A Multi-year Analysis of Renewable Energy Impacts in California: Results from the Renewable Portfolio Standards Integration Cost Analysis

Michael Milligan, Consultant
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401
USA
303-384-6927
Michael_Milligan@nrel.gov

Henry Shiu
California Wind Energy Collaborative
Department of Mechanical and Aeronautical Engineering
University of California, Davis
One Shields Avenue
Davis, CA 95616
530-752-2261
hjshiu@ucdavis.edu

Brendan Kirby
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831
USA
865-576-1768
kirbybj@ornl.gov

Kevin Jackson
Dynamic Design Engineering, Inc.
123 C St.
Davis, CA 95616
530-753-7961
KevinJacksonPhD@aol.com

Abstract

California's Renewables Portfolio Standard (RPS, Senate Bill 1078) requires the state's investor owned utilities to obtain 20% of their energy mix from renewable generation sources. The California Energy Commission (CEC), in support of the CPUC, organized a team to study integration costs in the context of RPS implementation. The analysis team is collectively referred to as the Methods Group, and consists of researchers from the National Renewable Energy Laboratory, Oak Ridge National Laboratory, staff of the California Independent System Operator, and staff from the California Wind Energy Collaborative. This group performed an analysis of 2002 renewable generation in California. These efforts estimated the impact of renewable generation in the regulation

and load following time scales, and calculated the capacity value of each renewable energy source using a reliability model. Since that time, a multi-year analysis covering 2002-2004 has been done. In addition to providing results from the three-year period, improvements in the data collection process have been incorporated into the study. This paper presents results from the multiyear analysis and the Phase III recommendations.

Introduction

The California Renewables Portfolio Standard requires a "least-cost, best-fit" strategy for selecting new generation projects to fulfill its renewable energy supply goals. This explicitly includes indirect integration costs in the bid evaluation process. In previous work integration costs were identified, valuation methodologies were defined, and a one year analysis of 2002 was performed.

The multi-year analysis documented in this report applies the integration cost valuation methodologies detailed in Phase III to a new multi-year dataset. The new analysis spans 2002 to 2004 and provided opportunities to verify the consistency of the methodologies and to further examine the practical issues associated with integration cost analysis. The methodologies were originally developed to be straightforward and were applied with little or no modification to capacity credit, regulation, and load following. They are described herein along with the analysis results. The methodologies, however, require good quality data and the difficulties encountered in assembling an adequate dataset hampered the analysis. Because these data issues will remain relevant to any future study, they are also detailed below. Finally, based on the experiences garnered from performing the multi-year analysis and resolving the data quality issues, recommendations are provided for future analyses.

Capacity Credit

Based on discussion with utilities and other stakeholders throughout the RPS Integration Cost project, several refinements were made to the ELCC calculation of renewable technologies. The Phase I report modeled the renewable variable generation using a probabilistic approach. This method is similar to what is often done with conventional units that have multiple output settings, each with an associated partial forced outage rate. As a result of extensive feedback during public workshops, the probabilistic method was replaced with a more direct approach that uses actual hourly output of the renewable generators.

The probabilistic approach is more appropriate as an indicator of future performance, where there are considerable uncertainties surrounding the timing of the power delivery from certain resources. Directly using hourly output is more appropriate for measuring past contributions to capacity from a variable resource. It does not consider alternative timing of the power delivery from variable resources, as does the probabilistic method. However, when multiple years of data are analyzed, this is not a significant limitation. Therefore, single-year estimates should be considered as such, and would be expected to vary somewhat from year to year. This was discussed in detail and applied in the Phase III update to the one year capacity credit study. In the multi-year study, we continued to use the direct hourly method.

Other improvements were made in the input data. For the multi-year analysis we utilized renewable generation data directly from the IOUs. This allowed us to bypass some data from CaISO's Plant Information (PI) system that suffered from data errors. Those errors were sometimes difficult to detect because the renewable generation data was aggregated, which tended to obscure the errors. The data errors caused artificial offsets to actual generation and injected unrealistic ramping behavior over long time periods into the data set. The CaISO data also had related problems with the reported nameplate capacity of the generator aggregates. The IOU data aggregates used for the multi-year analysis were the ones that most closely matched the CaISO data used in the regulation and load following analyses, below.

We were able to obtain one-minute hydro data from CaISO and used hourly averages of this data directly in the multi-year analysis. This is an improvement over the hydro modeling previously used. In Phase III, an optimal dispatch of hydro was used based on CEC information on monthly minimum and maximum flows and rough estimates of pond-storage and pumped hydro data. However, a significant portion of hydro energy is run-of-river, which is uncontrollable and subject to nature. This is similar to wind and solar, although hydro is less variable than wind and has different characteristics than solar. But ultimately, these forms of generation are not dispatchable. As discussed further in some of the workshops and the Phase III report, the impact of the hydro system on the hourly risk profile is significant. The results below support this view and also show the significant effect of interchange to neighboring balancing authorities.

The outcome of the public workshops during the Phase I work suggested that scheduled maintenance from conventional units should be eliminated from the modeling and was excluded in the one year and multi-year analyses. As was stated in the Phase III report, whether this should continue is a policy question. Workshop participants in the earlier phases of this project suggested that in principle, the capacity value of renewable generators should be independent from conventional maintenance scheduling.

Multi-Year Reliability Modeling and Discussion

Power systems experience a wide variety of conditions from year to year. Because load is generally sensitive to weather, unusually warm or cool temperatures can cause the load profile in a given year to diverge from "normal." Generation does not always respond in the same way to nearly identical load conditions. Because loads can change significantly from year to year, both in magnitude and timing, one would expect that reliability indicators such as LOLP would also change, perhaps significantly. Because LOLP is a key ingredient in calculating capacity credit, we began the analysis by collecting the results of the base case reliability model runs for each of the three year periods (as discussed below, note that 2004 is represented by data from September 2003 to September 2004). Figure 1is a LOLP-duration curve for each year, plotted on the same graph. We can see from the graph that 2004 exhibits a relatively sharp decline in LOLP as loads drop off from the annual peak. Much of the annual risk occurs in a smaller number of hours, whereas the curves for 2002 and 2003 indicate a more gradual decline. In 2002 the risk is spread over more hours. The significance of this graph is that the risk profile of the CaISO system, as measured by LOLP, changes from year to year. It is not possible *a priori* to determine which hours will have the highest risk, or even to predict the risk profile with certainty.

For a closer view, we generated a series of graphs for the three-year period that show not only the relationship between load and LOLP, but the overall impact that the hydro system and interchange have on risk. In general (ignoring hydro and interchange), the highest annual LOLP would be expected to occur during the peak hour. However, there are many factors that can cause LOLP in near-peak hours to exceed the LOLP on the system peak. Generator schedules, exchange schedules, and hydro generation are capable of responding to the high prices that accompany peak or near-peak loads, subject to operating constraints. It is therefore possible that real-time reserves are higher during system peak than at near-peak. These and other factors can contribute to a LOLP profile that is similar to, but does not match, the peak load profile.

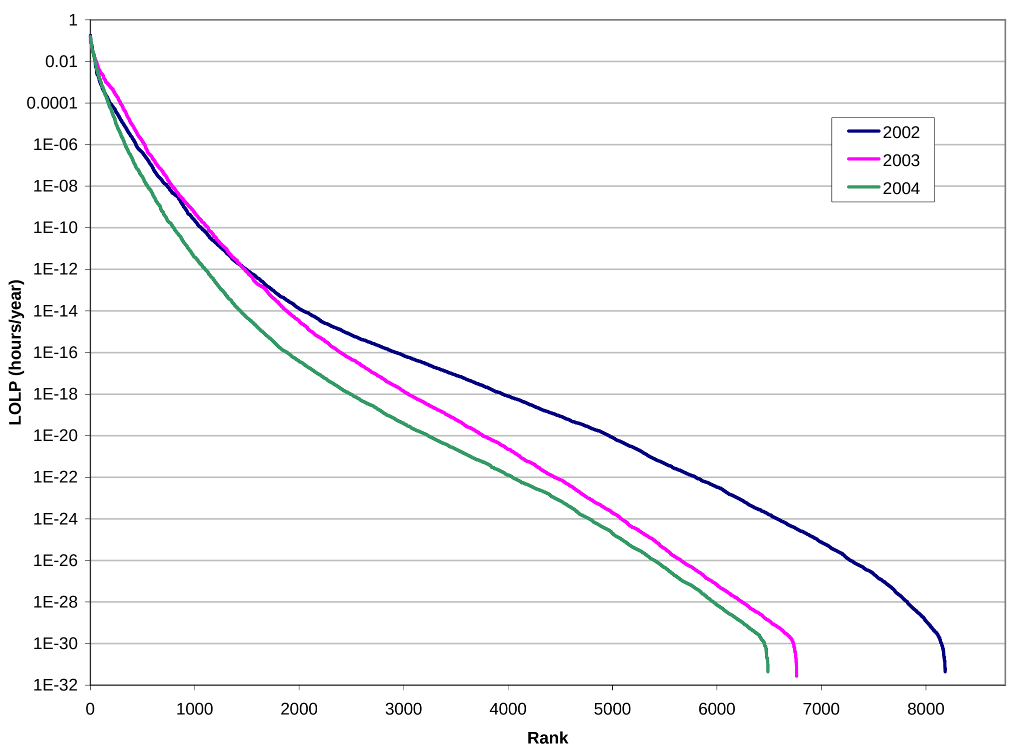

Figure 1. Hourly LOLP, ranked, 2002-2004

In Figure 2, the graph shows a typical load duration curve, in this case for 2002 (the blue line with the smooth characteristic). Superimposed on this graph are two additional rankings. The first shows the ranking of load by hourly LOLP (red). What the graph shows is that high load hours may generally be correlated with high LOLP, but the correlation is weak when we view the top 271 hours (the somewhat arbitrary cutoff point was LOLP >= 0.000001 days/year).

The final ranking on the graph (green) is based on the load that remains to be served after hydro and interchange have been taken into account. We refer to this as the load, net of hydro and interchange. Because hydro's and imports' forced outage rates are very low and/or cannot be objectively assessed, standard practice is to ignore forced outage rates for these resources. The implication is that the primary impact that hydro and imports have on system risk is to shift the timing of risk. For variable resources such as wind this further implies that for the generator to reduce annual LOLE, it must provide power during periods of high LOLP after taking account of hydro and imports/exports. This can have a significant impact on the LOLP profile, which is apparent from the figure.

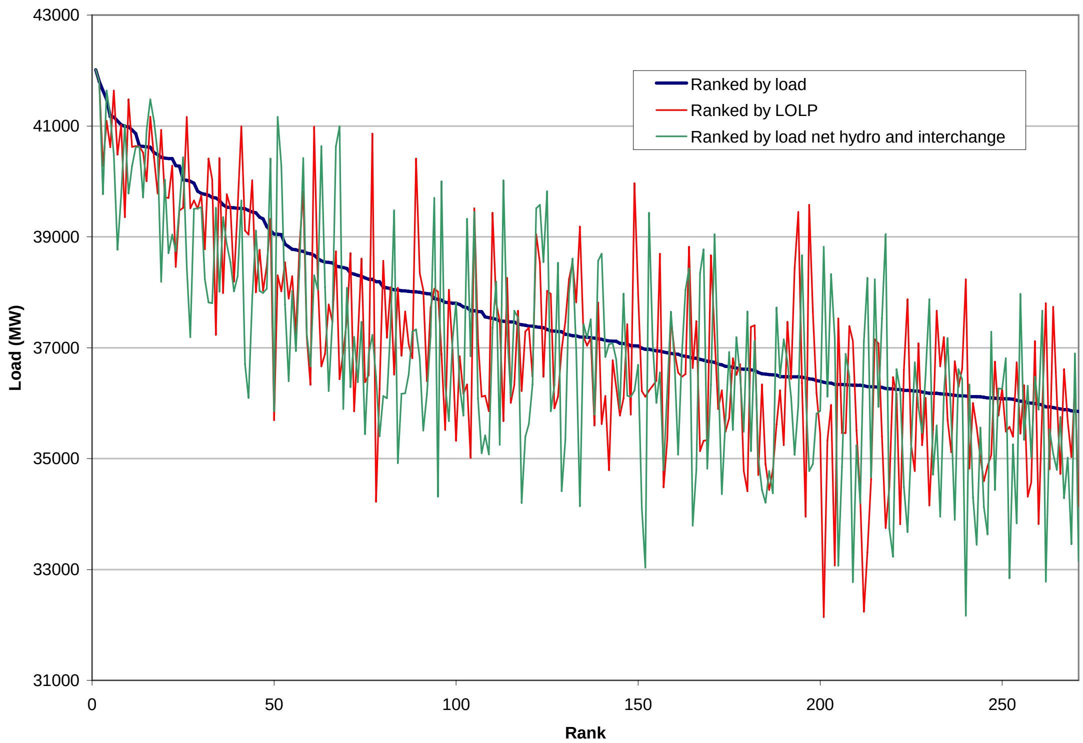

Figure 2. Load in 2002 during top risk hours, ranked by load, LOLP, and load net hydro and interchange.

Figure 3 takes a closer look at the load net hydro and interchange. The LOLP duration curve is not monotonically decreasing as a function of net load. If a variable resource delivers its energy during the high LOLP events, it will achieve a relatively high capacity credit. The timing of these high LOLP events will not necessarily correspond to highest load events. Similar characteristics were found in the 2003 and 2004 data sets, and are described in the original report.

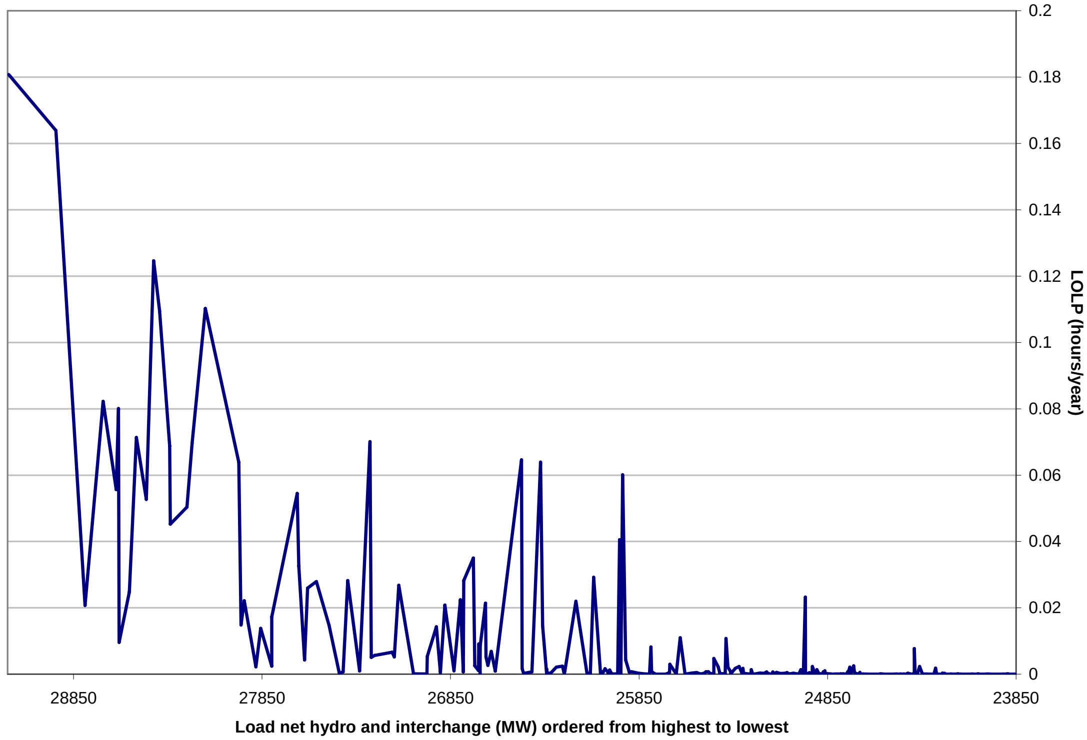

Figure 3. LOLP in 2002 at top hours of load net hydro and interchange.

ELCC results

All ELCC results were calculated based on the method described in the Phase III report, and with data received from CaISO and the IOUs. The analysis requires a complete year of data for each calculation. The input datasets used had complete years for 2002 and 2003, but not 2004. 2004 is represented by data from mid-September 2003 to mid-September 2004. It is simply referred to as "2004" for convenience.

As in the prior work, ELCC is measured relative to a benchmark unit, a gas combined cycle generator with a 4% forced outage rate and a 7.6% annual maintenance rate. All wind resources were modeled as time series, using the actual hourly generation provided by the IOUs for the full year. Transactions (interchange) and hydro were also represented by actual hourly data, obtained from CaISO. We note that in February 2002 there were some errors in the hydro data, which we patched through a combination of interpolation and pattern matching. Because LOLP during the month of February is so close to zero, this will not impact the results.

During the processing of the data for the analysis, some discrepancies were uncovered in the reported nameplate capacities of some of the generation aggregates. In prior work we reported capacity value as a percent of the annual maximum hourly generation for the resource in question. In the results below we have represented capacity value in three ways: (1) MW, (2) percent of maximum hourly output for the year, and (3) as percent of rated capacity as indicated by the IOU providing the generation data. In the case of wind, the relatively large discrepancy between actual generation and nameplate generation is probably an artifact of the older technology that still exists in some areas in California.

We believe that modern and future wind turbine technology will be more reliable than some past technology has been, minimizing this capacity discrepancy. If wind generation were to receive capacity payments, the wind operator would have an incentive to keep the turbines running and in good repair, especially during high load or LOLP events. Although we generally believe that capacity value should be represented as a percentage of nameplate capacity, this depends on having accurate nameplate values. The PG&E nameplate estimates do not match the maximum wind generation as well as those from SCE. Although this is not conclusive, it suggests that caution should be used in interpreting these capacity values. All of the data issues introduced above are discussed in further detail in a later section of this paper.

Table 1 shows the capacity value results expressed in terms of annual peak generation. To clarify, to calculate the relative ELCC for this table, the ELCC (in MW) is divided by the maximum hourly generation, for the resource in question, over the year.

Table 1. Capacity credit analysis results, based on annual peak generation.

Resource	2002			2003			2004		
	ELCC (MW)	Annual peak generation (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Annual peak generation (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Annual peak generation (MW)	Capacity credit (relative ELCC)
Wind (Northern Cal)	160	489	33%	170	463	37%	205	462	44%
Wind (San Gorgonio)	138	325	42%	89	317	28%	89	332	27%
Wind (Tehachapi)	168	584	29%	191	568	34%	167	571	29%

The capacity results for wind are different than those of the Phase III report. There are a couple of reasons for these differences. The hydro dataset used in this analysis is actual hydro, hourly, for the full year. In the Phase III work we were constrained to work with modeled hydro. Second, the generation aggregates used for the Phase I and Phase III one-year analyses differ somewhat from those in the current work.

Table 2 shows the results in terms of the reported nameplate capacity. As can be seen in the table, the percentage capacity values are generally lower than in Table 1. For the wind resources, we would expect the capacity value to decline when we use reported capacity as the basis of the capacity value, and we believe that using an accurate measure of nameplate capacity is the most appropriate metric.

Table 2. Capacity credit analysis results, based on rated capacity reported by the IOUs.

Resource	2002	2003	2004

	ELCC (MW)	Reported rated capacity (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Reported rated capacity (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Reported rated capacity (MW)	Capacity credit (relative ELCC)
Wind (Northern Cal)	160	679	24%	170	679	25%	205	680	30%
Wind (San Gorgonio)	138	357	39%	89	362	24%	89	362	25%
Wind (Tehachapi)	168	652	26%	191	659	29%	167	659	25%

To get an idea of the impact that hydro and interchange have on the LOLP profile, we removed them and re-ran the analysis. We show the results in terms of annual peak generation (Table 3) and in terms of reported rated capacity (**Error! Reference source not found.**).

Table 3. Capacity credit results with hydro and interchange removed; results based on annual peak generation.

Resource	2002			2003			2004		
	ELCC (MW)	Annual peak generation (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Annual peak generation (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Annual peak generation (MW)	Capacity credit (relative ELCC)
Wind (Northern Cal)	129	489	26%	129	463	28%	179	462	39%
Wind (San Gorgonio)	124	325	38%	69	317	22%	93	332	28%
Wind (Tehachapi)	175	584	30%	167	568	29%	178	571	31%

A comparison of the wind capacity values from Table 3 with the Phase III results shows a much closer correspondence. For example, in the Phase III report Altamont (Northern California) had a capacity value of 26% (based on maximum generation), San Gorgonio 31%, and Tehachapi 29%. The obvious outlier is San Gorgonio. The relatively good correspondence between some of these values may however be spurious, since there are substantial differences in the data sets used in the two analyses. Assuming accurate data, Table 4 provides the most accurate assessment of the capacity values that would have occurred in the absence of interchange and hydro.

Table 4. Capacity credit results with hydro and interchange removed; results based on rated capacity reported by the IOUs.

Resource	2002			2003			2004		
	ELCC (MW)	Reported rated capacity (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Reported rated capacity (MW)	Capacity credit (relative ELCC)	ELCC (MW)	Reported rated capacity (MW)	Capacity credit (relative ELCC)
Wind (Northern Cal)	129	679	19%	129	679	19%	179	680	26%
Wind (San Gorgonio)	124	357	35%	69	362	19%	93	362	26%
Wind (Tehachapi)	175	652	27%	167	659	25%	178	659	27%

Based in part on comments received by Solargenix during the Phase I discussions, we calculated the capacity factor for each renewable based on SCE's definition of the peak period: weekdays during the months of June through September (except holidays) between 12:00 p.m. and 6:00 p.m. As an alternative, we also included the month of May.[i] Table 5 shows the results of these calculations based on annual peak generation, and Table 6 shows the same information based on rated capacity.

Table 5. Capacity factor over peak hours based on annual peak generation.

Resource	2002		2003		2004	
	May through September	June through September	May through September	June through September	May through September	June through September
Solar	85%	90%	70%	76%	85%	89%
Wind (Northern Cal)	27%	27%	29%	30%	37%	35%
Wind (San Gorgonio)	41%	39%	28%	26%	34%	30%
Wind (Tehachapi)	36%	33%	28%	28%	33%	29%

Table 6. Capacity factor over peak hours based on rated capacity reported by the IOUs.

Resource	2002		2003		2004	
	May through September	June through September	May through September	June through September	May through September	June through September
Wind (Northern Cal)	19%	19%	20%	20%	25%	24%
Wind (San Gorgonio)	37%	36%	25%	23%	31%	28%
Wind (Tehachapi)	32%	30%	24%	24%	29%	25%

Table 7 collects results from Table 2 and Table 6. All ELCC values in the table are expressed as a percentage of the rated capacity and the peak capacity factors are all calculated based on the period from June through September and use rated capacity in the denominator. In each case we also calculated the three-year average. For some of the wind resource areas we have an excellent match between the three-year average ELCC and the three-year average peak capacity factors. Unfortunately this close match does not extend to the Northern California wind resource, which differs by about 5%.

Table 7. ELCC compared to peak capacity factors (June through September, weekdays excluding holidays, 12:00 p.m. to 6:00 p.m.) for three years, based on rated capacity reported by the IOUs.

Resource	2002		2003		2004		3-Year Average	
	ELCC	Peak	ELCC	Peak	ELCC	Peak	ELCC	Peak

	(% of rated capacity)	capacity factor	(% of rated capacity)	capacity factor	(% of rated capacity)	capacity factor	(% of rated capacity)	capacity factor
Wind (Northern Cal)	24	19	25	20	30	24	26	21
Wind (San Gorgonio)	39	36	24	23	25	28	29	29
Wind (Tehachapi)	26	30	29	24	25	25	27	26

Table 8 summarizes some of the key results as above, but instead uses ELCC values from the runs that exclude hydro and interchange. All ELCC values in the table are expressed as a percentage of the rated capacity and the peak capacity factors are all calculated based on the period from June through September and use rated capacity in the denominator. In this case we have a match between the results for the Northern California wind area and have a 2% difference in San Gorgonio. Because the hydro and interchange data were removed from these ELCC calculations, the ELCC results are not quite as accurate because of the missing resources. However, because of the lack of interchange and hydro, the relationship between load and LOLP is more straightforward.

Table 8. ELCC with hydro and interchange excluded compared to peak capacity factors (June through September, weekdays excluding holidays, 12:00 p.m. to 6:00 p.m.) for three years, based on rated capacity reported by the IOUs.

Resource	2002		2003		2004		3-Year Average	
	ELCC (% of rated capacity)	Peak capacity factor	ELCC (% of rated capacity)	Peak capacity factor	ELCC (% of rated capacity)	Peak capacity factor	ELCC (% of rated capacity)	Peak capacity factor
Wind (Northern Cal)	19	19	19	20	26	24	21	21
Wind (San Gorgonio)	35	36	19	23	26	28	27	29
Wind (Tehachapi)	24	30	25	24	27	25	26	26

Discussion of results

During the prior phases of this project, there have been numerous discussions regarding whether nameplate capacity assignments for existing wind resource areas were correct. Although it is useful to measure capacity value in MW, it is difficult to properly interpret the effectiveness of the resource if its rated capacity is unknown. Although this has been an issue in this project, we believe that it will be less of a problem with new wind facilities. If capacity payments are to be made to renewable (or other) generators, the incentive provided by the payment should be an inducement to ensure generator availability. Perhaps even more important is the evolution in wind turbine technology. Modern turbines are quite unlike many older turbines currently installed in California. Combined with taller towers and larger rotors, energy can be generated at lower wind speeds than with older technology. We expect that the capacity credit, however calculated, will be significantly different for modern/future wind turbines. Going forward, we do not believe that large numbers of turbines will be unaccounted for if good engineering and business practices are followed.

With the uncertainties surrounding data quality during this project, it is hard to know the extent to which data inaccuracies influence the results. We have much better confidence in the revised data sets used for this analysis than in the past. Data confidentiality issues have made it difficult to fully assess the results, particularly given the confidential aggregations of renewable generators.

The ELCC for the wind generators that were calculated for this project indicate the reliability contribution to the generator fleet in California. There are many moving parts that are captured by the model as a snapshot. For example, there may be significant synergies between hydro operations and wind (and other renewable technologies). Based on discussions during this project it appears that the hydro system is dispatched independently of the variable generation. With improvements in forecasting, especially for wind, it is possible that some incremental reliability can be gained by exploiting these potential synergies.

We ran several alternative scenarios to calculated ELCC. It is clear that hydro and interchange make a difference in the LOLP profile, and therefore on the ELCC of variable generators. It is also evident that ELCC results are not necessarily transparent. We found a reasonably good correspondence between ELCC and capacity factors that were calculated over the peak period. Whether to use ELCC or a capacity factor approximation is a policy decision. The overriding factor that would seem to favor ELCC is that it is a rigorous method that explicitly considers risk via the LOLP equation. Any approximation method will fall short. Conversely, a simpler method such as that considered above can come close and our examples showed that over three years, differences in methods may become less important. Simple methods also have the advantage of transparency and ease of reproduction.

Regulation

The method for calculating regulation costs was developed by Brendan Kirby et al at Oak Ridge National Laboratory (ORNL). The methodology and its results are described below with background material presented in the original Multiyear Report. The regulation analysis methodology has been applied to a variety of other control areas to quantify the ancillary service impacts of loads and variable resources. It determines the regulation and load following impacts to the control area. These impacts are the result of fluctuations in aggregate load and/or uncontrolled generation that must be compensated. Once the requirements are quantified, the method then determines the costs incurred in terms of greater amounts of purchased regulating capacity and greater use of the short-term energy markets.

Ancillary Services Terminology and Overview

Terminology associated with ancillary services has not been standardized across the utility industry and this sometimes has led to confusion. It is important to distinguish between the *impacts* imposed upon the power system and the *resources* or *services* the

CaISO utilizes to compensate for these impacts. The impacts in the regulation time frame are imposed upon the power network by loads, uncontrolled generators, and transactions. The resources or services that compensate for these impacts are supplied by generators responding to *automatic generation control* (AGC) and the *automated dispatch system* (ADS).

Regulation and load following are intimately related; both continuously balance aggregate load and generation within the control area. The two services differ in the time frame over which they operate with regulation operating minute-to-minute while load following operates over a ten minute or longer time frame. In 1996 the Federal Energy Regulatory Commission (FERC), defined six ancillary services in its Order 888. This order did not discuss load following. Perhaps because of this omission, most utilities and independent system operators (ISOs) do not include load following in their tariffs. The absence of this service required some ISOs to acquire much more regulation than they otherwise would need. Perhaps because of these problems, FERC, in its notice on regional transmission organizations (RTOs), proposed to require that RTOs operate real-time balancing markets.[ii] The responsive resources for these supplemental energy markets are generators that can change output every ten minutes as needed to follow load.

The CaISO obtains responsive resources to achieve the required real-time balancing of generation and load from the hourly regulation markets and the short-term energy markets. The alignment between the impacts that the CaISO must meet and the services it procures to meet those impacts is not perfect. Resources procured through the regulation markets, for example, could be used to provide load following, accommodate energy imbalance, or even supply base energy if there were no other alternatives. Load following itself is not a service which the CaISO procures directly. The CaISO meets its load following needs through short-term energy transactions, including both AGC generators and the supplemental energy market. Load following results are discussed in a later section of this paper.

Definition of Regulation and Load Following

Loads within a control area can be decomposed into three components: base energy, load following, and regulation, as shown for a hypothetical weekday morning in Figure 4. Starting at a base energy of 3566 MW, the smooth load following ramp (blue) is shown rising to 4035 MW. Regulation (red) consists of the rapid fluctuations in load around the underlying trend, shown here on an expanded scale to the right with a ±55 MW range. Combined, the three elements serve a total load (green) that ranges from 3539 MW to 4079 MW during the 3 hours depicted.

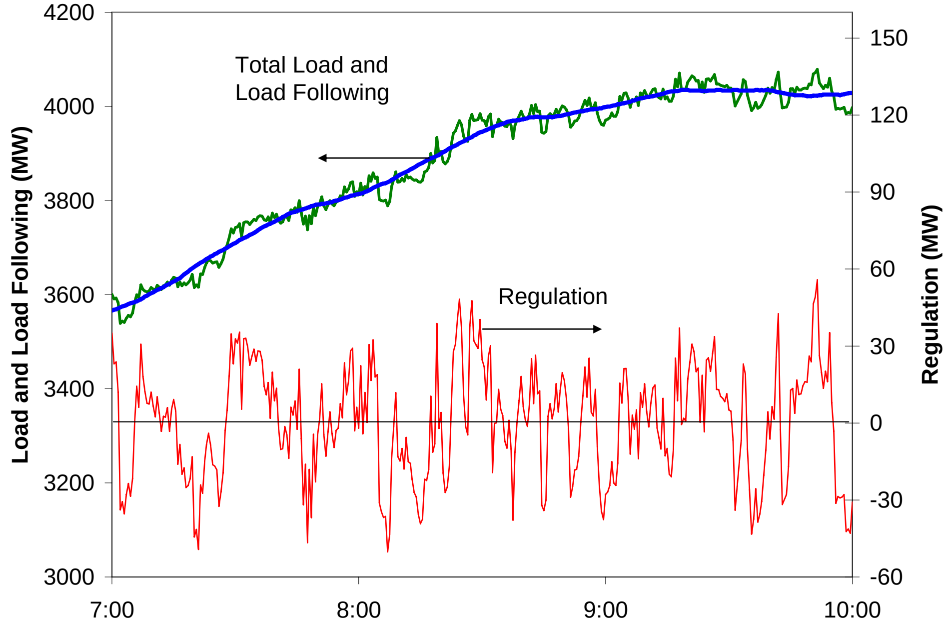

Figure 4. Decomposition of hypothetical weekday morning load.

The system responses to the second and third components are called load following and regulation. These two services ensure that, under normal operating conditions, a control area is able to balance generation to load. The two services are briefly defined[iii,iv,v] as follows:

- *Regulation* is the use of online generating units that are equipped with automatic generation control (AGC) and that can change output quickly (MW/minute) to track the moment-to-moment fluctuations in customer loads and to correct for the unintended fluctuations in generation. In so doing, regulation helps to maintain interconnection frequency, manage differences between actual and scheduled power flows between control areas, and match generation to load within the control area. This service can be provided by any appropriately equipped generator that is connected to the grid and electrically close enough to the local control area that physical and economic transmission limitations do not prevent the importation of this power.

- *Load following* is the use of online generation equipment to track the intra- and inter-hour changes in customer loads. Load following differs from regulation in three important respects. First, it occurs over longer time intervals than does regulation, 10 minutes or more rather than minute to minute. Second, the load-following patterns of individual customers can be highly correlated with each other, whereas the regulation patterns are largely uncorrelated. Third, load-following changes are often predictable (e.g., because of the weather dependence of many loads) and have similar day-to-day patterns.

There is no hard-and-fast rule to define the temporal boundary between regulation and load following. If the time chosen for the split is too short (e.g., five minutes), too much of the fluctuations will appear as load following and not enough as regulation. If the boundary is too long (e.g., 60 minutes), too much of the fluctuations will show up as regulation and not enough as load following. But in each case, the total is unchanged and is captured by one or the other of these two services. A 15-minute rolling average is recommended here to separate regulation from load following. The rolling average for each 1-minute interval should be calculated as the mean value of the seven earlier values of the variable, the current value, and the subsequent seven values. For load:

$$\text{Load Following}_t = \text{Load}_{estimated\text{-}t} = \text{mean}\ (L_{t-7}, L_{t-6}, \ldots, L_t, L_{t+1}, \ldots, L_{t+7}) \qquad \text{Equation 1}$$

$$\text{Regulation}_t = \text{Load}_t - \text{Load}_{estimated\text{-}t} \qquad \text{Equation 2}$$

This method is somewhat arbitrary and imperfect. It is arbitrary in that the time-averaging period (15 minutes as recommended here) and the temporal aggregation of raw data (1 minute) cannot be predetermined. In principle, the control-area characteristics (dynamics of generation and load and the short-term energy market interval) should determine these two factors.[vi] The 15-minute rolling average is recommended because it provides good temporal segregation and captures the characteristics of California's supplemental energy market.

In practice, system operators cannot know future values of load. They generally produce short-term forecasts of these values to aid in generation-dispatch decisions. While aggregate load forecasts are typically well developed, and a short-term energy market now operates in California, short-term forecast methodologies for non-dispatchable conventional and renewable generators are not. The rolling average has proven to be a reasonable analytical substitute in studying other control areas. The rolling average, like the system operator through the use of the short-term energy market, is constantly moving the regulating units back to the center of their operating range. If consistent, robust short-term forecasts are available and verified for all of the renewable generation technologies, this analysis can be performed without the use of a rolling average.
The use of the rolling average rather than the short term forecasts can impact the allocation of variability between the regulation and load following services slightly. Significantly, the method assures that total variability is captured in one or the other service and that there is no double counting. The distinctions between regulation and load following are discussed in another section of this paper.

Regulation Analysis Methodology

The regulation analysis methodology quantifies the regulation impacts of loads and generating resources within a control area. These impacts are the result of fluctuations in aggregate load and/or uncontrolled generation that must be compensated. Once the requirements are quantified, the method then determines the costs incurred in terms of greater amounts of purchased regulating capacity.

The regulation requirement of the entire system is first determined by taking the standard deviation of the 1 minute regulation values (applying Equation 2) for the total system. This is done hourly because the regulation market clears hourly. It is then possible to calculate individual contributions to that total requirement. Regulation aggregation is nonlinear; there are strong aggregation benefits. It takes much less regulation effort to compensate for the total aggregation than it would take if each load or generator compensated for its regulation impact individually. An allocation method should:

- Recognize positive and negative correlations
- Be independent of sub-aggregations
- Be independent of the order in which loads or resources are added to the system
- Allow dis-aggregation of as many or few components as desired

The method used here, meets these criteria. It was developed to analyze the impacts of nonconforming loads on power system regulation and works equally well when applied to non-dispatchable or uncontrolled generators. The allocation method does not require knowledge of each individual's contribution to the overall requirement. Specific individual contributions can be calculated based upon the total requirement and the individual's performance. Because regulation is composed of short, minute-to-minute fluctuations, the regulation component of each individual is often largely uncorrelated with those of other individuals. If each individual's fluctuations (represented by the standard deviation, σ_i) is completely independent of the remainder of the system, the total regulation requirement (σ_T) would equal:

$$\sigma_T = \sqrt{\sum \sigma_i^2} \qquad \textit{Equation 3}$$

where i refers to an individual and T is the system total

For the case of uncorrelated contributions, the share of regulation assigned to each individual is:

$$Share_i = \left(\frac{\sigma_i}{\sigma_T}\right)^2 \qquad \textit{Equation 4}$$

The more general allocation method, presented in Equation 5, accommodates any degree of correlation and any number of individuals. This allocation method is more complex but no more data-intensive than the previous method. This method yields results that are independent of any sub-aggregations. In other words, the assignment of regulation to generator (or load) g_i is not dependent on whether g_i is billed for regulation independently of other non-AGC generators (or loads) or as part of a group. In addition, the allocation method rewards (pays) generators (or loads) that reduce the total regulation impact.

$$Share_i = \frac{\sigma_T^2 + \sigma_i^2 - \sigma_{T-i}^2}{2\sigma_T} \qquad \textit{Equation 5}$$

The general allocation method (Equation 5) is recommended for analysis of the impacts of various individual renewable generators on the overall system's regulation requirements.

Calculated hourly regulation requirements are compared with actual hourly regulation purchases by the CaISO and hourly regulation self-provided by scheduling coordinators. Typically, three to five standard deviations of regulating reserves are carried to assure adequate CPS (Control Performance Standards) performance. Total regulation requirements are then allocated back to individuals. Hourly regulation costs are used to allocate the cost of regulation back to individuals. All of the CaISO's regulation requirements are allocated based upon the short-term variability impacts of the loads and renewable generators.

Data Requirements

Studying regulation requires one-minute, synchronized, integrated-energy, time series data for total control area load and the individual renewable resources of interest.
At a minimum, the data list must include time series data for:

- Total load
- Each renewable generator of interest

Experience has shown that it is also wise to perform an energy balance around the control area to assure data integrity. This requires 1-minute data for total generation, net actual imports/exports, net scheduled imports/exports, system frequency (and the frequency bias), and ACE. The data list should include one minute, synchronized, integrated-energy, and time-series data for:

- Total generation
- Net actual imports/exports
- Net scheduled imports/exports
- Area control error (ACE)
- Frequency (and frequency bias) – often provided as a deviation from scheduled frequency

Regulation analysis requires only one system data element plus one for each renewable generator of interest, each minute. Verifying data integrity requires an additional five system data elements each minute.

The CaISO runs hourly markets for regulation up and regulation down. Price and quantity data from these markets are used to determine practical quantities and costs of procured regulating resources. Scheduling coordinators are also allowed to self-provide regulation. The amount of self-provided regulation must be added to the amount of purchased regulation to obtain the total regulation amount. There is no price associated with self-provided regulation so the market price of the purchased regulation for the same hour is used to calculate the total dollar value of regulation for each hour.

- Hourly regulation-up price
- Hourly regulation-down price
- Hourly MW of regulation-up procured (hour ahead and real-time)
- Hourly MW of regulation-down procured (hour ahead and real-time)
- Hourly MW of regulation-up self-provided
- Hourly MW of regulation-down self-provided

Analysis Changes from Phase I and Phase III

In the spring of 2005, an independent review* of the Phase I report revealed that the calculation of the total system compensation requirement did not include the renewable generators' variability along with the total load variability. Only the total load variability was included. The methodology implementation description above now explicitly includes the individual generators as well as the load.

Later, a one minute data misalignment was discovered in the wind data for San Gorgonio used in the Phase I analysis. The misalignment only affected the regulation results because its effect is suppressed by the hourly and ten minute averaging used by the capacity credit and load following calculations. A revised set of results for the Phase I regulation analysis is presented below. This includes the complete calculation of the total system compensation requirement and synchronized data for San Gorgonio.

Table 9. Original and corrected results of the Phase I (one year, 2002) regulation analysis. Negative values are costs to the system.

Resource	Regulation Cost ($/MWh or mills/kWh)	
	Original	**Corrected**
Total System	-0.42	-0.44
Total Load	-0.42	-0.41
Wind (Altamont)	0.00	-0.22
Wind (San Gorgonio)	-0.46	-0.08
Wind (Tehachapi)	-0.17	-0.53
Wind (Total)	-0.17	-0.33

The results for the total system and for load remain approximately the same because the load represents the majority of variability in the entire system. However, because the variability of the individual generators was not originally included in the total system regulation requirement, the amount of variability allocated to each generator was

* The independent review was performed by Matthew Barmack of Analysis Group, Inc. When he could not duplicate the Phase I regulation results, we investigated further and found the omission in the total system compensation calculation. We are grateful to Matthew

understated. The decrease in San Gorgonio is not a result of including its variability in the total regulation requirement, but because of the correction of the one minute misalignment in its generation data (a calculation with the original misaligned data indeed results in a cost increase). The cost for San Gorgonio is several times lower than the other wind regions. This may be an anomaly, as shown in the multi-year results for San Gorgonio, below. The results are discussed further along with the multi-year analysis results in the following section.

The datasets used in the multi-year analysis vary somewhat from the datasets used in the Phase I one year analysis. The CaISO multi-year dataset has expanded aggregates in an attempt to better represent the generators being studied. However, the multi-year dataset exhibited new types of errors. To address these errors, the multi-year dataset was reviewed and checked for errors using data from PG&E and SCE as bases of comparison.

The multi-year analysis replaced the Altamont aggregate with an aggregate including plants from Altamont, Solano, and Pacheco; this was necessary to more closely match the corresponding PG&E data aggregate that it was compared against. Because of gaps in the 2002 biomass and solar data, the 2002 biomass and solar regulation analyses were run normally, but the runs for the other generation aggregates excluded biomass and solar from their calculation of the total system compensation requirement. This was considered a reasonable approximation because results from the 2002 one-year analysis are available for comparison. All of the data issues are detailed in the original report.

Multi-Year Regulation Analysis Results and Discussion

The methodology described above was applied to the CaISO multi-year dataset. The results of the multi-year analysis appear below.

Table 10. Results of regulation analysis of multi-year dataset. Negative values are a cost.

Resource	Regulation Cost ($/MWh or mills/kWh)[1]		
	2002	**2003**	**2004**
Total System	-0.42	-0.47	-0.39
Total Load	-0.41	-0.46	-0.36
Wind (Northern California)	-0.24	-0.40	-0.33
Wind (San Gorgonio)	-0.09	-0.43	-0.58
Wind (Tehachapi)	-0.57	-0.70	-0.56
Wind (Total)	-0.36	-0.53	-0.47

The 2002 results from the multi-year analysis and the one year analysis (Table 9) match well. There is some minor variation, but this is expected as the composition of the generation aggregates are not exactly identical.

In general, regulation costs increased slightly from 2002 to 2003 and then fell again in 2004, although not to previous levels. The calculated regulation purchase amount and costs are scaled from actual regulation commitment and purchase data from the CaISO OASIS database, which is shown below in Table 11.

---

1 Using $/MWh as a metric for regulation is both useful and dangerous. It is useful because what we really want to know is how much this ancillary service (something we are forced to buy but don't really want) adds to the cost of electricity (something that does useful work for us and we do want to purchase). In that sense a metric that is in the same units ($/MWh) as the commodity we are purchasing is very useful. It is dangerous because the amount of regulation required and the price have almost nothing to do with the amount of energy consumed or produced. The amount of regulation depends upon the short-term volatility of the generation or load, not the energy consumption or production. Use $/MWh in reference to regulation with great caution.

Table 11. Actual regulation amounts committed in the CaISO control area, 2002-2004.

	2002	2003	2004
Regulation up, self provided (MW-hr)	1,855,270	1,769,493	1,972,175
Regulation down, self provided (MW-hr)	2,078,057	1,797,975	2,073,533
Regulation up, procured (MW-hr)	1,659,438	1,116,009	1,109,265
Regulation down, procured (MW-hr)	1,627,342	1,488,440	1,255,973
Total regulation (MW-hr)	7,220,107	6,171,916	6,410,947
Total value ($)	98,270,561	109,357,025	88,141,708
Average regulation price ($/MW-hr)	13.61	17.72	13.75

In Table 11 above, note that MW-hr is the commitment of one MW of capacity for one hour and is not the same as MWh, a unit of energy. Also, as stated above, there is no price associated with self-provided regulation so the market price of the purchased regulation for the same hour is used to calculate the total dollar value of regulation for each hour.

Between 2002 and 2003, the actual amount of regulation committed over the entire CaISO control area decreased by 15%. However, the average price increased by 30%, resulting in a net increase in cost of 11%. From 2003 to 2004, the amount of regulation committed stayed approximately the same with a 4% increase. The price returned to 2002 levels resulting in a net cost decrease of 19% between 2003 and 2004.

The calculated regulation costs for the total system requirement and total load follow this pattern closely. In all three years, the regulation costs of the total load are very close to that of the total system requirement, a result of the sheer size of the load. The results could have been different only if one or more of the other studied resources had a dramatic regulation impact. A single large arc furnace, for example, would have sufficient impact to alter the cost of regulation for the rest of the load. None of the resources studied have that sort of regulation impact. In fact, the generating resources studied have quite minor impacts on total system regulation requirements.

Ignoring the outlying low value of San Gorgonio in 2002 for now, the regulation costs of the wind aggregates range from $0.24/MWh to $0.70/MWh. Not unexpectedly the wind plants impose a small regulation burden on the power system within the same order of magnitude as load when evaluated on a per MWh basis. This was expected because there is no apparent mechanism that would tie the wind plant performance to the power system's needs in the regulation time frame. The regulation burden is low because there is no mechanism that ties wind plant fluctuations to aggregate load fluctuations in a compounding way either. Wind and load minute-to-minute fluctuations appear to be uncorrelated. Hence they greatly benefit from aggregation.

The variation in regulation costs across the three wind regions may be a result of geography, technology, and turbine numbers. The Northern California wind aggregate, for example, has lower costs all three years than the other two regions (again, ignoring San Gorgonio in 2002), possibly because it is composed of the largest numbers of turbines[vii].

The inter-annual changes in regulation costs for Tehachapi follow the overall trend of actual regulation commitment in the CaISO control area. The Northern California wind aggregate does too, but to a lesser extent between 2003 and 2004 when the cost increased 67%. San Gorgonio is unique among all the resources studied, showing a 378% jump between 2002 and 2003 and then further increase instead of a decline between 2003 and 2004. The $.09/MWh value for 2002 is significantly lower than any of the other annual wind regulation results. San Gorgonio's individual variability, as defined in Equation 3.17, is not significantly lower in 2002 than 2003. There are also no known mechanisms that would correlate (or not correlate) its fluctuations in the regulation time frame to the rest of the system any differently in 2002 than in any other year. The 2002 value therefore remains anomalous. It was confirmed with the results from the analysis of the 2002 one year dataset, but it may be possible that there are underlying, undetected issues with the 2002 San Gorgonio data in both the one year and multi-year datasets. The 2003 and 2004 results are more consistent with the results of the other regional wind aggregates.

Overall, the regulation analysis results are reasonable. Because (1) inter-annual variations exhibited by some resources were disproportionate to changes in actual purchases amounts, (2) large amounts of new capacity will be installed in the future, and (3) technology and operation changes may have a significant effect, the continued understanding of regulation impacts and costs would benefit from more analysis over future years. Analysis with the methodology as described remains straightforward, given the availability of sufficient quality data.

Load Following

In this section we focus on the renewable resource impacts in the load following time frame, which generally encompasses periods ranging from ten minutes up to a few hours.

Overview

Load and generation must be continuously balanced on a nearly instantaneous basis in an electric power system. This is one of the characteristics that makes supplying electricity different from providing any other public good such as natural gas, water, telephone service, or air traffic control. It is a physical requirement that does not depend on the market structure. How load and generation are balanced does depend, in part, on the structure of the electricity markets. One benefit of interconnecting multiple control areas is that balancing load and generation within a single control area does not have to be perfect. The North American Electric Reliability Council (NERC) has established rules governing how well each control area must balance load and generation. Control Performance Standards 1 and 2 (CPS1 & CPS2) establish statistical limits on how well each control area must balance minute-to-minute fluctuations. Inadvertent interchange accounts track longer term differences. In all cases the total system remains in balance (otherwise blackouts occur). When one control area fails to balance its load with its generation, generation in another control area provides the balance.

The balancing of aggregate load with aggregate generation is accomplished through several services that are distinguished by the time frame over which they operate. As discussed above, regulation and load following (which, in competitive spot markets such as in California, is provided by the intra-hour workings of the real-time energy market) are the two services required to continuously balance generation and load under normal conditions[vi]. There is no hard-and-fast rule to define the temporal boundary between regulation and load following. In the PJM region, New York, New England, and Ontario, load following is defined as the 5 minute ramping capability of a generator. In Texas it is a 15 minute service, and in Alberta and California it is a 10 minute service.

Interestingly, control area operators do not need to specifically procure load following; it is obtained from the short-term energy market with generators responding to real-time energy prices. In the CaISO control area, this is known as the supplemental energy market. Regulation, however, requires faster response than can be obtained from units responding to market signals alone. Instead, generators (and potentially storage and/or responsive load) offer capacity that can be controlled by the system operator's AGC system to balance the power system.

Control areas are not able and not required to perfectly match generation and load. CPS1 measures the relationship between the control area's area control error (ACE) and the interconnection frequency on a 1 minute average basis. CPS1 values can be either "good" or "bad." When frequency is above its reference value, undergeneration benefits the interconnection by lowering frequency and leads to a good CPS1 value. Overgeneration at such times, however, would further increase frequency and lead to a bad CPS1 value. CPS1, although recorded every minute, is evaluated and reported on an annual basis. NERC sets minimum CPS1 requirements that each control area must exceed each year.

CPS2, a monthly performance standard, sets control-area-specific limits on the maximum average ACE for every 10 minute period. Control areas are permitted to exceed the CPS2 limit no more than 10% of the time. This 90% requirement means that a control area can have no more than 14.4 CPS2 violations per day, on average, during any month.

Methodology Description

Integration of large amounts of renewable generators could potentially increase errors between scheduled and actual generation. Increases in scheduling error combined with the existing error in load forecasting could change the composition or size of the "generator stack" which responds to load following needs. If such a distortion of the generator stack occurred it could shift the market to marginal generators, whose costs are higher. That could increase the price of energy across the market and thus create implicit costs which were imposed on the entire system by the renewable generators.

The analysis focused on the potential impacts to the generator stack caused by scheduling error. The methodology looks at the impact of renewable generators on the total system scheduling error. If renewable generators create systematic errors that significantly increase the need for generation resources, then they could have a material effect on the composition of the generator stack or the ex-post price for energy.

The analysis methodology first determines system forecasting and scheduling errors for a benchmark case without renewable generators. CaISO prepares hour-ahead forecasts of its generation requirements, which represent its best estimate of actual system load. The scheduling coordinators provide schedules for generation which are designed to economically meet the forecasted needs. The scheduling coordinators typically schedule significantly less generation than is needed during peak demand periods and rely upon the hour ahead market to provide the balance. The difference between the forecasted load and the scheduled load is defined as the scheduling bias. Forecast and scheduling errors in the benchmark case provide an indication of the variability inherent in operating the utility grid and are important because they define the normal range of errors without renewable generation impacts.

The scheduling errors for each renewable generator under study are then calculated. The difference between the actual and forecasted load is the load forecasting error. Worst case scheduling was used to estimate the impacts of the renewable generators. Bids for the hour ahead market are due 150 minutes prior to each market cycle. The scheduled output for the hour ahead market was defined by a simple persistence model, assuming that output 150 minutes in the future would be equal to output at the present time. For solar generators it was assumed that scheduled output was equal to what it had been on the previous day at the same time period.

The total system error including the renewable resources was calculated by combining the system forecast error (without renewables) with the additional scheduling error produced by the renewable resources. The forecasting error including renewable generators was then compared against the benchmark case and reviewed to identify significant differences. The goal of this analysis was to determine if the renewable resources significantly changed the total system error, thereby potentially modifying the generator bid stack.

Multi-Year Analysis Results and Discussion

The load forecasts prepared by CaISO provide the best estimate of the upcoming system load conditions. Figure 5 presents a graphical comparison of the hour ahead forecast load and the actual load for an example period of several days. Since it is not possible to perfectly predict the load in the hour ahead time frame, there will always be some forecast error.

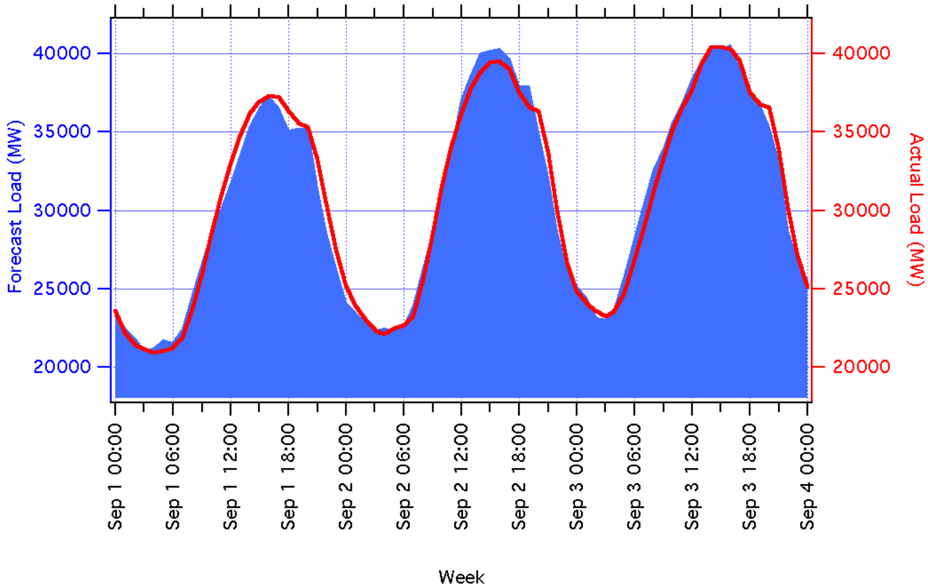

Figure 5. Forecast and actual load over a three day sample period.

The load schedule is created by the scheduling coordinators based on forecast information from CaISO and conditions in the energy markets. The hour ahead schedule as compared to the actual load is presented in Figure 6 for several example days in September. During peak hours the scheduled load is typically well below the actual load with the difference made up by the hour-ahead market. This indicates that the hour ahead market can be relied upon for large amounts of power to meet short term needs.

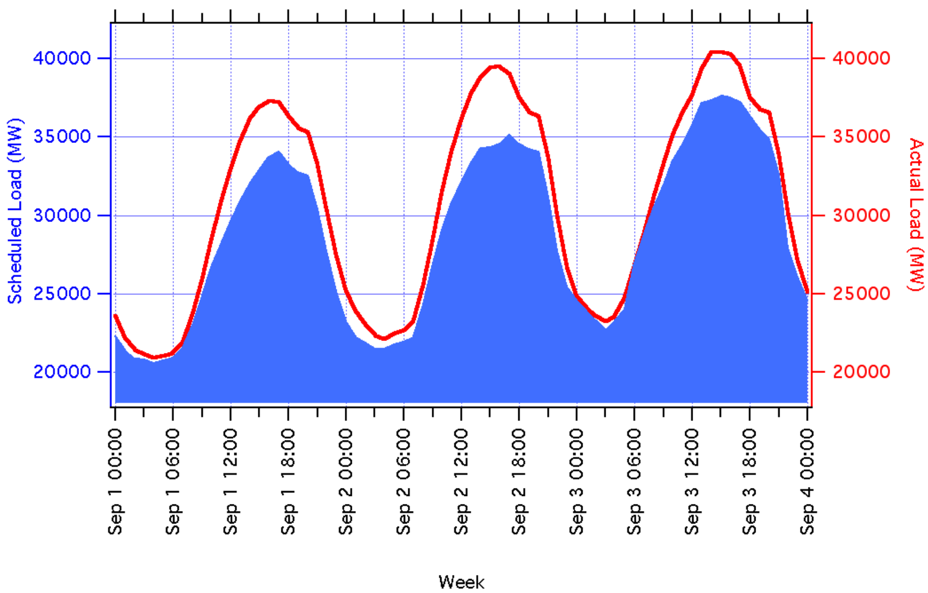

Figure 6. Scheduled and actual load over a three day sample period.

The difference between the scheduled load and the forecast load is the scheduling bias. It is typically negative (scheduled generation is less than forecast load) and, interestingly, reaches the largest negative values during peak summer hours when the power system is typically under the most stress. The scheduled load provided by the scheduling coordinators is often thousands of megawatts less than the forecast load provided by CaISO. Over the three year analysis period, the scheduled generation was as much as 5832 MW less than forecast load during peak hours. The average minima and maxima of the scheduling bias during peak hours are shown in **Error! Reference source not found.** over the three year analysis period. The large negative bias of the hour-ahead schedules provides an indication of the amount of generation assets available in the short term energy market. The data implies that the scheduling coordinators are comfortable with the depth of the generator stack; they can call up several thousand megawatts of generation whenever it might be needed. The scheduling bias was used as a proxy for estimating the depth of the generator stack. It was used for comparison purposes in determining the significance of renewable impacts on the system error.

The hour-ahead schedules for each renewable generation resource were developed using a simple persistence model. This model provides a schedule of renewable output for the hour ahead market and is a conservative (worst-case) approach. Use of true forecasting models will reduce scheduling error and reduce the significance of renewable impacts from those calculated here. Figure 7 presents an example of actual output and scheduled output for a wind generator using the simple persistence model to calculate the schedule. The resource scheduling error was calculated as the difference between the resource's scheduled generation and its load following component of generation; with the hourly data used in this analysis, the hourly generation values were used directly as the value of the resource's load following component. The forecasting error including the scheduling error was then calculated by adding the resource scheduling error to the load forecasting error.

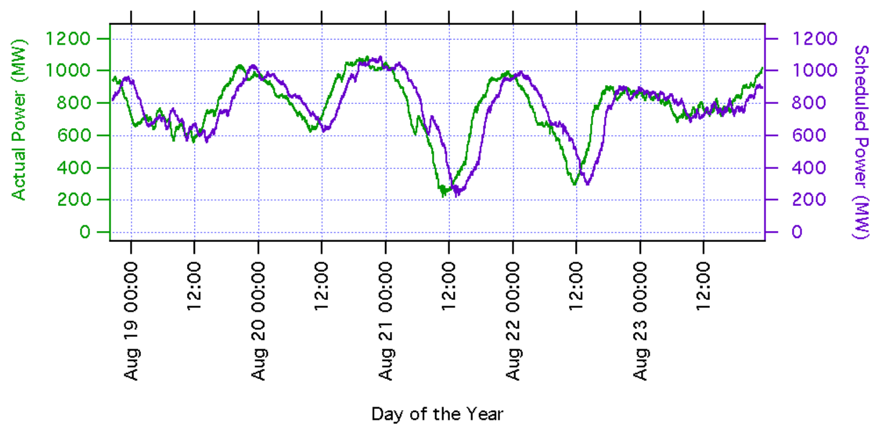

Figure 7. Actual and scheduled wind generation over a three day sample period. A simple persistence model was used to produce the schedule.

We compared the average minimum and maximum forecasting error during peak hours (noon to 6 p.m.) as a means of evaluating the significance of the renewable generator impacts. The results for the three analysis years are presented in Table 12. Negative values indicate that incremental energy purchases were required to compensate for undergeneration or unexpected load. Positive values indicate over-generation or lower demand than expected, requiring generators in the short term energy market to decrement their output. The minimum forecasting error was changed by no more than two percentage points by any of the renewable resources with slight improvements in some cases. The impact on the maximum forecasting error was similarly small. This indicates that at current penetration levels, the scheduling error of the renewables do not have a significant effect on the total energy requirements from the short term market. The minimum scheduling bias reduced over the years but remained more than 200% greater than the load forecast error. This implies ample depth in the generator stack to handle incremental energy requirements. The analysis concluded that the impact on unit commitment was too small to measure.

Table 12. Results of multi-year analysis of forecast and scheduling errors during peak hours.

	2002				2003				2004			
	AVERAGE MINIMUM		AVERAGE MAXIMUM		AVERAGE MINIMUM		AVERAGE MAXIMUM		AVERAGE MINIMUM		AVERAGE MAXIMUM	
ERROR	MW	Compared to load forecast error alone (%)	MW	Compared to load forecast error alone (%)	MW	Compared to load forecast error alone (%)	MW	Compared to load forecast error alone (%)	MW	Compared to load forecast error alone (%)	MW	Compared to load forecast error alone (%)
Load forecast alone	-1945	100%	2112	100%	-1600	100%	2151	100%	-1439	100%	1529	100%
Load scheduling alone	-4747	244%	1302	62%	-4021	251%	2158	100%	-3700	257%	1776	116%
Scheduling bias	-5337	274%	1708	81%	-3336	208%	1534	71%	-3016	210%	1634	107%
Combined load forecast and renewable resource scheduling error												
Biomass	-1944	100%	2115	100%	-1603	100%	2157	100%	-1432	100%	1536	100%
Geothermal	-1947	100%	2112	100%	-1599	100%	2149	100%	-1442	100%	1529	100%
Solar	-1897	98%	2055	97%	-1631	102%	2153	100%	-1467	102%	1541	101%
Wind (Northern Cal)	-1946	100%	2148	102%	-1591	99%	2203	102%	-1419	99%	1554	102%
Wind (San Gorgonio)	-1930	99%	2142	101%	-1581	99%	2163	101%	-1443	100%	1545	101%
Wind (Tehachapi)	-1931	99%	2177	103%	-1569	98%	2181	101%	-1435	100%	1544	101%

Discussion of the Ramping Capability Analysis results

It is possible to calculate a lower bound to the ramping capability within a given control area using public databases. In our experience some significant capabilities could not be estimated and more ramping capability exists than we were able to measure.[viii]

It appears that there is a very large amount of ramping capability in the CaISO control area during most hours of the 2002 analysis year we studied. This ramping capability is a natural result of the resource mix that has developed. Because each increase or decrease of renewable generation does not need to be matched one-for-one by another generator, the ability to absorb moderate or even large quantities of wind, solar, and other renewables appears significant for most of the year.

The CaISO control area appears to have significant ramping resources available from thermal generation that is partially loaded and physically able to respond. CaISO, like most ISOs, operates energy markets that clear several times an hour, providing access to the ramping capabilities of the generators active in the energy markets. Control areas that do not have access to fluid intra-hour markets still have the physical capabilities of the generators but may not have *access* to that capability simply based on the hourly market structure. This lack of access denies the generators the ability to position themselves (ramp) to sell as much energy as customers want, forces the control area operator to use additional regulating resources instead, and forces consumers to pay for the inefficiency. There may be significant opportunities for neighboring control areas to assist each other in the load following time frame as well. This is partly a natural consequence of the ability of larger control areas to better manage variability, whether caused by load, wind, or a combination with other resources. It is also a consequence of additional capability being inherently available from a larger pool of generators.

Assessing the ramping capability of a control area with public data presents some challenges. Because some data are unreported, and because of the shortcomings of our method, it is not possible to obtain an accurate measure. However, having said that, we think that this type of analysis can be useful in several ways. The estimates provided by this approach provide a lower bound on the load following capability in a control area. The approach is transparent, which makes it possible to more easily understand how the more complex methods embodied in production simulation models work. The approach could easily be extended to include data from non-CEMS-reporting resources. For entities that have access to such data, a more detailed analysis would be possible, and would provide a better estimate of the load following capability of the control area.

Data Issues

A variety of data issues were encountered in the various datasets used in the analysis. They are discussed below along with the methods used to address them.

Confidentiality

Although the need to preserve the confidentiality of much of the study data is recognized, data confidentiality significantly impeded the study at several occasions. Establishing the initial data nondisclosure agreement (NDA) with CaISO was a very lengthy process. The experience garnered from the completion of this first NDA was valuable later in the study, as new study participants were able to receive draft NDAs from CaISO quickly. Some other NDA processes were not as successful. In particular, SCE and NREL were unable to reach a confidentiality agreement even after numerous exchanges between their lawyers. Consequently, another analyst had to be trained to perform the capacity credit analysis, delaying the progress of the study.

Even with NDAs in place, the data released was aggregated because of concerns about the proprietary nature of power generation data from individual plants. Data aggregation aggravated data issues in the CaISO one-year and multi-year datasets, as discussed below. Later in the study, CaISO made a notable effort to allow the study analysts to view non-aggregated data while on-site at the CaISO offices; again, this is discussed further below.

Manageability

The sheer size of the data is a problem, particularly with one minute data as in the CaISO one-year and multi-year datasets. To assemble the renewable aggregates, CaISO had to extract more than eighty pieces of raw data, each with 525,600 values per year. Even with automated retrieval scripts, extensive computer time was required to query such a large volume, especially in the case of the three year dataset. Because the disk space requirement for storing all of the individual data items was considered to be too great, CaISO calculated aggregated values as the individual data items were being retrieved; only the aggregated value was stored and individual data values were immediately discarded. The lack of ready availability of non-aggregated data later hindered the data review process.

Performing the data review and error checks for so much data was also a time intensive process. Because of the difficulties introduced by aggregation, the effectiveness of automated data checks was limited and all of the CaISO one minute data required manually review. The errors discovered in the one-year and multi-year datasets revealed an underlying problem. Because much of CaISO's data is stored automatically and is never used for operations or in any other way, it does not undergo any inspection except for generic automated tests by the PI system. Much of the data is therefore recorded without any verification of the quality of the data or the actual recording process.

Lossy Compression

CaISO's PI system records over 180,000 pieces of data, some sampled many times a minute. To store so much data, a lossy compression scheme is used. Lossless compression uses algorithms that reduce the size of data while maintaining complete fidelity; when the data is uncompressed, it is exactly identical to what it was before compression was applied. Lossy compression sacrifices some accuracy for large improvements in size reduction; when the data is uncompressed, it is not exactly identical to what it was originally, but the changes should be negligible. The PI system uses the "Swinging Door" algorithm, a lossy scheme with configurable settings that trade off data fidelity and size. Ideally, information removed by compression is insignificant. However, the regulation analysis tracks even small fluctuations over short time periods. Data compressed without consideration for this type of calculation may affect the analysis when regulation impacts are small. Inspection of the data and regulation results suggests that the effects of compression might be significant only at impact levels when the regulation cost is negligible anyway.

Recommendations

The Phase III report made several recommendations about the implementation of integration cost analysis. Based on experiences from the multi-year analysis, the following additional recommendations pertaining to data reporting/collection and an Integration Cost Analyst (ICA) are proposed.

Data Reporting and Collection

The majority of time and effort required for the multi-year analysis was dedicated to data collection and processing. The actual calculations and review of the results were relatively straightforward. Specific recommendations are therefore made for the handling of data for future integration cost analysis.

In Phase III of the study, it was proposed that data collection should be performed by an Integration Cost Analyst, a CEC or CPUC staff tasked with performing and reporting on regular integration cost analysis. Given the complex data quality issues described in Section 0 and the need for similar data in other recent and current studies such as the CEC's Strategic Value Analysis and Intermittency Analysis Project, it is now recommended that data handling and integration cost analysis be separated into two distinct tasks. A data handling entity would be responsible for collecting, reviewing, storing, and providing data for integration cost analysis and, possibly, associated data for other studies. In Phase III, it was assumed that data collection and processing was essentially an accounting function which would be highly automated. While this eventually may become true, given the data issues described in this report, data handling is more appropriately an engineering task. The data handling entity would have to meet the following requirements and perform the following duties:

- Satisfy confidentiality requirements of CaISO, IOUs, and other sources to access data.

- Provide a database that securely stores data and that can be easily queried for both manual and automated data input and retrieval.

- Coordinate with CaISO, IOUs, and other sources to receive data on a frequent, regular basis; a one month basis is recommended. Jointly develop a reporting standard with the data sources for incoming data and, as necessary, tools to process various data types and formats. Also, jointly develop an automated reporting system so that data is transferred from the sources to the data handling entity automatically. Update data requests as necessary as new generators come online and other changes occur.

- Review and verify the quality of incoming data and flag and/or correct bad data.

- Coordinate with CaISO, IOUs, and other sources as necessary to ensure that the quality of data they are collecting and recording is sufficient for the intended analyses. As ongoing integration cost calculation is presumed for the future, this process should begin immediately.

- Coordinate with the Integration Cost Analyst to ensure that the required data is collected with sufficient quality and provided to the ICA on a frequent, regular basis; again, a one month basis is recommended. Jointly develop a reporting standard with the Integration Cost Analyst and an automated system for transfer of data from the data handling entity to the ICA.

One of the key aspects of the proposed data handling process is that the assurance of data quality is a shared responsibility between the data sources (CaISO, IOUs, etc.), the data handling entity, and the Integration Cost Analyst. The task otherwise becomes disproportionately difficult to manage and complete.

It is also important to collect and review data on a frequent and regular basis. Many of the difficulties encountered with the processing of the datasets for the multi-year analysis were the result of working with such a large, lumped amount of data at once. As originally proposed in Phase III, it is recommended that data be documented monthly in arrears for the previous month. Processing data on a frequent basis not only keeps the task more manageable, but allows errors and issues to be identified and corrected before they propagate into a larger amount of data over an extended period. Automated data reporting would simplify the collection process, but the data review will always include some manual inspection.

Integration Cost Analyst

An Integration Cost Analyst (ICA) was introduced in Phase III and is recommended again with some revisions to the original description of qualifications and responsibilities. The function of the ICA is to perform regular analysis and reporting of integration costs. It is proposed that the CEC or CPUC designate one or more staff to assume this role. Specifically, the ICA would have to meet the following requirements and perform the following duties:

- Satisfy confidentiality requirements of CaISO, IOUs, and other sources to access data.

- Coordinate with the data handling entity previously described and, as necessary, the various data sources to ensure that all required data is of sufficient quality and is received on a frequent, regular basis in a consistent format. Again, it is recommended that data be received on a monthly basis.

- Review incoming data as it is received to verify data quality.

- Annually perform integration cost analysis.
- Prepare annual reports documenting the results of the integration cost analysis.

Assuming the availability of good data, the calculations involved in integration cost analysis are relatively straightforward and can be highly automated. Once procedures are established and refined, it is estimated that the ICA will require approximately one to two days per month to perform data handling tasks and approximately two additional weeks each year to conduct the integration cost calculations, perform an analysis of the results, and generate a report.

Conclusions and Summary

The results from this analysis indicate that

- Wind capacity value of the existing fleet is in the mid-20's range as a percent of rated capacity. The 3-year modified ELCC values (excluding hydro and interchange) match peak period capacity factors quite well.
- Regulation impacts of wind are the same relative order of magnitude as the regulation impact of load. Costs are moderate, and vary somewhat from year to year and by resource location
- Load following impacts of wind in California appear to be very small, and are dwarfed by the magnitude of unscheduled generation
- Data issues are significant in a project of this size. Ongoing quality assessment and regular data sampling would significantly improve the data quality, and would result in a higher quality assessment of the impact of renewable generation on indirect costs

The methods developed for this project can be applied as the penetration of wind energy increases in the CaISO footprint. At higher penetrations, it is likely that the impact of wind in the load following and unit commitment time frames will become significant. We suggest ongoing analysis in an effort to capture these impacts as they occur.

References:

[i] Piwko, Osborn, Gramlich, Jordan, Hawkins, and Porter, "Wind Energy Delivery Issues", *IEEE Power and Energy*, Vol. 3, No. 6, Nov/Dec 2005.

[ii] U.S. Federal Energy Regulatory Commission, *Regional Transmission Organizations, Notice of Proposed Rulemaking*, Docket No. RM99-2-000, Washington, DC, May 13 1999.

[iii] U.S. Federal Energy Regulatory Commission, *Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities, Final Rule*,

Docket Nos. RM95-8-000 and RM94-7-001, Order No. 888, Washington, DC, April 24 1996.

[iv] Hirst, E. and B. Kirby, “Unbundling Generation and Transmission Services for Competitive Electricity Markets: Examining Ancillary Services”, NRRI 98-05, National Regulatory Research Institute, Columbus, OH, January 1998.

[v] North American Electric Reliability Council, *NERC Operating Manual*, Princeton, NJ, November 2002.

[vi] E. Hirst and B. Kirby, “Measuring Generator Performance in Providing the Regulation and Load-Following Ancillary Services”, ORNL TM/2000-383, Oak Ridge National Laboratory, Oak Ridge, TN, December 2000.

[vii] Electronic Wind Performance Reporting System, http://wprs.ucdavis.edu/, accessed in October 2005.

[viii] B. Kirby and M. Milligan, “A Method and Case Study for Estimating The Ramping Capability of a Control Area or Balancing Authority and Implications for Moderate or High Wind Penetration”, NREL CP-500-38153, National Renewable Energy Laboratory, Golden, CO, June 2005. http://www.nrel.gov/docs/fy05osti/38153.pdf, accessed May 17, 2006.