REFLEX: An Adapted Production Simulation Methodology for Flexible Capacity Planning

Jeremy Hargreaves, *Member, IEEE,* Elaine K. Hart, *Member, IEEE,* Ryan Jones and Arne Olson

***Abstract*—As intermittent energy resources become more significant in power production, traditional capacity planning may be insufficient to ensure reliable system operation. A system planner must ensure that flexibility solutions are available to respond to large and uncertain ramping events. These solutions may be operational, such as improved unit commitment and dispatch, curtailment of renewables, or demand response; procurement based, such as new fast ramping resources or batteries; or involve market reform. This paper outlines a new methodology for modeling the economic tradeoffs in implementing flexibility solutions for integrating renewables. The proposed model includes both a stochastic treatment of system states to account for a wide range of operating conditions and an adapted production simulation methodology that weighs the cost of reliability and subhourly flexibility violations against the cost of the operational flexibility solutions available to mitigate them. The model's functionality is demonstrated with a case study of California at a 50% RPS in 2030. The model tests the value of 1,088 MW of generic flexible units, relative to the same capacity of must-run resources, finding an expected annual value of \$347±42 million/yr. Potential applications of the model for resource planning and procurement are also discussed.**

I. Nomenclature

Symbol	Description
Sets	
T	Set of time periods in forecast window
I_{flex}, I_{firm}	Set of all units that can/cannot change unit commitment in forecast window
I_{intra}	Set of all units that can provide intra-timestep flexibility
J	Set of renewable technology types, or individual renewable plants
F^+, F^-	Set of all facets forming the expected upward/downward flexibility deficiency response surface
Variables	
x_{it}	Generation of unit i in time t (MW)
n_{it}	Commitment of unit i in time t, $\{0,1\}$
c_{jt}	Curtailment of renewable type j in time period t (MW)
$\overline{u}_t$, $\underline{u}_t$	Unserved energy/overgeneration in time t (MW)
$\overline{rr}_{it}$, $\underline{rr}_{it}$	Upward/downward ramp rate of unit i in time t (MW/timestep)
$ramp_t^+$, $ramp_t^-$	Intra-timestep upward/downward ramp policy
res_t^+, res_t^-	Intra-timestep upward/downward reserve policy
EFD_t^+, EFD_t^-	Expected upward/downward flexibility deficiency (MWh)
Parameters	
L_t	Forecasted load in time period t
R_{jt}	Forecasted generation from renewable type j in time period t
N_{it}	Locked in commitment of unit i in time t from previous window, $\{0,1\}$
$\overline{RR}_i$, $\underline{RR}_i$	Maximum upward/downward ramp rate for unit i (MW/timestep)
$\overline{X}_{it}$, $\underline{X}_{it}$	Maximum/minimum operating level for unit i in time period t
β_k^f	k^{th} coefficient in facet $f \in F^+$
γ_k^f	k^{th} coefficient in facet $f \in F^-$

II. Introduction

Renewable portfolio standards, feed in tariffs, and other policy incentive mechanisms, are increasing the penetration of intermittent renewable resources on electricity systems around the world. As a result, traditionally structured power systems must adjust to accommodate these new resources. Intermittent resources increase both the forecast error and the variability in net loads across multiple time scales [1]. This presents a challenge to system planners tasked with deciding which new resources to build, subject to forecasted trends in load growth and future renewable development. Traditionally, planning to a reliability standard involved ensuring sufficient system capacity to meet peak load. However, increasing renewable penetrations can cause system reliability problems due to inadequate flexibility rather than capacity. In this context, system planners must account for generator characteristics that are not typically modeled in capacity planning. These include start-up and shutdown time, minimum up and down time, minimum operating levels, and maximum ramp rates. Flexibility analysis for systems with high penetrations of renewables must therefore include significantly more operational detail than has been required by traditional reliability analysis.

Energy & Environmental Economics, San Francisco, CA

The requirement for greater system flexibility with higher renewable penetrations has been well established in numerous integration studies, for example [2], [3], [4], and as summarized in [5], [6], [7]. These analyses have attempted to quantify flexibility needs on a system planning time scale by looking at system operations under greater renewable penetrations. Regional studies have been typical in recent years for system operators facing the coming challenge of integrating significant amounts of renewables. For example, in California the system operator developed a methodology to assess integration issues for a 20% RPS, finding that thermal resources may not always be sufficiently flexible at low loads and recommending a reduction in self scheduling [4]. NREL's western wind and solar integration study examines the needs of the western states and concluded that 35% penetration of renewables was possible with increased balancing area cooperation, improved day-ahead forecasting, and sub-hourly scheduling [3]. These studies identify flexibility as an important component of the planning problem, but make only general recommendations regarding planning for flexibility.

Lannoye et al. recognize that system planning with flexibility targets increasingly includes elements of system operations and that those targets can be met by sources other than conventional generation [8], [9]. They define metrics by which to judge flexibility needs. Analogous to loss of load probability (LOLP) and effective load carrying capability (ELCC) used in traditional capacity planning, they propose inadequate ramp resource probability (IRRP) and effective ramping capability (ERC).

Capacity planning in the US generally allows an expectation of unserved load due to generator outages of 1 event every 10 years. However unlike traditional capacity planning, there are no accepted standards for reliability to determine adequate system flexibility. The "1-in-10" standard implicitly values the level of reliability customers are willing to accept, as would any new reliability standard for flexibility. The value of service reliability has been explicitly treated in the capacity planning problem in the past, for example in a recent study of the ERCOT reserve margin [10], though the "1-in-10" standard is well accepted and will likely continue to be so in the future. Rather than determine a new flexibility reliability standard however, we propose that the cost of various types of system failures due to flexibility deficiencies be explicitly considered in system planning, balancing gains in system reliability against the cost of the measures taken to achieve them.

The costs of system failures have previously been poorly defined. Survey studies have been done on value of lost load (VOLL), for example [11], [12], however the results range widely. Our motivation for including explicit system failure costs is to open the discussion between stakeholders in system planning on what mutually acceptable best estimates of failure costs might be, rather than relying on implicit values contained in the definition of a new flexibility standard that may not be exposed to stakeholder scrutiny. We apply this explicit cost framework to both system operations and system planning.

In system operations, reserves are held for capacity and flexibility needs. Ela, Milligan, and Kirby show that the scheduling of typical reserve products has to change as renewable penetrations increase [13]. They include a review of work done in determining reserve requirements for high penetration systems including [14], [15], [16], [17], [18]. Ortega-Vasquez and Kirschen[14] introduce the idea of trading off the marginal cost of spinning reserve against the social cost of load shedding. They use a spinning reserve requirement differentiated by net load level and determined offline. Morales et al [15], Yong et al [16], Bouffard et al [17], and Wang et al [18], solve for operating reserve levels endogenously using two-stage stochastic unit commitment problems. These are short term models solving for the operations of a day or less. We build on this work by linking system reserves to an expectation of system imbalances depending on system state, and defining a cost for each imbalance. Reserves can therefore be solved for endogenously when performing least cost commitment and dispatch.

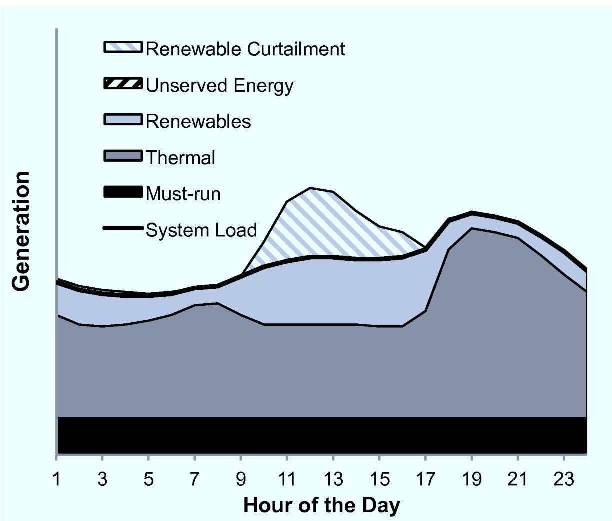
In system planning, the least cost portfolio for meeting capacity and flexibility needs may include non-traditional solutions [9], [19]. Examples of these include demand response, storage devices, improved forecasting, changes in market structure, renewable curtailment, and informed renewable procurement. We present a modeling methodology for determining the least cost portfolio of flexibility solutions for a given system. This method is based on a value-based reliability planning framework for flexibility, which is developed in more detail in a companion paper [20]. The modeling methodology described in this paper builds on traditional loss of load probability (LOLP) reliability analysis with an adapted mixed integer programming (MIP) production simulation model to address system flexibility. The MIP approach was chosen because of its extensive use in the energy sector today, the detailed and current datasets available for various energy systems, and the scalability of commercial solver packages. The planning framework is designed to help system planners answer the following question: which portfolio of capacity and flexibility solutions serve system needs at the lowest total cost?

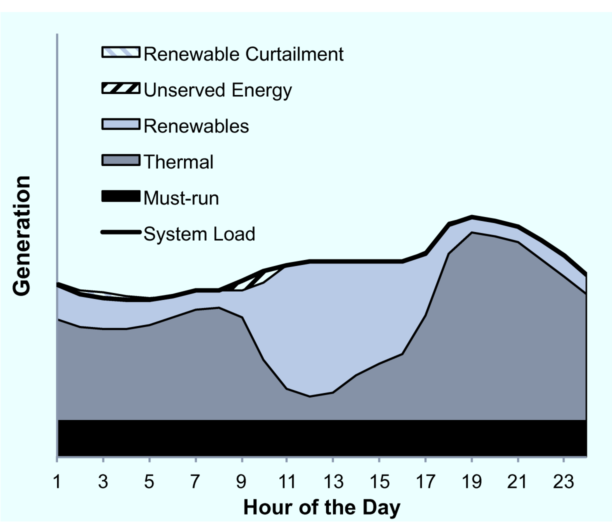
III. Theory

In value-based reliability analysis, the costs of reliability solutions can be justified when the solutions adequately avoid more costly reliability failures. This concept can be applied in both planning and operations. In planning, investments in capacity might be made to avoid firm load curtailment if the fixed costs do not exceed the avoided costs associated with firm load curtailment due to inadequate capacity [21], [22]. In operations, reserve levels can be adjusted based on the cost of holding additional reserves and the expected cost of reliability failures associated with inadequate reserves. In this analysis, we incorporate the cost trade-offs involving both traditional reliability failures and flexibility-based failures and their solutions in both the operations model and the planning decision.

To illustrate the tradeoff between flexibility violations and flexibility solutions, consider the case in which renewable curtailment is used as a flexibility solution. When the cost penalty for renewable curtailment is small, flexibility challenges brought on by renewables can be managed by curtailing the renewables. This is illustrated for an example day in Figure 2(a)). In this example, units are kept online to provide operating reserves throughout the day. These units cannot generate less than their minimum stable levels, so the system curtails renewables to balance the load. At the other extreme, curtailment costs can be set so high that the system prioritizes delivering renewables over holding operating reserves. This is shown for the same day in Figure 2(b). In this example, thermal plants are shut down in the morning to allow maximum delivery of solar power as the sun rises. When unexpected fog or cloud cover results in lower than anticipated renewable output, the operating reserves are not adequate to meet demand and the system experiences unserved energy.

Fig. 1. Example day in system with renewable curtailment penalty cost that is (a) significantly lower, and (b) higher than other violation penalty costs. The relative costs in this example determine whether the system experiences renewable curtailment or violations like unserved energy under high penetrations of renewables.





In the modeling framework described in this paper, these economic tradeoffs are treated explicitly in the economic unit commitment problem via cost penalties attributed to both flexibility violations and renewable curtailment. The various types of flexibility violations are discussed in more depth in [20].

IV. Methodology

Unit scheduling through solving security constrained unit commitment (SCUC) on a day ahead and hour ahead basis is common practice in deregulated competitive electricity markets[18]. These short time horizon deterministic models can incorporate many features of the system while maintaining acceptable run times. System planning, on the other hand, requires detailed examination of potential future system configurations, ideally under the full range of potential net load and system outage conditions that could be encountered. Traditional capacity planning can incorporate the full distributions of unit outages, and stochastic renewable production and load because modeling operations is unnecessary. Very infrequent loss-of-load events can still be captured without being limited by computing resources.

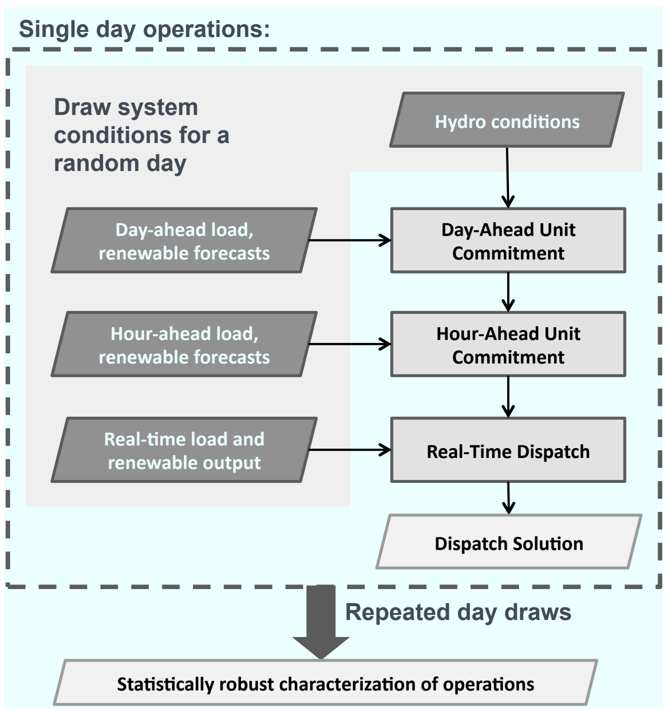
System planning for both capacity and flexibility, however, requires that operations are modeled to capture the capability of the system to avoid flexibility violations. With limited computing resources, the planning problem has to include the components of operations that are most important in evaluating flexibility, while stripping out components that have less of an impact on flexibility results. This tradeoff should be made on a system by system basis, however two main model features are necessary to capture the flexibility of a system. Firstly, unit characteristics, including ramp rates, start up and shutdown times, minimum up and down times, and minimum and maximum operating levels, should be captured. Secondly, the set of model runs chosen must be representative of the distributions of system conditions, such that the likelihoods of flexibility violations are properly captured.

The methodology described in this paper assumes that the number of MWs needed to meet the "1-in-10" capacity planning standard has already been identified and included in the resource stack. This number can be found using a loss of load probability (LOLP) approach, such as defined in [23]. We assess different resource planning portfolios by comparing the expected operating costs, including the expected costs of system failures, using multiple runs of an adapted production simulation methodology. The portfolios vary in the flexibility characteristics assigned to the MWs identified for capacity (i.e. the types of generating units used), and the additional procured resources or flexibility solutions. We can then determine the least cost portfolio solution to meeting future capacity and flexibility needs. This methodology is described in the following section.

A. Production simulation model

Multistage production simulation has been used in recent years to capture the effects of forecasting errors that are experienced between the day-ahead, hour-ahead, and real-time markets [4]. The adapted production simulation method presented in this paper adopts this approach and makes two additional contributions to production simulation. The first is a probabilistic framework to ensure that the results are sufficiently representative of input distributions, including low probability tail events, for planning analysis. The results of the model are therefore expected values of operating costs, violations, and costs of violations, which reflect the distributions of expected system conditions. The number of runs necessary for convergence to acceptable tolerances on expectations will vary by system. Convergence analysis is demonstrated in the case study later in this paper. The second is valuation of flexibility violations with endogenous treatment of reserve scheduling that captures the economic trade-offs in system operations with high penetrations of renewables. The multistage structure of the modeling approach is summarized at a high level in Figure 2. The two primary contributions are described below.

Fig. 2. Schematic of REFLEX operations modeling methodology.



1) Probabilistic Framework: One challenge in using production simulation in a planning process is the need for high resolution historical data. In traditional production simulation, a single year of historical load, renewable, and hydro data is adjusted to represent the simulation year (by scaling if necessary). An "average" year is often selected for this purpose so that the simulation results characterize typical operations. This limits the range of phenomena encountered in the simulation to the types of events that were experienced in that historical year. While this can provide a useful look at system operations under normal conditions, this approach does not provide enough operational information for a planning analysis, which must also ensure that the system is adequate to meet demand during unlikely events.

Rather than simulating a single year of operations, the model simulates a collection of days that are randomly generated from historical data in order to better represent the full range of system conditions that may be encountered. Each scenario is formed by combining randomly selected load, wind, and solar profiles from the historical record. In order to preserve seasonal and meteorological correlations in these datasets, the historical data is first binned and then daily profiles of each variable are selected from within the same bin. Binning strategies may vary across systems, but initial tests have been performed with a binning scheme that first separates days based on month and then bins the days in each month according to whether the system experiences high load or low load conditions for that month. The demarcation between high vs low load days is determined by minimizing the variance in daily average wind and solar conditions within each bin. The binning takes into account whether each day falls on a weekday, weekend, or holiday so that only meteorological effects (rather than human effects) determine which days are deemed high vs low load days. If n days of coincident load, wind, and solar shapes are available from the historical record and the days are broken into m bins, this approach increases the number of daily net load shapes that can be simulated by the model by a factor of approximately $(n/m)^2$ (the actual factor depends on the sizes of each bin). One additional benefit of the sampling approach is that each modeled day is independent from the other days in the simulation, allowing for parallelization to improve runtimes.

In modeling a single day's operations, dependence on system conditions prior to, and after, the modeled day should be accounted for. Dependencies include unit commitment obligations of long up or down time units, hydropower and future large storage energy budgets, unit outages prior to the modeled day, and potentially ramp rate constraints. To partially capture these dependencies, we model a period before the focus day and a period after. The appropriate length of these periods will depend on the system modeled and the tradeoff between model accuracy and runtime.

In addition to the long-term uncertainty in net load conditions over time, the model takes into account short-term stochastic phenomena like forecast errors and forced outages. When each simulation day is generated, load and renewable forecasts specific to the commitment windows being modeled, day ahead or hour ahead for example, are either pulled from historical forecast data or generated using statistical models. The unit commitment problems are solved using the associated forecasts (see Figure 2). Forced outages are incorporated via Monte Carlo simulation by randomly drawing the outage state in the first time step of each day based on the forced outage rate and applying an outage/repair model (e.g. representing the time to failure and time to repair with exponential distributions) for the remaining time steps in each day.

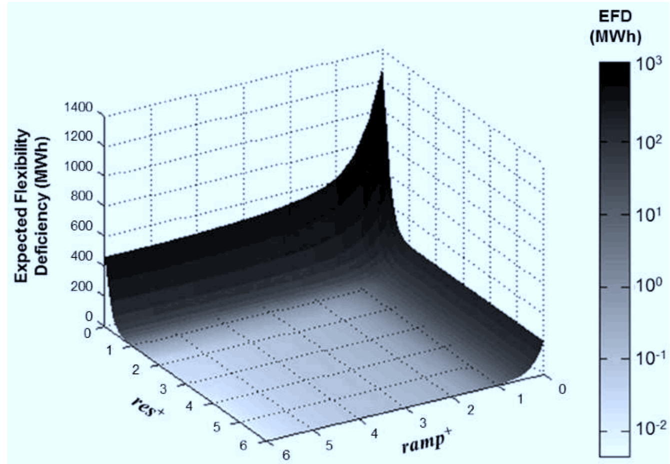
2) Valuing Flexibility Violations and Endogenous Reserve Scheduling: As discussed in Section III, the adapted production simulation methodology incorporates the costs of flexibility violations. Four types of violation are explicitly treated in the adapted production simulation methodology: unserved energy (UE) and overgeneration (OG) from not meeting forecasted net load at each simulation time step, and the costs of upward and downward flexibility violations expected to occur within the simulation timestep (EFD+ and EFD-). EFD violations arise when reserve levels are not adequate to accommodate sub-timestep net load fluctuations and/or forecast errors. While traditional production simulation methods might register a load following violation (in MW), which is difficult to value, this adapted methodology will register the expected energy shortfall (or overgeneration) within the hour (in MWh) associated with inadequate reserves. Shortfalls can be multiplied by their expected costs and solutions to the shortfall can be evaluated on a cost basis. Small EFD violations may be manifested as area control error (ACE), while larger EFD violations might actually bring about lost load. In the simplest formulation, the user assigns each violation a constant cost penalty that reflects the relative severity of the violation. In more complicated formulations, non-linear cost penalties can be included, provided that the relationship between the cost penalty and magnitude of violation in each timestep is monotonic and non-decreasing.

By including the EFD$^+$ and EFD$^-$ terms in the objective function, the model is capable of endogenous reserve scheduling, which trades the cost of expected sub-timestep flexibility violations against the cost of holding additional reserves to avoid the violations. This method relies on the development of response surfaces that relate the EFD$^+$ and EFD$^-$ to the system state. These EFD surfaces are developed prior to the unit commitment simulation. As will be shown, the surfaces are convex and monotonic with respect to the total MW of reserves held and the combined MW/timestep of ramping available from those reserves. The EFD surfaces can therefore be represented with a series of inequality constraints in the model formulation. The model chooses the system cost minimizing solution, including the decision variables representing a) upward and downward system reserves (res^+/res^-), and b) upward and downward system ramp ($ramp^+$/$ramp^-$).

The EFD surfaces can be built in a number of ways. For the case study presented in Section V, a simple high-resolution time-sequential simulation step was performed prior to running the full production simulation model. Within each hour of historical data, and given the demand and renewable forecasts, the corresponding actual 5-min demand and renewable shapes, and characteristics of a generic set of committed units, the simulation tracks the ability of the fleet to ramp up and down with the net load on a 5-min basis. The fleet's response is limited by the aggregated minimum stable level, maximum generation, and maximum upward and downward ramp rates, all represented by the system res and $ramp$ variables. The simulation is performed over a wide range of these state variables to build a multidimensional surface describing the ability of various unit commitment solutions to accommodate net load fluctuations and forecast errors.

In a real system, the EFD surface is expected to change for each state of the system (renewable output, R, and load, L). To reduce the number of dimensions in the EFD surface, the res and $ramp$ variables are normalized by the traditional load following requirements in each time step, $f(R_t, L_t)$, a function of the renewable output and load, so that the same EFD upward and downward surfaces apply to all possible system states. Separate EFD surfaces are built for each commitment window so that reserve levels reflect the improvement of forecasts with decreasing forecast horizon. An example EFD surface for upward flexibility violations in the hour-ahead market is shown in Figure 3. This approach of parameterizing short time-scale operations (which cannot be resolved with traditional production simulation) for inclusion in the unit commitment problem may be extended to shorter time scales provided that the user has data with high temporal resolution and an accurate method of simulating the response of various technologies to net load fluctuations on those time scales. Alternatively the model can be run with pre-determined reserve constraints to provide traditional production simulation analysis.

Fig. 3. Example EFD surface for upward flexibility in the hour-ahead unit commitment problem. Color scale shows that EFD spans several orders of magnitude.



Note that the output of the real-time dispatch stage of the production simulation provides the same information required to build the EFD surface. In future analyses, the subhourly unserved energy observed in real-time dispatch may be fed back into the day-ahead and hour-ahead unit commitment problem to improve the treatment of dynamically scheduled reserves. Alternative normalization and binning schemes may also be appropriate for each system under study.

The following model constraints show the adaptations made to a traditional production simulation formulation. These apply to all commitment windows. The load balance includes all generation as well as unserved energy, overgeneration, and renewable curtailment:

$$\sum_i x_{it} + \sum_j (R_{jt} - c_{jt}) + \overline{u}_t - \underline{u}_t = L_t \quad \forall t \in T \tag{1}$$

Unserved energy and overgeneration are represented by $C_{\overline{u}} \times \sum_{t\in T} \overline{u}_t$ and $C_{\underline{u}} \times \sum_{t\in T} \underline{u}_t$ in the objective function, respectively. The cost coefficients $C_{\overline{u}}$ and $C_{\underline{u}}$ are per MWh penalties set as inputs to the model.

The model also includes an option to economically schedule and dispatch renewables in the downward direction to meet flexibility requirements. This downward dispatchability is incorporated into the unit commitment problem with the renewable curtailment variable c_{jt} :

$$0 \le c_{jt} \le R_{jt} \qquad \forall j \in J, t \in T \tag{2}$$

Where renewable curtailment introduces an additional cost in the objective function: $C_c \sum_{t\in T, j\in J} c_{jt}$. The renewable curtailment option is exercised by setting the renewable curtail-

TABLE I
SUMMARY OF CONVENTIONAL UNITS USED IN THE CASE STUDY, WITH OPERATING CHARACTERISTICS LISTED FOR AN AVERAGE (WEIGHTED BY CAPACITY) UNIT OF EACH TECHNOLOGY TYPE.

	Fleet Statistics			Average Unit Characteristics					
Technology	**No. of Units**	**Total Capacity (MW)**	**No. Rescheduled in Hour-Ahead**	**Max Gen (MW)**	**Min Gen (MW)**	**Max Ramp (MW/min)**	**Min Up Time (hrs)**	**Min Down Time (hrs)**	**Heat Rate (Btu/kWh)**
CCGT	61	22,833	3	374	150	4.53	6.93	6.04	7,678
CT	141	8,760	103	62.1	31.4	5.66	1.73	1.51	10,538
Cogen	90	3,867	0	43.0	42.7	0.01	8.47	8.47	(must run)
ICE	16	213	10	13.3	4.70	1.39	1.94	1.47	9,208
Nuclear	3	3,077	0	1,026	899	3.44	168	168	(must run)

ment cost, C_c, to a value that is lower than the overgeneration penalty cost.

The total upward ramping ability of the system within the hour is described by the variable $ramp_t^+$ times a parameterized function of the system state in terms of the forecasted demand and available renewables (eg. the load following requirement), $f(R_t, L_t)$. Examples of reserve requirement functions are described in [24]. The upward ramping capability in each time step is incorporated into the traditional unit commitment problem with the following constraint:

$$ramp_t^+ \times f(R_t, L_t) \leq \sum_{i \in I_{intra}} \left[\overline{rr}_{it} - (x_{it+1} - x_{it})\right] \quad \forall t \in T \tag{3}$$

Similarly, the upward reserves available within the hour (which is limited by the difference between each unit's set point and maximum output), is described by $res_t^+ \times f(X)$ and is incorporated into the problem with the following constraint:

$$\begin{aligned} res_t^+ \times f(R_t, L_t) \leq & \sum_{i \in I^w_{firm} \cap I_{intra}} \left[N_{it}\overline{X}_i - x_{it}\right] \\ & + \sum_{i \in I^w_{flex} \cap I_{intra}} \left[n_{it}\overline{X}_i - x_{it}\right] \quad \forall t \in T \end{aligned} \tag{4}$$

Units with firm commitment cannot change commitment status due to the length of their start up or shut down times, or other constraints on their operation. Their status may be set in the previous commitment window when start up and shut down times are not limiting. The EFD^+ associated with each $(ramp_t^+, res_t^+)$ policy is approximated in the model by a piecewise linear function so that:

$$\text{EFD}_t^+ \geq \beta_0^f ramp_t^+ + \beta_1^f res_t^+ + \beta_2^f \quad \forall t \in T, f \in F^+ \tag{5}$$

The EFD^+ is incorporated into the objective function with a cost term equal to $C_{\text{EFD}^+} \times \sum \text{EFD}_t^+$. The same approach is used for downward flexibility:

$$ramp_t^- \times f(R_t, T_t)) \leq \sum_{i \in I_{intra}} \left[\underline{rr}_{it} + (x_{it+1} - x_{it})\right] \quad \forall t \in T \tag{6}$$

$$\begin{aligned} res_t^- \times f(R_t, T_t)) \leq & \sum_{i \in I^w_{firm} \cap I_{intra}} \left[x_{it} - N_{it}\underline{X}_i\right] \\ & + \sum_{i \in I^w_{flex} \cap I_{intra}} \left[x_{it} - n_{it}\underline{X}_i\right] \quad \forall t \in T \end{aligned} \tag{7}$$

$$\text{EFD}_t^- \geq \gamma_0^f ramp_t^+ + \gamma_1^f res_t^+ + \gamma_2^f \quad \forall t \in T, f \in F^- \tag{8}$$

An additional cost term reflects the downward flexibility violation cost: $C_{\text{EFD}^-} \times \sum_{t \in T} \text{EFD}_t^-$.

V. CASE STUDY

To demonstrate the functionality of the model, a case was developed to examine flexibility challenges in California under a high RPS. The case described in this section was modified from the 2030 50% Large Solar Case presented in [25] and was run on the ProMaxLT™ production simulation platform, modified according to the formulation described in Section IV. The study area includes the CAISO, SMUD, and LADWP balancing areas. The case was modified for this analysis by converting 1,088MW of CCGTs and CTs into generic inflexible resources that operate as must-run at the maximum generating level and can therefore provide no ramping or reserve services. Table I summarizes the conventional resources that are modeled in the case and Table II summarizes the renewable resources. Details regarding the modeling assumptions for imports/exports and hydropower can be found in [25].

Input prices and cost penalties are listed in Table III. The curtailment cost penalty depends on the system being analyzed and the goal of the study. For a planning exercise, the curtailment cost penalty may reflect an opportunity cost. For example, if renewable resources are curtailed and a system fails to meet its renewables portfolio standard or other binding

TABLE II
RENEWABLE RESOURCES IN THE 50% RPS CASE STUDY. ANNUAL ENERGY REFERS TO AVAILABLE ENERGY, BEFORE CURTAILMENT. SOLAR PV INCLUDES ROOFTOP PV NOT CONTRIBUTING TO THE RPS. WIND EXCLUDES CONTRACTED OUT-OF-STATE RESOURCES.

Renewable Resource	Annual Energy (GWh)
Biogas	2,133
Biomass	7,465
Geothermal	16,231
Small Hydro	4,525
Solar PV	75,829
Solar Thermal	4,044
Wind	29,948
Total	**140,175**

TABLE III
INPUT COST ASSUMPTIONS FOR THE 50% RPS CASE STUDY. ALL VALUES LISTED IN 2012$.

Cost Parameter	**Input Value**
Natural Gas Price	$6.06/MMBtu
Carbon Price	$50/t$CO_2$
Renewable Curtailment Penalty	$150/MWh
Unserved Energy Penalty	$39,000/MWh
EFD_t^+ Penalty	$39,000/MWh
EFD_t^- Penalty	$150/MWh

target as a result, then the penalty for renewable curtailment should reflect the cost of procuring additional resources (which can be delivered) to meet the target. A value of $150/MWh was selected for this cost to approximately reflect a renewable PPA price that has been scaled up to account for curtailment of incremental resources. Further discussion of system-appropriate penalty costs is beyond the scope of this paper, and will constitute important future work. The cost of unserved energy was approximated as $39,000/MWh based on a load-weighted average of the value of service by customer type as summarized in [12].

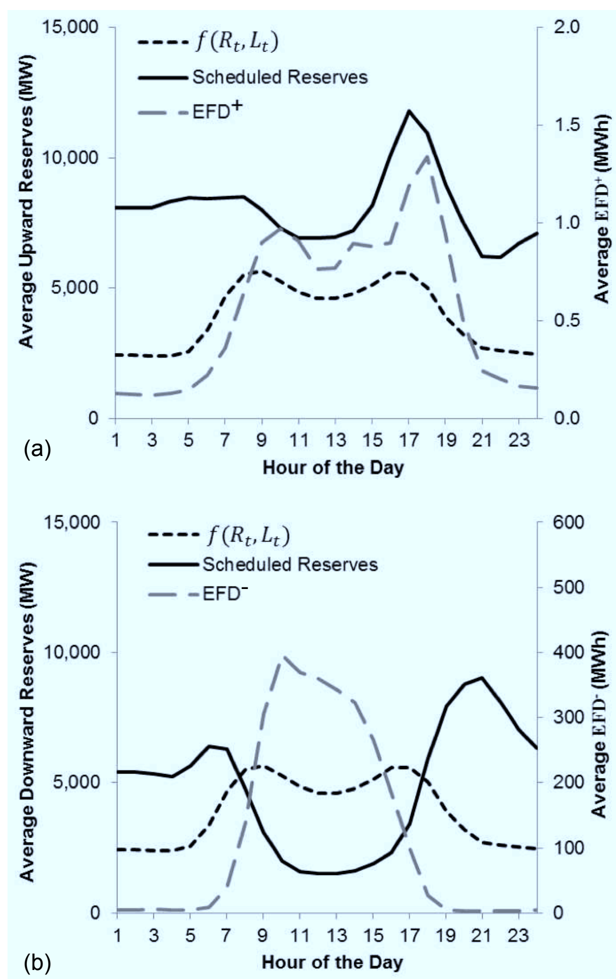
The proposed approach was used to model both the day-ahead and hour-ahead unit commitment, where the commitment schedules for long-start units are locked in based on the day-ahead commitment decision. Separate EFD surfaces were built from day-ahead and hour-ahead forecasts, which were simulated to reflect plausible forecast errors in 2030 (see [25] for more detail). The case study explores three crucial aspects of the proposed methodology: endogenous reserve scheduling; the application of Monte Carlo methods to production simulation; and the economic framework for flexible resource planning. These analyses are described below.

For this case, the parameterization, $f(R_t, L_t)$ was calculated for each hour using the method described in [2] for calculating load following requirements. The cost penalties selected for expected subhourly imbalances (EFD$^+$ and EFD$^-$) determine how the model schedules reserves and should therefore reflect real operational consequences of subhourly imbalance. It is assumed in the case study that subhourly upward imbalances lead to unserved energy and that subhourly downward imbalances must be mitigated with real-time renewable curtailment. The cost penalties therefore reflect the cost of unserved energy and the cost of renewable curtailment, respectively. In practice, some subhourly imbalance may contribute to area control error (ACE) without serious consequence. Consideration of ACE may be incorporated in future analyses by subtracting an allowable (ie. unpenalized) amount of imbalance from each EFD surface.

Because the consequences of upward imbalances are significantly more costly than downward imbalances, the optimization chooses reserve policies that prioritize upward flexibility at the expense of downward flexibility when the system is flexibility constrained. This is demonstrated in Figure 4, which shows the hourly average load following requirements ($f(R_t, L_t)$) throughout the day, the available reserves scheduled by the model, and the resulting EFD. In general, large EFD is observed when the scheduled reserves do not exceed $f(R_t, L_t)$, as shown in Figure 4(b) for downward reserves. In the modeled scenario, which relies heavily on solar power, thermal fleet flexibility is constrained during daytime hours, when units shut down to accommodate large amounts of generation from solar resources. Figure 4 shows that in these hours, the set points of the remaining online resources are generally low enough to provide adequate upward reserves, but not high enough to provide adequate downward reserves. The consequence is substantial EFD$^-$ during daytime hours, which in this case operationally translates into real-time renewable curtailment.

The convergence behavior of various model outputs was investigated to inform the selection of appropriate convergence criteria. Model outputs include expected daily values of: operating cost, including both fuel and variable O&M costs; renewable curtailment; upward and downward subhourly imbalances; and total cost, including production cost, renewable curtailment costs, and subhourly imbalance costs. For this analysis, random days were drawn sequentially from each

Fig. 4. Average (a) upward and (b) downward reserves scheduled in hour-ahead unit commitment for the 50% RPS case study with average EFD shown throughout the day.



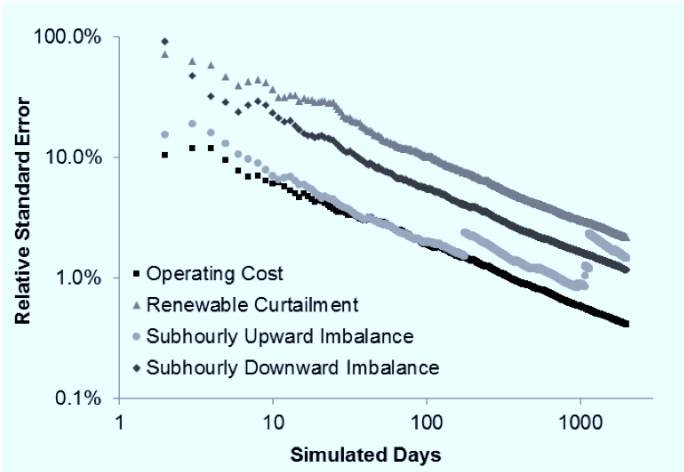


Fig. 5. Convergence behavior of the key model outputs in the 50% RPS case study.

TABLE V
ECONOMIC ANALYSIS OF THE VALUE OF FLEXIBILITY IN THE TEST UNITS. ALL VALUES ARE LISTED IN 2012\$. UNCERTAINTIES LISTED ARE EQUAL TO THE STANDARD ERROR CALCULATED FROM THE SIMULATED DISTRIBUTION OF EACH CLASS OF COSTS.

Expected Cost ($million/yr)	**Test Units Inflexible**	**Test Units Flexible**	**Cost Savings**
Operations (All Units)	$8,403 \pm 34$	$8,287 \pm 35$	116.1 ± 49
Operations (Excluding Test Units)	$7,961 \pm 34$	$7,965 \pm 35$	-4.3 ± 48.9
Renewable Curtailment	$2,068 \pm 44$	$1,756 \pm 41$	312 ± 60
Subhourly Upward Imbalance (EFD_t^+)	193.3 ± 2.8	179.7 ± 2.7	13.5 ± 3.9
Subhourly Downward Imbalance (EFD_t^-)	159.1 ± 1.8	133.2 ± 1.6	25.9 ± 2.4
Total Cost (Excluding Test Units)	$10,381 \pm 31$	$10,034 \pm 29$	347 ± 42

month of the year and this process was repeated until 2,000 days were drawn in total. After simulating operations on each day, the complete set of days was used to calculate expected annual values for each of the listed model outputs, assuming equal weighting of all the days.

The updated relative standard error of the expected daily model outputs (standard error divided by the sample mean) was also calculated after each simulation day. This is shown in Figure 5 as a function of the number of simulated days. The analysis shows that while the expected operating cost can be approximated to within 1% in less than 1,000 days for the modeled system, renewable curtailment and subhourly imbalances require longer simulations to converge. In particular, phenomena that are strongly penalized in the model, like upward reserve shortages, may occur infrequently and therefore require simulation of significantly more days to achieve the desired precision. Despite this observation, the precision required for each model output will depend on the application. In the application described below, only the total production cost (including operating cost, renewable curtailment cost, and subhourly imbalance penalties) is required to make a determination of flexible resource value. Expected values and standard errors are summarized for the primary model outputs in Table IV.

To investigate the economics of flexible resource procurement, the case was also run with the 1,088 MW of inflexible units replaced with 1,088 MW of generic new flexible combined cycle resources (ie. two identical 544 MW units with minimum up and down times of one hour, minimum stable levels of 130MW, and the ability to ramp up to maximum output in less than an hour). The model was run with both the inflexible and flexible versions of these "Test Units" to approximate the value of the flexibility provided by these resources. The economic analysis, which is summarized in Table V suggests that the flexibility of the Test Units provides up to $347million/yr of value, with an uncertainty of $42million/yr based on the standard errors of each model output. Note that this value does not include any operational cost savings of the test units themselves, only cost savings through avoided curtailment, improved subhourly operations, and more efficient operation of the rest of the generation fleet. For this case, approximately 90% of the flexibility value is in avoided renewable curtailment, approximately 10% of the value is associated with avoiding subhourly imbalances, and there is no value associated with more efficient operation of the rest of the generator fleet.

TABLE IV
KEY MODEL OUTPUTS FOR 50% RPS CASE STUDY. ALL DOLLARS ARE LISTED IN 2012\$.

Model Output	**Expected Value**	**Standard Error**
Operating Cost ($million/yr)	8,403	34.3
Renewable Curtailment (GWh/yr)	13,786	294
EFD_t^+ (GWh/yr)	4.96	0.072
EFD_t^- (GWh/yr)	1,061	12.2

In future analyses, this approach can be used to quantify the relative flexibility value of a wide range of generating resources. A system planner could, for example, select the resource or portfolio of resources with the lowest net cost (fixed costs minus flexibility value as calculated above). The completely inflexible units used in this case serve as an economically unintuitive resource, but may be a useful baseline that can be commonly compared to various procurement options. An important note regarding this type of analysis is that comparing the flexibility value of resources with similar operating constraints and costs using this method may require a substantial number of simulation days to ensure appropriate levels of precision. The requisite number of simulation days must be determined on a case-by-case basis.

The case study also highlights areas for further investigation in the future. In this case, three additional hours were modeled both before and after each simulation day to lessen the effects of edge constraints on the dispatch in the simulation day. In other analyses, this period has been increased to a full day on either side [26], however a more rigorous analysis of the impact of increasing the length of these buffers has not yet been undertaken. An additional reason for increasing the length of each simulation relates to the modeling of resources that are constrained over periods longer than a day. Modeling hydropower, for example, typically requires an energy budget constraint and imposing this constraint on a daily level potentially over-constrains the flexibility of the hydropower fleet, which has some ability to shift generation between days if needed. Prior production simulation analyses have typically used a simulation length of one week to account for inter-day flexibility, but the authors know of no analysis that specifically investigates the impacts of varying the simulation length. This analysis is beyond the scope of the current paper, but will be an important line of research as production simulation methods continue to inform planning decisions.

VI. Conclusions

This paper presents a new methodology for analyzing the economics of flexibility solutions in systems with high penetrations of renewables. The model combines a stochastic treatment of loads and resources with an adapted production simulation formulation that trades the cost of holding additional reserves against the cost of experiencing flexibility violations. The utility of the model was illustrated with a case study of California with a 50%RPS in 2030. The model was used to test the value of 1,088 MW of flexible units in the resource stack, relative to the same capacity of must-run resources. Comparison of two runs of the model in which only the operating characteristics of the test units were adjusted yielded an expected annual value of $347±42 million/yr. In real systems, the cost trade-offs will likely be between various options, including flexible CTs, energy storage, flexible loads, and new market mechanisms. It is also likely that some systems that have relatively inflexible existing resources reach lowest cost by procuring capacity beyond what would be identified in the traditional capacity planning paradigm. This complexity will complicate the identification of a single least cost portfolio. However, the modeling methodology presented in this analysis will be a useful means of comparing between specific procurement options.

In addition to procurement and development of new market mechanisms, renewable dispatchability via curtailment was identified as a potential flexibility solution. The extent to which the system operator relies on this solution will depend on the value of delivering renewable energy relative to the cost of experiencing flexibility violations and the costs of other operational flexibility solutions. These trade-offs suggest that quantification of the social cost of renewable curtailment will be an important step in integrating larger shares of renewable resources on to the grid. Furthermore, agreement among stakeholders regarding the costs of reliability and flexibility violations may be crucial to modeling operations in future systems that seek to reliably and economically integrate renewables.

The model presented in this paper has been commercialized under the name REFLEX (Renewable Energy FLEXibility model). For large systems, the REFLEX approach has been implemented on commercial production simulation platforms including ProMaxLT™ and PLEXOS® to improve scalability and runtimes as well as to incorporate more advanced modeling options. The formulation can be extended to include new functionality on a system-by-system basis. In systems with binding transmission constraints, the formulation can be adapted for zonal or nodal treatment. However, additional model complexity leading to longer runtimes must ultimately be weighed against the improvement in model accuracy.

At a high level, this analysis highlights the need for new modeling tools to plan for systems with increasing levels of renewable generation. The electric power industry, which has benefited from operational experience spanning many decades, is now faced with a rapidly changing energy supply landscape. In the coming years, more work will need to be devoted to planning for a low-carbon grid. This will require ongoing model development, benchmarking to new operational experience, and flexibility on the part of all stakeholders in adapting to new planning paradigms.

Acknowledgments

The authors would like to thank Rod Frowd at ECCO International for running the simulations described in the case study.

References

[1] A. Von Meier, "Integration of renewable generation in California: Coordination challenges in time and space," in *2011 11th International Conference on Electrical Power Quality and Utilisation (EPQU)*, 2011, pp. 1–6.

[2] EnerNex Corporation, "Eastern Wind Integration and Transmission Study," National Renewable Energy Laboratory, Subcontract Report NREL/SR-550-47086, 2010.

[3] GE Energy, "Western Wind and Solar Integration Study," National Renewable Energy Laboratory, Subcontract Report NREL/SR-550-47781, 2010.

[4] CAISO, "Integration of renewable resources: Operational requirements and generation fleet capability at 20% RPS," Tech. Rep., 2010. [Online]. Available: http://www.caiso.com/2804/2804d036401f0.pdf

[5] L. Soder and H. Holttinen, "On methodology for modelling wind power impact on power systems," *Int. J. of Global Energy Issues*, vol. 29, no. 1/2, pp. 181–198, 2008.

[6] J. Smith, M. Milligan, E. DeMeo, and B. Parsons, "Utility wind integration and operating impact state of the art," *IEEE Trans. Power Systems*, vol. 22, no. 3, pp. 900–908, 2007.

[7] E. Ela, M. Milligan, B. Parsons, D. Lew, and D. Corbus, "The evolution of wind power integration studies: Past, present, and future," in *2009 IEEE Power Energy Society General Meeting*, 2009, pp. 1–8.

[8] E. Lannoye, M. Milligan, J. Adams, A. Tuohy, H. Chandler, D. Flynn, and M. O'Malley, "Integration of variable generation: capacity value and evaluation of flexibility," in *IEEE Power and Energy Society General Meeting*, Jul. 2010.

[9] E. Lannoye, D. Flynn, and M. O'Malley, "Evaluation of power system flexibility," *IEEE Trans. Power Systems*, vol. 27, no. 2, pp. 922–931, 2012.

[10] S. A. Newell, K. Spees, J. Pfeifenberger, I. Karkatsouli, N. Wintermantel, and K. Carden, "Estimating the economically optimal reserve margin in ERCOT," http://www.brattle.com/system/news/pdfs/000/000/613/original/Estimating_the_Economically_Optimal_Reserve_Margin_in_ERCOT.pdf, Technical Report, 2014.

[11] Astrape, "The economic ramifications of resource adequacy," http://www.naruc.org/grants/Documents/Economics%20of%20Resource%20Adequacy%20WhitePaper_Astrape_Final.pdf, NARUC, White Paper, 2013.

[12] M. J. Sullivan, M. Mercurio, and J. Schellenberg, "Estimated value of service reliability for electric utility customers in the united states," http://certs.lbl.gov/pdf/lbnl-2132e.pdf, Ernest Orlando Lawrence Berkeley National Laboratory, Tech. Rep. LBNL-2132E, 2009.

[13] E. Ela, M. Milligan, and B. Kirby, "Operating reserves and variable generation," NREL, Technical Report NREL/TP-5500-51978, 2011. [Online]. Available: http://www.nrel.gov/docs/fy11osti/51978.pdf

[14] M. Ortega-Vazquez and D. Kirschen, "Estimating the spinning reserve requirements in systems with significant wind power generation penetration," *IEEE Trans. Power Systems*, vol. 24, no. 1, pp. 114–124, 2009.

[15] J. Morales, A. Conejo, and J. Perez-Ruiz, "Economic valuation of reserves in power systems with high penetration of wind power," *IEEE Trans. Power Systems*, vol. 24, no. 2, pp. 900–910, 2009.

[16] T. Yong, R. Entriken, and P. Zhang, "Reserve determination for system with large wind generation," in *Power Energy Society General Meeting*, 2009, pp. 1–7.

[17] F. Bouffard and F. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE Trans. Power Systems*, vol. 23, no. 2, pp. 306–316, 2008.

[18] C. Wang and S. M. Shahidehpour, "Effects of ramp-rate limits on unit commitment and economic dispatch," *Power Systems, IEEE Transactions on*, vol. 8, no. 3, pp. 1341–1350, 1993.

[19] D. Kirschen, A. Rosso, J. Ma, and L. Ochoa, "Flexibility from the demand side," in *Power and Energy Society General Meeting, 2012 IEEE*, 2012, pp. 1–6.

[20] A. Olson, R. Jones, E. K. Hart, and J. Hargreavest, "Optimal investment in flexible capacity," www.ethree.com, In Preparation, 2013.

[21] M. L. Telson, "The economies of alternative levels of reliability for electric power generation systems," *Bell Journal of Economics*, vol. 6, no. 2, pp. 679–694, 1975.

[22] M. Munasinghe and M. Gellerson, "Economic criteria for optimizing power system reliability levels," *Bell Journal of Economics*, vol. 10, no. 1, pp. 353–365, 1979.

[23] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems*. Plenum Press, 1996.

[24] M. Milligan, P. Donohoo, D. Lew, E. Ela, B. Kirby, H. Holttinen, E. Lannoye, D. Flynn, M. O'Malley, N. Miller, P. B. Eriksen, A. Gøttig, B. Rawn, M. Gibescu, E. G. Lázaro, A. Robitaille, and I. Kamwa, "Operating reserves and wind power integration: An international comparison," in *9th Annual International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants Conference*, no. NREL/CP-550-49019, October 2010, Conference Paper.

[25] Energy & Environmental Economics, Inc., "Investigating a higher Renewables Portfolio Standard in California," https://ethree.com/documents/E3_Final_RPS_Report_2014_01_06_with_appendices.pdf, Tech. Rep., 2014.

[26] CAISO, E3, "Renewable energy flexibility (REFLEX) results," http://www.caiso.com/Documents/RenewableEnergyFlexibilityResults-Final_2013.pdf, California Independent System Operator, Webinar, 2013.