas Turbines or Reciprocating Engines	$4.80
Biomass or Waste	$5.00

Source: CAISO Resource Modeling website link:
http://www.caiso.com/market/Pages/NetworkandResourceModeling/Default.aspx

6. Imports and Direct Sales

The WECC interconnect is a very complicated region, with power flowing over numerous transmission interfaces. Several large plants provide energy to multiple regions, and provide valuable reliability service across WECC. Some regions are more dependent on direct sales from outside the region than others, and it is very important to link regions with the generating plants that supply them with power. For example, Southern California Edison relies on imported power from among other facilities, the Palo Verde Nuclear Station in Arizona and Hoover Dam in Nevada. Via the “direct-sale” variable, SERVM allows users to identify a unit and the region to which it directly sells the power.

A drawback with the “direct_sale” variable however is that the imported generator is dispatched as a must run facility, without economic dispatch considerations. Thus there is the possibility of unrealistic dispatch patterns. For those external facilities that are imported into a region but are dispatched economically, those facilities were listed as being within the regions they are imported into. This preserved the economic dispatch function. For generators that re dispatched as must run, however, such as nuclear facilities or intermittent renewable facilities, the “direct_sale” variable did not produce unrealistic dispatch.

The CEC provides capacity supply forms for all LSEs within California, listing for all LSEs (including SMUD and LADWP) the unit specific sources of capacity that the LSE is relying on to meet energy needs. These Utility Capacity Supply Forms are updated annually, public, and posted to the CEC website.[17]

ED staff used these forms to ensure that facilities across WECC are properly providing capacity to all LSEs in California.

D. Renewable Resources – Type R

The major distinction in SERVM between Type R resources and other types (such as F, T, or N) is in how resources are dispatched. Type R facilities (whether renewable or not) are modeled with production that is dependent on weather, and not dependent on economic logic. Type R facilities (loosely here called renewable) include wind and solar photovoltaic (PV) facilities. Other renewable resources, such as geothermal, biomass, and biogas generation facilities, are more accurately modeled economically via production cost dispatch; thus, the term "renewable" is really shorthand for weather-dependent intermittent must-take resources. Thus facilities that are going to be modeled with prices and startup costs, eventually including solar thermal facilities, will be modeled as Type F or T units. The two facility type options enable SERVM to model resources either as peakers which operate at a single operating level or as a thermal resource capable of operating at various operating levels.

This section details the inputs and assumptions utilized in modeling type R resources, including the methodology for creating weather-based wind and solar photovoltaic generation profiles.

1. Wind and Solar Generation Profiles

Wind and solar facilities have significant dependence on ambient weather conditions, which must be taken into account to correctly predict their output. Their output is a function not just of wind speed and solar irradiance, respectively, but also of other weather parameters such as cloud cover and temperature. Complicating this correlation is the fact that publicly available weather data is restricted to standardized locations (generally airports), and is not specific to the exact location (including altitude/height and orientation) of individual renewable energy facilities.

Additionally, renewable energy projects employ a multitude of different technologies, each of which may have a different response to the same weather conditions. For example, tracking and non-tracking PV will generate different amounts of electricity under the same weather conditions. Panel orientation also contributes to significant differences between non-tracking facilities. Solar thermal technology has an even more divergent weather response, relative to solar photovoltaic technologies. Because of the unique features of solar thermal facilities, they are not addressed individually in this document at this time; they are lumped together with solar PV facilities. The methodology for modeling solar thermal facilities will instead be addressed in a future document.

[17] These forms are posted to the CEC website here: http://energyalmanac.ca.gov/electricity/s-1_supply_forms_2013/

To accurately reflect the variability in wind and solar photovoltaic generation, modeling of solar and wind facilities requires mapping of the power output of existing and new facilities utilizing various technology types to the 33 years of historical weather that are modeled in SERVM. This mapping results in hourly performance profiles for each year of weather data, representing the overall variability of wind and solar production related to weather.

There are multiple possible approaches to developing such hourly performance profiles. One approach is to utilize generation profiles created by key stakeholders who are already conducting similar facility performance modeling. For example, developers need to forecast the generation profiles of their facilities in order to predict potential energy revenues and inform bids into RFOs or energy markets. Thus they could be helpful in developing similar production profiles for use in SERVM. Utilities also have an interest in predicting potential generation for resources that they are considering for contracting, operation, or management. Both developers and utilities may be able to create annual synthetic production profiles based on the same publicly available NOAA weather data utilized in SERVM synthetic load profile generation. Thus, there are several potential sources of wind and solar generation profiles that could be used.

However, there could be drawbacks to utilizing manufacturer, developer, or utility-supplied data for reliability modeling. It might be difficult to match potential production to load profiles or weather profiles, as the manufacturer curves or utility information may predict performance based on other factors, or may be based on single-year weather projections that cannot be extrapolated to the entire 33 years of weather history required for consistency with other weather-based SERVM inputs and algorithms. Data for performance of wind and solar facilities external to California may also be much more difficult to access, complicated by different utility service areas, regulatory jurisdictions, and information access guidelines.

Staff has pursued an alternative approach for the 2016 RA compliance year, mapping standard, publicly available weather information to the power output of wind and solar facilities using either normalized profiles based on output from the NREL PVWatts[18] calculator (for PV facilities) or off-the-shelf neural network modeling software (for wind facilities) discussed in Section V.A.3 and already used for development of load shapes. The neural network determines the relationships between weather/facility input variables and wind facility production, and outputs a predictor file. With this predictor file, ED staff, together with Astrape Consulting, constructed synthetic wind production profiles for existing and new facilities that correspond to the 33 years of weather history and associated synthetic load shapes utilized by SERVM. The large sample of weather years will enable SERVM to capture realistic variability in generation from wind and solar facilities. However, creating these wind and solar facility profiles required extensive performance, technology, and weather data.

It is expected that the synthetic production profiles (and the predictor file, for wind facilities) will be reconstructed at least every two years to reflect the evolving relationships between weather and

[18] http://pvwatts.nrel.gov/

production (considering such issues as technology improvement and locational clustering of installed capacity). Intra-hour variation and forecasting uncertainty is addressed separately in Section H of this paper. The section below describes:

1) the sources for performance data,
2) the weather data and regions modeled,
3) the development of technology categories to group similar responses to weather inputs,
4) neural network modeling or PVWatts-based calculation to be utilized to create weather response predictions for each technology category, and
5) how these predictions are input into and used by the SERVM software.

ED staff expects these generation profiles to be very important in modeling overall reliability of the electricity grid, and expects variability in production of wind and solar facilities to be one of the more important drivers of reliability risk in the future, as wind and solar resources continue to account for an increasing share of the California generation mix. Thus, while this area of data development has required significant effort, the current generation profiles and any future refinements will also pay off in greater modeling accuracy.

a) Performance Data Sources and Assumptions

ED staff receives hourly settlement data (in hour-ending or "HE" format, representing average output over the hour) from all facilities represented by scheduling resource IDs on the CAISO Master Generating Capability Data List. These data have been supplied for facilities since 2008 for use in QC calculations, and were used to validate the synthetic shapes that were developed.

For the construction of synthetic wind profiles for facilities both inside and outside of the CAISO service territory, 2004-2006 hourly wind speed and generation profiles were taken from the NREL Western Wind Resources Dataset.[19] The dataset includes over 30,000 potential wind sites nationwide, with generation profiles for each site assuming a 100-meter hub height and 100-meter rotor diameter. In modeling facility performance, wind facilities within each SERVM region were assumed to have the same geographic distribution as RPS-certified wind facilities in that region, as reported by the CEC.[20]

Solar PV profiles were calculated based on several performance assumptions, as shown in Table 10, below.

[19] http://wind.nrel.gov/Web_nrel/

[20] http://www.energy.ca.gov/portfolio/documents/List_RPS_CERT.xls

Table 10. Solar PV Facility Performance Inputs

Performance Input	Assumption
Reference Efficiency	14.94%
Nominal Operating Cell Temperature (NOCT)	45°C
Temperature Coefficient	0.0045
Short Circuit Coefficient	0.000545
Solar Radiation Coefficient	0.12
Reference Temperature	25°C
Inverter Efficiency	97%

b) Weather and Regions Modeled

All solar PV weather data are sourced from the NREL National Solar Radiation Database (NSRDB).[21] Solar data from 1980-1990 come from 58 unique sites, while 1991-2010 data come from 225 unique sites. These are referred to as the TMY2 and TMY3 datasets. These data are comprised of the following inputs:

- Local solar time (calculated based on latitude, longitude, day of year, and other variables)
- Direct radiation
- Diffuse radiation
- Air temperature
- Wind speed

As with wind, the solar facilities within each SERVM region were assumed to have the same geographic distribution as RPS-certified solar facilities in that region, as reported by the CEC. [22] In other words, when developing solar generation profiles, staff utilized the weather data for the NSRDB sites nearest to each CEC-certified RPS facility. Because NSRDB data extends only to 2010, previous years' weather is used to determine the 2011 and 2012 generation profiles. The modeled solar profile for 2011 is identical to the pattern seen in 2008 and the modeled solar profile for 2012 is identical to the pattern seen in 2009 for all regions. A more rigorous approach to developing solar profiles for 2011 and 2012 may be used in the future.

Wind weather inputs are sourced from 33 years of NOAA data, as described in Section IV.A above. Specific weather inputs are: hour of day, wind speed, temperature, dew point, and cloud cover.

Because weather data are available at limited locations, and because modeling time increases dramatically as granularity increases, one weather profile was compiled per wind or solar technology for each modeling region, for each historical weather year being modeled. To create each region's weather profile, staff calculates a weighted average hourly weather profile based on one to three weather stations that are selected as indicative of a given renewable technology's generation capacity in the

[21] http://rredc.nrel.gov/solar/old_data/nsrdb/

[22] http://www.energy.ca.gov/portfolio/documents/List_RPS_CERT.xls

region. In other words, if capacity of a particular technology type is primarily located in the northern part of a region, the weather modeled for that region in SERVM will be more heavily weighted towards the northern weather station(s) selected for that region. The location of each facility is sourced from it's the CEC RPS Certification Report[23]. Alternative approaches to weather station weightings may be considered if SERVM is utilized for longer-term modeling in the future; sensitivity to weather station selection will also be tested.

An important exception to the above methodology is the treatment of wind in the SCE TAC Area modeling region. Because most wind resources are in either the San Gorgonio or the Tehachapi areas, and because these areas have very distinct weather, wind in the SCE TAC Area is modeled with two separate weather profiles, one for each of these two sub-regions. Individual wind facilities in the SCE region are also separated into San Gorgonio and Tehachapi sub-regions, with facilities located in the Big Creek/Ventura Local Area assumed to be in Tehachapi and all others assumed to be in San Gorgonio. Wind and solar resources outside of the United States are treated differently, due to data limitations. Neural network modeling and PVWatts calculations are not being conducted for Canadian and Mexican wind and solar resources. Rather, the results from similar US regions – the Pacific Northwest and the average of IID and New Mexico, respectively – are applied to the Canadian and Mexican weather shapes in order to develop weather-based wind and solar production profiles for these regions. In addition, Wyoming wind was ascribed weather from Colorado, as there were no production profiles developed for Wyoming.

In developing wind technology and weather response relationships in the neural network software, the representative regional weather was input. In the future, the model performance will be tested using more local weather for individual facility locations, where available; however, neural networks generally yield better predictive capability when developed with a more limited set of parameters. Too many variables involved in the creation of the predictor file can create muddied correlations that lead to bad predictions of weather and generation relationships.

c) Technology Categories

Wind and solar resources are grouped into five technology categories that output at similar levels under identical weather conditions (as a percentage of their maximum output). Wind is divided into two categories: smaller and larger turbines. Smaller turbines are modeled assuming 80 meter hub height and 85 meter rotor diameter, while larger turbines are modeled assuming 100 meter hub height and 100 meter rotor diameter. Solar facilities are assigned to one of two categories: fixed tilt PV, and tracking (single axis) PV. Although better modeling methods will be developed for solar thermal facilities in the next phase of the modeling project, for the time being they are included in the category of tracking solar facilities. Additional possible categories that could be explored in the future include PV with storage, wind with storage, or south facing versus west facing PV.

[23] The CEC RPS Certification Report is linked here:
http://www.energy.ca.gov/portfolio/documents/List_RPS_CERT.xls

Each technology category and region (or sub-region, in the case of wind in the SCE region) is evaluated separately to develop weather response predictions within that region, and across that category type. SERVM models each facility's generation based on both its technology category (indicative of response to weather) and weather region (the relevant weather input).

For the 2015 RA compliance year, staff proposes to divide wind facilities into two size-based categories: (1) "large wind" and (2) "smaller wind". Large wind is modeled based on generation profiles in the NREL Western Wind Resources Dataset, which assumes 100-meter hub height and 100-meter rotor diameter. Smaller wind is modeled based on adjusted generation profiles that assume 80-meter hub height and 85-meter rotor diameter. There are formulas that can extrapolate generation profiles for differing turbine heights.

However, data are limited. The ISO Generating Capability Data List places units in Local Areas, which are translatable to regions in the SERVM database, while for facilities outside of ISO, the WECC TEPPC database provides the primary reference for location, generation type (solar, wind, tracking, fixed, etc.), and date of commercial operation.[24]

Therefore, when conducting modeling, individual facilities will be assigned one of the two production profiles based on vintage. Facilities built in 2011 or earlier will be assigned to the "smaller wind" category, while facilities built in 2012 or thereafter will be assigned to the "larger wind" category. This categorization is based on the average hub heights and rotor diameters reported in the U.S. Department of Energy 2013 Wind Technologies Market Report.[25] In future years, staff recommends considering switching to wind facility categories based on individual generator size in MW, or considering alternative hub height and rotor diameter assumptions.

Regardless, direct category assignment (rather than imputing technology category based on vintage) will be possible for facilities in California either from the monthly RPS Project Development Status Reports (PDSRs) or via data gathered from the CEC Wind Performance Reporting System. Information for facilities outside of California will continue to be derived from TEPPC 2024 Common Case information.

Solar facilities were assigned to either fixed or tracking PV generation profiles. For facilities with CAISO settlement data, this was determined by analyzing late-afternoon generation on a sunny day and assessing whether generation levels (normalized for facility capacity) were indicative of fixed PV (lower generation in the late afternoon) or tracking PV (higher generation in the late afternoon). To ensure accuracy, facilities with known technology types (as submitted by the CPUC-jurisdictional IOUs) were analyzed to develop an appropriate cut-off point in assigning those facilities with unknown technology to one of the two possible categories. For facilities outside of the CAISO service territory, the TEPPC

[24] For more information on TEPPC, see https://www.wecc.biz/committees/BOD/TEPPC/Pages/TEPPC_Home.aspx.

[25] http://energy.gov/sites/prod/files/2014/08/f18/2013%20Wind%20Technologies%20Market%20Report_1.pdf

2024 common case specified which of the two categories is more appropriate for modeling individual PV facilities.

d) PV Production Profile Development

Because full weather data are available from 1980 through 2010, PV production profiles could be developed by inputting the weather and facility data discussed above into the NREL PVWatts[26] calculator. The calculator output a facility-specific hourly generation profile for each facility in each year, which was then scaled to reflect that facility's capacity (sized in MW). The fixed-tilt PV facility curves within each region were summed to create aggregate generation profiles for the fixed-tilt PV technology category, with one hourly profile (8760 consecutive hours/year) per region per year. The tracking PV facility curves within each region were similarly summed for each year to create regional hourly production profiles for that category. Different sites were available for 1980 - 1990 and 1991 - 2010, so some calibration of the datasets was required to ensure continuity. To ensure consistency, the range in output duration curves of the 1980 - 1990 was shaped to match the range in duration curves seen in 1991 - 2010 data. This entailed calibrating the number of hours at max output, moderate output, low output and every point in between. Since NSRDB weather data only runs to 2010, 2011 is assumed to be the same as 2008 and 2012 is identical to 2009.

The result of the above exercise is 33 years of regional generation profiles for fixed-tilt PV and 33 years of regional generation profiles for tracking PV. With 17 regions modeled in SERVM, this yields 1122 hourly generation profiles for solar PV facilities. In light of this already-high level of modeling complexity, staff does not recommend creating more granular generation profiles at this time. Additionally, facility-specific profiles utilizing confidential information (such as facility-specific historical hourly generation data) could not be shared with the public, reducing transparency. Nevertheless, this is a refinement that could be considered in future years. The current solar PV results are not confidential, and they have been posted for stakeholder review and validation on the RA website.[27] Staff looks forward to receiving feedback from parties on these generation profiles.

e) Wind Production Profile Development

(1) Neural Network Modeling

Because weather (NOAA), expected resource generation (Western Wind), and actual resource generation (CAISO) data are all available for the 2004-2006 period, generation data can be directly compared to the weather data on which it depends. However, generation depends on many aspects of weather, complicating the relationship. The fact that SERVM weather region inputs are not specific to the precise resource location further obscures the relationship between weather and generation output. To create a reasonably accurate prediction of generation output in response to weather, a neural

---

[26] http://pvwatts.nrel.gov/, http://www.nrel.gov/docs/fy14osti/60272.pdf

[27] http://www.cpuc.ca.gov/PUC/energy/Procurement/RA/Probabilistic+Modeling.htm#Modeling

network can be used to map weather to output and create a relationship file that can be used for new facilities and weather years. This process is similar to the use of a neural network to create synthetic load shapes, which are used elsewhere in the SERVM model and described in Section V.A above.

First, the regional weather data are placed into a spreadsheet for a given technology category.[28] One variable is chosen as the primary predictor of generation output, and is placed in the left-hand column. In the case of wind technologies this is the region's wind speed (from NOAA). The other key weather inputs (from NOAA, as discussed previously) are included as additional columns. These data are paired with actual facility and potential generation data from facilities of the given technology type, sourced from the Western Wind Resources Dataset, as previously discussed. The hourly generation data from the Western Wind Resources Dataset (WWRD) were first scaled by a ratio of each facility's nameplate capacity in MW to the facilities modeled in the WWRD, and then summed to create a single aggregate generation profile for a given region and year. It was this aggregate profile that was utilized in neural network modeling.

This neural network model trained itself to see the underlying relationships between the hourly generation data and the other columns of input data (much like a dynamic iterative regression model). It developed predictive relationships between the columns of data (the variables mentioned previously such as temperature and humidity), and produced an algorithm that was able to predict relationships between regional wind speed, secondary weather variables, and generation facility output.[29] Once that was completed, the algorithm predictor model could then produce a generation forecast from any set of NOAA weather data, for any facility that falls under the given technology category (including new facilities).

However, because of significant volatility and randomness in wind data, neural network models tend to predict average values more frequently than they actually occur. For this reason, there was some adjustment to the distribution of wind predictions after the initial neural network modeling. Staff performed validation on the resulting performance shapes to ensure accuracy, by comparing the resulting shapes for the CAISO regions to actual historical generation patterns and normalizing to ensure that the predictor files output similar generation magnitudes and duration curves, compared to historical generation. Staff also checked that the capacity factors output by the predictor files were reasonable. Once the predictor files and additional processing were finalized, regional production profiles were created for all wind facilities in each weather year, using hourly NOAA weather data from 1980 to 2012.

[28] As previously mentioned, wind resources in the SCE TAC Area region are modeled separately, by San Gorgonio/Tehachapi sub-region.

[29] The algorithm was trained until the distribution of peaks and valleys was adequate. Further calibration was performed in a later step.

After the wind profiles were created and validated to history, they were further modified to distinguish between large wind facilities and smaller wind facilities. Large wind facilities were assumed to have a 100 meter hub height and 100 meter rotor diameter, while smaller wind facilities were assumed to have an 80 meter hub height and 85 meter rotor diameter. Because the neural nets were based on 100 meter hub height NREL data, these were adjusted down based on wind speeds at lower hub heights to develop 80 meter hub height generation profiles.

Staff has shared these non-confidential aggregate shapes with stakeholders, providing them the opportunity to analyze and validate the results.[30] Staff expects to have a high amount of confidence in the resulting profiles due to the significant checking and validation that are to be performed.

(2) Category Normalization

In order to compare across wind facilities of varying sizes, output was normalized relative to sum of all the capacity in a particular category currently installed in the region prior to neural net "training". Additionally, because the neural network develops aggregate production profiles for a given technology category in a given region, differences in installed capacity over the training years must also be accounted for and normalized. One option is to assume that the smaller capacity installed in earlier years is representative on average of the larger total capacity in future years regardless of facility location or technology installed. However, this may be imprecise due to differences in the generation profile shape and volatility as more capacity is installed in new locations. To mitigate this problem, new facilities are instead assigned an hourly production profile from one representative existing facility determined to be "similar" in location and technology, and scaled to the MW size of the facility being modeled.

E. Demand Response – Type C

Demand response inputs and assumptions in SERVM are primarily based on the DR program filings/tariff and Load Impact Reports (LIRs).[31] Key inputs currently incorporated into the model are listed in Table 11, below.

Table 11. Current Demand Response Resource-Specific Inputs

Input (as applicable to the program)[32]	Units	Source
Maximum capacity	MW	LIR portfolio-adjusted load

[30] http://www.cpuc.ca.gov/PUC/energy/Procurement/RA/Probabilistic+Modeling.htm#Modeling

[31] The Load Impact Protocols followed in developing Load Impact Reports were specified by Decision 08-04-050, and modified by Decision 10-04-006.

[32] Different DR programs have different design constraints; as a result, different inputs will apply to different programs. If a program lacks a certain constraint (for example, no maximum number of dispatches per week), then the associated input will not be included in the specification of that program in SERVM.

		impacts, 1 in 2 weather[33]
Maximum dispatch days per week	days	Program tariff
Maximum consecutive dispatch days	days	Program tariff
Maximum dispatch hours per day	hours	Program tariff
Minimum minutes per dispatch	minutes	Program tariff
Maximum number of dispatches per day	dispatches	Program tariff
Maximum dispatch hours per month	hours	Program tariff
Maximum number of dispatches per month	dispatches	Program tariff
Maximum dispatch hours per year	hours	Program tariff
Maximum number of dispatches per year	dispatches	Program tariff
Minimum number of dispatches per year	dispatches	Program tariff
Look-Ahead	hours	Energy Division studies
Notification period	Hours/minutes	Either DA (10am), HA,
First month available each year	month	Program tariff
Last month available each year	month	Program tariff
Period Availability (i.e., weekdays from 2-6 pm)	days and hours	Program tariff
Curtail (Dispatch) price	$/MWh	CAISO Plexos assumptions or program tariff[34]
Emergency-only dispatch	Yes/No	Program tariff
Region[35]	Region name	Program tariff
Program in-service dates	mm/dd/yyyy – mm/dd/yyyy	Program tariff
Ramp Rate	MW/min	Program tariff
Program performance degradation (customer fatigue)	Percent degradation factor per day	

a) Dispatch Notice and Response Time

DR programs have different dispatch notice requirements (day-ahead, 30-minute-ahead, etc.), which are described in their tariffs. Once dispatched, they also have varying response times. These requirements,

---

[33] In the future, staff plans to modify the maximum capacity to account for both 1-in-2 and 1-in-10 weather conditions. See the "Potential Future Expansions/Changes to Inputs" section below.

[34] Most DR programs do not have a set price trigger. The assumptions adopted by the CAISO for its Plexos modeling are an approximation of a price trigger that corresponds to the actual dispatch criteria.

[35] These regions are used throughout the SERVM model, and are described further in the Weather Data and Regions section of this document.

whether a time of day cut-off or a minimum advance notice period, could be incorporated into the model in the future.

b) Resource Capacity

Currently, the maximum capacity for a given DR resource is set to its Load Impact Report (LIR) portfolio-adjusted monthly system peak values for 1-in-2 weather conditions. However, under more extreme weather conditions, performance for weather-dependent resources may exceed the 1-in-2 value, potentially reaching the LIR 1-in-10 capacity values. Apart from weather impacts, a DR resource may underperform or overperform relative to expectations due to variation in customer load and response.

To address the possibility of DR resources performing beyond the 1-in-2 value, staff plans to ultimately incorporate 1-in-10 values into the model as well. This can be accomplished by creating a "technology response curve" that maps regional temperature to changes in DR capacity. For 90^{th} percentile temperatures (the conditions under which the 1-in-10 LIR is calculated) and above, the LIR portfolio-adjusted monthly system peak values for 1-in-10 weather conditions can be used. For 50^{th} percentile temperatures (the conditions under which the 1-in-2 LIR is calculated) and below, the 1-in-2 LIR capacity values can be used. Linear interpolation can be used to approximate DR response between these two temperature bounds.

To address the possibility of over- or underperformance relative to expectations, three years of program history could be used to create a likely distribution of responses. The difference relative to expectation for a given dispatch can be defined as the percentage difference between the ex-post load impact found in the LIR and the daily forecast capacity predicted day-ahead. Each historical dispatch can be weighted according to the magnitude of the daily forecast capacity, so that larger dispatches are more heavily weighted. When a DR program is dispatched by SERVM, its response magnitude would then be adjusted upwards or downwards by selecting one of the historical performance data points. The performance point selected would be random, but weighted as previously discussed. While the necessary data for such adjustments have not yet been input into the model, the modeling functionality is in place, and staff plans to incorporate this performance uncertainty in the future. This could be accomplished with a variable that allows for randomly drawn output. For instance, if a DR resource has three performance levels of 90%, 100%, and 110%, and each is entered into the database, then one third of the time when it is dispatched it will operate at 90% of maximum, one third at 100% of maximum, and one third at 110% of maximum capacity.

c) Triggers

Most existing DR programs do not have a set price trigger. The model currently adopts assumptions used by the CAISO in its modeling to support the 2012 LTPP as an approximation of the price trigger that most closely corresponds to the actual dispatch criteria. Currently, a number of the DR programs are triggered via heat rate or emergency stage triggers, which are difficult to translate to price points; ED staff continues to explore alternative approaches to fit the current panoply of DR programs into the economic dispatch model in SERVM. The CAISO trigger price assumptions currently in use in SERVM are listed in Table 12, below.

Table 12. Demand Response Program Price Triggers Assumed

IOU	DR Program	Trigger Price ($/MWh)
PG&E	AMP-DA	1,000
	AMP-DO	1,000
	BIP	600
	CBP-DA	1,000
	CBP-DO	1,000
	DBP	1,000
	PDP	1,000
	PeakChoice	1,000
	SmartAC-Non-Residential	600
	SmartAC-Residential	600
	SmartRate	1,000
SCE	API	600
	BIP	600
	CBP-DA	1,000
	CBP-DO	1,000
	CPP	1,000
	DBP	1,000
	DRC-DA	1,000
	DRC-DO	1,000
	SDP-COM	600
	SDP-RES	600
	SPD	1,000
SDG&E	BIP	1,000
	CBP-DA	1,000
	CBP-DO	1,000
	CPPD	1,000
	DBP	1,000
	PTR	1,000
	SCTD	1,000
	Summer Saver Commercial	600
	Summer Saver Residential	600

Source: CAISO, September 2013.

d) Customer Fatigue

The SERVM simulations did not currently consider the impacts of customer fatigue on long-duration or consecutive dispatches. With appropriate data, such impacts could be incorporated in the future.

e) Look-Ahead

For DR programs with dispatch limitations, demand response providers may occasionally refrain from dispatching if they believe that the resource could be better dispatched at a later time. For example, if a week is expected to have steadily increasing temperatures, a DR resource may not be dispatched earlier in the week, even if the price trigger has been reached, in order to preserve the possibility of operating later in the week. This "look-ahead" dispatch decision is not incorporated into the SERVM model, but could be in the future.

F. Energy Storage Resources - Type P

While there are numerous different energy storage technologies, most can be described according to several key variables such as available energy, maximum output, maximum draw, and efficiency. This section describes these modeling inputs. However, because very little energy storage has been deployed to date, the testing protocols and sources will need to be developed over the coming months. Staff looks forward to parties' input on the current and potential future modeling inputs listed below.

1. Current Inputs

Input	Units	Source
Maximum rated discharge	MW	CAISO MasterFile
Total usable storage volume (given allowable depth of discharge)	MWh	Calculated based on testing: Maximum rated discharge * (discharge test duration)
Maximum rated charge	MW	CAISO MasterFile
Round trip efficiency	% efficiency	Calculated based on testing submitted to the CAISO: (discharge MW*duration) ÷ (charge MW*duration)
Capable of supplying non-spinning reserves	Y/N	Start time testing submitted to the CAISO demonstrating < 10 minute startup
Facility in-service dates	mm/dd/yyyy – mm/dd/yyyy	CAISO MasterFile, unless utilities have more current information
Scheduled maintenance and maintenance outage periods	% of month/year, date range, and/or hours to repair	Historical data from the CAISO, to be collected over time for new facilities
Able to provide regulation	Y/N	Ability to provide regulation, from CAISO MF

2. Potential Future Expansions/Changes to Inputs

Input	Units	Source

Ramp rate	MW/min	Testing submitted to the CAISO
Advance notice requirement	minutes	CAISO Master File
Startup, shutdown, or charge-discharge transition profiles	TBD	Testing submitted to the CAISO
Forced (full and partial) outage rates, time to failure, time to repair	various	Historical data from the CAISO, to be collected over time for new facilities

G. Hydropower Resources – Type H

All hydropower (hydro) resources that are not pumped storage[36] are modeled as Type H units. SERVM classifies these hydropower resources according to three subtypes: run of river (ROR), scheduled, and emergency hydro. Each of these resources can have capacity and energy levels that vary by month and year, in order to reflect the seasonal variability of this resource type.

Run of river hydro represents the minimum output that is expected to occur regardless of electricity system needs or economic dispatch. Scheduled hydro represents the portion of the hydropower fleet that can be economically dispatched, in light of monthly resource availability. Emergency hydro represents the capacity and energy that can be "borrowed" from scheduled hydro to address occasional, short-term electricity system emergencies. Table 13 lists sources for particular data inputs.

Table 13. Data Sources for Hydropower Inputs

Data	Source
Facility generation per month (MWh/month), 1980-2012	Form EIA-923: Power Plant Operations Report[37]
Facility locations (model region)	TEPPC 2024 Common Case, Form EIA-923, CEC Energy Almanac[38], and miscellaneous other sources such as the US Bureau of Reclamation[39]
Regional maximum capacity (MW)	TEPPC 2024 Common Case
Monthly hydro dispatch	EIA form 923
Hourly hydro flows within California	CEC historical monitoring data

Hydro resources[40] are modeled in aggregate, by subtype and modeling region, although granularity may be improved in future RA compliance years.[41] For example, all ROR hydro facilities in SCE service

[36] Pumped storage is modeled as Type P, as discussed in the energy storage section of this document, above.

[37] These data are available for download at http://www.eia.gov/electricity/data/eia923/.

[38] Data can be downloaded from http://energyalmanac.ca.gov/renewables/hydro/hydro.xls and http://energyalmanac.ca.gov/powerplants/Statewide_PP_8.5X11_hydro.pdf.

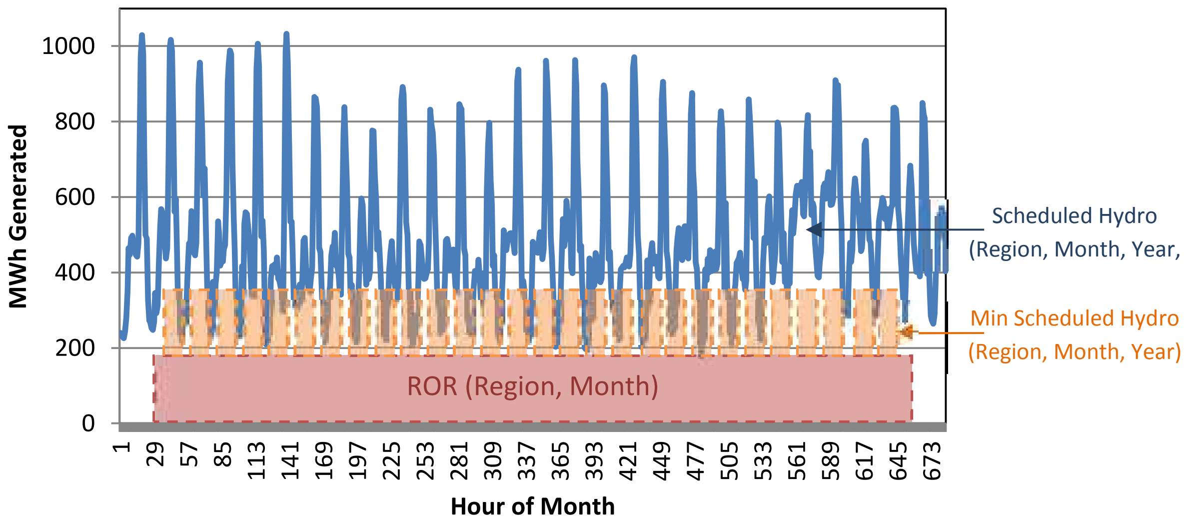
[39] Searchable database at http://www.usbr.gov/projects/.

[40] In this document, "hydro resources" refers to all hydropower resources excluding pumped storage facilities.

[41] See the Weather Data and Regions section for more information on the regions modeled in SERVM.

territory are modeled as one "unit" in SERVM. Before these "units" can be input, the aggregated energy and capacity for each region must be calculated, and then allocated across the three subtypes. An intuitive visualization of the resulting allocation can be seen in the randomized sample hydro generation shape below. The methodology used will be described in more detail in the following sections.

Figure 3. Sample hydro generation shape and sub-type allocation (based on randomized historical data)



For facilities in Canada and Mexico, hydropower generation shapes are based on the closest neighboring US region (the Pacific Northwest and Arizona, respectively), and simply scaled to the region's maximum capacity. Parties are encouraged to comment on whether different US regions (or a weighted average of multiple regions) might be more appropriate.

1. Regional Aggregation of Energy

Monthly hydro generation (MWh/month) for all existing hydro resources in WECC is listed in Form EIA-923 for the years 1980-2012. This actual historic generation by month is used to determine the energy available from ROR, scheduled, and emergency-only hydro profiles generated for each region. However, because it is reported to EIA on a facility-specific basis, each facility must first be assigned to its region in SERVM based on the particular facility's location in the TEPPC Common case. The generation from all facilities within a given region is then summed to yield the total energy generated in that region, in each historical month and year.

2. Run of River (ROR) Hydro Resources

The energy generation and capacity for ROR resources within a region are unique to each month, but uniform across all weather years. This is because the ROR unit represents a minimum output that is always present, regardless of weather variability and dispatch choices, and because there have been very few new hydro facilities developed in WECC over the last 33 years. The available energy is set to be the fifth percentile of MWh generated by all hydro resources for a given month and region. In other words, the value is set such that in 19 out of 20 years, hydro facilities in that region in that month produce more than that number of MWh.

The capacity of ROR units is calculated as the available energy value in MWh divided by the total number of hours in that month. ROR units are assumed to operate at this calculated capacity for all hours of the month, meaning there is no hourly or daily variation in output within a given month. In other words, ROR production is flat across all hours of a given month, across all years modeled.

3. Scheduled Hydro Resources

Once ROR energy and capacity have been subtracted from the total energy and capacity available to a region, the remainder must be allocated across the two dispatchable hydro subtypes: scheduled and emergency hydro.

The energy allocated to the scheduled block is equal to the total regional monthly generation less the ROR generation. A portion of the scheduled energy is allocated to a minimum daily schedule. This minimum schedule or generation (flow) per day is a variable that is unique to each month and year. This value is set to the tenth percentile of daily MWh generation in that month and year, and is sourced from CEC historical generation data. Regions outside of California for which data is lacking are modeled data generated for the most similar region for which we have sufficient data. In some months, the minimum generation per day may be very close to zero; if selecting the tenth percentile results in more generation being dispatched than is available, SERVM will flag the issue and the value will be reduced. The minimum daily schedule for each scheduled hydro profile is spread across a specified number of hours each day in equal amounts.

The remainder of the energy in the scheduled block is used to shave the peaks off net loads; in other words, higher output is scheduled in hours with higher net load. The capacity used to shave the peaks is related to the monthly generation. Available hydro capacity is allocated between emergency, scheduled, and run or river hydro based on higher or lower levels of hydro available generation and typical historical usage in each month. Scheduled, run of river and emergency hydro capacity always sums to total month specific capmax of the hydro fleet.

All scheduled hydro is dispatched one week in advance. The minimum generation quantity is scheduled to be centered on the anticipated peak load hour of each day. The number of hours over which that minimum generation is spread is set with a monthly variable. This variable is determined by observing CAISO settlement data and estimating the typical number of hours over which hydro facilities are scheduled in a given region and time of year. Non-CAISO regions use values based on the nearest CAISO region. Scheduled hydro above the minimum is economically dispatched, up to the maximum capacity calculated for that month.

4. Emergency Hydro Resources

Because emergency hydro resources are not intended for regular dispatch, they are triggered only by high market prices (currently set to $2,500) or load-shedding contingencies. These units allow a region's fleet to reach full nameplate capacity for approximately twenty hours. When emergency hydro is dispatched, the energy must be replaced by lowering scheduled hydro in some future hour. In this way the total energy for the month never violates the input energy. If no energy is available to borrow from future schedules, the emergency hydro capacity is unavailable. The sum of total capacity of emergency, scheduled, and run of river hydro is equal to the total capacity of the hydro fleet in each area.

The full nameplate capacity is sourced from the TEPPC 2024 Common Case. The available energy comes from the scheduled hydro unit in the region, to which the emergency unit is linked. The emergency unit is given the ability to borrow a MWh amount equivalent to a specified number of hours of full operation from the scheduled hydro unit.

H. Variability and Uncertainty in Load, Wind ,and Solar Profiles

In order to fully simulate the real behavior of the electric system, it is important to accurately simulate how actual people dispatch the electric system, and to accurately simulate how variable factors behave. Actual human dispatchers make forecasts of hourly profiles of load, wind generation, and solar generation one day ahead, then organize commitment and dispatch to meet those forecasts. Several hours later, as the operational day approaches, dispatchers again make forecasts, revising earlier forecasts, and adjust earlier commitments and dispatch in order to be increasingly accurate and updated. Finally, one hour ahead of an actual operational hour, dispatchers again revise and refresh their forecasts of load levels, wind generation levels, and solar levels and adjust dispatch to meet updated forecasts. In real life, day-ahead and hour-ahead dispatches do not exactly match actual load or reflect actual wind and solar generation, due to forecasting uncertainties. To account for this mismatch, resources must be brought on- or off-line on short notice, in order to ensure that real-time supply and demand are matched. In short, without incorporation of forecasting uncertainty, the need for the operators to employ ramping or flexible resources would be systematically understated.

Users of SERVM can simulate these forecast uncertainty effects by creating uncertainty distributions that represent bands around how dispatchers create and revise forecast of load, wind, and solar levels on various time scales. Without performing this analysis, staff would be left with a simulation that assumed “perfect foresight,” as if the grid operator knew for certain the next day’s and the next hour’s demand and must-take generation, and could dispatch resources perfectly and efficiently. Perfect foresight would unrealistically minimize need for flexible and rampable generation; although not as essential for the initial ELCC analysis, staff has developed data to create uncertainty distributions and will begin simulating dispatch and forecast uncertainty during 2015 as staff uses SERVM for operational flexibility analysis.

To create these distributions, staff examined hourly forecasts and actual values for wind and solar generation, as well as hourly forecasts and actual values of electricity demand over the last few years. A distribution of forecast versus actual “errors” was created, which then was fed into the SERVM database, so that SERVM could mimic the behavior of actual operators as they commit generation units in real life. The following sections detail the data gathering, data cleaning, processing, and model import methodology utilized by staff. As more data are available for analysis, these distributions can be periodically revised to ensure maximum possible accuracy and correspondence with reality.

1. Data Gathering

Staff downloaded historical load, wind, and solar forecast data as well as actual load, wind and solar MW data from the CAISO OASIS database.[42] For the analysis of forecast error in wind and solar generation, only two years (2012 & 2013) of forecast and actual generation data were collected, since there was not very much renewable generation available online until 2012. Load data from 2009 through the end of 2013 were analyzed. Data are in hourly increments.[43] All data are reported in MW, aggregated to each of the regions listed in Table 14.

Table 14 Load and Generation Forecast Data Summary

Locations	Data Types	Date Range
Northern and Southern California (NP15 and SP15)	Actual Wind Generation Day and Hour Ahead Forecasts	1/1/2012-12/23/2013
Northern and Southern California (NP and SP)	Actual Solar Generation Day and Hour Ahead Forecasts	1/1/2012-12/23/2013
PG&E, SCE, and SDG&E Service Areas	Actual Load Day and Hour Ahead Forecasts	3/6/2009-12/23/2013

2. Data Processing

a) Solar Generation Forecast and Actual Generation

Staff eliminated all solar generation and forecast values for hours that corresponded to after sunset and before sunrise each day. Hours where the legitimate forecast would have been zero MW should not factor into an analysis of forecast error. This would have overly skewed the distributions towards lower forecast errors Staff tabulated the actual time of sunset and sunrise in 2012 on the 15th day of each month. Times from San Francisco were used to cover Northern California and times for Los Angeles were used for Southern California. They are tabulated in Table 15.

In addition to elimination of generation for hours that were before sunrise or after sunset each day, staff analyzed the very small MW values immediately after sunrise or before sunset, just to ensure that those patterns were accurate, and that forecast error corresponding to those hours did not overly skew the error distributions. For example, during February in Northern California, the hours from 7 am to 6 pm were considered for inclusion in the analysis, based on the sunrise and sunset times shown. However, a

[42] http://oasis.caiso.com/

[43] The day-ahead solar data exclude the hours from 12 am to 6 am. However, these data are not needed for the analysis because there is little to no solar generation during these hours (due to it still being dark or twilight). Solar generation hours are discussed in further detail in Section VI.H.2.a).

pattern was observed; this range of hours often included many twilight hours after 5 pm that have small generation (<1MW) and 0 MW forecast values. Including error percentages derived from these hours would result in extremely large percentage errors that approached 100% (Table 15). Therefore these hours were not included in the forecast error analysis. To ensure that only actual generation hours would be included in the analysis, all data from 5pm to 6 pm in February were excluded from the analysis, resulting in a data set for the month of February that includes solar generation data from 7 am to 5 pm only. Staff found this pattern in other time periods, and set the cutoff times accordingly.

Table 15 Sunrise and Sunset Times for Northern California (San Francisco)

Month	Sunrise[44]	Sunset	Dataset Start	Dataset End
Jan	7:25	17:01	7:00	17:00
Feb	7:14	17:33	7:00	17:00
Mar	6:40	18:04	8:00	18:00
Apr	6:54	19:33	8:00	20:00
May	6:13	20:01	7:00	20:00
Jun	5:49	20:26	7:00	20:00
Jul	5:52	20:35	7:00	21:00
Aug	6:14	20:18	8:00	20:00
Sep	6:40	19:38	8:00	20:00
Oct	7:06	18:52	8:00	19:00
Nov	7:36	18:10	7:00	18:00
Dec	7:07	16:51	7:00	19:00

[44] Source: 2013 NOAA sunrise and sunset data for the 15th of each month. http://www.esrl.noaa.gov/gmd/grad/solcalc/sunrise.html.

Note: DST started on 3/10/13 and ended on 11/3/13.

Table 16 Sunrise and Sunset Times for Southern California (Los Angeles)

Month	Sunrise	Sunset	Dataset Start	Dataset End
Jan	6:58	16:54	7:00	17:00
Feb	6:50	17:23	7:00	17:00
Mar	6:21	17:50	7:00	18:00
Apr	6:39	19:14	7:00	19:00
May	6:03	19:37	7:00	20:00
Jun	5:42	20:00	7:00	20:00
Jul	5:45	20:08	7:00	20:00
Aug	6:05	19:53	7:00	20:00
Sep	6:27	19:18	7:00	19:00
Oct	6:48	18:36	7:00	19:00
Nov	7:13	17:59	7:00	18:00
Dec	6:41	16:44	7:00	17:00

b) Wind Generation Forecast and Actual Generation

Day-ahead wind forecast values are only available from 6 a.m. to midnight. Therefore, the day-ahead analysis that follows excludes wind forecast error for the hours from midnight to 6 a.m. However, hour-ahead forecast data and actual generation levels are available and were analyzed for all hours. The wind forecast data did not require any processing like the solar data did.

The actual wind generation data, included a significant number of negative values. These accounted for approximately 10% of the collected data. For example, on January 8, 2013 at 9 a.m., the wind hour-ahead forecast was 124.99 MW in the SP15 region, while the actual generation was reported as -6.68 MW. The negative generation did not appear to show any pattern in the distribution of hours or dates, so it was hard to ascertain the reason for it. As in the case of the negative solar generation values, these anomalous, negative generation data are cause for concern. For the moment, the negative generation data and the forecast data corresponding to those hours are excluded from the analysis, but in the future that data can be reintegrated into the forecast error distributions.

c) Load Forecast and Actual Values Data Processing

Day-ahead load forecast values were downloaded from the CAISO OASIS system and analyzed, following a process identical to the solar and wind values discussed earlier. Data was formatted and organized for creation of forecast error percentage distributions. Load values are never near zero or negative, thus there are very few outliers and very little processing was required.

3. Creation of Forecast Uncertainty Distributions

Forecast uncertainty is represented in SERVM as a percent error between forecast and actual generation or load:

$$Percenta \quad E = (F \quad V - A \quad V)/(A \quad V)$$

The above formula can generally be applied to forecasted solar generation and wind generation. An over-forecast (forecast > actual) is represented as positive error, and an under-forecast (forecast < actual) as negative error, assuming positive input values.[45]

The discussion above illustrates a more general feature of the percentage error calculation: as generation approaches either 0 MW or the maximum possible MW, the error distribution changes and becomes skewed. It is not physically possible for a 0 MW forecast to be too high, and it is likewise not physically possible for a forecast of the maximum possible generation to be too low. Moreover, it is also conceivable that certain mid-range generation values are more likely to be under- or over-forecast for other reasons. To account for such differences, SERVM is able to incorporate multiple forecast error distributions, selecting error values from the appropriate distribution depending on the actual generation versus max possible generation level being simulated.

To create multiple forecast error distributions, Staff separated error data according to generation level:

Generation Level = (*H R G*) */(Maximum MW generation value seen in the entire generation profile)*

Staff generated distribution graphs over a wide range of different solar generation levels. Based on the distributions observed for solar generation, 70% was selected as the cut-off point between medium and high solar generation levels. A cut-off point of 4% was selected to distinguish between low and medium solar generation levels, because the lowest quartile (25%) of the data set is comprised of generation values that are below 4% of maximum observed solar generation. SERVM distributions are more continuous, and effectively create distributions for more "buckets" of generation percentage levels; the following charts highlight a portion of the range however, to highlight uncertainty in various ranges of operating levels. For wind generation, the forecast percent error also varies depending on generation level, though not as much as for solar generation. Based on observed distributions, three generation levels, 0-4%, 4-70%, and 70-100%, were selected to represent low, medium, and high generation levels.

4. Use of Distributions in SERVM

To simulate the variability and uncertainty that strains the operator's ability to manage the system, distributions of forecast uncertainty for wind, solar, and load are generated and input into SERVM. Based on the slope and magnitude of hourly wind/solar/load hourly profiles figures, SERVM interpolates the minute by minute variation via an algorithm internal to SERVM that intentionally creates variation. The generating resources that are available to dispatch are incrementally or decrementally dispatched according to their capabilities to ramp.

The uncertainty distributions allow the model to generate uncertainty that compounds the challenges

[45] When actual solar generation is 0 MW, to avoid an undefined error due to dividing by zero, the error is set to 0%; these data points generally represent twilight or nighttime hours when forecasted generation is under 3 MW, and any errors under these conditions are negligible in absolute terms.

presented by variability. SERVM will always dispatch perfectly to load. There is uncertainty about the actual level of load, wind, and solar MW levels, and SERVM dispatches to meet forecasts, then seeks to adjust during the interval. If SERVM is unable to redispatch to actual levels of wind, solar, and load, then firm load is shed. SERVM will log reserve shortages as well as overgeneration due to the variability and uncertainty logic, and report that to the user as a separate class of outage events.

The SERVM model makes use of the inputs to create distributions out of which the model will randomly draw in order to simulate commitment and dispatch by an operator working in a real life electric system. The model will simulate commitment of generation to meet a day ahead forecast of hourly wind, solar, and load values. As opposed to a model which will dispatch day ahead to meet perfectly known MW values, staff intends to run SERVM with uncertainty distributions that will force the dispatch to be incorrect, and remedy with quick adjustments at the last minute. This will stress the system's ability to meet load by ramping generators up or down.

The data supplied to SERVM is drawn from the overall distribution of formatted forecast errors downloaded from the ISO. The user enters a 120 point distribution broken into 10 intervals. Each interval represents a set of percentage error values that apply to the forecasts when load/solar/wind MW levels are at a certain percentage level relative to max. This means that when load/wind/solar MW levels are at a low (small percent of max MW) level, SERVM draws from the set of forecast error percentages that are applicable to the low percent level interval. SERVM uses the intervals as a lookup table, such that when the generation profiles for wind or solar resources are at a certain level relative to the max generation in that profile, SERVM looks up the possible forecast uncertainty errors applicable and appropriate only to that level. For example, it would be inappropriate to apply a certain percentage error appropriate to low levels of generation to a higher level of generation; at a low level of generation, the forecast could be 100% deviated, which 100% deviation at a high level of generation would create a much larger MW error than is realistic.

5. Forecast Error Distributions Entered into SERVM

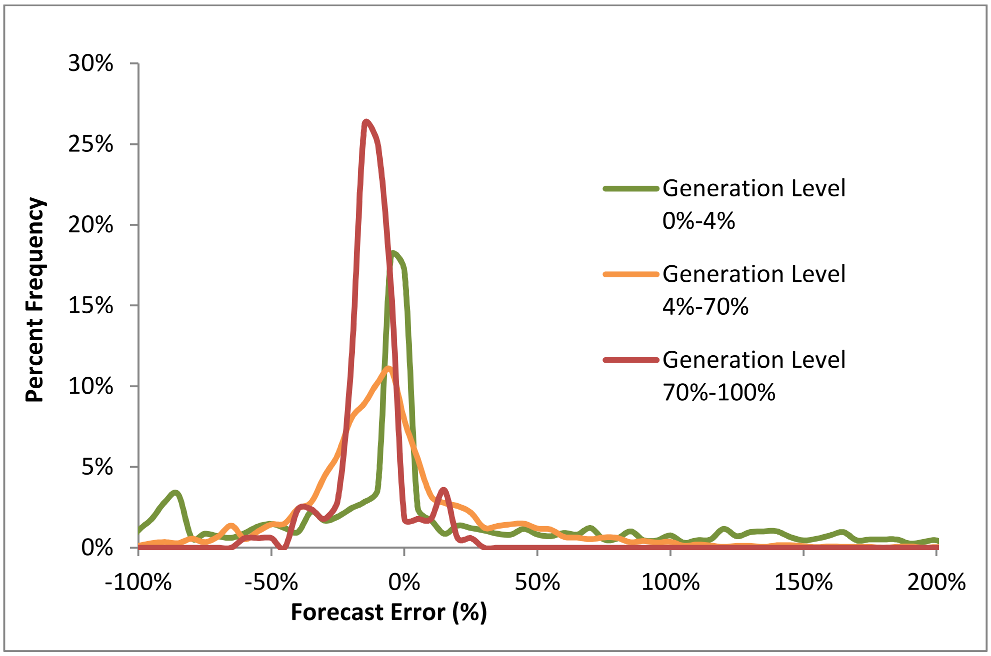
a) Solar Generation Forecast and Actual Generation

Figure 4 suggests that the percent error follows a roughly symmetrical distribution. However, the error at lower generation levels is very strongly skewed toward overforecasting, with a long tail of high positive forecast errors. This makes intuitive sense; it is more likely that 1 MW of actual generation would have a 100% forecast error (2 MW forecasted, 1 MW absolute error) than that 1000 MW of actual generation would have a 100% forecast error (2000 MW forecasted, 1000 MW absolute error). Therefore, the smaller the generation level, the higher the percent error.

On the negative side, because generation values are zero or positive only, -100% is the lower limit for the distribution. The negative bound of -100% occurs when 0 MW are forecasted but actual generation is above zero, which generally corresponds to times when actual generation values are also quite small.

Additionally, the overall distribution is shifted to the left, which means that 55% of forecast percentage errors are negative (i.e., under-forecast errors). Also, compared to the distribution with generation level from 4%-70%, the distribution with generation level from 70%-100% is more shifted to the left, which means that under-forecasting is more common at higher generation levels. This also makes intuitive sense. Operators will never forecast generation at 110% of maximum possible output, as that would be physically impossible, so generation at near-100% levels is likely to be either correctly forecast or under-forecasted, but only rarely (and even then only slightly) over-forecasted.

Figure 4 Northern CA Hour-Ahead Solar Forecast Errors Observed



The range of percent error was from -100% to over 500,000%, prior to the exclusion of outliers. However, the vast majority of over-forecast errors are within 200%, and it is likely that any values beyond that are erroneous (or that the generation levels are so small that the absolute magnitude of the error is negligible). To ensure that only reasonable, non-erroneous data are utilized in SERVM modeling, all percent error values beyond the -100% to +200% range are excluded from this analysis.

b) Wind Generation Forecast and Actual Generation

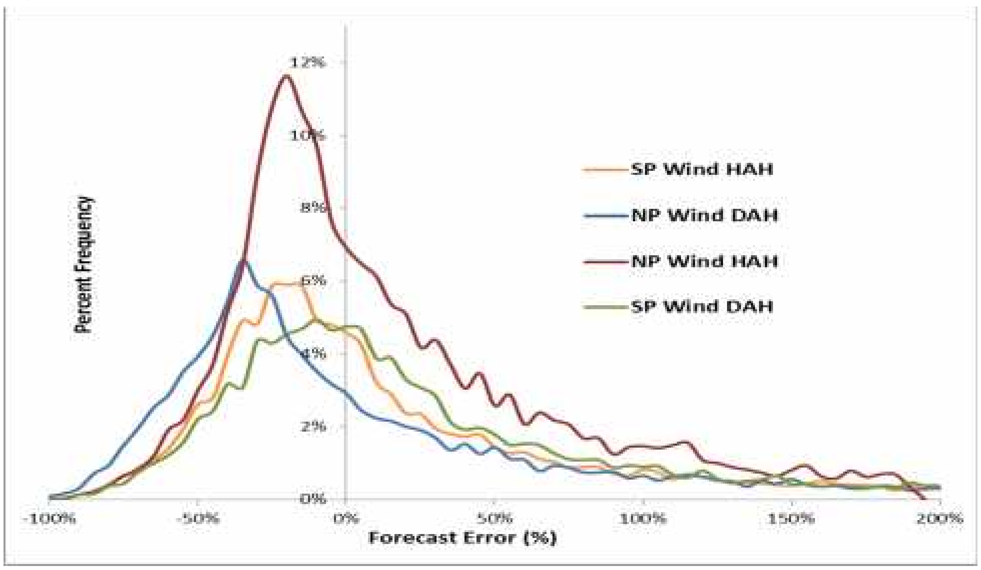
Figure 5 suggests that the distribution of percent error follows a roughly symmetrical distribution similar to solar generation. Similar to solar forecast error, the overall distribution is right-skewed, which means the majority (over 50%) of forecast percent error are under 0% (i.e. under-forecasting). The range of percent error is from -100% to 600,000%.[46] As with solar generation, the -100% errors are caused by near zero MW forecasts and non-zero actual generation, while extreme positive percent errors are

[46] Excludes potentially erroneous negative generation values, as previously described although more analysis of negative generation levels is warranted.

caused by over-forecasting at times of very small MW generation levels. Again, the extreme values beyond the -100% to +200% range are excluded from this analysis and are not imported into SERVM.

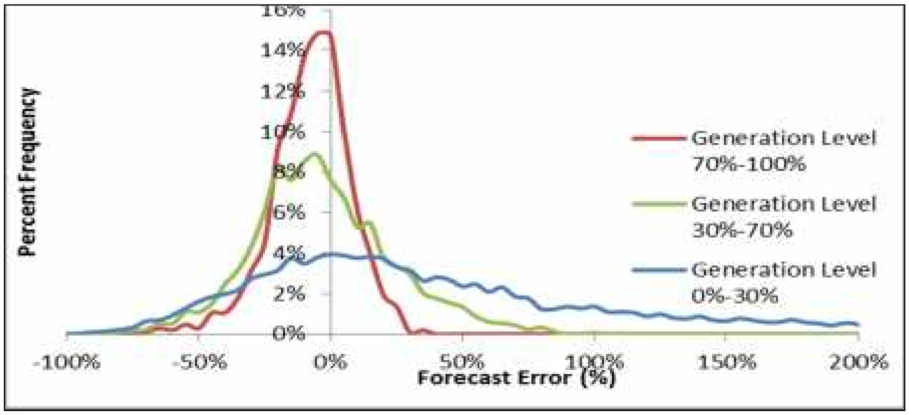
In geographic comparison, it is observable from Figure 5 that wind generation in Northern California is more often under-forecast than in Southern California since the distribution in the north is shifted farther to the left. This might be due to the fact that the Northern California wind resources analyzed for this study were more abundant although that is not currently the case; Northern California wind annual generation is around 7,100,000 MWh while annual wind generation in Southern California is only around 5,400,000 MWh. Additionally, when comparing day-ahead forecasting with hour-ahead forecasting, the hour-ahead forecasting percent error graphs are usually more stable. Finally, the generation level as a percent of capmax has an impact on the forecast percent error. As shown in Figure 6, the higher the generation level, the smaller the standard deviation, and correspondingly more accurate forecast value.

Figure 5 Wind Hour Ahead Generation Forecast Error Distribution (0%≤Generation Level≤100%)



Percentage magnitude of forecast error varies notably depending on the level of generation (percent of max) being forecasted. SERVM handles this significant variation by developing a setoff separate distributions, each specific to a range of output for a particular set of facilities. Figure 6 illustrates as an example the different ranges of generation for wind facilities in Southern California and the very different shapes of forecast uncertainty distributions at different generation levels.

Figure 6 Southern CA Wind Hour Ahead Generation Forecast Error Distribution



These results represent a good starting place, but more work is needed in the area of forecast uncertainty distributions in order to develop a more accurate study. Uncertainty inputs are critical to staff study of intra hour intervals and operational flexibility.

- **More analysis of negative generation values**

ED staff is confused by the proliferation of negative actual generation values in both wind and solar data. While it was relatively easy to relegate the negative solar generation data to error, given it was predominant during the night time, it was not as easy to understand the dynamics of wind generation data. It is possible that the negative values can be attributed to station service, when the facilities take more energy to operate than they produce. Greater understanding of the issues would help staff to create more accurate distributions with a wider set of data.

- **2014 Data and Data outside of California**

Data from CAISO has been relatively easy to collect. 2014 data has already been downloaded, and a next step is to add it to the forecast error distributions. Data for generation outside of the CAISO, or even outside of California, might also be available, and it would help to create more realistic dispatch scenarios for production cost simulations. ED staff has not found a good source for it yet. Now that the Energy Imbalance Market (EIM) is in place, it might be possible to cooperate more fully with the EIM partners to access further market data.

VII. Transmission Inputs

SERVM uses a transportation representation of the transmission system instead of an AC or DC representation. For a given region and a given connected region, the capacity limits in and out of the region (with respect to the connected region) are specified. These limits can vary by study year, by month, and by percent of peak load. Currently, these limitations are static, reflective of the values currently used by the CAISO for the modeling CAISO is doing to help inform the CPUC in the LTPP proceeding, otherwise referred to as the “PLEXOS dataset”. The values are included below.

Table 17 Transfer Limitations used for SERVM

Region	Connected Region	MW Limit In	MW Limit Out
PGE_Bay	TID	174.4	174.4
PGE_Bay	PGE_Valley	15000	15000
PGE_Valley	SCE	4000	3000
PGE_Valley	SMUD	15231.7	15231.7
PGE_Valley	TID	1465.1	1465.1
PGE_Valley	Nevada	150	160
PGE_Valley	Pacific Northwest	3720	4900
SCE	SDGE	4922.8	2500
SCE	Arizona	6000	7341.7
SCE	Nevada	1540.6	1540.6
SCE	IID	1500	1500
SCE	LADWP	3656	2156
SDGE	Arizona	2721.3	2721.3
SDGE	Mexico	800	408
SDGE	IID	702	702
IID	Arizona	478	478
LADWP	Arizona	11317	4206
LADWP	Nevada	3500	3050
LADWP	Pacific Northwest	3220	3100
LADWP	Utah	2600	2600
SMUD	TID	4664.2	4664.2
Arizona	New Mexico	7221.9	7221.9
Arizona	Utah	825	900
Canada	Montana	750	750
Canada	Pacific Northwest	3500	4350
Colorado	Arizona	485	485
Colorado	Montana	390.4	390.4
Nevada	Arizona	7899	7899
Nevada	Utah	1358.5	820.5
New Mexico	Canada	400	400
New Mexico	Colorado	1212	1212
New Mexico	Montana	290.8	290.8
New Mexico	Pacific Northwest	5685.4	5685.4
New Mexico	Utah	904	904
Pacific Northwest	Montana	2200	1350
Pacific Northwest	Nevada	622.6	622.6
Pacific Northwest	Utah	4364.8	3464.8
Utah	Colorado	1454	1454

Utah	Montana	337	337
Utah	Wyoming	4971.9	5512.4
Wyoming	Colorado	6875.5	6875.5
Wyoming	Montana	600	600

VIII. System Inputs

A. *System Periods*

SERVM allows for outputs to be available in specific periods of the day or week but not others. DR programs are given specific system periods when they are available. The system periods are defined according to the days of the week and hours of each day that are assigned to each period. Parties are invited to comment on these periods and suggest additional or alternative periods.

System Period	Day	Hours (*Hour Ending*, or HE)
OffPeak	Friday	23-24
	Saturday-Sunday	1-24
	Monday	1-6
Weekday	Monday-Friday	7-22
WeekdayOffPeak	Monday-Thursday	23-24
	Tuesday-Friday	1-6

B. *Ancillary Services*

Ancillary service requirements and incremental targets are input as a percentage of load, and are assumed to be consistent across all regions, months of the year, and hours of the day. Parties are invited to suggest alternative or more differentiated ancillary service requirements, along with documentation. Current staff assumptions are shown in Table 18, below.

Table 18. Ancillary Service Requirements and Targets

Ancillary Service	Requirement or Target	Value (Percent of Load)
Regulation Up	Requirement	1.5%
Regulation Down	Requirement	1.5%
Spinning Reserves	Requirement	1.5%
	Incremental Target	2.5%
Non-Spinning (Quickstart) Reserves	Requirement	3%
	Incremental Target	0%

C. *Operating Reserve Demand Curves (Scarcity Pricing)*

Regulation up, regulation down, spin, and non-spin scarcity prices are input into SERVM, specified according to the applicable remaining hourly reserve margin percentage. While values can vary by region, month, and hour, staff is not currently utilizing this feature. Data for reserve demand curves are in development.

IX. Next Steps

ED staff will continue to refine this paper, post additional data sets, and work to update the data that is in the SERVM database. Several parts of the database stand out as being in need of updating. In addition, the RA proceeding will continue to explore ELCC and make determinations as warranted. Parties are encouraged to contact ED staff with suggestions of better data sources, help in developing and formatting data, or checking for errors in datasets.