Energy Policy 73 (2014) 211–224

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Reliability payments to generation capacity in electricity markets

Fernando Olsina [a,b,*], Rolando Pringles [a,b], Carlos Larisson [b], Francisco Garcés [a,b]

[a] National Scientific and Technical Research Council (CONICET), Argentina
[b] Institute of Electrical Energy (IEE), National University of San Juan (UNSJ), Av. Lib. Gral. San Martín 1109(O), J5400ARL San Juan, Argentina

HIGHLIGHTS

- A new approach for remunerating supply reliability provided by generation units is proposed.
- The contribution of each generating unit to lessen power shortfalls is determined by simulations.
- Efficiency, fairness and incentive compatibility of the proposed reliability payment are assessed.

ARTICLE INFO

Article history:
Received 18 October 2013
Received in revised form
6 May 2014
Accepted 7 May 2014
Available online 6 June 2014

Keywords:
Adequacy
Supply reliability
Power generation
Price spikes
Investment
Regulation

ABSTRACT

Electric power is a critical input to modern economies. Generation adequacy and security of supply in power systems running under competition are currently topics of high concern for consumers, regulators and governments. In a market setting, generation investments and adequacy can only be achieved by an appropriate regulatory framework that sets efficient remuneration to power capacity. Theoretically, energy-only electricity markets are efficient and no additional mechanism is needed. Nonetheless, the energy-only market design suffers from serious drawbacks. Therefore, jointly with the evolution of electricity markets, many remunerating mechanisms for generation capacity have been proposed. Explicit capacity payment was the first remunerating approach implemented and perhaps still the most applied. However, this price-based regulation has been applied no without severe difficulties and criticism. In this paper, a new reliability payment mechanism is envisioned. Capacity of each generating unit is paid according to its effective contribution to overall system reliability. The proposed scheme has many attractive features and preserves the theoretical efficiency properties of energy-only markets. Fairness, incentive compatibility, market power mitigation and settlement rules are investigated in this work. The article also examines the requirements for system data and models in order to implement the proposed capacity mechanism. A numerical example on a real hydrothermal system serves for illustrating the practicability of the proposed approach and the resulting reliability payments to the generation units.

1. Introduction

Modern societies have developed a critical dependence on continuous delivery of electric power. Because of the vast impact —and potentially indeterminate reach—of power rationing events, supply reliability and generation adequacy (NERC, 1997) are matters of utmost concern for consumers and are deemed strategic by government bodies such as policymakers, regulatory authorities and agencies overseeing homeland security.

After restructuring of the electricity industry, centralized generation expansion planning has been replaced by decentralized investment decision-making following price (coordinating) signals and expectations on future returns. Capital allocation in generation capacity is now decided by multiple agents who aim at maximizing own profits while protecting themselves from risks. Since the very beginning of power markets, regulators and market designers have been reluctant to leave the market alone to warrant security of supply. Indeed, the pioneering electricity markets in Chile, UK and Argentina considered special mechanisms and provisions aimed at attracting timely investments in power capacity and sustaining supply reliability.

The rules governing electricity markets and their payment mechanisms should generate signals that produce efficient investments in terms of amount of installed capacity, mix of generation technology, and timing for being online. The power capacity that maximizes social welfare is regarded as the optimal adequacy level. However, determination of the optimal generation capacity

* Corresponding author at: Institute of Electrical Energy (IEE), National University of San Juan (UNSJ), Av. Lib. Gral. San Martín 1109(O), J5400ARL San Juan, Argentina. Tel.: +54 264 4226444; fax: +54 264 4210299.
E-mail address: olsina@iee.unsj.edu.ar (F. Olsina).

http://dx.doi.org/10.1016/j.enpol.2014.05.014
0301-4215/

is not an easy task as requires an accurate estimate of the Value of Lost Load (VOLL).

The heart of the current debate regards the proper regulatory framework that sets efficient remuneration to generation capacity. The selection of the right capacity mechanism is perhaps the most contentious issue in the design of electricity markets. As in other fields of economics, most regulatory proposals for capacity remuneration can be classified as either price-based or quantity-based. Sometimes, these dichotomous regulatory views are misrepresented as interventionist or market-friendly, when actually both approaches rest on administratively setting key design parameters. Currently, there is a lack of ample consensus regarding the superiority of some approaches with respect to others, as well as towards which mechanism is better suited for a particular organization in a given electricity market and the characteristic of its underlying power system. Unfortunately, the presently available theoretical and empirical evidence on these matters is at best spare and ambiguous. Under these circumstances, advocacy and opposition to the different approaches have often followed ideological lines of discussion.

Despite the weaknesses pointed out in the literature, price-based regulation to remunerate peak and reserve generation capacity by explicitly setting administrative payments is still one of the preferred schemes by regulators in many countries. This fact can probably be explained by the mechanism's success in addressing the challenge of attracting continuous capital investment flows in power generation to keep track of high load growth rates in fast expanding economies. Many markets also rely on explicit capacity payments in order to deliver proper supply reliability in hydro-dominated power systems, which are much more risky and challenging than thermal-only generation systems.

Most of the implementations of the capacity payment approach have shown a number of drawbacks. First, capacity payments are often fixed and do not reflect the prevailing adequacy of the generation system. Second, objective procedures for establishing the administrative value of the capacity price are generally missing or overly simplistic. Third, the capacity product to be exchanged for these payments is commonly loosely defined. In fact, the product is usually defined in terms of the generator´s “firm capacity”, which is normally estimated by means of very arguable procedures. Consequently, payments resulting from such methods often do not necessarily correlate with actual contributions of generating units to system reliability and its ability to deliver energy during scarcity. This leads to inefficiencies such as the misallocation of payments, the distortion of investment signals, and unfairness.

The aim of this paper is to present a methodological contribution to improve the way capacity payments are currently established in order to overcome most of its pitfalls. To this purpose, a reliability-based approach for determining the payments that should be awarded to each individual generating unit has been developed. The proposed reliability payments intend to reward the real contribution of each generation unit in regards to overall system reliability. The method presents attractive properties regarding efficiency and fairness.

The reminder of the article is organized as follows. Section 1.1 briefly revisits the most relevant issues about adequacy in electricity markets; Section 1.2 reviews each existing capacity mechanism aimed at supply adequacy and discusses in detail the shortcomings of several existing approaches for setting capacity payments, our main focus. Section 2 presents the proposed reliability payment, and includes data and models required for practical implementation of the new remuneration method. Detailed results of the proposed methodology for a real power system are illustrated in Section 3 jointly with an in-depth discussion of the policy implications. Section 4 closes with the conclusions.

1.1. Generation adequacy in electricity markets

It has been proven theoretically that electricity spot markets operating under perfect competition provide the right incentives to deliver optimal investments regarding capacity level and generation technology mix to supply power and energy demand at minimal cost (Stoft, 2002; Schweppe et al., 1988; Caramanis, 1982). Under these conditions, price spikes during rationing periods lead to scarcity revenues sufficient enough to attract the needed investments in peak generation (Olsina et al., 2006; Oren, 2000).

However, several problems in real settings must be considered. First, electricity demand is nearly inelastic and can lead to sharp albeit infrequent price spikes, which may be seen by both customers and regulators as a legitimate signal of system inadequacy. Second, scarcity creates favorable conditions for exercising market power in small or concentrated markets. It can be very difficult to distinguish legitimate high prices due to scarcity from those artificially elevated by the exercise of market power.

Although scarcity rents can be very significant, they are sporadic, erratic and unpredictable by their very nature. As a consequence of the extreme volatility of scarcity revenues, risk-averse investors delay or simply abandon investment plans in peak capacity necessary to guarantee long-term adequacy, causing a lower than optimal adequacy level and eventually capacity shortfall conditions. This situation can drastically deteriorate depending on if regulators administratively limit the market prices below the VOLL, e.g. by introducing price caps, or even if market participants believe the regulator would do it in the future, when scarcity events arise, in order to protect the demand from paying politically inacceptable prices. Because of the rather lengthy lead construction times, such situations may take considerable time to overcome.

1.2. Market designs for generation adequacy

The problems of finding the optimal production capacity and pricing of non-storable commodities, like electrical energy, have long been an important problem and have received extensive treatment in economic literature during decades. After the seminal article by Boiteux (1949), classical works on this topic, often referred in the literature as “peak load pricing”, were published (Kleindorfer and Fernando, 1993; Chao, 1983; Crew and Kleindorfer, 1976). Reviews of economic literature on different features and approaches to this problem can be found in Crew et al. (1995) and Joskow (1976).

The most important findings of these works show that under the hypothesis of risk neutrality and maximization of social welfare under uncertainty, the optimal generation capacity is that for which the marginal cost of an additional unit of capacity equals the expected marginal cost of the unserved energy. Although these results suggest that no further payment for capacity are needed, further research and empirical evidence show that some form of capacity remuneration is necessary in order to ensure that enough generation capacity be timely built. Several proposals for complementary payments for power capacity are reviewed by Batlle and Rodilla (2010), Baldick et al. (2005), Wen et al. (2004) and Vázquez et al. (2002).

Table I summarizes four different approaches to market design for generation adequacy, includes their relevant features and attributes, and mentions systems in which they have been applied. Comparative studies of different capacity mechanisms based on dynamic simulation models show that fixed capacity payments perform nearly as well as reliability options and long-term forward markets (de Vries and Heijnen, 2008).

However, implementing capacity payments is not exempt from notable difficulties. Here, a key issue is determining the payment level that yields the right capacity. If capacity price is settled too low, adequacy will deteriorate and conversely, if fixed too high, overcapacity will likely arise. Often, obscure (or plainly discretionary) procedures are used to determine payments and the capacity entitled to receive them. Indeed, there is a lack of rigorous methodologies for establishing the capacity prices based on sound technical and economic foundations.

Some implementations of this mechanism may alter the efficiency of short-term system operation and price signals. For instance, the payments introduced in Argentina were initially set too high and were only paid to dispatched units. Therefore, many generators bid below their marginal costs in order to be entitled to receive the capacity payment. This behavior distorted the efficient operation and did tend to depress the energy prices, enhancing the need for additional payments to ensure recovery of fixed costs. In other implementations, all self-declared available generators receive the payment. Unavailable generators can still realize the payment by declaring themselves available and bidding too high to avoid being dispatched. Measures like minimum annual time of operation, random availability tests, and penalty procedures need to be introduced in order to prevent gaming.

In addition, capacity payments are typically fixed and constant over time. The capacity price does not adjust to actual supply conditions in the system. An invariable capacity price leads to pay generators under situations of excess capacity, which is inefficient and may be politically unacceptable (de Vries and Heijnen, 2008). A dynamic capacity payment was introduced in the England and Wales pool during the period 1990–2001 (Chuang and Wu, 2000; Allan and Navarro Sanchez, 2004). Under the pool rules, generators were paid the expected energy price in each hourly time interval, which is a signal sensitive to prevailing system adequacy. All possible future states of the generation system are assembled into two mutually exclusive sets: states without power deficit, with cumulative probability $(1-LOLP)$, and capacity shortfall (complementary) states with occurrence probability $LOLP$, where LOLP stands for the “Loss of Load Probability”. For normal states, the system marginal energy price SMP results from the usual market clearing. For deficit states, the price escalates to the scarcity price given by VOLL. The expected price λ_t that gets a dispatched generator at time t is therefore given by:

$$\mathrm{E}[\lambda_t] = (1-LOLP_t)SMP_t + LOLP_t VOLL \tag{1}$$

By rearranging Eq. (1), the well-known price formula prevailing in the old British pool before entering into effect in 2001 the New Energy Trading Agreements (NETA) is recognized:

$$\mathrm{E}[\lambda_t] = SMP_t + LOLP_t(VOLL - SMP_t) \tag{2}$$

An available but not dispatched generator still receives a capacity payment equal to the second term of (1). This mechanism requires a reliable estimation of VOLL in order to send long-term efficient signals. This method is simple, transparent, and based on LOLP calculations that are well known and understood. Capacity payments reflect the real marginal value of generation capacity by considering variable system conditions. Nevertheless, in highly concentrated markets, as was indeed the case of the duopoly in the UK market at the time, this approach is prone to suffer serious manipulation by generators. Under tight reserve margins, often some units can artificially be declared unavailable so that the computed LOLP is illegitimately inflated by a large extent and higher capacity payments are captured by all remaining available units. Finally, the question persists as to whether or not a capacity price that changes on an hourly basis delivers a better investment signal than the spot price for electrical energy.

2. Methodology

2.1. Desirable properties of a capacity payment scheme

The design of a well-functioning capacity payment mechanism should conform to efficiency criteria and comply with the following desirable features:

- Instead of paying for nameplate capacity, payments to generating units should remunerate its ability to lessen the occurrence of power shortfalls. The product in exchange of payments should be system reliability and not its proxy, e.g. the unit's installed capacity or the “firm capacity”.
- Remuneration to generation capacity should reflect the prevailing supply adequacy and should automatically adjust to the actual status of the power system.
- Economic signals provided by capacity payments should deliver the optimal amount of power capacity and lead to the least-cost technology mix in the long-run.
- The energy-only spot market is a design that ensures economic efficiency in short-term operation and in long-run investments. To preserve efficiency, any other alternative market design should not alter economic signals regarding the energy-only market. Hence, expected revenues accrued by generators and consumer payments must be kept unaltered when introducing new rules for remunerating capacity.
- The introduced remuneration rules should not distort the bidding behavior to energy markets in order to prevent inefficient clearing prices, which deteriorate efficiency of operation, consumption and investments decisions.
- Uncertainty and volatility of the revenue stream in peaking units should drastically be reduced in order to lessen risk-aversion by investors.
- The mechanism should be fair. Payments to each unit must reflect its effective contribution to system reliability. Primary resource uncertainty, location, unit size, operating flexibility, maintenance policies and unit availability should be properly accounted for in payment determinations.
- Rules for establishing reliability payments should be objective, well-founded in both economic theory and engineering criteria, without introducing arbitrary or discretionary considerations.

2.2. Proposed reliability payment

Under perfect competition, rational expectations and risk-neutrality, scarcity rents obtained by generators in energy-only markets during power deficits yield the long-run optimal capacity level and efficient generation mix (Schweppe et al., 1988; Caramanis, 1982). In long-run equilibrium conditions, inframarginal rents obtained by each generating unit during capacity shortfalls are the exact amount needed to cover all costs, including investment fixed expenses. This result holds even under uncertainty and price-inelastic demand (Chao, 1983).

During a capacity shortage, scarcity rents are captured by not only peaking units, but also by all available units supplying power to the system, provided sufficient transmission capacity is available. Under locational marginal pricing, only transmission-unconstrained generators would be paid the scarcity rents during the power shortfall, according to their marginal contribution to alleviate the power deficit. Scarcity rents collected by generators clearly reflect the adequacy of the generation system to meet the locational power and energy consumption, and thereby the need for capacity additions. These rents are a reliable signal for the entry of new generation investments.

Besides providing the right long-term investment signals, energy-only markets are inherently fair and efficient with respect to remunerate the effective contribution of each generating unit to supply reliability. On average, lower scarcity rents would be captured by generators as result of several factors: units with poor forced outage records (running and startup failures); extended or improper maintenance scheduling; long start times and slow ramping; lengthy minimum downtimes; and/or units located in unreliable or congested areas of the transmission grid. The same effect applies for

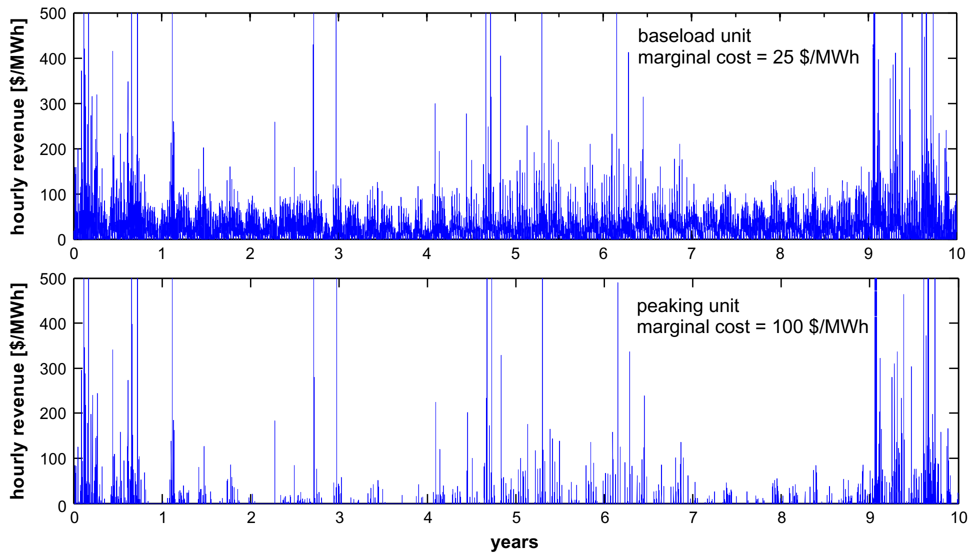

Fig. 1. Simulated 10-year hourly revenues earned in the energy market by a baseload and a peaking unit according to spot price dynamics observed on the German EEX.

generators relying on uncertain renewable primary resources (e.g. hydro, wind, solar, etc.). As capacity shortfalls and size of generation outages are correlated[1], larger power plants would contribute less to system reliability than smaller units on a per megawatt basis. Failure of large units or unavailability of renewable generators would be more likely involved in the causes of a power deficit. Therefore, they have lower probability of supplying energy during a power shortfall in order to collect the scarcity rents.

The main drawback of leaving capacity remunerated in energy-only spot markets is the extremely volatile and sporadic nature of scarcity rents. Because of the high number of intervening factors, and the random nature of some events (e.g. failure of system components), the occurrence time and duration of power deficits are exceedingly difficult to estimate with any degree of practical accuracy. Revenues earned during shortfall events can vary dramatically from one year to the next. This volatility has particularly severe implications for peaking and reserve units, as they strongly rely on scarcity rents to cover investment costs[2].

Fig. 1 illustrates the hourly revenues per MW that base and peak load units would earn in a 10-year period. Noticeably, the annual volatility of revenue flows for peaking units is remarkably elevated. Fig. 2 depicts the inter-annual variability of the yearly gross profits with respect to the long-term mean as a function of the unit's marginal generation cost. Annual revenue variability is six-fold higher for a peaking unit with a cost of 100 $/MWh than for a base load generator with a marginal cost of 25 $/MWh.

Even though these financial flows may be sufficient on average to cover fixed costs entirely, investment in new peaking units can hardly get funded under these circumstances. Shrinking investments and a bias towards less capital-intensive (and likely less-efficient) generating technologies will result in consequence of the higher risk premium required in peaking generation projects. If investors are not risk neutral, as indeed they are, the unpredictability of revenues

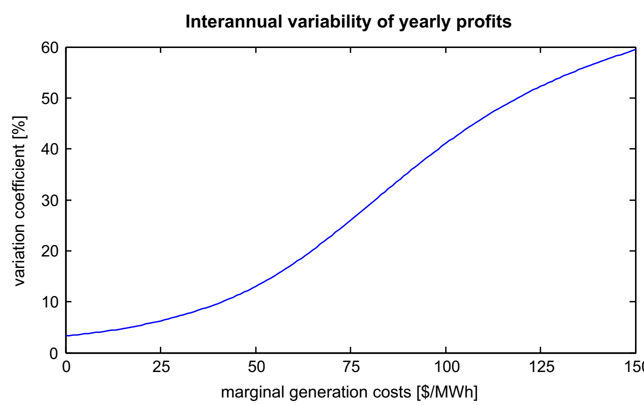

Fig. 2. Coefficient of variation, σ/μ (standard deviation-to-mean ratio) of annual gross profits computed for generators with increasing marginal cost of production. Sample gross margins were computed for a sample of 1000 annual realizations of hourly EEX spot energy prices simulated according to a spectral representation algorithm (Olsina and Weber, 2009).

originated in scarcity rents becomes the main obstacle for achieving the right generation capacity in energy-only market designs. Nevertheless, the theoretical results of the energy-only spot markets still constitute the efficiency benchmark.

Based on these observations and the premises developed in Section 2.1, we suggest replacing the particular annual realization of the stochastic revenue stream earned by each generating unit during scarcity in an energy only-market with its certainty risk-neutral equivalent. By exchanging the highly uncertain annual revenues during rationing with a fully-certain amount equal to its mathematical expectation, the related financial risk can be eliminated, while simultaneously preserving efficiencyof the energy-only design. By suppressing uncertainty in scarcity revenues, risk-averse investors should behave identically as if they were risk-neutral.

Let R_y^k be the annual cumulated revenue obtained by the k-th generation unit of the system during load shedding events occurring in year y. If the probability density function of these annual incomes can be estimated, denoted as $f(R_y^k)$, the annual reliability payment P_y^k for unit k can be computed as follows:

$$P_y^k = \mathrm{E}[R_y^k] = \int_0^\infty R_y^k f(R_y^k) dR_y^k \quad (3)$$

[1] Besides reducing reserve margins, sudden failures of large power stations or main tie lines often lead to load curtailment actions necessary to keep an instantaneous power balance and to prevent propagation through cascading outages.

[2] Typically, collected scarcity rents have negligible influence on the economic signals that govern investment decisions in baseload power plants. Indeed, price spike revenues represent only a small fraction of the gross sales by these generators in the energy market.

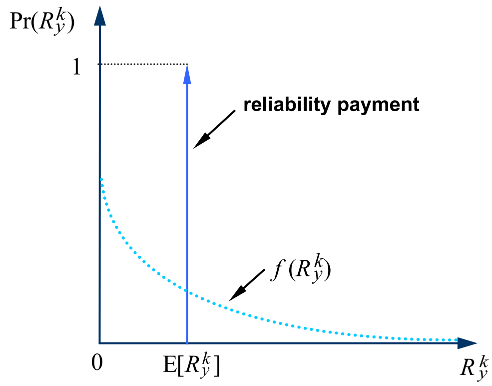

Fig. 3. Probability density functions of the annual scarcity rent and the proposed reliability payment.

Since the probability of a loss of load event is always positive, i.e. $LOLP > 0$ holds irrespective of the excess capacity in the system, the payment P_y^k is always constitutes positive cash flow, i.e. $P_y^k > 0$. Reliability payments change according to the prevailing reserve margin in the system. If the system has installed far more capacity than needed, generators would be paid a small amount, as capacity under these conditions still has some value. Contrarily, if system adequacy deteriorates, the reliability payments may rise considerably. Under equilibrium capacity and optimal technology mix, this payment is the exact amount necessary to cover fuel and investment fixed costs (Olsina et al., 2006; Stoft, 2002; Oren, 2000).

Fig. 3 schematically depicts the probability density function (PDF) for the annual scarcity revenues that a generator would accrue in an energy-only market design (dotted line). Note that this distribution is highly right-skewed and that scarcity revenues cannot be negative. Fig. 3 also shows the substitution of this probability density by the certainty-equivalent payment under the proposed reliability mechanism. For a guaranteed annual reliability payment $E[R_y^k]$, the corresponding probability density is a Dirac delta function with integral value $\Pr = 1$, indicating full certainty of occurrence[3].

Since this assured annual cash flow is by definition identical to the expected annual scarcity rent the generator would capture in the energy-only market, the efficiency of the economic signal is not altered. This fact preserves the long-term optimality of generation investments.

The per megawatt expected scarcity rent computed for each generating unit will yield different values according to technology, unit size, network location, outage rate, operating flexibility, scheduled maintenance, etc.—all reflecting the actual contribution of this capacity to the overall supply reliability. In fact, generating capacity is not a homogenous product, but only a proxy of the unit's contribution to system reliability[4]. Therefore, capacity itself cannot be priced uniformly. To price the traded product uniformly, in most implementations the nameplate capacity of generators is somewhat derated by ad hoc (and often rather obscure and discretionary) procedures to characterize the unit's "firm capacity". The proposed reliability payment is quite different from other capacity payments, which typically establish a uniform rate for all capacity, irrespective of their characteristics.

[3] In probability theory and statistics, Dirac delta functions are widely used to represent discrete probability density functions as is the case with the reliability payment. Dirac delta functions are conventionally represented by an arrow.

[4] Perhaps the notion of effective load carrying capability (ELCC) of the generating unit is a better definition for the reliability product provided by the power capacity (Garver, 1966).

Furthermore, the proposed reliability payments to generators completely substitute the individual annual realizations of the scarcity rents. Therefore, they are not supplementary to revenues collected in the energy market. This warrants that bidding behavior and short-term operating decisions remain efficient. These provisions avoid distortions such as the introduced in the Argentine wholesale power market, as generation companies had incentive to underbid in the energy market to get dispatched and become entitled to receive the capacity payment (Batlle and Rodilla, 2010).

2.3. Incentive compatibility and settlement

In order to set the corresponding reliability payments on a yearly basis, the ISO numerically computes the mathematical expectation of the annual revenues that each unit of the system would collect under deficit conditions over all future states. The proper models and data for accomplishing this task are discussed in the following section.

Under the proposed mechanism, generators participate as customary in the energy spot market and, if available, will capture price spikes during periods of rationing. This ensures incentive compatibility in order to keep unit availability as high as possible when most necessary. A generator has further incentives for keeping unit availability high. A generation company who strategically and systematically declares a unit unavailable will henceforth deteriorate its own availability statistics that are used in turn to compute the payment. Following a manipulation, the reliability payment the unit is entitled to receive in all subsequent periods after year y decreases[5]. For this reason, the proposed mechanism does not suffer from the shortcomings of the payment approach introduced in UK markets, where dominating generators withhold capacity for increasing the hourly *LOLP* on a daily basis.

In energy-only markets, when systems experience tight reserve margins, generators may exercise market power by economic or physical withholding. Since the reliability payment preserves the efficient investment signal entirely, market prices can be capped at a value substantially lower than the VOLL to curb economic withholding without endangering long-term incentives. At the same time, physical withholding occurs because companies with large capacity can retain a few MWs when reserve is tight to force clearing prices to escalate to the scarcity price (i.e. VOLL) and thereby collect the scarcity revenues for the remaining generators. The proposed mechanism mitigates the exercise of market power, making this strategy unprofitable. In fact, any scarcity revenues (legitimate or not) captured in the energy market by generators are accounted for in the final settlement of the reliability payment, as explained in the next paragraph.

If the particular yearly realization of the system yields less power deficit hours than average, the ISO at the end of the period will pay each unit the difference between the expected annual scarcity revenue $E[R_y^k]$ and the actual cumulated revenues R_y^k realized during rationing events. On the contrary, if in the course of a year the cumulated duration of the realized power shortfalls exceeds its expected value, generators reimburse the excess collected to the ISO at the end of the year, i.e. $R_y^k - E[R_y^k]$[6].

On the other side of the market we have the consumers. Demand plays into spot markets and will face price spikes during

[5] The inherent incentive to keep availability statistics as high as possible does not preclude the introduction of other common mechanisms to prevent capacity withholding, e.g. random availability tests, penalties for unavailability, etc.

[6] This settlement rule exposes the ISO to credit/counterparty risk, as some generators may enter into bankruptcy before refunding the obligation. However, as collected scarcity rents in excess of the reliability payment are relatively small and infrequent — and credit risk is diversified among all generators — the risk borne by the ISO on behalf of the demand is fairly low.

scarcity periods. The effectively served demand will pay generators the scarcity rents at times of capacity shortfalls. If the particular realization of the year results in lower deficit hours than average, the scarcity revenues collected by generators will be $R_y^k \le \mathrm{E}[R_y^k]$. At the end of the period, the demand will be charged an uplift to pay capacity, which is then allocated among generators according to the above settlement rules. Similar to the generators, this uplift will be allocated among consumers according to the expected annual payments each load must make during rationing periods. To determine the expected capacity charges, location of the load in the transmission grid, automatic load shedding schemes, etc. should all be taken into consideration when identifying those loads that benefit the installed generation capacity the most. At the end of a year, if the scarcity rents paid by the load in the energy market exceed the expected annual payment, the ISO can refund the load or can take any excess charged as credit towards paying next year's obligation.

2.4. Requirements for system data and models

Accurate estimation of expected scarcity revenues collected annually by each generation unit is not a trivial task, as it requires sophisticated simulation models and extensive system data (e.g. outage records; maintenance scheduling; unit size, fuel and technology; uncertainty on power demand and hydrological inflows, etc.) in order to reproduce the stochastic operation of the whole power system.

Even though the data and model requirement may at first appear overwhelming and costly, it is not generally the case, and the implementation of the methodology should not add much incremental effort or cost to the ISO. Indeed, one of the main tasks of ISOs is the periodical evaluation of overall reliability of the power system under its administration. To comply with this responsibility, ISOs employ stochastic simulation models of the system operation to evaluate the occurrence probability of loss of load events. These simulative models and the related input data are essentially identical to that which is necessary to calculate reliability payments to each generating unit. Hence, only minor adaptions of the models used by ISO are required for calculating the expected value of generator revenues during power shortfall conditions. Carrying out the necessary changes in these models is relatively simple and inexpensive. Furthermore, the involved computation effort is quite low as these models typically run on conventional desktop computers and results are obtained in a few hours, even for large-scale power systems.

The entire procedure for setting the reliability payments is explained as follows. A flowchart diagram of this process is illustrated in Fig. 4. On a yearly basis and ex-ante, the ISO generates an ensemble of synthetic realizations of hourly power loads, river flows and random failures of generating units applying proper stochastic models. Based on this information and system data, the ISO also computes, ex-ante, the time-varying expected future cost functions (FCF) of each water reservoir by means of a stochastic mid-term hydrothermal planning model. The FCF of a given reservoir not only depends on its own stored volume, but also on the status of all remaining system reservoirs. In the case of the present example, FCFs are computed with weekly resolution and are linearly interpolated to obtain hourly values.

Hourly optimal operation of the power system is detailed simulated for each annual realization from a representative ensemble of the driving stochastic variables (outages, load, water inflows, etc.). In the stochastic simulation model, the expected marginal water value is used as coordination variable to economically dispatch hydro units jointly with the thermal generators. For the prevailing FCF in each hourly interval, marginal water values of reservoirs are endogenously computed in the stochastic

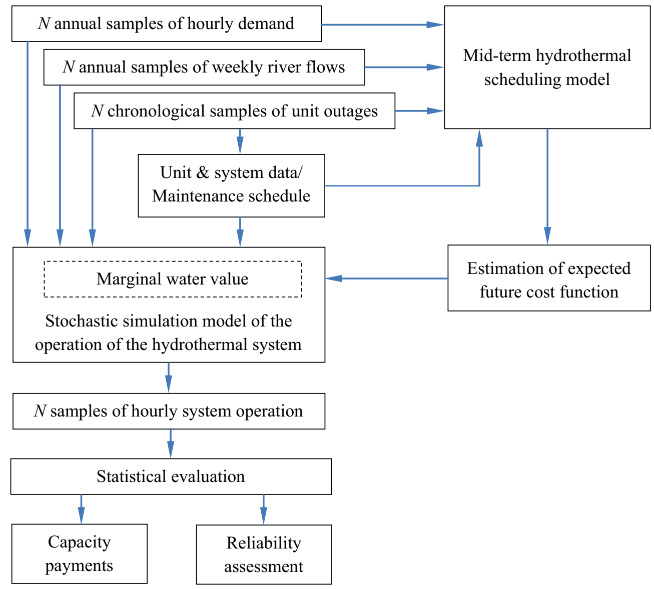

Fig. 4. Flowchart of the overall procedure for computing reliability payments.

simulation model, as they depend on reservoir levels and natural inflow, as well as availability state and binding constraints on hydro units downstream.

For each sample realization of the hourly operation, intervals with power shortfalls are identified to calculate the revenues obtained by each unit at those times. The scarcity revenue collected by the generator is computed as the power delivered by the unit (if available) at the intervals with power deficit multiplied by the prevailing market price, which is administratively set at VOLL. By repeatedly simulating hourly system operation for a year under a massive number of different realizations of the exogenous stochastic variables (Monte Carlo), a sample of the system states with power deficit can be obtained. The expected value of the collected scarcity revenues by each individual generation unit during the year is the reliability payment entitled to each generator. The expected value can be statistically estimated by averaging the annual collected scarcity revenues accrued by each generator over all simulated realizations.

For a large enough ensemble of possible, hourly sample realizations of the system states, the probability distribution of the annual scarcity rents captured by each generator can be estimated. The computation of the reliability payment in year y awarded to each generating unit k can be numerically estimated from the simulated sample of scarcity revenues as

$$P_y^k = \hat{\mathrm{E}}[R_y^k] = \frac{1}{N}\sum_{i=1}^{N} R_y^{k(i)} \tag{4}$$

where N is the sample size of simulated hourly trajectories of system operation, and $R_y^{k(i)}$ is the sample scarcity revenue collected by the unit k along the i-th realization. Eq. (4) represents the numerical estimation of the integral expression in Eq. (3) by means of Monte Carlo sampling. It can be demonstrated that the sample mean is an unbiased and efficient estimator of the expected value of a probability distribution (Fisher, 1990). Sampling must conform to random criteria to preserve statistical independence. The sample size must be large enough to ensure statistical convergence of the estimation. For a large enough sample size, the error of the estimation is lower than a prespecified threshold for a given confidence level (Hahn, 1972).

Sample realizations of stochastic variables, e.g. random operation and failure times of generating units, hourly power demand fluctuations, as well as water inflows and wind speed time series, are required as input data for the stochastic simulation model of the system operation.

Random fluctuations in power demand are important for simulating power shortfalls. A stochastic chronological hourly load model that incorporates long-term (uncertain) demand growth (drift), seasonality (daily, weekly and yearly regularities) and short-term random fluctuations, e.g. weather-driven shocks, is required (Breipohl et al., 1992).

Of further need is a detailed reliability model of each generating unit and transmission components (e.g. lines, transformers) to simulate the dynamics of the available generation capacity. Chronological multi-state reliability models for generators are preferred as they account for the cycling operating characteristics of each unit according to their production variable costs, as well as unit states with derated capacity, postponable forced outages, startup failures, operating inflexibilities, etc. (Billinton and Allan, 1996). This allows better estimation of each unit's ability to supply power during deficit episodes. For our purposes here, we have selected a chronological four-state model to describe the stochastic behavior of unit failures (Billinton and Ge, 2004; IEEE, 1972).

Calculation results of reliability models cannot be better than the data used to describe the reliability of the system components. Hence, reliability data that is to be used must represent, as accurately as possible, the individual failure behavior of each unit. Different generator availability performance leads to differential reliability payments. Hence, outage data must be based on extensive observations of each unit in order to have representative and individual statistics. For instance, collecting enough outage statistics may require several years for a peaking unit that operates only a few hours per year. Then, for new units entering the system, only availability data from manufacturers, or outage statistics of similar units with records of multiple years in operation, may be used as proxy while actual statistics are being collected. In any case, new outage data gathered in the real operation of units should be added to the sample used to calculate its reliability parameters. This warrants that the used reliability data reflects the actual failure behavior of the system components as accurately as possible.

If hydropower and/or wind power has a relevant contribution to generation, multivariate stochastic models for water inflows at the different rivers (Koutsoyiannis, 2000; Yevjevich, 1987) and/or wind speed at the wind farm locations (Olsina, 2013; Olsina et al., 2007b; Billinton et al., 1996) are also required in order to account for the uncertain availability of the aforementioned renewable resources.

Additionally, it is necessary to couple the stochastic demand model, the renewable resource model, and the reliability model of generation units with an optimization model that replicates the least-cost operating policy of each water reservoir and the optimal unit commitment and dispatch of the thermal system. For such a purpose, variable generation costs based on fuel prices and thermal efficiency of each fuel-fired unit must be estimated. The ISO can estimate this information from generator bids to the energy market. By assuming linear variable generation cost functions, i.e. constant marginal cost of production, efficient algorithms based on *priority list* techniques may be used to estimate the optimal unit dispatch (Senjyu et al., 2003). These methods provide good dispatch solutions while drastically reducing computation effort in the context of stochastic simulations.

In hydro-dominated generation systems, the probability of rationing events strongly depends on the volume stored in water reservoirs, which are themselves contingent on present and past water inflows, as well as prior operating decisions made on hydro units. This problem is exceedingly complex if the system has coupled (cascading) hydro reservoirs, as operating decisions on an upstream reservoir influence the state of other reservoirs and/or constrains hydro units downstream. Embedding the large-scale problem of optimal hydro scheduling in the framework of a stochastic hourly simulation is only possible by using dual information on the reservoirs at each time step. By using the marginal opportunity cost of the water contained in each reservoir (the so-called "water value") as coordination variable (Reneses et al., 2006), stochastic simulations of the hourly joint dispatch of thermal and hydro units can be performed.

Literature on chronologic reliability models of hydrothermal systems is quite spare. Simple stochastic simulation models of the chronologic operation used to assess the reliability of hydrothermal systems were first proposed in Ubeda and Allan (1994) and Greco (2000). These models contained many important simplifications, as well as some unrealistic assumptions. In order to properly address the challenges of this application, a detailed stochastic model for simulating the optimal operation of hydrothermal systems has been developed and is extensively described in Olsina et al. (2007a).

Lastly, to compute reliability payments, the ISO must be informed of the scheduled maintenance for all generating units and transmission components. Additionally, information regarding new capacity commission and power plant retirement during the considered year is also required. Topological changes and expansions of the transmission systems must also be considered. ISOs routinely collect and have most of these data and information available to properly operate the power system and monitor its reliability level.

3. Results and discussion

To illustrate the practicability of the proposed reliability payment mechanism under the complexities present in a real setting, let us use the hydrothermal generation system that supplies electricity to the power market of El Salvador as an example. This interconnected system has an installed generation capacity of 1254.28 MW, and is comprised of 69 thermal units of different technologies, fuels, and sizes, as well as 4 hydro power plants (HPP). Basic data of the thermal units are given in Table 2.

Table 2 also provides basic reliability parameters for the thermal units, i.e. steady-state failure probability $\Pr(F)$, mean time to failure $\mathrm{E}(T_O)$ and mean time to repair $\mathrm{E}(T_F)$. From these parameters, the failure rate λ and repair rate μ of each generation unit can be easily derived as $\lambda = \mathrm{E}(T_O)^{-1}$ and $\mu = \mathrm{E}(T_F)^{-1}$ respectively. As is customary in reliability studies, it is assumed that unit residence time in each state is exponentially distributed, according to parameters λ and μ. The stochastic failure behavior of each unit is represented with a 4-state Markovian reliability model (Billinton and Ge, 2004; IEEE, 1972). This more complex model is essential for capturing the behavior of cycling generators, as peaking and reserve units are generally down and the probability of a failure in these states is very low. This results in a very high availability of these units when requested to generate in order to prevent or mitigate a power shortfall. For the sake of simplicity, we have neglected unit failures during downtimes and startups.

Hydro capacity represents 36% of the total installed generation capacity, which substantially impacts the optimal operation of the system and the reliability of supply. The hydropower generation system encompasses four coupled hydro power plants located along the Lerma River. Two controllable upstream reservoirs of very different storage volume allow for the management of water resources to minimize the expected annual generation costs of thermal units. Production coefficients of each HPP are non-linear functions of the stored volume in the associated reservoir. Fourth-order polynomials have been fitted to the empirical power-discharge data of each HPP. The two remaining power plants are

Table 1
Market designs for generation adequacy.

Approach/type	Description and hypothesis	Pros	Cons	Implementation
Energy-only markets	Under perfect competition, rational expectations and risk neutrality, energy market prices are efficient and lead to long-term least-cost generation system Stoft (2002), Caramanis (1982).	Theoretical efficiency in short-term operation and long-term investments. Simplicity: no additional payments for capacity needed	Design vulnerable to market power Need reliable estimation of VOLL. Supply adequacy provisions are often still required Batlle and Rodilla (2010).	ERCOT (Texas), NEM (Australia), Alberta (Canada), European countries, Scandinavian Nordpool.
		Under risk-aversion, shifting to less-capital intensive investments Neuhoff and de Vries (2004)	Design prone to suffer investment cycles Arango and Larsen (2011), Ford, (1999), Olsina et al. (2006)	
Capacity markets (quantity-based)	ISO administratively determines the required capacity including a reserve margin. Each load serving entity (LSE) has to back-up its requirements with capacity credits purchased in the capacity market.	Simplicity. ISO sets the required capacity or establish a price-capacity demand curve, determining capacity prices in an annual auction.	Setting the right capacity amount is an interventionist measure, at a large extent arbitrary. Failed to guarantee the recovery of investment costs and led to extreme capacity price volatility Battle and Rodilla (2010).	Former ICAP markets in northeastern USA (e.g. PJM).
Long-term forward contracts/ Reliability options (quantity-based)	LSEs must hold long-term contracts or options for energy backed up by physical generation assets. Forward reserve markets with the obligation to offer energy in the day ahead and real time markets for a price not less than a given value in order to guarantee that these resources only seldom are called to supply energy Vázquez et al. (2002).	Reserve acquisition in a long-term market is effective for reducing volatility of spot prices. Deemed to be a market-friendly approach to secure capacity adequacy.	The strike price for exercising the option must be administratively set. Participants may have difficulties to value the options when bidding. Penalties are needed in the case the generator defaults the delivery obligation.	Proposal for California Baldick et al. (2005), Vázquez et al. (2002) New England reserve market Cramton and Stoft (2008).
Capacity payments (price-based)	Fixed payments for the installed generation capacity are administratively established. Payments are additional to revenues generators obtain in the energy spot market.	Capacity payments have proved to be effective for timely attracting enough investments, dampen construction cycles, keep reliability and reduce price volatility.	Lack of rigorous methodologies for awarding capacity payments based on sound technical and economic foundations. Payments are typically fixed and constant over time, not adjusted to the actual supply adequacy.	Colombia, Argentina, Bolivia, Chile, El Salvador, Panama, Peru. Formerly also Colombia and Ecuador. Spain, Italy, South Korea, the former UK market
		Energy markets can safely be capped without damaging the investment signal	Payments may alter bidding behavior in the energy market. Provisions for measuring actual availability are often needed.	

Table 2
Data of the thermal generation system of El Salvador.

Technology	Fuel	Units	Capacity [MW]	Pr(*F*)	E(T_O) [h]	E(T_F) [h]
ICE	Diesel	15	83.74	0.02	980	20
ICE	Bunker C	40	363.73	0.02	980	20
Geo	–	7	171.80	0.015	985	15
ST	HFO	2	59.72	0.02	735	15
ST	Biomass	4	56.88	0.04	960	40
GT	Diesel	1	64.55	0.02	1470	30
		69	**800.42**			

ICE: Internal Combustion Engine; **HFO**: Heavy Fuel Oil;
ST: Steam unit; **GT**: Gas turbine; **Geo**: Geothermal.

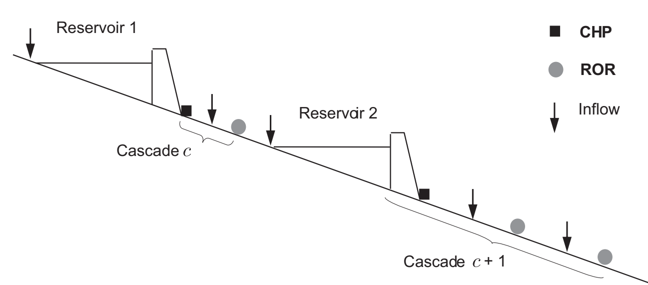

Fig. 5. Topological arrangement of the cascading hydropower system.

sited downstream. These are run-of-river power plants. The production of these generators depends mainly on the operating decisions of the upstream reservoirs. The topology of the hydro system is schematically depicted in Fig. 5. Basic data of the hydro units and the system reservoirs are given in Table 3.

The stochastic hourly load model considers uncertainty in the underlying consumption growth process (drift), a variety of deterministic patterns (daily, weekly and seasonal cycles; holidays and other calendar effects), as well as random fluctuations and shocks. The expected peak load of the system is 964.32 MW and the standard deviation is 13.6 MW. The probability distribution of the peak load departs from normality as it is right-skewed. An hourly load model for El Salvador was developed based on the historical records. The underlying energy consumption growth was identified by means an ARIMA model. Typical daily cycles, as well as weekly and seasonal regular patterns, were estimated from historical datasets as well. The effect of specific calendar days (New Year, Christmas, Easter, national holidays, etc.) was also considered. Random fluctuations in hourly power demand was represented by a multivariate ARMA model, in which each hour of the day is a separate random variable statistically dependent of previous intervals. This multivariate representation allows for

Table 3
Data of the hydropower generation system of El Salvador.

Type	V^R_{min}[hm^3]	V^R_{max}[hm^3]	V^R_0 [hm^3]	P_{max} [MW]	Units	q_{min}[m^3/s]	q_{max}[m^3/s]	κ [MW/m^3/s]
CHP	423	875	443	21	1	0	43	0.451
CHP	687	2042	859	171	2	0	255	0.672
ROR	94	94	94	92	5	0	231	0.398
ROR	307	307	307	170	2	0	590	0.288

CHP: Controllable hydro power plant; **ROR**: Run-of river power plant.
V: reservoir storage volume, *R*: reservoir, *P*: power output, *q*: water flow; κ: mean production coefficient.

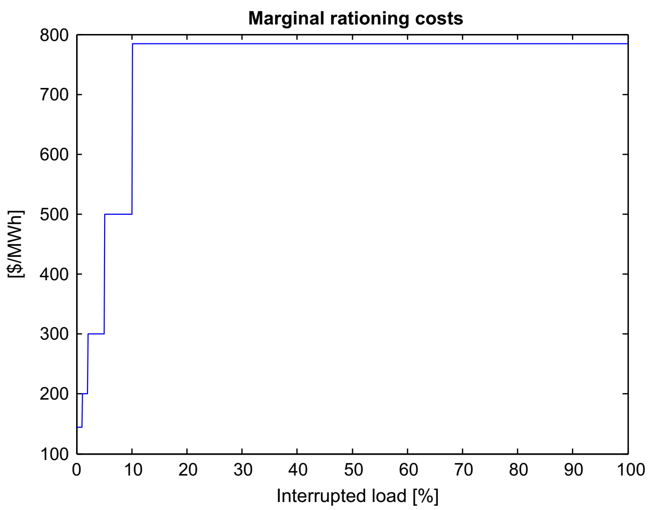

Fig. 6. Step-wise function of the marginal load rationing cost.

capturing of the observed covariance structure, as well as hourly changing (non-stationary) standard deviations.

Under expected peak load conditions, the generation gross reserve margin is 30.07%[7]. The Salvadoran ISO prescribes that a power capacity equal to 3% of the prevailing demand must be kept as spinning reserve at all times. Consumption has been regarded as price-inelastic in the short-term. Involuntary load curtailments are marginally priced according to an increasing function as illustrated in Fig. 6. This step-wise marginal value of the interrupted load is imposed by the regulator of El Salvador (SIGET) based on outage cost studies. This function is applied by the ISO to administratively set the market price when the market is unable to clear due to rationing being required to restore power balance.

The natural water inflows are variables subjected to considerable uncertainty and can dramatically vary from year to year. In hydrothermal-dominated systems, inflows have a decisive effect on supply reliability and risk of rationing. Water flows on geographically nearby rivers normally presents significant spatial cross-correlation, as they are subjected to a similar rainfall regime. Water flows on the same river also exhibit long-lasting autocorrelation structures extending months or even years. In addition, large-scale phenomena, like El Niño, have strong influence on climate and water availability in rivers of Central and South America. Such features and patterns must be accounted for in the stochastic inflow models when assessing the probability of rationing events since they have simultaneous impact on all regional basins and introduce a common-cause failure mode. To model natural water inflows, we rest on the multivariate ARMA model included in the SDDP software package for mid-term hydrothermal scheduling (PSR Inc., 2006). This tool identifies the optimal model order and fit model parameters to the 40 available years of observational hydrological records of the involved rivers.

The ISO determines the optimal mid-term hydrothermal policy based on system data, prevalent uncertainties, scheduled maintenance of equipment and system topological changes (e.g. generation upgrades and capacity retirements). The mid-term hydrothermal coordination problem is solved by stochastic dual dynamic programing (SDDP) (PSR Inc., 2006; Gorenstin et al., 1992).

For each system reservoir and weekly stage, expected future cost functions can be obtained from the optimal hydro-scheduling problem. SDDP is a hydrothermal planning tool for mid-term hydrothermal coordination that is well proven and widely used by ISOs in Latin America and other regions with dominant hydro share (PSR Inc., 2009). It can handle hundreds of controllable hydro power plants with several hundred natural inflows, and without limitations on the hydro topology.

The dual information delivered by the mid-term optimal hydro-scheduling can be further exploited as coordination variable to massively simulate the hourly dispatch and operating conditions of the power system. The marginal opportunity cost of releasing water from reservoirs is determined hourly in the stochastic simulation model. This marginal cost can be used to economically dispatch hydro units in the context of a priority list method. The hourly optimal operation of the hydrothermal system has been simulated for 1000 possible annual realizations of the stochastic driving variables (outages, load fluctuations and water inflows). This sample size entails solving 8.76 10^6 hourly dispatch problems. Simulations accurately consider the operating cycles of generation units, operating constraints, generation failures, random load deviations, stochastic water inflows to reservoirs, operating constraints on the cascading hydro system, etc. Fig. 7 illustrates 100 simulated samples of the stochastic hourly evolution and uncertainty characterization of the water stored in the system's largest reservoir.

The simulated sample allows for the reliability assessment of this power system. Reliability indices are LOLP$=5.978\times10^{-3}$, expected energy not served EENS$=$1814.50 MWh/yr and the expected interrupted load E[P_{def}]$=$30.39 MW. Fig. 8 shows the distributions of the annual cumulated durations of both power deficit events (left) and the magnitude of the interrupted load (right). From these histograms, it is easily inferred that annual scarcity rents are extremely variable from one year to another.

During capacity shortfalls, all generators are requested to be online irrespective of their marginal generation costs[8]. At those times, available units will collect time scarcity revenues given according to power delivered during the deficit period multiplied by the prevailing scarcity price, which itself depends on the

[7] The gross reserve margin is computed as the excess installed power capacity with regards to the expected peak demand divided by the expected peak demand, i.e. (1254.28 − 964.32)100/964.32 = 30.07%.

[8] This implicitly assumes that the marginal opportunity cost of consumption exceeds the marginal production cost of the most expensive generating unit.

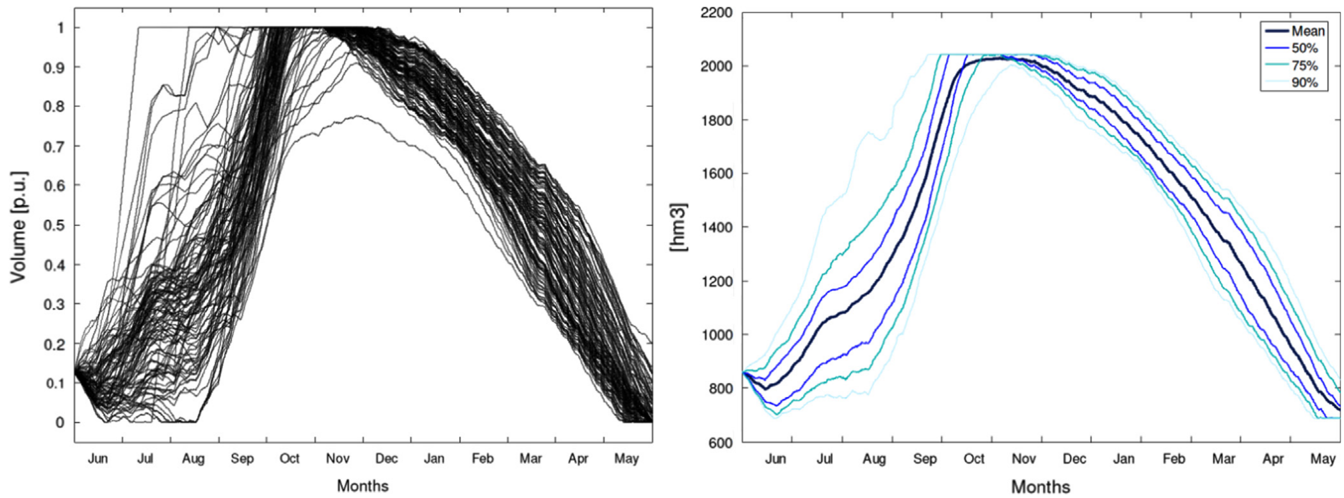

Fig. 7. Samples of stored water volume along the year (left) and confidence intervals of reservoir state (right).

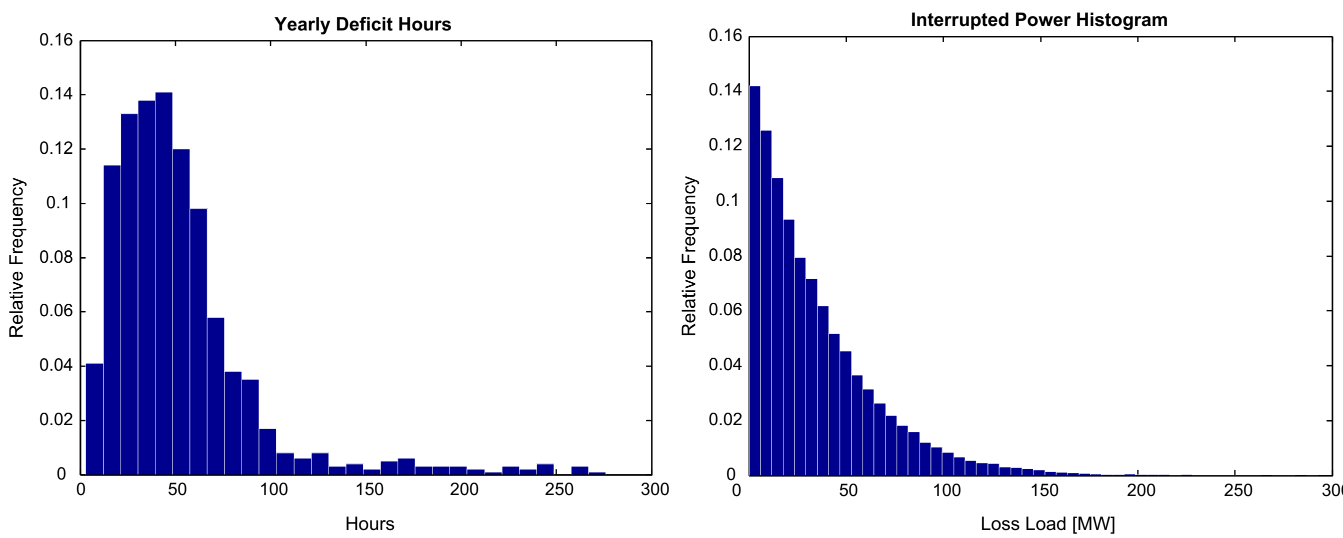

Fig. 8. Relative frequency histograms of the reliability indices.

magnitude of the rationing[9]. By averaging the obtained scarcity revenues over all simulated samples, the reliability payments awarded to each unit under the proposed rules can be estimated. Total payment to generators amounts to 7.09 M$/yr. Fig. 9 illustrates the normalized expected scarcity revenues per MW of installed capacity annually captured by each system unit.

For most of the thermal units, the typical reliability payment is about 7500 US$/MW/yr[10], as they are very similar with respect to unit size, availability and operating cycling characteristics. Noticeably, the remuneration for several others departs significantly from this typical value. This is evident for hydro-power generators (Unit 1 to Unit 4). Even though technical availability of hydro units is usually very high, hydro power plants are more frequently involved among causes of power shortages because of the significant uncertainty in the availability of water resources and the relatively large size of these generators. Dry years with unexpectedly poor water inflows have two effects on hydro power plants. On the one hand, total water volume in the period that is available for energy generation becomes exceedingly constrained. Often, hydro power plants only have water sufficient for a few hours of operation per day, leading to severe problems in supplying the aggregated electrical energy demand during these periods. And on the other hand, the maximum power output of hydro units becomes highly derated due to the low levels (head) in the reservoirs. As a result, available power generation capacity reduce drastically, giving rise to serious difficulties in meeting the peak load. On top of both of these issues, droughts reduce water availability in fairly large geographical areas, affecting many hydro reservoirs simultaneously and exacerbating the described problems[11]. Cross-correlation of hydrological resources significantly diminishes the contribution of hydro units to supply reliability in hydro-dominated power systems.

Under an energy-only market design, hydro generators will, on average, be able to capture only a fraction of the scarcity revenues that would capture a thermal unit with the same nameplate

[9] For some units, the supplied power during shortfalls can be significantly lower than its capacity because of operating constraints (e.g. ramping), derated states, depleted or low reservoir levels, transmission congestions, etc.

[10] Here, average payment is determined by current system reliability. If supply adequacy worsens, reliability payments would increase; the opposite is also true. The dynamic investment signal is a desirable property of the proposed remunerating mechanism.

[11] Severe droughts are the main risk factor in systems with a substantial share of hydropower. For instance, this was the cause behind power rationing and rolling blackouts in Argentina (in 1988/89), Brazil (in 2000/01), Ecuador (in 2009/10), Chile (in 1998/99, and again under severe shortfall risk by 2007/08 and 2011/12) and Panama (in 2013).

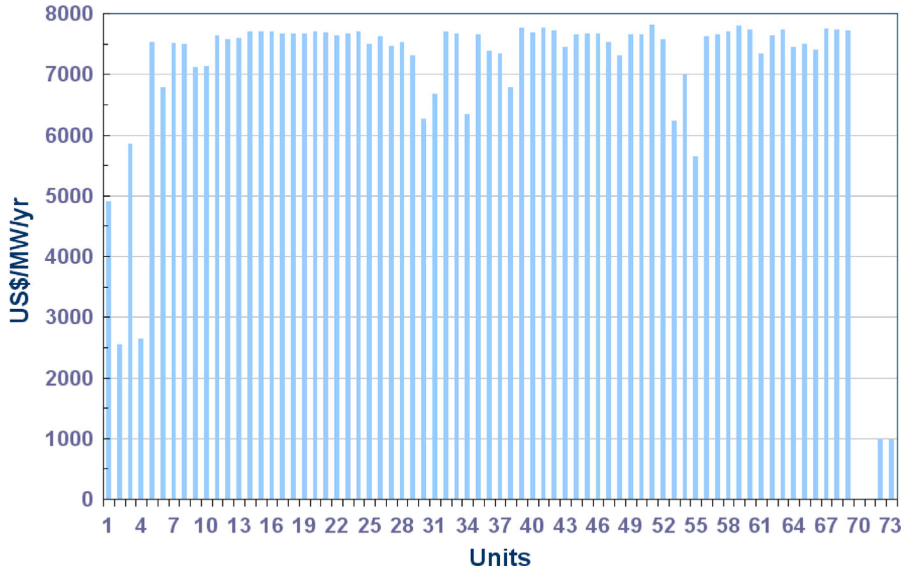

Fig. 9. Normalized annual reliability payment for each generation unit.

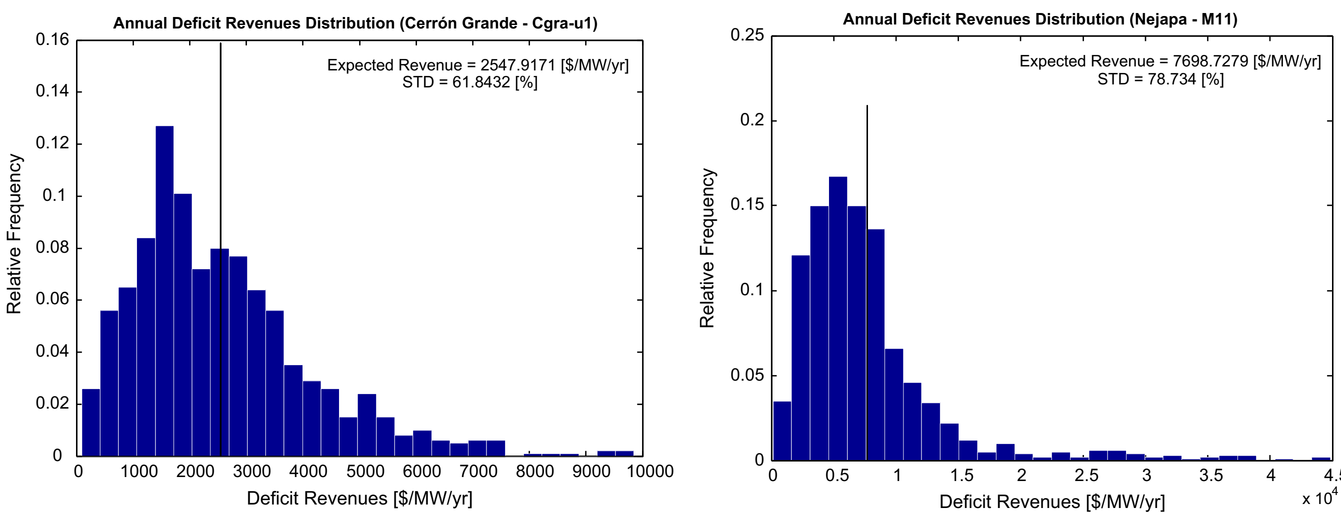

Fig. 10. Distribution of scarcity revenues and reliability payments for Unit 2 (left) and for Unit 15 (right).

capacity. Therefore, measured per MW of installed capacity, hydro units are typically awarded significantly lower reliability payments than thermal plants, properly reflecting the lower contribution these renewable generators have towards system reliability. In the simulation model, this fact is confirmed by the statistical evaluation of the revenues obtained by hydro units during shortfall events. Fig. 10 shows the frequency distribution of normalized scarcity revenues collected by a hydro and a thermal unit respectively.

The reliability payment to each individual generator is determined by its effective contribution in reducing the probability and extent of power deficit events. The stochastic simulation model does not explicitly penalize payments to generators based on renewable generation technologies. Reliability payments have clear advantages in terms of fairness and efficiency when compared to conventional capacity payments. In most implementations, renewable generators like wind or solar are regarded as unable to provide “firm capacity”, and rules typically do not entitle these generators to receive any payment amount. As seen earlier with regards to hydro and biomass units, the proposed approach recognizes the contribution of renewable power plants to system reliability. Moreover, renewable generation subject to low uncertainty in primary resource availability may receive reliability payments quite similar to typical thermal units. The power system of El Salvador comprises seven geothermal power units (Unit 32 to Unit 38). Geothermal generation has low uncertainty with regards to resource availability, thus making a significant contribution to supply reliability. This is reflected in the high normalized scarcity revenues collected by these generators, as is illustrated in Fig. 9.

The effect of unit size on the reliability payment becomes evident for Unit 3 and Unit 4 (cf. Fig. 9). Both generators are run-of-river power plants subject to the same water flows. The installed capacities are 92 MW and 170 MW for Unit 3 and Unit 4, respectively. Absolute payment amount awarded to each generator is similar. However, Unit 3 receives a reliability payment substantially higher than Unit 4 when measured per MW of installed capacity.

The substantial differences in the normalized payments reflect the fact that the capacity of each unit makes for a notably distinct contribution to overall system reliability. Unit 4 is the largest generator in the power system; as such, its outage is the most severe contingency, and the most likely cause of a power shortfall. In fact, the unavailability (either forced or planned) of this generator leaves the system with a very tight reserve margin

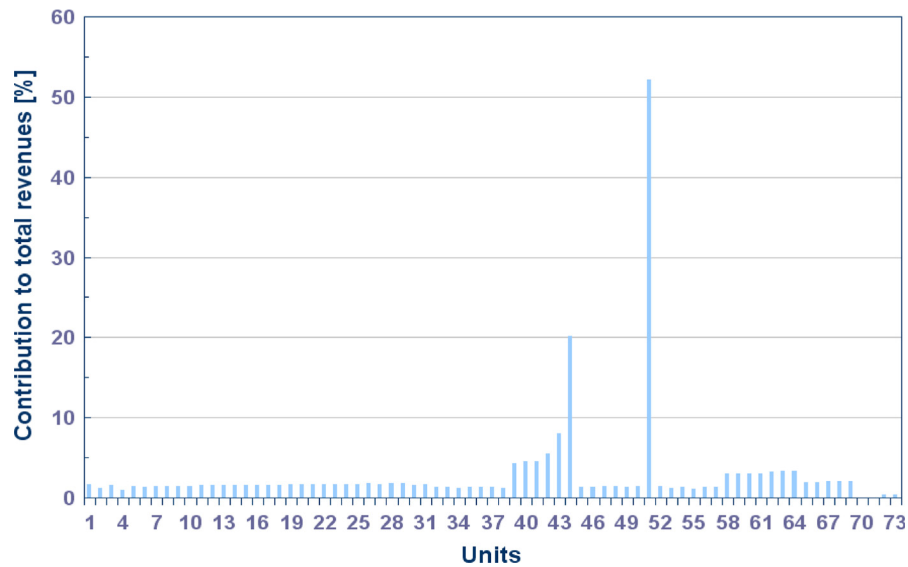

Fig. 11. Contribution of reliability payments to the expected total revenues.

and vulnerable to undergo a power shortage. On the other hand, Unit 3's outage is far less critical because of its smaller size. As the contingency of Unit 4 is more likely involved in the causes of a power shortfall event, Unit 4 would capture on average much less scarcity revenues per MW of installed capacity than any smaller generator under an energy-only market design. Likewise, the proposed reliability payments reflect precisely this effect.

Time and duration of scheduled maintenance may also play an important role on the reliability payment received in a particular year. For instance, Unit 5 to Unit 31 make for 27 identical 5.3 MW ICE-based generators. However, Unit 6, 30 and 31 receive maintenance on weeks with the lowest levels in system reservoirs, i.e. during the critical periods for supply demand. For this reason, these units are unavailable when needed most, and might not collect scarcity revenues if load curtailment becomes necessary. The computed payments properly account for the lower contribution of these units to system reliability. In the most extreme case, Unit 70 and 71 will not receive any payment because they would undergo a major overhaul and repowering, which are scheduled for the entire annual period.

Likewise, fuel availability can strongly impact payments. Unit 72 and Unit 73 are thermal generators owned by a sugar company who utilize sugar cane remnants as fuel. The availability of biomass used to fire these units is limited to the harvest period that spans only a few weeks a year. Accordingly, the capacity contribution of these units to reliability is quite modest, but still worth something, as reflected in the reliability payments[12].

Although Unit 51 has a capacity twice as large as the average thermal unit, with more than 7800 $/MW/yr, it receives the highest reliability payment. This fact can be explained in that this generator typically has the highest variable generation cost, so it serves only as standby reserve unit. Unit 51 is only dispatched to prevent or mitigate a power shortfall. Because of its highly infrequent and extremely short operating times, it poses the highest unit availability in the system, and can therefore capture the scarcity revenues with the highest probability. The calculated reliability payment clearly reflects this fact.

Fig.11 shows the contribution of expected scarcity revenues (or the certainty-equivalent payment) to the expected total revenues annually collected by each unit. For most units, reliability payments account for only a thin fraction of the gross revenues, as the primary source of revenues for most generators are energy sales in the spot market. Nevertheless, for reserve units, scarcity revenues can represent the main revenue source, as is the case with Unit 51 discussed above. The investment signal for peaking and reserve units is, therefore, more consistent and much less risky under the reliability payment design than under the energy-only market approach.

4. Conclusions and policy implications

Reliability is perhaps the most valued attribute in the delivery of electrical energy. A continuous and reliable power supply can only be achieved if adequate generation capacity is timely in place to meet increasing electricity consumption and ever-occurring outages, scheduled or unscheduled, of system equipment.

Under the hypothesis of risk neutrality, energy-only spot electricity markets are efficient and deliver the long-run optimal generation capacity. However, investors' non-neutral behavior, coupled with the dramatic volatility of scarcity rents, hinder efficiency and supply security in real market settings. For this reason, many regulators have introduced capacity mechanisms in order to ensure timely and sufficient generation investments. Rules for remunerating generation capacity are perhaps the most contentious issue when designing electricity markets. Although several capacity mechanisms have been proposed to overcome the shortcoming of price-based capacity approaches, explicit fixed capacity payments are still a preferred remunerating scheme by regulators in many systems.

In this work, a new probabilistic, price-based approach to remunerate supply reliability provided by generation units has been envisioned. The proposed scheme preserves the theoretical efficiency properties of energy-only markets while removing most of its deficiencies. The proposed reliability payments are aimed at introducing a significant methodological improvement in those markets currently relying on price-based regulation for securing supply adequacy. These reliability payments are intended as a superseding mechanism to replace the fixed capacity payment approach in such markets.

Under the proposed capacity remuneration approach, volatile and sporadic scarcity rents are replaced by a certainty-equivalent

[12] Please note in Fig. 10 that reliability payments represent only the 0.36% of the gross revenue for the biomass-fired Unit 72 and Unit 73. These generators rely on selling into the energy market for ensuring financial sustainability.

payment, drastically reducing volatility of scarcity revenues. By paying generators the expected value of the scarcity revenues that they would collect from the energy market, financial risk associated with these highly uncertain cash flows is removed. Thereby, conditions of risk neutrality necessary to attain efficiency equivalence to energy-only markets are thereby reestablished.

The remuneration approach is efficient and fair, as it inherently values the effective contribution of each individual generation unit to the overall supply reliability. When determining the individual contribution of each unit to system adequacy, which is thereby reflected in reliability payments, unit size, failure rate, cycling characteristics, operating constraints, resource availability, and unit maintenance scheduling are all important factors taken into consideration.

Payments to generation units automatically adjust to the prevailing, overall generation adequacy; they increase if system reliability deteriorates and vanish if capacity overbuild arises. Therefore, the proposed mechanism preserves efficiency of the long-term investment signal.

Reliability payments have advantages in terms of fairness when compared to conventional capacity payments. For instance, under the capacity payment design, renewable generators are typically regarded as unable to provide “firm capacity”, and rules often do not entitle these generators to receive any payment. Though less worthy than a typical thermal unit, renewable generation capacity is still valuable in terms of system reliability. The proposed reliability payments acknowledge this contribution.

In addition, the reliability payment mechanism has attractive incentive compatibility features with respect to market power mitigation during rationing, as well as maintaining high unit availability. Furthermore, the introduction of the suggested remuneration scheme does not distort the energy market and the short-term generators' operating decisions.

Computing the proposed reliability payments requires fairly sophisticated stochastic simulation models, particularly if system reliability depends mostly on operating decisions on hydro reservoirs. The stochastic model allows for identifying of a set of system states that lead to load curtailment actions to maintain the power balance together with the magnitude and occurrence probability of rationing events.

The methodology also needs extensive system data, which may constitute a difficulty if data is unavailable or unreliable. Furthermore, gathering these data may be expensive and time consuming. At first, these difficulties may seem a weakness of the proposal. However, in practice this is not the case, as ISOs routinely collect and have available most of these data in order to properly perform its function of operating the power system and monitoring the reliability level. ISOs routinely run hydro-scheduling models for mid-term planning and coordination, as well as stochastic simulations models to estimate the risk of a power shortfall. These planning and reliability models and the needed input data are essentially identical to that required for computing the proposed reliability payments. For this reason, the incremental effort and cost of implementing the proposed approach should be low.

The proposed approach does not restrict applicability to larger hydrothermal systems with more units and more watersheds. Computation of the expected future cost function and the marginal value of water in each reservoir in large-scale hydrothermal systems can be easily addressed by current mid-term hydro-scheduling software packages. Furthermore, the stochastic hourly simulation of a hydrothermal power system is a problem that only grows linearly with the number of units. In addition, Monte Carlo simulations are independent from each other. This allows for distributed computing techniques to be applied in order to exploit the multi-core architecture of modern processors and computer clusters, further reducing computation time.

The examination of locational considerations in reliability payments is an important avenue of research. The transmission network is not fully reliable and may constrain generators that would otherwise aid in alleviating deficit conditions. Spatial efficiency and fairness of the proposed reliability payment mechanism under LMP rules still needs to be investigated.

Acknowledgments

This work was supported in part by the National Scientific and Technical Research Council (CONICET), and the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) under Grants PIP 2009-11220090100796 and FONCyT-PICT-2008-2075 respectively.

The article has benefited from the feedback of colleagues of the academic and the industry sector. The constructive discussions with Prof. Osvaldo Añó, Dr. Diego Ojeda and Mr. Ricardo Rubio are gratefully acknowledged. The authors are the only responsible for any remaining error.

References

Allan, R.N., Navarro Sánchez, H., 2004. Uncertainty considerations in the pool purchase price in the England and Wales electricity supply industry. IEEE Proc. Gener., Transm. Distrib. 141 (2), 125–132.
Arango, S., Larsen, E., 2011. Cycles in deregulated electricity markets: empirical evidence from two decades. Energy Pol. 39 (5), 2457–2466.
Baldick, R., Helman, U., Hobbs, B., O´Neill, R., 2005. Design of efficient generation markets. Proc. IEEE 93 (11), 1998–2012.
Batlle, C., Rodilla, P., 2010. A critical assessment of the different approaches aimed to secure electricity generation supply. Energy Pol. 38 (11), 7169–7179.
Billinton, R., Allan, R., 1996. Reliability evaluation of power systems. Plenum Press, New York.
Billinton, R., Chen, H., Ghajar, R., 1996. Time-series models for reliability evaluation of power systems including wind energy. Microelectron. Reliab. 36 (9), 1253–1261.
Billinton, R., Ge, J., 2004. A comparison of four-state generating unit reliability models for peaking units. IEEE Trans. Power Syst. 19 (2), 763–768.
Boiteux, M.P., 1949. La tarification des demandes en pointe: Application de la théorie de la vente aucoût marginal. Revue générale del' électricité 58, Translated as “Peak Load Pricing”. J. Bus. 33 (157–179), 1960.
Breipohl, A.M., Lee, F.N., Zhai, D., Adapa, R., 1992. A Gauss–Markov load model for the application in risk evaluation and production simulation. IEEE Trans. Power Syst. 7 (4), 1493–1499.
Caramanis, M.C., 1982. Investment decisions and long-term planning under electricity spot pricing. IEEE Trans. Power Appar. Syst., PAS 101 (12), 4640–4648.
Chao, H.P., 1983. Peak load pricing and capacity planning with demand and supply uncertainty. Bell J. Econ. 14 (1), 179–190.
Chuang, A., Wu, F., 2000. Capacity payments and the pricing of reliability in competitive generation markets. In: Proceedings of the 33th Hawaii International Conference on System Sciences, pp. 1–7.
Cramton, P., Stoft, S., 2008. Forward reliability markets: less risk, less market power, more efficiency. Utility Pol. 16 (3), 194–201.
Crew, M.A., Kleindorfer, P.R., 1976. Peak load pricing with a diverse technology. Bell J. Econ. 7 (1), 207–231.
Crew, M.A., Fernando, C.S., Kleindorfer, P.R., 1995. The theory of peak-load pricing: a survey. J. Regul. Econ. 8 (3), 215–248.
de Vries, L., Heijnen, P., 2008. The impact of electricity market design upon investment under uncertainty: The effectiveness of capacity mechanisms. Utility Pol. 16 (3), 215–227.
Fisher, R.A., 1990. Statistical Methods, Experimental Design and Scientific Inference. Oxford University Press.
Ford, A., 1999. Cycles in competitive electricity markets: a simulation study of the western United States. Energy Pol. 27 (11), 637–658.
Garver, L., 1966. Effective load carrying capability of generating units. IEEE Trans. Power Appar. Syst. PAS 85 (8), 910–919.
Gorenstin, B., Campodonico, N., da Costa, J., Pereira, M., 1992. Stochastic optimization of a hydro-thermal system including network constraints. IEEE Trans. Power Syst.s 7 (2), 791–797.
Greco, D., 2000. Reliability assessment of interconnected power systems considering relevant transmission constraints (PhD dissertation). National University of San Juan (UNSJ), Argentina.
Hahn, G.J., 1972. Sample size in Monte Carlo simulations. IEEE Trans. Syst. Man Cybern. SMC 2 (5), 678–680.
Joskow, P.L., 1976. Contributions to the theory of marginal cost pricing. Bell J. Econ. 7 (1), 197–206.
Kleindorfer, P.R., Fernando, C.S., 1993. Peak-load pricing and reliability under uncertainty. J. Regul. Econ. 5 (1), 5–23.

Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resour. Res. 36 (6), 1519–1533.
IEEE Task Group on Peaking Service Units, 1972. A four state model for estimation of outage risk for units in peaking service. IEEE Trans. Power Appar. Syst. PAS 91, 618–627.
Neuhoff, K., de Vries, L., 2004. Insufficient incentives for investment in electricity generation. Util. Pol. 12 (4), 253–267.
North American Electric Reliability Council (NERC), 1997. Operating Policies and Planning Standards.
Olsina, F., 2013. Simulation of wind velocity for stochastic power market and power system analysis. Ingenerare 27 (12), 67–78.
Olsina, F., Garcés, F., Haubrich, H.-J., 2006. Modeling long-term dynamics of electricity markets. Energy Pol. 34 (12), 1411–1433.
Olsina, F., Larisson, C., Garcés, F., October 26–27th, 2007a. Analysis of uncertainties in hydrothermal generation systems . In: Proceedings of the VII Latin American Congress on Electricity Generation & Transmission (CLAGTEE), paper C021, Viña del Mar, Chile (in Spanish).
Olsina, F., Roscher, M., Larisson, C., Garcés, F., 2007b. Short-term optimal wind power generation capacity in liberalized electricity markets. Energy Pol. 35 (2), 1257–1273.
Olsina, F., Weber, C., 2009. Stochastic simulation of spot power prices by spectral representation. IEEE Trans. Power Syst. 24 (4), 1710–1719.
Oren, S., May 21–26, 2000. Capacity payments and supply adequacy in competitive electricity markets. In: Proceedings of the VII Symposium of Specialists in Electric Operational and Expansion Planning (SEPOPE), Curitiba, Brazil.
PSR Inc., May, 2006. SDDP Model. Methodology Manual, Version 9.0.
PSR Inc., November, 2009. SDDP Folder. 〈http://www.psr-inc.com/psr/download/folders/sddpfoldereng.pdf〉 (Online).
Reneses, J., Centeno, E., Barquín, J., 2006. Coordination between medium-term generation planning and short-term operation in electricity markets. IEEE Trans. Power Syst. 21 (1), 43–52.
Senjyu, T., Shimabukuro, K., Uezato, K., Funabashi, T., 2003. A fast technique for unit commitment problem by extended priority list. IEEE Trans. Power Syst. 18 (2), 882–888.
Schweppe, F.C, Caramanis, M.C., Tabors, R.D., Bohn, R.E., 1988. Spot Pricing of Electricity. Kluwer Academic Publishers, Boston, MA, USA.
Stoft, S., 2002. Power system economics: designing markets for electricity. IEEE-Wiley Press, Piscataway, NJ, USA.
Ubeda, J.R., Allan, R.N., 1994. Reliability assessment of composite hydrothermal generation and transmission systems using sequential simulation. IEEE Proc. on Gener., Transm. Distrib. 41 (4), 257–262.
Vázquez, C., Rivier, M., Pérez-Arriaga, I.J., 2002. A market approach to long-term security of supply. IEEE Trans. Power Syst. 17 (2), 349–357.
Wen, F., Wu, F., Ni, Y., 2004. Generation capacity adequacy in the competitive electricity market environment. Electr. Power Energy Syst. 26 (5), 365–372.
Yevjevich, V., 1987. Stochastic models in hydrology. Stoch. Hydrol. Hydraul. 1 (1), 17–36.