JHU Working Paper
May 16, 2014

New Bounding and Decomposition Approaches for MILP Investment Problems: Multi-Area Transmission and Generation Planning Under Policy Constraints

Francisco D. Munoz
Analytics Department, Sandia National Laboratories, Albuquerque, NM 87123,
fdmunoz@sandia.gov

Benjamin F. Hobbs
Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, MD 21218,
bhobbs@jhu.edu

Jean-Paul Watson
Analytics Department, Sandia National Laboratories, Albuquerque, NM 87123,
jwatson@sandia.gov

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 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 new lower bound, which we also 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. Attaining tight optimality gaps, however, requires the decomposition phase. 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.

Key words: Benders decomposition, stochastic programing, investment planning, transmission and generation planning, policy constraints

1. Introduction

The electric power industry is a major area of applications of optimization (Hobbs 1995). This sector comprises over 2% of the U.S. economy, and recent restructuring has strengthened incentives for electric utilities to plan and operate power infrastructure efficiently. Increasing amounts of generation from renewable resources make optimization of short-term operations and long-term planning more challenging, and so promote 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 et al. 2013, Papavasiliou and Oren 2013). Investment planning models, because they consider both investment and operations, 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 and Webster 2011). To achieve computational tractability, investment planning instead utilized deterministic or probabilistic models for calculating production costs based on load-duration 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 et al. 1982), to account for the effect of non-dispatchable generation technologies, such as wind and solar, on the optimal generation mix. A simple approach is to select operating hours to be simulated by performing moment matching on demand, wind, solar, and hydro data (van der Weijde and 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 et al. (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 et al. 1984, Pereira et al. 1985, Sherali et al. 1987, Sherali and Staschus 1990, Huang and Hobbs 1994). Such approaches separate the investment problem (i.e., master problem) from the production cost problems (i.e., subproblems), which can then be solved 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 are an ongoing subject of research (McDaniel and Devine 1977, Magnanti and Wong 1981, Sahinidis and 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.

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 and Ji 1999) and Benders decomposition, both of which provide 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 et al. 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 the 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 companion to this paper.

2. Abstract Planning Model

We focus on investment planning models that can be formulated as linear or mixed integer linear programs.[1] Examples of such models include: Caramanis et al. (1982), Bloom (1983), and Sherali and Staschus (1990) for generation expansion planning; Binato et al. (2001) for transmission expansion planning; and Pereira et al. (1985), Dantzig et al. (1989), van der Weijde and Hobbs (2012), and Munoz et al. (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 et al. 2001) , ReEDS (Short et al. 2011), Haiku (Paul and Burtraw 2002), and MARKAL (EPA 2013).

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;
(Ω, p)	:	Discrete probability space composed of the sample space Ω and the probability measure $p(\cdot)$ over Ω. For planning purposes, this space can, for example, be constructed using 8,760 historical observations of hourly demand, wind, solar, and hydro data from a representative year (i.e., $\|\Omega\| = 8,760$). Each event ω_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 ω;
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 ω in Ω;

2.2. Two-Stage Investment-Operations Model

Broadly, the 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 (Ω, p). We formulate the planning problem as the following stochastic mixed-integer linear program:

$$TC((\Omega,p)) = \min_{x} \; e^T x + f(x,(\Omega,p)) \tag{1}$$

$$\text{s.t. } Ax \leq b \tag{2}$$

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

The function $f(x,(\Omega,p))$ denotes the minimum expected operating costs for a given set of investments x and scenarios described by (Ω,p). 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 et al. (2013), our application models transmission investments using binary variables and generation capacity as continuous. The formulation in (1)-(3) presumes that there is a single investment planning stage, i.e., all investments are here-and-now-variables, while all recourse variables are operations variables. However, 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 (Ω, p). The problem is formulated as a linear program:

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

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

$$\mathbb{E}_\omega[Ky(\omega)] \le d \tag{6}$$

$$y(\omega) \ge 0 \qquad \forall \omega \in \Omega \tag{7}$$

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).

3. Phase 1: The Bounding Algorithm

One alternative to solving large-scale stochastic programs (Birge and 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 Ω, until a certain optimality gap is achieved (Huang et al. 1977, Birge and Louveaux 1997, Hobbs and Ji 1999). However, expectation constraints in our case 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

and 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. New 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 Section EC.1 of the electronic companion.

PROPOSITION 1. *Given a discrete probability space Ω with measure p and a partition $S_1, ..., S_m$ of Ω, a sample space $\Psi^m = \{\xi_1, ..., \xi_m\}$ is defined with measure q^m such that the probability of each event ξ_i in Ψ^m equals the probability of each subset S_i, defined as $q^m(\xi_i) = p(S_i)$, $\forall i \in \{1, ..., m\}$. 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 $S_1, ..., S_m$ 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, $f(x, (\Psi^m, q^m)) \leq f(x, (\Omega, p))$.*

The interpretation of this result is that if the space Ω is partitioned or clustered into subsets, and if 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., $f(x, (\Psi^m, q^m)) \leq f(x, (\Psi^{m+1}, q^{m+1}))$, $\forall x$, $\forall m \in \{1, ..., |\Omega| - 1\}$) and convergent to $f(x, (\Omega, p))$ (Birge and Louveaux 1997). Finally, 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 $f(x, (\Psi^m, q^m)) \leq f(x, (\Psi^{m+1}, q^{m+1}))$, $\forall x$, $\forall m \in \{1, ..., |\Omega| - 1\}$, then $TC(\Psi^m, q^m) \leq TC(\Psi^{m+1}, q^{m+1})$, $\forall m \in \{1, ..., |\Omega| - 1\}$.

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 implementation 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. Further, 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 (Ω, p) and, therefore, large operations models that would be difficult to solve even in the absence of expectation constraints.

Other solution 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 et al. (2011), and within Benders decomposition in Infanger (1992) and Higle and Sen (1991). The quality of these approximations is progressively improved by increasing the sample size between iterations (Birge and Louveaux 1997, Pierre-Louis et al. 2011) or by combining information from multiple independent samples through Benders' cuts (Higle and 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 and 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 and 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 to reduce 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 Ω as Ω^M, and define a new probability measure $p^M(\cdot)$ such that all observations have the same probability of occurrence $p^M(\omega_i) = 1/M$, $\forall i \in \{1,...,M\}$. To approximate $f(x,(\Omega,p))$, we draw N independent samples of M observations each, denoted $\{\Omega_1^M,...,\Omega_N^M\}$, and solve N independent operations problems, denoted $f(x,(\Omega_1^M,p_1^M)),...,f(x,(\Omega_N^M,p_N^M))$. 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 and Wilson 2001). An advantage of this method over other approaches that rely on unique, large-sample results is that problems $f(x,(\Omega_1^M,p_1^M)),...,f(x,(\Omega_N^M,p_N^M))$ 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 Ω and, therefore reduce the variance of the sub-samples $f(x,(\Omega_1^M,p_1^M)),...,f(x,(\Omega_N^M,p_N^M))$. Stratified sampling has been previously used for electricity production cost modeling (Marnay and Strauss 1991), but has yet to be utilized yet 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 Ω is partitioned into disjoint subsets $S_1,...,S_M$, 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 $T(\omega)$). A stratified sample of M observations $\{\omega_1,...,\omega_M\}\subset\Omega$ is such that $\omega_i\in S_i$, $\forall i\in\{1,...,M\}$. We weight each observation ω_i in the operations problem with probability $p^M(\omega_i)=p(\omega_i)/p(\omega_i|S_i)$ $\forall i\in\{1,...,M\}$.

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 and Louveaux 1997, Hobbs and 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. A experimental analysis of the effects of sample sizes and sampling methodologies is given in Section EC.3 of the electronic companion to this paper. 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 M, 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 (Ψ^k,q^k) in (1), instead of the full space (Ω,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 $e^Tx_k^*+f(x_k^*,(\Omega,p))<UB_{k-1}$, update the upper bound $UB_k=e^Tx_k^*+f(x_k^*,(\Omega,p))$ and the incumbent solution $x^*=x_k^*$; otherwise, $UB_k=UB_{k-1}$.
3. Compute the optimality gap, defined as $GAP_k=100\%\times(UB_k-LB_k)/UB_k$. 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 (Ω,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 $k\to|\Omega|$. A formal proof of convergence is given in the electronic companion to this paper.

It is often observed in clustering algorithms that only a few partitions explain a large fraction of the variance of the full dataset (the "elbow" phenomenon), but that the remaining fraction of variance explained converges asymptotically to 1 as the partitions are refined (Tibshirani et al. 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 and Ji 1999). This is particularly challenging for planning problems that consider binary decision variables (e.g., transmission investments), since solving the lower-bound planning problems $TC(\Psi^k,q^k)$ using a branch-and-bound type of algorithm for large values of k can 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 and Wong 1981, Sahinidis and Grossmann 1991). Other techniques address the computational challenge of solving multiple mixed-integer linear master problems by initially computing cuts from linear (McDaniel and Devine 1977) and Lagrangian relaxations (Hoang Hai 1980, van Roy 1983, Cote and Laughton 1984, Aardal and Larsson 1990), as well as from feasible sub-optimal solutions found by prematurely stopping branch-and-bound type algorithms (Geoffrion and 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 et al. 1987, Sherali and Staschus 1990). However, approximate solution methods may obviate the conditions needed to guarantee convergence of Benders decomposition to the 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 and 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:

$$\min_{x,\theta} \; e^T x + \theta \tag{9}$$

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

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

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

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

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 Equations (4) - (7) for the sample space (Ψ^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 Ω. For $k = 0$, no constraints on the value of θ are imposed through constraint (12), so that the problem defined by Equations (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 suitable 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.[2]

Finally, as in the L-shaped method for stochastic programs (Birge and Louveaux 1997), we compute a single cut at each Benders iteration using the expected value of the dual multipliers $\pi_1(x^*), ..., \pi_N(x^*)$ and the operating costs $f(x^*, (\Omega_1^M, p_1^M)), ..., f(x^*, (\Omega_N^M, p_N^M))$ 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 and 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.

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 Section EC.2 of the electronic companion. The planning problem formulated using $8{,}736^3$ 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. 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 $1,000 per MWh (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 MWh.

The clustering, bounding, and Benders decomposition algorithms are all implemented using the Pyomo algebraic modeling package (Hart et al. 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 Optetron processors of 2.2 GHz and 112 GB 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 and Ji 1999). 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 and 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% of the variance of the normalized data set of 8,736 observations of loads, wind, solar, and hydro (see Figure 1). The point of diminishing returns (i.e., the elbow) is reached at approximately 50 clusters, capturing 46.4% of variance. After this point, capturing an extra 10% of variance requires partitioning the space of loads, wind, solar, and hydro levels using 100 additional 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 allow 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 EPA (2013). Figure 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. The lower bound (LB) is computed solving the planning problem using one representative hour from cluster (i.e., average of all observations of loads, wind, solar, and hydro

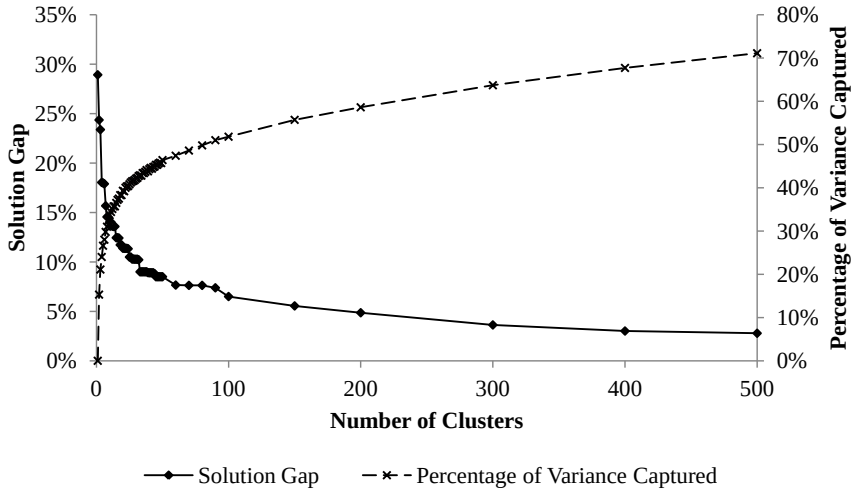

Figure 1 Optimality gap for the linear relaxation and percentage of variance captured as a function of the number of clusters.

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 8,736 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% of the global optimum (Figure 1), but more than 200 clusters are required to reduce the gap further to 5%. The "knee" in the optimality gap mirrors the elbow associated with the percentage of variance captured from the 8,736 hour dataset, which decreases at a much slower rate after the first 39 clusters. The optimality gap is reduced from 28.9% to 8.9% with the first 39 clusters and increasing the number of partitions to 500 only reduces the gap to 2.8%, but 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 ϵ 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% for a 1% MILP gap and that 90 additional clusters only reduce the optimality gap to 7.4%. Optimality gaps and bounds as a function of the number of clusters are shown in Figures EC.4 and EC.5, respectively, of the electronic companion. 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%, 3%, and 1% MILP gaps, respectively. The reduction of 5.5 hours in solution time achieved from loosening the MILP gap from 1% to 5% is, however, contrasted with an increase in the total optimality gap from 7.4% to 11.1%. Further attempts to solve mixed-integer linear problems with 200 or more clusters and a MILP gap of 1% 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

To summarize the foregoing, the bounding algorithm can efficiently find investment solutions whose total system costs have a 2.8% optimality gap for the linear relaxation and a 7.4% 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 and 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., subproblems).

5.3.1. Linear Problem (LP) Figure 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 Figure 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 $TC_{LP}(\Psi^{150}, q^{150})$ 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.

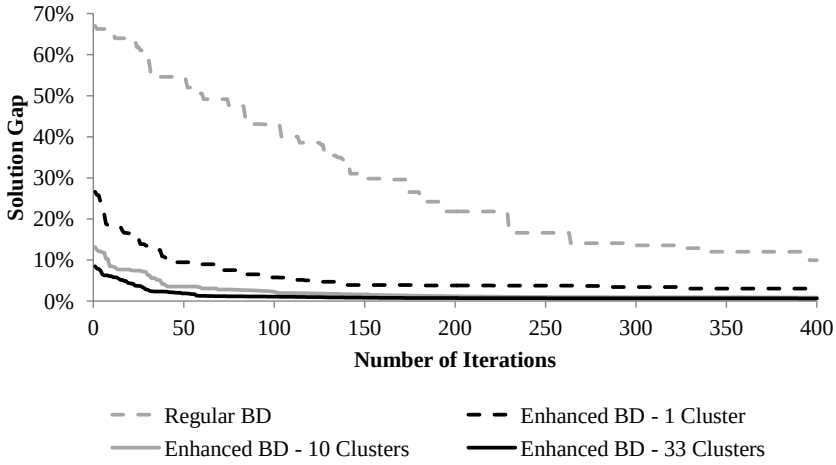

Figure 2 Optimality gap versus the number of iterations for different Benders decomposition schemes of the linear relaxation. Upper and lower bounds are in Figure EC.6 in the electronic companion.

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	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	6,026
3.0%	329	95,053	69	21,018	29	10,359
1.0%	>400	>117,332	246	78,917	116	44,001

In Figure 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% to 3% 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% 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 solution time to achieve a 1% 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).

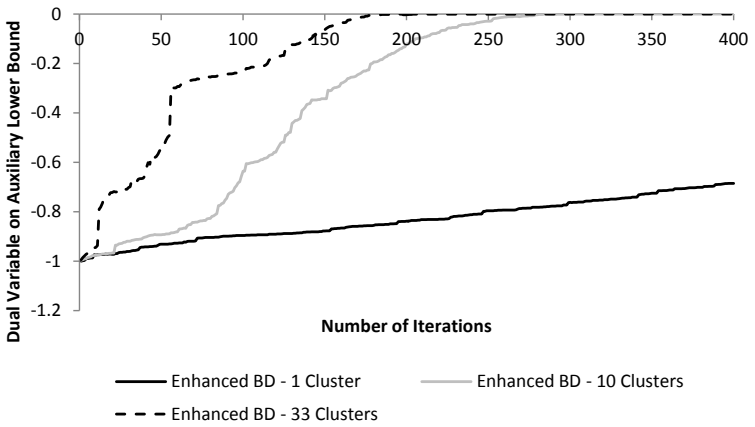

Figure 3 Dual multiplier on auxiliary lower bound versus number of iterations of Benders decomposition for the linear relaxation.

Larger speedups can be attained 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% 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 towards convergence of the UB and LB: Figure 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 modified 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%.

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% 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% 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%. In contrast, the best investment plan found without pre-computed cuts results in a 7% 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 $TC_{LP}(\Psi^{150}, q^{150}) = \$624.3\text{B}$ to initialize the lower bound ($LB_0 = \$624.3\text{B}$). 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%) and in the highest lower bound ($635.5B) after 400 iterations and 44 hours of computation time (see Figure EC.3 in the electronic companion). If this value is used as an initial lower bound ($LB_0 = \$635.5\text{B}$), the optimality gap after 200 iterations is reduced from 4.1% to 2.4%, which is arguably sufficient for long-term planning studies. Attaining even tighter optimality gaps requires more iterations of the algorithm, but these come at the expense of longer solution times beyond the scope. 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 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 again the high-resolution operations problem, which is used to compute upper bounds[4]

We also observe that the lower-bound problem builds too little transmission capacity relative to near-optimal solutions for small cluster counts, which increases penalties due to curtailed load and noncompliance fines in the upper bound. Modifications to the clustering algorithm could potentially address this issue by weighting (Tseng 2007) or constraining (Wagstaff et al. 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 and Oren 2013). However, these are all subjects of future research.

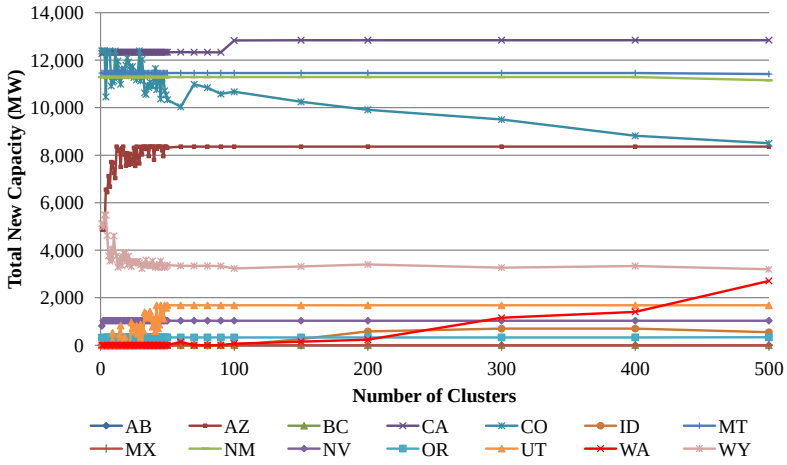

Figure 4 Aggregate investments in wind capacity per state as a function of the number of clusters for the linear problem.

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. 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% and by 42% in the states of Utah (UT) and Arizona (AZ), respectively, and overestimates investments in Wyoming (WY) by 60% with respect to the optimal levels of the 500-cluster problem (see Figure 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%, 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 Figure 1). The elbow criterion, often used to determine the number of clusters in a dataset (Tibshirani et al. 2001), can be used in our case to identify the point when the clustering algorithm becomes inefficient, and when it may be better 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 approaches proposed in 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% to 6% 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 Jensens' 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 (Palmintier and Webster 2011, Nweke et al. 2012, Shortt et al. 2013). This would, however, require including binary variables in the operations problems which would then 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.

Endnotes

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

2. 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.

3. The sample is weighted by 8,760/8,736 in the objective function and expectation constraints of the operations problem. The sample size of 8,736 hours results from considering 52 weeks of hourly solar data for a typical year.

4. Upper and lower bounds for the linear relaxation are in Figure EC.3 in the electronic companion. 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% (Figure 1)

Acknowledgments

The research in this article was supported by the Consortium for Electric Reliability Technology Solutions (CERTS) funded by the U.S. DOE, the Department of Energy's Office of Advanced Scientific Computing

Research, and the Fulbright Foundation. The authors are grateful by 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.

References

Aardal, K., T. Larsson. 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, Mihai, John R. Birge. 2008. Convergence of Stochastic Average Approximation for Stochastic Optimization Problems with Mixed Expectation and Per-Scenario Constraints. Tech. rep., Argonne National Laboratory.

Baringo, L., A. J. Conejo. 2012. Wind Power Investment: A Benders Decomposition Approach. *IEEE Transactions on Power Systems* **27**(1) 433–441.

Bertsekas, D. P., N. R. Sandell. 1982. Estimates of the Duality Gap for Large-Scale Separable Nonconvex Optimization Problems. *21st IEEE Conference on Decision and Control*, vol. 21. 782–785.

Bertsimas, D., E. Litvinov, X. A. Sun, J. Zhao, T. Zheng. 2013. Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem. *IEEE Transactions on Power Systems* **28**(1) 52–63.

Bertsimas, Dimitris, John Tsitsiklis. 1997. *Introduction to Linear Optimization*. Nashua, NH:Athena Scientific.

Binato, S., M. V. F. Pereira, S. Granville. 2001. A New Benders Decomposition Approach to Solve Power Transmission Network Design Problems. *IEEE Transactions on Power Systems* **16**(2) 235–240.

Birge, J. R. 2011. Uses of Sub-sample Estimates in Stochastic Optimization Models. Working Paper, The University of Chicago.

Birge, J. R., F. V. Louveaux. 1988. A Multicut Algorithm for 2-Stage Stochastic Linear-Programs. *European Journal of Operational Research* **34**(3) 384–392.

Birge, J. R., S. W. Wallace. 1986. Refining Bounds for Stochastic Linear-Programs with Linearly Transformed Independent Random-Variables. *Operations Research Letters* **5**(2) 73–77.

Birge, John, Francois Louveaux. 1997. *Introduction to Stochastic Programming*. New York, NY:Springer.

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., M. Caramanis, L. Charny. 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–&.

Caramanis, M. C., R. D. Tabors, Kumar S. Nochur, F. C. Schweppe. 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., M. A. Laughton. 1984. Large-Scale Mixed Integer Programming - Benders-Type Heuristics. *European Journal of Operational Research* **16**(3) 327–333.

Dantzig, G.B., P. W. Glynn, M. Avriel, J. Stone, R. Entriken, M. Nakayama. 1989. Decomposition Techniques for Multi-area Generation and Transmission Planning Under Uncertainty: Final Report. Electric Power Research Institute.

de Sisternes, F. J., M. Webster. 2013. Optimal Selection of Sample Weeks for Approximating the Net Load in Generation Planning Problems. Working paper ESD-WP-2013-03, Massachusetts Institute of Technology.

EPA. 2013. MARKAL Technology Database and Model. Retrieved March 20, 2012, from http://www.epa.gov/nrmrl/appcd/climate_change/markal.htm.

Gabriel, S. A., A. S. Kydes, P. Whitman. 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., G. W. Graves. 1980. Multicommodity Distribution-System Design by Benders Decomposition. *Management Science* **26**(8) 855–856.

Hart, William E., Carl Laird, Jean-Paul Watson, David L. Woodruff. 2012. *Pyomo - Optimization Modeling in Python*. New York, NY:Springer.

Heejung, Park, R. Baldick. 2013. Transmission Planning Under Uncertainties of Wind and Load: Sequential Approximation Approach. *IEEE Transactions on Power Systems* **28**(3) 2395–2402.

Higle, Julia L., Suvrajeet Sen. 1991. Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse. *Mathematics of Operations Research* **16**(3) 650–669.

Hoang Hai, Hoc. 1980. Topological Optimization of Networks: A Nonlinear Mixed Integer Model Employing Generalized Benders Decomposition. *19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, vol. 19. 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., Y. D. Ji. 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., W. T. Ziemba, A. Ben-Tal. 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., B. F. Hobbs. 1994. Optimal SO2 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, Gerd. 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. 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.

Kall, P., J. Mayer. 2010. *Stochastic Linear Programming: Models, Theory, and Computation*. New York, NY:Springer.

Kazempour, S. J., A. J. Conejo. 2012. Strategic Generation Investment Under Uncertainty Via Benders Decomposition. *IEEE Transactions on Power Systems* **27**(1) 424–432.

Krokhmal, P, J. Palmquist, S. Uryasev. 2002. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* **4**(11-27).

Kuhn, D. 2009. Convergent Bounds for Stochastic Programs with Expected Value Constraints. *Journal of Optimization Theory and Applications* **141**(3) 597–618.

Law, A. M., J. S. Carson. 1979. Sequential Procedure for Determining the Length of a Steady-State Simulation. *Operations Research* **27**(5) 1011–1025.

MacQueen, James. 1967. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 281–297.

Madansky, A. 1960. Inequalities for Stochastic Linear-Programming Problems. *Management Science* **6**(2) 197–204.

Magnanti, T. L., R. T. Wong. 1981. Accelerating Benders Decomposition - Algorithmic Enhancement and Model Selection Criteria. *Operations Research* **29**(3) 464–484.

Marnay, C., T. Strauss. 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., M. Devine. 1977. Modified Benders Partitioning Algorithm for Mixed Integer Programming. *Management Science* **24**(3) 312–319.

Munoz, F. D., B. F. Hobbs, J. Ho, S. Kasina. 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., E. E. Sauma, B. F. Hobbs. 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., F. Leanez, G. R. Drayton, M. Kolhe. 2012. Benefits of Chronological Optimization in Capacity Planning for Electricity Markets. *IEEE International Conference on Power System Technology (POWERCON)*. 1–6.

O'Brien, M. 2004. Techniques for Incorporating Expected Value Constraints Into Stochastic Programs. Ph.D. thesis, Stanford University.

Palmintier, B., M. Webster. 2011. Impact of Unit Commitment Constraints on Generation Expansion Planning with Renewables. *IEEE Power and Energy Society General Meeting* .

Papavasiliou, A., S. S. Oren. 2013. Multiarea Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network. *Operations Research* **61**(3) 578–592.

Paul, A., D. Burtraw. 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., L. M. V. G. Pinto, S. H. F. Cunha, G. C. Oliveira. 1985. A Decomposition Approach to Automated Generation Transmission Expansion Planning. *IEEE Transactions on Power Apparatus and Systems* **104**(11) 3074–3083.

Pierre-Louis, P., G. Bayraksan, D. P. Morton. 2011. A Combined Deterministic and Sampling-Based Sequential Bounding Method for Stochastic Programming. *Proceedings of the 2011 Winter Simulation Conference (Wsc)* 4167–4178.

Sahinidis, N. V., I. E. Grossmann. 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, Suvrajeet. 2013. Discussion About the Use of Stabilization Techniques for the Stochastic Decomposition Algorithm in the NEOS Solver, Personal Communication.

Sherali, H. D., K. Staschus. 1990. A 2-Phase Decomposition Approach for Electric Utility Capacity Expansion Planning Including Nondispatchable Technologies. *Operations Research* **38**(5) 773–791.

Sherali, H. D., K. Staschus, J. M. Huacuz. 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., P. Sullivan, T. Mai, M. Mowers, C. Uriarte, N. Blair, D. Heimiller, A. Martinez. 2011. Regional Energy Deployment System (ReEDS). NREL/TP-6A2- 46534. Golden, CO: National Renewable Energy Laboratory.

Shortt, A., J. Kiviluoma, M. O'Malley. 2013. Accommodating Variability in Generation Planning. *IEEE Transactions on Power Systems* **28**(1) 158–169.

Steiger, N. M., J. R. Wilson. 2001. Convergence Properties of the Batch Means Method for Simulation Output Analysis. *Informs Journal on Computing* **13**(4) 277–293.

Tibshirani, R., G. Walther, T. Hastie. 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.

van der Weijde, A. H., B. F. Hobbs. 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.

van Roy, T. J. 1983. Cross Decomposition for Mixed Integer Programming. *Mathematical Programming* **25**(1) 46–63.

Wagstaff, K., C. Cardie, S. Rogers, S. Schr. 2001. Constrained K-means Clustering with Background Knowledge. *In 18th International Conference on Machine Learning*. 577–584.

Electronic Companion for "New Bounding and Decomposition Approaches for MILP Investment Problems: Multi-Area Transmission and Generation Planning Under Policy Constraints" by Francisco D. Munoz, Benjamin F. Hobbs, and Jean-Paul Watson

EC.1. Supporting Material for Section 3.1

For the development of new lower bounds, we first define the function $g(x,(\Omega,p))$ as a relaxation of $f(x,(\Omega,p))$ that only includes per-scenario constraints (thus omitting (6)). Because $g(x,(\Omega,p))$ is a function that involves solving a stochastic linear optimization program with stochastic right-hand-sides and no cross-scenario constraints, the standard lower bound based on Jensen's inequality can be invoked as in the Lemma EC.1 below.

LEMMA EC.1. *Given a discrete sample space Ω with measure p and a partition $S_1,...,S_m$ of Ω, a sample space $\Psi^m=\{\xi_1,...,\xi_m\}$ 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, i.e., $q^m(\xi_i)=p(S_i)$, $\forall i \in 1,...,m$. If the right-hand-side vector of parameters $r(\cdot)$ and the transition matrix $T(\cdot)$ are computed using the expected value of these parameters over the partitions, 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, $g(x,(\Psi^m,q^m)) \leq g(x,(\Omega,p))$.*

Proof of Lemma EC.1. The result follows from the convexity of the optimal objective function of linear programs on the right-hand-side vector of constraints and the application of Jensen's inequality (Huang et al. 1977, Birge and Louveaux 1997) $\square$

If the sample space Ω is partitioned using a hierarchical clustering algorithm (i.e., Ψ^{m+1} is derived from Ψ^m by subdividing one (or more) of the subsets $S_1,...,S_m$ that define Ψ^m), then the bound always improves as the partitions are refined (i.e., $g(x,(\Psi^m,q^m)) \leq g(x,(\Psi^{m+1},q^{m+1}))$ $\forall m \in \{1,...,|\Omega|\}$) (Birge and Louveaux 1997). Convergence of the lower bounds $g(x,(\Psi^m,q^m)) \rightarrow g(x,(\Omega,p))$ is guaranteed as $m \rightarrow |\Omega|$ (Birge and Wallace 1986, Kall and Mayer 2010). Comparisons of the effect of different partitioning rules on convergence rates are presented in Birge and Wallace (1986) and Hobbs and Ji (1999). Unfortunately these properties are not directly applicable to our operations problem $f(x,(\Omega,p))$, which has both per-scenario and expected-value constraints.

By definition, the relaxation $g(x,(\Omega,p))$ provides a lower bound on $f(x,(\Omega,p))$, and thereby $f(x,(\Psi^m,q^m)) \leq f(x,(\Omega,p)), \forall x=(x_1,x_2),\ x_1 \in \{0,1\},\ x_2 \geq 0, \forall m \in 1,,|\Omega|$. However, these bounds, which relax the expectation constraints in $f(x,(\Omega,p))$, are likely to be loose if these constraints are active in an optimal solution to $f(x,(\Omega,p))$. In Munoz et al. (2014), for instance, renewable goals and emissions limits, both formulated as expectation constraints, are the main drivers of

transmission and generation investments in the model; their relaxation would therefore be expected to result in substantial underestimation of costs.

To develop tighter bounds we, first define the partial Lagrangian dual function $\phi(\lambda, x, (\Omega, p))$ of the minimization problem $f(x, (\Omega, p))$ as:

$$\phi(\lambda, x, (\Omega, p)) = \phi_c(\lambda, x, (\Omega, p)) = \underset{y(\omega)}{Min}\ \mathbb{E}_\omega[c^T y(\omega)] + \lambda^T(\mathbb{E}_\omega[Ky(\omega)] - d) \tag{EC.1}$$

$$\text{s.t. } Wy(\omega) \leq r(\omega) - T(\omega)x \qquad \forall \omega \in \Omega \tag{EC.2}$$

$$y(\omega) \geq 0 \qquad \forall \omega \in \Omega \tag{EC.3}$$

The weak duality theorem states that $\forall \lambda \geq 0$, $\phi(\lambda, x, (\Omega, p)) \leq f(x, (\Omega, p))$, while strong duality ensures that $\exists \lambda^* \geq 0$ such that $\phi(\lambda^*, x, (\Omega, p)) = f(x, (\Omega, p))$ (Bertsimas and Tsitsiklis 1997). Strong duality holds because the objective function and constraints are all affine functions (Bertsimas and Tsitsiklis 1997). The following proposition extends Lemma EC.1 to stochastic linear programs with both per-scenario and expectation constraints.

PROPOSITION 1. *Given a discrete sample space Ω with measure p and a partition $S_1, ..., S_m$ of Ω, a sample space $\Psi^m = \{\xi_1, ..., \xi_m\}$ 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, i.e., $q^m(\xi_i) = p(S_i)$, $\forall i \in \{1, ..., m\}$. If the right-hand-side vector of parameters $r(\cdot)$ and the transition matrix $T(\cdot)$ are computed using the expected value of these parameters over the partitions, 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, $f(x, (\Psi^m, q^m)) \leq f(x, (\Omega, p))$.*

Proof of Proposition 1. From the weak duality theorem it follows that $\forall \lambda \geq 0$ $\phi_c(\lambda, x, (\Omega, p)) \leq f(x, (\Omega, p))$. Defining a new cost vector $c_\lambda^T = c^T + \lambda^T K$ and re-arranging terms in the objective function of $\phi_c(\lambda, x, (\Omega, p))$, we have $g_{c_\lambda}(x, (\Omega, p)) - \lambda^T d = \phi_c(\lambda, x, (\Omega, p))$. Now Lemma 1 (Jensen's lower bound) can be applied to $g_{c_\lambda}(x, (\Psi^m, q^m))$, implying that $\forall m \in \{1, ..., m\}$ $g_{c_\lambda}(x, (\Psi^m, q^m)) - \lambda^T d \leq g_{c_\lambda}(x, (\Omega, p)) - \lambda^T d$. Replacing c_λ^T by $c^T + \lambda^T K$ in the objective function of $g_{c_\lambda}(x, (\Psi^m, q_m))$ and re-arranging terms, we obtain $\phi_c(\lambda, x, (\Psi^m, q^m)) = g_{c_\lambda}(x, (\Psi^m, q^m)) - \lambda^T d$. Using the strong duality theorem, we pick λ^{*m} such that $\phi(\lambda^{*m}, x, (\Psi^m, q^m)) = f(x, (\Psi^m, q^m))$, which implies:

$$f(x, (\Psi^m, q^m)) = \phi(\lambda^{*m}, x, (\Psi^m, q^m)) \leq \phi_c(\lambda^{*m}, x, (\Omega, p)) \leq f(x, (\Omega, p)) \tag{EC.4}$$

□

From left to right in (EC.4): the equality results from the strong duality theorem; the first inequality follows from Jensen's inequality through Lemma 1; the last inequality derives from the weak duality theorem. Similar arguments were used in Kuhn (2009) to develop convergent bounds for multi-stage stochastic problems with expectation constraints, but our proof is much simpler. Convergence of $f(x, (\Psi^m, q^m))$ to $f(x, (\Omega, p))$ as $m \to |\Omega|$, for any trial investment plan x, follows directly from propositions 3.7 and 4.6 in Birge and Wallace (1986).

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

Proof of Proposition 2 Assume x^* and x^{m*} are the investment plans that yield the minimum total system costs for the full-resolution problem $TC(\Omega, p)$ and the partitioned/clustered problem $TC(\Psi^m, q^m)$, respectively. From Proposition 1 we have that for x^*, $e^T x^* + f(x^*, (\Psi^m, q^m)) \leq e^T x^* + f(x^*, (\Omega, p)) = TC(\Omega, p)$, $\forall m \in \{1, ..., |\Omega|\}$. From the definition of x^{m*}, $e^T x^{*m} + f(x^{*m}, (\Psi^m, q^m)) = TC(\Psi^m, q^m) \leq e^T x^* + f(x^*, (\Psi^m, q^m))$, which implies that $TC(\Psi^m, q^m) \leq TC(\Omega, p)$, $\forall m \in \{1, ..., |\Omega|\}$ □

PROPOSITION EC.1. *Assume the conditions described in Proposition 2 and consider an optimal investment plan* x^m *that yields total system cost* $TC((\Psi^m, q^m))$. *Then, both the lower bound* $TC((\Psi^m, q^m))$ *and the upper bound* $e^T x^m + f(x^m, (\Omega, p))$ *are* ϵ*-convergent to* $TC((\Omega, p))$ *as* $m \to |\Omega|$.

Proof of Proposition EC.1 From Propositions 3.7 and 4.6 in Birge and Wallace (1986), $\forall \epsilon > 0$ $\exists N_\epsilon \in [1, |\Omega|]$ such that $\forall m \geq N_\epsilon$, $|f(x, (\Omega, p)) - f(x, (\Psi^m, q^m))| < \epsilon$ for any candidate investment plan x. In particular, for the optimal investment plan x^m, $|f(x^m, (\Omega, p)) - f(x^m, (\Psi^m, q^m))| < \epsilon$. Adding and subtracting the capital investment costs $e^T x^m$, the last inequality is equivalent to:

$$|e^T x^m + f(x^m, (\Omega, p)) - (e^T x^m + f(x^m, (\Psi^m, q^m)))| < \epsilon \tag{EC.5}$$

From Proposition 2, $TC((\Psi^m, q^m)) = e^T x^m + f(x^m, (\Psi^m, q^m)$ is a lower bound on the optimal total system cost $TC((\Omega, p))$. This directly implies:

$$|e^T x^m + f(x^m, (\Omega, p)) - TC((\Omega, p))| \leq |e^T x^m + f(x^m, (\Omega, p)) - (e^T x^m + f(x^m, (\Psi^m, q^m)))| \tag{EC.6}$$

From Section 3.2 of the paper, $e^T x^m + f(x^m, (\Omega, p))$ is an upper bound on the optimal total system cost $TC((\Omega, p))$. Consequently:

$$|TC((\Omega, p)) - (e^T x^m + f(x^m, (\Psi^m, q^m)))| \leq |e^T x^m + f(x^m, (\Omega, p)) - (e^T x^m + f(x^m, (\Psi^m, q^m)))| \tag{EC.7}$$

By replacing the expressions EC.6 and EC.7 into EC.5, we obtain that $\forall \epsilon > 0$ $\exists N_\epsilon \in [1, |\Omega|]$ such that $\forall m \geq N_\epsilon$:

$$|e^T x^m + f(x^m, (\Omega, p)) - TC((\Omega, p))| < \epsilon \tag{EC.8}$$

$$|TC((\Omega, p)) - (e^T x^m + f(x^m, (\Psi^m, q^m)))| < \epsilon \tag{EC.9}$$

Equations EC.8 and EC.9 imply ϵ-convergence of the upper bound to the optimal total system cost $TC((\Omega, p))$ □

EC.2. Description of the WECC test case

Our numerical experiments were performed using a modified version of the two-stage transmission and generation investment model used in Munoz et al. (2014). The network, depicted in Figure EC.1, consists of 240 existing buses, 448 transmission elements, and 157 aggregated generators. Generation build limits are given in Table EC.1.

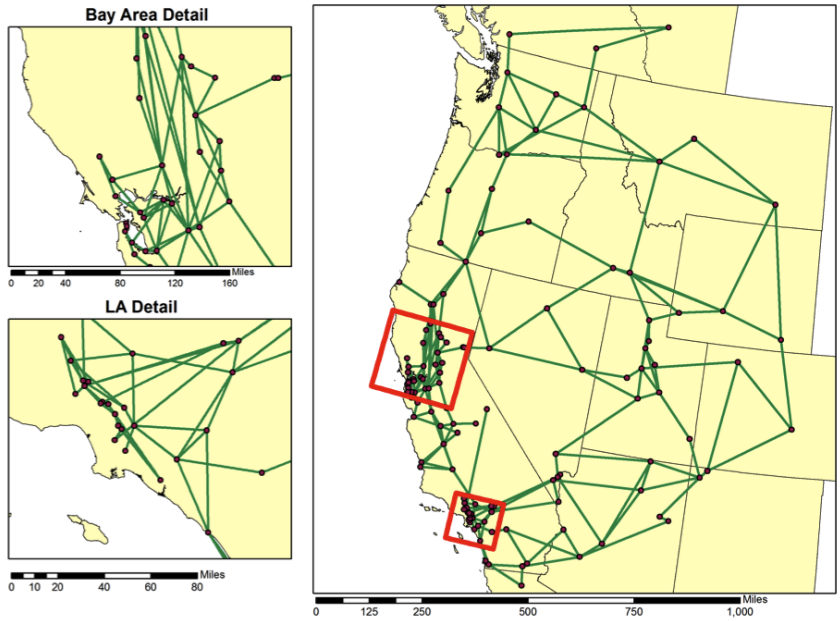

Figure EC.1 Map of the 240-bus representation of the WECC.

Table EC.1 Maximum installed generation capacity per technology and state.

State	Maximum Installed Generation Capacity per Technology (MW)									Total per State (MW)
	Biomass	CCGT	CCGT CCS	Coal	CT	Geothermal	Hydro	Solar	Wind	
AB	269	20,000	20,000	5,275	20,000		1,890		1,959	69,393
AZ	489	40,000	40,000	2,685	40,000			74,361	8,363	205,898
BC	880	20,000	20,000		20,000	308	5,926			67,114
CA		340,000	340,000		340,000			62,569	14,777	1,097,346
CO	141	40,000	40,000	3,978	40,000			4,943	12,394	141,456
ID	358	20,000	20,000	13	20,000	279	8		1,603	62,261
MT	147	40,000	40,000	2,225	40,000				11,458	133,830
MX		20,000	20,000		20,000					60,000
NM	44	40,000	40,000	5,889	40,000			183	11,290	137,406
NV	318	60,000	60,000	1,037	60,000	1,368	2	56,303	1,228	240,256
OR	349	60,000	60,000	452	60,000	832	557		4,443	186,633
UT	3	20,000	20,000	5,091	20,000	225		15,868	1,678	82,865
WA	448	60,000	60,000	1,290	60,000	32	165		5,463	187,398
WY	29	20,000	20,000	4,675	20,000				22,093	86,797
Total per Technology (MW)	3,475	800,000	800,000	32,610	800,000	3,044	8,548	214,227	96,749	2,758,653

Time-varying resources were simulated using 151 historical profiles of hourly demand, wind, solar, and hydro levels across multiple regions representing operating conditions for a typical year. Table EC.2 shows the population moments of a sample of 18 profiles of load, wind, and solar parameters. The full dataset consists of 8,736 observations of 22 demand profiles, 67 wind profiles (18 existing + 49 candidate locations), 31 solar profiles (2 existing + 29 candidate), 29 hydro profiles (27 existing + 2 candidate), and 2 profiles of existing biomass powered generators.

Table EC.2 Means, standard deviations, and correlations for a sample of load, wind, and solar profiles.

		Demand profiles per region					Wind profiles per state							Solar profiles per state			
		SOUTHWST	BAYAREA	NORTHWST	CANADA	ROCKYMT	CA	AZ	NM	CO	NV	AZ	WA	CA	AZ	NM	CO
Demand	SOUTHWST	1.00															
	BAYAREA	0.71	1.00														
	NORTHWST	0.42	0.70	1.00													
	CANADA	0.25	0.68	0.91	1.00												
	ROCKYMT	0.67	0.82	0.86	0.79	1.00											
Wind	CA	-0.27	-0.28	-0.21	-0.21	-0.26	1.00										
	AZ	-0.21	-0.18	-0.15	-0.13	-0.18	0.49	1.00									
	NM	-0.27	-0.15	-0.01	0.06	-0.09	0.20	0.38	1.00								
	CO	-0.31	-0.19	-0.03	0.04	-0.14	0.23	0.41	0.77	1.00							
	NV	-0.30	-0.12	0.04	0.08	-0.04	0.46	0.37	0.20	0.20	1.00						
	AZ	-0.33	-0.16	-0.02	0.03	-0.11	0.48	0.55	0.22	0.26	0.83	1.00					
	WA	-0.16	-0.11	-0.03	0.06	-0.06	0.09	0.08	0.18	0.27	-0.03	0.03	1.00				
Solar	CA	0.23	0.40	0.32	0.32	0.36	-0.07	-0.10	0.02	-0.01	-0.01	-0.07	0.05	1.00			
	AZ	0.25	0.43	0.32	0.34	0.37	-0.08	-0.10	0.01	-0.02	-0.01	-0.07	0.03	0.96	1.00		
	NM	0.24	0.42	0.33	0.34	0.37	-0.07	-0.10	0.02	-0.01	-0.01	-0.07	0.04	0.96	0.98	1.00	
	CO	0.23	0.41	0.33	0.34	0.37	-0.06	-0.09	0.02	0.00	-0.01	-0.06	0.04	0.96	0.97	0.98	1.00
Mean		15,077	4,246	21,299	14,160	9,593	0.37	0.26	0.31	0.34	0.27	0.26	0.23	0.24	0.22	0.23	0.22
Std. Dev.		3,328	741	3,289	1,571	1,348	0.29	0.25	0.26	0.27	0.25	0.21	0.21	0.32	0.30	0.30	0.29

For illustration purposes, we assume that market and regulatory conditions are deterministic and focus instead on capturing the variability of intermittent resources. The planning model used to illustrate our algorithm is a single-scenario version of the two-stage stochastic model described in Munoz et al. (2014) without disjunctive constraints.[5]

EC.3. Analysis of Upper Bound Estimates

A formulation of the operations problem $f(x,(\Omega,p))$ for 8,736 observations (i.e., $|\Omega| = 8,736$) of demand, wind, solar, and hydro data results in a linear program with only a few thousand variables and constraints smaller than the original planning problem. Solving this problem with currently available hardware is computationally infeasible due to memory limitations. To validate the sub-sample approximation method, we instead use a large sample of intermittent data that minimizes the sum of the square difference of means, standard deviations, and correlations between the sample and the full dataset (van der Weijde and Hobbs 2012). A 5,000 hour operations problem is the largest manageable approximation of the 8,736 hour operations problem and best matches its statistical characteristics. Although decomposition approaches could be used to solve $f(x,(\Omega,p))$ exactly (e.g., Hobbs and Ji (1999) and O'Brien (2004)), their implementation is beyond the scope of this paper.

Table EC.3 Summary of results for 100 random and stratified samples for different sub-sample sizes.

Sub-Sample Size	Random Samples				Stratified Samples				5000-hr problem
	25	50	100	200	25	50	100	200	
Mean ($B)	543.9	526.3	535.7	526.3	527.0	532.6	528.2	529.7	527.7
Standard Deviation ($B)	100.2	70.9	52.2	37.5	59.3	43.5	27.8	16.6	-
Mean Solution Time (s)	23.4	48.3	103.1	208.6	23.4	48.7	104.3	211.5	8,814

We choose a sub-sample size M and a sub-sample count N taking into account the quality of the approximation and hardware restrictions. We first draw 100 independent samples of intermittent data, using both random and stratified sampling, for sub-samples considering 25, 50, 100, and 200 observations. This allows us to study the effect of the data stratification and sub-sample sizes on the quality of the estimator. All operations problems are run using the investment solution of the linearized expected value (i.e., single cluster) planning problem.

As shown in Table EC.3 and Figure EC.2, the sub-sample standard error[6] decreases dramatically with the number of sample hours in the operations problems for both random and stratified samples. The effect of data stratification using the k-means clustering method yields approximately one half of the standard deviation observed in the random samples and is, in this case, more effective

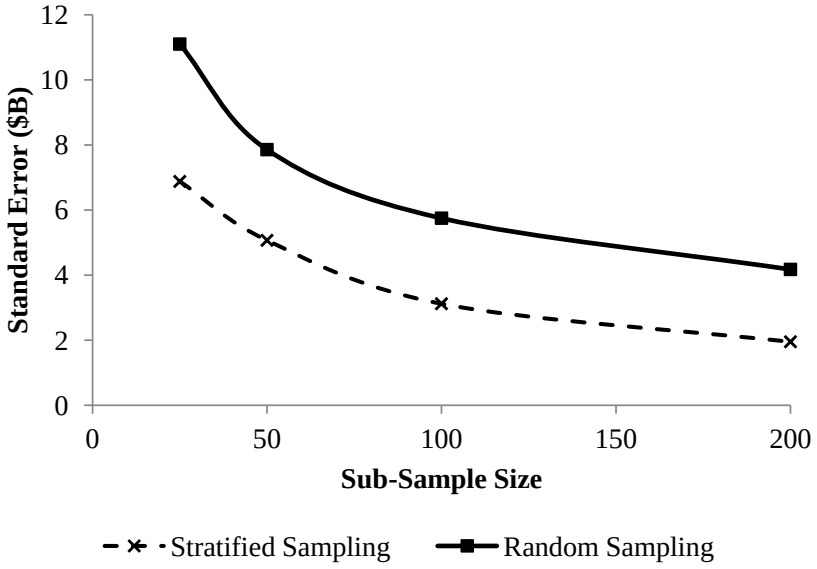

Figure EC.2 Sub-sample standard error versus sub-sample size for 100 replications of random and stratified samples ($N = 100$).

than doubling the sub-sample size.[7] Solution times also scale linearly. The same investment plan yields \$527.7B when tested against the 5,000 hour operations problem, but requires nearly two and a half hours of computation time. In our implementation, we limit the sub-sample count to 20 ($N = 20$) in an attempt to balance computer resources for the estimation of the upper and lower bounds. Because the sub-sample count is rather small in comparison to the 100-sample analysis performed in this section, we choose to use sub-sample sizes of $M = 200$ stratified observations, which yield the lowest sub-sample standard error among the four sub-sample sizes considered. Much larger sub-sample counts and sizes can be used to ensure tight confidence intervals on the optimal operating costs $f(x, (\Omega, p))$, but at the expense of higher computational complexity. Other potential extensions include re-sampling within the bounding phase (Pierre-Louis et al. 2011) or within the Benders phase (Higle and Sen 1991, Infanger 1992), but we leave these as subjects of future research in the context of our application.

Endnotes

5. Disjunctive constraints are used to enforce Kirchhoff's Voltage Laws in candidate lines. In this paper, Kirchhoff's Voltage Laws are only enforced in existing lines and relaxed for candidate lines. As discussed in Munoz et al. (2014), disjunctive constraints can cause numerical difficulties in mixed-integer solvers. However, the bounding and decomposition procedures proposed here are equally applicable to the formulation of the problem with disjunctive constraints.

6. The standard error (SE) is defined as $SE = \frac{s}{\sqrt{N}}$, where s corresponds to the sample standard deviation and N is the sub-sample count.

7. For instance, the standard error of the 50 hour random sub-samples is approximately \$8B. Doubling the sample size to 100 hours reduces the standard error to approximately \$6B. However, stratifying the sample space in the 50-hour sub-samples yields a standard error of approximately \$5B.

Appendix A: Phase 1, Supporting Figures

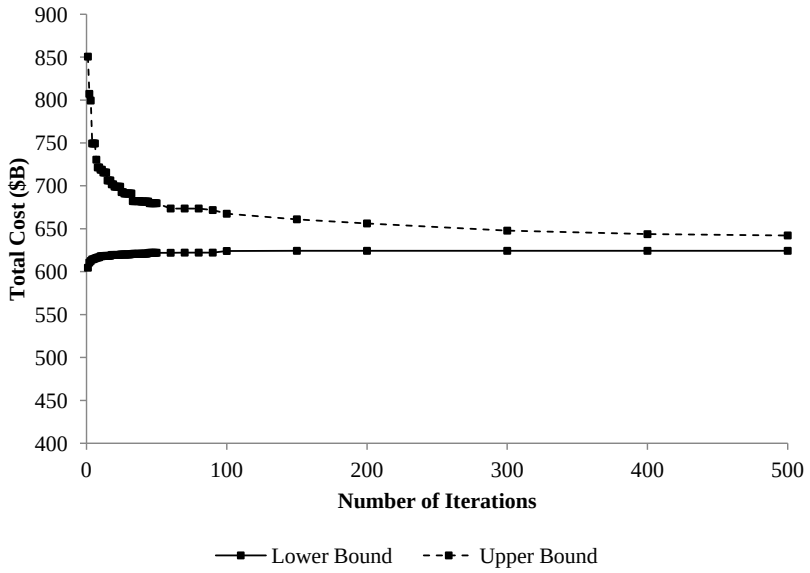

Figure EC.3 Upper and lower bounds for the linear problem versus the number of clusters.

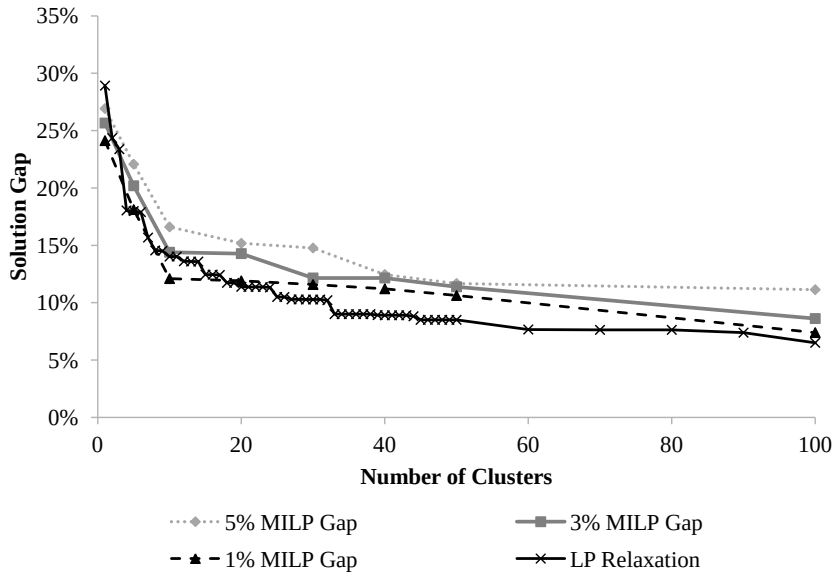

Figure EC.4 Optimality gap versus the number of clusters for different MILP gaps and the linear relaxation.

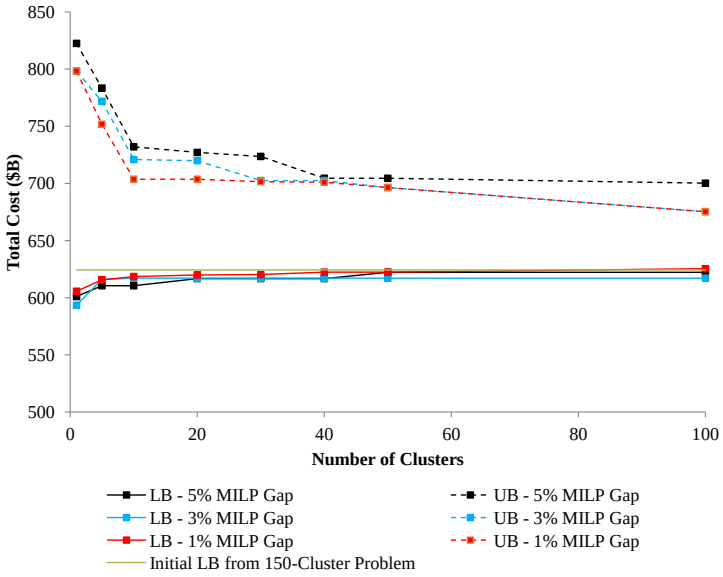

Figure EC.5 Upper and lower bounds for the mixed-integer linear formulations versus the number of clusters.

Appendix B: Phase 2, Supporting Figures

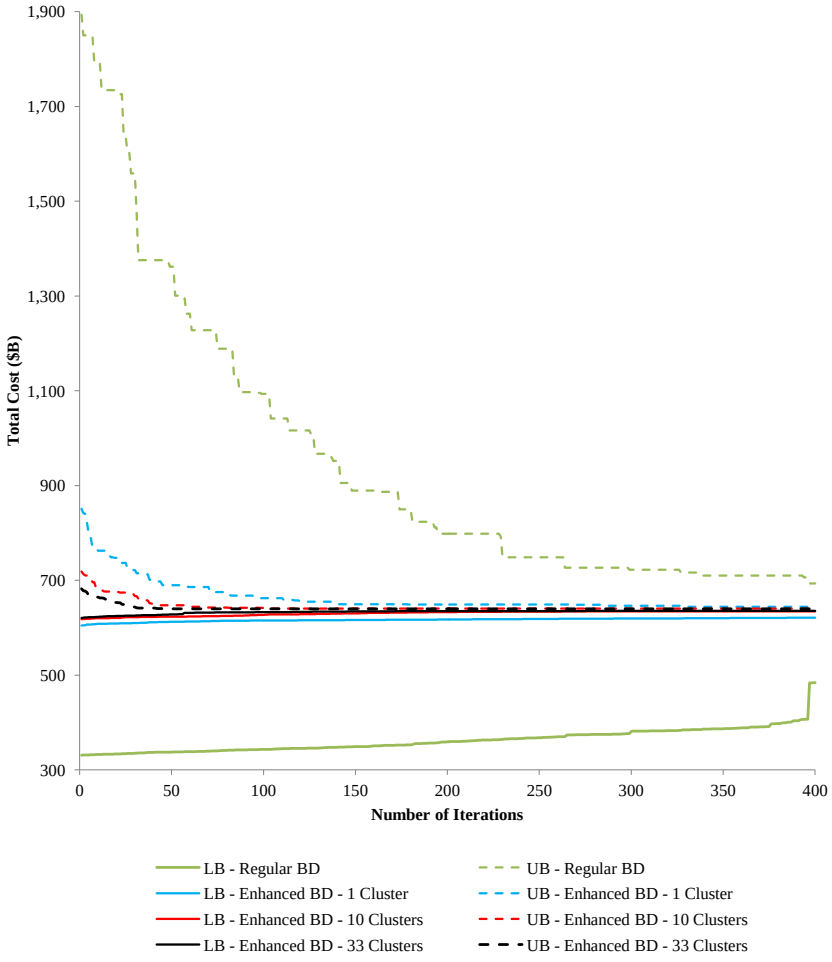

Figure EC.6 Upper and lower bounds for the linear problem versus the number of iterations.

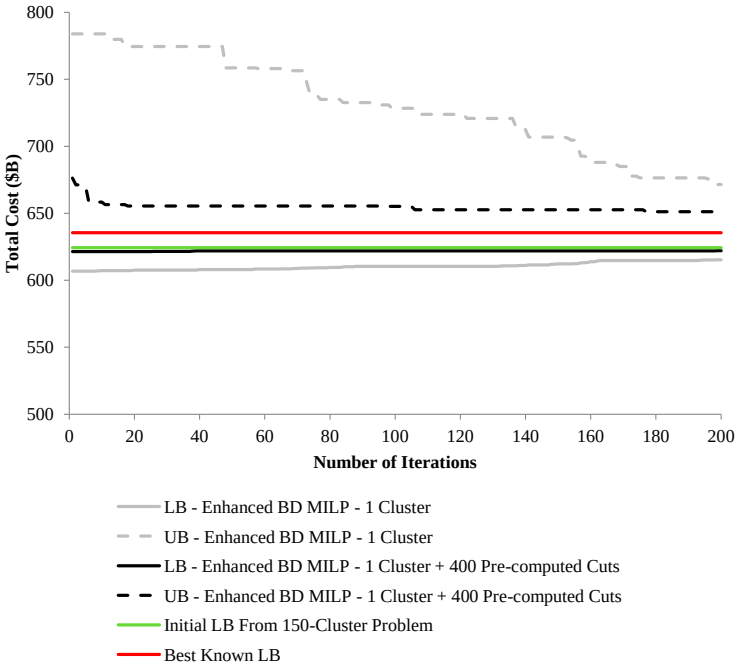

Figure EC.7 Upper and lower bounds for the mixed-integer linear problem versus the number of iterations.