



## Discrete Optimization

# New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints

F. D. Munoz<sup>a,\*</sup>, B. F. Hobbs<sup>c</sup>, J.-P. Watson<sup>b</sup><sup>a</sup> Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile<sup>b</sup> Discrete Math and Optimization Department, Sandia National Laboratories, Albuquerque, NM 87123, USA<sup>c</sup> Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

## ARTICLE INFO

## Article history:

Received 5 August 2014

Accepted 29 July 2015

Available online 14 August 2015

## Keywords:

OR in energy

Stochastic programming

Benders decomposition

## ABSTRACT

We propose a novel two-phase bounding and decomposition approach to compute optimal and near-optimal solutions to large-scale mixed-integer investment planning problems that have to consider a large number of operating subproblems, each of which is a convex optimization. Our motivating application is the planning of power transmission and generation in which policy constraints are designed to incentivize high amounts of intermittent generation in electric power systems. The bounding phase exploits Jensen's inequality to define a lower bound, which we extend to stochastic programs that use expected-value constraints to enforce policy objectives. The decomposition phase, in which the bounds are tightened, improves upon the standard Benders' algorithm by accelerating the convergence of the bounds. The lower bound is tightened by using a Jensen's inequality-based approach to introduce an auxiliary lower bound into the Benders master problem. Upper bounds for both phases are computed using a sub-sampling approach executed on a parallel computer system. Numerical results show that only the bounding phase is necessary if loose optimality gaps are acceptable. However, the decomposition phase is required to attain optimality gaps. Use of both phases performs better, in terms of convergence speed, than attempting to solve the problem using just the bounding phase or regular Benders decomposition separately.

© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

## 1. Introduction

The electric power industry is a major applications area for optimization (Hobbs, 1995). This sector comprises over 2 percent of the U.S. economy, and recent restructuring has strengthened incentives to plan and operate power infrastructure efficiently. Increasing amounts of generation from renewable resources makes optimization of short-term operations and long-term planning more challenging, thus promoting the development of new decision-support tools to account for renewable variability and unpredictability. For instance, stochastic unit commitment models that explicitly factor in uncertainty in the availability of supply from wind and solar generators often yield lower dispatch costs when compared to traditional deterministic unit commitment approaches. However, these

cost reductions come at the expense of higher computational complexity (Bertsimas, Litvinov, Sun, Zhao, & Zheng, 2013; Papavasiliou & Oren, 2013). Because they consider both investment and operations, investment planning models present even greater computational challenges. First, resource-specific characteristics, such as locational constraints and distance from load centers and the existing transmission grid, require analysis of both transmission and generation investment alternatives on a system-wide basis. Second, failure to capture the variability and spatial correlations among intermittent resources will likely result in suboptimal investment recommendations (Joskow, 2011). In this paper, we develop practical approaches to solve multi-area generation and transmission investment planning problems that account for the aforementioned challenges.

Because of computational limitations, as well as uncertainty in long-term forecasts of demand and capacity factors of intermittent resources, investment planning models have traditionally avoided fine-grained representations of short-run production costs (Palmintier & Webster, 2011). To achieve computational tractability, investment planning instead utilized deterministic or probabilistic models for calculating production costs based on load-duration

\* Corresponding author. Tel: +1 720 2520 424.

E-mail addresses: [fdmunoz@uai.cl](mailto:fdmunoz@uai.cl), [fdmunoz@sandia.gov](mailto:fdmunoz@sandia.gov), [elpanchomunoz@gmail.com](mailto:elpanchomunoz@gmail.com) (F.D. Munoz), [bhobbs@jhu.edu](mailto:bhobbs@jhu.edu) (B.F. Hobbs), [jwatson@sandia.gov](mailto:jwatson@sandia.gov) (J.-P. Watson).

curve approximations (Hobbs, 1995; Kahn, 1995). These models usually approximate the load or net-load duration curves using a small number of categories (e.g., peak, shoulder and off-peak demand), ignore spatial correlations between demand zones and intermittent generation across multiple regions, and do not model time dependencies in operations, thereby ignoring ramping constraints and start-up costs. Early planning models only considered single-area load duration curves based on time-series of historical and forecasted data (Anderson, 1972; Booth, 1972). These were later improved, e.g., through the use of Gram–Charlier series (Caramanis, Tabors, Nochur, & Schweppe, 1982), to account for the effect of non-dispatchable generation technologies, such as wind and solar, on the optimal generation mix. A simple approach to the latter determines operating hours to be simulated via moment matching on demand, wind, solar, and hydro data (van der Weijde & Hobbs, 2012). In this approach, the sample of hours that best approximates the means, standard deviations, and correlations of the data is selected to determine the optimal portfolio of transmission and generation investments. Palmintier and Webster (2011), Shortt, Kiviluoma, and O'Malley (2013), and de Sisternes and Webster (2013) proposed further refinements of the use of load-duration curves in planning models considering unit commitment variables and constraints. Yet these were only applied to generation and not transmission planning. None of these approximation methods, however, provide metrics (e.g., bounds) to quantify the effect of the quality of the approximations on the resulting investment plans and total system costs. Therefore, they can only be deemed as heuristics.

Large-scale applications and computational limitations have historically motivated researchers to solve generation and transmission planning models using Benders decomposition (Bloom, 1983; Bloom, Caramanis, & Charny, 1984; Huang & Hobbs, 1994; Pereira, Pinto, Cunha, & Oliveira, 1985; Sherali & Staschus, 1990; Sherali, Staschus, & Huacuz, 1987). Such approaches separate the investment problem (i.e., master problem) from the production cost problems (i.e., subproblems), which can then be solved by independently taking advantage of parallel computer systems. The quality of the investment plans proposed by the master problem is improved by iteratively evaluating their performance against the production cost models, which also provide marginal cost information that is subsequently used in the master problem. Benders decomposition also provides bounds upon the optimal system costs for each candidate investment and its convergence is guaranteed under certain conditions (Geoffrion, 1972). However, these bounds cannot be guaranteed as valid if only a few observations of demand, wind, solar, and hydro data are considered in the subproblems, as is often the case in planning studies. Furthermore, convergence of the algorithm is often slow, which has prevented its widespread utilization among practitioners, although acceleration techniques have been an important subject of research (Magnanti & Wong, 1981; McDaniel & Devine, 1977; Sahinidis & Grossmann, 1991). Finally, consideration of environmental constraints, such as imposition of minimum annual amounts of generation from renewable resources, impedes the parallel solution of the subproblems. These constraints couple the solutions for distinct hours, which then all need to be considered simultaneously in the optimization problem. This imposes a computational restriction on the level of granularity of the market operations representation.

Transmission expansion is a particularly challenging aspect of infrastructure planning in power systems that has received significant attention from researchers. Latorre, Cruz, Areiza, and Villegas (2003), Lee, Ng, Zhong, and Wu (2006), and Hemmati, Hooshmand, and Khodabakhshian (2013) provide comprehensive surveys of the most significant advancements. Recent contributions in the area focus on enhancing models to account for uncertainty and optionality (Konstantelos & Strbac, 2014; Moreno, Pudjianto, & Strbac, 2013; Orfanos, Georgilakis, & Hatziargyriou, 2013). New

modeling approaches with potentially efficient solution algorithms for large-scale applications have been proposed by Lumbreras and Ramos (2013), Asadamongkol and Eua-arporn (2013), Desta Zahlay, De Cuadra, Olmos, Rivier, and Perez-Arriaga (2013), Ozdemir, Muñoz, Ho, and Hobbs (2015). However, most of these algorithms have only been applied to small test-cases or to large instances but with a very small number of dispatch scenarios.

In this paper, we develop a computationally-tractable algorithm to generate candidate transmission and generation investment plans, as well as bounds upon the minimum system costs. We propose a two-phase approach based on a bounding algorithm (Hobbs & Ji, 1999) and Benders decomposition, both of which provide bounds on the expected system costs. In Phase 1, a lower bound is computed by solving a low-resolution planning problem using clustered observations of time-dependent demand, wind, solar, and hydro data, based on an extension of Jensen's inequality for stochastic programs with expectation constraints. Upper bounds are estimated using a sub-sampling method to approximate the operations costs for each candidate investment plan proposed by the lower-bound planning problem. These bounds are progressively tightened by refining partitions of the space of time-dependent load and renewable energy data. Due to the asymptotic properties of our algorithm, however, tight optimality gaps may only be achieved in the limit, requiring very fine partitions of the data that result in computationally expensive lower-bound planning problems. To overcome this difficulty, we propose a second phase (Phase 2) to the bounding approach that uses Benders decomposition with an auxiliary lower bound to close the optimality gap. This is the first solution approach capable of solving multi-area transmission and generation planning problems with expectation constraints, while providing bounds on the optimal total system costs. We apply that algorithm to a realistic large-scale representation of a power system in the U.S. Our approach can be generalized to other stochastic programs with both per-scenario and expectation constraints, such as optimization problems with CVaR constraints in finance (Krokhmal, Palmquist, & Uryasev, 2002).

The rest of this paper is organized as follows. In Section 2, we describe an abstract planning model that is formulated as a stochastic mixed-integer linear program with per-scenario and expectation constraints. In Section 3, we extend Jensen's inequality in order to compute lower bounds for a stochastic problem with expected-value constraints and describe a statistical method to compute upper bounds that takes advantage of parallel computer systems. Section 4 describes our implementation of Benders decomposition, including the introduction of auxiliary lower bounds in the master problem to accelerate convergence. In Section 5, we illustrate the performance of the proposed bounding and decomposition algorithms on a transmission and generation planning study of a 240-bus representation of the Western Electricity Coordinating Council (WECC). The WECC is the largest synchronized power system in the U.S., comprising 14 western states as well as portions of Alberta, British Columbia, and Mexico. Conclusions are presented in Section 6. Proofs of all propositions together with details of derivations are provided in the electronic supplementary material.

## 2. Abstract planning model

We focus on investment planning models that can be formulated as linear or mixed integer linear programs.<sup>1</sup> Examples of such models include: Caramanis et al. (1982), Bloom (1983), and Sherali and Staschus (1990) for generation expansion planning; Binato, Pereira, and Granville (2001) for transmission expansion planning; and Pereira et al. (1985), Dantzig et al. (1989), van der Weijde and Hobbs

<sup>1</sup> Modeling unit commitment variables and constraints or AC optimal power flows yields non-linear and non-convex operations models. Their use in long-term investment models has been limited to research applications on small test-cases.

(2012), and Munoz, Hobbs, Ho, and Kasina (2014) for composite transmission and generation expansion planning. Other electricity investment planning market simulation models that are commonly used for energy and environmental policy analysis include IPM (ICF, 2013), the Electricity Market Module of NEMS (Gabriel, Kydes, & Whitman, 2001), ReEDS (Short et al., 2011), Haiku (Paul & Burtraw, 2002), and MARKAL (EIA, 2013).<sup>2</sup>

### 2.1. Notation

We now define the main notation used in the paper. Additional parameters and variables will be introduced as needed.

#### Parameters

$A$	Coefficient matrix associated with investment constraints
$b$	Right-hand-side vector associated with investment constraints
$c$	Vector of marginal generation and curtailment costs
$d$	Right-hand-side vector associated with expected-value constraints
$e$	Vector of transmission and generation capital costs
$K$	Fixed recourse matrix associated with expected-value constraints
$T(\omega)$	Coefficient matrix associated with investment variables in operations problem. Also known as a transition matrix. This matrix includes scenario- or time-dependent parameters such as hourly levels of wind, solar, and hydro power production
$(\Omega, p)$	Discrete probability space composed of the sample space $\Omega$ and the probability measure $p(\cdot)$ over $\Omega$ . For planning purposes, this space can, for example, be constructed using 8760 historical observations of hourly demand, wind, solar, and hydro data from a representative year (i.e., $ \Omega  = 8760$ ). Each event $\omega_i$ would then have probability of occurrence $p(\omega_i) = 1/8,760, \forall \omega_i \in \Omega$
$r(\omega)$	Right-hand-side vector of constraint parameters for scenario $\omega$
$W$	Fixed recourse matrix

#### Decision variables

$x$	Vector of generation and transmission investment variables. Some investment decisions are often modeled as binary (e.g., transmission investments) while others are modeled as continuous (e.g., generation investments)
$y(\omega)$	Vector of power generation levels, power flows, phase angles, and demand curtailment variables for each realization of $\omega$ in $\Omega$

### 2.2. Two-stage investment-operations model

The basic goal of a planning tool is to provide a recommendation of where and when to invest in new transmission and/or generation infrastructure, given a distribution of forecast operating conditions that we model with the probability space  $(\Omega, p)$ .<sup>3</sup> We formulate the

<sup>2</sup> There also exist more sophisticated modeling approaches based upon game theory that consider the strategic interactions between transmission and generation investments (Pozo, Contreras, & Sauma, 2013a; Pozo, Sauma, & Contreras, 2013b). However, the development of solution approaches for such models is beyond the scope of this paper and is left to future work.

<sup>3</sup> The loss of model fidelity in assuming that we can anticipate the full horizon of uncertain load and renewable generation supply will, in general, depend on the penetration level of time-dependent resources, the quality of forecasts actually available to operators, and the practical consequences of those forecasts in terms of, e.g., ex-post sub-optimal generator unit commitment decisions. Although we have not quantified the economic implications of this simplification, a relatively simple approximation to account for this error is to enforce a percentage of operating reserves as a function of renewables available in the system. This approximation is used in the ReEDS model by the National Renewable Laboratory (Short et al., 2011).

planning problem as the following stochastic mixed-integer linear program:

$$TC((\Omega, p)) = \min_x e^T x + f(x, (\Omega, p)) \quad (1)$$

$$\text{s.t. } Ax \leq b \quad (2)$$

$$x = (x_1, x_2), \quad x_1 \in \{0, 1\}, \quad x_2 \geq 0. \quad (3)$$

The function  $f(x, (\Omega, p))$  denotes the minimum expected operating costs for a given set of investments  $x$  and scenarios described by  $(\Omega, p)$ .<sup>4</sup> By adopting a discrete two-stage stochastic programming model, we explicitly assume the availability of a finite, fixed number of operations scenarios. This is not an unrealistic assumption in practice, particularly when operations scenarios are (as in our case) associated with historical observations. However, we consequently ignore issues relating to scenario sample selection and validation. The function  $TC(\Omega, p)$  denotes the minimum total system cost. The matrix  $A$  and vector  $b$  define investment constraints such as generation build limits, installed reserved margins per area, and limits on the maximum number of transmission circuits per corridor. The elements in the vector of investment variables  $x$  can be defined as discrete ( $x_1$ , counts of plants or transmission lines at a particular location) or continuous ( $x_2$ , generation capacity variables, in mega-watts). As in Binato et al. (2001), van der Weijde and Hobbs (2012), Munoz, Sauma, and Hobbs (2013), our application models transmission investments using binary variables and generation capacity using continuous variables. The formulation in (1)–(3) assumes a single investment planning stage, i.e., all investments are “here-and-now” decision variables, while all recourse variables are operations variables. More generally, the solution methods of this paper can be applied to multi-stage planning models, such those described in van der Weijde and Hobbs (2012), Munoz et al. (2013), and Munoz et al. (2014), which can be represented using minor variants of formulation (1)–(3).

### 2.3. Operations model

The objective of the operations problem is to minimize operating costs for a given discrete probability space, denoted  $(\Omega, p)$ . The problem is formulated as a linear program:

$$f(x, (\Omega, p)) = \min_{y(\omega)} \mathbb{E}_\omega[c^T y(\omega)] \quad (4)$$

$$\text{s.t. } Wy(\omega) \leq r(\omega) - T(\omega)x \quad \forall \omega \in \Omega \quad (5)$$

$$\mathbb{E}_\omega[Ky(\omega)] \leq d \quad (6)$$

$$y(\omega) \geq 0 \quad \forall \omega \in \Omega. \quad (7)$$

In our application, the per-scenario (e.g., hourly) constraints (5) consist of Kirchhoff's first and second law, maximum generation limits for both conventional and intermittent units, maximum power flow limits, flowgate limits, and ramping constraints. The expectation constraints (6) are used to enforce policy objectives, such as renewable targets or emission limits on a yearly basis (Munoz et al., 2014).

<sup>4</sup> The focus of the algorithms proposed in this paper is on models for economic planning, which aims at finding the most cost-effective combination of generation and transmission investments to meet demand and policy objectives assuming risk neutrality (minimization of expected system cost). Although reliability requirements are core drivers of infrastructure investments—particularly transmission—they need to be evaluated using non-economic tools that focus on the ability to serve load, which differs from the economic-focused tools we consider here (Munoz et al., 2014). Regarding economic risks, there is empirical evidence that suggests that both public and private investors are inherently risk averse. Some alternatives to account for such behavior in planning models include the use of utility functions, a constraint or weight on the conditional value at risk (CVaR), or the use of the Savage minimax regret criterion discussed in van der Weijde and Hobbs (2012). The linearity of the CVaR implementation of Rockafellar and Uryasev (2000) make this metric an attractive alternative to model risk aversion in generation and transmission planning models. A study of the economic implications and algorithmic challenges that result from using a CVaR is, however, beyond the scope of this paper and we leave it as a subject of future research.

### 3. Phase 1: the bounding algorithm

One alternative to solving large-scale stochastic programs (Birge & Louveaux, 1997) is to use computationally-tractable approximations that provide lower and upper bounds on the optimal objective function value. Well-known bounds for problems with stochastic right-hand sides include Jensen's inequality for lower bounds (Jensen, 1906) and the Edmunson–Madansky inequality for upper bounds (Madansky, 1960). These bounds can be progressively tightened by refining the partitioning of the space  $\Omega$ , until a certain optimality gap is achieved (Birge & Louveaux, 1997; Hobbs & Ji, 1999; Huang, Ziemba, & Ben-Tal, 1977). However, in our case expectation constraints prevent the direct application of Jensen's inequality, because it is only applicable to separable problems with per-scenario constraints. Computation of upper bounds still involve the solution of large optimization problems, which are sometimes facilitated by the application of decomposition algorithms (Hobbs & Ji, 1999). We now introduce an extension of the Jensen's-inequality-based lower bound to problems with both per-scenario and expectation constraints (Section 3.1), and describe a sub-sampling method that provides a statistical estimate of the upper bound problem, which we implement on a parallel computer system (Section 3.2).

#### 3.1. Lower bounds

Proposition 1 extends the Jensen's-inequality-based lower bound to stochastic linear programs with expected value constraints, as is the case in the operations model of Section 2.3. Detailed derivations and proofs are provided in the electronic supplementary material.

**Proposition 1.** *Given a discrete probability space  $\Omega$  with measure  $p$  and a partition  $S_1, \dots, S_m$  of  $\Omega$ , a sample space  $\Psi^m = \{\xi_1, \dots, \xi_m\}$  is defined with measure  $q^m$  such that the probability of each event  $\xi_i$  in  $\Psi^m$  equals the probability of each subset  $S_i$ , defined as  $q^m(\xi_i) = p(S_i)$ ,  $\forall i \in \{1, \dots, m\}$ . If the vector of right-hand-side parameters  $r(\cdot)$  and transition matrix  $T(\cdot)$  are computed using the expected value of these parameters over the partitions  $S_1, \dots, S_m$  such that  $r(\xi_i) = \mathbb{E}_\omega[r(\omega)|S_i]$  and  $T(\xi_i) = \mathbb{E}_\omega[T(\omega)|S_i]$ ,  $\forall \xi_i \in \Psi^m$ , then for any vector of investments  $x$ ,  $f(x, (\Psi^m, q^m)) \leq f(x, (\Omega, p))$ .*

We can interpret this result as follows. If the space  $\Omega$  is partitioned or clustered into subsets, and if the expected values of these parameters, conditioned on each subset/cluster, are used in the optimization problem and weighted in the objective function in proportion to the cluster sizes, then solving the operations problem  $f(x, (\Psi^m, q^m))$  provides a lower bound on the operations problem  $f(x, (\Omega, p))$ —which considers the full distribution of time-dependent parameters. If a hierarchical clustering algorithm is used then the bound can be guaranteed to be nondecreasing (i.e.,  $f(x, (\Psi^m, q^m)) \leq f(x, (\Psi^{m+1}, q^{m+1}))$ ),  $\forall x, \forall m \in \{1, \dots, |\Omega| - 1\}$  and convergent to  $f(x, (\Omega, p))$  (Birge & Louveaux, 1997). Next, Proposition 2 shows how this bound can be used to compute bounds on the optimal total system cost  $TC(\Omega, p)$  associated with the investment planning problem.

**Proposition 2.** *Given the conditions described in Proposition 1,  $TC((\Psi^m, q^m))$  is a lower bound on  $TC((\Omega, p))$ .*

Further, if a hierarchical clustering algorithm is used, such that  $f(x, (\Psi^m, q^m)) \leq f(x, (\Psi^{m+1}, q^{m+1}))$ ,  $\forall x, \forall m \in \{1, \dots, |\Omega| - 1\}$ , then  $TC(\Psi^m, q^m) \leq TC(\Psi^{m+1}, q^{m+1})$ ,  $\forall m \in \{1, \dots, |\Omega| - 1\}$ .

#### 3.2. Upper bounds

For any feasible investment plan  $x$ ,  $e^T x + f(x, (\Omega, p))$  clearly provides an upper bound on the optimal total system cost  $TC(\Omega, p)$ . However, computing  $f(x, (\Omega, p))$  could be prohibitive due to the presence of expectation constraints that link all scenarios within the operations problem and which consequently impede a direct parallel im-

plementation on a per-scenario basis. Relaxation of these constraints through Benders or Dantzig–Wolfe decomposition methods can address this difficulty (O'Brien, 2004), but this approach then requires implementation of nested decomposition algorithms within our proposed approach. Furthermore, some investment planning problems require consideration of multi-year time-series, sub-hourly resolution of intermittent data, and scenarios including component failures. These additional features would result in extremely large sample spaces  $(\Omega, p)$  and, therefore, large operations models that would be difficult to solve even in the absence of expectation constraints.

Other approaches that can reduce the computational complexity of large-scale stochastic optimization problems involve the use of samples of uncertain parameters, as opposed to their full distributions. Examples of this approach to compute upper bounds are described in Birge and Louveaux (1997) and Pierre-Louis, Bayraksan, and Morton (2011), and within Benders decomposition in Infanger (1992) and Hige and Sen (1991). The quality of these approximations is progressively improved by increasing the sample size between iterations (Birge & Louveaux, 1997; Pierre-Louis et al., 2011) or by combining information from multiple independent samples through Benders' cuts (Hige & Sen, 1991; Infanger, 1992). In particular, convergence of the Sample Average Approximation method (SAA), which relies on large-sample results, is guaranteed for stochastic linear programs with per-scenario and expectation constraints (Anitescu & Birge, 2008). However, a major drawback of these approximation methods is the poor quality of the estimation during initial iterations as a consequence of their asymptotic convergence properties. Recent results from Birge (2011) show that in some cases a combination of sub-sample estimates, as in the batch-means method (Law & Carson, 1979; Schmeiser, 1982), can achieve faster convergence and more robust results than when considering a single large sample of equivalent size. In the spirit of Birge (2011), our method to compute estimates of  $f(x, (\Omega, p))$  utilizes a sub-sampling approach that is enhanced through stratified sampling, which reduces the estimate variance.

##### 3.2.1. Sub-sample estimation

Our estimation method relies on using the means of  $N$  independent groups, or batches, of  $M$  observations each, instead of using a single large sample of equivalent size  $N \times M$ . We denote a random sample of  $M$  observations from the space  $\Omega$  as  $\Omega^M$ , and define a new probability measure  $p^M(\cdot)$  such that all observations have the same probability of occurrence  $p^M(\omega_i) = 1/M$ ,  $\forall i \in \{1, \dots, M\}$ . To approximate  $f(x, (\Omega, p))$ , we draw  $N$  independent samples of  $M$  observations each, denoted  $\{\Omega_1^M, \dots, \Omega_N^M\}$ , and solve  $N$  independent operations problems, denoted  $f(x, (\Omega_1^M, p_1^M)), \dots, f(x, (\Omega_N^M, p_N^M))$ . An estimate of  $f(x, (\Omega, p))$  is then calculated as:

$$f(x, (\Omega, p)) \cong \frac{1}{N} \sum_{j=1}^N f(x, (\Omega_j^M, p_j^M)). \tag{8}$$

Using our sub-sample method to approximate  $f(x, (\Omega, p))$ , the term  $e^T x + \frac{1}{N} \sum_{j=1}^N f(x, (\Omega_j^M, p_j^M))$  provides an upper bound on  $TC(\Omega, p)$  for large values of  $M$  and  $N$ . Convergence of this method as  $N$  is increased is assured for stochastic linear programs (Birge, 2011), as it is in our case. However, small systematic biases might arise depending on the structure and stringency of the constraints in the operations problem. Previous research on the traditional batch-means estimator shows that small batch sizes (i.e., small  $M$ ) can be a source of bias for the sample-mean estimator (i.e., sample size), although an asymptotically normal distribution of errors is guaranteed for large samples and batch counts (Schmeiser, 1982; Steiger & Wilson, 2001). An advantage of this method over other approaches that rely on unique, large-sample results is that problems  $f(x, (\Omega_1^M, p_1^M)), \dots, f(x, (\Omega_N^M, p_N^M))$  can be solved in parallel. Therefore, the extra computational load that results from increasing the sample size  $M$  or batch count  $N$  to reduce bias, and to

ensure tight confidence intervals on the sample mean, can be efficiently distributed among multiple independent processors, instead of being given to a single optimization problem of comparable size (e.g.,  $f(x, (\Omega^{N \times M}, p^{N \times M}))$ ).

### 3.2.2. Reducing the variance through stratified sampling

To further reduce the computational cost of approximating  $f(x, (\Omega, p))$  through sub-sample estimations, we propose the utilization of a stratification technique to select samples that would more accurately match the characteristics of  $\Omega$  and therefore reduce the variance of the sub-samples  $f(x, (\Omega_1^M, p_1^M)), \dots, f(x, (\Omega_N^M, p_N^M))$ . Stratified sampling has been previously used for electricity production cost modeling (Marnay & Strauss, 1991), but has yet to be utilized in the context of the sub-sampling method proposed by Birge (2011). Our stratified sampling algorithm proceeds as follows. For a given predetermined sample size  $M$ , the space  $\Omega$  is partitioned into disjoint subsets  $S_1, \dots, S_M$ , with the objective of grouping the events into clusters with similar characteristics (e.g., observations are grouped based on load, wind, solar, and hydro levels included in  $r(\omega)$  and  $T(\omega)$ ). Since in our application each observation is defined by a spatially disaggregated vector of load, wind, solar, and hydro levels for one representative hour, the clustering algorithm preserves the correlations among time-dependent parameters. For instance, high-load and low-wind hours are grouped separately from low-load and high-wind scenarios. The stratified sampling method selects one representative hour from each of these clusters (e.g., one representative observation from a cluster with high-load and low-wind scenarios and one from the low-load and high-wind group). Therefore, a stratified sample of  $M$  observations  $\{\omega_1, \dots, \omega_M\} \subset \Omega$  is such that  $\omega_i \in S_i, \forall i \in \{1, \dots, M\}$ . We weight each observation  $\omega_i$  in the operations problem with probability  $p^M(\omega_i) = p(\omega_i)/p(\omega_i|S_i) \forall i \in \{1, \dots, M\}$ .

### 3.3. Updating the upper and lower bounds

Bounding methods that rely on clustering algorithms, or sampling, require either progressive refinement of the sample space partition or increasing sample sizes to decrease the optimality gap and improve the accuracy of the upper bound estimate (Birge & Louveaux, 1997; Hobbs & Ji, 1999; Pierre-Louis et al., 2011). To avoid poor initial estimates of the upper bound, we propose selecting and fixing both the sample size  $M$  and batch count  $N$  prior to initialization of the bounding or decomposition phases of our algorithm. An experimental analysis of the effects of sample sizes and sampling methodologies is given in the electronic supplementary material. Throughout the rest of this paper, we will assume that a computationally efficient method to approximate  $f(x, (\Omega, p))$  for any candidate investment plan  $x$  is available.

The bounding algorithm, or Phase 1 of our methodology, proceeds as follows. For a given sample size  $M$  and batch count  $N$ , we initialize the iteration counter by setting  $k = 0$ , and the lower and upper bounds as  $LB_0 = -\infty$  and  $UB_0 = +\infty$ , respectively. The incumbent solution is denoted  $x^*$ .

1. Set  $k = k + 1$ , solve the lower-bound planning problem using the partitioned space  $(\Psi^k, q^k)$  in (1), instead of the full space  $(\Omega, p)$ . Find a trial investment plan  $x_k^*$ , and a lower bound on the optimal total system costs  $TC(\Psi^k, q^k)$ . If  $TC(\Psi^k, q^k) > LB_{k-1}$ , update the lower bound to  $LB_k = TC(\Psi^k, q^k)$ ; otherwise,  $LB_k = LB_{k-1}$ .
2. Compute the operating costs  $f(x_k^*, (\Omega, p))$ . If  $e^T x_k^* + f(x_k^*, (\Omega, p)) < UB_{k-1}$ , update the upper bound  $UB_k = e^T x_k^* + f(x_k^*, (\Omega, p))$  and the incumbent solution  $x^* = x_k^*$ ; otherwise,  $UB_k = UB_{k-1}$ .
3. Compute the optimality gap, defined as  $GAP_k = 100 \text{ percent} \times (UB_k - LB_k)/UB_k$ . If  $GAP_k$  is less than or equal to a pre-determined optimality gap, stop and use  $x^*$  as the proposed investment plan. Otherwise, go to step 1.

For a sequence of partitions defined such that its limit is the full discrete probability space  $(\Omega, p)$ , and a large sample size  $M$  and batch count  $N$ , the lower ( $LB_k$ ) and upper bounds ( $UB_k$ ) are convergent to  $TC((\Omega, p))$  as  $k \rightarrow |\Omega|$ . A formal proof of convergence is given in the electronic supplementary material.

Empirically, clustering algorithms often exhibit the following behavior: only a few partitions are required to explain a large fraction of the variance observed in the full dataset (the “elbow” phenomenon), but the remaining fraction of variance explained converges asymptotically to 1 as the partitions are refined (Tibshirani, Walther, & Hastie, 2001). A potential implication of this observation for the bounding phase of our algorithm is that moderate optimality gaps might be achieved using only small numbers of representative hours from the sample space (clustered load, wind, solar, and hydro levels), but that tight optimality gaps might be only attained for large values of  $k$  (Hobbs & Ji, 1999). This is particularly challenging for planning problems that consider binary decision variables (e.g., transmission investments), because solving the lower-bound planning problems  $TC(\Psi^k, q^k)$  using a branch-and-bound type of algorithm for large values of  $k$  become increasingly difficult. In the following section, we describe the use of Benders decomposition (Phase 2 of our methodology) to close the residual optimality gap from the bounding phase through the addition of cuts to the lower-bound planning problem.

## 4. Phase 2: enhanced Benders decomposition

An alternative to the bounding approach described in the previous section is to take advantage of the decomposable structure of the planning problem and solve it iteratively using Benders decomposition (Bloom, 1983). However, the main drawback of this method is its slow convergence speed and the increasing size of the master problem as the algorithm iterates due to the accumulation of cuts. Multiple techniques have been proposed to accelerate the convergence of Benders decomposition when applied to mixed-integer linear optimization problems. For example, speed improvements can be achieved from tighter mixed-integer formulations and the selection of Pareto optimal cuts for subproblems with degenerate solutions (Magnanti & Wong, 1981; Sahinidis & Grossmann, 1991). Other techniques address the computational challenge of solving multiple mixed-integer linear master problems by initially computing cuts from linear (McDaniel & Devine, 1977) and Lagrangian relaxations (Aardal & Larsson, 1990; Cote & Laughton, 1984; Hoang Hai, 1980; van Roy, 1983), as well as from feasible sub-optimal solutions found by prematurely stopping branch-and-bound type algorithms (Geoffrion & Graves, 1980). Trust regions combined with high-quality initial solutions have been proposed to reduce the magnitude of changes in the master problem solution from iteration to iteration, resulting in quicker convergence (Sherali & Staschus, 1990; Sherali et al., 1987). However, approximate solution methods may prevent achievement of the conditions required to guarantee convergence of Benders decomposition to an optimal solution, resulting in either premature convergence to a suboptimal solution or failure to converge (Holmberg, 1994).

Phase 2 of our algorithm uses a modification of Benders decomposition master problem, which we augment with a polyhedral lower bound on the optimal operating costs  $f(x, (\Omega, p))$ , based on the results we introduced in Section 3.1. This approach can be interpreted as a generalization of the stabilization scheme for the stochastic decomposition algorithm (Higle & Sen, 1991) currently implemented in the NEOS Solver (Sen, 2013), which utilizes a partition with a single subset (i.e., the expected-value of all time-dependent parameters) to construct an auxiliary lower bound in the master problem.

Our modified Benders decomposition master problem is:

$$\min_{x, \theta} e^T x + \theta \quad (9)$$

$$\text{s.t. } Ax \leq b \tag{10}$$

$$\theta \geq f(x^*, (\Omega, p)) + \pi^T(x^*)(x - x^*) \tag{11}$$

$$\theta \geq f(x, (\Psi^m, q^m)) \tag{12}$$

$$x = (x_1, x_2), x_1 \in \{0, 1\}, x_2, \theta \geq 0. \tag{13}$$

Constraint (11) represents the Benders' cuts, which are computed using the sub-sampling method described in Section 3.2.1. The Lagrange multipliers  $\pi(x^*)$  in (11) result from imposing  $x = x^*$  in the calculation of  $f(x^*, (\Omega, p))$ . Constraint (12) captures the operations problem defined by Eqs. (4)–(7) for the sample space  $(\Psi^m, q^m)$ . The Benders master problem in electricity capacity expansion typically only considers investment variables (Bloom, 1983). In contrast, our master problem corresponds to a planning problem with an embedded low-resolution operations problem. The fidelity of the operations problem can be improved by increasing the number of clusters  $k$  used to approximate the sample space  $\Omega$ . For  $k = 0$ , no constraints on the value of  $\theta$  are imposed through constraint (12), so that the problem defined by Eqs. (9)–(11) and (13) corresponds to the standard master problem. For  $k = |\Omega|$ , all observations are considered and the master problem is equivalent to the original planning problem, which converges in a single iteration.

A second improvement on the auxiliary lower bound is the utilization of the objective function value of a relaxed linear programming version of the problem (1)–(3), denoted  $TC_{LP}(\Psi^k, q^k)$ . For a suitably large value of  $k$ ,  $TC_{LP}(\Psi^k, q^k)$  might provide a tight initial lower bound ( $LB_0$ ) for the Benders' iterations. This auxiliary bound is introduced to address a potential limitation of bounding algorithms that might find high-quality solutions during initial iterations, but optimality cannot be proven until the difference between the upper and lower bounds is below a certain tolerance.<sup>5</sup>

Note that unlike the methods proposed by Infanger (1992) and Higle and Sen (1991), we utilize a unique large sample across all iterations. This means that the quality of the approximation of the recourse function  $f(x^*, (\Omega, p))$  is predetermined and it does not depend on the number of iterations of the Benders decomposition algorithm. The main advantages of this approach are its simplicity, since it does not require resampling between iterations, and the possibility of obtaining high-quality approximations of the recourse function during early iterations. However, the disadvantage is that we cannot prove that the batch size and count we select are large enough to guarantee convergence of the algorithm to  $TC((\Omega, p))$ . Much larger batch sizes and batch counts could potentially improve the quality of the solution, and re-sampling between iterations could be implemented to build on the results of Infanger (1992) and Higle and Sen (1991). However, implementation of those variants is beyond the scope of this paper.

Finally, as in the L-shaped method for stochastic programs (Birge & Louveaux, 1997), we compute a single cut at each Benders iteration using the expected value of the dual multipliers  $\pi_1(x^*), \dots, \pi_N(x^*)$  and the operating costs  $f(x^*, (\Omega_1^M, p_1^M)), \dots, f(x^*, (\Omega_N^M, p_N^M))$  of the  $N$  sub-samples. An extension of this method, known as the multi-cut L-shaped algorithm, requires the addition of one cut per scenario (instead of a single "expected" cut) at each Benders iteration (Birge & Louveaux, 1988). However, improved convergence of Benders decomposition due to the multi-cut method is at least partially offset by the increased growth in the size of the master problem. Thus, we leave implementation of this extension to future research.

<sup>5</sup> This is often observed in branch-and-bound type of algorithms, where even if an optimal solution is found within the first iterations, optimality cannot be guaranteed until the algorithm has completed all the nodes.

## 5. Numerical example

In this section, we describe an application of our two-phase bounding and decomposition algorithms to a large-scale generation and transmission planning problem using a 240-bus representation of the Western Electricity Coordinating Council (WECC) in the U.S. (Munoz et al., 2014). This model consists of 240 existing buses, 448 transmission elements, and 157 aggregated generators. We model intermittent resources using 151 historical profiles of hourly demand, wind, solar, and hydro levels across multiple regions representing operating conditions for a typical year. More details of this model are given in the electronic supplementary material. The planning problem is formulated using 8736<sup>6</sup> observations of time-dependent data results in a mixed-integer linear program with 56 million constraints and 31 million variables; 1020 of these variables represent discrete (integer) transmission investment options.<sup>7</sup> We consider both the mixed-integer linear formulation, suited for real-world planning studies, and its linear relaxation, which resembles high-level models used for policy analysis. In order to obtain feasible operations problems for any candidate investment plan, we allow for load curtailment at a cost of \$1000 per megawatt hour (the price ceiling used in most electricity markets in the U.S.). Noncompliance with annual renewable energy targets is penalized at a rate of \$500 per megawatt hour.

The clustering, bounding, and Benders decomposition algorithms are all implemented using the Pyomo algebraic modeling package (Hart, Laird, Watson, & Woodruff, 2012). All optimization problems are solved with the CPLEX 12.4 solver and parallelized through the Message Passage Interface (MPI) on a 32-core computer system with 2 AMD Opteron processors of 2.2 gigahertz and 112 gigabyte of RAM.

### 5.1. Clustering algorithm

We partition the space of time-dependent parameters using the k-means algorithm (MacQueen, 1967), although other partitioning schemes have been used in similar applications (Hobbs & Ji, 1999). The k-means algorithm has been used to model load levels in production cost models (Wogrin, Duenas, Delgado, & Reneses, 2014) and for wind generation investment (Baringo & Conejo, 2013), but not to compute lower bounds upon total system cost as we do in this article. K-means is a nonhierarchical clustering algorithm, implying that the total cost associated with the planning problem might decrease rather than increase as the partitions are refined (Birge & Louveaux, 1997). Despite this property, Hobbs and Ji (1999) report that k-means yielded the best clustering efficiency in their application (probabilistic production cost modeling), measured as the fraction of variance captured from the full data set, compared to several other partitioning methods—including hierarchical methods. Our implementation considers up to 500 clusters, which capture 71.1 percent of the variance of the normalized data set of 8736 observations of load, wind, solar, and hydro (see Fig. 1). The point of diminishing returns (i.e., the elbow) is reached at approximately 50 clusters, capturing 46.4 percent of variance. After this point, capturing an extra 10 percent of variance requires partitioning the space of load, wind, solar, and hydro levels using 100 additional clusters.

<sup>6</sup> The sample is weighted by 8760/8736 in the objective function and expectation constraints of the operations problem. The sample size of 8736 hours results from considering 52 weeks of hourly solar data for a typical year.

<sup>7</sup> We model transmission investment alternatives as binary decision variables because network infrastructure is "lumpy", e.g., a corridor is populated or not (Joskow & Tirole, 2005). Modeling generation investments as continuous variables is obviously an approximation. However, there exist relatively small scale economies for new renewable and conventional generation technologies, such that the minimum economic sizes for those new resources is well below the capacity of a single extra high voltage line. This issue is discussed in Munoz et al. (2013).

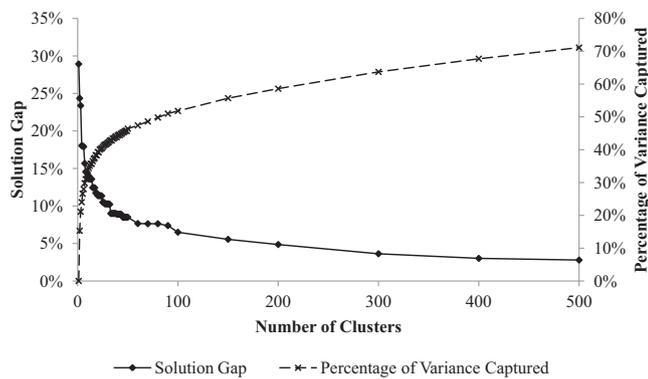


Fig. 1. Optimality gap for the linear relaxation and percentage of variance captured as a function of the number of clusters.

## 5.2. Phase 1: performance of the bounding algorithm

### 5.2.1. Linear problem (LP)

First, we consider a linear relaxation of the original mixed-integer linear problem. This allows us to study the efficiency of our algorithm when applied to linear electricity investment planning and market simulation models such as the ones described in ICF (2013), Gabriel et al. (2001), Short et al. (2011), Paul and Burtraw (2002), and EIA (2013). Fig. 1 shows the optimality gap as a function of the number of clusters, calculated as  $100\% \times \frac{UB-LB}{UB}$  for the linear relaxation. We compute the lower bound (LB) by solving the planning problem using one representative hour from cluster (i.e., average of all observations of load, wind, solar, and hydro parameters within each cluster). The upper bound (UB) equals the sum of the investment costs associated with the lower-bound problem, plus a statistical estimate of the 8736 hour operations problem  $f(x, (\Omega, p))$  computed using the sample mean of  $N = 20$  sub-samples of  $M = 200$  stratified samples of time-dependent parameters. For this test-case, only 33 representative hours are needed to obtain a solution within 10 percent of the global optimum (Fig. 1), but more than 200 clusters are required to reduce the gap further to 5 percent. The “knee” in the optimality gap mirrors the elbow associated with the percentage of variance captured from the 8736 hour dataset, which decreases at a much slower rate after the first 39 clusters. The optimality gap is reduced from 28.9 percent to 8.9 percent using 39 clusters. Increasing the number of partitions to 500 only reduces the gap to 2.8 percent, which is acceptable for high-level planning models associated with policy analysis. The optimality gap for a single cluster, also known as the expected-value problem, provides an upper bound on the Value of the Stochastic Solution (VSS) as defined in Birge and Louveaux (1997). This is a measure of the potential cost savings that could be achieved by considering the full distribution of hourly load levels and capacity factors of wind, solar, and hydro resources across all regions, instead of planning a system using the expected value of these parameters.

### 5.2.2. Mixed-integer linear problem (MILP)

For the mixed-integer linear case we estimate the optimality gap as  $100\% \times \frac{UB-(1-\epsilon)LB}{UB}$ , where  $\epsilon$  corresponds to the MILP optimality gap for the lower bound model. Note that  $(1-\epsilon)LB$  provides a lower bound on the optimal system costs for a zero MILP gap ( $\epsilon = 0$ ). Therefore,  $100\% \times \frac{UB-(1-\epsilon)LB}{UB}$  is an upper bound on the optimality gap that could be achieved if  $\epsilon = 0$ . As the MILP gap is increased, the optimality gap increases as a consequence of both the deterioration of the lower bound (the  $(1-\epsilon)$  factor) and the suboptimality of the investment decisions, reflected as higher operating costs in the upper bound. We observe that 10 clusters are sufficient to achieve optimality gaps of 12.1 percent for a 1 percent MILP gap and that 90 additional clusters only reduce the optimality gap to 7.4 percent. Optimality gaps and bounds as a function of the number of clusters are

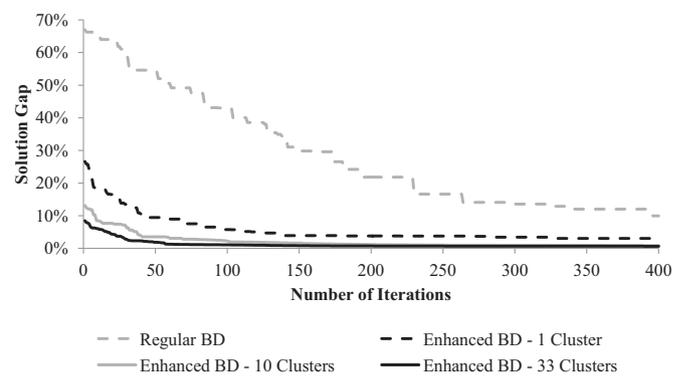


Fig. 2. Optimality gap versus the number of iterations for different Benders decomposition schemes of the linear relaxation. Upper and lower bounds are in Fig. C.6 in the electronic supplementary material.

shown in Figs. B.4 and B.5, respectively, of the electronic supplementary material. For practical implementations, the linear relaxation of the mixed-integer planning problem could be used as a screening tool to assess the minimum number of clusters needed to approximate the operations problem with a pre-specified optimality gap.

Solution times for the mixed-integer linear formulations are sensitive to the choice of the MILP gap and are orders of magnitude larger than solution times for the linear relaxation. While the linear relaxation of the 100-cluster problem requires only 5 minutes to solve, the mixed-integer formulations requires approximately 1.3, 6.6, and 6.8 hours to achieve 5 percent, 3 percent, and 1 percent MILP gaps, respectively. The reduction of 5.5 hours in solution time achieved from loosening the MILP gap from 1 percent to 5 percent is, however, contrasted with an increase in the total optimality gap from 7.4 percent to 11.1 percent. Further attempts to solve mixed-integer linear problems with 200 or more clusters and a MILP gap of 1 percent did not yield a solution within the optimality gap after more than 30 hours of computation time, and execution was stopped.

## 5.3. Phase 2: enhanced Benders decomposition

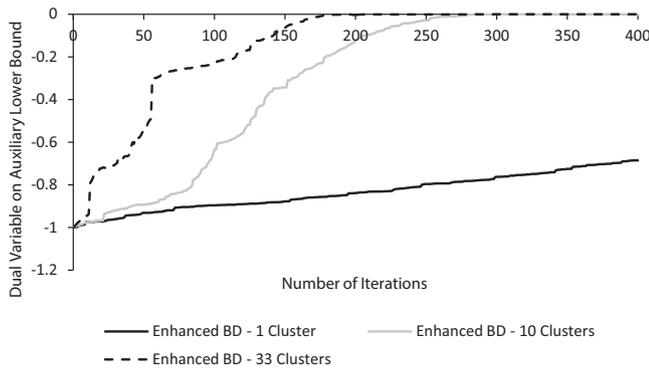
Our bounding algorithm can efficiently find investment solutions whose total system costs have a 2.8 percent optimality gap for the linear relaxation and a 7.4 percent gap for the mixed-integer linear case. Achieving tighter optimality gaps, however, requires significant refinements of the clustering scheme due to the slow convergence of the bounding method (Hobbs & Ji, 1999). To overcome this limitation, we use Benders' cuts to further reduce the optimality gap by iterating successively between the lower-bound investment planning problem (i.e., master problem) and the operations problems (i.e., sub-problems).

### 5.3.1. Linear problem (LP)

Fig. 2 shows the optimality gap for the linear relaxation of Benders decomposition using different number of clusters as auxiliary lower bounds in the master problem. The figure also illustrates the convergence behavior of the traditional Benders decomposition algorithm without auxiliary bounds on the operations costs. The number of clusters for the experiments was selected based on changes in the convergence rate of the optimality gap of the linear relaxation (see Fig. 1). No experiments are considered beyond 33 clusters due to the significant decrease in the convergence rate of the bounding algorithm after that point. Although more clusters could further reduce the number of Benders' iterations required to achieve tight optimality gaps, their use causes significant increases in the time required to solve the model in each iteration, particularly in the mixed-integer linear case. We leave those experiments as a subject of future research. We use the objective function value

**Table 1**  
Number of cuts and total computation time for different optimality gaps and auxiliary lower bounds of the linear relaxation.

Solution gap LP relaxation (percent)	Number of clusters used in the auxiliary lower bound					
	1 Cluster		10 Clusters		33 Clusters	
	Number of cuts	Total time (s)	Number of cuts	Total time (s)	Number of cuts	Total time (s)
5.0	117	31,140	38	11,307	17	6026
3.0	329	95,053	69	21,018	29	10,359
1.0	>400	>117,332	246	78,917	116	44,001



**Fig. 3.** Dual multiplier on auxiliary lower bound versus number of iterations of Benders decomposition for the linear relaxation.

$TC_{LP}(\Psi^{150}, q^{150})$  to initialize the lower bound ( $LB_0$ ), similar to the method proposed by van Roy (1983) to compute initial bounds for Benders decomposition using Lagrangian relaxation.

In Fig. 2 we observe that including the expected-value problem (i.e., single partition) in the master problem, as done in Sen (2013), results in shrinkage of the optimality gap from 28.9 percent to 3 percent in 329 iterations after 26.4 hours of computing time (Table 1). In contrast, the regular Benders decomposition implementation in which no such auxiliary constraint is included only yields a 10.0 percent optimality gap after 400 iterations and 30 hours of computing time. Including more clusters results in further reductions in the number of iterations and solution time to achieve tight optimality gaps (Table 1). In this case, using 33 clusters reduces the time to achieve a solution with a 1 percent optimality gap by more than half (to 12.2 hours), compared to including the expected-value solution (1 cluster) in the master problem (more than 32.6 hours).

Larger speedups can be achieved for looser optimality gaps by including more clusters in the auxiliary lower bound of Benders decomposition, but the resulting larger master problem eventually caused solution times to become higher than those achieved by increasing the number of clusters using the bounding algorithm in the linear case. Attaining a 3 percent optimality gap for the LP relaxation requires 400 clusters and 30 minutes of computation through the bounding method, but 29 iterations and 2.8 hours using our enhanced Benders' approach with 33 clusters. Hence, the bounding method may be more efficient and practical than the modified Benders' algorithm if loose optimality gaps for the linear relaxation are acceptable for planning purposes.

*Contribution of the auxiliary lower bound toward convergence of the UB and LB:*

Fig. 3 depicts the value of the dual variable on constraint (12) of the modified master problem. The dual provides a measure of the contribution of the auxiliary lower bound to the objective function value of the lower-bound planning problem ( $f(x, (\Psi^k, q^k))$ ). No cuts are available the first time the master problem is solved in the modi-

fied Benders decomposition and the only lower bound upon the operating costs is  $f(x, (\Psi^k, q^k))$ . The dual variable is in that case equal to  $-1$ . As Benders' cuts are incorporated into the master problem, the contribution of the auxiliary lower bound  $f(x, (\Psi^k, q^k))$  to the total system cost is progressively reduced, reflected in the smaller magnitude of its dual. When Benders' cuts alone provide a tighter (i.e., higher) lower bound on the operating costs than  $f(x, (\Psi^k, q^k))$ , the value of the dual variable for the auxiliary constraint (12) is equal to zero, as it is the case for the 10- and 33-cluster experiments after 300 and 279 iterations, respectively. Note that the contribution of the lower bound is far beyond what a good initial solution or warm start of the master problem can provide. This behavior can be observed in the 33-cluster experiment, for instance, where constraint (12) is still binding after 278 iterations, at which point the enhanced Benders' algorithm has attained an optimality gap of 0.67 percent.

**5.3.2. Mixed-integer linear problem (MILP)**

For the mixed-integer linear case, we follow the approach proposed by McDaniel and Devine (1977) and use cuts computed from the linear relaxation of the mixed-integer master problem, in conjunction with the auxiliary lower bound (12), to warm start the algorithm.

For illustration purposes, we only consider further iterations of the enhanced Benders decomposition using the expected-value problem (1 cluster), a 0.5 percent MILP gap, and 400 pre-computed cuts from the linear relaxation. Computing the 400 cuts required 32.5 hours, but results in significant improvements in the optimality gap in the mixed-integer linear problem. Only 11 iterations are needed to attain a 5 percent optimality gap (after 1.9 hours), and letting the algorithm run for 166 more iterations and 39.4 hours only reduces the optimality gap to 4.1 percent. In contrast, the best investment plan found without pre-computed cuts results in a 7 percent optimality gap after 200 iterations and requires 43 hours of computation time.

As in the previous experiments with the linear relaxation, we utilize the objective function value  $TC_{LP}(\Psi^{150}, q^{150}) = \$624.3B$  to initialize the lower bound ( $LB_0 = \$624.3B$ ). However, a tighter lower bound can be obtained from solving the linear relaxation close to optimality using the enhanced Benders decomposition. In our experiments, the 33-cluster implementation results in the tightest optimality gap (0.64 percent) and in the highest lower bound ( $\$635.5B$ ) after 400 iterations and 44 hours of computation time (see Fig. C.6 in the electronic supplementary material). If this value is used as an initial lower bound ( $LB_0 = \$635.5B$ ), the optimality gap after 200 iterations is reduced from 4.1 percent to 2.4 percent, which is arguably sufficient for long-term planning studies.<sup>8</sup> Attaining even tighter optimality gaps requires more iterations of the algorithm, but these come at the expense of longer solution times. Consequently, the auxiliary lower bound, pre-computed cuts, and a tight lower bound can be used to attain tight optimality gaps for large-scale planning problems with integer variables.

**5.4. Can the lower-bound problem be used for planning?**

The lower-bound investment planning problem can provide useful information about total and marginal system costs required to meet forecasted demand and environmental goals, but these results may only be meaningful for large cluster counts. Strictly using the

<sup>8</sup> We have necessarily ignored a number of uncertainties in this model, choosing to focus on specific issues relating to renewables integration. Examples include fuel price uncertainty, political uncertainties (which could, e.g., impact the feasibility of certain transmission corridors), and future demand. Consequently, an overly precise computation given a limited fidelity model is unnecessary in practice. In our experience with real-world planning problems in a variety of domains, decision makers are comfortable with such an approach. Ultimately, outputs of such a model would be just one of many inputs to decision making, meaning that a highly precise solution will not be directly realized in practice.

objective function value of the lower-bound planning problem (i.e., total cost) as an indicator of convergence, e.g., as in Heejung and Baldick (2013), could result in premature termination of the algorithm. Changes in the objective function value of the lower-bound planning problem only reflect improvements of the fidelity of the embedded operations problem that utilizes clustered data, but do not necessarily guarantee improvements in the quality of the investment plan when tested against the high-resolution operations problem, which is used to compute upper bounds.<sup>9</sup>

We also observe that the lower-bound problem builds too little transmission capacity relative to near-optimal solutions for small cluster counts, which causes an increase in penalties due to curtailed load and noncompliance fines in the upper bound. This occurs because in the case of small cluster counts, high-impact and low-probability scenarios (i.e., peak load) are not accurately represented by the cluster centroids (i.e., they are grouped with and therefore averaged together with off-peak hours). This can cause underinvestment in transmission or generation infrastructure, of which we observe the former. Modifications to the clustering algorithm could potentially address this issue by weighting (Tseng, 2007) or constraining (Wagstaff, Cardie, Rogers, & Schr, 2001) the clustering algorithm to include peak-load hours, for instance, as individual clusters. Alternatively, the identification of hours that drive transmission and generation investments could also be automated by first solving the planning problem for each hour independently. The resulting vector of total costs could be used to bias an hour-selection algorithm, as done in importance sampling (Infanger, 1992; Papavasiliou & Oren, 2013). However, these are all subjects of future research.

The amount of generation investments by technology, on the other hand, remain roughly constant as partitions are refined and more representative hours are introduced into the lower-bound problem.<sup>10</sup> However, increasing the resolution of the time-dependent parameters shifts the optimal geographical distribution of investments for certain technologies. For instance, the expected-value problem, which considers a single representative hour, underestimates wind capacity by 100 percent and 42 percent in the states of Utah (UT) and Arizona (AZ), respectively, and overestimates investments in Wyoming (WY) by 60 percent with respect to the optimal levels of the 500-cluster problem (see Fig. 4). Although the aggregate capacity of wind investments in these three states becomes stable after only 50 clusters, there is a striking shift of wind investments from Colorado (CO) to Washington (WA) as the number of clusters is increased to 500. This result highlights the importance of using fine-grained representations of variability within investment planning models to capture the true economic value of intermittent electricity generating technologies (Joskow, 2011).

In summary, unless large cluster counts are considered, we do not recommend using the lower-bound planning problem to find investment plans without assessing their quality against the full resolution operations problem (upper bound). In the linear case, more than 200 clusters were needed to attain a optimality gap below 5 percent, which is nearly four times the number of clusters needed to achieve the elbow on the fraction of variance explained from the full dataset of load, wind, solar, and hydro levels (see Fig. 1). The elbow criterion, often used to determine the number of clusters in a dataset

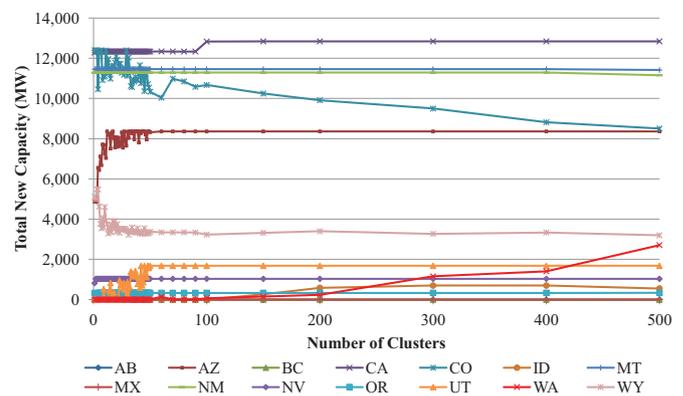


Fig. 4. Aggregate investments in wind capacity per state as a function of the number of clusters for the linear problem.

(Tibshirani et al., 2001), can be used in our case to identify the point when the clustering algorithm becomes inefficient. In this situation, it may be advantageous to switch to Phase 2 (Benders decomposition) of our proposed two-phase approach if tighter optimality gaps are required. However, the elbow itself provides no information regarding the potential quality (i.e., optimality gap) of the investment plan that results from solving the lower-bound planning problem.

## 6. Conclusions

We propose a two-phase algorithm to find investment plans with bounds upon the optimal system costs for large-scale planning models with numerous operating subproblems, with a focus on transmission and generation planning. The bounding phase is an extension of the approaches proposed by Huang et al. (1977) and Hobbs and Ji (1999) for stochastic problems with environmental restrictions that we model with expectation constraints. The decomposition phase is a modification of Benders decomposition that includes a low-resolution operations problem in the master problem as an auxiliary lower bound upon the operating costs. We compute upper bounds for both algorithms using a sub-sample estimation of the true operating costs for a given investment plan implemented in a parallel computer system. From our numerical experiments, we find that the bounding phase can be more efficient than the traditional Benders decomposition to find investment plans within moderate optimality gaps (i.e., 3–6 percent approximately) for both linear and mixed-integer cases. For implementation purposes, the bounding phase is far more practical than Benders decomposition since improving the quality of the investments only requires refining the clustering of the time-dependent data. However, for applications where the bounding method fails to converge sufficiently rapidly, a combination of the bounding algorithm with Benders decomposition, as demonstrated in Phase 2, can be used to attain tight optimality gaps, and is more efficient than using either of these two algorithms separately.

Our enhancement of the Benders algorithm is based on a lower bound that can be progressively improved by refining the partitioning of the space of load, wind, solar, and hydro levels, but that requires the planning problem to have all stochasticity limited to the right-hand-side of the constraints so that we can apply Jensen's inequality. An interesting direction for future research would be to explore the effect of including other valid lower bounds in the master problem of Benders decomposition. This could be done, for example, by relaxing constraints that complicate the solution of the planning model and iterating between loosely constrained (lower bound) and highly constrained (upper bound) problems.

Another potential extension of our algorithm is the inclusion of unit commitment variables and constraints in long-term planning models, an area of growing attention among operation researchers

<sup>9</sup> Upper and lower bounds for the linear relaxation are in Fig. B.3 in the electronic supplementary material. The rate of improvement of the lower bound deteriorates rapidly after the first 20 clusters, which could meet the convergence criterion described in Heejung and Baldick (2013), even though the optimality gap is still above 10 percent (Fig. 1).

<sup>10</sup> Investment costs for transmission and generation represent 3.6 percent and 48.2 percent, respectively, of the total system cost for the 500-cluster experiment (\$642.2B) – which yields a 2.8 percent optimality gap. Aggregate generation cost remains relatively constant with respect to the number of clusters. However, transmission costs increase from \$16.3B (one cluster, expected-value problem) to \$23.4B (500-cluster problem).

(Nweke, Leanez, Drayton, & Kolhe, 2012; Palmintier & Webster, 2011; Shortt et al., 2013). This would require inclusion of binary variables in the operations problems, which would as a result become nonconvex. Our bounding algorithm would still be applicable by relaxing all binary variables to compute lower bounds; however, we would not be able to guarantee convergence of the bounds to the true optimal system costs. Baringo and Conejo (2012) and Kazempour and Conejo (2012) have recently implemented and shown convergence of Benders decomposition including integer variables in the subproblems. Their results are based on Bertsekas and Sandell (1982), who proved that for a certain class of stochastic mixed-integer optimization problems, the duality gap converges to zero as the number of scenarios and integer variables is increased to infinity. A future step in our research is to study the implications of this result for a planning problem with unit commitment variables and to verify convergence of the Benders' algorithm with nonconvex subproblems.

Finally, as we state in Section 5.3, the optimality gaps we obtain are arguably sufficient to answer practical questions. A combined transmission-generation optimization model with a multidecadal time horizon can be useful for both long-range planning and policy analysis. For planning, the method identifies transmission developments that potentially have high economic value as a result of increasing power trade within and between regions. For instance, new lines might link areas whose load and renewable outputs are imperfectly correlated so that when region is short of power it might take more advantage of surpluses elsewhere. If additions between two regions are made for most or all solutions within a gap, this gives confidence that such links are worth investigating by planners with more detailed production costing models. Meanwhile, the model can also be used to investigate policy issues in more detail and with greater accuracy than is now possible using models such as IPM (ICF, 2013) that aggregate generation into large zones while modeling transmission using simplified transportation models that disregard Kirchhoff's laws. Many environmental and energy policies differ in stringency or type among subregions. Consequently, the geographical detail that our model allows for more realistic representation of local effects, and interactions among the regions. As an example, elsewhere (Perez, Sauma, Munoz, & Hobbs, 2015) we have used this model to examine the impact of diverse state renewable energy requirements in the western US, with a focus on restrictions that states impose on trade in renewable credits. By providing more realistic representation of between-state power flows, more credible estimates have been obtained of the benefits, in terms of reduced costs and pollution, of lessening such trade restrictions. The model can assess the effect of those restrictions on the economic value of power trade and new transmission, as well as the interactive impacts of transmission investment and renewable credit trade limits upon the cost of integrating renewables into a region's various power markets.

## Acknowledgments

The authors would like to thank the anonymous referees for their feedback and suggestions to improve earlier versions of this manuscript. The research in this article was supported by the Consortium for Electric Reliability Technology Solutions (CERTS) funded by the U.S. DOE, the Fulbright Foundation, and the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contract number KJ0401000 through the Project "Multifaceted Mathematics for Complex Energy Systems," and the grant CONICYT/FONDAP/15110019. The authors are grateful for useful comments by John Birge, Mihai Anitescu, and Suvrajeet Sen. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

## Supplementary materials

Supplementary data associated with this article can be found, in the online version, at [10.1016/j.ejor.2015.07.057](https://doi.org/10.1016/j.ejor.2015.07.057).

## References

- Aardal, K., & Larsson, T. (1990). A Benders decomposition based heuristic for the hierarchical production planning problem. *European Journal of Operational Research*, 45(1), 4–14.
- Anderson, D. (1972). Models for determining least-cost investments in electricity supply. *Bell Journal of Economics and Management Science*, 3(1), 267–299.
- Anitescu, M., & Birge, J. R. (2008). Convergence of stochastic average approximation for stochastic optimization problems with mixed expectation and per-scenario constraints. *Technical Report*. Argonne National Laboratory.
- Asadamongkol, S., & Eua-arporn, B. (2013). Transmission expansion planning with AC model based on generalized Benders decomposition. *International Journal of Electrical Power & Energy Systems*, 47, 402–407.
- Baringo, L., & Conejo, A. (2013). Correlated wind-power production and electric load scenarios for investment decisions. *Applied Energy*, 101, 475–482.
- Baringo, L., & Conejo, A. J. (2012). Wind power investment: A Benders decomposition approach. *IEEE Transactions on Power Systems*, 27(1), 433–441.
- Bertsekas, D. P., & Sandell, N. R. (1982). Estimates of the duality gap for large-scale separable nonconvex optimization problems. In *21st IEEE conference on decision and control*: vol. 21 (pp. 782–785).
- Bertsimas, D., Litvinov, E., Sun, X. A., Zhao, J., & Zheng, T. (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1), 52–63.
- Binato, S., Pereira, M. V. F., & Granville, S. (2001). A new Benders decomposition approach to solve power transmission network design problems. *IEEE Transactions on Power Systems*, 16(2), 235–240.
- Birge, J., & Louveaux, F. (1997). *Introduction to stochastic programming*. New York, NY: Springer.
- Birge, J. R. (2011). Uses of sub-sample estimates in stochastic optimization models. In *Working Paper*. The University of Chicago.
- Birge, J. R., & Louveaux, F. V. (1988). A multicut algorithm for 2-stage stochastic linear programs. *European Journal of Operational Research*, 34(3), 384–392.
- Bloom, J. A. (1983). Solving an electricity generating capacity expansion planning problem by generalized Benders decomposition. *Operations Research*, 31(1), 84–100.
- Bloom, J. A., Caramanis, M., & Charny, L. (1984). Long-range generation planning using generalized Benders decomposition – Implementation and experience. *Operations Research*, 32(2), 290–313.
- Booth, R. R. (1972). Power-system simulation model based on probability analysis. *IEEE Transactions on Power Apparatus and Systems*, Pa91(1), 62–8.
- Caramanis, M. C., Tabors, R. D., Nochur, K. S., & Schweppe, F. C. (1982). The introduction of non-diispatchable technologies a decision variables in long-term generation expansion models. *IEEE Transactions on Power Apparatus and Systems*, PAS-101(8), 2658–2667.
- Cote, G., & Laughton, M. A. (1984). Large-scale mixed integer programming – Benders-type heuristics. *European Journal of Operational Research*, 16(3), 327–333.
- Dantzig, G., Glynn, P. W., Avriel, M., Stone, J., Entriken, R., Nakayama, M. (1989). Decomposition techniques for multi-area generation and transmission planning under uncertainty: Final Report. Electric Power Research Institute.
- Destz Zahlay, F., De Cuadra, F., Olmos, L., Rivier, M., & Perez-Arriaga, I. (2013). A formulation for large-scale transmission expansion planning problem and a solution strategy. In *European energy market (EEM), 2013 10th international conference on the* (pp. 1–8). IEEE.
- EIA (2013). MARKAL technology database and model. Retrieved March 20, 2012, from [http://www.epa.gov/nrmrl/appcd/climate\\_change/markal.htm](http://www.epa.gov/nrmrl/appcd/climate_change/markal.htm).
- Gabriel, S. A., Kydes, A. S., & Whitman, P. (2001). The national energy modeling system: A large-scale energy-economic equilibrium model. *Operations Research*, 49(1), 14–25.
- Geoffrion, A. M. (1972). Generalized Benders decomposition. *Journal of Optimization Theory and Applications*, 10(4), 237–260.
- Geoffrion, A. M., & Graves, G. W. (1980). Multicommodity distribution-system design by Benders decomposition. *Management Science*, 26(8), 855–856.
- Hart, W. E., Laird, C., Watson, J.-P., & Woodruff, D. L. (2012). *Pyomo – Optimization modeling in python*. New York, NY: Springer.
- Heejeung, P., & Baldick, R. (2013). Transmission planning under uncertainties of wind and load: Sequential approximation approach. *IEEE Transactions on Power Systems*, 28(3), 2395–2402.
- Hemmati, R., Hooshmand, R.-A., & Khodabakhshian, A. (2013). Comprehensive review of generation and transmission expansion planning. *IET Generation, Transmission & Distribution*, 7(9), 955–964.
- Higle, J. L., & Sen, S. (1991). Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16(3), 650–669.
- Hoang Hai, H. (1980). Topological optimization of networks: A nonlinear mixed integer model employing generalized Benders decomposition. In *19th IEEE conference on decision and control including the symposium on adaptive processes: vol. 19* (pp. 427–432).
- Hobbs, B. F. (1995). Optimization methods for electric utility resource planning. *European Journal of Operational Research*, 83(1), 1–20.
- Hobbs, B. F., & Ji, Y. D. (1999). Stochastic programming-based bounding of expected production costs for multiarea electric power systems. *Operations Research*, 47(6), 836–848.

- Holmberg, K. (1994). On using approximations of the Benders master problem. *European Journal of Operational Research*, 77(1), 111–125.
- Huang, C. C., Ziemba, W. T., & Ben-Tal, A. (1977). Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. *Operations Research*, 25(2), 315–325.
- Huang, W., & Hobbs, B. F. (1994). Optimal SO<sub>2</sub> compliance planning using probabilistic production costing and generalized Benders decomposition. *IEEE Transactions on Power Systems*, 9(1), 174–180.
- ICF (2013). Integrated planning model brochure. Retrieved June 10, 2012, from <http://www.icfi.com/insights/products-and-tools/ipm>.
- Infanger, G. (1992). Monte Carlo (Importance) sampling within a Benders decomposition algorithm for stochastic linear programs. *Annals of Operations Research*, 39(1), 69–95.
- Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1), 175–193.
- Joskow, P., & Tirole, J. (2005). Merchant transmission investment. *The Journal of industrial economics*, 53(2), 233–264.
- Joskow, P. L. (2011). Comparing the costs of intermittent and dispatchable electricity generating technologies. *American Economic Review*, 101(3), 238–241.
- Kahn, E. (1995). Regulation by simulation – The role of production cost models in electricity planning and pricing. *Operations Research*, 43(3), 388–398.
- Kazempour, S. J., & Conejo, A. J. (2012). Strategic generation investment under uncertainty via Benders decomposition. *IEEE Transactions on Power Systems*, 27(1), 424–432.
- Konstantelos, I., & Strbac, G. (2014). Valuation of flexible transmission investment options under uncertainty. *IEEE Transactions on Power Systems*, PP(99), 1–9.
- Krokhmal, P., Palmquist, J., & Uryasev, S. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4, 11–27.
- Latorre, G., Cruz, R. D., Areiza, J. M., & Villegas, A. (2003). Classification of publications and models on transmission expansion planning. *IEEE Transactions on Power Systems*, 18(2), 938–946.
- Law, A. M., & Carson, J. S. (1979). Sequential procedure for determining the length of a steady-state simulation. *Operations Research*, 27(5), 1011–1025.
- Lee, C., Ng, S. K., Zhong, J., & Wu, F. F. (2006). Transmission expansion planning from past to future. In *2006 IEEE PES power systems conference and exposition, 2006. PSCE'06*. (pp. 257–265). IEEE.
- Lumbreras, S., & Ramos, A. (2013). Transmission expansion planning using an efficient version of Benders' decomposition. A case study. In *PowerTech (POWERTECH), 2013 IEEE Grenoble* (pp. 1–7). IEEE.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability: vol. 1* (pp. 281–297). University of California Press.
- Madansky, A. (1960). Inequalities for stochastic linear-programming problems. *Management Science*, 6(2), 197–204.
- Magnanti, T. L., & Wong, R. T. (1981). Accelerating benders decomposition – Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3), 464–484.
- Marnay, C., & Strauss, T. (1991). Effectiveness of antithetic sampling and stratified sampling in Monte Carlo chronological production cost modeling [Power Systems]. *IEEE Transactions on Power Systems*, 6(2), 669–675.
- McDaniel, D., & Devine, M. (1977). Modified Benders partitioning algorithm for mixed integer programming. *Management Science*, 24(3), 312–319.
- Moreno, R., Pudjianto, D., & Strbac, G. (2013). Transmission network investment with probabilistic security and corrective control. *IEEE Transactions on Power Systems*, 28(4), 3935–3944.
- Munoz, F. D., Hobbs, B. F., Ho, J., & Kasina, S. (2014). An engineering-economic approach to transmission planning under market and regulatory uncertainties: WECC case study. *IEEE Transactions on Power Systems*, 29(1), 307–317.
- Munoz, F. D., Sauma, E. E., & Hobbs, B. F. (2013). Approximations in power transmission planning: Implications for the cost and performance of renewable portfolio standards. *Journal of Regulatory Economics*, 43(3), 305–338.
- Nweke, C. I., Leanez, F., Drayton, G. R., & Kolhe, M. (2012). Benefits of chronological optimization in capacity planning for electricity markets. In *IEEE international conference on power system technology (POWERCON)* (pp. 1–6).
- O'Brien, M. (2004). *Techniques for incorporating expected value constraints into stochastic programs* (Ph.D. thesis). Stanford University.
- Orfanos, G. A., Georgilakis, P., & Hatzigiorgyriou, N. D. (2013). Transmission expansion planning of systems with increasing wind power integration. *IEEE Transactions on Power Systems*, 28(2), 1355–1362.
- Ozdemir, O., Munoz, F. D., Ho, J. H., & Hobbs, B. F. (2015). Economic analysis of transmission expansion planning with price-responsive demand and quadratic losses by successive LP. *IEEE Transactions on Power Systems* (in press).
- Palmintier, B., & Webster, M. (2011). Impact of unit commitment constraints on generation expansion planning with renewables. In *Power and Energy Society General Meeting* (pp. 1–7). San Diego, CA: IEEE.
- Papavasiliou, A., & Oren, S. S. (2013). Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research*, 61(3), 578–592.
- Paul, A., Burtraw, D. (2002). The RFF Haiku electricity market model. Resources for the future. Retrieved March 10, 2012, from <http://www.rff.org/RFF/Documents/RFF-Rpt-Haiku.v2.0.pdf>.
- Pereira, M. V. F., Pinto, L. M. V. G., Cunha, S. H. F., & Oliveira, G. C. (1985). A decomposition approach to automated generation transmission expansion planning. *IEEE Transactions on Power Apparatus and Systems*, 104(11), 3074–3083.
- Perez, A. P., Sauma, E. E., Munoz, F. M., & Hobbs, B. F. (2015). The economic effects of interregional trading of renewable energy certificates in the WECC. *The Energy Journal*. accepted for publication.
- Pierre-Louis, P., Bayraksan, G., & Morton, D. P. (2011). A combined deterministic and sampling-based sequential bounding method for stochastic programming. *Proceedings of the 2011 Winter Simulation Conference (Wsc)*, 4167–4178.
- Pozo, D., Contreras, J., & Sauma, E. (2013a). If you build it, he will come: Anticipative power transmission planning. *Energy Economics*, 36, 135–146.
- Pozo, D., Sauma, E. E., & Contreras, J. (2013b). A three-level static MILP model for generation and transmission expansion planning. *IEEE Transactions on Power Systems*, 28(1), 202–210.
- Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2, 21–42.
- van Roy, T. J. (1983). Cross decomposition for mixed integer programming. *Mathematical Programming*, 25(1), 46–63.
- Sahinidis, N. V., & Grossmann, I. E. (1991). Convergence properties of generalized benders decomposition. *Computers & Chemical Engineering*, 15(7), 481–491.
- Schmeiser, B. (1982). Batch size effects in the analysis of simulation output. *Operations Research*, 30(3), 556–568.
- Sen, S. (2013). Discussion about the use of stabilization techniques for the stochastic decomposition algorithm in the NEOS Solver, Personal Communication.
- Sherali, H. D., & Staschus, K. (1990). A 2-phase decomposition approach for electric utility capacity expansion planning including nondispatchable technologies. *Operations Research*, 38(5), 773–791.
- Sherali, H. D., Staschus, K., & Huacuz, J. M. (1987). An integer programming approach and implementation for an electric utility capacity planning problem with renewable energy-sources. *Management Science*, 33(7), 831–845.
- Short, W., Sullivan, P., Mai, T., Mowers, M., Uriarte, C., Blair, N., et al. (2011). Regional energy deployment system (ReEDS). NREL/TP-6A2-46534. Golden, CO: National Renewable Energy Laboratory.
- Shortt, A., Kiviluoma, J., & O'Malley, M. (2013). Accommodating variability in generation planning. *IEEE Transactions on Power Systems*, 28(1), 158–169.
- de Sisternes, F. J., & Webster, M. (2013). Optimal selection of sample weeks for approximating the net load in generation planning problems. In *Working paper ESD-WP-2013-03*. Massachusetts Institute of Technology.
- Steiger, N. M., & Wilson, J. R. (2001). Convergence properties of the batch means method for simulation output analysis. *Inform Journal on Computing*, 13(4), 277–293.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 63, 411–423.
- Tseng, G. C. (2007). Penalized and weighted K-means for clustering with scattered objects and prior information in high-throughput biological data. *Bioinformatics*, 23(17), 2247–2255.
- Wagstaff, K., Cardie, C., Rogers, S., & Schr, S. (2001). Constrained K-means clustering with background knowledge. In *18th international conference on machine learning* (pp. 577–584).
- van der Weijde, A. H., & Hobbs, B. F. (2012). The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty. *Energy Economics*, 34(6), 2089–2101.
- Wogrin, S., Duenas, P., Delgado, A., & Reneses, J. (2014). A new approach to model load levels in electric power systems with high renewable penetration. *IEEE Transactions on Power Systems*, 29(5), 2210–2218.