CALIFORNIA PUBLIC UTILITIES COMMISSION: ENERGY DIVISION

FILED
11-16-15
11:17 AM

Proposed Revisions to LTPP Modeling Methodology

Energy Division Staff Proposal

Neal Reardon, Carlos Velasquez, Keith White & Patrick Young

11/6/2015

The California Public Utilities Commission's Energy Division staff feels it is imperative to revise the Long-Term Procurement Planning Modeling Methodology to develop more accurate and transparent information for decision makers. This "Staff Proposal" documents Energy Division staff's views on refining, validating, and standardizing the Long-Term Procurement Planning Models. Staff looks forward to receiving formal feedback on these views, which we expect will be adopted via Commission Decision.

TABLE OF CONTENTS

1. INTRODUCTION

In April of 2015, Governor Brown set a new goal under Executive Order B-30-15 that implemented a new interim statewide target to reduce greenhouse gas emissions to 40 percent below 1990 levels by 2030. In October 2015, Governor Brown signed Senate Bill (SB) 350, which, among other things, requires 50 percent of the electricity to come from renewable sources by 2030. These goals will help California reach its ultimate goal of reducing greenhouse gas (GHG) emissions to 80 percent below 1990 levels by 2050. Increasing the levels of renewable resources has the potential of increasing over-generation and flexibility needs in the electricity system. The current tools for determining the extent of these challenges and testing possible solutions are inadequate.

Billions of dollars will be spent procuring new renewable resources, new efficiency resources, and new resources to maintain grid reliability. Making decisions based on the current modeling tools or a best guess of what is expected to be "no regrets" procurement is unlikely to achieve the state's goals at an acceptable cost and may not maintain the level of reliability required by the state's economy. Therefore, the Energy Division proposes improving the modeling tools and validating the analysis results to better inform Commissioners and other decision-makers.

As part of the 2014 Long-Term Procurement Plan (LTPP) proceeding, Energy Division staff recommends revisions to the deterministic and stochastic modeling methodologies. Specifically, Energy Division staff's updated recommendations address methods to validate, refine, and standardize the models that will be used to examine the need for flexible and generic system resources in the 2016 and future LTPP proceedings. Staff feels it is imperative that the Commission utilize the standard modeling methodologies proposed in this document. Specifically, staff feels the recommendations details in Sections 3 and 4 should be adopted in Ordering Paragraphs of a Decision. We believe that parties will continue to act in good faith pursuing the model validation activities proposed in Section 5.

2. PROCEDURAL BACKGROUND

Biennially, the LTPP proceeding sets rules for utility procurement activities and evaluates the need for additional generation resources necessary to maintain reliability on a system-wide basis and in transmission-constrained areas (local capacity areas). Decision (D.) 14-03-004 in the 2012 LTPP[1] Rulemaking (R.) 12-03-014, authorized the procurement of resources in the LA Basin and San Diego local capacity areas in order to alleviate the local area needs resulting from the permanent closure of San Onofre Nuclear Generating Station and expected retirements of Once-Through-Cooling facilities located along the Southern California coast.

R.13-12-010 evaluated the need for additional flexible resources. Such resources are able to ramp up or down in response to fluctuating supply and demand to manage the increasing levels of distributed energy resources and variable energy resources (VERs) – e.g. wind and solar. In order to perform this evaluation, the LTPP proceeding made use of technical studies assessing the ability of the expected future generation fleet to meet future electricity demand. Studies were submitted by Southern California Edison (SCE) and the California Independent System Operator (CAISO). While the models used in these studies have been under development for several years, the 2014 LTPP proceeding sought to use them to inform flexible resource procurement decisions for the first time.

[1] Decision (D.) 14-03-004, issued March 14, 2014, is available online at: http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M089/K008/89008104.PDF

The results highlighted a potential need for additional flexibility in the electric system to follow both load fluctuations and variable energy resource output under a range of possible futures. However, the majority of the parties to the proceeding – including those who took part in the modeling efforts – cautioned the Commission about making procurement-related decisions based on the initial results of their studies. Instead, these parties advised the Commission to focus on altering the models to enhance their transparency and accuracy.

In March of 2015, the assigned Administrative Law Judge (ALJ) issued a Ruling setting forth issues for the remainder of the LTPP proceeding. The Ruling concluded that "there is not sufficient evidence at this time to authorize additional flexible or system capacity through 2024" and "there is both sufficient time and a critical need to further develop modeling efforts to inform the 2016 LTPP proceeding regarding the need for flexible capacity through 2026."[2] Pursuant to these findings, the Ruling recommended that the remainder of the 2014 LTPP focus on developing and validating models that can "accurately highlight and distinguish needs for both flexible and generic system resource attributes to maintain reliability, to investigate efficient solutions to potential operational flexibility events (such as over-generation events), and to set the stage for expanded future analyses which will balance the cost-effectiveness and GHG impacts of measures to ensure system reliability."

Furthermore, the Ruling directed Energy Division staff to host technical working groups on three topics:

1) Developing common definitions, metrics and standards

2) Identifying standard outputs

3) Validating stochastic and deterministic models and making technical improvements.

Energy Division staff invited all interested parties to participate in a series of technical discussions held from April through June of 2015. For each of the topic areas, Energy Division staff solicited input from the working group participants in order to establish a roadmap for using the models in Commission decision-making. Staff also requested that the participants identify necessary improvements that could be made to the models within the 2016 LTPP timeframe. A preliminary Staff Proposal was shared with parties to the LTPP proceeding on July 27, 2015 and was presented and discussed during an August 4, 2015 Energy Division staff workshop. Staff solicited informal comments from parties via an August 6, 2015 email and received informal feedback by the August 13, 2015 due-date.

This document describes the outcome of these technical working group discussions, and the resulting, revised, Energy Division staff recommendations on the three topic areas identified above. The post-workshop revisions that are incorporated into this document are based on 1) the reconsideration of the general feedback that we received during the working group calls, 2) the feedback received at the August 4, 2015 workshop in which we presented our initial Staff Proposal and 3) the informal post-workshop comments. **This Staff Proposal does not attempt to describe all positions and issues that were raised during the working group discussions, at the workshop, or via informal comments; rather, it focuses on the issues staff believes are appropriate for Commission consideration.**

[2] Administrative Law Judge's Ruling discontinuing Phase 1a and setting forth issues for Phase 1b, Issued March 25, 2015, available online at: http://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M148/K825/148825409.PDF

3. DEFINITIONS, METRICS, AND STANDARDS

3.1. Background

The LTPP proceeding is tasked with planning for the long-term reliability of the electric system by evaluating the need for new resources to maintain a sufficiently reliable level of service.[3] Given the expectation of increasing levels of VERs (e.g. wind and solar generation) as well as distributed energy resources to achieve California's long-term policy goals, the electric system needs to be evaluated to determine if it will have sufficient flexibility to maintain reliable service. Flexibility may be regarded as the ability of electric resources to continuously balance electricity supply and demand given the uncertainty and variability of net load. Net load is the load remaining after accounting for wind and solar generation. Flexibility needs differ by season, hours of the day, and even within hours – and the electric system may require more flexibility at higher levels of VERs, absent other possible changes in operations, electricity infrastructure, and electricity consumption patterns.

The 2014 LTPP proceeding, Phase 1A, served as the forum in which the Commission continued investigating how to supplement existing planning paradigms in order to evaluate both generic system reliability and operational flexibility needs in a future with increasing levels of VERs.[4] This investigation continues in Phase 1B. In Phase 1A, parties provided testimony on several deterministic and stochastic technical studies that simulated future system performance under a range of conditions in order to determine if the system needed additional capacity and/or flexibility to integrate higher levels of VERs while continuing to reliably serve future demand.

The deterministic technical studies were conducted using production cost simulations. These simulations modeled electric grid operation by economically dispatching resources to balance supply and demand for all hours of a study year under an assumed, fixed (i.e. deterministic) set of expected future conditions and operational constraints. During a production cost simulation, for any hour during which the model could not dispatch the system to balance supply and demand, that hour and the magnitude of imbalance was recorded as an "event." In the case of an over-supply of energy, the simulation recorded the event as an "over-generation"[5] event. In the case of an under-supply of energy, the simulation recorded the event as a "loss of load" event. The magnitude of the largest "event" and/or the cumulative total of "events" over the study year were viewed as potential measures of system deficiencies. Varying definitions of the conditions under which an "event" gets recorded affected the study's assessment of system resource deficiency, which subsequently affected the amount of necessary new resource procurement to fill the deficiency. LTPP Parties in Phase 1A disagreed on the appropriate definition to use for an "event."

The production cost simulations included granular operational details for generation resources across the entire WECC[6] in order to reflect realistic grid operations as closely as possible, given the fixed assumptions. Because the deterministic technical studies only modeled fixed assumptions, they are not able to provide an analytical assessment of the probability of its outcome.

Stochastic technical studies were also reviewed in Phase 1A of the 2014 LTPP. Stochastic modeling techniques were generally viewed as more robust than deterministic techniques because they analyze the probability of outcomes under both average and extreme conditions. The computational intensity required by stochastic

---

[3] See "Scoping Memo and Ruling of Assigned Commissioner and Administrative Law Judge" in docket R.13-12-010, May 6, 2014, pp. 3-5

[4] Consideration for the need of operational flexibility began in the 2010 LTPP.

[5] Over-generation may be caused by inability to dispatch the system downward to balance supply and demand or surplus non-dispatchable generation (e.g. wind and solar) within a particular time horizon.

[6] The Western Electric Coordinating Council (WECC) administers electric reliability standards in the Western part of North America. The acronym "WECC" is often used as shorthand to refer to the geographic area representing the interconnected synchronous electric grid in the Western part of North America.

modeling, however, meant that the modeler had to limit the amount of detail and granularity in the model in order to complete the studies in a reasonable amount of time.[7]

Like the deterministic studies, the stochastic technical studies also incorporated production cost simulations but additionally used Monte Carlo simulation techniques to simulate electric grid operation under a wide range of potential future conditions (ideally the universe of all physically possible future conditions). The stochastic studies were designed to explicitly analyze the probability of outcomes and therefore could be compared to a probabilistic (or risk-based) reliability standard such as the well-known "one day in ten years loss of load" industry standard.[8] If the studies reported that the probability of "loss of load" exceeds the "one day in ten years" standard, then that was an indication of potential system deficiency – meaning the reliability of the system is below standard, and a specific amount of new capacity would need to be added to the system to reduce the probability of "loss of load" (or increase the reliability) until it meets the "one day in ten years" standard.

Stochastic studies can quantify the reliability of the system using several different probabilistic metrics such as Loss of Load Expectation (LOLE), Loss of Load Hours (LOLH), Loss of Load Probability (LOLP), and Expected Unserved Energy (EUE).[9] Each metric differs on the dimension of reliability that it is best suited to measure, for example, the frequency of events, the hours of events, the amount of unserved energy from events, and the probability of events. While the electric utility industry has relied on the "one day in ten years" standard as a reliability guideline, industry consensus has not been established on how to apply the standard to any one of these probabilistic metrics, or whether the focus should be placed on the frequency, the hours, the amount of energy, or the probability of "events". Given that the choice of metric to compare to the standard affects the calculation of system resource deficiency and potential resource needs, naturally the LTPP parties differed on which metric should be compared to the standard.

Similar to the deterministic technical studies, the stochastic studies also needed to define the conditions under which an "event" gets recorded, and the choice of definition likewise affected the assessment of system resource deficiency. The LTPP Parties disagreed on the appropriate definition to use for an "event" in the stochastic studies as well.

In addition to choosing a metric for measuring reliability and defining the conditions under which an "event" occurs, stochastic technical studies must also define when a "day" has occurred in order to calculate a result that can be compared to the "one day in ten years" standard. For example, a model could choose to record a "day" has occurred when 24 cumulative hours of "events" have occurred, or alternatively record a "day" has occurred if any one hour of a day contained an "event." These substantially different definitions of when a "day" should be counted in a stochastic study materially affect the assessment of system resource deficiency. There is no industry consensus on how to count a "day," and different parties suggested different definitions to use for "day."

Working Group One was tasked with recommending common definitions, metrics, and standards as they relate to the deterministic and stochastic technical studies introduced above. The remainder of this section details the issues that were discussed and proposes a recommendation for each.

---

[7] For further background on the comparative features of deterministic and stochastic models, see this review of planning models related to studies done during the 2012 LTPP proceeding. http://www.cpuc.ca.gov/NR/rdonlyres/ECE43E97-26E4-45B7-AAF9-1F17B7B77BCE/0/CombinedLongTermProcure2014OIR_Report_CollaborativeReview.pdf

[8] The "one-day-in-ten-years" standard generally means that the electric system shall be planned such that electric service interruptions due to inadequate supply are expected to occur only once in ten years. An electric system that meets this standard generally has available capacity in excess of the highest expected electricity demand by a certain margin; that margin is known in the industry as the reserve margin or planning reserve margin (PRM). For decades, the electric utility industry has relied on a "one-day-in-ten-years" loss of load reliability standard when planning to maintain a reliable electric infrastructure

[9] For further background on probabilistic reliability metrics and the industry 1-in-10 resource adequacy standard, see this FERC report: http://www.ferc.gov/legal/staff-reports/2014/02-07-14-consultant-report.pdf

3.2. Metrics and Standards

Metrics and standards as used in the following discussion are defined as follows. A metric is a measurement of the reliability level expected from an electric system. A variety of outputs from a resource planning model can be chosen as metrics for system reliability. A standard is a chosen value of a metric deemed to represent reasonably reliable electric service at reasonable cost. If the calculated metric from a resource planning model does not meet the standard, then some amount of new capacity (or other changes to the electric system such as changing demand) must be added to the system to increase reliability until the calculated metric meets the standard.

3.2.1. Deterministic Metrics and Standards

For the deterministic studies in the 2014 LTPP, the metric is the production cost simulation's recording of the magnitude of "events" over the course of the year being simulated. For example the production cost simulation recorded the MW of system imbalance, if any, for all hours of the year. The standard would be set by deciding what number of "events" is allowed to occur in the simulation. For example, the standard could be that the production cost simulation must complete without any "events" occurring (or just a small limited number). If the simulation completed with some hours where "events" occurred, then the amount of system deficiency would be determined by the magnitude of the largest event. Adding this magnitude of capacity to the system would bring the system up to standard. This is the implementation of metric and standard generally used and recommended by the CAISO for deterministic technical studies, i.e. single production cost simulations for a study year.

Regarding choosing standards for the deterministic models, one must distinguish between the two types of "events" being captured: loss of load, and over-generation. Loss of load represents an actual system deficiency and reliability problem. The standard for maximum acceptable "loss of load events" may appropriately be that no such "events" should occur in a production cost simulation (or just a small limited number). Over-generation on the other hand is not necessarily a reliability problem if the simulation is allowed to curtail non-dispatchable generation as needed, particularly wind and solar generation. There is no standard for how much generation could be curtailed since that decision is generally based on economics. Therefore, while the standard for maximum acceptable "over-generation events" may appropriately be that no such "events" should occur in a production cost simulation (or just a small limited number), there is no standard for how much curtailment can be used to remove over-generation. The Over-generation Events section (3.3.2) later in this document expands on the distinction between over-generation and curtailment.

The CAISO and other parties using deterministic technical studies have generally followed the framework just outlined above, but there is significant disagreement on when to count that an "event" has occurred, regardless of whether that "event" was loss of load or over-generation. The Working Group One discussion on the appropriate "event" definition is covered in the Event Definitions section (3.3) below.

3.2.2. Stochastic Metrics and Standards

Stochastic studies of the type used in the 2014 LTPP can report various metrics that are comparable to the probabilistic reliability standard "one day in ten years." However, the comparison is subject to varying interpretations depending on how one defines a "day" and depending on which metric is used for comparison. Stochastic studies can measure system reliability using the following metrics: Loss of Load Expectation (LOLE), Loss of Load Hours (LOLH), Loss of Load Probability (LOLP), and Expected Unserved Energy (EUE). Additionally, for comparison to a "one day in ten years" standard, one must decide how to count a "day." For example, one can count a single "event" within an hour[10] as the occurrence of a "day" or one can count a "day" when 24 cumulative hours of events have occurred. There was no consensus in Working Group One regarding which metric and interpretation of a "day" should be compared to the "one day in ten years" standard. But this is an important area

[10] Note that all models considered and discussed in this document are hourly dispatch models. Each model makes assumptions about what is likely to happen within a given hour, but results are recorded on hourly intervals. So if the model records that an event occurred for a given hour, it is as if the event persisted for the entire hour.

to understand and seek agreement on because the choice of metric and interpretation of a "day" can significantly affect the calculated reliability level and hence the necessary reserve margin of generating capacity to meet a reliability standard such as "one day in ten years."[11]

The Loss of Load Expectation (LOLE) metric quantifies the expected frequency of "events". An example value is 0.1 "events" per year.[12] This example could be compared to the "one day in ten year" standard by multiplying by ten to yield a value of one "event" per ten years. One must also decide how to map "events" to "days." For example, a "day" could signify any day where one or more "events" happened within that day, regardless of the magnitude, duration, or number of "events" as long as occurring within the same day.[13] SCE's stochastic studies measured system reliability using the LOLE metric and used the above example definition of "day."[14] However, SCE uses the term "event" instead of "day" with respect to the "one day in ten years" standard, and appears to treat the two terms equivalently in its studies. The CPUC's Resource Adequacy proceeding (R.14-10-010) also uses the LOLE metric and the above example definition of "day" within its modeling framework for measuring system reliability performance.[15]

The Loss of Load Hours (LOLH) metric quantifies the expected number of hours of "events." An example value is 2.4 hours (of "events") per year. For comparison to the "one day in ten years" standard, one could interpret "day" to be any day where at least one hour of "event" occurred. One could also interpret "day" to have occurred only when 24 cumulative hours of "events" have occurred, regardless of the day or time those occurrences happened or the magnitude of those occurrences. The former interpretation implies that adherence to the "one day in ten years" standard means that measured LOLH should amount to no more than one hour in ten years, or no more than 0.1 expected hours per year. The latter interpretation implies that adherence to the "one day in ten years" standard means that measured LOLH occurrences should amount to no more than 24 cumulative hours in ten years, or no more than 2.4 expected hours per year.

The Expected Unserved Energy (EUE) metric quantifies exactly what its name says, the expected amount of loss of load (unserved energy). EUE is usually normalized to represent the total expected loss of load due to supply shortages, divided by the total system net energy for load (net electricity generation to serve load over a year), and therefore represents an overall expected percentage of system load that cannot be served over the study year. The typical standard for normalized EUE is 0.001%.[16] Normalized EUE is sometimes also called LOLP – Loss of Load Probability because it in fact represents a probability of loss of load for the system under study.

---

[11] This FERC report illustrates how different metrics and definitions of "day" lead to different reserve margins necessary to meet a "one day in ten years" standard: http://www.ferc.gov/legal/staff-reports/2014/02-07-14-consultant-report.pdf

[12] The concept of a fractional "event" per year may be difficult to grasp. Keep in mind that this value represents a probabilistic expectation of an "event" over one year. Stochastic simulations typically simulate thousands of years to calculate how many "events" occur across all those years. When the expected value is scaled down to "the expected number of events that might occur in any one year," the answer could be fractional.

[13] A "day" under such an interpretation would be deemed to have occurred even if just a single hour of "event" occurs, but another "day" would be deemed to have occurred only if at least one "event" occurred on a different day.

[14] See Second Revised Phase 1A Testimony of SCE on Resource Need in R.13-12-010, November 20, 2014, p. 11. http://www3.sce.com/sscc/law/dis/dbattach5e.nsf/0/59050DA733479E4288257D98000594CC/$FILE/R.13-12-010_2014%20LTPP-SCE%20Addtional%20Phase%201a%20Testimony.zip

[15] See "Probabilistic Reliability Modeling Inputs and Assumptions", July 15, 2015 here: http://www.cpuc.ca.gov/NR/rdonlyres/54510A14-B894-4E05-A933-3665DA72F060/0/ProbabilisticReliabilityModelingInputsandAssumptionsPartTwo.pdf

[16] The typical standard of 0.001% was described in the FERC report describing probabilistic reliability metrics cited earlier.

The EUE metric does not correspond well to the "one day in ten years" standard since it does not include a time dimension, and hence, does not incorporate a definition of a "day." However, the EUE metric measures the magnitude of an "event," which is a dimension not captured simply by examining LOLE or LOLH metrics. Therefore, it may be informative to consider the EUE metric alongside the LOLE metric (which measures frequency of "events") and the LOLH metric (which measures the hours of "events"). The results have the potential of revealing several "event" dimensions that would otherwise be lost relying solely on a single metric.

3.2.3. Working Group Views on Metrics and Standards

Working Group One participants presented the following opinions on how to define "day" and which metric to use for testing system reliability against the "one day in ten years" standard.

CAISO:
In the context of stochastic modeling, the California Independent System Operator (CAISO) proposed LOLH as a primary metric that should be used and defined a "day" as consisting of seven cumulative hours regardless of when the loss of load hours occurred. The rationale for defining a "day" as consisting of seven hours is that, when modeled in an hourly simulation, a loss of load occurrence would most likely occur in just some hours of the day, not during an entire day. The CAISO referenced Dr. Roy Billinton's work[17] stating that it used a ratio of reliability measurements from a hypothetical test system to estimate how many loss of load hours might be expected to occur on a day with loss of load events. The measurements were 9.39 hours/year of loss of load and 1.37 days/year of loss of load. The ratio of the two, 6.87 is close to seven.[18] Thus the recommended interpretation of the "one day in ten years" standard is that LOLH should not exceed "seven hours in ten years."

ORA:
The Office of Ratepayer Advocates (ORA) indicated that Working Group One "…provided a broad outline of the resource adequacy reliability metrics currently in use, including loss of load expectation (LOLE), loss of load hours (LOLH), loss of load probability (LOLP), and Expected Unserved Energy (EUE). While the electric power industry continues to use the "1 day in 10 year" rule of thumb as a reliability guideline, there has been no industry consensus as to which reliability metric is most useful or whether the focus should be on events, hours, energy, or a probability. The choice of metrics and its interpretation could easily swing the target reserve margin between 8 percent and 17 percent, depending on the characteristics of the system. Working Group One did not have sufficient time to analyze this issue in depth – further work is required to understand the most appropriate metric for modeling."[19] ORA also expressed concern that CAISO's proposal to use LOLH with a definition of seven cumulative hours to represent a "day" has not been thoroughly vetted or justified for use in CPUC decision-making. ORA also raised the importance of considering costs associated with any particular standards, and if those costs reasonably reflect the value placed on that level of reliability. ORA suggested that the discussion of adequate reliability standards would benefit from further discussion of how the "value of lost load" could be compared to the costs to procure capacity to prevent "loss of load." This would inform target values for the chosen metric, which the Commission could use as a benchmark for procurement.

TURN:
The Utility Reform Network (TURN) did not propose a specific interpretation of "day" but noted that "…the definition of 'day' in the 'one day in ten years' standard can range from an insufficiency of generation within a single hour to insufficiencies over 24 separate hours within a simulation. Some prefer the criterion be expressed

[17] R. Billinton, R. N. Allan, Reliability Evaluation of Power Systems, Plenum Press, New York and London (second edition, 1996), pp. 384, 484.
[18] CAISO Response to the Second Set of Data Requests Related to Phase 1A of the Office of Ratepayer Advocates in Docket No. R.13-12-010, July 1, 2015.
[19] See attachment to ORA's email to Working Group One participants on June 24, 2015.

as 'one event in ten years', in which an 'event' is an individual occurrence of loss of load that may range from one to many consecutive hours."[20]

Regarding metrics and measuring the reliability stressors of capacity, lack of flexibility, and over-generation, TURN suggests that it is "…important that reliability metrics reflect the specific model(s) being employed. For example, criteria based on deterministic simulations – which are based on a single view of a system's loads and resources – should be different from those based on stochastic simulations – which are intended to perform numerous simulations reflecting variations in key uncertainties about future reliability. As another example, some models attempt to consider all of the above possible reliability stressors in a single model. But it is not clear that the above three possible stressors – lack of capacity, lack of flexibility and over-generation – can be practically analyzed in a single model."[21]

TURN also suggests it is "…premature to adopt a rigid reliability metric related to instances of over-generation. TURN is yet unaware of any precedents for such criteria in the industry. Given the differences in the challenges posed by a lack of capacity and an over-abundance of energy, it also does not seem appropriate that the importance of the two events be measured using the same probability, that is by the terms '7 hours-in-10 years' or '1 hour-in-1 year'."[22]

TURN in fact proposes an alternate reliability modeling framework that uses metrics tailored to the specific model(s) being employed and the specific question being answered. The Appendix describes the details of TURN's proposal.

UCS:
The Union of Concerned Scientists (UCS) believes that "Expected Unserved Energy (EUE)[23] is an attractive reliability metric because it incorporates both hourly and sub-hourly loss-of-load events. EUE should be quantified and considered alongside metrics that quantify loss of load frequency. Reliability metrics that focus on a certain frequency of loss-of-load events encounter difficulties when faced with short timescale loss-of-load events. Recent LTPP modeling efforts have simulated the electricity system on an hourly timescale, requiring translation of the traditional one-day-in-ten-years reliability metric to an hourly metric. The use of EUE as a reliability metric would make it possible to change temporal resolution without determining a new reliability metric. For example, if 5 minute simulations are run in future LTPPs, then it would be necessary to discuss how many 5 minute shortfall events are equivalent to one hour or one day. EUE does not have this problem as it includes shortfall duration no matter the timescale of the simulation. This represents an important first step in being able to evaluate economic tradeoffs between different resources." [24]

Wellhead:
Wellhead Electric indicates that "the discussion of Reliability and Standards needs to include metrics that reasonably ensure the CAISO is able to operate the electric system in full compliance with "reliability/operating criteria" (ROC) *and* without contravening California's environmental policies and goals. This effort will require consideration/understanding of the various factors that are driving the need for significantly more flexibility in the generating fleet than exists today. Foremost is the inclusion of uncertainty/variability (also called "forecast error") of renewable generation. In this regard, intra-hour forecast error is most important because the response cannot

[20] See attachment to TURN's email to Working Group One participants on June 24, 2015.
[21] Ibid.
[22] Ibid.
[23] UCS defines EUE as a measure of the transmission system's capability to continuously serve all loads at all delivery points while satisfying all planning criteria. EUE is energy-centric and analyzes all hours of a particular year and results are calculated in Megawatt-Hours (MWh). EUE is the summation of the expected number of MWh of load that will not be served in a given year as a result of demand exceeding the available capacity across all hours. Additionally, this measure can be normalized based on the assessment area's total Net Energy for Load. Normalizing the EUE provides a measure relative to the size of a given assessment area.
[24] See attachment to UCS's email to Working Group One participants on June 24, 2015.

wait—it must be fast/now, not in the next hour or two. In fact, the problem can appear and disappear within an hour meaning an hourly analysis will completely overlook the problem that required a rapid response. As such, the criteria need to take account of all events when ROC is reasonably expected to be violated. The magnitude, duration, and frequency of the ROC violation events (are) important with regards to identifying/selecting the best solution(s). However, there can be no tolerance for planning that reasonably shows ROC violations or contravention of California policies."[25]

3.3. Event Definitions

In both deterministic and stochastic technical studies, precisely defining the conditions under which an "event" gets recorded in a given simulation can affect the study's assessment of system resource deficiency, and subsequently affect the amount of necessary new resource procurement to fill the deficiency. Unfortunately, there is no industry consensus on the precise system conditions under which an "event" should be triggered, whether that "event" is loss of load or over-generation.

In order to further discuss what should be the appropriate definition for both "loss of load events" and "over-generation events," we must first define operating reserves. Operating reserves in this context can be understood to be some amount of extra capacity held in reserve to handle the inability to perfectly predict and match supply and demand for a given hour and within the hour. Operational reserves generally include the following categories: contingency reserves, regulation up/down, and load following up/down. Contingency reserves consist of spinning and non-spinning reserves and are meant to handle "contingency" situations such as the sudden loss of a large generator or a transmission line. Regulation and load following reserves are meant to handle forecast error and intra-hour variability and uncertainty in net load (recall that net load is the load remaining after accounting for wind and solar generation). Spinning and non-spinning reserves, and regulation-up/down services are collectively also referred to as ancillary services.

3.3.1. Loss of Load Events

The condition for a "loss of load event" could be defined as when expected supply for a specific hour falls below expected demand for that same hour. Alternatively, the condition for a "loss of load event" could be defined as when expected supply for a specific hour falls below expected demand for that same hour plus a set level of operating reserves. In other words, this latter definition recognizes a grid operator's discretion to shed some load under stress conditions in lieu of fully exhausting operating reserves and perhaps compromising the operator's ability to balance the grid. Studies from other regions around North America have used a range of definitions for what constitutes a "loss of load event."[26] Some choose to define a "loss of load event" as any event requiring emergency operating procedures, such as calling demand response, implementing voltage control, or depleting operating reserves. (In other words, the "event" is triggered upon the condition of having to start to use some operating reserves.) Others choose to record a "loss of load event" only if operating reserves are depleted and firm load must be shed.

The typical production simulation model scheduling and dispatching with hourly granularity attempts to balance the system hourly by dispatching from least costly to most costly dispatchable resources. In the CAISO's implementation of deterministic studies using the PLEXOS production cost simulation modeling platform, the simulation co-optimizes for energy, ancillary services, and required load following reserves to achieve minimum cost solutions. Under stress conditions, the model follows a priority order (similar to that in the CAISO market scarcity pricing mechanism) for use of reserves and ultimately load shed under shortfall conditions (available supply is less than demand). Under such conditions, load following-up reserves are depleted first, followed by non-spinning, spinning, regulation-up, and finally load shed. For example, if there is a shortfall, the shortfall occurs

---

[25] See body of Wellhead Electric's email to Working Group One participants on June 24, 2015.
[26] See the FERC report http://www.ferc.gov/legal/staff-reports/2014/02-07-14-consultant-report.pdf at pp. 14-15 and A-3 for reliability event definitions used in different regions of North America.

first in load following-up reserves. If the shortfall is large enough, it will spill over to non-spinning, spinning, regulation-up and finally to load shed (implying actual unserved energy).[27] The threshold for deciding when a "loss of load event" occurs for the purpose of reliability planning could be set at any point in the above modeling sequence, for example, as soon as load following-up reserves need to be used, or only when actual load shed occurs.

3.3.2. Over-generation Events

Over-generation is the other type of "event" being measured in the production cost simulations used in the LTPP technical studies. In this case, the stress condition is not undersupply or shortfall, but a potential inability to balance the grid due to oversupply. If over-generation is projected for the next hour, the model will ramp down resources that can economically dispatch down. In general, most dispatchable resources are priced in models to dispatch down before non-dispatchable resources are curtailed (i.e. renewable curtailment). In the CAISO's implementation of deterministic studies using the PLEXOS production cost simulation platform, renewable curtailment is set with a penalty price of $ -300/MWh (representative of the bid floor in the expected future CAISO market). Load-following-down and regulation-down are priced at $ -650/MWh and $ -1000/MWh respectively in the model to clearly establish a dispatch down priority order.[28] Assuming renewable curtailment is sufficient to handle over-generation, load following-down or regulation down reserves would not be dispatched.

If renewable curtailment is restricted in part or completely (i.e., available economic renewable curtailment insufficient to handle over-generation), the model would then dispatch downward flexibility reserves (load-following-down and regulation-down in that order) and record the amount and type of downward flexibility reserves used. This is recorded as load following-down and regulation-down "shortfalls" (recall that downward flexibility reserves should normally be reserved to address intra-hour variability and uncertainty of net load). If all the downward flexibility reserves get used and there is still over-generation remaining, the model would record this as "unsolved over-generation" (the PLEXOS model records occurrences of this condition in a variable called "dump energy"). This condition means all dispatchable resources have been dispatched down, all curtailable renewable generation has been curtailed, and all load following-down and regulation-down has been used, but there is still excess supply.[29] Alternatively, the threshold for deciding when an "unsolved over-generation event" occurs for the purpose of reliability planning could be set for when load following-down reserves start to be used (in other words, including load following-down and regulation-down "shortfalls").

3.3.3. Working Group Views on Event Definitions

Within the Working Group One discussions, the CAISO and TURN provided detailed recommendations regarding when a "loss of load event" should be deemed to have occurred, and also when an "over-generation event" has occurred.[30] These recommendations are described below along with feedback provided by other parties:

---

[27] This model description was adapted from CAISO testimony of Dr. Shucheng Liu served in Docket No. R.13-12-010, August 13, 2014, pp. 10-12.

[28] But, these prices are only used to get the correct dispatch order in this type of model – the real CAISO market model uses a second run called the "pricing" run to calculate real resource costs. In fact, scheduling coordinators economically bid curtailment of renewable generation into the CAISO market at prices other than what is modeled in the PLEXOS resource planning exercise.

[29] Note that this case assumes renewable curtailment has been limited in some fashion, so if the over-generation condition is large enough, the model may not be able to remove all the over-generation with curtailment. The remaining over-generation gets captured in load following-down shortfall, regulation-down shortfall, and finally "dump energy."

[30] Working Group One participants differed on whether "over-generation events" could cause a reliability problem or whether they should be measured in the same manner as "loss of load" events. Nonetheless, there seemed to

CAISO:
In the context of deterministic modeling, the CAISO proposed a more stringent[31] definition of an "incident"[32] than for its stochastic modeling. "Incidents" of capacity shortfall under the CAISO's deterministic modeling methodology would include any loss of load following-up reserves greater than 50% of its requirement, any loss of contingency reserves, any loss of regulation-up, or any unserved energy. "Incidents" of over-generation would include any loss of load following-down reserves greater than 50% of its requirement, any loss of regulation-down, or any "dump energy."

Based on the CAISO's definition of an "incident," if a particular simulated hour had a load-following requirement of 500 MW, an "incident" would be reported for that hour if the available capacity[33] for that hour was less than the forecast load plus all regulation-up, all contingency reserves, and 50% of load following-up reserves (250 MW). Note that these simulations mimic the CAISO market scarcity pricing mechanism in terms of priority order for dispatching reserves when the system is short in a given hour.[34] For example, under the CAISO's deterministic modeling methodology, when there is an upward shortfall, resources providing load-following-up would be depleted first. If the shortfall is large enough, resources providing non-spinning, spinning, and regulation-up would be depleted, in that order, before unserved energy (loss of load) occurs. An "incident", however, would be triggered whenever the simulation encounters a shortage situation for a given hour and it uses up more than half the load-following reserves to deal with the shortage. If the situation was over-generation,[35] load-following-down resources would be depleted first, followed by the depletion of resources providing regulation-down; "dump energy" ("unsolved over-generation," as described earlier) would occur last.

In the context of stochastic modeling, the CAISO proposed that "incidents" of capacity shortfall would include any loss of contingency reserves greater than 3% of load,[36] any loss of regulation-up, or any unserved energy. "Incidents" of over-generation would include any loss of regulation-down or any "dump energy."

As Energy Division staff understands CAISO's proposal for stochastic modeling, a "loss of load event" definition incorporating any loss of contingency reserves greater than 3% of load essentially means any loss of spinning reserves (which are typically 3% of load). This choice of definition seems to derive from the need to comply with the Stage 3 Emergency operating instructions issued by the CAISO during "a System Emergency and any circumstances in which the CAISO considers that a System Emergency is imminent, anticipated or threatened, as per the CAISO Tariff."[37] The CAISO "Emergency Stage 3" operating procedures state "the CAISO issues an Emergency Stage 3 when the Spinning Reserve portion of the Operating Reserve depletes, or is anticipated to deplete below the WECC Operating Reserve requirement and cannot be restored. The *WECC Operating Reserve requirement states that Spinning Reserve shall be no less than 50% of the total Operating Reserve requirements."*[38]

---

be some level of agreement that technical studies assessing the frequency, magnitude, and duration of over-generation events should use a common definition of when an over-generation event has occurred.

[31] "Stringent" in the sense of requiring more capacity to be reserved to handle uncertainty and variability not explicitly modeled in a deterministic study with fixed inputs.

[32] The CAISO's proposal chose to use the term "loss of load incident" as opposed to a "loss of load event."

[33] "Available capacity" in this context means capacity available for dispatch to provide energy and ramping as needed to follow load plus the required operating reserves.

[34] See CAISO testimony of Dr. Shucheng Liu in docket R.13-12-010, served on August 13, 2014, pp.11-12.

[35] Assuming economic renewable curtailment was used up and there is remaining over-generation.

[36] The CAISO updated its stochastic modeling definition of a loss of load incident after the Staff Proposal was issued on July 27, 2015. The CAISO's original stochastic modeling definition included *"loss of contingency (operating) reserve, loss of regulation-up, or unserved energy"*; its revised definition now includes *"loss of contingency (operating) reserve **greater than 3% of load**, loss of regulation-up, or unserved energy."*

[37] http://www.caiso.com/FASTSearch2/Pages/Results.aspx?k=4420%20System%20Emergency%20-%20California%20ISO, PDF "4420 System Emergency", page 2 of 17.

[38] http://www.caiso.com/FASTSearch2/Pages/Results.aspx?k=4420%20System%20Emergency%20-%20California%20ISO, PDF "4420 System Emergency", page 11 of 17.

(Italicized emphasis added) Note that if the CAISO's tariffs deemed that 3% "operating reserve requirement" were sufficient for a reliable system, the effective spinning reserves could fall to 1.5% and CAISO would still comply with the WECC Operating Reserve requirement. Conversely if CAISO's tariff deemed 12% "Operating Reserve requirements" were sufficient for a reliable system, the effective spinning reserves would need to be 6%. Staff understands CAISO's operating reserves currently amount to roughly 6%: ~3% spinning and ~3% non-spinning; 6% does not seem to be directly mandated by WECC.

TURN:
According to TURN, the definition of an occurrence of "loss of load" can range from defining such occurrences whenever modeled capacity falls below "load plus zero percent" – a common, well-documented practice – to a situation in which the modeled capacity falls below "load plus 'X' percent," where "X" is greater than zero (a seemingly less common practice).[39] Regardless of the value of "X," the traditional application of the criterion assumes that firm load will not be met under such circumstances due to a lack of capacity and does not consider whether load might be interrupted due to a lack of flexibility.

Criteria for measuring a lack of flexibility in the system are not, to TURN's knowledge, well-established or accepted. As a first step the type of flexibility being sought must be defined, followed by defining the consequences of not having such flexibility on the system, which may not necessarily lead to a loss of load event (as would obviously happen when capacity is insufficient). It is conceivable that a lack of a particular type of flexibility could lead to a loss of load on a "one-to-one" basis, that is, that load will be curtailed on a "MWh-for-MWh" basis for every MW or MWh of flexibility not available. However, it is also conceivable that curtailments of load merely become more likely when flexible capacity is insufficient, and further that such likelihood could be quite small. Defining criteria for flexibility needs thus requires a more careful examination of the type of flexibility desired and the consequences of its absence.

TURN did not propose a specific definition for a "loss of load event," but does state that "the CAISO proposal would also define 'loss of load' in a manner far more stringent than any other definition known to TURN, seemingly far beyond current industry practice." Based on illustrative calculations provided by TURN, staff interprets CAISO's proposed event definitions to mean that "loss of load" occurs when resources fall below "load plus 4.0 percent" in stochastic simulations and below "load plus 8.8 percent" in deterministic simulations.[40] TURN recommends that this issue receive additional analysis and stakeholder input.[41]

TURN did not propose a specific definition of "over-generation event" but mentions the importance of assessing it, in the context of its statements about defining criteria for flexibility: "Over-generation in large quantities may pose operational challenges, but planning to manage over-generation may still be more of an economic problem than a reliability problem. As with defining flexibility requirements, a careful examination of the potential of over-generation, and its potential consequences for reliability and cost, are needed."[42]

Other Parties' Views:
During the Working Group One calls, TURN, City of Redondo Beach, California Wind Energy Association (CalWEA), and UCS each thought that the CAISO's definition of a "loss of load event" (or "incident" per CAISO's definition)

---

[39] Woodruff Prepared Testimony, Appendix G. The column labeled "SoCo" states that in its reliability modeling, the Southern Company will shed firm load to maintain both spinning and non-spinning reserves. See page A-3 from Resource Adequacy Requirements: Reliability and Economic Implications, prepared for the Federal Energy Regulatory Commission by The Brattle Group and Astrape Consulting, September 2013, available at https://www.ferc.gov/legal/staff-reports/2014/02-07-14-consultant-report.pdf.
[40] The estimated "stochastic" operating reserve requirement of 4.0% equals spinning reserves of 3.0% of load plus regulation-up of 1.0% of load. The estimated "deterministic" operating reserve requirement of 8.8% equals contingency reserves of 6.0% of load, plus regulation-up of 1.0% of load, plus 1.8% of load for load following-up (which is half of the peak hour "load following-up" requirement of 3.6% of load).
[41] See attachment to TURN's email to Working Group One participants on June 24, 2015.
[42] Ibid.

may be too conservative (e.g. lead to over-procurement). Other parties presented alternative opinions. Some of the more detailed feedback provided as informal comments from various parties[43] follows:

- CalWEA recommended that "loss of load events" should be defined to include loss of load following-up and regulation-up amounting to no more than 2% of hourly load. Contingency reserves, on the other hand, should not be included in the "loss of load event" definition in order to prevent "double-counting."[44]

- Jan Reid cited the Congress of the United States Congressional Budget Office Paper from September 2001 titled "Causes and Lessons of the California Electricity Crisis"[45] in which the CAISO's definition of a Stage 3 emergency was set to when "electric reserves fell below 1.5%," and argued that the operating reserves modeled in both deterministic and stochastic studied should be set at 1.5% of energy requirements.

- PG&E's feedback appeared to support at least some of CAISO's "incident" definition, explaining the need for both deterministic and stochastic models that do utilize intra-hour dispatch or recourse[46] to include regulation reserves and a certain amount of load following requirement (PG&E argued 50% of load following reserve requirements) in addition to 3% spinning reserves (consistent with CAISO's operating procedures), because there is a need to account for the contribution of forecast uncertainty and/or intra-hour variability of net load to "loss of load events."[47] For stochastic modeling that does include intra-hour dispatch and recourse, only regulation reserves ("about 1.5% of load") (plus the 3% spinning reserves consistent with CAISO's operating procedures) are needed because for stochastic modeling "it is possible to estimate net load uncertainty and recourse as well as the contribution of forecast uncertainty or intra-hour variability of net load to loss of load."

- City of Redondo Beach articulated a concern it shares with TURN and UCS. City of Redondo Beach is concerned about possible distortion of the flexibility need results due to the inclusion of additional reserve commitments in the model. "Inclusion of such reserves (regulation and load following up, and regulation and load following down) could result in, among other things, increasing the magnitude, frequency and duration of the shortfalls calculated by the model. Since the goal of the study is to determine the additional flexibility need due to the addition of intermittent resources (e.g., wind and solar), the increase in flexible capacity shortfall and over-generation as a consequence of adding reserve requirements, biases the results and deviates from the study goal." Staff understands this to imply a

---

[43] Emailed to staff and the LTPP service list on about August 13, 2015 following the public workshop held at the CPUC on August 4, 2015.

[44] CalWEA stated that "employing a lower margin above hourly load in long-term production simulation studies intended to study capacity needs by excluding so-called contingency reserves (spinning and non-spinning reserves) from that margin avoids redundantly accounting for the forced outage of available (existing plus planned) resources. This is because, in daily system operations, when the system operator commits and dispatches contingency reserves, it does so precisely to account for the potential forced outage of available resources. If, in the long-term production simulation studies, contingency reserves are added to load when determining capacity shortfalls, it would be tantamount to double-counting the impact of supplies' forced outages, hence leading to the identification of additional capacity shortfall events and likely triggering supply additions that far exceed a reasonable Planning Reserve Margin."

[45] http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/30xx/doc3062/californiaenergy.pdf

[46] The ability to adjust commitment or dispatch to meet intra-hour forecast uncertainty and variability.

[47] PG&E stated that "…(deterministic) modeling uses hourly dispatch (not intra-hour dispatch), does not model the load/wind/solar uncertainty, and does not have recourse (the ability to adjust commitment or dispatch to meet intra-hour forecast uncertainty and variability). By adding 50% of the load following requirements and counting load following shortfalls as part of loss of load events, the contribution of forecast uncertainty and variability of net load can be accounted for. We propose 50% load following reserves to cover 2/3 of the forecast and variability deviations of net load (one standard deviation)."

recommended "loss of load event" definition as when supply is less than load without regard to flexibility reserves (i.e. regulation and load following-up).

3.4. STAFF RECOMMENDATIONS

3.4.1. STAFF RECOMMENDATION ON LOSS OF LOAD EVENT DEFINITION

Staff recommends the following:

- "Loss of load event" in stochastic models shall be defined to occur when effective operating reserves deplete to 2.5% of hourly load or less (1.0% regulation + 1.5% spinning reserves).
- "Loss of load event" in deterministic models shall be defined to occur when effective operating reserves deplete to 3.5% of hourly load or less (1.0% regulation + 1.5% spinning reserves + 1.0% load following-up).

Working Group participants discussed the possibility of producing modeling results using alternative "event" definitions, for example, either including most operating reserves in the "event" definition or including no operating reserves. Due to the resource constraints that modeling parties face, staff has decided to include one definition of an "event" for modeling purposes. Staff made this decision in the context that the definitions and standards being recommended by staff are for modeling purposes – to produce consistent and comparable modeling results across different models and analyses produced by different parties. When the Commission needs to make a procurement decision, it will need to evaluate the modeling results in the context of the adopted definitions compared to alternative definitions, potential reliability risks that are not counted in the modeling, potential measures that were not explicitly modeled, and any other mitigating factors.

Staff's recommendation chooses a middle-of-the-road approach that recognizes (1) the historical expectation that the CAISO would not interrupt load until reserves fall below 1.5%, (2) the CAISO tariff appearing to necessitate a maintenance of 3.0% of spinning reserves absent emergency conditions, and (3) the possibility that including any contingency reserves in an "event" definition for long-term resource planning may be "double-counting."

The inclusion of some regulation and load following reserves in the staff recommended "event" definition recognizes the point that the hourly dispatch models have limited ability to account for forecast error and intra-hour variability and uncertainty. We also considered UCS' argument that any increment to the regulation and load following reserves that is used to define a loss of load event should address only the need for that type of reserve that is triggered by the additional MW of intermittent resources.[48] We judged the potential merit of UCS' argument in light of the seemingly implied notion that renewable energy inherently will not provide regulation or load following up/down, which necessarily dismisses the potential system load following and regulation up/down benefits that could be had from energy storage connected to renewable resources, and from the impacts of DERs without first closely considering them. We also note ORA's observation that voltage reductions should be considered as an emergency measure, which LTPP staff understands may be able to mitigate some intra-hour load uncertainty and generation variability in the event of an emergency.

---

[48] UCS states "if regulation reserve and load following capability are not historically included in the definition of a shortfall, then the variability of load and other historical factors should not be included in the CPUC's definition of shortfall. In this case, only the fraction of unmet regulation and load following reserve requirement related to the increasing penetration of variable renewable resources could be counted as a shortfall. For example, take a case where the load following requirement in a given hour would have been 1,000 MW without any variable renewables, but would be 2,000 MW when renewable variability is included in the calculation. If the simulation registers a load following shortfall of 1,500 MW in this hour, this shortfall would only count as a 1,500 MW - 1,000 MW = 500 MW shortfall." See attachment to UCS's email to Working Group One participants on June 24, 2015.

3.4.2. Staff Recommendation on Over-generation Event Definition

The Working Group discussions placed a great deal of focus on defining the conditions for a "loss of load event" in reliability planning, despite the lack of consensus on this issue. There was, however, little discussion if any on when "unsolved over-generation" should be recorded after all economic renewable curtailment has been exhausted within a production cost simulation.

Considering that the over-abundance of energy poses a different challenge than a lack of capacity, and to avoid overstating the potential for "unsolved over-generation," staff recommends the following:

- For both deterministic and stochastic hourly models with curtailable renewable generation, over-generation gets mitigated by renewable curtailment before downward flexibility reserves (load-following-down and regulation-down) get dispatched.
- When all available renewable curtailment is exhausted, then the model starts to use downward flexibility reserves (load-following-down and regulation-down in that order) and records the amounts and types of downward flexibility reserves used.
- If all the downward flexibility reserves get used and there is still over-generation remaining, the model would record this as "unsolved over-generation" ("dump energy" in the PLEXOS model).

3.4.3. Staff Recommendation on Metrics and Standards

Staff recommends the following:

Stochastic Modeling

- Stochastic modeling studies should report the LOLE, LOLH, and EUE metrics using the "loss of load" event definition recommended above. All of these metrics provide valuable information that the Commission should consider for authorizing new resource procurement. The EUE metric measures the magnitude of events; when considered alongside the LOLE metric (which measures frequency of events) and the LOLH metric (which measures the hours of events), the results have the potential of revealing several "event" dimensions that would otherwise be lost relying solely on a single metric. EUE may be a particularly useful metric because it measures the cumulative magnitude of "loss of load events," which is required information to evaluate tradeoffs of the cost of incurring loss of load verses the cost of mitigating measures to avoid the loss of load.
- Definition of a "Day:"
    - The LOLE metric shall define "day" as any day where one or more "loss of load events" occurred within that day – regardless of the magnitude, duration, or number of occurrences within that day. This definition appears to be consistent with the modeling approaches used or recommended by SCE and PG&E, and is consistent with the approach used in CPUC's Resource Adequacy proceeding. The LOLE result can be directly compared to the "one day in ten years" standard (i.e. measured LOLE should be no more than one expected day in ten years, or no more than 0.1 expected days per year).
    - When using the LOLH metric, it is not necessary to adopt a definition of "day" for comparison to the "one day in ten years" standard ahead of time. The LOLH result is reported in "expected hours per year" and can be compared to the "one day in ten years" standard by using any number of interpretations for "hours per day", for example, one hour, seven hours, or 24 cumulative hours.
    - The "day" definition does not apply to the EUE or LOLP metrics.

- Stochastic modeling studies should report alongside each metric the implied Planning Reserve Margin (PRM), and whether any additional capacity is needed when compared to the following standards:
    - The LOLE metric is directly comparable to the "one day in ten years" standard.
    - The LOLH metric shall be compared to the "one day in ten years" standard by assuming that 24 cumulative hours is equivalent to a "day". The PRM and additional capacity needed may also be reported using other assumptions for "hours per day" in addition.
    - The EUE metric shall be compared to a standard of 0.001%.[49]
- Stochastic modeling studies should report the expected curtailment and the expected "unsolved over-generation" in expected annual energy terms, similar in nature to the EUE metric. The curtailment and "unsolved over-generation event" definitions as recommended by staff above shall be used.

Deterministic Modeling

- Deterministic modeling studies should report the production cost simulation's recording of the magnitude of "events" over the course of the year being simulated, using the "loss of load" event definition recommended above. Deterministic results cannot be compared to a "one in ten years" standard like stochastic results. So, for deterministic modeling studies staff recommends using a "no hour in one year" standard for "loss of load", meaning if any one hour of the year had a "loss of load event," then the system has a deficiency. The size of the deficiency would be quantified by "event" magnitude of the hour with the largest "event" magnitude. This proposal does not intend to change the current deterministic modeling framework for measuring "loss of load," except for adjusting the "loss of load event" definition.
- Deterministic modeling studies should report curtailment and "unsolved over-generation" in terms of annual energy and annual maximum capacity (i.e. the hour of the year with the largest magnitude of curtailment or unsolved over-generation). The curtailment and "unsolved over-generation event" definitions as recommended by staff above shall be used.

The tables below summarize the definitions, metrics, and standards recommended by staff for use in LTPP deterministic and stochastic technical studies.

TABLE 1 LOSS OF LOAD METRIC FOR DETERMINISTIC MODELS

Metric	Event definition	Reliability standard	Additional capacity to meet standard
Loss of Load	Supply < Load + 1.0% Load following-up + 1.5% spinning reserves + 1.0% regulation-up [#]	No hours in one year	= "loss of load event" magnitude of the hour with the largest "loss of load event"
[#] Percentages refer to the percent of expected load for that hour			

[49] CalWEA in informal comments dated August 13, 2015 recommended that an EUE standard should be based on benchmarking EUE values to previous years of reliable system operation, which would provide a rational basis for identifying an acceptable EUE level. Determining the EUE standard by tabulating study results and picking an "acceptable" EUE level such as 0.001% may be subjective and arbitrary.

TABLE 2 LOSS OF LOAD METRICS FOR STOCHASTIC MODELS

Metric	Event definition	Day definition	Reliability standard	Additional capacity to meet standard
Loss of Load Expectation (LOLE)	Supply < Load + 1.5% spinning reserves + 1.0% regulation-up [#]	One "day" is any day where one or more events occurred	One day in ten years	= Capacity needed to reduce LOLE to 0.1 days per year
Loss of Load Hours (LOLH)	Same as LOLE	One "day" is 24 cumulative hours	One day in ten years	= Capacity needed to reduce LOLH to 2.4 hours per year
Normalized Expected Unserved Energy (EUE)	Same as LOLE	NA	0.001%	= Capacity needed to reduce EUE to 0.001%
[#] Percentages refer to the percent of expected load for that hour				

TABLE 3 OVER-GENERATION METRICS FOR DETERMINISTIC MODELS

	Event definition	Reliability standard	Additional capacity to meet standard
Curtailment	Supply + Downward flex reserves > Load [1]	NA	NA
Unsolved over-generation [2]	Supply > Load	Zero	Requires additional studies to test solutions
[1] Downward flex reserves refer to load-following-down and regulation-down [2] PLEXOS-based models record this in a variable called "dump energy"			

TABLE 4 OVER-GENERATION METRICS FOR STOCHASTIC MODELS

	Event definition	Day definition	Reliability standard	Additional capacity to meet standard
Curtailment expectation	Supply + Downward flex reserves > Load [1]	NA	NA	NA
Unsolved over-generation expectation [2]	Supply > Load	NA	NA	NA
[1] Downward flex reserves refer to load-following-down and regulation-down [2] PLEXOS-based models record this in a variable called "dump energy"				

As evident in the recommendations above, staff believes (as do some parties) that it is premature to rely solely on one reliability metric, and one interpretation of that metric, for comparison to a "one in ten years" standard. The Commission should consider other metrics that incorporate a different definition of a day/standard along with their corresponding implied PRMs and cost to the system. As described in Section 5 of this document, staff proposes a process for how both deterministic models with minor improvements, and stochastic models with major improvement and standardization, can help inform the CPUC's decision-making. The recommendations on definitions, metrics, and standards to apply to such modeling as outlined in this section are meant to complement the modeling work recommended in Section 5 of this document.

4. IDENTIFYING STANDARD OUTPUTS

Modeling results from the technical studies submitted in the 2014 LTPP were difficult to validate and compare to each other as the various models produced different outputs, driven both by variance in model methodologies and input assumptions. Working Group Two sought to identify which modeling methodologies should be used and detail which specific outputs should be generated by various models. Sections 5 and 6 of this document, describing the efforts of Working Group 3, elaborate on the development of more detailed guidance on key model methodologies and inputs to enable studies conducted by different parties to be more comparable. This section will address three topic areas: 1) whether deterministic, stochastic, or a combination of both modeling techniques should be used, 2) how greenhouse gas emissions should be modeled & a proposed method for bioenergy facilities, and, 3) what details should be included in iteration-specific results produced by models. Consensus within these areas can improve the ability of parties and decision-makers to compare outputs and draw more robust conclusions.

4.1. USE OF DETERMINISTIC, STOCHASTIC, OR A COMBINATION OF BOTH MODELING TECHNIQUES

Energy Division staff believes that these two modeling approaches have fundamentally different strengths. Deterministic models have been vetted through years of use, are relatively easy to understand, can model the entire WECC region with great detail, and can provide many details about system performance under a fixed set of assumptions. Deterministic models may be especially useful for reporting the cumulative values of system performance indicators such as production costs and emissions. On the other hand, stochastic models can account for a wide range of possible system conditions and are designed to report the probability of both average and extreme or rare system conditions. However, stochastic models are complex and require more computational resources, and may therefore model the system in less detail. A combination of these two approaches could use deterministic modeling to study system performance under the fixed set of assumptions in a given planning scenario, with stochastic analyses done to show a probability distribution of system performance around that scenario.

4.1.1. Staff Recommendation for Modeling Techniques

We suggest that for the upcoming LTPP proceeding and going forward, both models continue to be used in tandem. It is likely that stochastic models are not yet mature enough to be used for procurement-related decisions, but will continue to add value as they are refined. Parties performing modeling have cautioned the Commission against engaging in procurement activities based on the results. Thus, it is most appropriate to rely on deterministic models to inform procurement in the near term while stochastic models are more fully developed and validated. To the extent the recommendations of this report are adopted and the recommended modeling improvements are successful, the 2016 and future LTPPs can shift to rely more on stochastic model results. Many parties noted that deterministic modeling was appropriate for studying system performance related to over-generation, production costs, or emissions, while stochastic modeling, when more fully developed and validated, might be preferred for examining capacity (flexible and generic) needs. This topic is fully discussed in Section 5 and 6 of this document with specific recommendations on how to apply both deterministic and stochastic modeling techniques to evaluate system performance and inform Commission decision-making.

4.2. Accounting for Greenhouse Gas Emissions

As the state progresses on a path of continued Greenhouse Gas (GHG) emissions reductions, accurate forecasts of the emissions impact of various alternative scenarios will be critical for decision makers to make informed choices. Existing models do provide information on GHG emissions, but staff believes this output should be shown consistently across all models, with all corresponding inputs developed in a transparent manner. Working group participants were supportive of including the expected Greenhouse Gas (GHG) emissions of given model runs as an output. However, this may need to be done in different manners for deterministic and stochastic models. Deterministic studies embed detailed production cost models and are inherently able to report GHG emissions. Stochastic models can also be configured to report GHG emissions, though there are computational limits to stochastic modeling methods which need to be considered.

SCE noted that "GHG emissions for a stochastic model would be computationally intensive and might not yield usable results. This is particularly true in the case of modeling the WECC to get accurate GHG emissions. For example, at present both SCE and ISO's stochastic models do not perform granular dispatch of WECC generators (for example, units outside California are significantly aggregated). It could be reasonable to do stochastic modeling of GHG accounting according to Air Resources Board rules, but this would only include plants in the ISO control area, not WECC wide." SCE also noted challenges with compiling annual emissions information from stochastic models that used non-annual iteration runs as its fundamental Monte Carlo simulation. To the extent other annual results such as over-generation are aggregated, GHG emissions could be similarly reported.[50]

ORA suggested a way to overcome the difficulty of a simplified WECC model within stochastic models. "The existing stochastic models, however, will not account for emissions outside of California, despite the fact that they could show changing import patterns. This is not necessarily a problem, from the perspective of emissions

[50] See attachment to SCE's email to Working Group Two participants on April 24, 2015.

assessment from the rest of WECC. The stochastic models' output could be coupled to different WECC scenarios by assigning an emissions characteristic scenario matrix (i.e., different emissions coefficients by season, or month, and/or by peak/off-peak period, based on, e.g., "high", "med" or "low" coal retirement scenarios in the rest of WECC) to all imports of power into California that result from the stochastic model. This approach severs the stochastic modeling processes that focus on intra-California wind/solar/load probability distributions, from the rest-of-WECC modeling construct that can continue to be at a more aggregate level than the model's representation of California resources and load." Regarding how to interpret "expected value" for GHG emissions reported from stochastic models, ORA suggested they could be interpreted similarly to "expected value" for annual generation.[51]

City of Redondo Beach commented that it may be less important to focus on reporting GHG emissions from a stochastic model, but rather, GHG emissions should be reported from deterministic models studying a range of alternative futures. The statistical variations in GHG emissions that result from intra-hour and hour-to-hour statistical variations of load, solar and wind are not of interest given that issues around GHG emission levels are long-term in nature. Annual GHG emissions will be sensitive to long-term year-over-year statistical variation in load growth due to changes in economic activity, energy efficiency, and penetrations of behind-the-load meter distributed generation. Long-term GHG emissions could be evaluated through "scenario analysis" that uses deterministic inputs in a conventional production cost model, not via a statistical approach, for example, scenarios with 10% higher or 10% lower CO2 prices could evaluate sensitivity of GHG emissions to CO2 price.[52]

4.2.1. Staff Recommendation on accounting for Greenhouse Gas emissions

Staff believes that GHG emissions are a critical output to assess whether procurement-related decisions lead to achievement of GHG emissions reduction goals under a range of alternative futures, consistent with the Governor's recent Executive Order[53] calling for deep reductions in emissions. It is also important to consider the impact of increased integration across balancing areas in the West, and the ability to reduce overall emissions across the West, not just in California. We recognize there is considerable uncertainty around projecting future emissions from outside of California, especially considering evolving renewable procurement policies and EPA 111(d) compliance policies in other states. We therefore recommend that detailed deterministic modeling should continue to report monthly granular WECC-wide GHG emissions, in addition to California-specific emissions, with the best data available to characterize realistic granular dispatch of all emissions-producing units. Unit dispatch results including fuel use and generation are generally standard reporting for deterministic production simulation models and should continue to be reported because they aid model transparency and validation, and enable post-processing to calculate GHG emissions under alternative GHG accounting regimes. Deterministic model sensitivity analyses may be used to consider long-term uncertainties such as load growth, carbon price, and other states' renewable procurement policies. To the extent stochastic models can report GHG emissions, they should, but as discussed further in Section 5 and 6 of this document, we will depend on deterministic models to be the primary tool for reporting GHG emissions.

---

[51] See attachment to ORA's email to Working Group Two participants on April 24, 2015.
[52] See attachment to City of Redondo Beach's email to Working Group Two participants on April 24, 2015.
[53] 1 Executive Order B-30-15, available online at: http://gov.ca.gov/news.php?id=18938

4.2.2. The Green Power Institute's Proposal for Bioenergy Generators GHG Emissions

The Green Power Institute (GPI) recommends an alternative rationale for GHG accounting of generators using bioenergy, and they recommend two methods of implementation. They propose the following:

"The treatment of biogenic greenhouse-gas emissions from bioenergy production is a more complex matter. Bioenergy generators emit CO2 from the biomass or biogas fuels they consume to produce energy. However, unlike the case of fossil fuel use, in which the carbon in the fuels would remain geologically protected from the atmosphere if it not used for energy production, in the case of bioenergy production some or all of the carbon in the fuel would be released to atmosphere as a mixture of CO2 and CH4 in the absence of energy production, and the emissions from the generator have to be balanced against the emissions, both prompt and delayed, associated with the alternative fate of the fuel.

The California Air Resources Board (ARB), in formulating the rules for the state's AB 32 compliance programs, has determined that the production of energy from the kinds of bioenergy resources used in California today is carbon neutral or better, and therefore exempted bioenergy producers from compliance obligations under the state's cap-and-trade program for their emissions of biogenic CO2. Bioenergy generators report their biogenic CO2 emissions to the ARB, but the accounts for biogenic-carbon emissions are kept separate from the accounts for fossil-carbon emissions, and only the fossil emissions require offsetting allowances.

All of the greenhouse-gas emissions accounting that will be conducted in conjunction with the LTPP modeling work will depend on equipping the models with a set of greenhouse-gas-emissions factors for the generators included in the model's dataset. GPI proposes two alternatives with respect to how to handle the emissions factors for bioenergy generators. The simpler alternative is to declare bioenergy generators to be carbon neutral or better unless or until such time as the ARB determines that such a determination is no longer warranted. Until then, the generators would have a default greenhouse-gas emissions factor of zero, equivalent to wind and solar generators, for their biogenic-carbon emissions. Individual bioenergy generators, on a facility-specific basis, should be allowed to have a negative emissions factor assigned upon suitable demonstration that the avoided emissions of the alternative fate for the fuel exceed the emissions at the generator.

A more comprehensive alternative would be to equip the models to carry two greenhouse-gas emissions factors for each bioenergy generator, a gross emissions factor, which is the emissions emitted by the energy-production process itself, and a net emissions factor, which reduces gross by the avoided emissions of the alternate fate for the fuels. As above, the net emissions factor would be zero as a default value, and could be adjusted to a negative value on a generator-specific basis upon suitable demonstration of a net reduction of biogenic greenhouse-gas emissions."[54]

4.2.3. Staff Recommendation on Bioenergy GHG Accounting

Upon review of both a recent CPUC RPS procurement plan decision associating "net zero emissions" with electricity production from biomethane[55] and the ARB's regulations on emissions reporting under California's Cap

[54] See attachment to GPI's email to Working Group Two participants on June 30, 2015

[55] CPUC D.13-11-024 in referring to electricity production from biomethane having "net zero emissions" states: "The Commission has not fully explored how the "net zero emissions concept" will be put into practice in the context of RPS compliance. The Commission's implementation of SB 1122 (Rubio, Stats. 2012, ch. 612), adding 250

and Trade program for combustion of biomass-derived fuels,[56] it appears that accounting for GHG emissions from bioenergy facilities is both complex and unsettled at this time.

As described under the general recommendation for GHG emissions reporting above, staff recommends that models should report unit fuel and electricity production to enable post-processing of results under different GHG accounting frameworks. Given the complexities of accounting for the emissions from different biogenic fuel types, it would be analytically inappropriate to assign a single emissions factor of zero to combustion of such fuels, as GPI proposes. However, Staff understands that such granular calculations for bioenergy emissions may be resource-intensive, and the result may be precise, but not necessarily accurate. Therefore, staff recommends that bioenergy generators be accounted for as having "net zero emission" only for purposes of calculating an aggregate GHG emissions result to compare alternate LTPP planning scenarios. The underlying unit fuel and electricity production should still be available to perform more granular emissions accounting if necessary. Staff believes the simplified "net zero emission" accounting method is acceptable given that bioenergy electricity production and emissions are very small compared to all other emissions producing electricity production. We find this to be consistent with the current ARB guidance that GHGs may be reduced by substituting fossil fuels with those derived from biomass.

4.3. Using Iteration Specific Results

Deterministic models as used in the LTPP are essentially a single detailed production cost simulation for an entire study year and can be considered a single, detailed "iteration". Stochastic models as used thus far in the LTPP are based on running simplified production cost simulations many times as part of Monte Carlo simulations to capture statistical behavior of key parameters. Each "draw" of the Monte Carlo simulations is considered an "iteration" and may cover one month, one year, or some other temporal scope depending on the design of the stochastic model. The stochastic model post-processes outputs from all iterations to finally report aggregated results. Having key inputs and outputs from each iteration, or iteration-specific results, prior to any post-processing can facilitate a clearer understanding of the impact of underlying variables on model results and aid model validation and transparency. However, in order to save iteration-specific results for later post-processing, most models require that desired outputs must be specified in advance. Constraints in computing resources however, limit how many iteration-specific results can feasibly be saved. Therefore, it is important to focus only on the most critical inputs and outputs from each iteration and perhaps limit the requirement of capturing iteration-specific results for focused "deep-dive" efforts that examine specific iterations. The level of temporal granularity of iteration-specific results is also a key consideration. For example, results such as GHG emissions could be aggregated using monthly or perhaps annual granularity, while results related to measuring reliability might require hourly granularity.

4.3.1. Staff Recommendations of Iteration Specific Results

Staff recommends that certain iteration-specific results be generated to provide the most transparent data set for evaluation. In particular, we recommend that the data in the following list be reported for all deterministic models at the hourly level.[57]

MW of generation from bioenergy sources to the feed-in tariff program authorized by Section 399.20, is a logical opportunity to explore this concept further."

[56] See § 95852.1 of title 17, California Code of Regulations (CCR), describing Compliance Obligations for Biomass-Derived Fuels under the "Regulation for the California Cap on Greenhouse Gas Emissions and Market-Based Compliance Mechanisms to Allow for the Use of Compliance Instruments Issues by Linked Jurisdictions": http://www.arb.ca.gov/cc/capandtrade/capandtrade/unofficial_c&t_012015.pdf

1. Load
2. Upward reserve requirements (load following up, regulation up, spinning, non-spinning)
3. Ancillary services available separated by resource type
4. Energy generated separated by resource type
5. Magnitude of unserved energy and reserve shortfalls by category
6. Import limitations
7. Demand response dispatch
8. Generator outages or capacity not on outage. To make this more manageable, the available capacity of all conventional CAISO generators could be aggregated.
9. Magnitude of curtailment
10. Downward reserve requirements (load following down, regulation down)
11. Energy price for SCE, SDG&E, PG&E Bay, and PG&E Valley
12. Reserve/shadow prices for each reserve product.
13. Regional generation requirement shadow prices for CAISO, SCE, and SDG&E
14. Storage charge and discharge magnitude
15. Hydro and nuclear output
16. Net imports
17. CAISO combined cycle output (MWh) and CO2 emissions
18. CAISO combustion turbine / peaker output (MWh) and CO2 emissions
19. CAISO CHP MWh and CO2
20. Any other CO2 emissions in the CAISO
21. GHG emissions in the rest of WECC (if simulated)
22. Production costs
23. CO2 emission permit costs

For stochastic models, we recommend that at least the following list of data captured at the hourly level be saved for each iteration.

1. Load
2. Upward reserve requirements (load following up, regulation up, spinning, non-spinning)
3. Ancillary services available separated by resource type
4. Energy generated separated by resource type
5. Generator outages or capacity not on outage. To make this more manageable, the available capacity of all conventional CAISO generators could be aggregated.
6. Magnitude of unserved energy and reserve shortfalls by category

The following section of this document on the Models Validation Working Group provides full detail on the expected use of iteration-specific results to validate stochastic models and to conduct "deep-dives" investigating the circumstances around key events.

---

[57] This list was largely informed from a recommendation developed by UCS. See attachment to UCS's email to Working Group Two participants on June 24, 2015.

5. MODELS VALIDATION WORKING GROUP

5.1. OVERVIEW AND SUMMARY

Electric systems face the challenge of reliably and economically accommodating growing amounts of variable and only partly predictable wind and solar generation. Anticipating the timing, amounts, and types of flexible assets and operations needed to manage this challenge requires new, complex analyses. These analyses must realistically represent the wide range of operational conditions and requirements that could arise when variabilities and uncertainties associated with wind and solar generation interact with other variabilities and uncertainties such as regarding electric loads and equipment outages. Thus, to help inform planning decisions, new kinds of complex electric system flexibility analyses were prepared, evaluated, and discussed in the CPUC's 2014 LTPP proceeding.

While based in part on familiar electric system production simulation models, electric system flexibility analyses introduced into the 2014 LTPP proceeding also cover unfamiliar and challenging ground. While providing valuable information for the 2014 LTPP, these and similar analyses need to be better understood and perhaps modified before being relied upon for a variety of Commission decisions going forward. The overarching objective of 2014 LTPP Working Group Three has been to assess and develop recommendations regarding:

- What modeling methodologies are *fundamentally* appropriate (we have sufficient confidence in them) for informing what kinds of Commission decisions, and what modifications to their design or application would be desirable for these purposes?
- What modeling methodologies are worth pursuing but have not yet fundamentally achieved sufficient confidence, and what is needed to achieve such confidence?

This should inform use of modeling methods in the next LTPP cycle, for which specific study priorities would still need to be determined. The consensus view of Working Group 3 was that there is sufficient, fundamental confidence in familiar deterministic production simulation models (e.g., PLEXOS) for over-generation/curtailment, production costs, GHG emissions, and integration adder issues[58] and decisions. Furthermore, there was consensus that such models can also be used for capacity (flexible and generic) need studies at least until there is greater confidence in stochastic methodologies, at which point the stochastic methodologies might be found preferable for such studies. However, the dominant opinion of working group participants was that future applications of deterministic models especially for capacity need issues (driven strongly by extreme events) should utilize planning cases covering a range of conditions regarding loads plus wind and solar generation, and also covering a range of other conditions such as typically treated via planning scenarios. Additionally, the consensus view was that modeling of hourly flexible reserves commitment requirements in both deterministic and stochastic modeling needs to be more fully understood and accepted, especially if inability to meet such requirements is defined as a reliability event indicating capacity shortfall.

[58] Integration adders are additional costs (e.g., per MWh of output) attributed to incremental additions of variable wind and solar energy resources (and potentially to other kinds of resources) for planning and/or procurement purposes. These added costs account for costs that additions of the above resources impose on the system in ways not accounted for by valuation of those resources based on time-of-delivery energy values and on-peak capacity values. Generally, these added costs represent operation (and possibly addition) of system flexible response capabilities to accommodate the variable and partly unpredictable and/or undispatchable output patterns of the added resources to which the integration adders are assigned.

The near-consensus view of Working Group 3 was that stochastic methodologies (stochastic with regard to load, wind and solar conditions) are promising for capacity need studies and can also be valuable for other issues, such as over-generation. However, the consensus view was that sufficient confidence is lacking regarding the fundamental methodology of developing and sampling load, wind, and solar ("L/W/S") probability distributions to construct hourly (or other) profiles of L/W/S conditions for modeling. Furthermore, there was broad support both for validating overall probability distributions or profiles, and for conducting deep dives to assess particular combinations of circumstances modeled to produce capacity shortfall or over-generation events.

Working Group 3 developed near-consensus regarding a need for better clarification, understanding, and confidence concerning the modeling of regional generation requirements for various load areas, as an input to any model. There was also near-consensus that the realism[59] and consequences of the way that electric storage and hydro are modeled should be more fully and transparently examined going forward, but that this does not represent a fundamental barrier for moving ahead with an otherwise acceptable methodology.

In brief, CPUC staff's summary of preliminary recommendations (each described in a section below) regarding five model validation issues are as follows:

1. Rely primarily on deterministic modeling to inform decisions while continuing to develop, and gain confidence in stochastic modeling both to supplement deterministic analyses and potentially to rely more on stochastic analysis at some point in the future.
    a. Develop a range of appropriately diverse load+wind+solar (L/W/S) sub-cases for at least the most important deterministic study case(s), where each L/W/S sub-case represents a set of L/W/S profiles developed transparently from different historical years' coincident (co-occurring) L/W/S conditions.
2. Continue to apply, validate and gain confidence in stochastic modeling methods.
    a. Validate overall stochastic L/W/S profiles versus deterministic "benchmarking" profiles developed per 1a above.
    b. Also conduct "deep dives" into particular modeled capacity shortfall and over-generation events to assess the realism of L/W/S and other conditions being modeled, in part via comparison with the above benchmarking profiles.
    c. Obtain and assess clear descriptions (with numerical examples) of the methodologies and logic underlying development of L/W/S profiles used for stochastic modeling.
    d. Evaluate modeled solar profiles focusing on the relationship between the assumed mix of solar generator design parameters used to convert irradiance to modeled AC electric output profiles versus our current best information about the actual mix of designs being, and likely to be, deployed.
3. Improve understanding of modeled flexible reserve commitment requirements using a seven-part approach that combines improved explanation of the modeling logic with empirical assessment of the actual modeled MW requirements and their consequences such as via comparison among different information sources and modeling methods.
4. Assess and improve understanding of the basis of modeled regional generation requirements, which may require a detailed task force effort.

---

[59] "Realism" refers to the *synthetic electric system operation* as modeled (such as involving loads, resources and their characteristics) sufficiently matching the way the *real world electric system* (hence "realism") would be expected to operate based on empirical experience or reasonable extrapolation of that experience - - such as regarding commitment and dispatch of resources, costs incurred, and potential loss of load or need to curtail resources.

5. Validate modeling of storage and hydro resources within the course of 2016 LTPP studies including deep dives (see 2b) that may be conducted, but without separate modeling runs for this purpose.

Below, Working Group 3 views and CPUC staff recommendations are described more fully.

5.2. WORKING GROUP 3 VIEWS ON MODEL VALIDATION ISSUES

5.2.1. DETERMINISTIC METHODOLOGIES: WORKING GROUP 3 VIEWS

As summarized above, the consensus view of Working Group 3 was that there is sufficient, fundamental confidence in familiar deterministic models as used for over-generation/curtailment, production costs, GHG emissions, and integration adder issues and decisions. This in part reflects a perception that these issues are characterized by conditions expected to occur periodically over time and to be reflected mainly in *cumulative* consequences such as measured in MWh, dollars, and emissions. Thus, accurately capturing magnitudes and probabilities of rare events such as via stochastic modeling is less critical for these issues, unless reliability threats are also involved.

There was also near-consensus in Working Group 3 that deterministic models can also be used for study of capacity (flexible and generic) need issues at least until there is greater confidence in stochastic methodologies. The general view (stated explicitly by some) appears to be that appropriate deterministic and stochastic methodologies should both be used (in parallel) into the 2016 LTPP. This would provide opportunity to initially use deterministic methodologies both for illuminating capacity need issues and for benchmarking, evaluating and refining the stochastic methods.

Despite general acceptance of fundamental deterministic modeling methodologies there still remains, as pointed out by Working Group 3 participants, the important challenge of determining how such deterministic models should be applied, both regarding specific formal planning assumptions (e.g., low GHG, higher/lower load or hydro conditions) and regarding how an efficient *set of study cases* should be constructed and its aggregate results interpreted. Most Working Group 3 participants saw a need to apply deterministic methodologies across a greater *range* of conditions, ideally in a manner that assigns relative weights or probabilities to the different study cases, which is further discussed under CPUC staff recommendations. This is seen as especially necessary for examining capacity shortfall/need issues driven by extreme events occurring under infrequent but plausible conditions. There was near consensus that one or a very few deterministic cases are inadequate for such purposes, and that attempting to make one or a very few deterministic cases adequate for assessing need relative to probabilistic reliability standards *only* by using selective ways to define "reliability violations" is nontransparent and potentially misleading.

Thus, application of deterministic modeling faces the challenge of developing a set of study cases that appropriately capture the range of plausible future conditions, and that can be credibly probability-weighted or otherwise interpreted in the aggregate for purposes of informing decisions. This was the view of most Working Group 3 participants. One possible approach eliciting generally favorable reaction would be to develop, for one or more formal deterministic "cases" (e.g., trajectory, high loads or low GHG), multiple sets ("sub-cases") of load, wind and solar ("L/W/S") profiles based on L/W/S probability distributions developed in a manner that is simplified and transparent relative to more sophisticated stochastic methodologies being used. This is further described in staff recommendations. Besides producing more informative deterministic studies, this could be a useful step in understanding and gaining confidence in more sophisticated stochastic methodologies. Thus, as emphasized by several Working Group 3 participants, it is important to consider *both* a diversity of L/W/S conditions and a range of broader overall futures (formal planning cases). Such a set of formal planning cases achievable with a manageable level of effort would be established within the process of developing the next LTPP cycle.

5.2.2. Stochastic Methodology Validation Needs: Working Group 3 Views

Working Group 3 concluded that stochastic methodologies are promising and should continue to be pursued, but need to achieve additional validation and confidence before being more heavily relied upon to support Commission decisions. There was broad support both for validating overall load, wind and solar (L/W/S) probability distributions and profiles used for modeling and also for deep dives into particular modeled conditions producing shortfall and over-generation events. Most participants recommended that the Commission ultimately develop guidelines regarding L/W/S data sources or datasets, and possibly also guidelines (but not overly prescriptive) for applying these data for modeling purposes.

Working Group 3 participants had a range of recommendations regarding how, using what reliable data, L/W/S probability distributions and profiles should be assessed and validated. Recommendations included benchmarking and comparison versus publically available L/W/S data such as from NREL, CEC and CAISO (see staff recommendations below), as well as examining methods and data used by other studies such as those conducted by E3 and NREL. Several participants emphasized that validation should use comparison datasets in which load, wind and solar data[60] are explicitly synchronized, i.e., taken from conditions that were actually coincident in the real world.

For validating L/W/S probability distributions and profiles, Working Group 3 participants generally viewed correlations, central tendencies and bounding as important considerations. Specifically, many participants believed that there should be further assessment of L/W/S correlations in the actual data sources used for the 2014 LTPP modeling, in modeling results themselves, and/or in any other data specifically identified and used for benchmarking purposes, such as the “real world coincident” data mentioned above. Several participants supported testing for correlations occurring under some conditions but perhaps not occurring with statistical significance across the entire range of L/W/S conditions. For example, meteorology-driven potential wind and solar output levels (or even ramp magnitudes) might be especially high or low, or especially correlated with each other for high load days in certain months. There was also considerable agreement that (1) the central tendencies (mean, median) of L/W/S probability distributions should be checked for consistency with the corresponding established deterministic planning assumptions, and (2) for purposes of validating L/W/S distributions and perhaps ultimately for developing formal Commission guidelines, the extreme ends of L/W/S probability distributions used for modeling should have realistic bounds based on reliable information.

As part of overall assessment and validation of L/W/S distributions and profiles, there was some interest in examining whether the base solar generation output profiles (e.g., in the 2014 LTPP 33 percent RPS trajectory case) are an accurate representation of likely future conditions. This represents an *input* issue for operational flexibility modeling, but requires considering the exogenous modeling to develop these inputs. These profiles are influenced by what mixes of PV and solar thermal generating facility designs are assumed when converting irradiance data to electrical output profiles via modeling. It has already been noted during this proceeding that if penetration of tracking or of relatively high inverter loadings (ratio of panel capacity to inverter capacity) is understated, which may be the case, then “peakiness” (as opposed to flatness) of solar profiles will be overstated.

---

[60] which for wind and solar may be data for the underlying meteorological drivers

The consensus view was that, as described above, empirical validation comparing L/W/S distributions and profiles used for stochastic modeling with reliable and understandable information sources is more useful than "mechanistic" validation based on understanding and critiquing the rationale and logic for constructing the distributions and profiles. However, several Working Group 3 participants emphasized that there must be at least some minimum level of understanding of the rationale and logic, in order to have confidence in the overall methodology.

Most Working Group 3 participants supported (some strongly) using *deep dives* that document and analyze the particular system conditions being modeled during and possibly preceding particular hours for which significant shortfall or over-generation events occur. As a validation tool this would complement assessment and validation of *overall* L/W/S probability distributions and profiles. Furthermore, it could shed light on the plausibility of *other* factors modeled to combine with L/W/S in an electric "perfect storm" of conditions that produces shortfall or over-generation events. Most commenters appear to believe that such deep dives would be most valuable or necessary for examining shortfall events, which inherently represent rare, extreme combinations of circumstances. However, there was also considerable interest in using deep dives to examine at least the most extreme over-generation events. A few participants did not believe that deep dives into over-generation events would be an efficient use of analytic resources.

It is clear that deep dives should at a minimum capture one hour's worth of relevant modeling variable values for an extreme event. However, most commenters recommended obtaining a full day (or more, if applicable) of such modeling information for at least some events, to examine the chronological sequence of conditions leading to the event. Regarding what modeling variables should be extracted for deep dives: load levels, wind and solar generation levels, storage charging and discharging, imports and exports, and forced outages were all generally recommended. Some commenters recommended more comprehensive reporting of model output for other types of generation, and of up and down headroom/floor-room as well as uncommitted capacity for different resource categories. Over-generation deep dives might benefit from some information that is not as important for shortfall events.

Most Working Group 3 participants agreed that reliable and publicly available data sources regarding loads and wind/solar generation (or their driving meteorology) used to validate overall L/W/S probability distributions and profiles can also be used to assess the realism and credibility of conditions modeled to produce particular extreme (deep dive) events. A few participants noted that the combinations of conditions being modeled to produce extreme events may not correspond to the range of conditions represented by historical data potentially used for benchmarking.[61] Such differences (which are not necessarily "errors") are in fact part of what is to be evaluated via deep dives. Several participants pointed out that deep dives reveal the contribution to events of a wider range of conditions beyond L/W/S conditions.

Working Group 3 participants appeared to generally recognize that deep dives could be analytically laborious and time consuming. This depends in part on how many deep dives are conducted and what modeling variables are extracted. A two-pass approach to deep dives would use "standard" model runs to identify significant shortfall or over-generation events of interest, and then would conduct "second pass" runs for selected events. The second pass runs would extract and save additional detail (values of modeling variables) that is impractical to save for every standard model run, and if possible might simulate only the relevant portion (e.g., a day or month) of the

[61] In fact, L/W/S conditions developed by sophisticated stochastic methodologies are intended to sometimes fall outside the range of historically observed conditions, even after accounting for growth of load and of wind and solar generating capacity.

original run. This two-pass approach is conceptually appealing for deep dives, but may or may not be practical. Some Working Group participants including several who are experienced regarding production simulation studies express skepticism that a second pass approach would be efficient or desirable. It is possible that the pros and cons of a second pass approach differ across modeling methodologies, such as those using full-year or full-month simulations versus those simulating one randomly drawn day at a time. Overall, there is good reason to believe that deep dives will be a valuable validation tool, but that designing a deep dive approach requires careful informed consideration of the benefits of illuminating the interaction of L/W/S and other factors versus costs in terms of labor, time and data volumes.

5.2.3. Flexible Reserves Commitment Requirements: Working Group 3 Views

Working Group 3 identified a need for improved understanding and confidence regarding the establishment and modeling of flexible reserves commitment requirements, which is used in both deterministic and stochastic methodologies. Since reserve commitment levels could be adjusted if desired, this is not a barrier to confidence in the fundamental modeling methodologies themselves, but it could be a barrier to confidence in study results. If modeled failure to meet these commitment requirements was not by itself defined as a reliability "event," only about half of Working Group 3 participants would continue to see this issue as a high priority. However, virtually all participants expressed some interest (in some instances strong interest) in better understanding and validation of how flexible reserves commitment requirements are modeled. Several parties pointed to the potential impact of such requirements on over-generation, GHG emissions, and production cost results, and potentially also on other results including flexibility shortfalls in subsequent hours – due to the way that flexible reserves commitment requirements alter commitment and dispatch relative to what otherwise would have been modeled. It has also been pointed out that modeling 5-minute or other intra-hour dispatch would enable *explicit* treatment of flexibility need, provision and shortfalls over very short time horizons of a few minutes where starts and ramps are very constrained – which must be treated *indirectly* in hourly dispatch simulations.

How modeling of flexible reserve commitment requirements should actually be validated, especially with an efficient level of effort, is not easily resolved. Some parties suggested comparing these requirements to the CAISO's actual operating practices, while others pointed out that current practice do not deal with the higher future levels of wind and solar generation. At a minimum, some scaling or extrapolation of current practices to higher wind/solar levels would be needed. Some participants pointed out that the CAISO's method for developing reserves requirements ("Step 1", based on a Pacific Northwest National Laboratory methodology) is complex and nontransparent and needs to be understood and validated. Some participants supported comparing the requirements calculated by different methodologies under similar L/W/S circumstances (e.g., CAISO's method versus other methods[62]). Running sensitivity cases to examine the impacts of alternative reserve commitment levels was seen as valuable by some but not all participants. It was also noted that if hour-ahead L/W/S uncertainties are *explicitly* resolved by a model[63] to determine probability of unserved load or inadequate reserves, then perhaps we should just let that play out through the simulation of that hour. Then, the probability vs. magnitude of shortfall for that hour would be computed based on energy dispatch, committed reserves, forecast error distribution and assumed (if not modeled) intra-hour variability. How to proceed with the challenge of validating and achieving confidence in flexible reserve commitments is further discussed in staff recommendations.

---

[62] which could include methods used by SCE, E3 and/or NREL
[63] e.g., by calculating a forecast uncertainty-based probability distribution (MW versus probability) for net load exceeding supply from resources committed for energy plus ramp capability

5.2.4. Validating Regional Generation Requirements: Working Group Views

Among Working Group 3 participants there was broad support for clarifying, understanding, and validating the basis for setting regional minimum generation requirements, which were applied by CAISO but not by SCE in recent LTPP operational studies. Such requirements can be provided as an input assumption (a constraint) for any modeling methodology. Under these requirements, hourly energy dispatch summed over certain types of qualifying[64] resources within each of several load regions must amount to (i.e., offset or balance) at least 25 percent the hourly load within that region. These modeled requirements are a rough proxy for the provision of many different essential grid reliability services that are difficult to represent explicitly in production simulation models.

As pointed out by Working Group 3 participants, regional generation requirements will generally increase modeled amounts of over-generation/curtailment, production costs, and GHG emissions,[65] and yet the amounts of specific kinds of reliability services (e.g., inertia) required within specific electrical/geographic areas have not been clearly or formally documented or justified. CPUC staff suggested that an effort to understand and validate modeling of regional generation requirements may include the following sub-topics, which were generally supported by participants:

i. What different, clearly defined reliability services (such as inertia, primary frequency response, reactive power) are required to be provided, in what amounts, within what specifically defined load areas?
ii. How do these requirements vary across system conditions?
iii. What kinds of resources (characteristics and locations) are eligible to provide these services, both currently and potentially (realistically) in the future?
iv. What are the best sources of information for the above, including but not limited to current CAISO operating practices?
v. (How) can requirements for multiple kinds of regional reliability services (inertia, reactive supply/voltage support, etc.) be modeled in PLEXOS?

Regarding topic v. above, Working Group 3 commenters pointed out that it may be impractical or unnecessary to separately model (enforce) multiple kinds of regional requirements regarding reliability services,[66] such as for inertia versus primary frequency response versus reactive power versus simple dispatchability. Instead it may be reasonable to ask “can we identify which services or single service are/is the most limiting (likely to experience shortfall), and what amounts and types of resources and energy dispatch would be needed to ensure adequate provision of these services?”

Various Working Group 3 participants recommended that regional generation requirement questions be pursued by the CPUC working with the CAISO and with CAISO area transmission owners. There was general recognition that this effort could proceed apart from and in parallel with other modeling and analysis issues, and some participants believed that this could take considerable time.

---

[64] That is qualifying to provide the needed types of reliability services such as reactive power, primary frequency response, dispatchability, and inertia.
[65] The impact of regional minimum generation requirements on modeled capacity shortfall is less clear.
[66] Provision of some of these services is not linearly proportional to MW of dispatched output (e.g., reactive power, inertia), and in some instances may not require MW of real power output at all.

5.2.5. Validating Modeling of Storage and Hydroelectric Generation: Working Group 3 Views

Most Working Group 3 participants supported the idea of examining and gaining better understanding and confidence (thus, validation) of the modeling of storage and of hydroelectric generation, the latter including storage hydro not requiring pumping. Storage and hydro generally present production simulation modeling challenges requiring special attention, and this is even more important for studies focusing on shortfall and over-generation, and with inclusion of new kinds of storage for which there is little modeling experience. However, there is nothing fundamentally new about modeling storage and hydro, and most Working Group 3 participants appear to believe that while worth pursuing, validation of this aspect of modeling is not the highest priority and does not merit as much effort and time as other priorities discussed above. Some participants expressed interest in examining some of the hour-to-hour details of storage charge and discharge modeling.

6. PRELIMINARY CPUC STAFF RECOMMENDATIONS

The volume and specifics of analytic work implied by CPUC staff's preliminary recommendations below will need to be assessed and refined based on actual priorities and resources. The recommendations are intended to be reasonably if not precisely able to be pursued. For example, the level of effort required for running multiple sets of annual L/W/S hourly profiles for a single deterministic case implies more effort than required for a single deterministic case having a single set of L/W/S profiles, but less work than would be (and has already been) required for complex stochastic simulations.

6.1. Deterministic Modeling

CPUC staff recommends that going into the 2016 LTPP proceeding, there should be primary reliance on deterministic modeling to inform decisions related to over-generation, production costs, GHG emissions, and renewable generation integration adder issues, and also understanding of capacity (flexible and generic) shortfall/need issues. However, stochastic methodologies should continue to be developed, applied, improved, and used to initially supplement deterministic analyses and then to potentially play a greater role in the future especially for capacity issues. Furthermore, a range and diversity of deterministic modeling cases should be developed and analyzed that is appropriate for illuminating key uncertainties, drivers, and opportunities associated with the above and other issues. These cases should be selected as part of the 2016 LTPP process with input and interaction among CPUC, CAISO, CEC and parties. Cases should be informed by what has been learned in modeling and analyses supporting the 2014 LTPP (for example, regarding the importance of import/export constraints and regional generation requirements). The added value of sub-hourly (e.g., 5-minute) dispatch simulations has not yet been established, but deeper vetting of such approaches and their ability to bring added insights could be considered in the next LTPP cycle.

Where there is not strong reliance on stochastic analyses, deterministic analyses should capture a wider range of possible combinations of load, wind and solar (L/W/S) conditions than was captured in deterministic analyses for the 2014 LTPP. One way to do this, which is preliminarily recommended by CPUC staff, is to develop multiple "sub-cases" for one or more of the LTPP planning cases. These different sub-cases would represent different sets of within-California L/W/S profiles (different L/W/S conditions), all consistent with the particular load forecast and wind/solar generation portfolios established for the given planning case.

These sets of L/W/S profiles (sub-cases) should be developed from publicly available, reliable historical data on loads and the meteorological conditions driving wind and solar generation (e.g. wind speed, potential irradiance, and cloud cover). Furthermore, the L/W/S data used to produce any given set of profiles (any given sub-case) should all come from *coincident* real world conditions. For example, within a single sub-case (one set of L/W/S profiles), the load, wind and solar values for hour 15 of March 23 should all be based on (scaled or modified starting from) historical conditions that simultaneously existed in hour 15 of March 23, from whichever historical year was used as the basis for this particular sub-case. Thus, derivation of L/W/S profile sets (sub-cases) for deterministic modeling would have three desirable characteristics:

1. the resulting L/W/S profiles would be more diverse with increased potential for stressful combinations, relative to modeling a *single* set of L/W/S profiles for a deterministic planning case;
2. yet, the range of L/W/S conditions would be credible in that "we know these L/W/S *conditions* could co-occur, because they co-occurred historically"; and
3. the *methodology* for deriving the profiles would be based on public data, and be transparent and simpler, relative to methods used in more sophisticated stochastic studies.

Such an effort is enabled by wind and solar data from different historical years developed by and available from NREL, combined with load data for different historical years from the CAISO. NREL has developed (wind) and or is completing for release this fall (solar) profiles down to five-minute (wind) or one-minute (solar) granularity for each of about 40,000 sites WECC-wide, covering years 2007-2013, and possibly a few additional years for solar.[67] This includes profiles for basic meteorological data and also for corresponding AC electric output when assuming each of several wind and solar generating facility designs. Combined with hourly load data for the CAISO footprint for corresponding years, this can be used to produce multiple time-synchronized L/W/S profiles each based on a particular historical year, for constructing the L/W/S sub-cases of a given LTPP planning case.

If relying on full 8760-hour annual simulations (or month-by-month simulations) it would thus be possible to model each of the different historical years of L/W/S profiles, adjusting loads and wind/solar generating capacities to match the particular planning case. The level of effort and the magnitude of L/W/S profile data volumes would be less than for stochastic studies.

It would be straightforward to convert historical meteorological data into projected future hourly (or subhourly) wind and solar generation AC output to the grid. This conversion, such as conducted by NREL for an updated and expanded set of wind and solar profiles, is computed based on assumed generator locations (linked to meteorological data) and design configurations – scaled up to the MW capacities in a given planning portfolio. Attention would need to be paid to the assumed generator designs such as regarding the mix of tracking and panel/inverter capacity ratios for PV facilities in different locations. These assumptions can be consequential as has been pointed out in the 2014 LTPP proceeding, and as recommended in Section 6.2 below should be examined relative to available information about the deployed and anticipated mix of designs.

It is less straightforward to convert historical loads to multiple future load shapes for modeling, because future loads represent more than simple known physical responses to a few documented meteorological variables. It may not be necessary or desirable to "weather normalize" the historical loads since the intent is to combine load, wind and solar conditions that *actually co-occurred* in historical hours and days, reflecting whatever meteorology occurred at that time. For modeling a future year such as 2026 it may not be appropriate to simply scale any particular historical 8760-hour load profile so that it matches the specific annual energy and peak load forecasts adopted for the particular planning case. This would reduce the informative variability in historical loads over a

[67] http://www.nrel.gov/electricity/transmission/wind_toolkit.html and http://www.nrel.gov/electricity/transmission/sind_toolkit.html

series of successive years (e.g., 2007-2013), acting as if identical peak (and presumably near-peak) loads had occurred across all of those years. This is not what actually occurred, and such an approach would underrepresent historical variability.

The Commission has not generally specified precisely how load profiles, deterministic or stochastic, should be constructed for modeling. Yet, such modeling has been and is being done repeatedly. At this time CPUC staff does not have a precise proposal for constructing multiple historically-based deterministic load profiles for a future year that balances:

a. appropriate consistency with a single adopted load forecast for a given planning case,

versus

b. realistically capturing the range of variability of coincident L/W/S conditions experienced in recent years.

However, it should be possible to achieve such a balance for the 2016 LTPP. One general example of how this balance might be achieved is provided.[68]

The challenge of converting historical load profiles to load profiles for future conditions that may have fundamentally changed is neither new nor unique to the proposed development of multiple deterministic L/W/S sub-cases. For example, a single deterministic simulation uses a single historically-based load profile but this does not negate the obvious real world historical variability of yearly load profiles, and it is still necessary to confront the fact that fundamental load drivers may change significantly between the historical and future years (e.g., demand side measures and technologies, electric vehicles, perhaps rate designs). These challenges are also faced

---

[68] A simplified illustration of one possible approach for developing N different historically based load profiles for a future year planning case (e.g., for 2026) capturing historical load variability yet remaining consistent with the energy and peak load forecasts adopted for the planning case is as follows.

1. Produce growth-normalized load profiles for each of the N historical years of load data by adjusting the hourly historical loads by the rate of load growth over the N year historical period. This might, for example, raise slightly the loads in the first year.
2. Calculate a single composite load profile (8760 hours) for the N historical years by averaging the (growth-normalized) hourly loads for each hour, over the N years to produce a single ***composite historical load profile.*** This might include adjustment for weekends/holidays.
3. Develop the ***baseline future load profile*** (8760 hours, e.g., for 2026) for the particular planning case (a) by using current methods, (b) by scaling the ***composite historical load profile*** based on the energy and peak load forecast adopted for the future year planning case, or (c) by any other adopted method (e.g., if explicitly accounting for certain future demand-side measures, electric vehicle charging, etc),
4. Create the nth (out of N total) historically-based load profile for modeling the planning case (e.g., for 2026) by incrementing or decrementing each hourly load in the ***baseline future load profile*** by the same MW amount that the load for that same hour in the nth historical year (normalized for growth, as above) exceeded or fell below the corresponding load in the ***composite historical load profile***.

This captures via N different profiles the historical range of hourly load variability coincident with modeled hourly wind/solar variability, based on historical load variation from the ***composite historical load profile.*** This is in turn used to model future load variation from a future baseline load profile calibrated to the adopted load forecast. The result is greater load variability than if constraining each (future year) planning case load profile to have exactly the adopted annual energy and peak load forecasts. However, this expanded load variability is combined with wind and solar conditions (meteorology) that actually co-occurred historically with the load variability. It does not capture the possibility that, for example, future penetration of electric vehicles may change the pattern of year to year load variability, not just the baseline or average load pattern.

in developing future load profiles for *stochastic* modeling, even if they are overshadowed by concerns about realism of synthesized L/W/S combinations beyond those historically experienced.

To summarize, different formal planning cases established for the LTPP proceeding (such as a trajectory case, a high load case, a low GHG, a low hydro case, etc.) would represent alternative future conditions assumed to exist for the full study year, such as 2026. These cases might not be assigned quantitative probabilities or weights, although they might have different qualitative emphases. However, within a single formal case different L/W/S sub-cases would reflect alternative possibilities for L/W/S variability *within* the year, and might each have quantitative weights. This might permit calculation of quantitative performance relative to probabilistic reliability standards. However, if limited for example to 7 years of historical L/W/S conditions (the apparent extent of the new NREL wind data), it would not be technically possible to calculate “events in 10 years” without making additional assumptions such as regarding probabilities of different formal planning cases or their L/W/S subcases.

Apart from providing more robust and informative deterministic studies, a second *equally important* reason for developing multiple historically based L/W/S profiles for an individual formal planning case is to provide a basis for validating (developing confidence in) more sophisticated stochastic methodologies, as described in the next section.

6.2. STOCHASTIC MODELING

CPUC staff recommendations for validation of stochastic modeling are consistent with the Working Group 3 consensus that there needs to be better understanding and confidence regarding the development and modeling of L/W/S probability distributions and profiles. Also consistent with Working Group 3 views is the CPUC staff recommendation that this should involve both (1) evaluation of overall probability distributions and profiles and also (2) deep dives that extract and analyze key modeling variables during hours and in some instances entire days when generic or flexible capacity shortfall[69] or over-generation events are simulated to occur. Some of the validation methods discussed below may also be useful for assessing *deterministic* modeling studies brought into the LTPP proceeding.

The evaluation of overall L/W/S probability distributions and profiles used for stochastic modeling is expected to be a major data-intensive task unlikely to reach conclusion in time for full application to 2016 LTPP studies. However, it is desirable and possible that initial progress will inform 2016 LTPP studies.

CPUC staff recommends that overall L/W/S probability distributions and profiles used for stochastic modeling be compared and evaluated against analogous deterministic distributions and profiles developed in a more transparent (“we know it could happen”) manner using publicly available data - - “benchmarking” profiles. These benchmarking profiles should consist of the same extended set of multiple L/W/S profiles (“sub-cases” within individual formal LTPP planning cases) recommended for deterministic studies as described in the preceding section. The distribution of hourly load, wind, solar and resulting net load values over each month should be compared between benchmarking profiles versus profiles developed via sophisticated stochastic methodologies. This comparison should consider absolute upper and lower bounds as well as the tails (upper and lower percentiles) of the distributions, and should consider not only hourly MW levels but also frequency and magnitude of extreme ramps.

[69] Assessing system *flexibility* to address operational circumstances that dynamically evolve chronologically is more complicated and challenging than simply assessing the probability that total (generic) system *capacity not on outage* will meet or exceed total system load under a range of conditions not necessarily linked chronologically.

L/W/S correlations should be examined across the full ranges of the stochastically modeled profiles and the benchmarking profiles. Furthermore, there should be examination of whether there are L/W/S correlations occurring under specific conditions such as high or low load days in certain months – correlations that may be diluted or statistically insignificant when tested across the entire range of conditions. This should specifically include testing for L/W/S correlations under those L/W/S conditions modeled to produce extreme events. For example, if extreme events occur in the top 15 percent of load days in July or the bottom 15 percent of load days in March under stochastic modeling, do those days statistically have high or low wind or solar generation potential (meteorology) in either the data used to develop stochastic modeling profiles or in the benchmarking data?

If L/W/S profiles constructed via sophisticated stochastic methodologies contain a greater frequency of extreme values or extreme ramps compared to benchmarking profiles this does not necessarily mean that the sophisticated profiles are unrealistic or inappropriate. In fact, the sophisticated methodologies are designed to construct and model L/W/S conditions beyond those historically experienced. The question is whether such exceedance of historically experienced conditions is reasonable and appropriate for the kinds of studies and decisions being pursued.

CPUC staff also recommends using *deep dives* into stochastically modeled shortfall and over-generation events to evaluate whether specific combinations of L/W/S and other conditions giving rise to such events are credible. To aid consideration of what deep dive L/W/S conditions are "credible," we can compare those conditions with the benchmarking profiles described above. To the extent stochastically modeled L/W/S conditions associated with substantial shortfall or over-generation events are more extreme than the benchmarking profiles, we should ask if the magnitude of this difference is reasonable and explainable. Other conditions (besides L/W/S values) modeled to occur during deep dive extreme events such as forced outages, imports/exports, or use of storage and hydro facilities would also be evaluated based on reasonable expectations in these respective areas.

CPUC staff expects that deep dives will be most important and necessary for examining capacity shortfall events, which inherently occur in "perfect storms" of extreme conditions. However, it is recommended that deep dives also be pursued for over-generation events, but with lesser emphasis, such as by diving into fewer over-generation events.

Determination of what modeling variables should be extracted via deep dives that are not normally saved from a simulation will depend on the tradeoff between the value of retaining certain values versus the time, labor, and analytic complexity (including data volumes) for doing this. It is expected that selected modeling variables would be examined for at least the hour of an extreme event, and at least for a selected subset of such deep dive events values would be examined for an entire day. The minimum set of variables saved should include L/W/S values (MW per hour), forced outage amounts, net imports, unmet loads and/or reserve requirements, and the specified reserve commitment requirements. Other desirable information could include generation from different categories of resources (e.g., various renewable categories, gas-fired combined cycle and combustion turbine, run-of-river and storage hydro, other storage) as well as storage charging and available headroom (up to Pmax) and floor-room (down to Pmin) for dispatchable generation. (See the discussion of hydro and storage modeling validation below.)

Key considerations are how much labor and time are involved in extracting particular kinds of deep dive information, and whether a two-pass approach (defined in Section 5.2.2) is feasible for re-simulating selected events to save "non-standard" information. Working Group 3 discussions and comments produced useful perspective regarding potential difficulties and labor/time implications of deep dives as well as concerns regarding a two-pass approach. CPUC staff does not attempt to propose exactly how a practical deep dive process would be designed to balance benefits versus difficulties. However, we do recommend an early start to a collaborative effort to design a deep dive process integrated with other modeling activities. This design effort must consider benefit/cost tradeoffs in an informed manner and will clearly involve the modelers. It seems reasonable that informative results from deep dives may be available before assessment of overall L/W/S probability distributions and profiles has reached conclusions. Furthermore, it is possible that deep dives could also enhance understanding of circumstances producing "events" in *deterministic* simulations, perhaps as a starting point for subsequently analyzing those specific circumstances in greater detail, such as stochastically or with intra-hour dispatch.

As part of stochastic methodology validation, CPUC staff also recommends further "mechanistic"[70] clarification and validation of the stochastic methodologies for constructing L/W/S probability distributions and profiles. This could take the form of a workshop or a round of data requests from LTPP parties, but this may not be the most effective approach, especially since workshops and data requests have already occurred at some length. Instead, CPUC staff recommends proactively (early in the studies process) requiring fuller explanation of methodologies by the modelers, making greater use of numerical examples. This should be integrated with the future LTPP schedule, and would depend on what stochastic methodologies are applied and need to be understood going forward.

For both deterministic and stochastic modeling, CPUC staff also recommends evaluating solar (particularly PV) profiles especially in terms of assumed prevalence of different facility designs (such as regarding PV tracking and inverter loading ratios) since design assumptions impact the assigned output profiles. This evaluation will benefit from information recently requested regarding the mix of designs being deployed in the different utility areas. This information will be compared with the mix of designs assumed in construction of profiles used for modeling, including the new NREL wind and solar data that will likely inform studies in the next LTPP cycle. If necessary, it should be possible to adjust profiles based on this evaluation, for both stochastic and deterministic modeling.

6.3. Flexible Reserves Commitment Modeling

CPUC staff recommends that efforts be made to improve understanding and confidence regarding modeling of flexible reserve commitment requirements, possibly leading to some modifications. This applies to both deterministic and stochastic modeling. This issue does not represent as strong a barrier to using certain modeling methodologies to support Commission decisions as does the need for deterministic modeling to cover a wider range of conditions, or the need for better understanding and confidence in L/W/S probability distributions and profiles used in stochastic modeling. However, there is concern and uncertainty regarding both the justification of modeled flexible reserve commitment requirements and also their impact on key results such as production costs, over-generation, GHG emissions and identified flexibility problems.

It is challenging to validate or gain confidence in modeling of flexible reserves commitment requirements in part because the purpose of those requirements is to protect (yet not over-protect) against capacity shortfall events that are generally not explicitly modeled[71] and for which probability distributions of hourly shortfalls occurring subsequent to reserves commitments may not be transparent or not be computed. CPUC staff proposes the following seven-part approach to assessing and gaining confidence in modeling of flexible reserves commitment requirements. This could be pursued largely within the course of otherwise-planned studies. The first four parts involve clarifying the mechanics (specific computational logic and its rationale) of modeling flexible reserves commitment requirements and their shortfall consequences. The last three parts involve different ways of using

[70] "Mechanistic" in this context means concerning the actual computational logic and its underlying "real world" rationale.
[71] E.g., the amount of capacity shortfall may depend on divergence between explicitly modeled L/W/S values for an hour, versus other L/W/S outcomes that probabilistically could have occurred (e.g., reflecting forecast error) but were not modeled.

modeling runs (inputs and results) and other quantitative information to assess the quantitative reserves commitment requirements and their relation to identification of shortfall events.

The following explanations (parts 1-4) should be requested and provided in a proactive and systematic manner near the beginning of the next LTPP study cycle, rather than relying heavily on ongoing clarification questions raised in a more ad hoc manner during data requests or workshops.

1. There should be clear definition of what each kind of flexible reserves represents, such as X MW of load following up (LFup) being defined as the ability to ramp upward X MW in Y minutes and having the headroom to do so, and including whether/to what extent this can be provided by units not already committed to at least minimum capacity. This includes LF up and down, regulation up and down, and contingency reserves (spinning and nonspinning).
2. There should be clear explanation of what variabilities and uncertainties LF and regulation commitment requirements are intended to address and what data are used to compute these variabilities and uncertainties.[72] For example:
    a. Is LFup intended to address hour ahead (HA) net load downward forecast error (forecast too low) plus maximum intra-hour net load 5 minute (or other) upward excursions above the hour H to hour H+1 deterministic ramp trend - - or something else? At what probabilistic confidence level (e.g., 95 percent) is committed LFup intended to address the above (or other) errors/deviations? With what temporal granularity is the LFup commitment requirement recalculated and re-applied (for every hour, for each of 24 hours in a day but averaged over an entire month, etc.)? How does this LFup commitment requirement account for (add to, not double count…) positive or negative ramp needed to meet the next hour's deterministic ramp to meet deterministic hourly net load change, and where is commitment for the deterministic ramp accounted for? How does the calculation of LFup requirement account for the fact that up-ramp capability to address HA net load forecast error [forecast too low] also makes some contribution to meeting intra-hour 5 minute net load upward excursions?
    b. Similar clear explanation is also needed for LFdown and for regulation up and down.
3. To complement textual description for 1. and 2. above, there should be numerical examples from representative modeled hours that include:
    a. MW of deterministic load and of wind and solar generation, for that hour
    b. MW of HA forecast error for load, wind, solar, and net load [as probability distribution and/or confidence intervals]
    c. MW of intra-hour variability and forecast error for load, wind, solar and net load [probability distributions and/or confidence intervals] for whatever different temporal granularities were utilized such as 5-minute variations and forecast errors, or within-5-minute variations).]
    d. Resulting modeled LFup, LFdown, regulation up and regulation down commitment requirements for that example hour, quantifying (for each of the above) the contributions (MW and/or percentage) of the following drivers (or others, as applicable) to these computed reserve requirements:
        i. HA net load forecast error,
        ii. 5 minute intra-hour net load variability (excursions from hour trend),
        iii. 5-minute net load forecast error

[72] It should also be confirmed that spinning and nonspinning contingency reserves are intended to address outage contingencies, per current reliability standards and practices.

iv. sub-5-minute net load variability (excursions)

4. There should be clear explanation of how, mechanically (computational logic and its rationale) a model calculates or can calculate the *achieved* operating reserves (MW or percent) for any hour, as a probability distribution and/or a confidence interval. For example, suppose that the model is able to commit X percent (up to 100 percent) of the specified flexible reserves *commitment* requirement for a given hour. Then, how is the probability of *actually* keeping reserves above Y percent for that hour computed, when considering forecast error and intra-hour variability? For example, Y might be 0 percent in which case not meeting Y means loss of load, or Y might be 3 percent in which case not meeting Y means a stage 3 emergency. How does/can the model calculate and report the probability of failing to meet this Y percent reserve threshold as described above? H how does this account for reserves inadequacy occurring within some sub-hourly interval in the middle of the hour?

Parts 5-7 involve assessment of modeling inputs and outputs or other quantitative reserves and shortfall information, rather than explanation of methodology.

5. The MW amounts of modeled flexible reserves commitment requirements should be compared among different modeling methodologies, for comparable L/W/S conditions and comparable definitions for each kind of reserves. "Different" methodologies might include CAISO's Step 1 methodology, the methodology used in recent LTPP-associated modeling to develop wind and solar integration adders,[73] the methodology used by SCE for stochastic studies in the current LTPP cycle, methodologies used by different NREL studies, or other methodologies for which this information is available.
6. A sensitivity test should be conducted for a modeling case(s) initially having significant amounts of shortfall and/or over-generation, in which the amount of flexible reserve commitments is then substantially increased and/or decreased and no other inputs or assumptions are changed. One possibility is to limit load following and/or regulation requirements to those calculated to address only load uncertainty and variability.
7. Flexible reserve commitment requirements in a modeled planning case (e.g., for 2026) should be compared to reserve commitment requirements under the CAISO's current operating practices, such as by calculating reserves requirements as they would be calculated under current practices, but substituting the higher wind and solar levels from the planning case.

6.4. REGIONAL GENERATION REQUIREMENTS MODELING

For validating modeling of regional generation requirements, CPUC staff recommends pursuing essentially all of the issues and questions described above under Section 5.2.4, *Validating Regional Generation Requirements: Working Group 3 Views*. It may be effective to establish a task force comprised of staff from energy agencies (certainly including CAISO) and also from the transmission owners and from other knowledgeable and interested parties. Broadly, the inquiry should examine at least the following:

[73] CPUC staff notes that implications of modeling of regulation and load following requirements based on the CAISO's "Step 1" methodology versus the alternative methodology applied for integration adder studies have already been the subject of considerable comment by parties including requests for further evaluation. The requirements used for the integration adder studies were stated to be based on the methodology in NREL's Eastern Wind Integration and Transmission Study (http://www.nrel.gov/docs/fy11osti/47078.pdf), which appears to use a simpler methodology for calculating reserves requirements, compared to the CAISO's Step 1 (Pacific Northwest National Laboratory) methodology. A NREL source has indicated that NREL's Western Wind and Solar Integration Study 2 used a reserves requirements methodology similar but not identical to that used in the above NREL study.

1. the amount of and technical rationale for "need" regarding different individual kinds of reliability services (frequency response, reactive power, inertia, etc.) in different electrical/geographic areas, in different times or conditions;
2. what kinds of resources are able to provide each kind of service now, and also in the future (and what infrastructure or operational changes would be needed for future provision);[74] and
3. what can we learn regarding such regional requirements from (a) relevant CAISO operating practices or requirements, and (b) relevant reliability or other studies?

This inquiry should explore whether it is desirable and/or feasible (1) to explicitly model different individual kinds of reliability services (or their production simulation proxies) for a region (e.g., inertia versus reactive power versus primary frequency response), (2) to expand the range of resource types modeled as eligible to provide certain services (e.g., storage, demand response, resources not operating for energy provision), and (3) to identify whether only one or two types of services are most limiting in a given region such that committing or dispatching (or perhaps installing) sufficient resources for these purposes would effectively address the full range of reliability service needs not otherwise being accounted for in the production simulation. For example, reactive power might not be a most limiting service (i.e., would be sufficient whenever a more limiting service need was met), or inertia might be most limiting only in certain locations under certain infrequent conditions.

6.5. Storage and Hydro-Electric Modeling

CPUC staff recommends that the modeling of storage and of hydro be examined more fully, but that this be done in the course of running studies otherwise scheduled rather than requiring separate studies for this purpose. Deep dives into modeled shortfall and over-generation events should provide an efficient opportunity to extract and analyze useful information on how storage and hydro are being modeled. Those deep dives extracting and analyzing a *full day's* worth of modeling data may especially shed light on how storage operation and allocation of hydro energy are being modeled. At that point, need for additional investigations into storage or hydro modeling could be determined.

---

[74] For example, one Working Group 3 participant suggests the following eligibility criteria for provision of different services. Such criteria, or others, could be considered in the course of a Task Force effort to evaluate appropriate modeling of regional generation requirements, given that production simulation or modeling of comparable electrical detail cannot directly simulate AC reliability issues but must instead impose proxy requirements.

- For local ancillary service requirements, any generator able to provide the ancillary service of interest in the local area should be allowed to do so in a local ancillary service requirement constraint.
- For inertia, the inertia constant of generators can be used in a constraint that requires a certain amount of inertia to be online in specific areas. This constraint would presumably be enforced to avoid transient stability problems and help CAISO maintain adequate area control error performance.
- For primary frequency response, only resources with active governor or governor-like controls should be allowed to participate. The constraint here would force the model to reserve headroom on resources equal to the amount of primary frequency response needed.
- For production of electricity near to demand centers in the case of the failure of a key transmission line, it would seem that all resources in the relevant area that are producing electricity should be allowed to contribute, regardless of whether they are conventional generation or non-conventional.
- For voltage support, any requirement could be met by resources in the relevant area that are able to provide this service.

7. APPENDIX A

The next three subsections include CAISO's, TURN's, and PG&E's specific proposals for a comprehensive framework of reliability standards applicable to either stochastic and/or deterministic models; included immediately below CAISO's proposal is TURN's critique on the CAISO proposal. These three subsections have been retained as an Appendix to this version of Staff Proposal in order for the Commission to consider them, and contrast them to the Energy Division's proposal, if need be.

7.1.1. CAISO PROPOSAL FOR RELIABILITY STANDARD

Stochastic modeling study techniques have been used in this LTPP proceeding in order to incorporating a large set of possible system conditions in simulation results that help us investigate the need for additional system and flexible capacity necessary while integrating an RPS portfolio that delivers more than 33% of the energy used in California. In stochastic modeling, such as capacity shortfall and over-generation are presented in a probability (or frequency) distribution format. Based on the distributions, the probability of occurrence of a specifically defined incident – such as capacity shortfall greater than 1,000 MW or the number of hours with capacity shortfall greater than 350 – can be identified; thus, it can provide support for LTPP procurement decisions. In order to determine the need for capacity and system flexibility based on the probability distributions of the results, a criterion metric based on industry reliability standards needs to be established. Because the stochastic modeling approach in the CPUC LTPP proceeding has no precedent, there is no such reliability standard metric to be adopted. Working Group One's goal was, partly, to facilitate discussions that would lead to an agreement on such standard. The CAISO proposes a metric to serve as a starting point, with the rationale described below.

CAISO recognized several factors to consider including the following:

- *Loss-of-Load Incident Definition*: "… the production cost simulation based stochastic modeling approach used in the LTPP proceeding is vastly different than the traditional stacking-up capacity sufficiency study approach. The definition of the loss of load incident in LOLE calculation needs to reflect the new approach the LTPP studies. The reliability standard metrics should be able to apply to situations with insufficient capacity, as well as the situations with insufficient flexibility and over-generation. As demonstrated in the CAISO recent filing of no renewable curtailment study results, over-generation can cause reserve shortfalls and unsolved over-generation (dump energy). These are serious reliability issues.[75] Therefore, loss of load in the LTPP studies needs to be redefined may be as loss of certain level of load following capability, loss of certain level of ancillary services, and dump energy."
- *Metrics:* "The 1 day-in-10 years LOLE standard in traditional capacity sufficiency study was based on daily peak load. '1 day' of loss of load is a single occurrence. If it were hourly simulation, a loss of load incident would most likely be just some hours of the day, not the whole day.[76] In the LTPP studies, the simulations are hourly. The LOLE needs to be measured by number of hours-in-ten years. Then the duration of the incidents need to be counted. Also, if the cost of loss of load is to be considered in economically deciding the options to further reduce loss of load, the cumulative volume (MWh) of loss of load needs also to be counted."
- *Reliability Standards*: "The reliability standard metric should be based on existing industry standards that have been widely used. One of them is the 1 day-in-10 years loss of load expectation (LOLE), which is a standard widely accepted in the traditional capacity sufficiency studies."[77]
- *Reliability vs. Economic Trade-Off: "*It has been widely discuss about the trade-off between reliability and cost. It is important to understand that cost most likely cannot measure the bottom-line of reliability. The bottom-

---

[75] See the CAISO LTPP filing "Report of the No Renewable Curtailment Sensitivity Cases Studies" at http://www.caiso.com/Documents/May8_2015_DeterministicStudies_nocurtailment_ExistingTrajectory_40percentRPS_R13-12-010.pdf.
[76] R. Billinton, R. N. Allan, pp 484.
[77] R. Billinton, R. N. Allan, *Reliability Evaluation of Power Systems, Plenum Press*, New York and London (second edition, 1996), pp 38, 409.

line of reliability is the level of system reliability that should not be violated at any cost. It is the reliability standard the metrics represents. The trade-off can be considered only when the bottom-line reliability is observed. That is to determine the cost to further reduce the probability of loss of load."

- *Reliability Standard for Deterministic Modeling: "*Based on the reliability standard metric for stochastic modeling, a similar standard can be established for deterministic modeling. Considering that a loss of load incident in deterministic modeling would be in equivalent to ten incidents-in-ten years in stochastic modeling context, the definition of the loss of load incident for deterministic modeling should therefore be different than that for stochastic modeling."
- *Reliability Standard Metric Proposal: "*In the table below is the (CAISO's) proposed reliability standard metrics. The "loss of load" is first defined to cover the upward capacity shortfall situation and the situation with over-generation. The LOLE is calculated based on the number of hours of "loss of load" in ten years. This approach counts the combination the number incident and the duration of each incident. The (CAISO) recommends reporting the cumulative volume (MWh) of "loss of load", but does not set a standard for it. A similar standard is proposed for deterministic modeling."

The CAISO's proposed reliability standard metrics are illustrated in Table 5 Summary of CAISO's Proposed Reliability Standard Metrics

TABLE 5 SUMMARY OF CAISO'S PROPOSED RELIABILITY STANDARD METRICS

Modeling Method	Definition of "Loss of Load" Incident	Counting Method	Standard	Notes
Stochastic	• Capacity shortfall situation: loss of contingency (operating) reserve greater than 3% of load, loss of regulation-up, or unserved energy • Over-generation situation: loss of regulation-down or dump energy	Cumulative hours (not necessarily continuous) of incidents in each year each iteration	7 hours-in-10 years[78]	Report cumulative volume (MWh) of "loss of load" – is the system reliable if there are no more than 7 hours in 10 years of a capacity shortfall?
Deterministic	• Capacity shortfall situation: loss of load following-up greater than 50 percent of its requirement, loss of contingency (operating) reserve[79], loss of regulation-up, or unserved energy • Over-generation situation: loss of load following-down greater than 50 percent of its requirement, loss of regulation-down, or dump energy	Cumulative hours (not necessarily continuous) of incidents in a year	1 hour-in-1 year	

<u>TURN's Critique on CAISO "Preliminary Staff" Proposal</u>

TURN believes that the system conditions cited under the column titled "Definition of 'Loss of Load' Incident" (see Table 3, above), if they occur in modeling, will merit further investigation by the Commission in this and future LTPP investigations. Nevertheless, TURN cited the following concerns about the CAISO proposal.

The CAISO Proposal Muddies the Meaning of Standard Reliability Metrics: The CAISO's proposal would define "loss of load" incidents in a manner that is not consistent with the historic definition and application of the "one-day-in-10-years" reliability metric. The current reliability metric is used only to set peak capacity needs. For example, some ISO/RTO-managed eastern capacity markets run traditional loss-of-load models to establish reserve margins as the percentage at which Loss of Load Expectation (LOLE) meets a given reliability standard.

TURN Recommends That "Reliability Metrics" Be Relatively Simple to Compute and Apply: The CAISO proposal might muddy the use and application of reliability criteria because it mixes "capacity shortfall" concerns with

[78] See the discussion of conversion at R. Billinton, R. N. Allan, pp 484. The 7 hours-in-10 years is a proposed standard. It did not come directly from the reference.
[79] When comparing the "loss of contingency (operating) reserves between stochastic and deterministic models, the CAISO explained the following: "The scenarios in deterministic studies represent much milder conditions compared to that in stochastic studies. We need to reserve sufficient protection in planning based on such mild conditions. In other words, we cannot plan for black out based on the mild conditions. A stricter standard is needed in deterministic studies"

"over-generation" issues. Though over-generation may pose operational challenges, such challenges are of an essentially different nature than those posed by "capacity shortfalls." And as discussed below, it is premature to adopt a reliability metric related to over-generation. TURN recommends that "reliability metrics" be relatively simple to compute and apply and suggests that its proposal to create separate criteria for peak and flexibility needs provides such simplicity. TURN also believes the analysis of over-generation is still sufficiently immature that it is best performed through the focused use of deterministic modeling.

The CAISO Proposed Definitions of "'Loss of Load' Incidents" are Too Conservative: The CAISO proposal would also define "loss of load" in a manner far more stringent than any other definition known to TURN. For example, given data from its modeling in this docket, the CAISO proposal would define "loss of load" at the peak hour of 2024 as occurring when resources fall below "load plus 7.0 percent" (TURN's 7% calculation is in reference to the CAISO's original proposal; on July 27, 2015 the CAISO changed the amount of operating reserves in its stochastic "loss of load incident" definition to 3%, lowering the percentage calculation TURN estimated to 4%. The CAISO's original deterministic loss of load incident definition did not change) in stochastic simulations and below "load plus 8.8 percent" in deterministic simulations.[80] The CAISO's proposed definition goes far beyond current industry practice.[81] TURN recommends that this issue receive additional analysis and stakeholder input.

TURN Recommends That Lack of Capacity and Over Capacity Not Be Measured Using the Same Probability: As stated above, it is also premature to adopt a rigid reliability metric related to instances of over-generation. TURN is yet unaware of any precedents for such criteria in the industry. Given the differences in the challenges posed by a lack of capacity and an over-abundance of energy, it also does not seem appropriate that the importance of the two events be measured using the same probability, that is by the terms "7 hours-in-10 years" or "1 hour-in-1 year."

TURN Recommends That Flexible Capacity Be Measured Using a New, Separate Test Specific to This Need: TURN recognizes that the CAISO grid will face new flexibility challenges in the coming years and that some may argue the CAISO's proposed expansion of the definition of "events" is a means of addressing that challenge. But TURN's proposal would instead address such flexibility needs directly by proposing a new, separate test specific to this need, and explore over-generation issues through a series of deterministic production cost model simulations.

7.1.2. PG&E's Proposal for Reliability Standard

In California, the ongoing shift in portfolio composition from conventional to renewable energy resources requires an updated or new resource adequacy assessment to account for *system flexibility needs*; i.e., the need to have sufficient generation resources on the electrical system that enable the integration of the increasing number intermittent resources (such as wind and solar) onto the electric grid, while maintain a certain level of system reliability. In particular, the new resource adequacy framework must improve upon the existing framework in temporal granularity and provide clarity about reliability, policy, and cost tradeoffs associated with adding flexible resources.

Because the lack of system flexibility affects reliability, policy, and cost, the new resource adequacy assessment must also consider new criteria. For instance, a system with limited flexibility may find itself unable to respond within the hour to large forecast errors; as a consequence, this will affect reliability. In terms of policy, a system with less downward flexibility may need to curtail output from renewable resources during periods of high renewable production to prevent over-generation; this necessity could potentially impact a state's policy goals. When a system with limited flexibility is compared to a system with ample flexibility, the system with less flexibility

[80] Computed from data in *Woodruff Stochastic Reply Testimony*, Table 1 (p. 10) and *Woodruff Prepared Testimony*, Figure 2 (p. 16). The estimated annual peak hour "stochastic" operating reserve requirement of seven percent equals contingency reserves of six percent plus a "regulation up" need of one percent in the peak hour. The estimated annual peak hour "deterministic" operating reserve requirement of 8.8 percent equals the prior seven percent plus 1.8 percent, which is half of the "load following up" requirement in the peak hour of 3.6 percent.
[81] *Ibid*.

will undoubtedly incur more cost: some due to the cost of curtailing low variable cost renewable resources and others due to the inefficient dispatch.[82]

PG&E proposes a two-step assessment method that incorporates many elements of the CAISO and TURN proposals described in the table below.

Step One – Reliability Assessment: The assessment uses the industry accepted reliability standard, LOLE of no more than 1 day in 10 years, to measure both peak capacity and system flexibility events. The assessment records any loss-of-load events, regardless of whether it is due to lack of peak capacity or system flexibility. A peak-capacity-deficiency event occurs if the system does not have enough capacity in one or more hours in a day to cover its peak demand plus a minimum amount of reserves. A flexibility deficiency occurs if the system does not have sufficient flexible capacity to meet its hourly ramping requirement or the intra-hour forecast uncertainty and variability of its load or wind and solar generation. Step one addresses whether a given system has sufficient peak and flexible capacity to meet the basic reliability standard, regardless of cost. If the system fails this test, additional capacity is needed.

Step Two – Economic (And Policy) Assessment: The assessment uses a single economic criterion – a cost and benefit ratio – that measures the net value of additional system flexibility.[83] As discussed in the prior section, a system's flexibility needs show up under both reliability and economics assessment. Step two reveals whether additional flexibility – in the forms of resource upgrades, contract modifications, or new resources – may be cost-effective or needed to meet specific environmental targets. To the extent that resource upgrades, contract modifications, or new resource are needed, PG&E believes each individual Load Serving Entity (LSE), under the guidance and review of the CPUC, should be given the opportunity to pursue and procure additional flexibility that is cost-effective for its customers.

PG&E's preference is to use stochastic modeling for these two steps for two main reasons. First, the most commonly used reliability standard of a 1 day in 10 year LOLE is a stochastic metric, and stochastic modeling produces results that can be directly compared to the reliability standard. Second, a stochastic production simulation model also provides production cost, renewable curtailment, and carbon emissions results over a wide range of scenarios that are useful to understand not only in terms of expected or probability weighted outcomes, but also in terms of the uncertainty around those expected values.

7.1.3. TURN's Proposal for Reliability Standard

TURN is concerned that the stochastic methods and models that the Commission is reviewing are extremely complex, cumbersome to run and post-process, very difficult to audit and verify, and lack transparency – aside from being yet unproven. TURN is further concerned that the current Phase 1b Working Groups will address only the symptoms of these problems, not the problems themselves.

TURN suggests that the Commission and all stakeholders would benefit from a critical re-evaluation of the types of models that will be used to assess reliability and other planning needs in the 2016 LTPP proceeding and beyond. In particular, TURN recommends that simpler approaches be used for assessing some reliability criteria and for enabling more rapid and transparent resolution of such issues. More specifically, TURN recommends that stochastic modeling not be used to develop inputs for numerous production cost model simulations. Instead, TURN recommends that stochastic methods be used to estimate directly the probability of negative reliability outcomes based on the uncertainties of the major drivers of such outcomes, much as "traditional" Loss of Load Probability (LOLP) modeling has long been used to estimate the capacity needed to meet peak loads.

[82] This cost of inefficient dispatch and commitment is a subtler effect that is often overlooked. However, its impact in terms of dollars can be as significant, sometimes even more so than the cost of curtailment.
[83] This is a comprehensive measure that accounts for savings in total production cost, curtailment cost, and emission cost.

TURN proposes the following approach to address the three key planning issues identified in the modeling conducted over the past several years and particularly in the 2014 LTPP:

- Peak load needs should be analyzed using the traditional stochastic approach cited above,
- Flexibility needs should be analyzed using a stochastic variant of this traditional approach that compares the need for and availability of flexible capacity based on a single measure of such flexible capacity, and
- Over-generation and its mitigation should be assessed using deterministic production cost model simulations.

A chart showing the TURN proposed reliability standard metrics appears as Table 4, below.

TURN offers the following observations in support of its proposal:

- Consistent with industry practice, stochastic modeling would be performed for two key reliability metrics by focusing only on the most important variables affecting reliability
- Stochastic modeling would not be used to develop inputs for numerous production cost simulations, reducing simulation and analysis effort and increasing transparency and acceptance of the method and its results
- Deterministic modeling would be employed to its best advantage in the LTPP, which is creating snapshots of possible system operations that are valuable for analyzing over-generation and least-cost means for its management and for meeting other goals such as reducing Greenhouse Gas (GHG) emissions

TURN recognizes that substantial effort would still be necessary – including Commission resolution of major analytic issues – to implement the envisioned stochastic modeling of *peak* and *flexibility* needs. However, TURN believes the relative simplicity of the proposed stochastic modeling approach will allow development on a schedule that would enable the Commission to make well-founded decisions regarding long-term reliability needs in the 2016 LTPP proceeding, just as the current schedule envisions.

TABLE 6 Summary of TURN's Proposed Reliability Model

Planning Issue	Reliability Metric	Result	Modeling Method	Notes
Peak	Firm load interrupted not more than "1-day-in-10-years" due to lack of resources	Planning Reserve Margin of "X" percent above 1-in-2 peak load established to meet peak metric	Stochastic calculation using LOLP algorithm of existing model (e.g., PLEXOS, SERVM) or other purchased or custom model	1/ 2/
Flexibility	Criterion comparable to "1-day-in-10-years" metric to be developed for lack of flexible resources	Flexibility Range of "Y" percent (of (100 + X) percent) established to meet flexibility metric	Stochastic calculation using enhancement to LOLP algorithm in one of above models	3/ 4/ 5/ 6/
Over-Generation	Over-generation is economic issue, so no reliability metric applicable	LTPP will assess least cost means for managing over-generation and meeting other planning goals, such as reducing GHGs	Deterministic production cost modeling (e.g., PLEXOS)	7/

1/ Stochastic modeling of this metric could be conducted relatively quickly and easily using existing tools – possibly even using an existing PLEXOS database.

2/ Key inputs could be the expected values of forced outages, load, variable generation, and the uncertainty of each and correlations among them, and the expected generation portfolio,

3/ Enhancement to LOLP algorithm to estimate Flexibility needs could likely be made relatively simply to the same tool used to estimate Peak needs. (See note 1/ above.)

4/ Metric needs to be developed, particularly as to whether a lack of ramping – especially downward ramping – will result in a loss of load, or significant probability thereof.

5/ Existing "three-hour ramp" needs and capabilities may be a good proxy for flexibility needs in general.

6/ Key inputs could include the same types of variables identified in 2/ above, with focus on modeling "net load shapes" and the availability of flexible generation within the generation portfolio.

7/ A surplus of resources does not pose the same reliability challenges as a lack of resources; key instead is finding most cost-effective means of managing over-generation.

(End of Attachment)