Cost-Causation-Based Tariffs for Wind Ancillary Service Impacts

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

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

Yih-huei Wan
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401
USA
303-384-7025
yih-huei_wan@nrel.gov

Abstract

As the use of wind energy increases in the United States, there has been significant interest in assessing the integration cost of wind. With higher penetration, these costs have been shown to increase, but are generally modest relative to the price of wind energy. Although specific tariffs for wind generation for ancillary services are uncommon, we anticipate that balancing authorities (control areas) and other entities will move towards such tariffs. Tariffs for regulation and imbalance services should be cost-based, recognize the relevant time scales that correspond with utility operational cycles, and properly allocate those costs to those entities that cause the balancing authority to incur the costs. In this paper we present methods for separating wind's impact into regulation and load following (imbalance) time scales. We show that approximating these impacts with simpler methods can significantly distort cost causation, and even cause confusion between the relevant time scales. Correctly calculating and allocating cost impacts provides a market signal that can encourage economic efficiency, and avoids subsidies. We present results from NREL's wind data collection program to illustrate some of the dangers of linearly scaling wind resource data from small wind plants to approximate the wind resource data from large wind plants. Errors in scaling can cause a significant over-estimate of wind impacts in the regulation and load following time frames and result in cross-subsidies in the resulting tariff. Finally, we provide a framework for developing regulation and imbalance tariffs, we outline methods to begin examining contingency reserve requirements for wind plants, we provide guidance on the important characteristics to consider, and we provide hypothetical cases that the tariff can be tested against to see if the results are what are desired.

Introduction

In this paper we discuss general principles of tariffs, with the goal of pointing out how tariffs can address pricing for services based on cost-causation. Because the use of wind energy has expanded significantly over the past few years, some Balancing Authorities (BA) are considering how to develop their own tariffs for wind plant integration. If care is not used to develop tariffs that address the specific costs for regulation and imbalance, for example, then these tariffs may either over-charge or under-charge wind for its impacts.

Good Tariff Design Should Be Based On Operating Principles

Modern power systems are complex networks of machines. To operate effectively, system operation personnel have developed various procedures that assure the reliable operation of the power system. The fundamental task is to maintain balance between system loads and generation. Although this is simple in concept, the practice of power system operation is quite complicated. Electric loads vary over all time scales, from a few cycles up through longer periods such as days, weeks, or months. The task of maintaining system balance is divided according to the time scale. Figure 1 is a typical representation of the critical time scales. From the operational perspective, the unit commitment time scale, which can range from several hours to a few days, represents the time required to ensure adequate slow-start generating units are available to provide power. Performing the commitment function correctly requires knowledge of future loads, along with other relevant variables like the availability of other generating plants.

The Regulation Time Frame

Once the unit commitment process is completed, the operator has the ability to control this online generation, subject to the various physical and electrical constraints of the generators and the balance of system. Some of this control involves automatic generator control (AGC) in response to the changing load-resource mix, performed automatically by computer. This process typically occurs over several seconds to several minutes, depending on the system. AGC units respond in the regulation time scale, and this ancillary service is typically quite expensive. In the minute-to-minute regulation time frame loads move in uncorrelated patterns, and the generators that perform the regulation service do not need to match each individual load change—only the aggregate system balance needs to be chased by the regulating units.

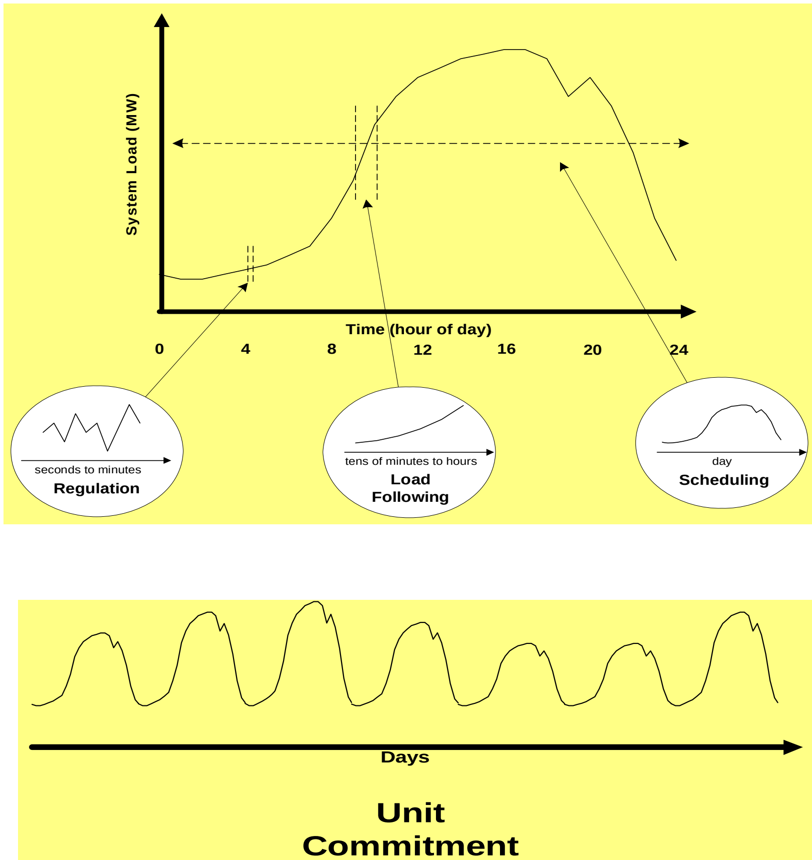

Figure 1. Representation of critical time scales for system operation

The Load Following Time Frame

The next time frame is load following. Changes in load in this time frame are generally correlated across electric customers, although there is typically a regulation "wiggle" that is superimposed on the correlated movements. The correlation typically occurs during the morning load pickup, when many customers increase their electrical demand, and in the evening when electrical demand falls off. The system operator must follow load with generation that is either already online, or generation that can be brought online quickly. In either case, the load following unit(s) must have the capability of changing output in response to operator instruction, and must be able to follow the instruction fairly closely though they do not necessarily have to be on AGC.

The Unit Commitment Time Frame

Significant time is required to prepare many generators for operation. The unit commitment decision, the process of deciding which generators are going to be used on a given day, is typically made the day before the operating day based on the load forecast and the schedule of available generators.

The total package of providing load following and regulation, along with voltage support (VAR support), contingency reserves, and imbalance service is called ancillary service provision. In some parts of the U.S. there are markets that cover some ancillary services. In other parts of the country, utilities still operate as regulated monopolies, and sometimes have access to wholesale markets to purchase energy, capacity, and ancillary services. However, this latter market is not widespread today in the U.S.

Regardless of whether electricity is provided by a market or regulated monopoly, the key performance criterion for power system operation is to ensure that generation and load are in balance. This means that the *aggregate* load must be matched by the *aggregate* generation within statistical limits that are prescribed by the North American Electric Reliability Council (NERC). This has a critical implication for wind: it is not necessary or useful to match each increase or decrease in wind power output by a corresponding movement of another generator. Instead, only the aggregate system increases or decreases in demand must be matched by changes in generation. This has important implications for tariffs related to wind ancillary service provision. The metric used to measure the impact of wind on ancillary service requirements for the system cannot be based solely on wind generation alone. It must instead be based on the contribution that wind makes to the overall level of ancillary services needed to balance the system.

Economic Efficiency and Ancillary Services

Because it is critical to operate the power system in a reliable and economic manner, one can deduce that society benefits if the economically efficient quantities of these ancillary services are provided, whether by regulated entities or markets, at their unique cost of service. Charging for these products at the cost of service sends the correct economic signal to users, who can then chose to use the services or not based on prices that reflect the value of the services to them at a particular, cost-based price. Economic efficiency is achieved only if the marginal benefit to society equals marginal cost to society. In perfectly competitive markets the marginal benefit curve coincides with the demand curve in the absence of externalities. Similarly, marginal costs coincide with the supply curve. Application of these concepts to ancillary service markets implies that the economically efficient quantity of the ancillary service should equal its marginal cost. The benefit to society includes reliable power system operation, and because ancillary services are required for reliable electricity supplies, the benefit of reliability is included in overall benefits.

Using regulation as an example, the left panel of Figure 2 illustrates the consequences of too much regulation capacity. The upward sloping red line represents the marginal cost (supply) of regulating resources. To obtain more regulation, buyers need to increase the price. The downward sloping blue line represents the marginal benefit (demand) for regulation. The intersection of these curves shows the socially optimal level of regulation. However, if more regulation is provided than is optimal, such as represented by the vertical green line, then a deadweight loss is incurred by society. This deadweight loss is represented by the black triangle, and can be interpreted as the economic value that is lost due to the oversupply of regulation. Although an oversupply of regulating resources is not harmful from the reliability standpoint it is costly and economically inefficient.

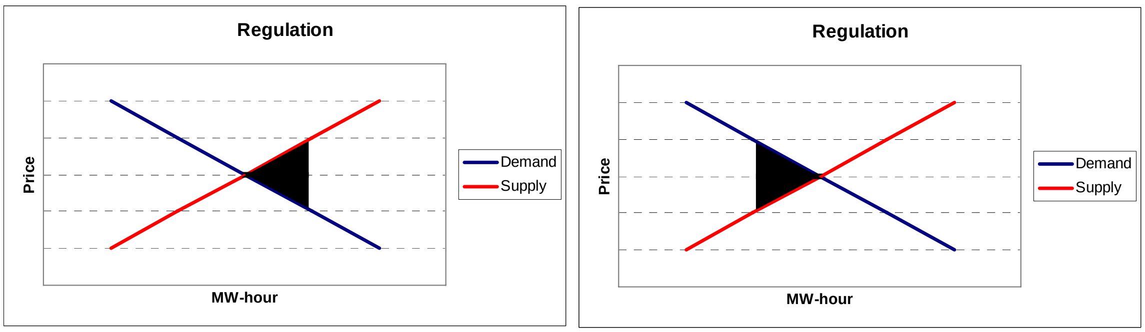

Figure 2. Deadweight loss for (a) too much regulation, (b) insufficient regulation

A similar deadweight loss occurs if there is not enough regulation. In this case, system reliability may be compromised. Buyers would be willing to buy more regulation than the market or tariffs provide at this inadequate level. But if there were some type of constraint preventing the acquisition of enough regulation, a deadweight loss, represented by the triangle in the right panel of Figure 2 would occur. Although these graphs are conceptual, they collectively illustrate why it is important to correctly determine the efficient quantity of regulating resources. This simple example could easily be extended for other ancillary services.

If ancillary services are provided by a market, one benchmark for the market would be to determine if the efficient level of ancillary services are provided. Markets that have either too much or too little ancillary service provisions are not efficient.[1] From this discussion we can conclude that

- There is a real benefit to acquiring sufficient ancillary services
- There is a real cost of acquiring excess ancillary services
- The objective is to balance aggregate system load with available aggregate generation, and it is therefore un-necessary and costly to balance individual loads or generators
- Markets, and therefore tariffs, should encourage market behavior that results in economically efficient outcomes (minimal deadweight loss)

---

[1] Throughout this discussion we assume that the cost and benefit curves have been adjusted for risk, so that a rational operator with imperfect foresight can prudently acquire ancillary services that may not be needed in hindsight, but whose cost is less than what might be incurred if insufficient regulation (or other ancillary service) is not available.

Tariffs Should be based on Cost Causation Principles[2]

These points, however, do not consider how costs should be shared among the various users of the products. We adopt the overall principle that tariffs should be based on cost-causation, as illustrated in the following more detailed points:

1. Because maintaining power system reliability is critical, tariffs should base prices on costs so that the costs of maintaining reliability can be obvious to users of the system and its reliability feature. .
2. Tariffs should be based on cost-causation and the cost of providing the service.
   a. Those individuals who cause costs to the system should pay for those costs
   b. Those individuals who mitigate costs to the system should either incur a lower cost or be paid for helpful actions
   c. Complex systems like electric grids produce both joint products and joint costs of production that must be allocated among users of the system.[3]
   d. Tariffs should allocate joint production costs on the basis of the use of joint products (the cost allocation principle of "relative use")
3. Tariffs should not collect revenue if no cost is incurred
4. Tariffs should be based on the physical behavior and characteristics of the power system
   a. Recognize the need to balance aggregate system load and aggregate system generation
   b. Recognize that balancing individual loads or resources is not necessary, is inconsistent with power system operations and is very costly
5. Tariffs should result in an efficient allocation of resources

Although our analysis is based primarily on wind energy, these principles should apply equally to all agents in the relevant markets; loads and generators.

Tariffs can be empirically tested, either thru real-world experience or through detailed modeling of the grid and the individual costs and behaviors in question. During the process of tariff design, hypothetical cases can be established to test each principle in question, and the performance of the tariff can be assessed against the principles of tariff design above. In a later section of this paper we propose a series of these tests that we call "thought experiments." Application of these thought experiments can reveal how well a tariff matches the principles set forth above.

There are some broader principles that tariffs should also support. The first is horizontal consistency. Horizontal consistency means that if two individuals (loads or generators) each cause equal increases in costs, then the tariff should assess each of them the same amount. A corollary to this principle is that if two individuals impose similar costs, then

---

[2] The classic text is Bonbright, James Cummings, "Principles of Public Utility Rates" (1961).

[3] The classic example is a sheep. A farmer raises a sheep. She sells mutton, hide, and wool. These are joint products. She incurs various costs for raising the sheep: feed, medicine, and a Shepard. These are joint costs of production. The electric system produces joint products: reliability, energy, capacity, convenient system access, ancillary services. The costs for producing these joint products (for fuel, engineers, capital, maintenance) must be allocated to the joint products. The most common allocation principle is relative use, the more you use, the more you pay.

they should be assessed similar payment amounts. We can extend the principle of horizontal consistency in cases where individuals contribute to cost mitigation. Equal cost mitigations or reductions should be matched by either identical reductions in cost assessment to the individuals, or equal payments to the individuals. If two individuals have similar cost mitigation impacts, then their payments should be similar.

Vertical consistency is the second additional principle. Vertical consistency implies that if individual A imposes a larger cost than individual B, then A should pay more than B. We can extend the concept of vertical consistency to cases where two individuals mitigate costs in a straightforward manner.

Horizontal and vertical consistency can be empirically tested, either thru real-world experience or through detailed modeling of the grid and the individual behaviors in question. Application of the tariff to the individual behaviors can determine whether horizontal and vertical consistency is achieved by the tariff.

Inherent Characteristics of Wind

Tariffs, reliability rules, and market design should include attention to the inherent characteristics of all of the market participants. This does not mean that any participant should receive preferential treatment. Most regulatory laws require that "undue" discrimination be avoided, not that all discrimination be avoided. Recognition of rational differences, as between rate classes for example, is "due" discrimination. Therefore, physical capabilities as well as physical limitations should be understood before designing rules. It makes little sense, for example, to impose imbalance *penalties* on wind plants. Wind output can generally not be controlled (at least under-generation can not be controlled). It can make sense to allocate imbalance *costs* to those that create them but penalties, which are designed to motivate behavioral change, do no good.

Aggregation

For nearly a hundred years, since the early days of electric power system development, it has been recognized that it is more reliable, easier and less expensive to serve an aggregation of loads rather than supplying each individual load separately. Balancing each load's variability with dedicated generation would be too expensive to consider. Instead, loads are aggregated and much of their variability cancels, greatly reducing everyone's costs.

This explains why "backing up" wind plants with dedicated generators is never done. Using the aggregation of generators available to a system operator is always a more economic approach to dealing with the variability of wind output, since the operator can chose the best option from a large menu of options for handling these tasks.

Wind generation itself also benefits immensely from aggregation. Individual wind turbines are limited in electrical size. Creating wind plants requires aggregating large numbers of individual wind turbines. Wind turbines are physically separated. The physical nature of the turbulent atmosphere interacting with the wind plant geography makes it impossible to get large numbers of wind turbines to synchronize their behavior, especially over the short time frames that are important for regulation and contingency reserves. So while the wind might be dying down in one part of a wind plant, it might be increasing in strength in another part. This effect is even more likely if wind plants are located across broad geographic areas. Aggregating multiple wind plants provides additional aggregation benefits. Separating wind plants by a few kilometers eliminates regulation correlation as shown in Figure 3.

Energy output from multiple wind generators adds linearly to equal the total output of a wind farm. The energy output of the wind plant is simply the sum of the outputs of the individual turbines. The variability of individual turbines, however, does not add linearly. The benefits of aggregation of numbers of wind turbines in the regulation time frame can be seen in Table 1 even for a small 11 turbine wind plant. The standard deviation of the

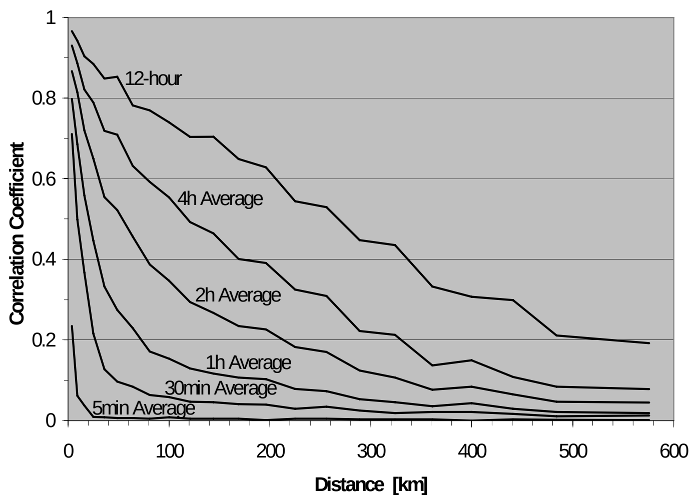

Figure 3. Wind generator variability looses corelation as the distance between machines increases and as the time frame of interst decreases. (Ernst, Wan, and Kirby 1999)

one minute output (after removing the load following component) provides a good metric for the regulation requirement. If the turbines' output variability was highly correlated, if the same wind turbulence hits each turbine at the exact same time, the variability of the array would be eleven times the average variability of each machine; 0.648 MW. But this is not the case. The total wind plant variability is less than half at 0.315 MW. Interestingly, this is 1.5 times the 0.197 MW that would be expected if the individual turbine outputs were completely uncorrelated. Note that the ratio of regulation to energy is cut in half by aggregating these eleven turbines. This why it is incorrect in modeling and in tariff design to assume (or assess) regulation requirements based on simple scaling.

Aggregating over larger numbers of turbines and greater geography produces more dramatic results (Table 2). Wind plant variability is reduced from 7.5 MW of regulation that would be required if all four sections of the example wind plant had to balance individually to 4.8 MW for the total wind plant. Aggregating wind plant variability with aggregate system load further reduces the amount of regulating reserves that are required to balance the power system and maintain reliability. The wind plant's share of the total regulation requirement is 0.8 MW. This is the same aggregation benefit and genuine savings that led to the original aggregation of loads and the creation of interconnected balancing authority areas (control areas).

Table 1. Regulation variability is greatly reduced through aggregation

	Total Plant	11 Individual Turbines Independent	Sum	Average
Energy (MW)	2.225	2.225	2.225	.202
Regulation StDev (MW)	.315	.197	.648	.059
StDev/Energy	14%	9%	30%	30%

Wind tariffs should recognize aggregation benefits, just the same way that load tariffs recognize these benefits. Charging for individual wind farm or wind turbine variability greatly over collects for regulation. It also encourages inefficient balancing on a smaller scale resulting in needless and costly AGC operations. Individuals (wind plants, individual loads, and other uncontrolled generators) should be charged for their properly allocated share of the total system variability based upon the impact their variability has on that total.

Table 2. The wind plant benefits by aggregating the four internal sections and with the balancing authority load.

	Interconnection Point				
	A	B	C	D	Total
Number of turbines	30	39	14	55	138
MW Rating	23	29	10	41	103
Stand-alone regulation requirement (MW)	1.8	2.2	1.0	2.5	4.8
Balancing Authority regulation allocation (MW)	—	—	—	—	0.8

The implications from these aggregation data is profound. As more electric utilities and other entities become involved with wind integration studies that evaluate impacts of wind on power systems operations and costs, there is a need to utilize wind generation data in these efforts. Often, wind data can be obtained from anemometers that are located in the proximity of the potential wind plants. Because the primary impacts of wind on the power system are because of the increased variability on the grid, any credible analysis must be based on a realistic modeling of the wind generation itself. We can't stress this point enough: **scaling from a single anemometer, or small wind plant, to a large wind plant that could be hundreds of MW in size, results in an inaccurate representation of the wind plant, and will overstate the impact of wind on the power system**. This will lead to overestimates of the ancillary service cost of wind. Such a tariff would clearly violate the principle of cost causation, and could lead to wind subsidizing other market participants' costs.

Contingency Reserves

Understanding wind plants' impacts on contingency reserve requirements is an area that needs additional research. Contingency reserves are required to protect the power system from the unexpected, sudden, loss of a large generator or transmission facility. When such a sudden loss occurs, a series of contingency reserves begins responding immediately to restore the generation and load balance. Governors on all generators respond autonomously to frequency deviations. Spinning reserves immediately follow with non-spinning reserves and replacement or supplemental reserves after that. The system operator is typically able to restore market operations or economic dispatch within two hours, often much sooner. When the reserves themselves are restored the system is back to normal. Figure 4 shows both a contingency event and the stacking of reserve response.

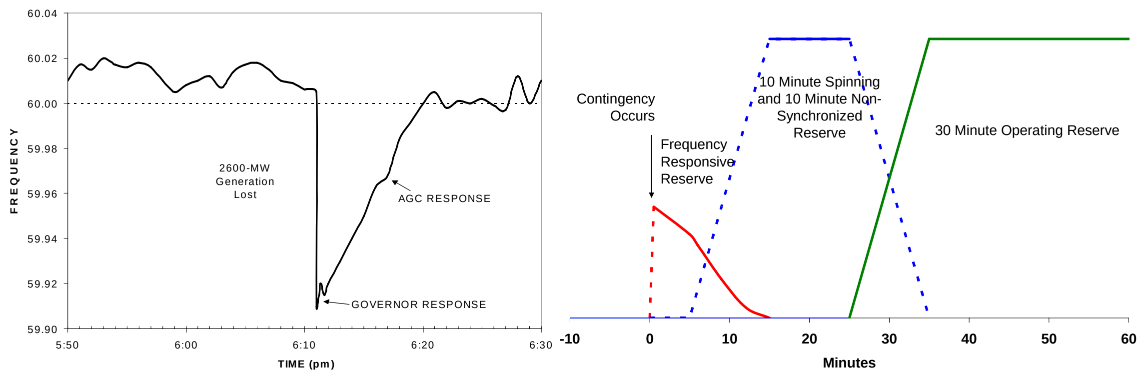

Figure 4 A series of contingency reserves are available to respond to the unexpected loss of a generator or transmission line.

How should contingency reserves be allocated to wind plants? This is a question that requires further investigation, but the outlines of that investigation can be specified here. The nature of wind plant design complicates the evaluation. Wind plants are composed of numerous small generators. Individual generators under 5 MW each do not warrant contingency consideration in any balancing authority area. So machine failures are not a concern. At the other extreme, failure of a radial transmission line that removes a few hundred MW wind plant from service clearly is a contingency that requires attention. Both of these type events should receive the same treatment as afforded other similarly sized and positioned generators. But what about wind driven events?

Examining the contingency characteristics of conventional power plants may help clarify the issue. We tend to think of contingencies as being based on the cause; something breaks and the power plant fails, lightning strikes and the transmission line trips. Underlying the various specific causes are the basic characteristics of a contingency:

- the event is relatively infrequent,
- large enough to need special treatment, and
- too fast for markets to accommodate.

The amount of contingency reserves that must be available usually depends on the size of the largest generating unit. The *minimum* size of a contingency event of concern depends on the size of the reserve sharing pool or balancing authority, but is typically in the 20 to 200 MW range.

The event speed is also not rigorously defined. Most contingency events are essentially instantaneous but something happening over a few minutes would also be too quick for a market response. Guidance concerning what would be too frequent an occurrence to be considered a contingency (and should be considered a normal part of the resources' market performance) can be found by looking at the behavior of conventional generators. Figure 5 shows the failure rates for 120 coal fired power plants in the 600 MW to 800 MW size range for 1996.

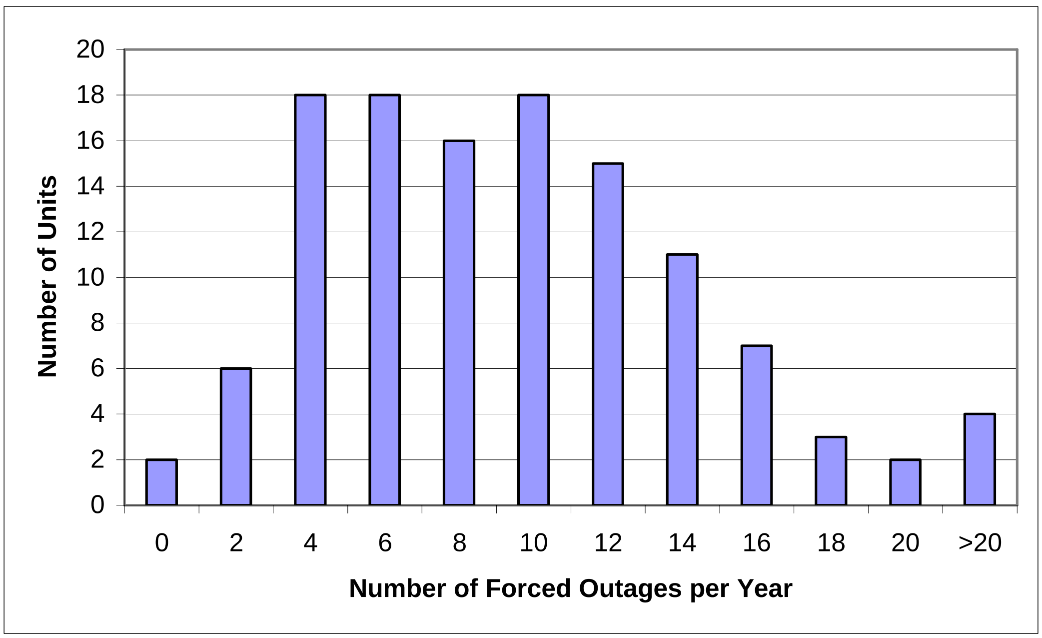

Figure 5. Conventional generators exhibit a range of acceptable outage rates for contingency reserve insurance.

Wind driven events require more study to determine if any have characteristics that are similar to those exhibited by conventional power plants and that should be treated as contingencies. This is complicated by the fact that wind plants exhibit variability across a broad spectrum of size and frequency. Aggregation typically reduces the magnitude of swings and lengthens the ramps. Only large, fast, and rare events should be considered contingencies. Slower or more frequent variability is better considered as load following and regulation. Figure 6 provides an example of events that one might treat as contingencies if it was found that they were infrequent enough. Significant statistical analysis is required to make this determination. Work is also needed to determine if a rapid return to service is typical (similar to transmission line reclosing) and if that is an important consideration.

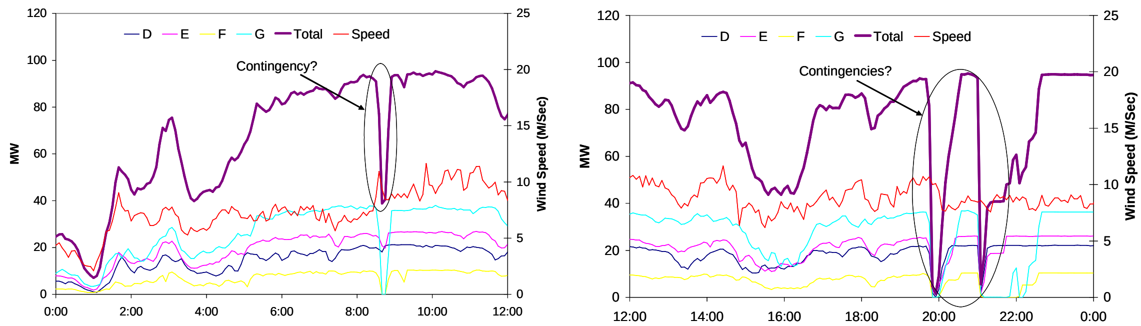

Figure 6. Some wind driven events may be large, fast, and infrequent enough to justify being treated as contingencies.

Impact of Wind on Power System Costs

Several wind integration cost studies have been done in the United States during the past few years, and are reviewed in Parsons et. al (2006). The more recent studies have benefited from ongoing technical review and increasing sophistication of methods to model the wind resource. Because these studies are prospective, evaluating the potential impact of various wind generation scenarios that do not yet exist, a significant effort has been placed on creating a realistic time series representation of wind generation. The overall approach involves a meso-scale meteorological model that recreates the state of the atmosphere for the study period. As this model runs, 10-minute wind speed can be extracted for as many locations as desired. Current practice indicates that a maximum of approximately 30 to 40 MW of wind generation can be represented by one of these “virtual” anemometers. A recent integration study for New York state collected data to represent more than 100 wind locations, and a recent study in Colorado collected data representing more than 1,000 MW of wind using 22 towers to represent 722 MW of wind generation.

Load following impacts have generally been assessed by analyzing the variability of load and wind together. This is the closest way to approximate the view of the system from the control room. At current wind penetration levels in the U.S. the impact of wind on this time scale is relatively small.

Regulation impacts are calculated using statistics from real wind plants and loads. Because these data are generally uncorrelated, the influence of wind on the regulation required to maintain system balance is small, and is generally the same order of magnitude as the regulation requirements for load.

Table 3 shows the impact of wind on ancillary service costs. It is important to note that there is not a one-size-fits-all cost impact. This is because systems can vary significantly, and because the studies have assumed different wind penetration rates.

Table 3. Wind Impact on System Costs

Date	Study	Wind Capacity Penetra-tion (%)	Regula-tion Cost ($/MWh)	Load Following Cost ($/MWh)	Unit Commit-ment Cost ($/MWh)	Gas Supply Cost ($/MWh)	Total Operating Cost Impact ($/MWh)
May ‘03	Xcel-UWIG	3.5	0	0.41	1.44	na	1.85
Sep ‘04	Xcel-MNDOC	15	0.23	na	4.37	na	4.60
July ‘04	CA RPS Phase III	4	0.46	na	na	na	na
June ‘03	We Energies	4	1.12	0.09	0.69	na	1.90
June ‘03	We Energies	29	1.02	0.15	1.75	na	2.92
2005	PacifiCorp	20	0	1.6	3.0	na	4.60
May ’06	Xcel-PSCo	10	0.20	na	2.26	1.26	3.72
May ‘06	Xcel-PSCo	15	0.20	na	3.32	1.45	4.97

Testing a Tariff with Thought Experiments

Thought experiments provide a means for testing a tariff to assure that it does what is intended and that it does not have undesired consequences. The behavior of the wind plants, other generators, loads, and power system components are carefully specified to test each tariff attribute of concern. Here we present five thought experiments that can be used to test how a regulation tariff assesses a volatile resource like wind. Each thought experiment is mapped to at least one of our tariff principles.

Thought Experiment #1: Perfect Following of a Volatile Schedule

In formal transactions both loads and generators forecast their expected behavior and establish a schedule for generation or consumption. Regulation tariffs often impose penalties if a resource does not follow its schedule. Some tariffs are based exclusively on schedule deviations. The reasoning is that the system operator must have a reserve of regulating resources available to immediately compensate for unexpected changes in a generator or load’s output or consumption. This is true. But does the regulation resource requirement go away if the resource follows its schedule perfectly? Figure 7 presents a typical system daily load with blocks of generation scheduled to meet that load. If the generation follows its schedule perfectly is there a regulation burden imposed on the system? What charge does the tariff impose?

A regulation tariff that is based exclusively on schedule deviations would impose no charge on the block scheduled generator. Indeed, many feel that scheduled imports and exports impose no regulation burden because the schedule is precisely known, often days in advance, and it is typically adhered to.

The right side of Figure 7 shows that block scheduling imposes severe ramping requirements on the system, adding $2.26 to the cost of each MWH delivered through the block schedule in this example (based upon modeling an example control area). The fact that these requirements always happen at the top of the hour and they are known well in advance does not reduce the amount of fast response capability the system operator needs to have to balance the system and meet CPS 1 & 2 requirements. The tariff needs to assess the individual's impact on total system variability.

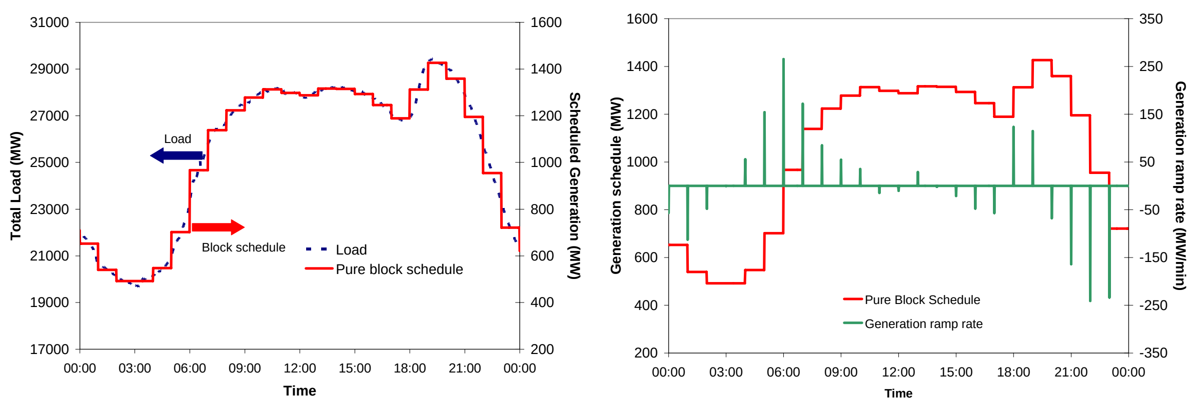

Figure 7. How does the tariff treat perfect following of a volatile schedule?

This example tariff would violate principle #2 (cost-causation) because under the (unlikely) scenario of a perfect wind forecast, the tariff would not assess any cost to the wind generator even though there is a cost of moving the regulating units to mitigate variability in the wind output signal. It also violates principle #4, which says that individual movements (or in this case schedule deviations) of individuals do not need to be matched by a responsive unit – only the aggregate variability of the entire system must be compensated. Extrapolating this type of tariff to a case when all schedules and loads are known perfectly in advance, the implication is that there is no cost to the system to chase the total system variability. This is clearly wrong, and would result in distortions in the market.

Thought Experiment #2: Reduced Ramping

It is tempting to design a regulation tariff that simply quantifies the peak-to-peak movements of the generator or load. But this ignores the speed at which the resource moves from one power level to another. If the block schedule used in Thought Experiment #1 (where the schedules changed abruptly at the top of each hour) is provided with 20 minute ramps (where schedules linearly ramp from ten minutes before the hour to ten minutes after the hour), as shown in Figure 8, the regulation costs imposed on the power system drop to $0.20 per MWH (again based on modeling an example control area).[4] Note that the ramp rate scale on the right axis of Figure 8 is one tenth of that in Figure 7.

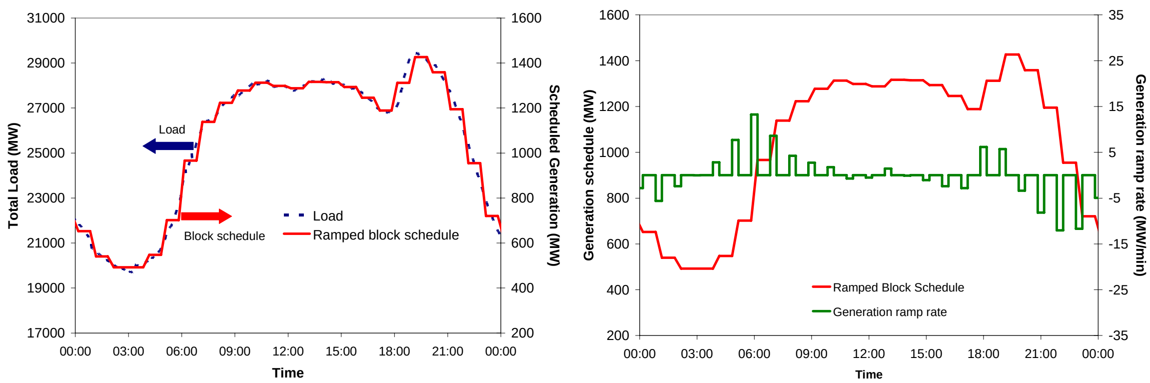

Figure 8. Ramping the block schedule does not impact the energy delivery or forecast accuracy but reduces regulation requirements.

This thought experiment violates principle #2, the principle of cost causation. Recognition of only the peak-to-peak ramp does not distinguish between the two behaviors illustrated here that have significantly different cost impacts. This also violates the principle of vertical consistency because there is a significant difference in imposed cost that would not be picked up in the tariff.

Thought Experiment #3: Ramp Rate or First Derivative Metrics

Another tempting regulation tariff simplification is to measure average ramp rate or the average first derivative of the minute-to-minute energy consumption. This can also be characterized as a “distance traveled” metric referring to the amount of “movement”. This attempts to quantify the amount of ramping or changing of output that a generator has to provide. The flaw in this simplification is that behaviors with very different system impacts can result in the same measured performance as shown in Figure 9.

Figure 9 compares the behavior of three hypothetical individuals (loads, wind generators, or balancing areas). The minute-to-minute change (“line slope”), integrated over the hour, is the same for all three; 60-MW-minutes. Clearly, however, the regulation burdens imposed by the three are radically different. In this very simple example the solid red entity requires 1 MW of regulation compensation. The dashed green entity requires 5 MW. The dotted blue entity requires a total of 60 MW but not of regulation. A sustained ramp is a load following requirement that can be, should be, and is (in most locations) supplied by moving the base load and intermediate generators. There is no regulation burden imposed by the dotted blue ramp.

Metrics based on average rate of change of an individual violate principle #2 (cost causation) and principle #4 (failure to recognize aggregation benefits).

---

[4] Ten minute ramps for interchange scheduling are standard in the eastern interconnection. The western interconnection benefits in reduced regulation costs from the use of twenty minute ramps.

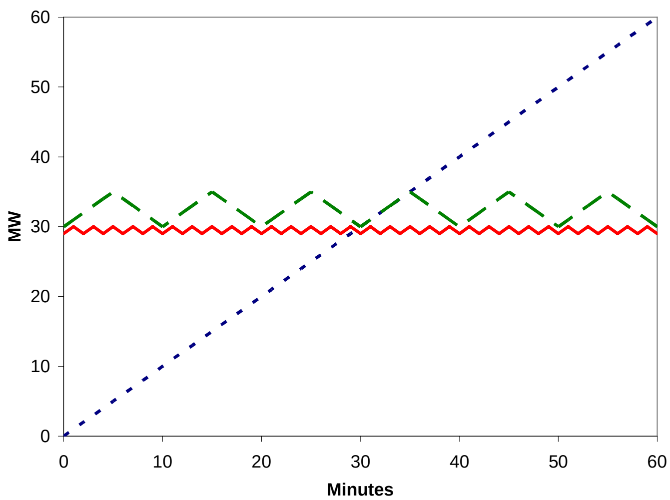

Figure 9 These three individuals impose radically different regulation requirements but have the same minute-to-minute-change metric performance.

Thought Experiment #4: Equal but Opposite Behavior

One very powerful feature of thought experiments is that they can be carefully tailored to examine specific behavior characteristics. They do not have to be realistic to be useful in determining if a tariff will produce desired results. Unrealistic examples can be useful in understanding the pieces of complex behavior that are often buried in the intricacies of actual operations.

When designing a regulation tariff it is tempting to assess the generator's or load's variability in isolation. This ignores the fact that the underlying reliability requirement to balance generation and load is imposed on the balancing authority (hence its name) rather than on the individuals. Figure 10 shows two mirror image wind plants and a total system load. If the wind plants were assessed for their variability in isolation of each other and the total system load they would both receive an identical regulation variability assessment. Together they present an absolutely constant output with no regulation burden.

This thought experiment is completely unrealistic but it illustrates an important point. A tariff that can not recognize complete compensation of one plant for another will not recognize more subtle interactions or uncorrelated behavior that, consequently, does not add linearly.

A tariff that does not recognize the impact of equal but opposite behavior would collect payment from both of these hypothetical wind plants. However, because their impacts net to zero, there would be no cost to the system. This type of tariff would therefore violate principle #3 (the principle that if no cost is incurred, the tariff should not collect revenue) and principle #4 (the recognition that only the aggregate system variability must be compensated for).

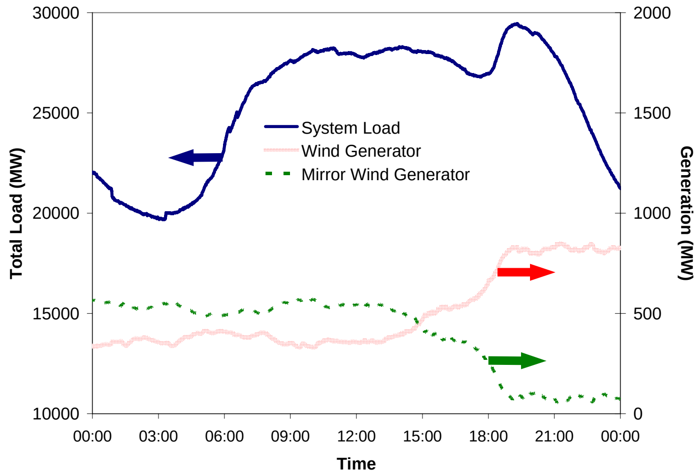

Figure 10 How is equal but opposite behavior treated?

Thought Experiment # 5: Beneficial Movement

The last thought experiment asks how the tariff treats movement that is beneficial. Regulation tariffs that only assess variability (total range, ramp rate, or adherence to a schedule) can penalize a resource that is actually helping reduce the total system aggregate variability. **Error! Reference source not found.** presents the measured variability of a number of generators and a total system. A tariff that simply charged for variability would penalize the AGC Generator that is deliberately balancing the system. Presumably the tariff would not be applied to this generator but the principal remains the same. A generator that inherently has favorable response characteristics for whatever reason should not be penalized.

A tariff that assesses a cost based on an individual's variability in isolation of what the system needs would discourage helpful behavior. Because this type of tariff would impost a cost when in fact the resource is providing a benefit and limiting costs by helping to mitigate system variability, this kind of cost in a tariff clearly violates principle #2 (cost causation), principle #3 (imposing a cost instead of paying the generator), and principle #4 (does not recognize system balance).

Although it is not directly testable, we emphasize that any of these tariff examples would cause an inefficient level of ancillary service to be acquired. Economic inefficiency occurs when marginal benefit diverges from marginal cost, as shown in Figure 2. This distortion in cost and price signals does not encourage behavior that helps the power system, and discourages helpful behavior

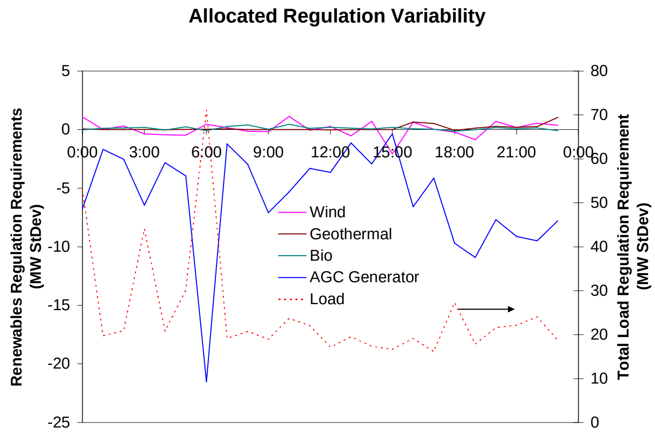

Figure 11. Tariffs should recognize and reward favorable behavior.

Load Following vs. Energy Markets

When FERC introduced electric industry restructuring in 1996 with Order 888 it did not establish a load following service. Instead, load following has been provided by energy markets. Some load following needs (the morning load pickup and evening load reduction) can be forecast and at least partially addressed by day-ahead markets. Markets clearing at five minute intervals can certainly respond quickly enough to meet the remaining load following requirements. But do fast energy markets capture load following costs? This is an interesting and underappreciated question requiring further research that we will explore a bit here.

Note first that the minute-to-minute regulation balancing ancillary service is a capacity service. It is generation (or responsive load) capacity held in reserve for use by the system operator to respond to variations in aggregate system load and uncontrolled generation. It is not fundamentally an energy resource. Any net energy which comes out of or goes into a regulation resource is incidental and is paid for separately. Presumably load following would also be a capacity service with (slower) responsive reserves held back to enable the system operator to balance aggregate generation with aggregate load. Any net energy into or out of the load- following resource would, presumably, be incidental and settled separately. Alternatively, fast energy markets are fundamentally *energy* markets that require an incidental response (ramp) so that the unit is correctly positioned to provide energy for the transaction. This distinction between the basic commodity and the incidental response is at the center of the problem. Is load following basically a capacity service like regulation or is the required capacity incidental to the energy being provided like hourly energy?

Figure 12 presents a typical daily load curve with four classes of generators serving the load. Nearly 20,000 MW can be served from base load generators that can run continuously. The lowest cost generators will be selected as base load units in both the vertically integrated economic dispatch environment and in the market environment. The base load units do not need to have any maneuvering capability in order to successfully meet their energy obligations. Nuclear plants, for example, can meet this need.

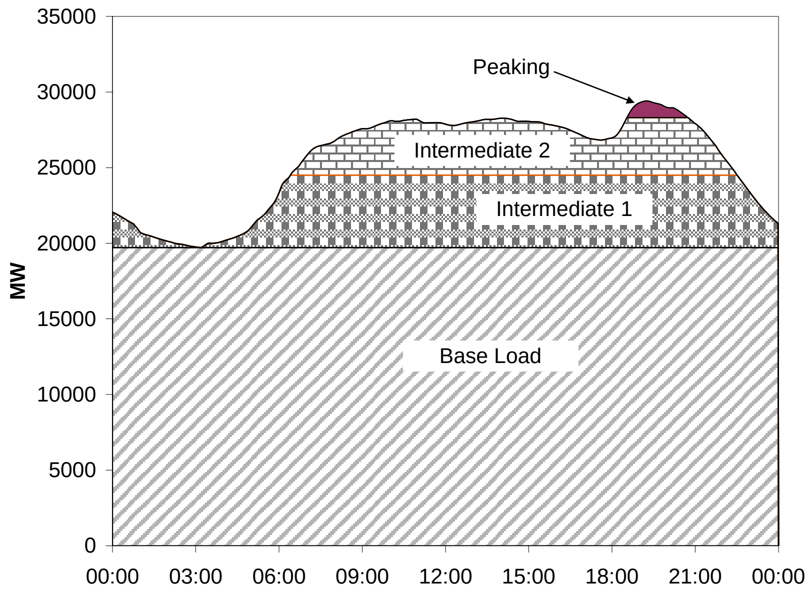

Figure 12. Participation in energy markets requires maneuverability for all but base load units.

Something interesting happens when the next generators are selected. Again this applies equally to both the vertically integrated and the market environments. Additional power is not needed all day long. In order to be selected to provide the next block of energy the intermediate units must be able to turn on for the hours when they are needed and off for the hours when they are not needed. It is probably a help to Intermediate Unit 1 that the requirement ramps up and down because the unit may not be able to turn on and off instantaneously. Once on, however, Intermediate Unit 1 has a flat output until it ramps off.

The requirements for Intermediate Unit 2 and the Peaking Unit are more interesting. They must have output flexibility simply to be in the energy market (or available for economic dispatch). The amount of output that will be required in any given hour depends on the overall system load which varies from hour-to-hour, day-to-day, season-to-season, and year-to-year. Regardless of the load following requirements, the last generators in the loading order, the most expensive units, must be flexible in the amount of power they can generate or they will be unable to successfully sell their energy output.

Let us belabor this point a bit. These last generators must be flexible concerning their output levels and run times not because of load following requirements but simply to be able to sell energy into a variable market. Flexibility would be needed even if the load was known in advance (it is still different from day to day). Flexibility would be needed even if the load made perfect step changes hourly (both run time and level would still vary). Inflexible units such as nuclear plants simply could not serve this part of the load or sell energy into this market. But a high capital cost, low operating cost resource like a nuclear plant probably could not economically survive in this market which has fewer operating hours than the base energy market.

The basic question of whether we need load following or fast energy markets can be looked at slightly differently now. Do the generators that are built to meet the intermediate and peaking energy markets (higher operating cost, lower capital cost) inherently have enough response capability to meet the system's load following needs? If so there is little point in creating or paying for a load following service. If the generators do not inherently have sufficient maneuvering capability then a load following service is required or the energy markets will be distorted.

A similar situation would exist for night-time contingency reserves. Contingency reserve prices are typically near zero at night because ample generation capacity is available waiting for the next day's peak energy need. If contingency reserves were only needed at night markets would not be created for them; contingency reserves would be a free byproduct of the energy markets. It is the on-peak requirement for contingency reserves, where the generators must forgo potential sales in the energy markets that make contingency reserve markets necessary.

We can hypothesize a system where ramping limits influence energy prices and a load following service seems necessary. Figure 13 shows a system with ample $10/MWh base load capacity. Unfortunately the base load units can only ramp at 1 MW/minute. When the load ramps from 2550 MW to 2850 MW in 30 minutes at 8:00 the base load units simply can not keep up. Peaking units costing $90/MWh (the only other generators in this example system) are required to serve load for five hours until the base load units can catch up. In a simple market with no load following service the *energy* price would jump from $10/MWh to $90/MWh for those five hours. An alternative would be to let the energy market clear at $10/MWh, purchase ramping capability (load following) from the fast responding generator, and also compensate the load following unit for the incidental energy it had to supply while following the load.

In Figure 12 and Figure 13 we see two different views of load following based upon the inherent capabilities of the generators that are trying to serve the energy market. If there are ample, reasonably flexible generators on the margin then a specific load following service is probably not justified. If a lack of ramping capability restricts which units can respond then energy markets will be distorted and a load following service would be beneficial.

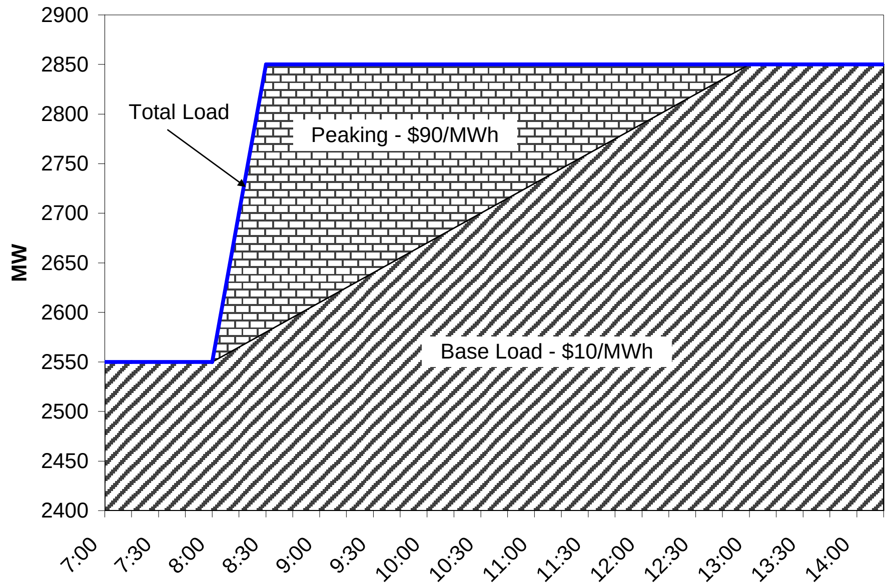

Figure 13. In this simple example load following is required from an expensive peaking generator but energy is only an incidental product.

It is difficult to obtain comprehensive public data concerning the ramping capabilities of generators within a balancing authority. To perform an initial scoping analysis we obtained one year of hourly data for total system load and individual thermal generator output for three balancing authorities (Kirby and Milligan, 2005). From this we calculated the total system ramping requirements for each hour. We also calculated the total ramping capability available in the direction of the load ramp from the currently on-line thermal generators. Thermal generation data is available because of emissions reporting requirements; data from hydro and nuclear units is not available. This data omission understates the system's available ramping capability. Figure 14 shows that significant excess ramping capability is typically available in all three balancing authorities in spite of the fact that hydro resources with significant ramping capabilities are excluded from the analysis. The same analysis showed that the load following requirements imposed by example wind plants was significantly smaller than that imposed by aggregate load.

In some cases there is significant existing ramping capability that is not currently available to the Balancing Authority system operators. This can happen if there is no market mechanism that can be readily utilized, whether the issue is to obtain ramping capability or energy in a very short time frame. If such capability is needed but not available, this can result in a deadweight loss to the system similar to what is shown in panel (b) of Figure 2.

These results tend to support FERC's ancillary service definitions which exclude load following and rely on fast energy markets. If, however, fast energy markets are found to exhibit times when prices are set by unit ramp rate constraints (rather than by simple availability to provide energy) then a load following service should be examined.

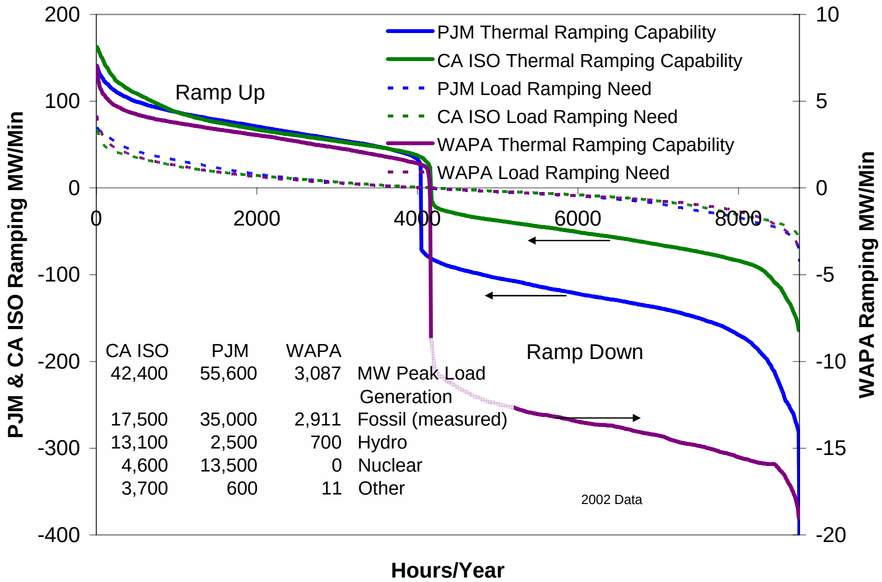

Figure 14. Preliminary analysis conducted with public data indicates that excess ramping capability typically exists in these three balancing authority areas.

A first solution should always be to assure that all physically available generating resources also have the opportunity to respond to real-time energy requirements. This will increase economic efficiency for loads and generators and mitigate the need for a separate load following service. If, however, fast energy markets do not provide for adequate ramping capability, and if this shortfall were to be rectified by a market for load following, then this would likely be a capacity service. In that regard, the characteristics of a capacity-based load following market would be similar to the way that regulation can be characterized.

Conclusions

In this paper we have proposed several principles that we believe are fair for all buyers and sellers of electricity. The basic goal is that the tariff should be based on cost causality, so that those who impose costs on the system bear those costs. Conversely, if an individual load or resource provides something of value, they should be paid. We develop several corollaries to these principles, along with a small number of tests that can help determine whether a proposed tariff is consistent with cost-causality principles. If the goal is to provide an economically efficient level of ancillary services (not only for wind, but for the entire system) then tariffs must provide the correct signals for buyers and sellers.

In the regulation time frame the requirements of individuals are not additive. This implies that a tariff that is intended to capture regulation impacts must carefully assess the physical characteristics of regulation, and that capturing variability does not recognize helpful variability vs. unhelpful variability. When the cost of unhelpful variability is assessed, it is important to distinguish it from the cost imposed by uncertainty. Tariffs that focus solely on uncertainty will not pick up the impact of variability, and will therefore not provide correct market signals. This also leads to a deadweight loss.

Based on a significant body of data from real wind power plants, we also show that scaling small wind plants or individual turbines to represent large wind plant output will overstate the variability of wind. If analyses based on this type of scaling are performed as a foundation for tariff development, then it is likely that the cost impact (and therefore tariff collection) will be too high. This also leads to a deadweight economic loss.

We also discuss load following and whether this service can be adequately provided by fast energy markets. More work needs to be done to separate the characteristics of this service and to determine whether and how markets should account for ramping and energy in this relatively short time frame.

References

Ernst, B.; Wan, Y.; Kirby, B. (1999), Short-term Power Fluctuation of Wind Turbines: Looking at Data from the German 250 Mw Measurement Program from the Ancillary Services Viewpoint, American Wind Energy Association Windpower ‘99 Conference, Washington, DC, June

Kirby, B.; Milligan, M. (2005). Method and Case Study for Estimating the Ramping Capability of a Control Area or Balancing Authority and Implications for Moderate or High Wind Penetration: Preprint. 19 pp.; NREL Report No. CP-500-38153. http://www.nrel.gov/docs/fy05osti/38153.pdf

Parsons, B., M. Milligan, J. Smith, E. DeMeo, B. Oakleaf, K. Wolf, M. Schuerger, R. Zavadil, M. Ahlstrom, D. Yen Nakafuji, *Grid Impacts of Wind Variability: Recent Assessments from a Variety of Utilities in the United States*. Presented at the Nordic Wind Power Conference, May 22-23, 2006. Espoo, Finland.