2014

Study of 2012 LTPP Base Scenario for CAISO System in 2022

Prepared for PG&E

Prepared by
Astrape Consulting
Kevin Carden
Nick Wintermantel

4/11/2014

Contents

Acknowledgements

This report was prepared for PG&E as part of a collaborative effort to review planning models. All results and any errors are the responsibility of the authors. Opinions expressed in this report, as well as any errors or omissions, are the authors' alone.
Nothing herein is intended to provide a legal opinion. The authors would like to thank the PG&E staff for their input, cooperation, and support. In particular, we would like to acknowledge the analytical, technical, and conceptual contributions of Antonio Alvarez and Larsen Plano of PG&E.

Executive Summary

This study was performed by Astrape Consulting at the request of Pacific Gas & Electric Company (PG&E) as part of a larger collaboration effort to review various planning models available to study the system reliability within the California Independent System Operator (CAISO) to meet the 33% RPS. This study was conducted using assumptions consistent with the Base Scenario (without San Onofre Nuclear Generating Station) used in the California Public Utilities Commission's (CPUC) 2012 Long-term Procurement Plan (LTPP) proceeding for 2022. Both the ability of the system to meet peak load as well as the ability to ramp intra-day and intra-hour to accommodate wind and solar generation were assessed.

Analytical Approach

The study was performed using the Strategic Energy and Risk Valuation Model (SERVM)[1]. Astrape Consulting has taken a stochastic approach in modeling the uncertainty of weather impacts on load and renewable generation, economic growth, unit availability, and unit commitment. Two separate analyses were conducted. The first analysis uses only the single load, solar, and wind shapes from the 2012 LTPP scenario assumptions. The second analysis uses 32 load, wind, and solar shapes representing 32 years of historical weather. The second analysis which utilizes multiple weather years performed 16,000 yearly simulations for the 2022 study year at 5 min intervals. For both analyses, Loss of load Expectation and Expected Unserved Energy due to generic capacity shortfalls ($LOLE_{GEN}$, EUE_{GEN}) was reported as well as Loss of Load Expectation and Expected Unserved Energy due to system flexibility deficiencies ($LOLE_{FLEX}$,EUE_{FLEX}) LOLE and EUE due to generic capacity shortfalls ($LOLE_{GEN}$, EUE_{GEN}) are calculated ignoring the flexibility constraints of resources such as start times and ramp rates, while $LOLE_{FLEX}$,EUE_{FLEX} are the difference between the total LOLE/EUE calculated considering these

[1] Astrape Consulting is the developer and exclusive licensor of the SERVM Model. SERVM is a chronological hourly and intra-hour unit commitment and dispatch model that is used to analyze system reliability and production cost analysis.

flexibility constraints less $LOLE_{GEN}$, $EUE_{GEN.}$ Renewable curtailment was also reported for every simulation. The following address the largest drivers of uncertainty.

1. Weather uncertainty: The first analysis used the single load, wind, and solar shapes scenario used in the 2012 LTPP. The annual load profiles are scaled up and down by load multipliers, which capture the long term weather uncertainty and economic growth uncertainty. The multipliers are applied uniformly across all hours of the annual profile.

   The multi weather shape analysis represented the uncertainty in weather by simulating 32 different shapes based on the last 32 years of historical weather for load shapes, wind shapes, and solar shapes. The 32 shapes were developed by Astrape Consulting as part of the current ELCC study being performed by the CPUC.[2] Each shape was given equal probability of occurrence.

2. Economic Growth Uncertainty: In the single shape analysis the annual load profiles are scaled up and down by load multipliers, which capture the long term weather uncertainty and economic growth uncertainty over a 4 year time horizon. The multipliers are applied uniformly across all hours of the annual profile to scale the load shape up and down.

   For the multi weather shape analysis, each of the 32 load shapes were scaled up and down with multipliers that only reflected the four year ahead load growth uncertainty since weather uncertainty is represented in the separate load shapes. The loads in all hours are scaled by the same multiplier. Each multiplier is given a distinct probability of occurrence and each load shape was assumed to have equal probability in the study. A sensitivity was performed with the economic growth uncertainty component excluded.

3. Unit availability uncertainty for conventional generation is modeled using a Monte Carlo approach which capture frequency and duration parameters. The model is provided time to fail

[2] Documentation of the process and shapes will be made public by the CPUC once finalized. Further discussion regarding methodology is included in Input Assumption Section of this report.

and time to repair distributions for each unit as well as an availability percentage for initialization purposes. At the start of the year, SERVM draws from an availability distribution for each unit to determine whether the resource is available at the start of the year. If the unit is available, SERVM then randomly draws from the time to fail distribution to determine how long the resource can run before it is forced out. Once the unit is online for its time to fail draw, then SERVM draws from the time to repair distribution to determine how long the resource will be unavailable. This continues until the entire year is simulated. Typically, SERVM is used to model more detailed outage events such as partial outages, start up failures, and maintenance outages. However, due to the inputs included in the 2012 LTPP dataset, only full forced outages and planned outages were modeled. Planned maintenance is based on rates in the 2012 LTPP dataset. Based on the annual load profile, SERVM schedules these planned maintenance events although fixed schedules can also be captured.

4. Unit commitment uncertainty is modeled using forecast errors at different time intervals to ensure the model does not have perfect knowledge when performing the unit commitment. Forecast error is drawn separately for gross load, wind, and solar and aggregated into net load uncertainty for each time interval. These time intervals include week ahead, day ahead, multi-hour ahead, and intra-hour time periods. SERVM allows for recourse at each of these time intervals by allowing the commitment to adjust subject to physical resource constraints as more certainty is gained about the net load as the prompt hour approaches. Intra-hour, the only recourse option available is to start up quick start resources within the region that has the need. Outside regions will not start up quick start resources intra-hour to serve a neighboring region, but will provide this support for all other time intervals. This assumption could be changed in future simulations.

Key Findings

The loss of load metrics and renewable curtailment results in Table ES1 represent the two analyses developed from the 2012 LTPP Base scenario without SONGS. The removal of economic forecast uncertainty was also shown to understand its impact on results. The following key findings were observed.

- Given the 2012 LTPP assumptions for 2022, the target reserve margin[3] for the CAISO system including imports is approximately 17%. This counts the renewable portfolio which includes all biomass, biogas, small hydro, solar, and wind at 35% of nameplate capacity.

- When simulating uncertainty around load forecasts, weather, and unit performance, the traditional $LOLE_{GEN}$ is greater than the 1 day in 10 year standard of 0.1 LOLE in events per year. $LOLE_{GEN}$ for the single shape analysis was 0.34 and 0.77 events per year for the multiple shape analysis. The simulations showed that almost all of the $LOLE_{GEN}$ events occurred after hour 18 of the day as demand response availability and renewable resources decreased. After hour 18, demand response is limited to approximately 700 MW versus 2,600 MW across the peak and the reliability contribution of renewable resources drops from an average of 35% to 20%. If demand response is allowed to provide 2,600 MW for all hours in the summer, then the single shape $LOLE_{GEN}$ shifts from 0.34 events per year to 0.13 events per year. The multi shape analysis shifts from $LOLE_{GEN}$ shifts from 0.77 events per year to 0.32 events per year.

- The total LOLE ($LOLE_{Gen+Flex}$) equals 0.426 for the single shape analysis and 1.064 when incorporating all 32 load, wind, and solar shapes. Again, allowing demand response resources to provide 2,600 MW for all hours of the day in the summer, $LOLE_{GEN+FLEX}$ reduces to 0.158 for the single shape analysis and 0.478 for the multi shape analysis. The majority of $LOLE_{FLEX}$ events occurred intra-hour during high load, but not during annual peak load periods as LOLE during

[3] The target reserve margin is calculated based on the forecasted peak load for 2022.

annual peak periods would be represented by $LOLE_{GEN}$. As previously noted, SERVM allows for additional CT commitment intra-hour with the assumption that quick start resources can start in 10 minutes. However, if load patterns are such that resources are only needed for a few hours, it is possible that CT resources are committed and some intermediate resources are not committed based on uncertain net load projections. During these higher load periods, if the intra-hour net load materializes substantially higher than projected during commitment, the region will not have recourse opportunities since the CTs were committed prior and the intermediate resources cannot start-up quickly enough. This situation primarily occurs in regions which are relatively short capacity. Regions with excess capacity typically have some spare CT capacity even on days when some intermediate resources are not committed. For the regions which did produce $LOLE_{FLEX}$, a significant portion of these events could potentially be eliminated if neighboring regions could provide intra-hour support. Currently, there is no market purchase recourse method available intra-hour in SERVM. Since $LOLE_{FLEX\ events}$ occur when additional capacity is available but not committed, additional load following requirements added during the commitment process would also remove most of the intra-hour flexibility problems. Higher load following requirements could also reduce some of the $LOLE_{GEN}$ since purchases would be made when available to preserve load following capability for unexpected changes in net load.

- Curtailment was substantial in the 2022 scenario due to significant must-take generation including renewable, hydro and dedicated imports and due to a restriction on exports to 0 MW from CAISO. Curtailment is highly dependent on the flexibility assumed for hydro and dedicated imports across the peak hours of day during shoulder months. For this analysis, the dedicated imports were treated as must-take generation so there was no flexibility to curtail these resources. The hydro was forced to meet weekly generation amounts assumed in the 2012 LTPP data set which meant a substantial amount of hydro was still being dispatched during peak renewable hours for the shoulder months.

- Analyzing multi-weather years versus a single shape had a significant impact on results. Loss of load metrics increased due to more severe net load shapes seen across all the weather years that are not recognized in a single shape analysis. Generation curtailment decreased substantially. This is likely due to lower capacity factor renewable shapes have during off-peak periods along with load shapes with more energy during off-peak periods.

- As expected, removing economic load forecast uncertainty shifted loss of load events down.

Table ES1. Summary of Results

		Single 2012 LTPP Hourly Shapes	Single 2012 LTPP Hourly Shapes: No Economic Load Growth Uncertainty	Multi Weather Year Hourly Shapes	Multi Weather Year Hourly Shapes: No Economic Load Growth Uncertainty
Reserve Margin	%	17%	17%	17%	17%
$LOLE_{GEN}$	Events/Yr	0.343	0.238	0.771	0.543
$LOLE_{Flex}$	Events/Yr	0.083	0.025	0.293	0.260
$LOLE_{GEN+FLEX}$	Events/Yr	0.426	0.264	1.064	0.803
EUE_{GEN}	MWh	141	51	211	69
EUE_{FLEX}	MWh	92	9	68	31
$EUE_{GEN+FLEX}$	MWh	233	60	279	100
Generation Curtailment (MWh): Inflexible Hydro and Dedicated Imports	MWh	1,738,919	1,693,088	385,785	377,795

While there are several solutions available that would shift the CAISO system reliability back to a 0.1 $LOLE_{GEN+FLEX}$[4], Table ES2 shows sensitivities in which quick start CT capacity was added to each scenario in 2% reserve margin increments. Using the single load shape analysis, approximately 1,100 MW are needed to achieve an $LOLE_{GEN+FLEX}$ of 0.1. For the multiyear weather shape analysis, approximately 1,900 MW are needed to achieve $LOLE_{GEN+FLEX}$ of 0.1. The incremental capacity requirements result in reserve margins of 19%-20.5% needed to achieve the one event in 10 year standard from both a peak and flexibility standpoint.

Table ES2. CT Capacity Additions Needed to Meet 1 Day in 10 year Industry Standard (0.1 $LOLE_{GEN+FLEX}$) Assuming no Load Growth Uncertainty

		Analysis1: Single Hourly Shapes	Base +1170 MW	Base +2340 MW
Target Reserve Margin	%	17%	19%	21%
$LOLE_{GEN}$	Events/Yr	0.238	0.065	0.015
$LOLE_{FLEX}$	Events/Yr	0.025	0.016	0.000
$LOLE_{GEN+FLEX}$	Events/Yr	0.264	0.082	0.015

		Analysis2: Multi-Year Weather Shapes	Base +1170 MW	Base +2340 MW
Target Reserve Margin	%	17%	19%	21%
$LOLE_{GEN}$	Events/Yr	0.543	0.123	0.042
$LOLE_{FLEX}$	Events/Yr	0.26	0.068	0.005
$LOLE_{GEN+FLEX}$	Events/Yr	0.803	0.191	0.047

Given the physical reliability results, supplemental economic analysis using the same SERVM setup should be performed to determine the cost-effectiveness of alternatives to achieve a specific reliability

[4] This analysis assumes that target reliability is equal to 0.1 LOLE whether due to capacity deficiencies or flexibility deficiencies. Traditional LOLE metrics do not typically include $LOLE_{FLEX\ as\ part}$ of the total.

metric such as 0.1 $LOLE_{GEN+FLEX}$. These should include increased load following, increase in capacity of different generation types, expansion of DR programs, and the replacement of inflexible generation.

I. Input Assumptions

A. Source Data and Study Year

All input data (i.e. load forecasts, generator data, etc) was based on CAISO's 2012 LTPP for the year 2022 unless otherwise specified in the following Input Sections. Two analyses were developed from the 2012 LTPP case.

1. The first analysis used only the single load, wind, and solar shapes within the 2012 LTPP study which would allow more direct comparison with the CAISO deterministic method.

2. The second analysis incorporates 32 years of weather history by simulating 32 synthetic load shapes, wind shapes, and solar shapes for the CAISO system. This is the recommended approach as the historical frequency and duration of severe weather is captured more accurately.

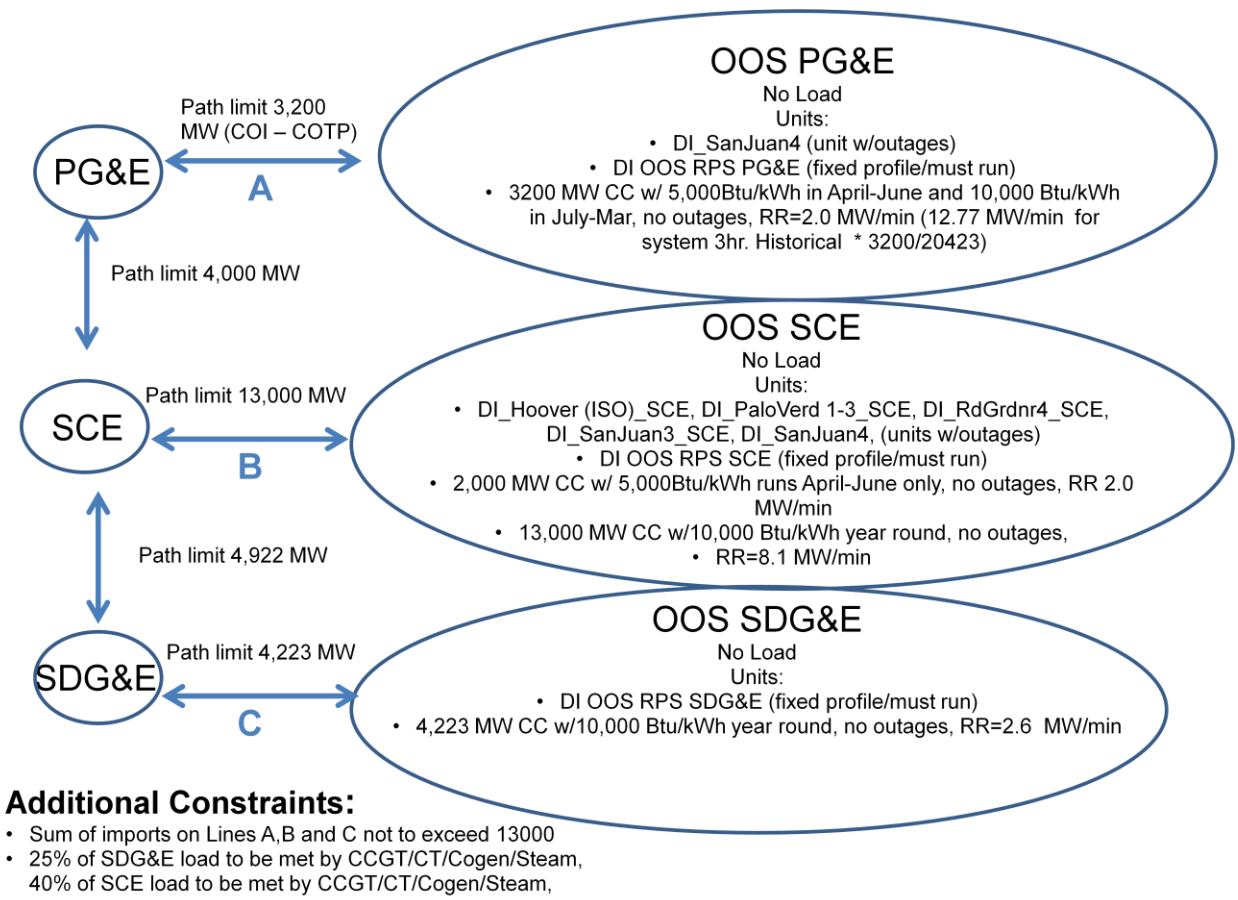
B. Study Topology

Figure 1 shows the study topology that was used for the Study. While SERVM provides the capability to model the entire WECC Region[5], due to various reasons including schedule constraints, the focus was on the CAISO region and the remainder of WECC was modeled in a simplified approach to approximate the imports into CAISO.[6] SERVM represents the regions in Figure 1 with a pipe and bubble representation allowing for regions to share capacity based on economics and subject to physical transmission constraints. Each of the external Out of State (OOS) regions was modeled with no load. All

[5] SERVM has been used to model large portions of the eastern interconnect and will be used to model all of WECC for the CPUC's ELCC analysis.
[6] Astrape Consulting recommends a more rigorous approach for future studies which would incorporate most or all of WECC. Only by modeling a more detailed representation of the outside regions' load and resources will a more accurate representation of imports be developed.

OOS RPS and Dedicated Imports (DI) were captured separately and were treated as must-take on the CAISO's system bubbles. Additional available capacity was modeled to capture the amount of imports expected into the CAISO system. The ramp rate of imports was controlled by the ramp rates of the external units. The maximum instantaneous import capability into CAISO was modeled as 13,000 MW. Additional minimum generation requirements were modeled in SDG&E and SCE: respectively, 25% and 40% of load in these areas was required to be served by conventional generation.

Figure 1. Study Topology

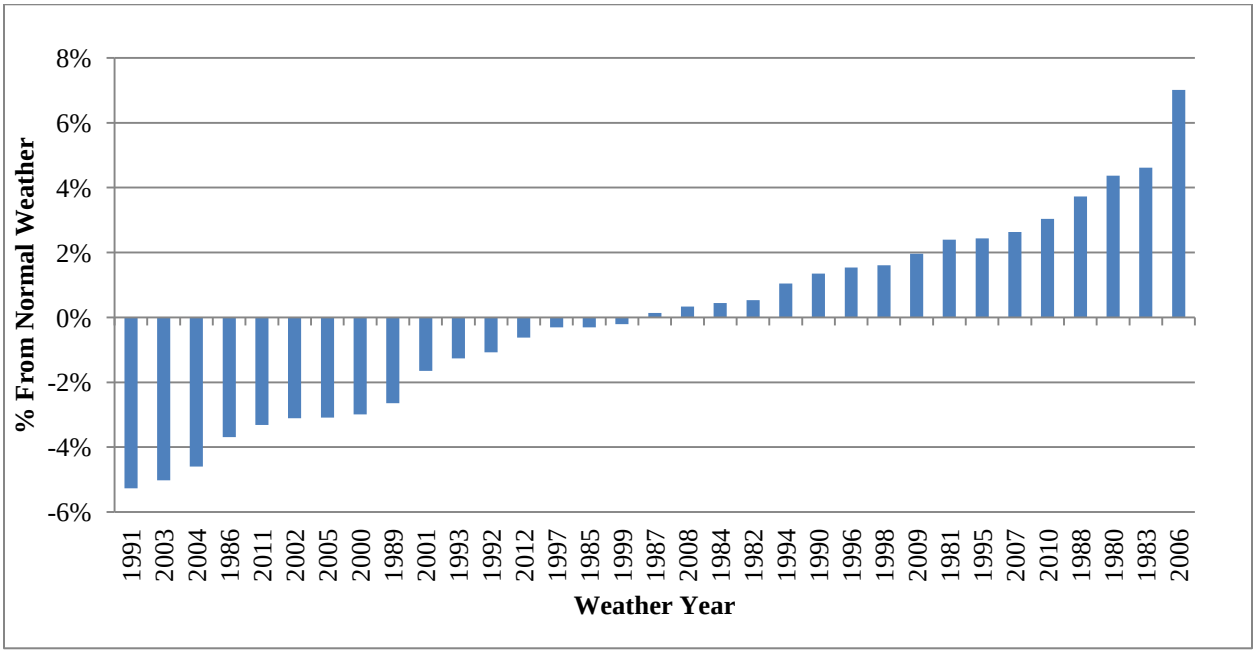


C. Load Modeling

Figure 2 displays the variance in peak load based on 33 years of historical weather.[7] This represents the simulated summer peak load for 2022 by weather year. The normal weather peak is assumed to be the average peak across all shapes and has been scaled to the peak seen in the 2012 LTPP. Compared to a normal weather year, peak loads across all three regions can be as high as 7% above normal and as low as 5% below normal as shown in the figure. This does not include any multi-year ahead economic load growth uncertainty, but only shows uncertainty due to weather being more severe or mild in a given year. The loads were developed using neural nets to develop a relationship between recent load and recent weather for each CAISO region. Next the relationship was applied to historical weather to develop multiple load shapes. This process not only captures the variability in peak but also captures the frequency and duration of severe weather seen in actual history and across each season. This type of modeling also captures the weather diversity among the regions within CAISO.

[7] Even though the variance in peak load was based on 33 years of historiy, SERVM only simulated 32 years since 1980 was inadvertently omitted from the simulation.

Figure 2. CAISO (PG&E, SCE, SDGE) Summer Peak Load Variance[8]



D. Multi Year Ahead Economic Load Growth Uncertainty

Economic load forecast error multipliers were developed to isolate the economic load growth uncertainty inherent in four year-ahead[9] load forecasts. Based on historical Congressional Budget Office (CBO) GDP forecasts 4 years ahead compared to actual data, a normal distribution of economic forecast error was calculated. Because electric load grows at a slower rate than GDP, a 40% multiplier was applied to the raw CBO economic forecast error to produce an economic load forecast error distribution. Table 1 shows the economic load forecast multipliers and associated probabilities. The table shows that 7.9% of the time, it is expected that load will be under-forecasted by 4% four years out. For the multi-weather shape analysis, the SERVM model utilized each of the 32 weather years and applied each of

[8] Ranking of years by peak load will not be perfectly correlated to the ranking of years by peak temperature. This is due to the possibility of peak temperatures falling on a weekend which would result in lower relative loads. That explains why the 2005 peak load in the multi-shape analysis is less than the normal peak load.

[9] Four year ahead forecast uncertainty was assumed for this analysis. It is noted that it likely takes more than four years to secure procurement authorization, procure competitively and actually license and build a powerplant in CA.

these five load forecast error points to create 160 different load scenarios. Each weather year was given equal probability of occurrence. The multipliers are applied to every hour in the year.

Table 1. 4 Year Ahead Economic Load Growth Uncertainty used in Multiple Weather Year Analysis

Load Forecast Error Multipliers	Probability %
0.96	7.9%
0.98	24.0%
1.00	36.3%
1.02	24.0%
1.04	7.9%

For the analysis performed using the 2012 LTPP single load shapes, Table 2 was developed to represent both weather and economic load growth uncertainty since multiple weather years weren't simulated. The peak load statistics around the 33 years of load shapes were used to create the weather uncertainty which was combined with the Table 1 economic load growth uncertainty multipliers. The peak load as a percentage of normal peak for each shape was multiplied by each of the probabilities from the values in Table 1. The resulting distribution including 165 points was simplified to a distribution of 10 points by combining points with similar multipliers resulting in the values in Table 2 below.

Table 2. Weather Plus 4 Year Ahead Economic Load Forecast Error Uncertainty used in Single Load Shape Analysis Only

Load Forecast Error Multipliers	Probability
0.93	7.20%
0.96	12.40%
0.98	10.20%
0.99	12.40%
1.00	12.40%
1.01	11.70%
1.02	12.40%
1.03	8.00%
1.05	9.10%
1.08	4.30%

E. Unit Outage Data

Unlike typical production cost models, SERVM does not use an Equivalent Forced Outage Rate (EFOR) for each unit as an input. Instead, historical Generating Availability Data System (GADS) data events are typically entered in for each unit and SERVM randomly draws from these events to simulate the unit outages. For this Study, the mean time to repair and EFOR values from PLEXOS were utilized and a mean time to failure value was calculated. Distributions around these values were then developed to be input into SERVM to represent the unit outage uncertainty. Data is typically entered into SERVM for the following variables. However, partial outage data and maintenance outage data was unavailable from the PLEXOS dataset so only the full outage inputs and planned outages were captured.

Full Outage Modeling
Time-to-Repair Hours
Time-to-Fail Hours

Partial Outage Modeling
Partial Outage Time-to-Repair Hours
Partial Outage Derate Percentage
Partial Outage Time-to-Fail Hours

Maintenance Outages
Maintenance Outage Rate - % of time in a month that the unit will be on maintenance outage. SERVM uses this percentage and schedules the maintenance outages during off peak periods

Planned Outages

Specific time periods are entered for planned outages. Typically these are performed during shoulder months.

As an example of how SERVM develops and uses unit outage statistics, assume that from 2008 – 2014, Unit A had 15 full outage events and 30 partial outage events reported in the GADs data. The Time-to-Repair and Time-to-Fail between each event is calculated from the GADS data. These multiple Time-to-Repair and Time-to-Fail inputs are the distributions used by SERVM. Because typically there is an improvement in EFOR across the summer, the data is broken up into seasons such that there is a set of Time-to-Repair and Time-to-Fail inputs for summer, shoulder months, and winter based on history. Further, assume Unit 1 is online in hour 1 of the yearly iteration. SERVM will randomly draw a Time-to-Fail value from the distribution provided for both full outages and partial outages. The unit will run for that amount of time before failing. A partial outage will be triggered first if the selected partial outage Time-to-Fail value is lower than the selected full outage Time-to-Fail value. Next, the model will draw a Time-to-Repair value from the distribution and be on outage for that number of hours. When the repair is complete it will draw a new Time-to-Fail value. The process repeats until the end of the iteration when it will begin again for the subsequent iteration. The full outage counters and partial outage counters run in parallel. This more detailed modeling is important to capture the tails of the distribution that a simple convolution method would not capture.

The most important aspect of unit performance modeling in reliability studies is the cumulative MW offline distribution. Most service reliability problems are due to significant coincident outages. The following figure shows the distribution of outages for CAISO based on the 2012 LTPP Dataset. The figure demonstrates that in any given hour, the CAISO system can have between 50 and 4,000 MWs of its generators offline due to forced outages. The figure shows that during 10% of all hours throughout the year, CAISO has greater than 2,000 MW in a forced outage condition. There are approximately 28,000

MW[10] of conventional generation modeled within in CAISO which is made up of nuclear, combined cycle, CHP, and peaking resources. Additionally, the figure shows that 50% of the time, approximately 1,000 MW are on outage which equates to 3.5% of the conventional generation.

Figure 3. Conventional Resources on Forced Outage as a Percentage of Time

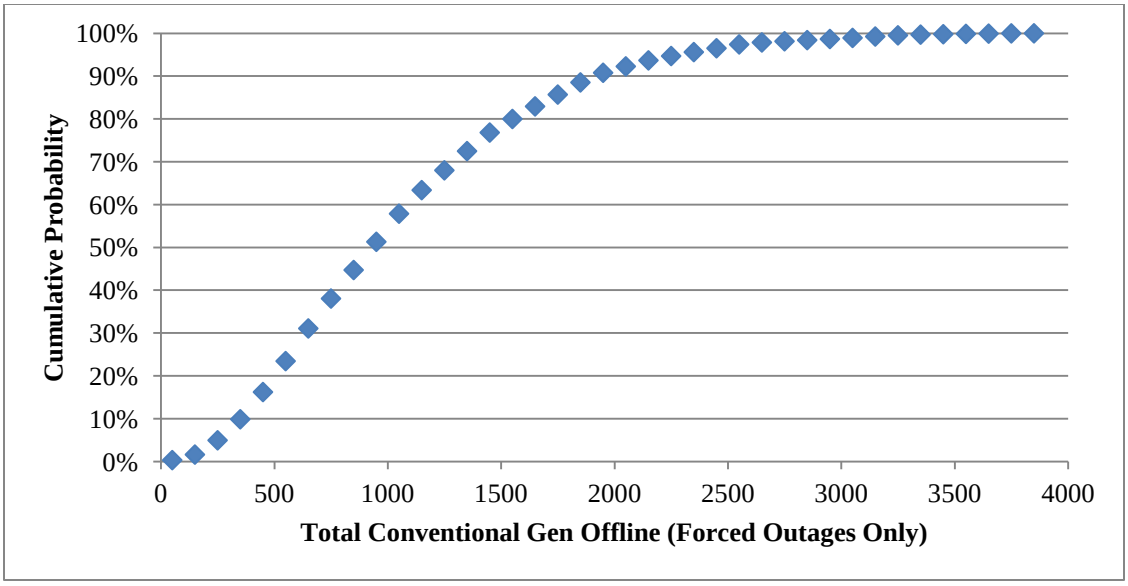
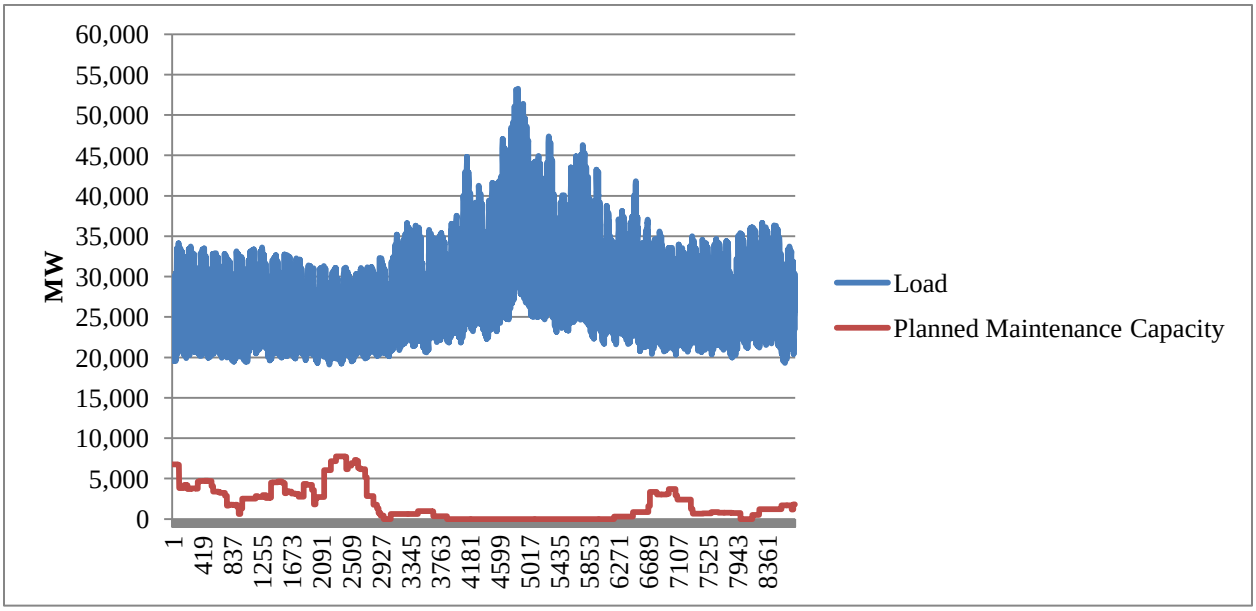


Figure 4 shows how SERVM takes the planned maintenance rates used in the 2012 LTPP and develops planned outage schedules across the year based on projected load periods.

[10] The figure does not include imports, pump storage, hydro, or other renewable resources.

Figure 4. Planned Maintenance



F. Hydro Modeling

Hydro resources are split into 3 categories within SERVM.:

1. Run of River Hydro: Dispatched as a fixed profile for the entire year based on the 2012 LTPP assumptions.

2. Scheduled Hydro: Dispatched to shave the peak but is forced to meet minimum gen requirements and max capacity levels. A weekly hydro generation is provided that must be fully used within the week. The hydro energies were based on the 33 historical hydro years. The max capacity levels were based on the realized hydro dispatched across the peak in the 2012 LTPP assumptions.

3. Emergency Hydro: Dispatched only in emergency events when prices meet a specific threshold and is energy limited. The assumed price threshold for this study was $2,500/MWh. These resources are linked to a scheduled hydro resource. When called, energy from the scheduled hydro resource is reduced. These estimated capacity levels were based on taking the total nameplate without the scheduled or run of river portions.

Figure 5 shows the total nameplate capacity of the hydro system modeled which is based on the 2012

LTPP Dataset. Based on 33 years of historical hydro energies the annual energies were developed and

shown in Figure 6. Depending on the weather year, hydro generation within the simulations varied significantly.

Figure 5. Hydro Capacity

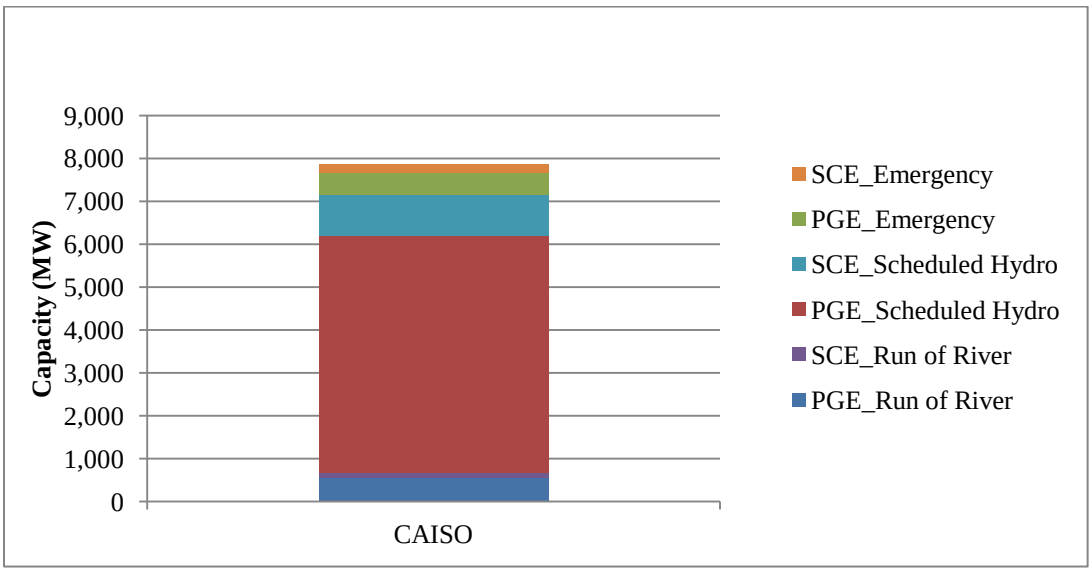
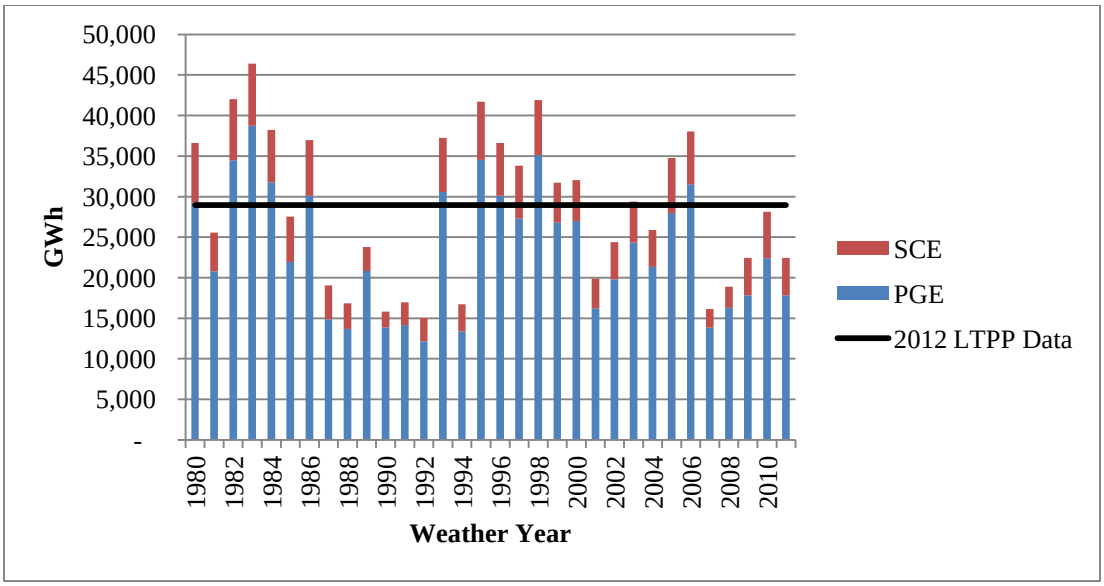


Figure 6. Hydro Energy



G. Operating Reserve Requirements and Operating Reserve Demand Curve

Table 3 shows the assumptions that were used by SERVM for regulation, spin, non-spin, and load following Targets.[11] These values represent target volumes which SERVM tries to provide if cost effective when comparing the value of regulation and load following against the unit commitment and dispatch costs. As noted in the table, the model will shed firm load to maintain 1.5% of regulation and 1.5% of spin. During peak periods, the load following requirements from SERVM represented load following requirement methods use in the modeling collaboration effort.

Table 3. Operating Reserve Targets[12]

	% of Load	Shed Firm Load to Maintain?
Regulation Up/Regulation Down	1.50%	Yes
Spin	3.00%	Yes for 1.5% of the total 3%
Load Following	On average totals approximately 4% - 5% Calculate based on the difference between instantaneous peak within the hour and average across the hour plus 1% of load	No
Non Spin	3%	No

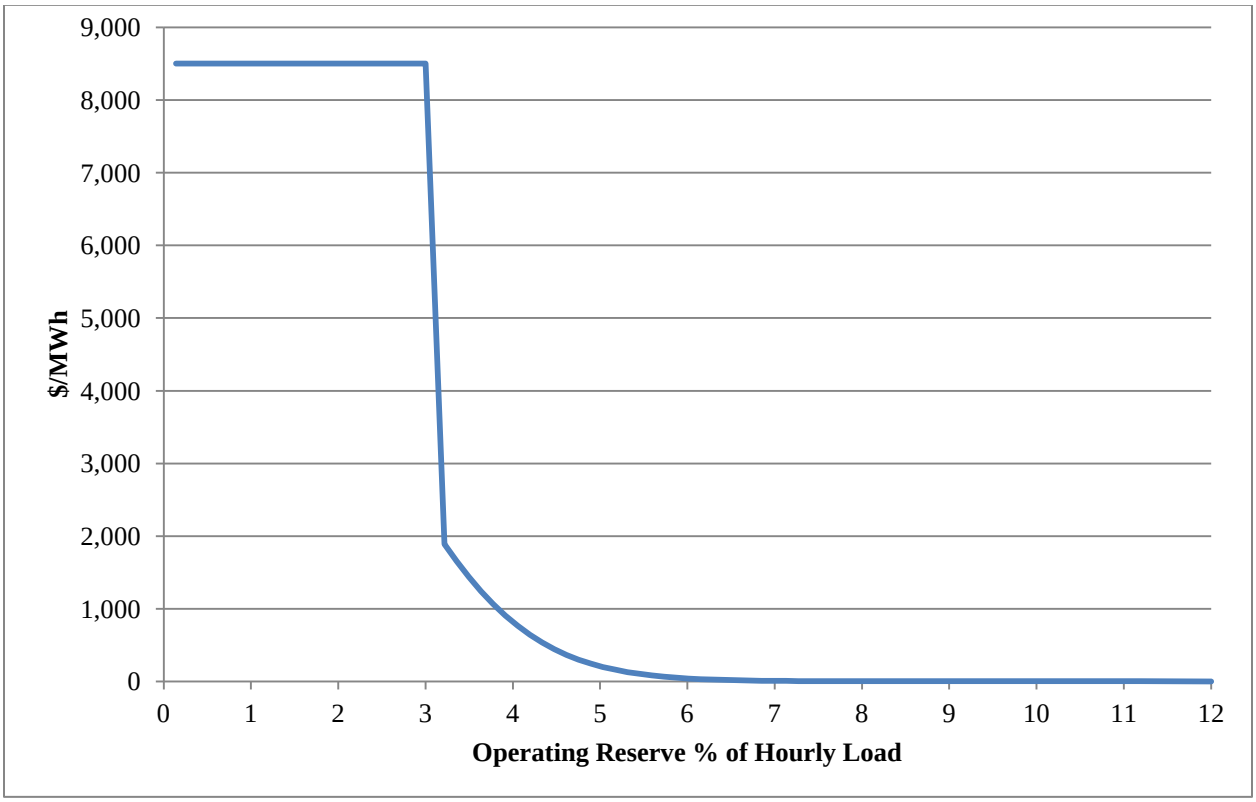
Figure 7 displays the operating reserve demand curve that was used in SERVM's unit commitment to determine how much additional spin and load following above the required 3% that is provided in any given hour. The prices in the curve represent incremental scarcity pricing above the marginal cost resource that meets the 1.5% regulation plus 1.5% spinning reserve. The curve is assumed to be flat for the first 3% at a value representing the Value of Lost Load (VOLL). At these prices, all resources in the system would be utilized to maintain the required 1.5% regulation and 1.5% spinning reserves. Then the

[11] The major distinction between SERVM and some of the other approaches used in modeling CAISO is that the load following target in SERVM is calculated based on the variability across the hour rather than a set value from an 8760 profile.

[12] The word "target" is used in this table to prepresent the desired amount of different reserves, which is purchased by SERVM if cost-effective.

curve drops down significantly and as long as the marginal cost resource of the next unit is below the operating reserve demand curve, then the system will achieve the full operating reserve requirement.

Figure 7. Operating Reserve Demand Curve

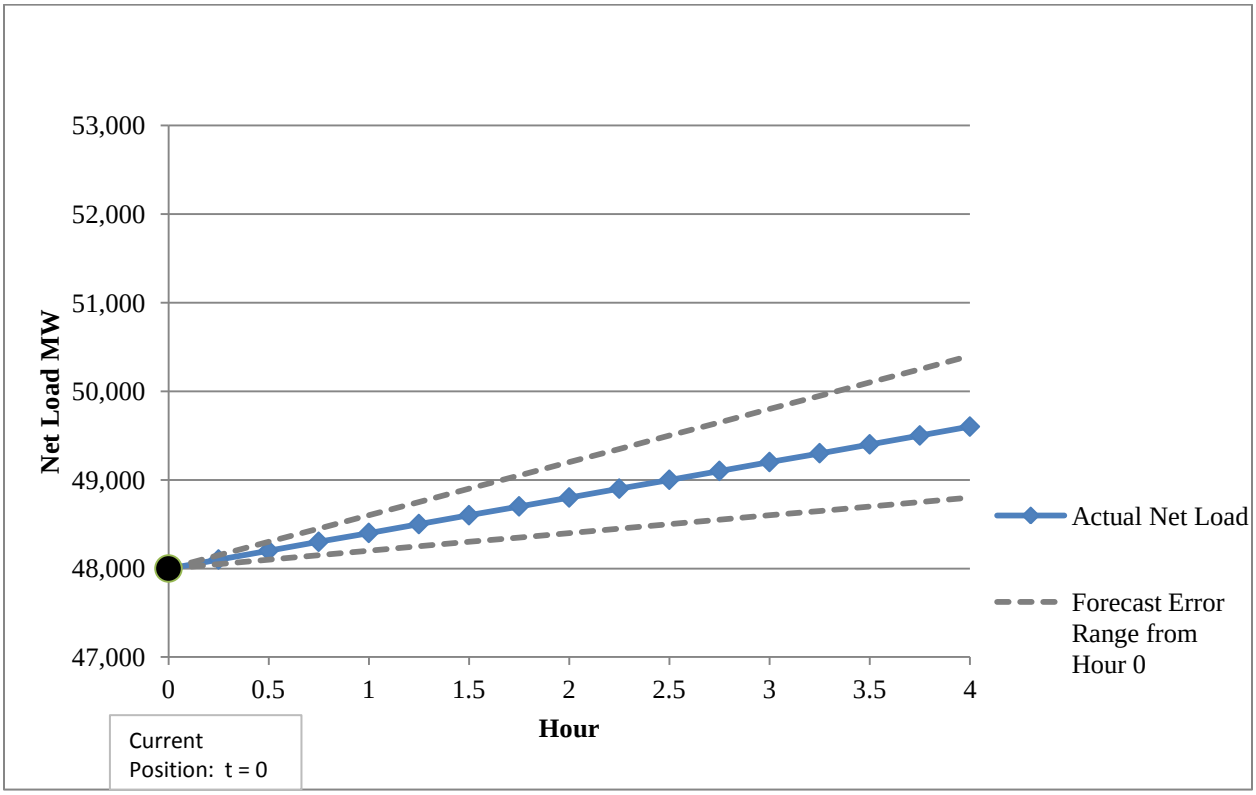


H. Unit Commitment Uncertainty & Recourse

SERVM's full economic unit commitment occurs over several time intervals. Each unit commitment is based on a forecasted net load that is calculated based on the uncertainty distributions at each time interval. First, a weekly commitment is done for the entire week. Then each day a day-ahead commitment is performed making adjustments to the original commitment to take into account unexpected outages that occurred since the weekly commitment was performed. Net load uncertainty is not re-drawn at this stage. Subsequent unit commitment decisions are made subject to resource constraints 4 hours ahead, 3 hours ahead, 2 hours ahead, and 1 hour ahead using new draws of forecast

uncertainty. Finally intra-hour commitment of quick start resources is allowed as the intra-hour load varies. As the actual hour is approached, the uncertainty is narrowed, and SERVM is allowed to make adjustments at each stage subject to physical constraints of the resources. Figure 8 provides an example of how the model adjusts its commitment each hour and how the uncertainty expands for long time intervals. At hour 0, SERVM draws from correlated load, wind, and solar forecast error distributions for intra-hour, 1 hour ahead, 2 hours ahead, 3 hours ahead, and 4 hours ahead uncertainties. SERVM then makes commitment and dispatch adjustments based on the uncertain forecast, but ultimately must meet the net load shape that materializes[13].

Figure 8. 1-4 Hour Ahead Forecast Error

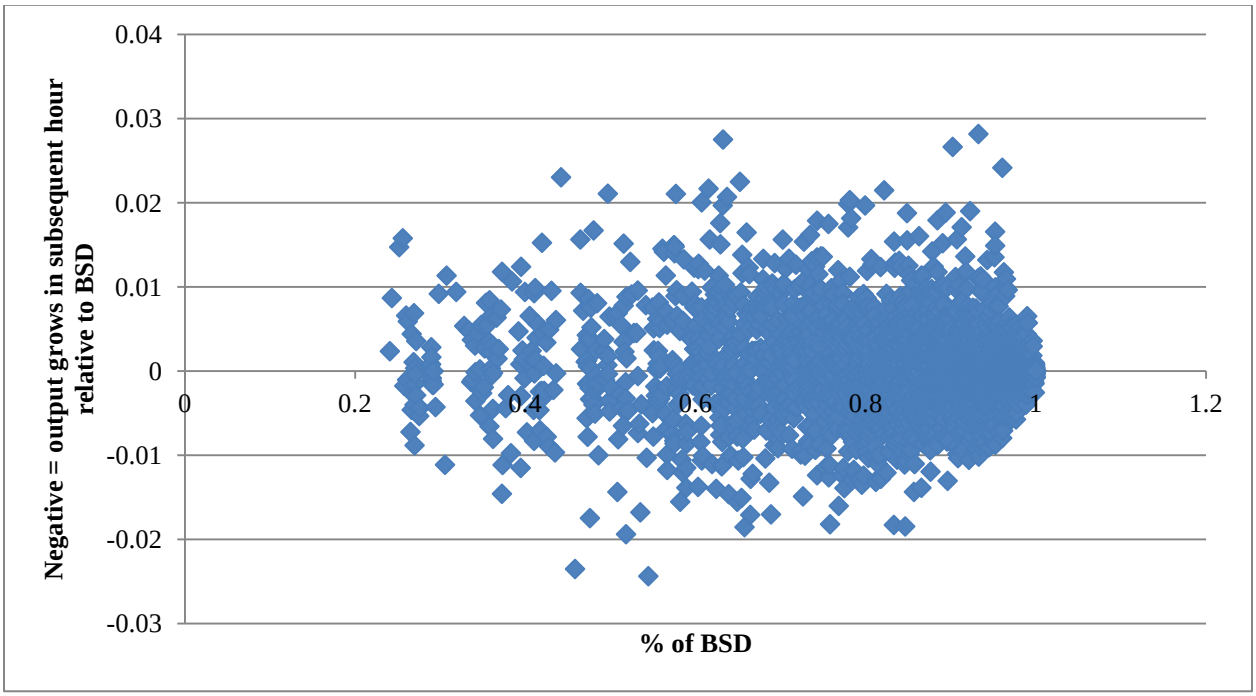


Forecast data one hour ahead, four hours ahead, and day ahead for aggregated wind and solar profiles for California along with corresponding actual wind and solar data was developed by EPRI and was used by

[13] The net load shape that materializes is always the original input shape into SERVM

Astrape to develop uncertainty distributions[14] for each of the same time intervals. The following figures represent samples of the distributions.

Figure 9. Intra-Hour Solar Uncertainty Example (Represents Hour 12)



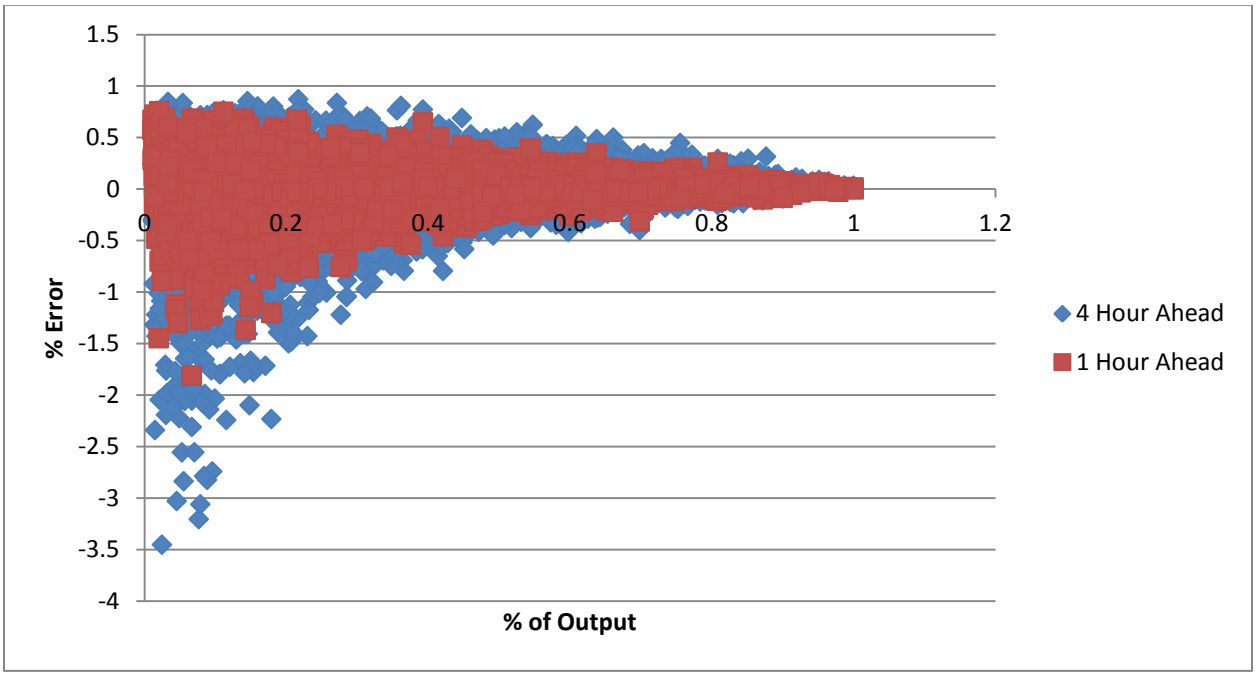
Solar forecast uncertainty[15] is represented as a function of time of day and the actual profile percentage of blue sky day (BSD). Blue sky day represents the theoretical maximum output given ideal weather conditions at each period throughout the year. At values equal to the theoretical maximum output, it is impossible to over-forecast the output. Similarly, at lower output values, the forecast error will be asymmetrical on the under-forecast side.

SERVM uses a similar technique of utilizing the actual profiles (wind, solar, and load) to select random draws of uncertainty that are reflective of the appropriate system conditions. This does not result in

[14] It is recommended that these distributions be updated in refined for future analysis.
[15] Uncertainty as used in this report only represents the deviation from the actual profile. The expected variability or ramping of load, wind, and solar is excluded from the uncertainty distributions for both inter-hour and intra-hour periods.

perfect knowledge bias since the distribution of forecast errors from the simulations match the forecast error distributions from actual history that were input into the model.

Figure 10. Multi-Hour Wind Uncertainty



The error bounds for forecast uncertainty are more strongly correlated to output level for wind resources as shown in the figure above. Again, some portions of the distribution are asymmetrical based on actual forecast error data.

Figure 11. Day Ahead Solar Uncertainty

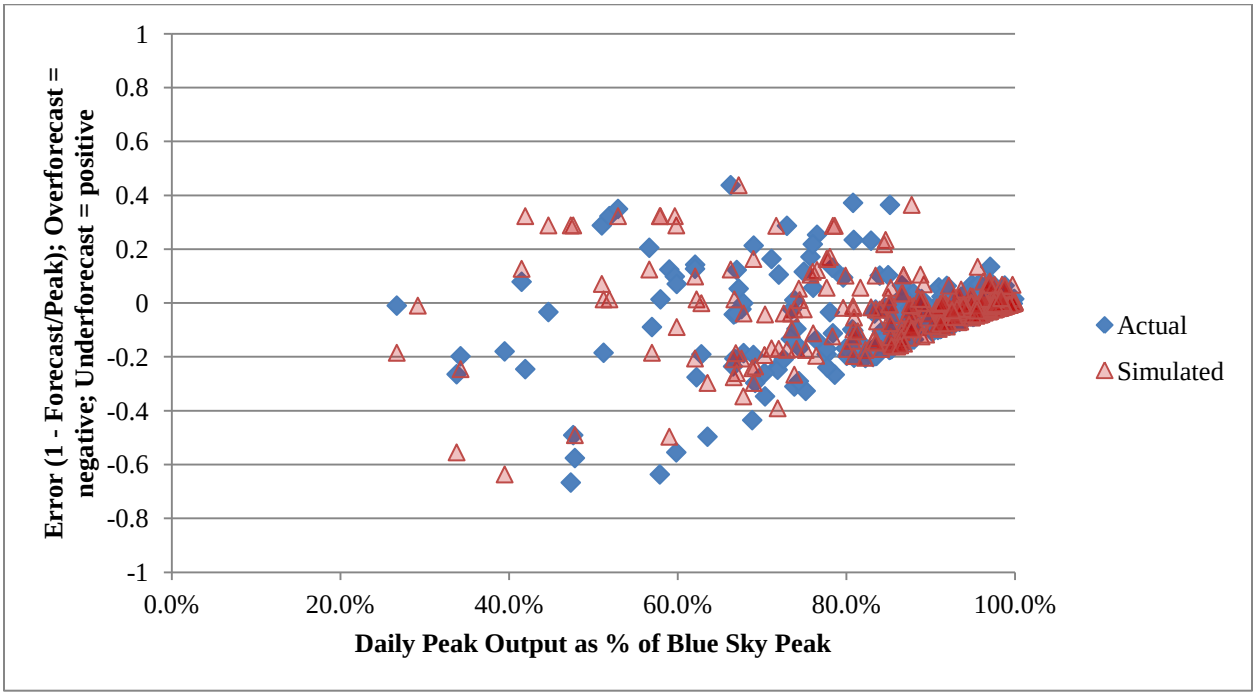


Figure 11 compares the input distribution for day ahead solar uncertainty with simulated solar forecast uncertainty.

I. CT Startup Times Intra-Hour

For this study, 10 minute startup times were assumed for quick start resources to be utilized intra-hour. This is the only unit commitment recourse option available intra-hour to each region, however online units can also be re-dispatched subject to ramp rate constraints. The model does not allow an external region to startup a quick start resource intra-hour to assist a neighbor; however, this can be done on all other time intervals (i.e hourly, day ahead, etc.) The quick start assumption import rules can be a significant driver in the intra-hour flexibility deficiency results.

J. Overgen Penalty

The overgen penalty or cost of renewable curtailment is an input. This economic assumption can affect the particular mix of resources selected for commitment. The commitment algorithm penalizes the selection of resources that would push minimum generation online above load by the input penalty amount. The lower the penalty, the more likely these situations occur. For this study, the overgen penalty was $200/MWh.

K. Demand Response Resources

Figure 12 shows the demand response assumptions used for the study. SERVM has the ability to model availability periods across the day as well as limit run hours across the day, month, season, or year. These assumptions are based on the 2012 LTPP. An import item to note in these is assumptions is that nearly 1,700 MW of the total 2,595 MW are unavailable after hour 18. This assumption has a significant impact on reliability results as solar profiles are declining during this period.

Figure 12. Demand Response Assumptions

Utility	Category	Price ($/MWh)	Max Capacity (MW)	Available	Daily Energy Limit (GWh)	Monthly Energy Limit (GWh)
PG&E	Low Cost	137	265	H12-18	1.08	
PG&E	Mid Cost	600	265	H12-18		5.37
PG&E	High Cost	1,000	136	All Hours		
PG&E Total			665		1.08	5.37
SCE	Low Cost	137	612	H12-18	2.51	
SCE	Mid Cost	600	612	H12-18		12.43
SCE	High Cost	1,000	572	All Hours		
SCE Total			1,797		2.51	12.43
SDG&E	Low Cost	137	62	H12-18	0.25	
SDG&E	Mid Cost	600	62	H12-18		1.25
SDG&E	High Cost	1,000	9	All Hours		
SDG&E Total			133		0.25	1.25
Sum			2,595		3.85	19.06

II. Simulation Methodology

A. Case Probabilities

Analysis 1: As discussed previously, the first analysis used the single load, solar, and wind shapes from the 2012 LTPP Dataset. To capture uncertainty in load for this analysis, the hourly load shapes were scaled up and down by 10 uncertainty multipliers which covered both weather and economic load growth uncertainty. For each of the 10 scenarios, 100 iterations were simulated which totaled 1,000 total iterations at 5 min intervals. The following table shows how the probabilities were calculated for each scenario.

Table 4. Case Probabilities for Single Weather Year Analysis

Weather Year	Load Multipliers due to Load Forecast Error	Load Multiplier Probability
Single Year	93%	7.20%
Single Year	96%	12.40%
Single Year	98%	10.20%
Single Year	99%	12.40%
Single Year	100%	12.40%
Single Year	101%	11.70%
Single Year	102%	12.40%
Single Year	103%	8.00%
Single Year	105%	9.00%
Single Year	108%	4.30%

Analysis 2: The second analysis used 32 load, solar, and wind shapes based on 32 years of actual history. SERVM utilized 32 years of historical weather and load shapes, 5 points of economic load growth forecast error, and 100 iterations of unit outage draws for each scenario to represent the full distribution of realistic scenarios. The number of yearly simulation cases at 5 min intervals that were simulated equates to 32 weather years * 5 load forecast errors * 100 unit outage iterations = 16,000 total iterations for the

base case. For the multi-weather shape analysis, an example of calculated probabilities for a few cases is shown in Table 5. Each weather year is given equal probability and each weather year is multiplied by the probability of each load forecast error point to calculate the case probability.

Table 5. Case Probability Example for Multi Weather Year Analysis

Weather Year	Weather Year Probability	Load Multipliers due to Load Forecast Error	Load Multiplier Probability	Case Probability
1981	3.03%	96%	7.90%	0.239%
1981	3.03%	98%	24.00%	0.727%
1981	3.03%	100%	36.30%	1.100%
1981	3.03%	102%	24.00%	0.727%
1981	3.03%	104%	7.90%	0.239%
1982	3.03%	96%	7.90%	0.239%
1982	3.03%	98%	24.00%	0.727%
1982	3.03%	100%	36.30%	1.100%
1982	3.03%	102%	24.00%	0.727%
1982	3.03%	104%	7.90%	0.239%

B. Physical Reliability Metric Outputs

The following reliability metrics are produced by SERVM for each of the 5 min interval simulations.

1. Loss of Load Expectation Generic ($LOLE_{GEN}$) – Events per year and only represents outage events that occur due to capacity shortfalls in peak conditions. If a resource is available but was not committed and can't meet load due to ramp rates or startup times, then the event is not counted.

2. Loss of Load Expectation Total ($LOLE_{GEN+FLEX}$) – Events per year and represents capacity shortfalls plus events caused from system ramping deficiencies and net load forecast error.

3. Loss of Load Hours Generic ($LOLH_{GEN}$) – Hours per year and only represents outage hours that occur due to capacity shortfalls during peak conditions.

4. Loss of Load Hours Total ($LOLH_{GEN+FLEX}$) – Hours per year and represents capacity shortfalls plus hours caused from system ramping deficiencies

5. Expected Unserved Energy Resource Adequacy (EUE_{GEN}) – Expected Unserved Energy only due to capacity shortfalls during peak conditions
6. Expected Unserved Energy Resource Adequacy ($EUE_{GEN+FLEX}$) – Expected Unserved Energy due to capacity shortfalls plus system ramping deficiencies

SERVM's ability to perform an accurate commitment and dispatch quickly, allows comprehensive reliability analysis to be performed without any seams issues between multiple models. All resource adequacy metrics from a generic capacity and flexibility standpoint can be produced from the same simulations. This avoids any estimation from one model to the next and provides meaningful incremental analysis between generic capacity shortfalls versus flexibility shortfalls. This, however, is not enough to determine how best to meet any system deficiencies. It is possible that a flexibility shortfall can be met with adding non-flexible capacity. Only by running several sensitivities to test the cost-effectiveness of adding different flexible and generic resources or changing operating guidelines, it is possible to tell which alternative is best.

III. Results: 2012 LTPP Single Profile Analysis

As previously discussed, initial simulations assumed single load shapes and renewable profiles included in the 2012 LTPP case. The single load shapes were scaled up and down based on the 10 economic forecast error multipliers discussed in the input section of the report. For this analysis, 1,000 yearly simulations at a 5 min interval were simulated in SERVM. The following table summarizes those results as well as a sensitivity that removes and economic load growth uncertainty. The $LOLE_{GEN}$ is 0.343 events per year for the expected 17% reserve margin case which is higher than the 1 event in 10 year (0.1 LOLE) industry standard. This assumes no loss of load events due to unit constraints such as ramp rates, startup times, minimum uptimes, and minimum downtimes or unit commitment uncertainty.

Table 6. Single Profile Analysis

		Base Case	Base Case Without Load Growth Uncertainty Component
Target Reserve Margin	%	17%	17%
$LOLE_{GEN}$	Events/Yr	0.343	0.238
$LOLE_{FLEX}$	Events/Yr	0.083	0.025
$LOLE_{GEN+FLEX}$	Events/Yr	0.426	0.264
EUE_{GEN}	MWh	155	51
EUE_{FLEX}	MWh	23	9
$EUE_{GEN+FLEX}$	MWh	178	60
$LOLH_{GEN}$	Hours/Yr	0.611	0.32
$LOLH_{FLEX}$	Hours/Yr	0.10	0.025
$LOLH_{GEN+FLEX}$	Hours/Yr	0.711	0.35
Generation Curtailment (MWh): Inflexible Hydro and	MWh	1,738,919	1,693,088

Dedicated Imports			

When unit constraints are captured, $LOLE_{GEN+FLEX}$ = 0.426 meaning that the loss of load contributed from flexibility problems is 0.083 events per year. The corresponding EUE is small in the Base Case even including the flexibility constraints. Based on the LOLH values, events on average are approximately 1.5 hours. Generation curtailment was seen as 1.7 million MWh in the Base Case. This case assumed that there was little flexibility in the hydro system across the peak and that all dedicated imports were treated as must-take generation. All loss of load events decreased marginally when load growth uncertainty was removed from the analysis

Figure 13 shows when EUE_{GEN} and EUE_{FLEX} are occurring by hour of day. The majority of EUE_{GEN} events occur in the later hours of the day (19-22) when a significant portion of the demand response is unavailable and solar generation is decreasing. Demand response is reduced from 2,600 MW to 700 MW after hour 18. This represents a 3.5% reduction in reserve margin for all hours after 18. The total reliability contribution of the renewable resources across the peak load is approximately 35% of nameplate capacity, but across the later hours of the day the contribution reduces to an average of 20% of nameplate capacity. This difference between renewable during hours 12-18 versus hours 19-22 represents approximately 4,000 MW on average which equates to a reduction of 7.5% in reserve margin. Based on this information, it is easy to see how a 17% system reserve margin that includes imports in its reserve margin calculation could become deficient in a few hours. If it is assumed that 2,600 MW of demand response is available in all hours of the day, then $LOLE_{GEN}$ is reduced from 0.343 to 0.133 events per year which is much closer to the 1 event in 10 year standard.

EUE_{FLEX} also occurs more in the peak net load hours because the majority of quick start resources which are the only recourse option available intra-hour area already committed during these periods. If load following requirements were increased or surrounding regions were allowed to commit quick start

resources intra-hour to meet a neighbor's need, the majority of EUE_{FLEX} would likely be removed within the simulations.

Figure 13. EUE by Hour of Day

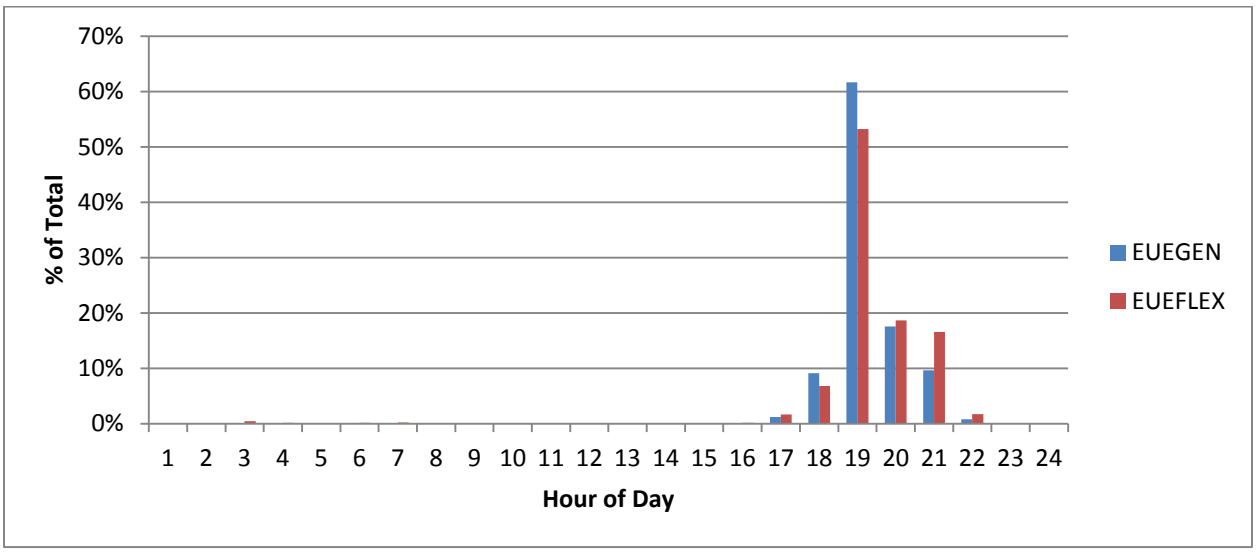
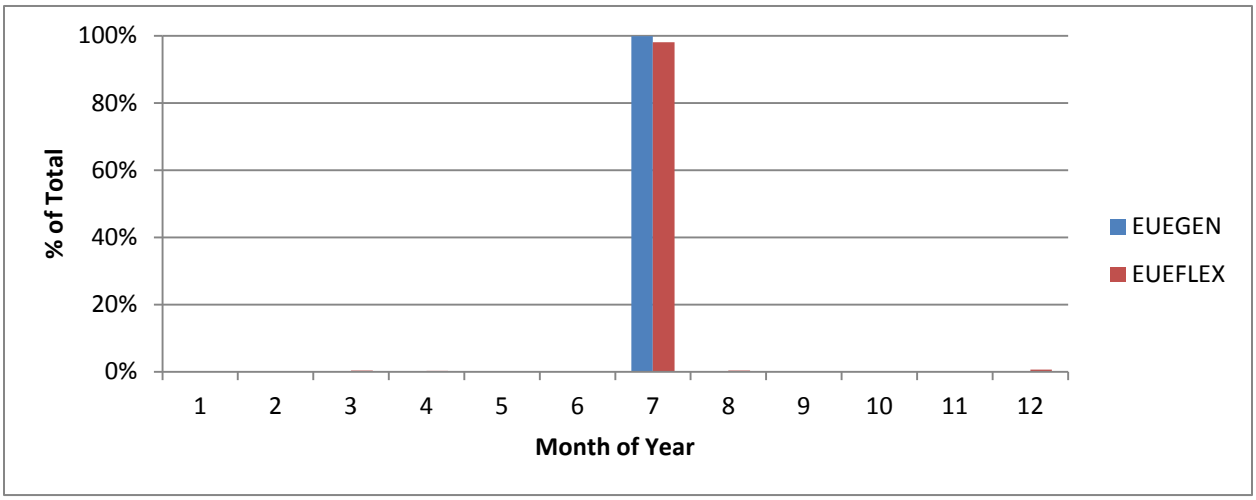


Figure 14 shows the same information by month of year. Again, the majority of events are seen during peak periods. Because the analysis is based on a single load shape, almost all loss of load events are occurring in the month of July which is when the system peaks.

Figure 14. EUE by Month



While there are many solutions to solving the flexibility problems, the solution used for this analysis was to add CT capacity with a 10 min startup time. Additional solutions would be to over commit resources or replace existing inflexible capacity with more flexible capacity. The results show that when reserve margin targets move from 17% to 19% the $LOLE_{GEN}$ decreases to 0.065, and $LOLE_{GEN+FLEX}$ decreases to 0.015. To achieve 0.1 for $LOLE_{GEN+FLEX}$ approximately 1,100 MW of additional capacity is needed resulting in a 19% reserve margin for the system.

Table 7. Base Case Results with Additional CT Capacity Assuming No Economic Load Growth Uncertainty

		Analysis1: Single Hourly Shapes	Base +1170 MW	Base +2340 MW
Target Reserve Margin	%	17%	19%	21%
$LOLE_{GEN}$	Events/Yr	0.238	0.065	0.015
$LOLE_{FLEX}$	Events/Yr	0.025	0.016	0.000
$LOLE_{GEN+FLEX}$	Events/Yr	0.264	0.082	0.015

IV. Results: 2012 LTPP Multi Weather Analysis (32 Years of Load, Wind, and Solar Shapes)

The next set of results introduces a more detailed approach to incorporating weather uncertainty. Load shapes, wind shapes, and solar shapes were developed based on the last 32 years of historical weather. To maintain correlation between the three, all 32 years were simulated within SERVM and given equal probability of occurrence. As discussed previously, each weather year was also simulated with 5 economic load forecast multipliers totaling 160 load scenarios. Each of these load scenarios was simulated with 100 iterations at a 5 min interval totaling 16,000 total years. Table 8 shows the results. Compared to the single load shape results, the loss of load metrics are higher as the multiple weather shapes introduce more variability within net load. However, the curtailment was reduced substantially when analyzing all weather years versus the single shape analysis. This is most likely due to the inclusion of lower wind and solar shapes as well as higher potential load shapes across the 32 weather years.

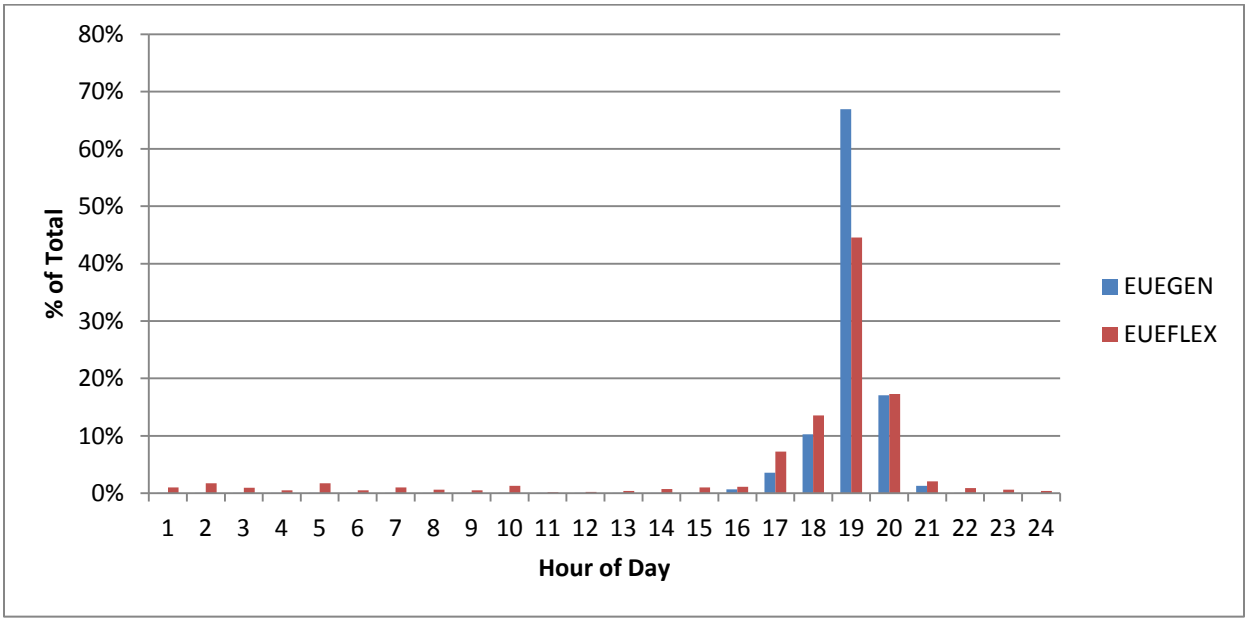
Table 8. Mutli-Weather Year Results

		Base Case	Base Case without Load Growth Uncertainty Component
TargetReserve Margin	%	17%	17%
$LOLE_{GEN}$	Events/Yr	0.771	0.543
$LOLE_{FLEX}$	Events/Yr	0.293	0.26
$LOLE_{GEN+FLEX}$	Events/Yr	1.064	0.803
EUE_{GEN}	MWh	211	69
EUE_{FLEX}	MWh	68	31
$EUE_{GEN+FLEX}$	MWh	269	100
$LOLH_{GEN}$	Hours/Yr	1.30	0.74
$LOLH_{FLEX}$	Hours/Yr	1.23	0.66
$LOLH_{GEN+FLEX}$	Hours/Yr	2.53	1.40

Generation Curtailment (MWh): Inflexible Hydro and Dedicated Imports	MWh	385,785	377,795

As seen in the single profile analysis, the base case which assumes approximately a 17% CAISO reserve margin doesn't meet a 1 day in 10 year standard. As seen in the previous analysis, $LOLE_{GEN}$ occurs mostly in the later hours of the day. Figure 15 shows the results by time of day and Figure 16 shows the events by month of year.

Figure 15. Multi Weather Year EUE by Hour of Day



Because we are evaluating different load shapes and not the single load shape, the events are more spread across the summer months and even some of the shoulder periods. There is a significant difference in evaluating multiple load, solar, and wind profiles compared to only analyzing a single shape.

Figure 16. Multi Weather Year EUE by Month

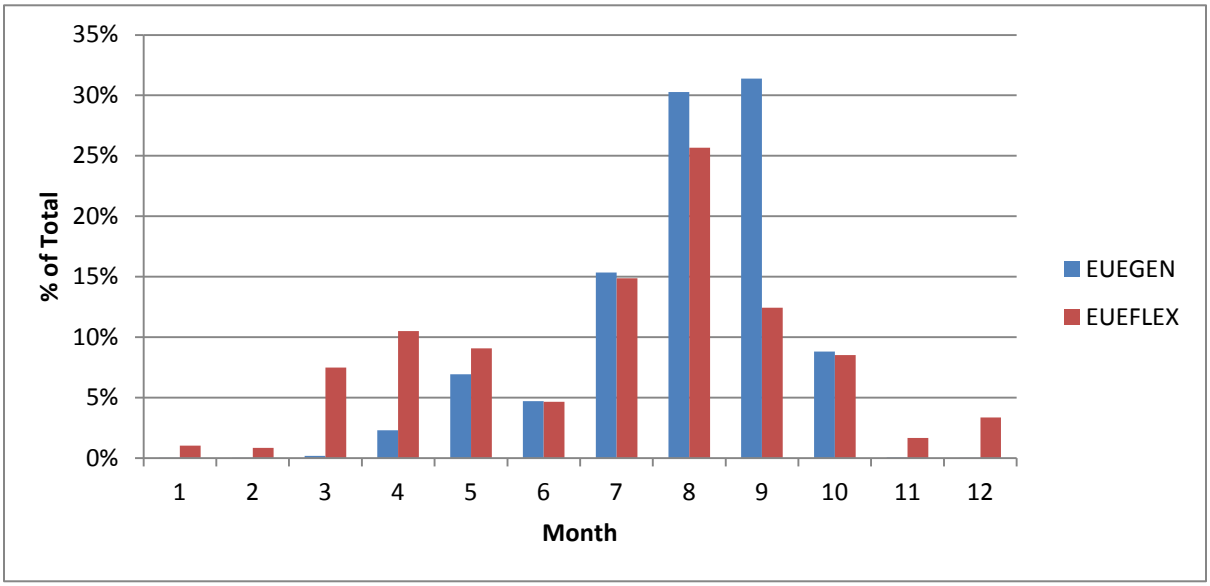


Figure 17 shows the EUE and LOLE across all scenarios for the CAISO Region. While $LOLE_{GEN}$ is 0.771 on a weighted average basis, there were scenarios modeled where it was as high as 5 events per year. Figure 18 shows the distribution of curtailed generation.

Figure 17. Distributions of LOLE and EUE Metrics

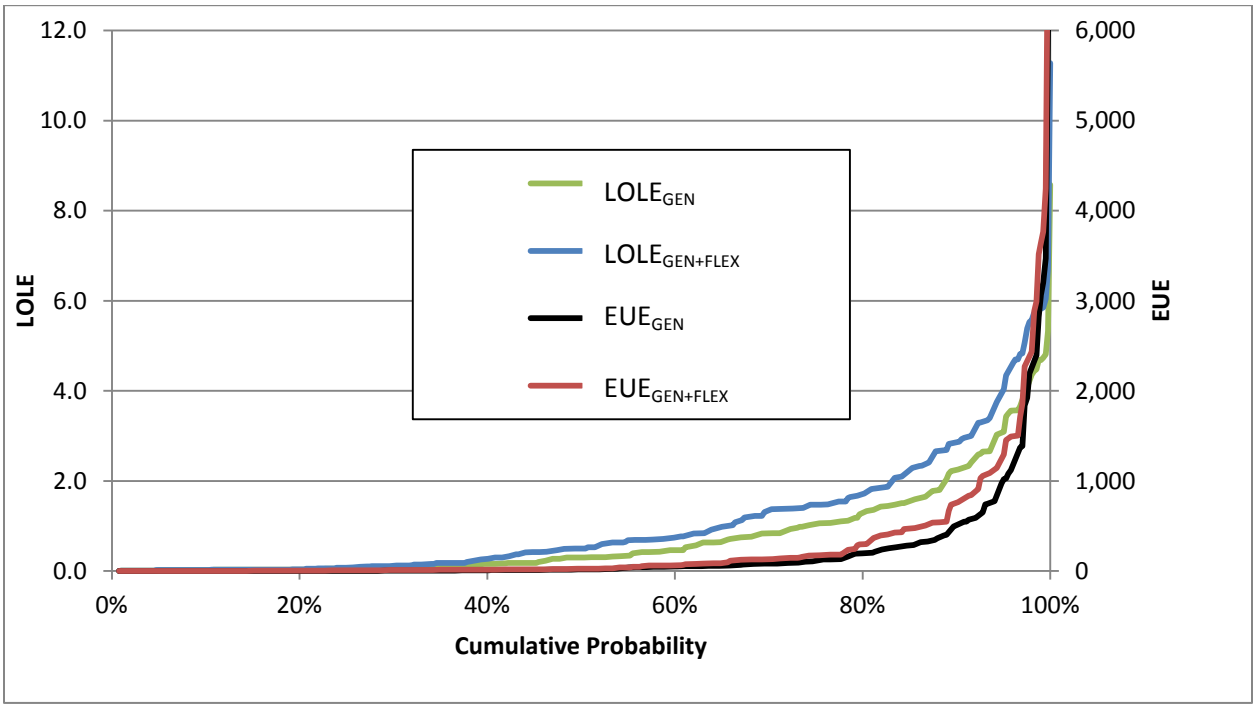
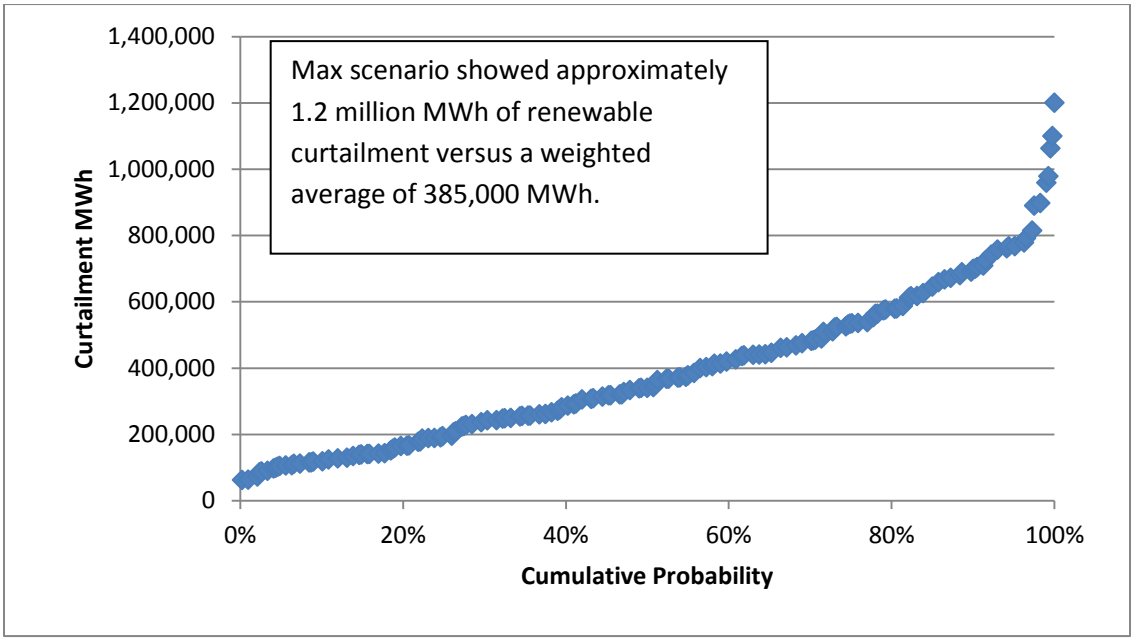


Figure 18. Distributions of Renewable Curtailment Based on Inflexible Hydro and Dedicated Imports



To achieve an $LOLE_{GEN+FLEX}$ of 0.1, approximately 1,900 MW of quick start CT capacity is needed which equates to an approximate 20.5% reserve margin. These results are shown in Table 9.

Table 9. Multi Weather Year Results with Additional CT Capacity Assuming No Economic Load Growth Uncertainty

		Analysis2: Multi-Year Weather Shapes	Base +1170 MW	Base +2340 MW
Target Reserve Margin	%	17%	19%	21%
$LOLE_{GEN}$	Events/Yr	0.543	0.123	0.042
$LOLE_{FLEX}$	Events/Yr	0.26	0.068	0.005
$LOLE_{GEN+FLEX}$	Events/Yr	0.803	0.191	0.047

V. Conclusions and Next Steps

Key Findings

- At a 17% reserve margin which includes all 13,000 MW of imports, $LOLE_{GEN+FLEX}$ in events per year is higher than 0.1[16]
- A significant proportion of the loss of load events occur in late afternoon hours after solar production has declined substantially and some demand response capacity has become unavailable. The loss of load metrics reduce significantly if demand response resources are available in hours 19-23.
- Assuming the no load growth uncertainty cases, a total of approximately 1,100 MW – 1,900 MW of CT capacity within CAISO was needed to achieve an $LOLE_{GEN+FLEX}$ of 0.1 which equates to approximately a 19-20.5% reserve margin.
- Increasing load following or allowing regions to commit quick start capacity intra-hour for a neighbor's need (especially within CAISO) likely removes the majority of the $LOLE_{FLEX}$ seen in the simulations.
- Renewable curtailment exists and can become problematic dependent on flexibility assumptions on dedicated imports and the hydro system across solar peaking hours

Next Steps

- Input Data Refinement
  - Hydro and Dedicated Import flexibility
  - Incorporate rest of WECC rather than simplified import method
  - Net Load Uncertainty Distributions (intra-hour, hour-ahead , and day ahead distributions for wind, solar, and load)

[16] Relaxing modeling constraints such as availability of intra-hour market purchases could impact this result.

    - Utilize separate multipliers for peak load and energy to reflect non-linear growth.

- The economic analysis produced by SERVM will be insightful when determining the best method to solve the reliability deficiencies found in these scenarios.
    - Increase reserve margin
    - Replace existing resources with flexible resources
    - Over commit resources
    - Change demand response program availability

VI. Appendix

Based on the 2012 LTPP Case without Songs the generation and capacity factor by unit type were compared between SERVM and PLEXOS to understand how the commitment and dispatch algorithms between the two models compared.

Figure A.1. Generation and Capacity Factor by Unit Type

	Generation (GWh)		Capacity Factor (%)	
	SERVM	PLEXOS	SERVM	PLEXOS
CCGT	54,801	57,958	40.0%	42.3%
GT	2,278	3,261	4.1%	5.8%
Demand Response	3	6	0.01%	0.03%
Nuclear	17,546	17,280	89.4%	88.1%
Coal	288	267	92.8%	85.9%
CHP	31,483	27,372	90.4%	78.6%
Renewable	67,012	67,008	28.4%	28.4%
Hydro*	35,776	32,847	58.1%	53.4%
Pump Storage	1,520	1,955	10.5%	13.5%
Net Imports	44,344	45,577	0.0%	0.0%

*The inputs related to total hydro generation were slightly higher in SERVM for this case. It is expected that if the corresponding PLEXOS case was simulated with the same hydro inputs that were used in SERVM, that the CCGT and GT generation would likely converge.

Figure A.2. Generation by Unit Type