# Probabilistic Reliability Modeling Inputs and Assumptions

**DRAFT STAFF RECOMMENDATIONS , PART ONE**
**RESOURCE ADEQUACY PROCEEDING R.11-10-023**
**CALIFORNIA PUBLIC UTILITIES COMMISSION – ENERGY DIVISION**

November 25, 2013

# Contents

# 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.11-10-023). 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 derating factor that is applied to a facility's maximum output in order to determine its QC. These subjects will be addressed in detail in a forthcoming, formal proposal. This document, however, focuses exclusively on the system reliability modeling that is a prerequisite to calculating a resource's ELCC and QC, and does not address the ELCC or QC algorithms. In particular, it lays out some of the inputs and assumptions that Staff proposes to use when populating the reliability model.

ED Staff plans to issue approximately three such papers documenting modeling inputs and assumptions, of which this paper is the first. Staff will also conduct workshops to discuss each paper with parties and receive informal feedback. Informal party comments on the proposed modeling inputs and assumptions will help staff in the development of a formal ELCC proposal, but will not become part of the rulemaking's record. Official party comments on the formal ELCC proposal and associated workshop will be solicited later and will become part of the rulemaking's record for Commission decision(s) on the use of probabilistic modeling.

This paper includes the following key components:

- Key data sources, coordination with other modeling initiatives
- SERVM (Strategic Energy Risk Valuation Model), software which is being developed for ED Staff use in probabilistic reliability modeling
- Weather data sources and weather region definitions
- Proposed methodology for modeling generation from wind and solar resources based on region, weather, and technology type
- Methodology for synthetic load shape creation

[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.

- Information on conventional, demand response, and storage resource inputs and modeled use limitations
- Natural gas price forecasts

These assumptions will be discussed further with parties in a workshop to be held on November 26, 2013. Informal comments are requested to be sent to Donald.Brooks@cpuc.ca.gov by December20, 2013. Additional modeling inputs and assumptions will be covered in future papers and workshops.

# Energy Division Reliability Modeling 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, however, addresses selected inputs and assumptions used in SERVM. Description of the ELCC calculation will be provided in a future paper.

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

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

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 public and are posted on the CAISO website.[3]

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 2012 and 2022, and is in the process of developing a dataset for 2014 and 2024. The Common Case dataset is publicly available, and can be downloaded from the WECC website.[4]

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 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

[3] 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.

[4] WECC TEPPC 2022 Common Case datasets are available for download here: http://www.wecc.biz/committees/BOD/TEPPC/Pages/TAS_Datasets.aspx.

approach. For near term modeling (such as for determining wind and solar ELCC values for the 2015 RA Compliance Year), ED staff proposes to use the TEPPC 2022 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 CAISO MasterFile and the TEPCC Common Case 2022 datasets, 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.

# Weather Dataand 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.

## *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 reliability modeling utilizes the NOAA Integrated Surface Data - Lite (ISD-Lite, DS3505) data product for the 30 year period from 1981 to 2010.[5] 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

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

Weather station metadata (latitude, longitude, and elevation) are sourced from the U.S. Automated Surface Observing System (ASOS) Station Listing.[6]

Solar irradiance data is supplemented by the National Solar Radiation Database (NSRDB), which also covers the period from 1981 to 2010.[7] To account for differences in wind capacity factors at different heights, NREL wind resource data may also be incorporated into the analysis.[8]

## *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, 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 |
|---|---|
| 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 |
| | Wyoming |

**Source: CEC staff work, based on TEPPC 2022 Common Case**

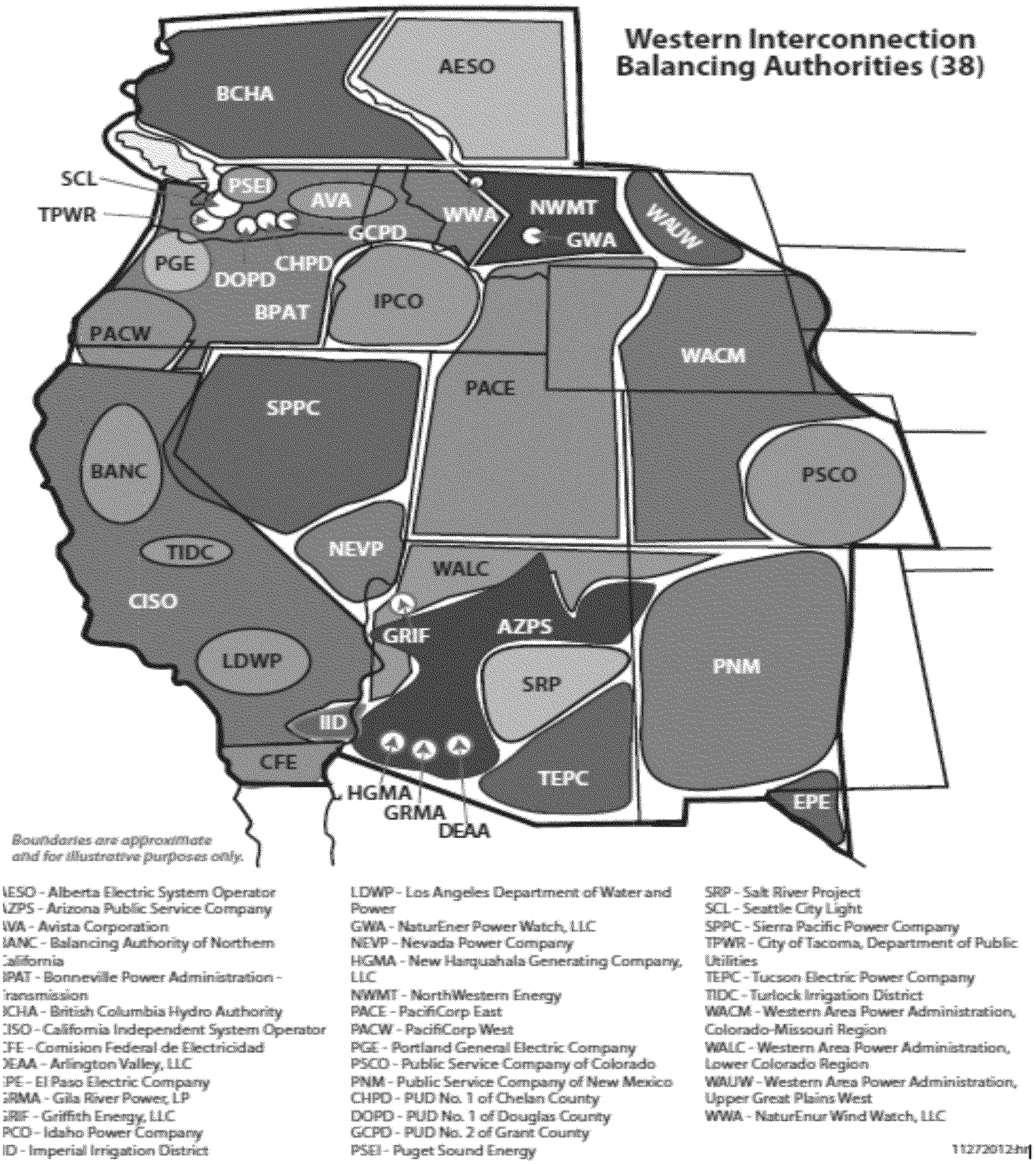
The above regions were developed by the CEC, through aggregation of several of the Local Areas within the CAISO and of the balancing authorities outside of the CAISO. The balancing authorities outside of the CAISO are shown in Figure 1, below.

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

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

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

**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

### *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.

## 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.

### *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.[9] 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[10] | 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 | | |
| Mexico | San Diego | 1 | | | | | | |
| 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 | | | | |
| Wyoming | Cheyenne | .33 | Denver | .33 | Colorado Springs | .33 | | |

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

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

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

## *Processing*

### 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. This is done to adjust a year of load from the past to represent a shape indicative of what load would have been in a future year, adjusting the past for load growth that has happened in the meantime. Each of the 8760 hours in the year is adjusted by the same percentage factor. 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. In essence, staff removed whatever impact demand response had on the historical load shapes, in order to study the unmitigated load shapes. 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.

## 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.

## *Neural Network Modeling*

The training data are then entered into a proprietary neural network model developed by Ward Systems called the NeuroShell predictor model.[11] The neural net model "trains" a network file to identify the 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 load from the five years of temperature and load files, the thirty 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 load years are then entered into the SERVM database for reliability modeling.

---

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

### 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.

# Resource Inputs and Use Limitations

## *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 3. 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)ydro** | Hydro 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 4, 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 4. 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 2022 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 2022 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 2022 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 2022 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 2022 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.

## *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 based on the PLEXOS 2022 data) 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.

### 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 2022 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 2022 Common Case, which is the most current and complete source of information for all balancing authorities within WECC. The 2022 dataset also includes incremental development in between 2012 and 2022.

## Ramp Rates

ED staff proposes to source ramp rates from the TEPPC 2022 Common Case set for generators outside of CAISO and to use the MasterFile ramp rate segments to calculate the MW weighted average ramp rate (maximum ramp rate for each segment MW weighted by the number of MW that apply to that segment) for generators inside of CAISO. Sensitivity analyses will also be conducted to gauge the effect of using ramp rates at minimum load or ramp rates at maximum load. 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. Sensitivity analysis can be done to see if higher or lower ramp rates produce significantly different results, and if the use of MasterFile data produces significantly different results than the use of TEPPC data.

For initial modeling, ED staff will provide links to the public TEPPC database, but will also utilize MasterFile information for generators within CAISO. Since SERVM utilizes one ramp rate value for each generator (although that variable can vary by month or year), ED staff will calculate maximum, minimum and MW weighted average ramp rates based on each generator's unit specific information in the CAISO MasterFile, and input those values into SERVM for modeling and sensitivity analysis.

## 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 5 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 5. 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 analysis of CAISO SLIC data, and will generate these statistics based on 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. Although the outage data (Generator Availability Data Systems or GADS data) is commonly used for this purpose in other states and jurisdictions, that is not a reasonable option for modeling in California. After analysis by ED staff and attempts to format the GADS data into inputs usable in SERVM, the data for California was found to be incomplete. A different option is required.

In modeling to support the Long Term Procurement Plan (LTPP) during 2012, the CAISO generated outage statistics based on its internal outage logging system. The CAISO does not make use of GADS data in its modeling. Instead, the CAISO uses data it gathers from generators via the Scheduling and Logging Interface for California (SLIC) database to generate class average summary statistics. While having the advantage of being public, class average values fail to meaningfully differentiate between generators that in reality perform quite differently. There is in fact quite significant variance in the actual performance of individual generators in California. 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.

While the class averages developed by CAISO staff do not have this level of granularity, they provide a starting point, and do allow for publishing of those averages for stakeholders. Alongside that effort, ED staff will work with CAISO to generate unit-specific indices similar to that produced by GADS data to assess the variance and sensitivity of modeling results to the more accurate generator outage indices. It is also possible that effort can be undertaken to refine and manage GADS data so as to make the data usable for this purpose. Due to the current limitations of GADS data, ED staff proposes to apply the CAISO class averages to generators inside CAISO and to process that data into SERVM inputs. The TEPPC Common Case includes similar class averages for generators external to CAISO, and those averages will also be processed into SERVM inputs.

## 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). Table 6 below provides the burnertip prices utilized in the model, including the natural gas hub and transportation prices for 2015, in nominal dollars. The CEC will finalize its forecast in December 2013; to the extent that the forecast changes, staff will update the SERVM model at that time.

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. To address this concern, staff may run natural gas price sensitivity scenarios to determine the effect on base case modeling results.

**Table 6. Burnertip prices utilized in SERVM**

| 2015 Location | Hub Price | Transportation Price | Burnertip Price |
|---|---|---|---|
| N. AZ | 3.96 | 0.34 | 4.30 |
| S. AZ | 4.06 | 0.48 | 4.54 |
| CO | 3.89 | 0.13 | 4.02 |
| Idaho (Kingsgate) | 3.86 | 0.09 | 3.95 |
| Kingsgate (Sumas) | 3.85 | 0.18 | 4.03 |
| Montana | 3.81 | 0.13 | 3.94 |
| N. NV | 4.19 | 0.30 | 4.49 |
| S. NV | 4.36 | 0.37 | 4.73 |
| N. NM | 3.78 | 0.36 | 4.14 |
| S. NM | 3.86 | 0.36 | 4.22 |
| Oregon | 4.09 | 0.00 | 4.09 |
| Malin | 4.06 | 0.00 | 4.06 |
| Utah | 3.86 | 0.00 | 3.86 |

| | | | |
|---|---|---|---|
| WA – Seattle | 4.07 | 0.27 | 4.34 |
| WA - Kingsgate | 3.92 | 0.00 | 3.92 |
| WY | 3.89 | 0.13 | 4.02 |
| PG&E Backbone | 4.43 | 0.13 | 4.56 |
| PG&E Local TX | 4.43 | 0.38 | 4.81 |
| SMUD < 85 MMcfd | 4.43 | 0.13 | 4.56 |
| SMUD > 85 MMCfd | 4.43 | 0.13 | 4.56 |
| Kern River | 4.33 | 0.00 | 4.33 |
| Mojave | 4.33 | 0.00 | 4.33 |
| Coolwater | 4.33 | 0.00 | 4.33 |
| SoCalGas | 4.77 | 0.31 | 5.08 |
| Blythe | 4.33 | 0.00 | 4.33 |
| SoCal Production | 4.45 | 0.00 | 4.45 |
| SDG&E | 4.91 | 0.30 | 5.21 |
| Otay Mesa | 4.91 | 0.30 | 5.21 |
| Alberta | 3.81 | 0.11 | 3.92 |
| British Columbia | 3.85 | 0.11 | 3.96 |
| Rosarito | 4.57 | 0.00 | 4.57 |

**Source: CPUC staff analysis of CEC NAMGas report**

## *Renewable Resources – Type R*

The major distinction for purposes of SERVM between Type R resources and other types (such as F, T, or N) is in how resources are dispatched in the model. 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, solar photovoltaic (PV), solar thermal, geothermal, biomass, and biogas generation facilities. However, there are also renewable facilities (such as biogas, biomass, or geothermal facilities) that are modeled economically via production cost dispatch; thus, the term "renewable" is really shorthand for weather-dependent must-take resources. Renewable facilities that are going to be modeled with prices and startup costs will be modeled as Type F units. The two facility type options enable SERVM to model diverse resources accurately, considering weather patterns and/or energy prices as appropriate.

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. Solar thermal generation is not addressed in this document, and will be covered in future staff publications.

### 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 in this document at this time; the methodology for modeling solar thermal facilities will instead be addressed in a future document.

To accurately reflect the variability in wind and solar photovoltaic generation, modeling of solar and wind facilities requires mapping of the multiple years of historical weather information to the power output of existing and new facilities utilizing various technology types. This mapping will create hourly performance profiles for each year of weather data, representing the overall variability of wind and solar production related to weather. Staff is currently considering multiple approaches to the development of these performance profiles.

One approach is to utilize generation profiles created by key stakeholders who are already conducting similar facility performance modeling. For example, manufacturers need to predict the potential generation profiles of their facilities in order to predict potential energy revenues and inform bids into RFOs or energy markets. Thus they would be helpful in developing potential production profiles. Utilities also have an interest in predicting potential generation for resources that they are considering for contracting or are attempting to operate and manage. Both manufacturers 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 at least two sources of wind and solar generation profiles that could be used, in addition to any information that state planning agencies use for study purposes.

However, there could be drawbacks to utilizing manufacturer 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 be predicting performance of generation related to other factors, and the data they develop may be in a different format or based on different weather projections that cannot be extrapolated to the entire 33 years of weather history. Data for performance of wind and solar facilities external to California may be much more difficult to access, complicated by different utility service areas and information access guidelines.

Alternatively, standard, publicly available weather information can be mapped to the power output of wind and solar facilities using off-the-shelf neural network modeling software discussed in earlier sections that has already been used for development of load shapes. The neural network would determine the relationships between the input variables and facility production, and output a predictor file. With this predictor file, synthetic wind and solar production profiles could be constructed for existing and new facilities to correspond to the 33 years of weather history and associated synthetic load shapes utilized by SERVM in modeling future years. The large number of years will enable SERVM to capture realistic variability in generation from wind and solar facilities.

It is expected that the synthetic production profiles (and the predictor file, if the neural network approach is adopted) will be reconstructed at least every two years to reflect the evolving relationships between weather and production (considering such issues as technology improvement and locational clustering of installed capacity). Intra-hour variation and forecasting uncertainty will be addressed in a future paper, which will outline the required data and sources, and explain how SERVM models forecast uncertainty.

If the neural network approach is adopted, ED staff will need to develop a model and predictor file, which will require extensive performance data and technology information. In either case, it will be necessary to collect relevant weather data. Focusing on the neural network approach, 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 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 may require significant effort, it will also pay off in greater accuracy.

### Performance Data Sources

ED staff has received 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 development of QC calculations. CPUC staff retains this data, and has used it to develop some performance profiles that can be applied. ED staff can utilize settlement data from 2008 to 2012, although each year there is a large amount of capacity coming online, so each year of data will need to be normalized to installed MW. ED staff has only begun to gather data for areas external to CAISO.

## Weather and Regions Modeled

Weather information is sourced from NOAA and NREL; specific inputs include temperature, wind speed, irradiance, and cloud cover. These inputs are formatted for neural net "training" as described previously.

Because weather data is available at limited locations, and because modeling time increases dramatically as granularity increases, one weather profile is compiled 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 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 its RPS Compliance Report. Alternative approaches to weather station weightings may be considered if SERVM is utilized for longer-term modeling; sensitivity to weather station selection will also be tested.

In developing technology and weather response relationships in the neural network software, the representative regional weather will be input. Later, 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.

## Technology Categories

Wind and solar resources will be grouped into approximately four to six technology categories that output at similar levels under identical weather conditions (as a percentage of their maximum output), with possible examples shown in Table 7, below. These categories will be developed based on the correlations observed. Each technology category and region will be considered separately by the neural network model to develop weather response predictions within that region, across that category type. SERVM will model each facility's generation based on both its technology category (indicative of response to weather) and weather region (the relevant weather input). Technology category and location are both sourced from a facility's RPS compliance report.

**Table 7. Possible Technology Categories for Wind and Solar Generation**

| Wind | Solar |
|---|---|
| Above/Below 80 Meters | Solar Thermal (with/without storage)[12] |
| Above/Below 50 Meters | Rooftop (residential/larger scale) |
| Older/Newer Vintage | Fixed Tilt (over/under 20 MW) |
| Utility Scale/Distributed | Tracking |

***Category Normalization***

In order to compare across facilities of varying sizes, output will have to be normalized relative to facility size prior to neural net "training". Additionally, because the neural network will develop 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 of the larger total capacity in future years. However, this may be imprecise due to differences in the generation profile shape and volatility as more capacity is installed in new locations. Alternatively, new facilities can be "trained" by assigning production profiles from existing facilities, chronologically matching the weather years being "trained" to maintain the correlation. New facilities are assigned an hourly production profile from an existing facility determined to be "similar" in location and technology, and scaled to the MW size of the facility being modeled.

## Neural Network Modeling

Because weather and resource generation data are available for the same years, 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 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.

First, the regional weather data is placed into a spreadsheet for a given technology category. 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 will be wind speed. In the case of solar technologies this will be irradiance. The other key weather inputs (discussed previously) will be included as additional columns. This data is paired with actual facility and generation data from facilities of the given technology type. The data may also be separated based on seasonal patterns, if deemed necessary.

[12] As previously mentioned, solar thermal generation is not being addressed in this document, and will be modeled using a different methodology.

Ten percent of the data will be excluded from the neural network prediction modeling in order to test the predictor relative to actual weather and generation data. The remaining data will be used to "train" the NeuroShell predictor model, a proprietary neural network model developed by Ward systems.

This neural network model will train 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 will develop predictive relationships between the columns of data (the variables mentioned previously such as temperature and cloud cover), and produces an algorithm that predicts relationships between regional wind speed or irradiance, secondary weather variables, and generation facility output. The model will then be able to produce a generation forecast from any set of weather data, for any facility that falls under the given technology category (including new facilities). To validate the tool's predictions, the data excluded from the predictor training will be input, and the resulting predictions compared to actual historical weather and generation. The resulting $R^2$ value will be calculated and reported. This process will be repeated for each combination of technology category and region, as each has a different relationship to weather inputs.

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 will likely be some adjustment to the distribution of wind predictions post processing. Additionally, the randomness and volatility will obscure some of the correlation that wind output has between regions, which will have to be accounted for and re-introduced in the production profile development post processing. This is less of a consideration for solar resources, but staff will monitor neural network output to verify reasonableness.

Staff will perform validation on the resulting performance shapes to ensure accuracy, by comparing the resulting shapes to historical patterns, as well as by sharing non-confidential aggregate sharing the resulting shapes with stakeholders. Staff will also test the predictor file created by the neural net modeling by "predicting" generation for historical weather that has been purposefully withheld from the historical information fed into the neural net modeling, to see how accurately the predictor file predicts generation and weather relationships. 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.

### Weather and Technology Response Curves in SERVM

Once the predictor file has been validated, its output can be imported into SERVM for use in reliability modeling. Because facility output depends on multiple factors, not simply wind speed or irradiance, a blended "weather input" curve will be created by the neural network predictor for each technology category in a given weather region. This weather input will correspond to the actual hourly weather data for the weather year being modeled. However, it will not consist of physical properties such as wind speed, but rather will indicate the percentage of maximum capacity output that can be expected for facilities of a given technology category exposed to a given region's weather, for each hour of the year.

The weather input for a given weather year and region will be imported into SERVM as a single spreadsheet column with 8760 rows, where each row represents a single hour for the given weather year and region. The contents of each cell will be a percentage that indicates the expected output from any facility in the technology category, relative to its maximum. During modeling, each facility's available capacity will then be adjusted for each hour of the year by these percentages.

## *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).[13] Key inputs currently incorporated into the model are listed in Table 8, below.

**Table 8. Current Demand Response Resource-Specific Inputs**

| Input (as applicable to the program)[14] | Units | Source |
|---|---|---|
| **Maximum capacity** | MW | LIR portfolio-adjusted monthly system peak values for 1-in-2 weather conditions[15] |
| **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 |
| **First month available each year** | month | Program tariff |
| **Last month available each year** | month | Program tariff |

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

[14] 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.

[15] 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.

| | | |
|---|---|---|
| **Availability window (i.e., weekdays from 2-6 pm)** | days and hours | Program tariff |
| **Dispatch price** | $/MWh | CAISO Plexos assumptions or program tariff[16] |
| **Emergency-only dispatch** | Yes/No | Program tariff |
| **Region**[17] | Region name | Program tariff |
| **Program in-service dates** | mm/dd/yyyy – mm/dd/yyyy | Program tariff |

## Potential Future Expansions/Changes to DR Inputs

To address the complexity of DR programs, several changes to the above inputs are under consideration. They are summarized in Table 9, below.

**Table 9. Possible Future Demand Response Resource-Specific Inputs**

| Demand Response Resource Input[18] | Units |
|---|---|
| **Notification period** | Option of either day-ahead (10am) hour-ahead, # minutes' notice, etc. |
| **Max number of dispatches per day** | dispatches |
| **Max number of dispatches per month** | dispatches |
| **Max number of dispatches per year** | dispatches |
| **Min number of dispatches per year** | dispatches |
| **Customer Fatigue** | % degradation in response per consecutive hour or day of dispatch |
| **Look-Ahead**<br>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. | hours |
| **Ramp Rate** | MW/min |

[16] 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.

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

[18] These inputs would mostly be optional, because different programs have different constraint types.

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

### 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.

### 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 10, below.

**Table 10. 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.**

### Customer Fatigue

The SERVM model does 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.

### 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.

### Ramp Rate

Ramp rates are currently not considered for DR resources in SERVM. This functionality could be added in the future.

## *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.

### Current Inputs

| Input | Units | Source |
|---|---|---|
| **Maximum rated discharge** | MW | Testing submitted to the CAISO. Test should go from fully charged to minimum allowable charge. Duration should also be reported. |
| **Total usable storage volume (given allowable depth of discharge)** | MWh | Calculated based on testing: Maximum rated discharge * (discharge test duration) |
| **Maximum rated charge** | MW | Testing submitted to the CAISO. Test should go from minimum allowable charge to fully charged. Duration should also be reported. |
| **Round trip 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 |
| **Real-time price at which storage is dispatched** | $/MWh | PPA variable O&M terms plus appropriate ROI value |

### 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 |

## Next Steps

ED staff will continue to revise, augment, and post future versions of this paper, adding sections on other modeling assumptions and datasets as they are developed. ED staff will also convene further workshops in December and January to inform stakeholders as to the progression of the modeling effort. ED staff also intends to develop a page on the CPUC website to house the public datasets that are permitted to be posted. Stakeholders are encouraged to contact ED staff with further questions and ideas as to modeling issues, and in sending questions are encouraged to suggest additions or augmentations to this report.