

# Scenario Reduction for Stochastic Unit Commitment with Wind Penetration

Yonghan Feng and Sarah M. Ryan  
Industrial and Manufacturing Systems Engineering  
Iowa State University  
Ames, IA, USA

**Abstract**— Uncertainties in the day-ahead forecasts for load and wind energy availability are considered in a reliability unit commitment problem. A two-stage stochastic program is formulated to minimize total expected cost, where commitments of thermal units are viewed as first-stage decisions and dispatch is relegated to the second stage. Scenario paths of hourly loads are generated according to a weather forecast-based load model. Wind energy scenarios are obtained by identifying analogue historical days. Net load scenarios are then created by crossing scenarios from each set and subtracting wind energy from load. A new heuristic scenario reduction method termed forward selection in recourse clusters (FSRC) is customized to alleviate the computational burden. Results of applying FSRC are compared with those of a classical scenario reduction method, fast forward selection (FFS) by evaluating the expected dispatch costs when the commitment decisions derived from each subset of scenarios are applied to the whole scenario set. In an instance down-sampled from data of an Independent System Operator in the U.S., the expected dispatch costs for both scenario reduction methods are similar, but FSRC improves reliability.

**Index Terms**-- Stochastic programming, Scenario reduction, Unit commitment.

## I. INTRODUCTION

The unit commitment (UC) problem identifies on/off decisions and generation levels of thermal units over a planning horizon to satisfy a load forecast, while minimizing total cost consisting of startup cost and generation (no-load and fuel) cost under restrictions on unit operation and transmission capacity. With the increasing penetration of renewable energy, such as wind and solar, uncertainties occur not only in the demand side because of load forecast errors, but also in the supply side because of the intermittence of renewable energy and the difficulty of predicting it hours in advance.

Two-stage stochastic programming is a promising approach for solving UC under uncertainty. In papers such as [1], [2], commitment of units is decided in the first stage before the real time information is realized, while decisions on dispatch amounts are delayed until the second stage after realizing actual load and the availability of uncertain resources. A number of scenarios will be created to represent future possible realizations. However, it is computationally

expensive to directly solve a stochastic program based on such large set of scenarios, even if decomposition methods, such as Benders decomposition [3] or a progressive hedging algorithm [4] is used. The size of the Benders master problem will increase dramatically if many scenarios are included. Progressive hedging requires some heuristic strategies to improve convergence if integer decision variables appear in the first-stage, as mentioned in [5]. Therefore, reducing the number of scenarios with good approximation becomes an attractive way to alleviate the computational burden. Over the past decade, several papers have discussed ways to select a modest number of scenarios from a large set. Among them, [6] derives scenario reduction approaches from probability metrics, and faster variants were proposed by [7]. However, these scenario reduction methods may not well identify which scenarios significantly influence the commitment of units in a power system with high wind penetration [8]. One possible reason is that these methods only account for transformation between probability metrics, but ignore the context of scenarios in a stochastic program. Instead of only focusing on the parameters of scenarios, [9] proposed a heuristic scenario reduction method, FSWC, a specific version of FSRC, to incorporate the impact from each scenario on first-stage decisions. This paper will discuss a customization strategy for FSRC applied to the UC problem with wind energy to provide a more reliable commitment of units even when only a modest number of scenarios can be employed in the two-stage stochastic program. We assess the performance of competing scenario reduction methods by evaluating the costs and reliability measures of the resulting commitment decisions against the whole set of scenarios.

The remainder of this paper is organized as follows. Section II briefly introduces a two-stage stochastic program for unit commitment. Section III describes the process to generate net load scenarios for day-ahead stochastic unit commitment. Section IV discusses a strategy to customize FSRC and Section V shows numerical results for a case study. We present conclusions in Section VI.

## II. STOCHASTIC UNIT COMMITMENT MODEL

Notation:

$T$ : Set of time periods  
 $B$ : Set of buses

---

This material is based on work sponsored by the U.S. Department of Energy under the ARPA-E Green Energy Network Integration (GENI) Program.

$S$ :	Set of scenarios
$\mathcal{G}$ :	Set of thermal units
$\pi_s$ :	Probability of scenario $s$
$D_{bts}$ :	Demand at bus $b$ in period $t$ in scenario $s$ (MWh)
$\zeta_s$ :	Second stage objective value of scenario problem $s$ (\$)
$\Gamma^+, \Gamma^-$ :	Penalties on shortage and excess respectively, (\$/MWh)
$p_g^{max}, p_g^{min}$ :	Max/min output of unit $g$ , (MW)
$c_{gt}^u(\cdot)$ :	Startup cost of unit $g$ in period $t$ (\$)
$c_{gts}^p(\cdot)$ :	Generation cost of unit $g$ in period $t$ in scenario $s$ (\$)
$v_{gt}$ :	Binary variable, 1 if unit $g$ is on, 0 otherwise in period $t$
$p_{gts}$ :	Decision variable, generation level of unit $g$ in period $t$ in scenario $s$ (MW)
$\bar{p}_{gts}$ :	Decision variable, maximum available generation level of unit $g$ in period $t$ in scenario $s$ (MW)
$\alpha_{bts}^+$ :	Decision variable, shortage at bus $b$ in period $t$ in scenario $s$ (MWh)
$\alpha_{bts}^-$ :	Decision variable, excess at bus $b$ in period $t$ in scenario $s$ (MWh)

Our formulation of stochastic reliability unit commitment (SRUC) follows the deterministic UC modeling strategy in [10], and extends it to two-stage stochastic program [5].

#### A. Objective function

$$\min \sum_{t \in T} \sum_{g \in \mathcal{G}} c_{gt}^u(v_{gt}) + \sum_{s \in S} \pi_s \zeta_s \quad (1)$$

The objective function (1) consists of two parts: the cost related to commitment of units, the first-stage decisions; and the cost related to generation and penalties for load imbalance, the second-stage decisions, upon realization of a scenario, as shown in (2).

$$\zeta_s = \sum_t (c_{gts}^p(p_{gts}) + \Gamma^+ \alpha_{bts}^+ + \Gamma^- \alpha_{bts}^-) \quad (2)$$

Generation cost  $c_{gts}^p(\cdot)$  in (2) is assumed to be a piece-wise linear function of generation level  $p_{gts}$  in this paper.

#### B. Constraints

$$\sum_{g \in \mathcal{G}(b)} p_{gts} + \alpha_{bts}^+ - \alpha_{bts}^- = D_{bts} \quad \forall b \in B, t \in T, s \in S \quad (3)$$

$$p_g^{min} v_{gt} \leq p_{gts} \leq \bar{p}_{gts} \leq p_g^{max} v_{gt} \quad \forall g \in \mathcal{G}, t \in T, s \in S \quad (4)$$

$$p_{gts}, \bar{p}_{gts}, \alpha_{bts}^+, \alpha_{bts}^- \geq 0, v_{gt} \in \{0, 1\} \quad \forall b \in B, g \in \mathcal{G}, t \in T, s \in S \quad (5)$$

Formula (3) is the demand balance constraint, while (4) represents operational constraints. In addition, each thermal unit must satisfy other constraints mentioned in [10], such as minimum up/down time constraints, ramp limitations and so on.

### III. SCENARIO GENERATION

In this paper, fluctuations in load and intermittence of wind energy are uncertainties represented by scenarios.

Load scenario generation follows an idea of approximation rather than sampling, starting from a day-ahead weather forecast. For details of the load scenario generation process depicted in Fig. 1, see [11].

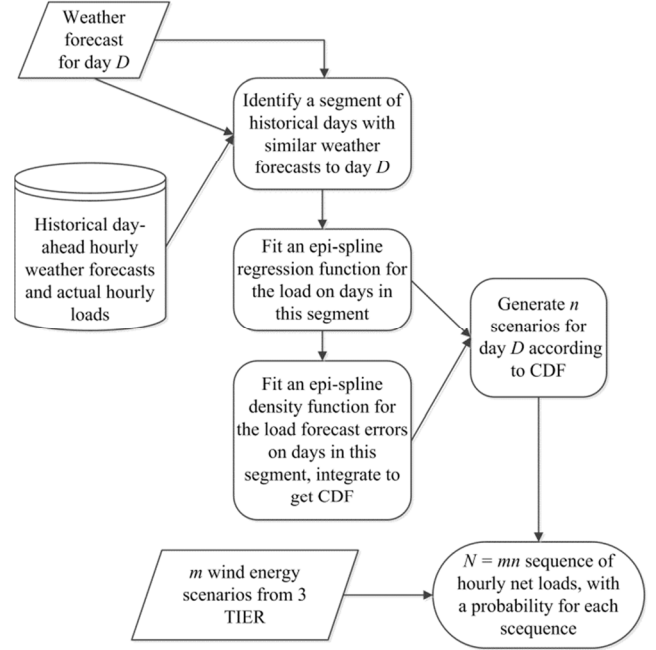


Fig.1 Scenario generation procedure

Hourly wind scenarios were obtained from a commercial vendor [12], according to their analogue method. In this paper, wind energy is assumed to be nondispatchable generation, and thereby considered as negative load. Net loads, obtained by subtracting wind energy from load, are then viewed as demands in model (1)-(5). Net load scenarios result from crossing load scenarios and wind energy scenarios.

### IV. SCENARIO REDUCTION

As already discussed in Section III, a finite number of scenarios have been generated to describe the uncertain parameters in the stochastic program. However, the computational effort strongly depends on the number of scenarios even if decomposition techniques, like progressive hedging [5] or Benders decomposition [3], will be applied to solve the stochastic mixed integer program. Therefore, it is natural to explore methods to approximate the generated scenarios with a modest sized subset of scenarios while keeping main features as a good approximation. Analysis and discussion of scenario reduction methods will be presented in this section.

#### A. Fast Forward Selection (FFS)

Over the past decades, several scenario reduction methods have been developed based on random sampling, clustering and analysis of probability metrics. The most widely used scenario reduction methods include forward selection (FS), backward reduction (BR) and their variants, such as FFS [7], which are based on analysis of the Fortet-Mourier probability metric. The FFS method is usually applied to select a modest-sized subset of scenarios  $S'$  from the original set of scenarios  $S$

because numerical results in [6], [7] indicate that FFS yields a reduced set more similar in distribution to the original set than BR does when the reduction is substantial. The FFS method aims to minimize the distance between a subset  $S'$  of the prescribed size and the remaining scenarios in  $S \setminus S'$ . Because exact computation of distance between scenario sets according to probability metrics is a hard problem, FFS is a tractable way to minimize an upper bound on the distance instead.

FFS only accounts for the scenario parameters  $S$  and corresponding probabilities  $\pi_s$ , but not for the possible influences from scenarios on decision variables. We conjecture that better performance could be achieved by considering the scenarios' impact on decisions and costs in the selection process. Therefore, we present a heuristic scenario method called forward selection in recourse clusters (FSRC) [9], which not only considers distances between selected scenarios and the remaining scenarios, but also measures the influences from scenarios on decision variables.

### B. Forward Selection in Recourse Clusters (FSRC)

To measure the impacts from each scenario on decision variables, a scenario subproblem is defined for each scenario  $k$  by setting scenario probability  $\pi_k = 1$  and  $\pi_i = 0, \forall i \neq k$ . By solving each scenario subproblem, values of an optimal solution corresponding to each scenario will be realized and can be customized to reveal features of each scenario as discussed below. Before introducing the FSRC algorithm, we discuss how to measure scenario impacts on decisions. A solution sensitivity index is created from the decision variables in a mathematical program or part of the objective function that could quantify differences among scenarios. For instance, in a UC problem, the hourly on/off status of each unit through the whole schedule horizon may be different for each scenario; thus, these decision variables could be considered as a solution sensitivity index. However, if there are hundreds of generators, considering the whole decision vector would be unwieldy, and may allow features of scenarios to be blurred because of the inherent difficulties in high dimensional data analysis. Instead, total generation cost could serve to distinguish among scenarios because higher demand often results in higher production cost. Further discussions on forming solution sensitivity indices will be presented in Section III.C.

For a generic stochastic program, the FSRC method is given as follows.

#### Forward selection in recourse clusters (FSRC):

Suppose the prescribed cardinality of selected scenarios is set to  $n$ , then

1. Define solution sensitivity indices, and get values of solution sensitivity indices vector  $\mathcal{N}^s$  by solving the scenario subproblem for each  $s \in S$ ;
2. Scale elements of vectors  $\{\mathcal{N}^s\}$  into similar magnitudes, denoted as  $\{V^s\}$ ;
3. Form  $n$  clusters of  $\{V^s\}$  by  $k$ -means method based on an appropriate norm (e.g.  $L_2$  norm), and create the corresponding  $n$  clusters of original scenarios at the same time;

4. Use the FFS method to select one scenario from each cluster of original scenarios.

The above algorithm presents the general process of FSRC. Because characterization of scenario impact on decision variable is often problem-dependent, it is necessary to customize FSRC for different applications. The customization specifies how to identify solution sensitivity indices, and then create clusters according to them. Discussion on customization strategies of FSRC for the SRUC problem follows.

### C. Customization Strategy of FSRC

Evaluation of similarities among scenarios is the engine to start FSRC. One of the intuitive ways is to solve a deterministic UC problem for each scenario directly, and get related values to create solution sensitivity indices, as in [9]. However, this strategy will lead to time consuming implementation when each scenario subproblem consists of a large mixed integer linear program (MILP). Instead of directly evaluating solutions of deterministic scenario subproblems, measuring the relative performance for each individual scenario given fixed first-stage decisions may suffice to distinguish among scenarios. Because the net load is the only uncertain factor in this model, generation cost, excess  $\alpha_{bts}^-$  and shortage  $\alpha_{bts}^+$  will reveal how hard it is to satisfy net loads in each scenario with the given UC strategy, and therefore directly distinguish among scenarios.

Following this computational strategy, processing a UC problem reduces to solving a dispatch problem that is a pure linear program given the UC strategy. Hence, a UC strategy will be obtained first by solving a deterministic UC problem for expected net load, and then this UC strategy will be applied to each scenario to get values of second-stage decision variables and second-stage costs, which will be further used to form solution sensitivity indices.

Creating solution sensitivity indices follows solving the sequence of dispatch problems. Notice that higher net loads often require more generators to be set on, and as a direct consequence, higher generation level and potential higher production cost will be realized; otherwise, fewer generators will be committed, and lower generation levels and production cost will result. Because the UC strategy (first-stage decision) is fixed for each scenario subproblem, the hourly generation level, generation cost of each generator and hourly generation mismatches through the schedule horizon will be the elementary entries to create solution sensitivity indices. To reduce the dimension we compute total generation cost, total excess and total shortage of a scenario subproblem. Once solution sensitivity indices have been created, a clustering algorithm will be applied on the solution sensitivity indices to cluster scenarios. Since different magnitudes of values may distort appropriate clustering, all values in solution sensitivity indices must be scaled to similar magnitudes. In addition, excess and shortage are weighted differently according to their effects on the power system. Excess could be alleviated by de-committing generators; e.g., curtailing renewable energy generation, and charging storage devices, such as batteries and pumped-storage hydro plants. Shortage will require more electric power to be transmitted from other areas, or even load

curtailment, which could impose high costs. Thus, shortage is assigned a higher weight in the clustering procedure.

From the above discussions, elements comprising solution sensitivity indices consist of total generation cost, excess and shortage for each scenario. The customization strategy of the scenario reduction method is summarized as the following.

*Customization of step 1 in FSRC:*

1. Compute cumulative generation cost through the whole scheduling horizon over all generators in scenario  $s$ , denoted as:  $C_s = \sum_t \sum_g c_{gts}^p(p_{gts})$ ;
2. Calculate total excess and shortage separately over all periods for each scenario  $s$  denoted as  $\Lambda_s^- = \sum_t \sum_b \alpha_{bts}^-$  and  $\Lambda_s^+ = \sum_t \sum_b \alpha_{bts}^+$ .

*Customization of step 2 in FSRC:*

1. Obtain average generation cost  $\bar{C}_s$  over all scenarios, and scale each scenario generation cost by dividing  $\bar{C}_s$ , as  $\mathfrak{C}_s = C_s/\bar{C}_s, \forall s \in S$ ;
2. Obtain average value  $\bar{\Lambda}$  over all nonzero generation mismatches  $\Lambda_s^+$  and  $\Lambda_s^-$  through all scenarios, scale them as  $\Theta_s^+ = \Lambda_s^+/\bar{\Lambda}$  and  $\Theta_s^- = \Lambda_s^-/\bar{\Lambda}$ ;
3. Set weights for scaled generation cost and scaled generation mismatches, as  $w_c, w_+$  and  $w_-$ ;
4. Create solution sensitivity indices vector  $V_s$ , where  $V_s = [w_c \mathfrak{C}_s, w_+ \Theta_s^+, w_- \Theta_s^-]$ ;

*Customization of step 3 in FSRC:*

1. Use the  $L_2$  norm in the  $k$ -means method.

Although transmission constraints are not included in this paper, the proposed scenario reduction procedure can be extended to that case by grouping buses in specified zones together, and using cumulated shortage, excess and generation cost over each group as solution sensitivity indices. The dimension of the sensitivity index vector  $V_s$  increases accordingly. If wind generators are dispatchable then in (3),  $\mathcal{G}(b)$  includes wind units and  $D_{bts}$  represents load rather than net load. In constraint (4), we can fix  $v_{gt} = 1$  for any wind plant  $g$ , and allow  $p_g^{max}$  to vary by scenario. In this case, the excess amounts  $\alpha_{bts}^-$  are likely to be much smaller overall in the UC evaluation because wind spillage will reduce the impact of underestimating wind power on the day ahead.

## NUMERICAL RESULTS

The customized FSRC method is applied to a modified test system down-sampled from the Independent System Operator of New England (ISO-NE) during a summer week. Hourly zonal load data were collected from ISO-NE to generate day-ahead total hourly load scenarios. All 8 zones in ISO-NE were treated as a single bus in this case study. 50 hourly wind energy scenarios for each day were obtained from a commercial vendor [12], according to the analogue method [13], and these wind scenarios were designed to correspond to a ‘‘scenario’’ representing 20% penetration of wind energy in 2024 in [14]. We scaled the total hourly load scenarios for ISO-NE in 2011 according to a 2.27% increase per year as

mentioned in [14] to approximate demand levels in 2024. This section will compare scenario reduction effects of FSRC to FFS. Such investigation starts from applying UC strategies obtained by solving the SRUC problems with selected scenarios to the whole set of scenarios, solving a dispatch problem for each scenario. Because of physical limitation on RAM, 20 generators are selected to keep computation manageable. In addition, the 10 highest probability wind energy scenarios are selected from 50 wind energy scenarios to cross with 8 load scenarios which have been generated as in [11], forming 80 hourly net load scenarios. The net load scenarios were scaled down to match the reduced generation capacity. As mentioned in Section IV.C, more effort is taken to avoid shortage, and to emphasize the effect of generation shortage during clustering. Therefore, penalties set on shortage and excess are  $10^7$  \$/MWh and  $10^5$  \$/MWh respectively, which are four and two, respectively, orders of magnitude larger than the marginal cost of the most expensive unit. In addition, weights  $w_c, w_+$  and  $w_-$  were set to 0.3, 0.4 and 0.3, respectively.

Net load scenarios of a single summer day are presented in Fig.2, and only 20 scenarios selected from FFS and FSRC methods are displayed in Fig.3 and Fig.4, respectively, because of limited space.

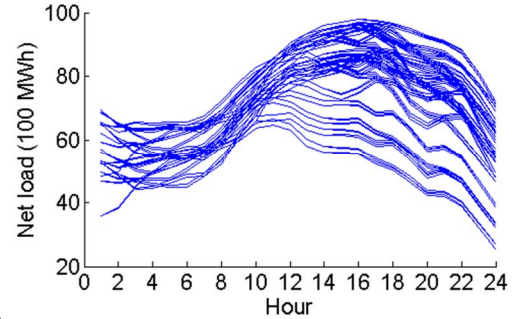


Fig.2 All 80 net load scenarios

UC strategies obtained from solving selected scenarios based SRUC are applied to the whole set of scenarios to get resulting expected dispatch costs and load imbalances of the scenario reduction methods. These expected values with respect to all scenarios for FFS and FSRC are accumulated through the summer week, and savings of FSRC in cumulative expected costs (including penalties) compared to FFS are plotted in Fig.5 for each cardinality,  $n$ , of the selected scenario sets. According to Fig.5, expected costs from FSRC are less than those from the FFS method. Savings in load imbalance of FSRC from FFS are plotted for each  $n$  in Fig.6 over the week. From this figure, FSRC results in less shortage than FFS does, and provides similar levels of excess generation. Comparing Fig.6 to Fig.5, it seems that shortage penalty mostly accounts for the differences shown in Fig.5. In this sense, FSRC selects scenarios in line with the decision maker’s concern to minimize shortages.

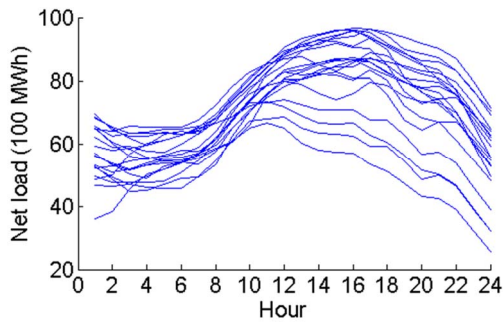


Fig.3 20 scenarios selected by FFS

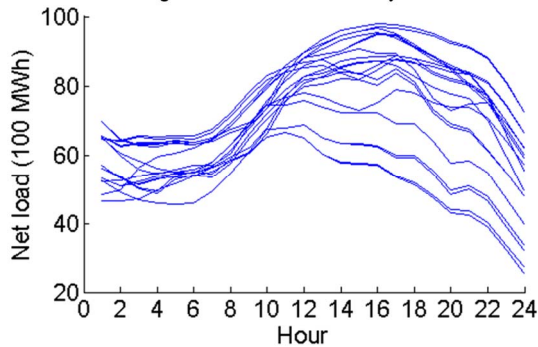


Fig.4 20 scenarios selected by FSRC

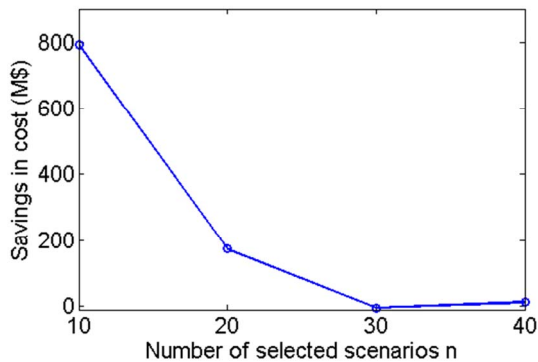


Fig.5 Savings in cost of FSRC from FFS

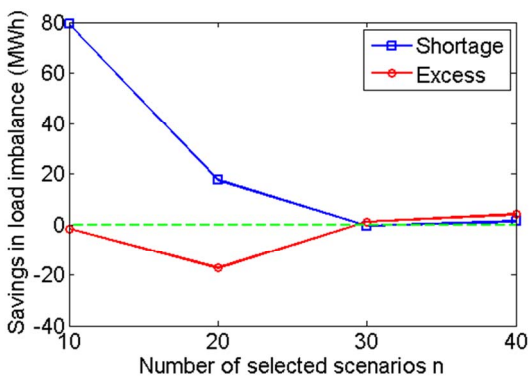


Fig.6 Savings in load imbalance of FSRC from FFS

## V. CONCLUSION

In this paper, a customization strategy of FSRC is presented, and applied to investigate its performance. Compared to the classical scenario reduction method FFS, the customized FSRC pays more attention to the performance aspects on which the decision maker focuses, and thereby leads to less shortage. From the perspective of the relationship between supply and demand, the FSRC method seems to perform more reliably than FFS does. In addition, FSRC will give a more economic schedule when shortage is given a higher cost penalty, compared to FFS. In ongoing work more numerical study will be performed on day-ahead SRUC, including larger sets of scenarios for full-scale power systems, and other necessary strategies of customizing the FSRC method will be investigated to improve its performance.

## REFERENCES

- [1]. E. Constantinescu, V. Zavala, M. Rocklin, S. Lee and M. Anitescu. "Unit commitment with wind power generation: integrating wind forecast uncertainty and stochastic programming," <http://www.mcs.anl.gov/papers/TM-309A.pdf>.
- [2]. A. Papavasiliou and S. Oren. "Multi-area stochastic unit commitment for high wind penetration in a transmission constrained network," *Operations Research*, vol.61, no.3 pp. 578-592, 2013.
- [3]. Q. Zheng, J. Wang, P. Pardalos and Y. Guan. "A decomposition approach to the two-stage stochastic unit commitment problem," *Annals of Operations Research*, vol.210, no.1, pp. 387-410, 2013.
- [4]. S. Takriti, J. Birge and E. Long. "A stochastic model for the unit commitment problem," *IEEE Trans. Power Syst.* vol.11, no. 3, pp.1497-1508, 1996.
- [5]. K. Cheung, D. Gade, C. S. Monroy, S. Ryan, J. Watson, R. Wets and D. Woodruff. "Toward scalable stochastic unit commitment part 2: assessing solver performance," *under review*.
- [6]. J. Dupacova, N. Growe-Kuska and W. Romisch. "Scenario reduction in stochastic programming – an approach using probability metrics," *Math. Program.* vol.95 no.3, pp. 493-511, 2003.
- [7]. H. Heitsch and W. Romisch. "Scenario reduction algorithms in stochastic programming," *Computational optimization and applications*, vol.24, no. 2, pp. 187-206, 2003.
- [8]. A. Papavasiliou, S. Oren and R. O'Neill. "Reserve requirements for wind power integration: a scenario-based stochastic programming framework," *IEEE Trans. Power Syst.* vol.26, no. 4, pp.2197-2206, 2011.
- [9]. Y. Feng and S. Ryan. "Scenario construction and reduction applied to stochastic power generation expansion planning," *Computers & Operations Research*, vol.40, no.1, pp. 9-23, 2013.
- [10]. M. Carrion and J. Arroyo. "A computationally efficient mixed formulation for the thermal unit commitment problem," *IEEE Trans. Power Syst.* vol.21, no. 3, pp.1371-1378, 2006.
- [11]. Y. Feng, I. Rios, S. Ryan, K. Spurkel, J. Watson, R. Wets and D. Woodruff. "Toward scalable stochastic unit commitment part 1: load scenario generation," *under review*.
- [12]. 3TIER Inc. <http://www.3tier.com/en/>.
- [13]. W. Mahoney, K. Parks, G. Wiener, Y. Liu, W. Myers, J. Sun, L. Monache, T. Hopson, D. Johnson, and S. Haupt, "A wind power forecasting system to optimize grid integration," *IEEE Transactions on Sustainable Energy*, vol.3, no. 4, pp.670-682, 2012.
- [14]. National Renewable Energy Lab. "Eastern wind integration and transmission study," <http://www.nrel.gov/docs/fy11osti/47078.pdf>.