REVIEW OF SUPPLY ADEQUACY CRITERIA IN THE NORTHWEST

Prepared for

NWPCC AND BPA

August 2011

FINAL VERSION

PSR

Table of contents

1 INTRODUCTION

Ideally, power system supply should be both *economic* – i.e. provide electricity at the lowest possible cost to consumers – and *reliable* – i.e. there should be no interruptions in power supply. However, these objectives are conflicting: if supply reliability is improved by the construction of additional generation, there will be an increase in electricity tariffs.

As a consequence, one tries to determine the best *tradeoff* between generation capacity and supply reliability. This tradeoff is illustrated in the conceptual figure below.

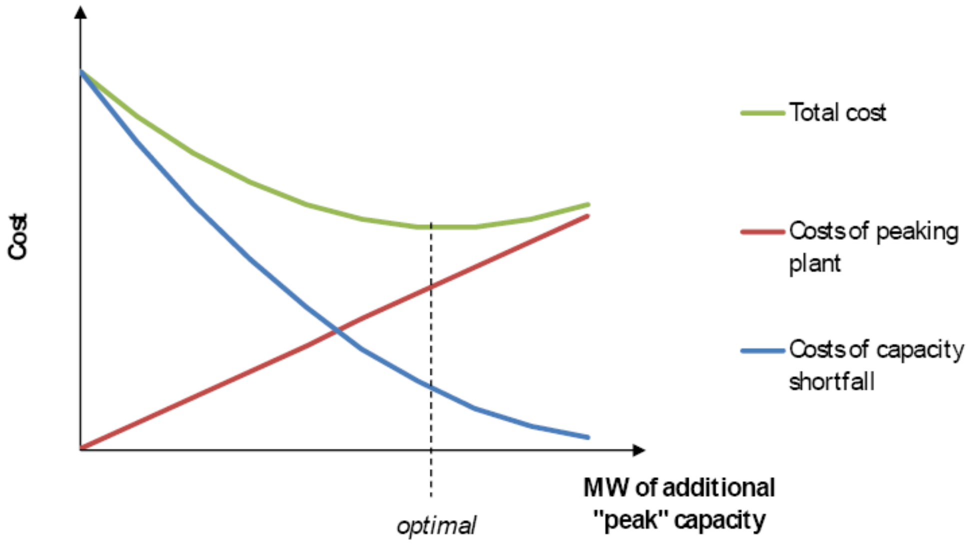

Figure 1.1 – Definition of optimal capacity margin

We see in the figure that investment cost increases with additional capacity, whereas the capacity shortfall costs decrease. Therefore, the "optimal" generation capacity would be the one that minimizes the sum of both costs.

In the past, when generation systems were planned in a centralized basis, the fact that supply reliability is a "*system*" attribute (i.e. which depends on the entire set of generation units and on the total load) was not an obstacle: the planning group would devise a trial expansion plan, calculate the resulting system reliability and, if the value was found to be inadequate under the "planning criterion", the expansion plan would be adjusted.

However, the situation became more complex when power systems were deregulated, because investors were only concerned with the physical and financial performance of their *individual* generating plants, and there was no mechanism to ensure that the resulting overall supply reliability would be adequate. Although it was argued at the time that "market signals" would ensure this reliability, many countries experienced underbuilding and underwent rationings during the 1990s.

As a consequence of these supply problems, many countries started to require an overall measure of system supply reliability, together with some mechanism to induce the entrance of new generation capacity (for example, contract auctions) if the supply shortfalls were higher than a given reference value (reliability standard).

1.1 Scope of this report

The objectives of the present study are:

1. Survey and then summarize the types of probabilistic power supply adequacy metrics that other energy-limited (or hydro dominated) utilities are using.
2. Critique the Northwest's current method of assessing power supply adequacy and, if needed, suggest ways to improve it or suggest a more useful alternative.
3. Critique the use of deterministic metrics and minimum thresholds (derived from probabilistic analyses) as a screening tool for adequacy assessments.
4. Critique the Bonneville Power Administration's proposed adequacy metric(s) and corresponding minimum threshold(s) for the federal power system.
5. Survey and then summarize how other utilities incorporate power supply adequacy standards into their long-term resource planning methods.
6. Aid in the development of a method to improve the way that the Northwest's adequacy standard is incorporated into resource planning processes and analytical models. This method must ensure that both energy and capacity adequacy concerns are addressed.

2 METHODOLOGY OVERVIEW

2.1 Supply security × supply adequacy

First of all, it is necessary to distinguish between two components of supply reliability: (i) *security* and (ii) *adequacy*. As defined by NERC, security refers to the ability of the system to withstand sudden disturbances, such as an electric short circuit or unanticipated loss of system elements. In turn, adequacy refers to the ability of the power system to meet the aggregate power and energy requirements of the consumption at any time. In this work, we will refer only to the "adequacy" component of supply reliability.

2.2 Energy- and peak-constrained systems

2.2.1 Peak-constrained systems

Historically, systems could be classified as thermal-dominated or hydro-dominated. For a given generation capacity, the main issue in thermal-dominated systems is *peak supply* shortfalls due to generation outages. This means that supply adequacy evaluation in such systems has to be carried out in an hourly basis, taking into account both equipment outages and temperature-driven load fluctuations. However, as illustrated in the figure below, this evaluation can be carried out *separately* for each hour of the each day, which allows the use of analytical techniques (convolution in load duration curves) in the probabilistic simulation of plant outages. Examples of "peak constrained" systems include the UK and the East Coast of the US.

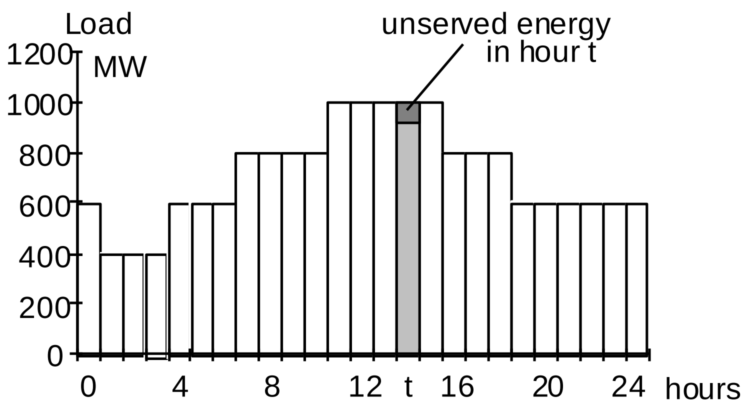

Figure 2.1 Supply adequacy evaluation in thermal systems

2.2.2 Energy-constrained systems

In turn, the main issue in systems with significant hydroelectric generation is *energy supply* shortfalls, usually due to severe droughts (i.e. lack of "fuel" for the hydroelectric system). Examples of "energy-constrained" systems include Brazil, Colombia, Norway and the province of Quebec. Because hourly load variations or equipment outages do not usually affect the risk of energy shortages, adequacy evaluations for energy-constrained systems are carried out through chronological simulations of system operation in weekly or monthly steps, for a large number of hydrological conditions across the entire year.

In some systems such as Norway and Quebec, historical inflow records (around 80 years) are used in the simulations. In other systems such as Brazil and Colombia, "synthetic" inflow sequences produced by a stochastic streamflow model are used (typically 2 thousand sequences). The figure below illustrates the chronological simulation of a hydro-dominated system.

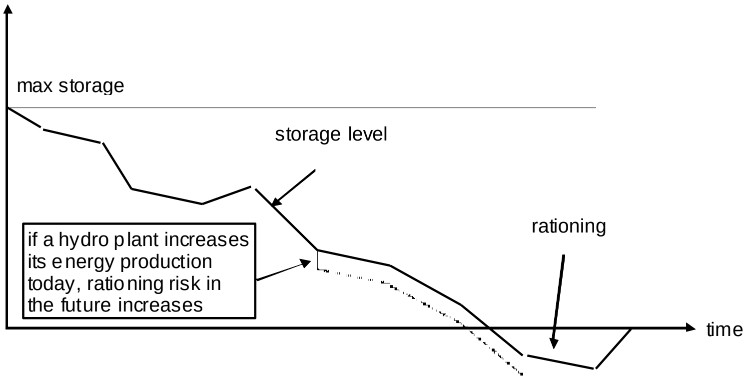

Figure 2.2 Supply adequacy evaluation in hydro-dominated systems

2.2.3 Energy- and peak-constrained systems

The increasing penetration of non-dispatchable and fluctuating generation sources, in particular wind power, combined with limited reservoir storage, has made an increasing number of systems both energy- and peak-constrained. Examples of such systems are the Pacific Northwest, Turkey and Guatemala. The supply adequacy evaluation in such systems is more complex, because it requires both chronological simulations for a large number of inflow sequences across the year and the detailed modeling of hourly events: equipment outages, temperature-driven load fluctuations and the variable production of non-dispatchable generation.

2.3 Supply adequacy metrics

The cost of a supply shortfall usually depends on its duration, severity and predictability. For example, peak supply interruptions are usually caused by a combination of generation outages and high load. They are unpredictable and last between minutes and several hours. In turn, energy rationings are usually caused by a combination of extreme droughts and the lack of sufficient non-hydro generation and/or inter-regional transmission capacity. An energy supply shortfall is usually more predictable, but lasts for a longer time (weeks to months, depending on the reservoir storage capacity and hydro share). These different types of supply shortfall are usually solved by different types of generation reinforcements. For example, a diesel generator may be an economic option as a peaking reserve, because of its low investment cost and the fact that its high variable operating cost is compensated by the limited hours in operation, whereas a gas-fired combined cycle plant, despite its higher investment cost, can be more competitive to solve energy shortfalls (base loaded).

As a consequence, separate peak and energy supply adequacy metrics have historically been used. Also, each metric can be broadly classified as "risk-oriented" or "severity-oriented".

2.3.1 Metrics for peak supply adequacy

2.3.1.1 "Risk" metrics

In the case of peak supply adequacy, the "risk" metrics are usually the expected number of hours in which there are supply interruptions in a given year ("loss of load expectation", LOLE, typically 2 or 3 hours per year). In some countries, this expected number of hours is divided by the total number of hours in the year, and referred to as a "loss of load probability", LOLP.

2.3.1.2 "Severity" metrics

One well known limitation of "risk" metrics is that they provide no information on the "depth", or "severity" (amount of energy curtailed) of supply interruptions. For this reason, some systems use "severity" metrics for peak supply adequacy such as the *expected unserved energy* (EUE, or EUSE), calculated as the total MWh interrupted along the year divided by the total MWh energy load.

2.3.2 Metrics for energy supply adequacy

In the case of energy supply adequacy, there are also "risk" and "severity" adequacy metrics.

2.3.2.1 "Risk" metrics

The "energy risk" metric is usually related to the probability of having any energy shortage in that year. This type of metric is used, for example, in the Brazilian system, where 95% of the simulated scenarios (or "games", in the Pacific Northwest jargon) should have no shortages across the whole year.

This energy risk metric (probability of having any shortage along a simulated year) is sometimes referred to as LOLP, the same name as the unrelated peak risk metric (number of hours interrupted along the year). This is particularly confusing when supply adequacy standards of energy and peak-constrained systems are compared.

2.3.2.2 "Severity" metrics

In turn, the "energy severity" metric is usually the total MWh curtailed along the year, or the ratio between these curtailed MWh and the total load (i.e. the fraction of energy load that is curtailed in the scenarios with supply shortfall).

2.3.3 Metrics for energy- and peak-constrained systems

At first sight, systems that are both energy- and peak-constrained should use both energy- and peak- metrics. In the reviewer's opinion, however, this use is not straightforward. The reason is that there may be "energy" and "peak" production shortfalls which, *separately*, would not cause supply problems but, when combined, may lead to load curtailment. One example is a reduced inflow that leaves the reservoir storage – and thus the production coefficient (MWh/outflow volume) – lower than usual. When combined with, for example, an unusually low wind situation and/or some generation outages, there may not be enough water for the hydro plants to meet the peak demand. This type of situation is better described as a “composite” energy and peak shortage, rather than a “pure” energy or peak problem.

As will be described later in this document, a possible “unified” energy & peak metric is the system *load supplying capability* (LSC), which is the maximum load profile that can be supplied for each simulated scenario (or game) with no shortages.(The LSC calculation, advantages and limitations will be discussed later.)

2.4 Is there a metric that combines the best features of risk and severity?

As mentioned previously, the risk and severity metrics “see” *complementary* aspects of the overall supply adequacy “picture”. One possibility is, of course, to use both metrics in the supply adequacy evaluation. However, this means that it becomes necessary to establish two separate risk and severity adequacy standards. As will be seen, these standards have little economic justification in most, if not all, countries. As a consequence, it is likely that the standards will not be very consistent, i.e. one of them will be more “strict” than the other.

Additionally, risk metrics such as LOLP and LOLE are not *coherent* in a mathematical sense, i.e. they are not convex functions of generation capacity. As a consequence, they are difficult to use in optimization-based expansion planning models (minimize investment plus expected operation costs, subject to a supply adequacy constraint). This coherence requirement has become a core issue in the recent stochastic optimization literature.

2.4.1 Conditioned Value at Risk - CVaR

In the reviewer’s opinion, a new adequacy metric that may be of interest to the Pacific Northwest is the *Conditional Value of Risk*, CVaR, defined as the expected unserved energy over the α% worst scenarios (i.e. those with the highest load curtailment). The parameter α is defined by the planner (for example, α=5%). The CVaR measure is being extensively used in the stochastic portfolio optimization area because of its ease of interpretation and attractive mathematical properties (coherence and representation as linear programming constraints in optimization models).

2.4.1.1 Very simple CVaR example

The calculation of LOLP, EUE and CVaR will be illustrated for the very simple system shown in the table below.

Table 2.1 – Simple example for the calculation of adequacy indices

scenario	UE
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	1
15	5
16	8
17	20
18	30
19	50
20	70

The table shows the hypothetical results of hourly simulations of system operation across one year for 20 different scenarios comprising, for example, inflows, temperature (which affects load, wind power production and others. The result of each simulation is the total unserved energy (in GWh or equivalent). We observe in the table that 7 out of the 20 scenarios presented some curtailment. Therefore, the rationing risk is 7/20 = 35%. In turn, the expected unserved energy is calculated as (1+5+8+20+30+50+70)/20 = 9.2 GWh. Finally, the CVaR for a α value of 20% - in this case, the 4 worst results – is calculated as (20+30+50+70)/4 = 42.5 GWh.

This simple example illustrates the main characteristic of each metric:

(i) The rationing risk indicates that only 65% of the simulations will not result in load curtailment; however, it does not “see” that the energy curtailed in some scenarios is much smaller (i.e. less severe) than in others (1 GWh in the least severe versus 70 GWh in the most severe).

(ii) The expected unserved energy does take into account the amount curtailed, but the mean value is “diluted” by the scenarios with load curtailment[1].

(iii) CVaR “captures” the mean severity of the worst 20% scenarios, but does not take into account the smaller load curtailments in the other scenarios; in this sense, CVaR can be interpreted as a type of “fuzzy worst case” adequacy metric.

Later in this document, we will discuss some important questions for the application of CVaR to real-life systems, such as, what α value to use, and others.

[1] It is interesting to observe that the most obvious way of “undiluting” the expected unserved energy, which would be to calculate the expected value energy over only the scenarios in which there was some unserved energy does not result in a convex function of generation capacity.

3 SUPPLY ADEQUACY STANDARDS

Once the adequacy metric is defined, the next step is to define an adequacy *standard*, i.e. a minimum "threshold" for a system supply to be considered adequate.

3.1 International experience

The table below[2] shows the adequacy standards used in some countries in Europe, Oceania and North America.

Table 2.2 Examples of adequacy standards

Country	LOLE	EUSE	Basis for standard	Scope
Australia (NEM)		.002%	Defined by Reliability Panel in 1998 at market start and confirmed by AEMC in 2007 when reviewed	Generation and bulk transmission. Single credible contingencies of generation or bulk transmission
Australia (WA)	MW margin (largest unit) or .002% EUSE		Unclear	Generation and bulk transmission. Prudent peak demand
Ireland	8 hours per annum		"Appropriate and acceptable" on historic basis	Ability of grid connected generation to meet GIP demand. Transmission limits, but not risks
UK	No formal standard		Monitor and contemplate economic trade-off, but" far from straightforward in practice"	Nominal capacity and "average cold spill" peak demand.
France	3 hours per annum		Government decision	Grid connected supply. Transmission outages and imports
PJM	2.4 hours per annum (one day in 10 years)		Formulation of NERC in 1965	Generation capacity. No transmission risk. LOL = invoking emergency operations procedures beyond demand resources and IL for reliability
Ontario	2.4 hours per annum		Follows NPCC standard (Northeast Power Coordinating Council)	Ad hoc transmission adjustments on supply

[2] New Zealand Electricity Commission, *Development of a Capacity Adequacy Standard*, October 2008.

As mentioned in the NZ Commission document, it is clear from this sample that different countries adopt different approaches to developing and specifying a standard. One common feature, however, is that none of them seems to have an explicit economic justification. Another common feature is that transmission and distribution uncertainties are typically excluded, and supply / demand variation included.

Finally, the level of LOLE standards seems to have a fairly small range, from 2.4 to 8 hours per year (the LOLE corresponding to the .002% EUSE standard employed in Australia was recently estimated as 3.5 hours per annum). However, it is important to observe that the LOLE calculation procedure varies from country to country and, as mentioned, may refer to very different attributes in peak or energy constrained countries.

3.2 "Translation" of adequacy standards into required reserve margins

In many countries, there is a perception that the best way to communicate the supply adequacy concepts is through a supply × demand balance. In a simplistic way, if the *generation reserve margin*, given by the difference between generation capacity and peak demand, exceeds a given percentage (for example, 20%), the system supply is considered adequate. Otherwise, the MW difference between the (inadequate) reserve margin and the target value provides an indication of how much capacity should be built in order to restore the adequacy standard.

One concern with the reserve margin criterion is that it does not take into account the *type* of generation being constructed, i.e. the addition of 1000 MW of wind power may have a substantially different impact on system reliability than the addition of 1000 MW of combined cycle natural gas plants. In the past, this was not particularly important because of the limited number of generation technologies. However, the appearance of renewable sources such as wind power makes it more important to differentiate between technologies.

3.3 Firm capacity

A possible alternative is to carry out the supply × demand balance in terms of *firm capacity*. Returning to the example of Table 2.1, suppose that the adequacy standard is a risk of 20%, i.e. only 4 out of the 20 scenarios can have a load curtailment. Given that the actual risk is 35% (7 scenarios) we can conclude that the current set of generators is unable to supply the projected load with the desired supply adequacy. The question is then: how much additional capacity is required to meet the adequacy standard? The answer would be to build a firm capacity that is able to produce 8 GWh, which is the highest unserved energy in the three scenarios (14, 15 and 16) that would have to be "fixed".

The advantage of this procedure is that it is applied directly to the probabilistic simulation results, and translates into a supply × demand balance that looks similar – while being more sophisticated – than a reserve margin requirement.

3.4 Load supplying capability

One limitation of the firm capacity balance is that it only provides useful information when the system is under-installed, but *not* when the system is over-installed, i.e., when the actual adequacy indices are better than the standards. In this case, one would like to know how much the load could *increase* while still meeting the adequacy standards. This *load supplying capability* (LSC) can be calculated by an iterative procedure in which: (i) the load is incrementally increased; (ii) the probabilistic simulations are repeated; and (iii) the adequacy indices are recalculated.

3.4.1 Example of LSC calculation - Brazil

The LSC × demand balance is at the core of Brazil's supply adequacy evaluation. The LSC, called "physical guarantee" in Brazil, is calculated as follows:

Repeat steps 1-5 below for each year of the planning horizon:

1. Define the *generation configuration* for that year as the set of all generators that already exist or will start operation in that year[3].

2. Define a *candidate* energy load, which in the case of Brazil is "flat" (the same along the whole year, used in Brazil). We will see in a later section that the LSC calculation can be generalized to represent demand seasonality (winter / summer) and daily variation (peak / off-peak).

3. Given the generation configuration and the candidate load, calculate the hydrothermal dispatch strategy (in the case of Brazil, a stochastic dynamic programming model is used for that) and simulate the system operation for a large number of scenarios (in the case of Brazil, the major uncertainty are the inflows. We will see in a later section that uncertainties on equipment availability and load fluctuations can be represented in a straightforward way). This simulation is carried out for a "steady state" situation, i.e. although the results refer to one year, we define a study horizon of, say, 5 years: the first two years are used to "erase" the effect of initial reservoir storage, and the last two serve as end-horizon "buffers", i.e. to prevent the reservoirs from emptying completely. All simulation statistics are then calculated for the remaining third year of the study.

4. Estimate the supply adequacy index, which in the case of Brazil is the *risk of rationing* (fraction of simulated scenarios which presented a supply shortfall in at least one hour across the year). If the calculated risk matches (within a given tolerance) the adequacy standard (5% for Brazil), the candidate energy load is the LSC, and the procedure is terminated.

5. If the calculated risk is *higher* than the adequacy standard (for example, 7%), the candidate energy load is *decreased* (for example, all values are multiplied by 0.98). If the cal-

[3] The generators which start operating in the first semester are anticipated to January; those which start operating in the second semester are postponed to January of the next year.

culated risk is smaller than the standard (for example, 4%), the candidate load is increased (e.g. all values multiplied by 1.1). We then return to step 3 for a new probabilistic simulation with the adjusted candidate load.

The figure below shows the balance of LSC and average yearly load for years 2010-2014.

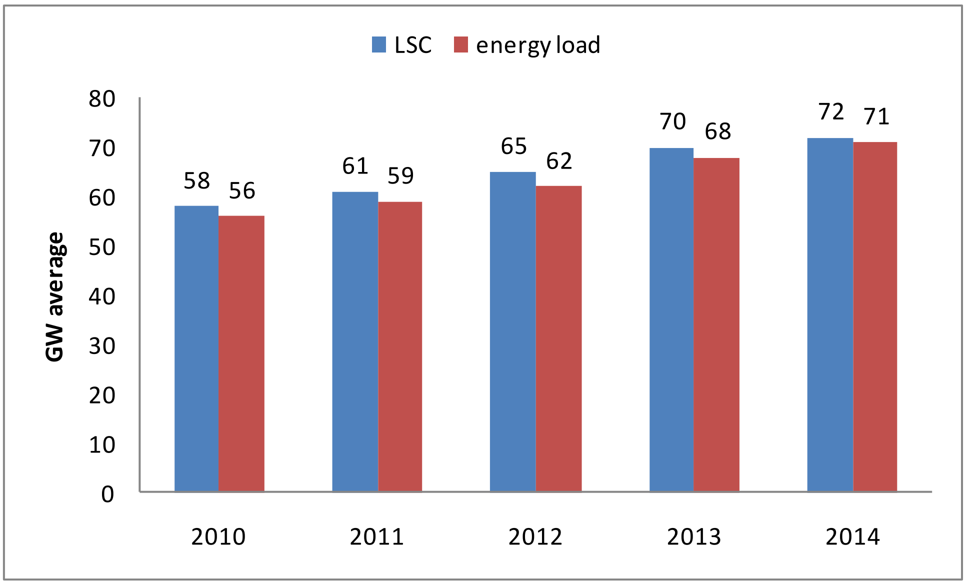

Figure 2.3 – LSC x load balance in Brazil

The red bar in the figure represents the energy demand in each year, in GW average. For example, the energy demand in 2010 is 56 GW average. In turn, the blue column represents the system LCS in each year, also in GW average. For example, the LSC for year 2012 is 65 GW average. The interpretation of the LSC × energy load balance is straightforward: if the LSC is smaller than the energy demand, the adequacy standards will not be met; the "firm capacity shortfall", i.e. the additional generation capacity that has to be built, is given by the difference between energy demand and LSC. Conversely, if the LSC is higher than the load, there is a surplus firm capacity given by (LSC – demand).

4 THE CURRENT NORTHWEST SUPPLY ADEQUACY METHOD

4.1 Overview of the current scheme

4.1.1 Supply adequacy metrics

In the reviewer's understanding, there are four adequacy metrics, given by the combination of energy and capacity shortfall risks in the summer and winter seasons. The values for these metrics are extracted from a large number of chronological simulation runs (the drivers for variation in each run are the historical inflows, equipment outages, temperature-driven load variation and wind power production). A given simulation run is tagged as a winter (summer) energy shortfall if the sum of curtailments over the winter (summer) months exceeds 28.8 GWh. In turn, it is tagged as a winter (summer) capacity shortfall if the curtailment in *any* hour of the same period exceeds 3000 MW.

The energy (capacity) summer (winter) shortfall risks (known as LOLPs) are given by the fraction of tagged runs over the total number of simulations. The adequacy standard in all cases is 5%, i.e. the supply will be considered adequate if none of the four LOLPs exceeds 5%.

4.1.2 Deterministic "translation" of supply adequacy

In the reviewer's understanding, the supply adequacy analysis is "translated" into the following deterministic requirements: (i) the system "available average annual generation" should not be smaller than the average annual load; and (ii) the peak capacity reserve margin should not be smaller than 23% (winter) and 24% (summer).

There is also a resource × load balance, in which the system available average annual generation is defined as the sum of four parcels, all in MW average. The first is the mean availability (in MW average) of non-hydro resources, taking into account maintenance, forced outage rates and other constraints (fuel supply, environmental etc.). The second refers to the mean availability of uncommitted IPP resources (market resources within the region), with full capability from October through May and a fraction of that in the remaining months. The third is the firm hydro generation, based on the worst drought in the historical record. Finally, the fourth parcel is said to be a *planning adjustment*, which is added to the previous parcels and represents the amount of "non-firm" resources to be counted on for planning purposes (a combination of better-than-critical hydro and out-of-region market supply). In the reviewer's understanding, this planning adjustment can be interpreted as the capacity that has to be acquired in order to ensure the supply adequacy.

4.2 Reviewer comments

4.2.1 Simulation procedure

In the reviewer's opinion, the current simulation procedure includes the modeling requirements for adequacy evaluation in a system with an important hydro component – therefore, energy-constrained – but that also may have peak capacity problems due to limited storage capacity and to the introduction of wind power: (a) hourly chronological simulation; and (b) a Monte Carlo simulation scheme, where the main uncertainties are hydrology, temperature-driven load fluctuations and wind power production. The simulation scheme also allows the representation of complex interdependences such as variation of both wind and load with temperature, and of the frequency and duration characteristics of generation outages.

One suggestion for improvement would be the use of stochastic inflow models, which would have the following potential benefits: (i) allow a wider range of scenarios, in particular ones with droughts which are more severe than the historical record; (ii) start modeling the effects of climate change in the Pacific Northwest inflows.

4.2.2 Energy and capacity shortfalls

The use of an energy adequacy metric related to the cumulative curtailment (GWh) for the simulation period and of a capacity metric related to the worst hourly curtailment (MW) in the same period also seem to be appropriate. However, it is not entirely clear why two separate measures for summer and winter are necessary, instead of the cumulative curtailed energy (and the worst curtailed hour) for the whole year. The reason is that, if supply adequacy is inadequate, it will be necessary to contract new capacity, which presumably would be available permanently. One possible explanation is that, instead of acquiring new generation capacity, reservoir operation could be adjusted to, for example, cover a winter shortfall, but is not clear whether this would be possible.

4.2.3 Management of small curtailments

Another observation concerns the "deductibles" of 28.8 GWh and 3000 MW. In the reviewer's understanding, this tries to capture the fact that those curtailments are not "real" because, if necessary, they could be eliminated by, for example, smaller local generation such as diesel plants used as back-up in shopping malls or industries. A similar scheme has been used in other countries, for example Brazil: only runs with cumulative curtailments smaller than 1% of the total (cumulative GWh) energy demand were tagged as "shortfall" years.

Note that these "deductibles" are only necessary when *risk* metrics are used, because a small curtailment counts as much as a large one. If metrics that take into account the depth of curtailments, such as EUE or CVaR, are used, the small curtailments would not affect the results.

4.2.4 Adequacy standard

The reviewer admits that the use of a 5% standard for all metrics has the same (lack of) economic justification as the standards applied in other countries. However, it is possible to carry out a "sanity check", which is to calculate the *implicit unit cost* of load curtailment associated to the 5% risk criterion. This can be done by running an optimal generation expansion model (minimization of investment cost plus expected value of operation cost and un-served energy cost) for different values of unit load curtailment cost until the same capacity as the risk-constrained plan was obtained. (In mathematical terms, it is as if we were calculating the Lagrange multipliers associated to the supply adequacy constraint.) The figure below illustrates this concept for a similar study carried out with the Colombian system. The horizontal axis shows different values for the cost of load curtailment. The curves show the total expansion plan cost (with increases as the load curtailment cost increases) and the expected energy not supplied (which decreases with the cost of load curtailment).

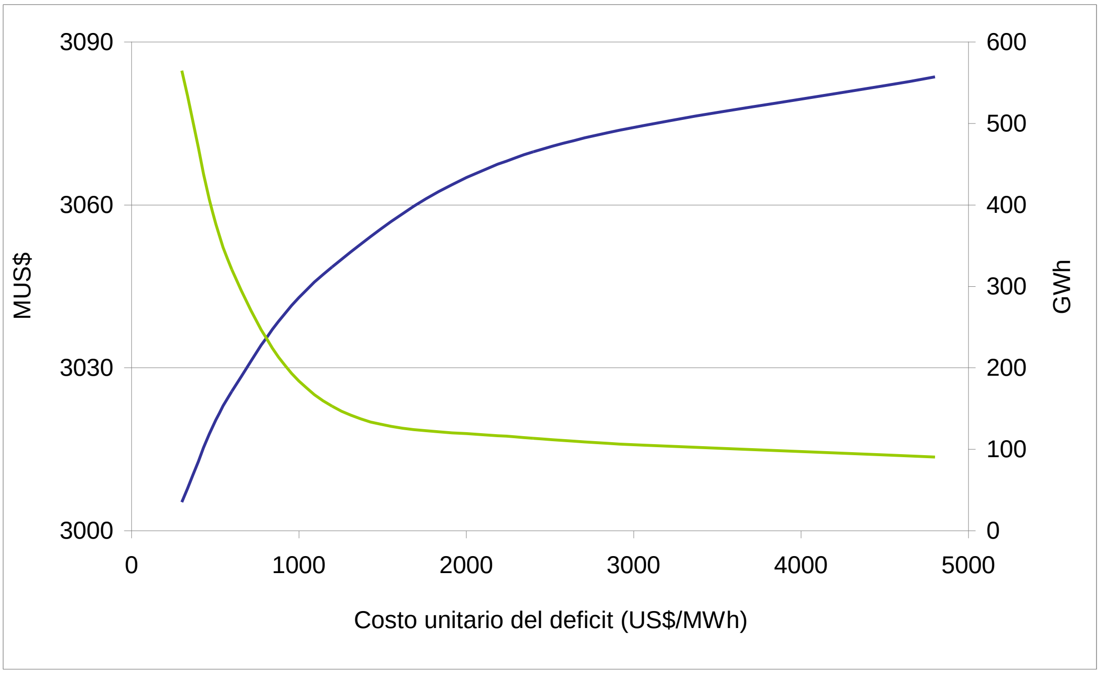

Figure 4.1 – Calculation of the implicit cost associated with a supply adequacy threshold

Although the “true” cost of load curtailment cannot be known, the implicit cost can be used to check if the “ballpark figure” is correct. For example, if the implicit cost is 2,000 US$/MWh, it probably means that the adequacy threshold is too “loose”, i.e. that the system is under-invested. Conversely, if the implicit cost is 200,000 US$/MWh, it probably means that the adequacy threshold is too “tight”, i.e. the system will be “over-invested”.

4.2.5 Use of deterministic metrics as screening tools for planning

As mentioned previously, the use of a generation reserve margin as a deterministic metric has the disadvantage of not recognizing that different types of plants with the same capacity may have different effects on supply adequacy. These differences will tend to become more significant as more wind power is built.

In the reviewer’s opinion, there is a more effective way of “translating” probabilistic criteria into deterministic ones. As discussed previously, the Load Supplying Capability (LSC) concept makes it is possible to retain the “flavor” of deterministic metrics such as the balance between energy resources and load while using a full probabilistic framework. Also as seen previously, the LSC can be used in both shortfall and surplus situations. Finally, the current resource x load balance procedure is not very different from the LSC.

5 THE PROPOSED BPA METRICS

In the reviewer's understanding, BPA proposes the use of a chronological Monte Carlo simulation scheme that represents a large number of uncertainties, such as hydrology, wind power fluctuations, load variation and others.

Two types of information will be extracted from these simulations: (i) the percent of total hours simulated that had load curtailment occurrences (referred to as LOLP by BPA); and (ii) analysis (frequency, duration, and magnitude) of infrequent but highly significant "tail" events.

In the reviewer's understanding, the adequacy criterion would be based on the LOLP measure; the additional "tail" information would provide deeper insights into the modes of supply failure. For example, BPA suggests the use of these measures: (i) Average, median, and percentiles for event duration and magnitude – determine the event spread; (ii) Event timing – heavy load hours vs. light load hours, seasonality; (iii) Event analysis: given that events occur, what is the worst that can be expected? Examination of the worst 5% of events by duration or magnitude; Average, median, and percentiles for tail event duration and magnitude – determine the spread; EUE for exclusive set of tail events – expected magnitude of a tail event.

In the reviewer's opinion, the use of a chronological Monte Carlo simulation is the correct way to capture the complex interplay and time/spatial interdependence of such phenomena as load, temperature, outages, inflows, wind production and others.

It is also very useful to have instruments for more detailed analysis of "tail" events, as they will provide very important insights into the system behavior.

The only area which, in the reviewer's opinion, there are opportunities for further developments is in the proposed use of a LOLP measure based on the expected number of hours with shortages. As discussed previously, the use of common names (LOLP, in this case) for supply adequacy indices of peak-constrained systems, as in most of the US, and for supply indices of a more complex energy- and peak-constrained system, as the Pacific Northwest, has the potential disadvantage of signaling that they are more comparable than they actually are. It is perhaps more important to show, as the BPA personnel has done in the presentations to reviewer on this topic, how different they are, and how important it is to develop indices that are as suitable as possible for the PNW situation.

For this reason, the reviewer suggests that other indices, such as those discussed in the previous chapter, could be considered by BPA.

6 USE OF CVAR AS A METRIC

In this section, we discuss some practical aspects for the application of CVaR to the Pacific Northwest system.

1. *What α value should be used? Are there any guidelines that will help us determine that value?*

As seen in the review of international experience, there seems to be little economic justification for the adequacy thresholds that are used. In the case of Brazil, we applied the "sanity check" mentioned above, of comparing the *implicit* cost of rationing that results from our 5% rationing risk criterion (i.e. we ran successive optimal expansion planning models with increasing rationing costs until the rationing risk of the optimal plan was 5%) with the "explicit" cost of rationing that was obtained from an input-output Leontieff matrix of the whole Brazilian economy. Because the implicit and explicit costs of rationing were similar, we concluded that our 5% risk criterion was adequate.

A simpler (but less precise) alternative to the above procedure is to use the approximate relationship: long-run marginal cost ≈ cost of rationing × risk of rationing (this relationship results from the optimality conditions of the expansion planning problem – minimize the sum of investment + operation + load curtailment costs). Suppose, for example, that your LRMC is 70 US$/MWh and that your proposed risk criterion is 5%. From the expression, we conclude that your implicit cost of rationing is approximately 70/0.05 = 1400 US$/MWh. If you feel that the "real" cost of rationing is higher than that, it means that you should "tighten" your risk value to, for example, 4%; and vice-versa.

Obviously, none of the above procedures is able to produce the "optimal" risk value, but they can help avoiding values that are too tight or too loose.

2. *Once the α value has been determined, how do we set a threshold (to differentiate between adequate and inadequate supplies)?*

In the case of Brazil, we used the probabilistic simulation with the generation configuration that was adjusted for the 5% risk level to calculate the resulting levels of load curtailment. In our case, the average load curtailment was around 8% of the yearly load. In other words, we used the risk metric to "calibrate" the CVaR metric.

As in the previous case, this procedure is far from perfect, but is helpful for avoiding gross mistakes.

3. *Should we define both an energy and capacity CVaR metric?*

In the case of the Pacific Northwest, the use of two CvaR metrics (energy and capacity) is probably the most prudent path. As more experience is acquired with the metrics, it may be interesting to analyze the possibility of using a single metric, related to load supplying capability. This topic is discussed later in this section.

4. *To assess a capacity CVaR value, should we count only the worst hour curtailment per game? Or, should we calculate the worst sustained-period curtailment? And, what happens if a curtailment doesn't occur over a full sustained period?*

That depends on whether the capacity curtailment was caused by lack of water or by lack of peaking capacity itself (due for example to generation outages). If it was caused by lack water, it is possibly better to use the total curtailed MWh along the peak period. If it was caused by generation outages, it is better to use the worst hour. The reviewer fully understands that this is easier said than done, because of the previously mentioned interaction between lack of water and equipment outages.

A possible way to get a "feel" for which of the causes (lack of water or equipment outages) is dominant is to run a generation reliability evaluation model, in which we assume that there is ample water. If the resulting expected load curtailment is close to what is obtained when an integrated water and outage Monte Carlo simulation model is used, then outages are dominant, and vice versa: if the load curtailment is small compared to the integrated result, then water is dominant.

As mentioned previously, there is a third possibility, that neither lack of water nor equipment outages by themselves suffice to cause load curtailment, but their "composite" effect does cause it. In this case, one option is to use the LSC scheme, discussed later in this section.

5. *To assess an energy CVaR value, should we count the total curtailment per game or the worst event curtailment per game? The thought is that if we find a solution for the worst event, it should suffice for all the others.*

In the reviewer's opinion, the total curtailment per game is possibly a better measure for the energy inadequacy. The reason is that energy problems are likely to be related to lack of water, which has an impact across the time stages. In other words, a 300 MWh curtalment in stage 10, followed by a 100 MWh curtalment in stage 25 (total 400 MWh), should not be very different, in terms of energy inadequacy, from a 200 MWh curtailment in both stages, also totaling 400 MWh.

6.1 Example of Load Supplying Capability

In this section, we illustrate the use of a load supplying capability (LSC) methodology for assessing the impact of inflow variation, equipment outages etc. on the Pacific Northwest supply adequacy.

The computational tool used was PSR's hourly hydrothermal scheduling model, NCP. The PNW data was provided by Dr. Mike McCoy, and includes: (i) Plant operational data for 88 hydro and 55 thermal generators; (ii) Demand profile (hourly values); and (iv) Water years file.

We concentrated on the modeling of the hydrological uncertainty. As mentioned previously, the representation of the other sources of uncertainty is straightforward.

For each hydrological year *y* and month *m*, we read from the BPA simulation output file the natural inflows and the initial & final storage levels. We then solved the following mixed integer programming (MIP) optimization problem using NCP:

Max λ

Subject to:

$$\sum_{i=1}^{I} g_{it} \geq \lambda d_t \; for \; t = 1, \dots, T$$

+ other operational constraints, such as: water balance, water routing times, minimum and maximum flows limits, initial and final storage levels, flood control levels, hydropower modeling (ex: forebay elevation curve, tailwater elevation curve, encroachment, available capacity, fuel constraints, etc.).

The scalar variable λ represents what % of the reference load can be supplied with no shortages. A value of λ > 1 at the optimal solution means that the system can safely supply an even higher load. Conversely, a value of λ < 1 means that the supply is inadequate.

In turn, the variables $\{g_{it}\}$ represent the energy production of each plant i in each hour t. Finally, d_t represents the reference load in hour t.

The figure below shows the optimal result for the first week of September 2010 and one hydro year.

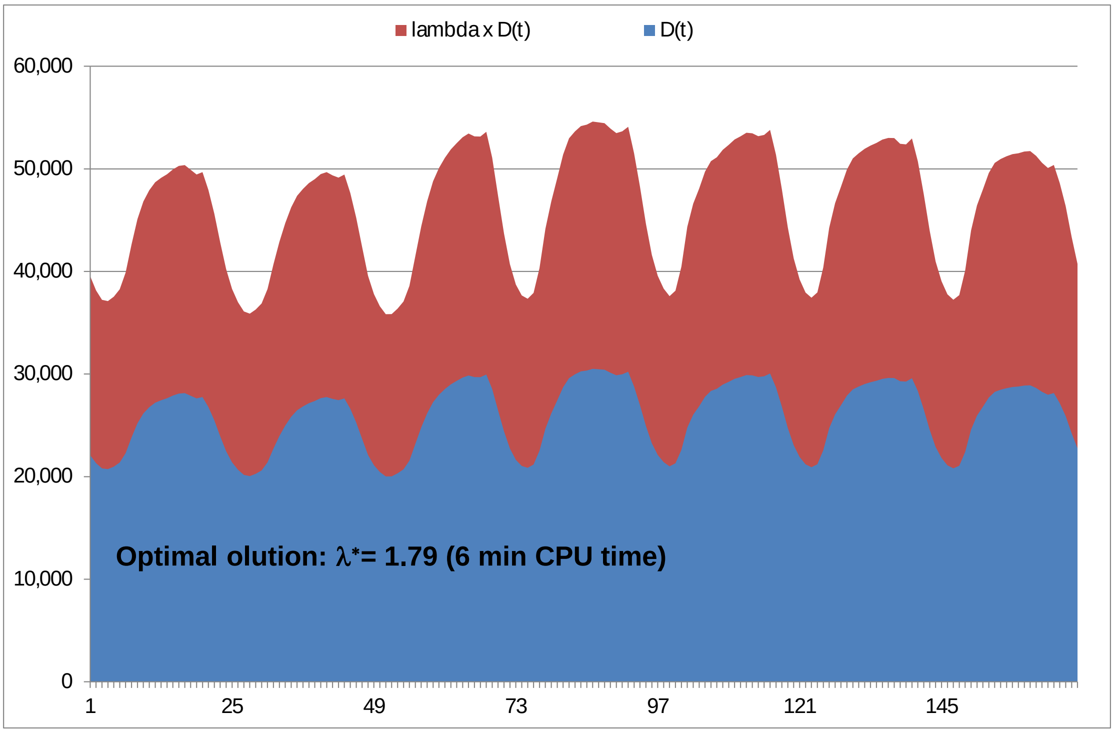

Figure 4.2 – Illustration of LSC calculation in the PNW – first week of September 2010

We see in the figure that the optimal λ in this case was 1.79. As discussed, this means that the generation system could supply a load that was 80% higher than the reference.

The figure below shows in more detail the hourly operation of each hydro and thermal plant along that week.

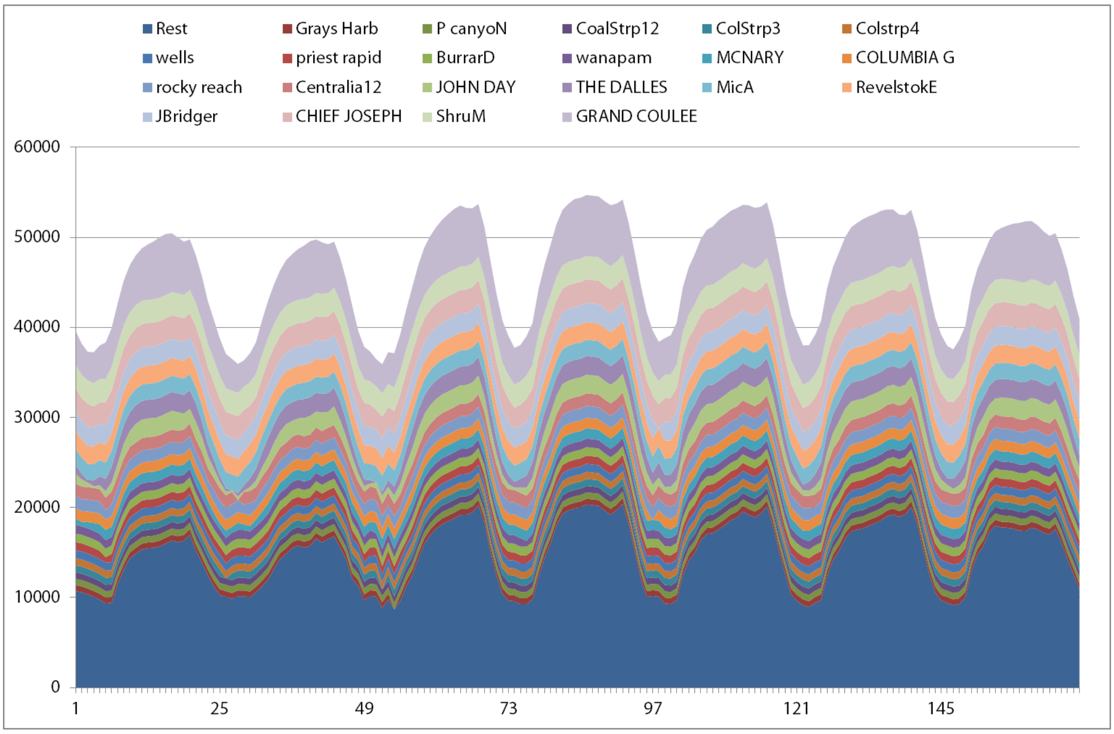

Figure 4.3 – Illustration of LSC calculation in the PNW – Hourly generation profiles

7 CONCLUSIONS

The use of supply adequacy criteria is an essential tool in the search for the best tradeoff between electricity costs and supply adequacy.

Historically, separate methodologies and supply criteria were developed for capacity-constrained and energy-constrained systems, and there was limited communication between both "worlds". With the acceleration of renewable power integration, in particular wind, the separation between both types of systems is becoming less sharp. In the reviewer's opinion, this is the case of the PNW system, which can now be characterized as both peak- and energy-constrained.

Another challenge is to choose between "risk" and "severity" indices, as they capture complementary aspects of the overall supply adequacy picture. Although it is, in principle, possible to use both types of indices for the peak and energy measures, this is not desirable, because it requires the definition of four different sets of standards and thresholds. Given the difficulty of having an economic justification for these standards, it becomes likely that they will not be very consistent, i.e. one of them will be more "strict" than the other.

In the reviewer's opinion, the recently developed Conditioned Value at Risk (CVaR) measure has the potential of combining the best features of "risk" and "severity" indices. The CVaR is widely used in the financial area and, as shown in the references, is beginning to be applied in power system planning and operations.

Also in the reviewer's opinion, the Load Supplying Capability (LSC) may be a potentially attractive path for integrating peak and energy adequacy measures. As illustrated in the body of this report, the LSC is able to capture the interdependencies between equipment outages, load fluctuations and streamflow volatility. In addition, the LSC provides useful information not only when the supply is inadequate, but also when there is "surplus" capacity, as it indicates how much additional load can be safely supplied.

Even if the "risk" and "severity" indices are merged into one CVaR index and the capacity and energy measures are integrated into a single LSC measure, there remains the challenge of defining the "threshold" between adequate and inadequate supply. As discussed in the document, there is no "right" way of doing this, but some "sanity checks" such as the implicit cost of rationing and the approximate relationship between long-run marginal cost (LRMC), cost of unserved energy and risk of load curtailment can be used as a support tool to avoid major inconsistencies.

8 CVAR REFERENCES

8.1 Methodology

[1] ROCKAFELLAR, R. T., URYASEV, S., "Optimization of Conditional Value-at-Risk", Journal of Risk, v. 2, pp. 21-41, 2000.

8.2 Applications in Finance

[1] ACERBI, C., "Coherent measures of risk in everyday market practice", Quantitative Finance, v. 7, pp. 359-364, August 2007.

[2] ACERBI, C., NORDIO, C., SIRTORI, C., "Expected Shortfall as a Tool for Financial Risk Management", arXiv, February 2001.

[3] ACERBI, C., TASCHE, D., "Expected Shortfall: a natural coherent alternative to Value at Risk", arXiv, May 2001.

[4] ACERBI, C., TASCHE, D., "On the coherence of Expected Shortfall", Journal of Banking & Finance, v. 26, n. 7, pp. 1487-1503, 2002.

[6] ARTZNER, P., DELBAEN, F., EBER, J.-M., et al., "Thinking coherently", The Journal of Risk, v. 10, n. 11, pp. 68-71, 1997.

[7] ARTZNER, P., DELBAEN, F., EBER, J.-M., et al., "Coherent measures of risk", Mathematical Finance, v. 9, n. 3, pp. 203-228, 1999.

8.3 Applications in power system planning and operations

[1] GUIGUES, V., RÖMISCH, W., "Sampling-based decomposition methods for risk-averse multistage stochastic programs", Optimization Online, October 2010.

[2] JABR, R. A., "Robust self-scheduling under price uncertainty using conditional value-at-risk", Power Systems, IEEE Transactions on, v. 20, pp. 1852-1858, Nov. 2005.

[3] PHILPOTT, A. B., MATOS, V., "Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion", Optimization Online, December 2010.

[4] SHAPIRO, A., "On a time consistency concept in risk averse multistage stochastic programming", Operations Research Letters, v. 37, n. 3, pp. 143-147, 2009.