

1-2013

Scenario Construction and Reduction Applied to Stochastic Power Generation Expansion Planning

Yonghan Feng
Iowa State University

Sarah M. Ryan
Iowa State University, smryan@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/imse_pubs

 Part of the [Industrial Engineering Commons](#), and the [Systems Engineering Commons](#)

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/imse_pubs/15. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Industrial and Manufacturing Systems Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Industrial and Manufacturing Systems Engineering Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Scenario Construction and Reduction Applied to Stochastic Power Generation Expansion Planning

Abstract

A challenging aspect of applying stochastic programming in a dynamic setting is to construct a set of discrete scenarios that well represents multivariate stochastic processes for uncertain parameters. Often this is done by generating a scenario tree using a statistical procedure and then reducing its size while maintaining its statistical properties. In this paper, we test a new scenario reduction heuristic in the context of long-term power generation expansion planning. We generate two different sets of scenarios for future electricity demands and fuel prices by statistical extrapolation of long-term historical trends. The cardinality of the first set is controlled by employing increasing length time periods in a tree structure while that of the second set is limited by its lattice structure with periods of equal length. Nevertheless, some method of scenario thinning is necessary to achieve manageable solution times. To mitigate the computational complexity of the widely-used forward selection heuristic for scenario reduction, we customize a new heuristic scenario reduction method named forward selection in wait-and-see clusters (FSWC) for this application. In this method, we first cluster the scenarios based on their wait-and-see solutions and then apply fast forward selection within clusters. Numerical results for a twenty year generation expansion planning case study indicate substantial computational savings to achieve similar solutions as those obtained by forward selection alone.

Keywords

power generation expansion planning, scenario generation, scenario reduction, stochastic programming

Disciplines

Industrial Engineering | Systems Engineering

Comments

NOTICE: This is the author's version of a work that was accepted for publication in Computers & Operations Research. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Computers & Operations Research, 40 (2013): 9, doi: [10.1016/j.cor.2012.05.005](https://doi.org/10.1016/j.cor.2012.05.005). Posted with permission.

Abstract:

A challenging aspect of applying stochastic programming in a dynamic setting is to construct a set of discrete scenarios that well represents multivariate stochastic processes for uncertain parameters. Often this is done by generating a scenario tree using a statistical procedure and then reducing its size while maintaining its statistical properties. In this paper, we test a new scenario reduction heuristic in the context of long-term power generation expansion planning. We generate two different sets of scenarios for future electricity demands and fuel prices by statistical extrapolation of long-term historical trends. The cardinality of the first set is controlled by employing increasing length time periods in a tree structure while that of the second set is limited by its lattice structure with periods of equal length. Nevertheless, some method of scenario thinning is necessary to achieve manageable solution times. To mitigate the computational complexity of the widely-used forward selection heuristic for scenario reduction, we customize a new heuristic scenario reduction method named forward selection in wait-and-see clusters (FSWC) for this application. In this method, we first cluster the scenarios based on their wait-and-see solutions and then apply fast forward selection within clusters. Numerical results for a twenty year generation expansion planning case study indicate substantial computational savings to achieve similar solutions as those obtained by forward selection alone.

Keywords: Stochastic programming, scenario generation, scenario reduction, power generation expansion planning

1. Introduction

A stochastic program is a mathematical program that involves some uncertain data. These parameters may be most accurately described as continuous random variables but, except in some trivial situations, it is difficult to optimize directly in terms of those distributions. Hence, in most applications, the continuous distributions are approximated by discrete distributions with a finite number of scenarios, or possible realizations, for the random variables. This discretization procedure is often called *scenario generation*. The stochastic programming model then can be specified by combining a deterministic formulation of the mathematical program, which could be solved for any scenario, with a set of scenarios for the uncertain data. Thus, a crucial step in the application of stochastic programming is to obtain a set of scenarios that realistically represents the distributions of the random parameters but is not too large.

Uncertainty is inescapable in long-term capacity planning applications. The random parameters can be conceived to follow a multidimensional stochastic process over the planning horizon, so that discrete scenarios represent sample paths. Even a moderate dimension for the process, when combined with a reasonable number of time steps in the planning horizon, can result in a very large number of such paths. In a two-stage stochastic programming formulation, the first-stage variables represent “here-and-now” decisions, such as facility investments, that must be taken before the realization of the uncertain data is known. The second stage, “wait-and-see,” decisions represent system operation, within the constraints dictated by the investment decisions, as the random parameters evolve over time. In the extensive form of a two-stage stochastic program, a separate set of second stage decision variables is used for each possible scenario path and the problem can be solved as a (very high dimensional) deterministic mathematical program. Assuming scenario independence, the constraint coefficients form a large block-diagonal matrix in which each block describes the same constraint structure with scenario-specific parameters. Because of this special structure, several decomposition methods have been used to solve the two-stage model [10]. However, a large number of scenarios may limit the tractability of solution, even when using decomposition techniques. To attain reasonable computation times, it is natural to try to approximate the original scenario set with a much smaller subset that retains essential features, a process often called *scenario reduction*.

* Corresponding author. Tel.: 515-294-4347. Email: smryan@iastate.edu

Power generation expansion planning (GEP) is a complex multi-year problem that aims to determine the timing and technology choices for generation investments over a long planning horizon under economic criteria while ensuring that expected electricity demand growth is met [18]. The traditional GEP involves solving centralized planning problems to identify plans that minimize costs for the whole system even under adverse conditions. The typical objective has been to minimize the expected sum of discounted costs, including investment cost for generating units, operation and maintenance costs, and penalties for unserved energy [19]-[22]; more recent models have also considered risk [30]. Recently, the electricity industry has undergone significant restructuring towards deregulation and competition, and attention has shifted to maximizing profits, which are composed of the revenues based on market prices and the costs, including both investment and operation, by individual generating companies in the market. However, restructuring has not yet taken hold in all regions of the US or in many parts of the world. The centralized GEP problem remains relevant, though it must increasingly account for growing uncertainties in fuel prices, availability of intermittent forms of renewable generation, possible environmental regulations, and demand stemming from increasing electrification (for example, of transportation) and implementation of smart grid technologies that increase consumer choice.

This paper investigates the application of a recently-developed scenario reduction heuristic, termed forward selection in wait-and-see clusters (FSWC) [40] to a two-stage stochastic programming formulation of GEP with uncertainty in demand growth and fuel prices. We generate scenarios for a case study by fitting a stochastic process model to historical data for electric load and natural gas price in the Midwest region of the US and construct scenario sets in the form of a tree or a lattice by a moment-matching method [5]. The FSWC heuristic first solves a deterministic “wait-and-see” subproblem for each scenario, in which both first- and second-stage decisions are chosen to minimize cost in that particular scenario. Then scenarios are clustered based on similarity of their “key” first-stage decisions. Identifying these key decisions is the application-specific aspect of FSWC. Finally, it applies the well-known forward selection heuristic [11], [12] to choose one representative scenario from within each cluster. In our GEP case study with a 20 year planning horizon, the overall computational time required by applying FSWC can be up to two orders of magnitude smaller than that required by applying classical forward selection, depending on the cardinality of the reduced scenario set, while similar first-stage solutions are obtained.

The contributions of this paper include statistically verifying that historical annual electricity demands and natural gas prices in the US have been consistent with correlated geometric Brownian motion (GBM) processes, and then generating scenario trees to extrapolate those processes by solving nonlinear programs derived from the moment matching method [5] under two methods of subdividing the planning horizon. In addition, the customization of FSWC for this application is described and tested.

The paper is organized as follows. In Section 2 we place our research in context of the literature. Section 3 presents a simplified long term centralized GEP model. Section 4 reviews the general moment matching method for scenario tree generation. Section 5 describes a novel heuristic scenario reduction method based on wait-and-see clustering that applies the forward selection heuristic within clusters. For our case study, Section 6 first verifies the historical fit of correlated GBM processes for annual electricity demands and average natural gas prices in the US. Then, a general scenario tree for unequal length sub-periods and a tree with lattice structure for equal-length sub-periods are generated for a twenty-year time horizon. Next, the FSWC heuristic is implemented for the two-stage stochastic programming model of GEP after defining the key first-stage variables. The related numerical results and comparisons between forward selection and FSWC are also provided. Finally, Section 7 provides conclusions.

2. Relation to the literature

Many researchers have discussed approaches for generating appropriate scenarios [1]. One of the most intuitive approaches is to sample directly from the specified marginal distributions and correlation matrix [2]. Another scenario generation method is based on optimal discretization. This method aims to approximate the stochastic process with finite scenarios by minimizing the pertinent error in the stochastic programming objective function of a discrete optimization model [8]. One of its advantages is to generate a whole multiple-period scenario tree at once. In addition, scenarios can be generated by path-based methods. Instead of generating tree-structured scenarios, these methods evolve the stochastic process to generate complete paths [7] in a “fan” structure. A procedure called clustering or bucketing must be employed [1] to transform a fan into a scenario tree if the tree structure is preferred in an application. Alternatively, scenarios can be generated by a moment matching method even without knowing the marginal distribution functions exactly [3]-[6]. This method summarizes the marginal distribution functions by their moments (mean, variance, skewness, etc.) or other statistical properties. In particular, Høyland and Wallace [5]

proposed a moment matching method, which we use in this paper, to generate a finite number of scenarios with a tree structure. The underlying idea is to construct scenarios and their corresponding probabilities by approximating the specified statistical properties according to the available degrees of freedom. This technique has also been applied to generate scenarios for spot prices and prediction errors based on the first four central moments [35].

Several scenario reduction methods have been developed in the past decade. The idea in [11], [12] derives from quantitative stability results in terms of a particular probability metric, resulting in two scenario reduction heuristics, termed forward selection and backward reduction. Reference [12] extends the original idea addressed in [11], and improves the iterative procedure, achieving a new version of forward selection and simultaneous backward reduction with considerable performance improvement. These scenario reduction techniques have been used in many power systems applications of stochastic programming, such as for power management [13], [14]. However, they are based on upper bounds of the particular probability metrics in form of certain mass transportation problems rather than on the particular metrics themselves. Reference [15] extends the earlier work by relying directly on the probability metrics. Based on stability behavior of multi-stage stochastic programming models, reference [16] argues that scenario tree reduction in a multi-stage model should not only rely on L_r -norms, and [17] derives a new scenario reduction method motivated by the idea mentioned in [16].

The probability-based scenario reduction method focuses on only the scenario parameters, but fails to account for where the uncertainties appear in the mathematical formulation and their impacts on the solution. To incorporate the influences of scenarios on the first-stage decisions in the reduction process, a new heuristic scenario reduction method FSWC [40] is applied in this paper. An implementation challenge is to track the influences of the stochastic parameters on every first-stage variable in a large-scale problem. Thus, it is reasonable to concentrate on a subset of first-stage decisions that dominate the others or can best reflect the essences of different scenarios. Accordingly, identification of key first-stage decisions is required. The details of identifying these key first-stage decisions in our GEP application are discussed in Section 6.

A number of papers focus on GEP, with scopes that range from problem formulation to solution algorithms. Due to the progress towards deregulation and competition, the conventional GEP models in some studies have been replaced by decentralized ones. Several papers have compared the optimal investment in the centralized environment to that in the decentralized environment, and developed new models and algorithms by including game theory and uncertain factors [25-28]. However, even in a decentralized environment, a centralized planning process can provide targets. Once these targets are identified, incentives and regulations can be designed to encourage decentralized decision makers to approach one of them [28]. Because the restructuring procedure has not taken hold in all regions of the US and large utilities still develop their own expansion plans, we maintain focus on this paper on a cost-minimization formulation that involves several uncertain factors. Among the papers addressing the centralized GEP issues, [19] is one of the earliest papers to discuss minimizing the investment and operational cost under uncertainties. However, it only covers the formulation of a stochastic program for minimizing the expected cost, and ignores the computational issues related to scenarios. References [21], [22] serve the same purpose, but account for more factors -- available capacities of each generation technology, loss of load probability and reserve margin. Nevertheless, they not only ignore the uncertainties of power demands and fuel prices, but also fail to guarantee that their solutions resulting from metaheuristic methods are optimal. To achieve a more reliable expansion strategy minimizing expected total cost for long-term centralized GEP under uncertainties, this paper shares a similar two-stage stochastic GEP model with [30]. The model minimizes the sum of expected investment, operation cost and penalty for unmet energy while considering the balance of system demand over the whole planning horizon and some limitations on the number of generators added of each type. Rather than randomly sampling scenarios as in [30], we explore the performance of the FSWC heuristic in this paper.

Because a GEP model often includes a number of discrete or nonlinear components, implicit enumeration may be required to obtain the optimal investment strategy. Dynamic programming (DP) has been widely used for solving such GEP models [31], [32], but the so-called "curse of dimensionality" often hinders its direct application in GEP especially when uncertainties are involved. As does stochastic programming (SP), DP requires finite sets of discrete states, which correspond to nodes in the scenario tree, and can capture the dynamic aspects of a problem. DP allows general nonlinear costs [32], but SP is more flexible with respect to state variables [10]. Large scale SP formulations of GEP are also amenable to decomposition methods that can exploit the special structure of subproblems [33], [34].

3. Generation expansion planning model

The traditional GEP model aims to determine the ideal technologies, expansion sizes, and timing of new generator construction in an economic manner, while ensuring sufficient capacity to satisfy the projected demand growth [18]. As explained above, we investigate generation expansion from the centralized planning point of view in this paper. Our primary focus is to investigate the performance of our new scenario reduction method in this context.

Developing a long term investment plan requires a model that includes extensive uncertainties. Among these uncertainties, electricity demand growth is often modeled as a stochastic process. Several additional sources of uncertainty such as the fuel cost, construction time, equipment outage rates, and government policies may also affect future planning activities. We neglect the equipment outages in our GEP model, and assume that government policies are stable. Thus, fuel cost is the other major uncertainty considered in this paper. According to the information provided by the Energy Information Administration (EIA) of the US Department of Energy, coal and natural gas are the major fuels for electricity generation in the US, with coal accounting for 45% and gas used to generate 23% of electricity generated in 2009 [36]. The price of coal has been relatively stable for the past 20 years but natural gas price has fluctuated considerably more [37], [38]. In this paper, we consider annual electricity demand and average natural gas price as the only uncertain quantities. Following the above assumptions, a simplified GEP model will be formulated after presenting the notations used in the whole paper.

Indices:

i	Index for scenarios
t	Index for years in the planning horizon
g	Index for generator types
ℓ	Index for branches from a node

Decision variables:

U_{gt}	The number of generators of type g to be built in year t , integer
E_{gti}	Energy provided by generator type g in year t under scenario i , MWh
UE_{ti}	Unserved energy in year t under scenario i , MWh

Parameters:

b_g	Total cost to build a generator of type g , discounted to beginning of construction period, \$/MW
fm_g	Fixed O&M cost of generator type g , \$/MW
I_g	The total generation capacity of generator of type g at the beginning of the planning horizon
u_g^{\max}	The maximum number of generators of type g to be built over the planning horizon
m_g^{\max}	Installed capacity of a new type g generator, MW
h_t	Total hours in year t
r	Annual interest rate for cost discounting
N	The total number of generated scenarios
n_g	Capacity factor of type g generator
P_i	The probability that scenario i occurs
Pc	Penalty for unserved energy, \$/MWh

Scenario dependent parameters:

c_{gti}	Generation cost of generator type g in year t under scenario i , \$/MWh
-----------	---

d_{it} Annual electricity demand in year t under scenario i , MWh

A simplified two-stage GEP model is formulated as (1)-(5):

$$\min_{U_{gt}, E_{gti}, UE_{it}} \sum_i p_i \xi_i \quad (1)$$

$$\xi_i = \left(\sum_t \sum_g \left(b_g + f m_g \right) m_g U_{gt} + \sum_t \sum_g \left(c_{gti} E_{gti} + Pc UE_{it} \right) \right) / (1+r)^{t-1}, \forall i \quad (2)$$

$$\text{s.t.} \quad \sum_g E_{gti} + UE_{it} = d_{it}, \forall t, i \quad (3)$$

$$E_{gti} \leq h_t \left(n_g m_g \sum_{k \leq t} U_{gk} + I_g \right), \forall g, t, i \quad (4)$$

$$\sum_t U_{gt} \leq u_g^{\max}, \forall g \quad (5)$$

Formula (1) is the objective function which indicates the purpose of identifying an expansion and generation plan that achieves the minimum expected cost over all possible scenarios. The discounted cost in formula (2) includes investment cost, generation (including maintenance) cost and penalty cost from unserved energy over the whole planning horizon. The investment decisions, U_{gt} are the first-stage "here-and-now" decision variables, while E_{gti} and UE_{it} are the second-stage "wait-and-see" decisions that depend on the scenario realization. Equation (3) requires that the annual electricity demand is the sum of provided energy and the unserved energy in each scenario. Constraint (4) represents capacity constraints of existing and new generators. Due to financial capacity, environmental impacts, and other drivers, we bound the total number of each type of generator added, as formulated in (5).

4. Scenario generation

It is generally intractable to solve a mathematical program with uncertain parameters described by continuous distributions. The common approach is to form an approximation of the original continuous stochastic process or underlying distribution by discretization. The procedure of discretizing the continuous distribution or stochastic process is termed scenario generation. We choose a tree or lattice rather than a fan structure in this paper because the size of a tree or lattice can be controlled by its structure.

In general, a scenario tree is a set of nodes and branches used in models of decision making under uncertainty. Every node in the tree represents a possible state of the world at a particular time point and a position where a decision can be made. Each tree node has only a single predecessor, but can have multiple successors, whereas a lattice node can have multiple predecessors. An arc emanating from a node indicates a possible realization of the uncertain variables from that state. The first node in the scenario tree is defined as the root node, representing the initial situation, being in this paper the values of annual electricity demand and natural gas price in the current year. Any possible forward path from the root node to a node at the last time point is defined as a scenario which describes a plausible realization of annual electricity demands and natural gas prices over the time horizon. The number of nodes in a tree at the final time point equals the total number of scenario paths. The nodes shared in a lattice illustrate that a plausible state may be achieved from the initial condition by multiple realization paths. Although the number of paths in the lattice is similar to the number of branches in the tree, the nodes at the final time point will be much fewer than the scenarios.

4.1 Optimization model for scenario tree generation

The moment matching method of generating scenarios [5] was adopted to generate the immediate successors from each node. The procedure to generate the whole scenario tree can be recursively applied from the root node to the nodes at the end of the final period. Since the proper statistical properties can partially describe the underlying distribution of data, we minimized a measure of distance between statistical properties of the fitted discrete distribution and the *statistical specifications*, subject to a constraint restricting the conditional branching

probabilities to sum to one. Examples of statistical specifications are mean, variance, and skewness. We assume the statistical specifications of random variables have been obtained. The details of their computation for the case study are shown in Section 6. Let S be the set of all specified statistical specifications, and S_{VAL_j} be the value of the j th statistical specification. The random variables corresponding to the same period are grouped into a vector, called the scenario node vector, denoted by x_ℓ , where ℓ indicates the ℓ th branch. The corresponding conditional probability of the ℓ th branch is denoted by π_ℓ . The symbol x_0 represents annual electricity demand and natural gas price at the root node, and π_0 is 1. The j th statistical property of the fitted approximating distribution is denoted as $f_j(x, \pi)$. For example, if $f_j(x, \pi)$ represents mean, then it is computed as $\sum_\ell x_\ell \pi_\ell$. Thus, the general scenario generation model based on optimization may be formulated as follows [5], where the common square norm is adopted to measure the distance.

$$\min_{x_\ell, \pi_\ell} \sum_j w_j \left(f_j(x, \pi) - S_{VAL_j} \right)^2 \quad (6)$$

$$\text{s.t.} \quad \sum_\ell \pi_\ell = 1 \quad (7)$$

$$\pi_\ell \geq 0 \quad (8)$$

A set of weights w_j represent the importance of satisfying different statistical specifications, which can be manually specified by individual preference. In general, model (6)-(8) must be solved at every node in the scenario tree except those at the last time point. However, if the stochastic processes describing the uncertain data are stationary, the solution obtained at the root node can be applied at every node. The nonconvex nonlinear program (6)-(8) was solved by the Matlab *fmincon* function. Because we often found solutions with nonzero objective values (indicating that the statistical properties of the fitted distribution were not close to the statistical specifications), we tried multiple starting points and selected the solution with the smallest objective value, stopping when we found one with objective sufficiently close to zero.

4.2 Controlling the size of scenario tree

Høyland and Wallace discussed potential pitfalls, such as *underspecification* and *overspecification*, when using the moment matching method to generate scenarios, and proposed an idea based on counting the degrees of freedom to control the size of scenario tree while avoiding those pitfalls [5]. The idea is expressed in formula (9), where the symbol " \sim " denotes "close to":

$$(D+1)y-1 \sim \text{the number of specifications}, \quad (9)$$

where D indicates the dimension of each scenario node vector, so that $D+1$ represents the number of random variables plus the branch probability; and y represents the number of branches from each node. Since constraint (7) eliminates one degree of freedom, $(D+1)y$ is reduced by one to get the final degrees of freedom. A simple example of calculating the number of branches is illustrated. Suppose a period includes one year, and the two random variables, the annual electricity demand and natural gas price, are specified by their means, variances, skewnesses and correlation. Then the number of specifications is 7. According to (9), we have $(2+1)y-1 \sim 7$, so y should be 3.

If the branches cover a two-year period, the number of specifications is 14, since four random variables are involved — two for annual electricity demands and two for average annual natural gas prices. The number of branches from a node should still be 3, as $(4+1)3-1=14$. Because the long time horizon with annual periods results in a large scenario tree, we use longer periods in the later years of the case study to decrease the number of generated scenarios, as described in detail later.

5. Scenario reduction

The computational effort for solving scenario-based optimization depends mainly on the number of scenarios even if decomposition techniques are used. The huge number of scenarios that frequently result from the scenario generation process limits tractability. Thus, it is essential to approximate the original scenarios with a much smaller subset that can well approximate the original scenario set. Existing scenario reduction methods control the approximation's goodness-of-fit according to probability metrics. The common scenario reduction methods, forward selection (FS), backward reduction and their variants, obtain the selected scenarios by recursive pairwise comparisons of the distances between scenarios. We review the *fast forward selection* (FFS) method as follows [12].

A scenario here is defined as a path from the root node to a node in the last stage, denoted as ω_i , $i = 1, \dots, N$, and its corresponding probability by p_i , which is the product of conditional probabilities over the path. In the following, $J^{[s]}$ represents the scenario subset from which the next scenario will be selected after the s th selection and $\eta(\cdot)$ is a nonnegative, continuous and symmetric function, usually defined as some norm on R^n . We use the L_2 -norm. The intuitive idea of the FFS method is to select a subset Ω' from the original finite scenario set Ω so that Ω' is the subset of the prescribed size that has the shortest distance to the remaining scenarios. The FFS method, which focuses on the selection of n scenarios beginning from an empty set, includes the following steps:

Fast Forward Selection (FFS):

- Step 1. Let $s = 1$, calculate the distances of all scenario pairs $\delta_{i,k}^{[1]} = \eta(\omega_i, \omega_k)$, $i, k = 1, \dots, N$, and compute the weighted distance of each scenario to the other scenarios $z_k^{[1]} = \sum_{i \neq k} p_i \delta_{i,k}^{[1]}$, $k = 1, \dots, N$. Select $k = \arg \min_{k \in \{1, \dots, N\}} z_k^{[1]}$, and set $J^{[1]} = \{1, \dots, N\} \setminus \{k\}$;
- Step 2. Let $s = s + 1$, and replace the scenario pair distance with the smaller of the original pair distance and the distance to the selected scenario in $(s - 1)$ th selection, $\delta_{i,k}^{[s]} = \min\{\delta_{i,k}^{[s-1]}, \delta_{i,k_{s-1}}^{[s-1]}\}$, $i, k \in J^{[s-1]}$. Compute $z_k^{[s]} = \sum_{i \in J^{[s-1]} \setminus \{k\}} p_i \delta_{i,k}^{[s]}$, $k \in J^{[s-1]}$ for every candidate scenario in the s th selection, and choose $k = \arg \min_{k \in J^{[s-1]}} z_k^{[s]}$, $J^{[s]} = J^{[s-1]} \setminus \{k\}$, the one that minimizes the weighted distance to the remaining scenarios;
- Step 3. If the number of the selected scenarios is less than n , return to Step 2;
- Step 4. Add to the probability of each selected scenario the sum of the probabilities of all unselected scenarios that are close to it; i.e., $q_j = p_j + \sum_{i \in L(j)} p_i$, for any $j \in \Omega'$, where Ω' is the set of selected scenarios, and $L(j) := \{i \in \Omega \setminus \Omega', j = j(i)\}$, $j(i) = \arg \min_{j \in \Omega'} \eta(\omega_i, \omega_j)$ for any $i \in \Omega \setminus \Omega'$.

Note that FFS accounts only for the paths ω_i and their corresponding probabilities p_i , but not for the decision variables, cost function or where the uncertainties appear in the mathematical programming formulation. Thus, FFS is a problem independent scenario reduction method, which will lead to the same subset of scenarios no matter whether the uncertain parameters exist in cost coefficients, constraint right-hand sides or the constraint coefficient matrix. Moreover, dramatic differences among the magnitudes of the uncertain parameters may distort the distance metric. Therefore, we use an alternative new heuristic scenario reduction method that considers the impacts of scenarios on first-stage decision variables instead of directly selecting scenarios based only on distances among them.

The heuristic scenario reduction method we use here is called forward selection in wait-and-see clusters (FSWC) [40]. As mentioned before, FSWC aims to incorporate the impact of each scenario on first-stage decision variables and, thus, may be especially effective when problems are solved repeatedly over with a rolling horizon. To measure the impacts of scenarios on the first-stage decision variables, a deterministic wait-and-see problem is defined for an individual scenario, k , by setting scenario probability $p_k = 1$, and $p_i = 0$, $i \neq k$, in (1)-(5), and deleting constraints (2)-(4) for $i \neq k$. The solution includes scenario-specific optimal values of the first-stage decisions.

This deterministic wait-and-see problem is solved for each scenario and the corresponding first-stage decisions are recorded. Because a large scale stochastic program often includes a large number of first-stage variables, it is difficult to track the impact on each of them. Therefore, it is reasonable to focus on the impact on the so-called *key first-stage variables* which were originally defined in [40] to be those variables that determine the other first-stage variables. As a simple example, if there are only two kinds of variables involved in first-stage variables, where one is the fuel purchased from the market in each period and the other is the power generated from a given power plant, then fuel amounts purchased will be key variables since the amount of fuel limits the amount of generation. The definition of key first-stage variables is extended here to include those variables that differ significantly according to scenarios. For example, if the planned numbers of power plants to exist in each year are the only first-stage variables in a stochastic program, and the electricity demand in every year is the only uncertain parameter, then the planned number of power plants in the last few years of the planning horizon can be considered as key first-stage variables, as they represent the targets of the expansion plan. The ending numbers of power plants can reflect the trend of the electricity demand more clearly than the planned numbers of plants in earlier years. The FSWC method implements forward selection within clusters of scenarios that are formed according to the similarity of those key first-stage variables. The clustering process is repeated as necessary to force the number of clusters to satisfy the desired cardinality of the reduced scenario set. The details of the FSWC procedure are presented below:

Forward Selection in Wait-and-See Clusters (FSWC):

As in the FFS method, the prescribed cardinality is set to n .

- Step 1. Solve the deterministic wait-and-see problem for each scenario, and retain the values of the key first-stage variables;
- Step 2. Group the original scenarios into the same cluster if their key first-stage decisions are the same; if the number of clusters n_m is less than or equal to n , go to step 4;
- Step 3. Cluster the n_m groups of key first-stage decisions into n clusters by k -means method based on the L_2 -norm [41], and create the corresponding n clusters of original scenarios at the same time;
- Step 4. Use the FFS method to select one scenario from each cluster of original scenarios.

Clustering the original scenarios before performing forward selection imparts two advantages. One is that less time is required to select one scenario from each group of original scenarios. The computational savings result not only from selecting from smaller clusters, but also because we select only one scenario from each cluster. In addition, the probability assigned to each selected scenario is simply the total probability of scenarios in the cluster from which it is selected. Hence, FSWC is more efficient than FFS because it selects scenarios from smaller sets and bypasses the process of recalculating probabilities.

The other advantage is that the scenarios with similar impacts on the first-stage decisions will be grouped together, avoiding the situation that an original scenario is represented by a scenario that is probabilistically close to it but has dramatically distinct impact on the first-stage decisions. Additional savings accrue if FSWC is used to obtain reduced subsets with different prescribed cardinalities n , such as to determine a good value for n . Not only does the procedure of obtaining wait-and-see decision variables need to be implemented just once, but also the clusters covering the same wait-and-see variables can be used repeatedly.

6. Case study

We implemented the GEP model (1)-(5) on a hypothetical system involving six types of generators over a 20 year horizon. The data for annual electricity demands and natural gas prices from 1973 to 2010 were collected from the Energy Information Administration of the US Department of Energy [43], [44]. The related data of each type of generators in the system were collected from reports [45]-[47], and are given in Table 1. The generator types include baseload (coal), gas combined cycle (CC), gas combustion turbine (CT), nuclear, wind (farm) and integrated gasification combined cycle (IGCC). The building costs were obtained by following the computation method illustrated in [30]. Because there is no evident intention to expand hydropower in the US Midwest in the next decades [49], [50], we did not include it as an expansion option in our case study. The maximum numbers of planned units of each type, u_g^{\max} , were set to reflect some practical considerations such as difficulty with obtaining permits for nuclear plants and hesitation to include too much wind generation because of its variable and intermittent

nature. Restrictions on the other types of generators are relaxed in [48], where additional characteristics of wind units, temporal variability in load, and risk considerations are also included.

Table 1.

The parameters for candidate generators & existing generation capacity

Generator type g	Base-load (coal)	CC	CT	Nuclear	Wind	IGCC
m_g^{\max} (MW)	600	400	230	1350	50	380
n_g (%)	84.6	84.6	80.0	90.2	30.0	81.0
c_{g0} (\$/MWh)	14.5	22.6	37.8	0.4	0	11.1
Growth rate of c_{g0} (%)	3	*	*	3	0	3
u_g^{\max}	45	65	90	8	105	45
I_g (MW)	21375	11000	11000	8800	2450	11250
f_{mg} (\$/MW)	28150	11960	10770	92040	30980	47150
b_g (\$/MW)	964922	371102	312963	1336759	895283	1078495

*Generation costs of CC and CT are scenario dependent parameters

As mentioned in Section 3, only annual electricity demand and average natural gas price were considered as random variables. To generate plausible scenarios, it is necessary to assume appropriate underlying distributions or stochastic processes for these two random variables. Similar to [30], but using updated data sets, correlated GBM processes were selected based on historical goodness of fit. The GBM process has often been invoked as a model for disparate qualities in financial markets, such as stock prices, natural resource prices, and the increasing rate of demand for products [39]. The following subsections describe verification of the assumption that GBM fits both the annual electricity demands and natural gas prices, scenario generation for these two random variables, and a comparison between the FFS and FSWC scenario reduction methods.

6.1 Geometric Brownian motion assumption and verification

According to [9], if $Y(t)$ is a Brownian motion with drift coefficient μ and variance parameter σ^2 , then the stochastic process $X(t) = e^{Y(t)}$ follows a GBM. Because the GBM has a property that its log ratio, $\log(X(t+1)/X(t))$, follows a normal distribution, it is necessary to verify normality of the annual growth log ratios of each variable's time series. Second, we can verify the independent increments of the underlying Brownian motion process by examining the autocorrelations of the log ratios. The JMP statistical software package was employed for these tests. The relevant parameters are presented in Table 2.

Table 2.

Statistical properties of log-ratios of annual electricity demand and natural gas price

	Mean	Std Dev	Shapiro-Wilk	p-value
Electricity demand	0.0214	0.0227	0.9571	0.1753
Natural gas price	0.0782	0.2503	0.9473	0.0782

The last two columns in Table 2 present the Shapiro-Wilk test statistics computed by JMP. The p-values for both series indicate failure to reject the null hypothesis that each data set is sampled from a normal distribution. Autocorrelation functions were computed to check the independence among successive log-ratios of each variable as well. As Fig.1 illustrates, the insignificant spikes at all lags for the sample auto-correlation function (ACF) indicate that there is no significant autocorrelation among log-ratios of either time series. Hence, it is reasonable to assume both electricity demands and natural gas prices satisfy GBM process. In addition, we used JMP to compute a correlation between electricity demand and natural gas price over the past years as 0.7774. According to these models and parameters, scenarios of electricity demands and natural gas prices could be generated based on the

method described in Section 4. The following subsection illustrates the steps of specifying statistical properties and then generating scenarios accordingly.

< Fig. 1(a) here > < Fig.1(b) here >
(a) (b)

Fig.1 Sample ACF of log growth ratios: (a) annual electricity demand (b) natural gas price

6.2 Important properties of GBM for scenario generation

The first step in generating scenarios by the moment matching method is to specify statistics, such as moments and correlations. According to formulas (6)-(8), for each scenario tree node, the values of important statistical specifications of its successors must be pre-calculated given the information about that node, including the lengths of the time periods between successive nodes. We included mean, variance, and skewness of each random variable as well as the correlation between them in the set of statistical specifications. According to the GBM process properties, the conditional central moments can be obtained explicitly. The moment generating function of a Brownian motion with drift is [42]:

$$M_{Y(t)}(s) = E(e^{sY(t)}) = e^{\mu s + \sigma^2 s^2 t / 2}, \quad -\infty < s < \infty. \quad (10)$$

It follows that the first two moments are given by:

$$E(X(t)) = X_0 e^{(\mu + \sigma^2 / 2)t}$$

$$E(X^2(t)) = X_0^2 e^{2(\mu + 2\sigma^2)t}$$

Thus, the variance and skewness of a GBM process are:

$$Var(X(t)) = X_0^2 e^{2\mu + \sigma^2 t} (e^{\sigma^2 t} - 1) \quad (11)$$

$$skew(X(t)) = (e^{\sigma^2 t} + 2)\sqrt{e^{\sigma^2 t} - 1} \quad (12)$$

Given the history of a GBM process up to time s , for $s < t$, the conditional expectation of $X(t)$ can be found as:

$$\begin{aligned} E[X(t) | X(s), 0 \leq u \leq s] &= E[e^{Y(t)} | Y(u), 0 \leq u \leq s] \\ &= E[e^{Y(s) + Y(t) - Y(s)} | Y(u), 0 \leq u \leq s] \\ &= e^{Y(s)} E[e^{Y(t) - Y(s)} | Y(u), 0 \leq u \leq s] \\ &= X(s) E[e^{Y(t) - Y(s)}]. \end{aligned} \quad (13)$$

From the normal distribution of $Y(t) - Y(s) \sim N(\mu(t-s), \sigma^2(t-s))$, equation (13) can be rewritten as

$$X(s) E[e^{Y(t) - Y(s)}] = X(s) e^{(\mu + \sigma^2 / 2)(t-s)}, \quad (14)$$

which expresses the stationary increments of the underlying Brownian motion process. Similarly, the conditional variance and conditional skewness of $X(t)$ can also be derived, given a previous value $X(s)$:

$$Var[X(t) | X(s), 0 \leq u \leq s] = Var[e^{Y(t)} | Y(u), 0 \leq u \leq s]$$

$$\begin{aligned}
&= \text{Var}[e^{Y(s)+Y(t)-Y(s)} | Y(u), 0 \leq u \leq s] \\
&= e^{2Y(s)} \text{Var}[e^{Y(t)-Y(s)} | Y(u), 0 \leq u \leq s] \\
&= X(s)^2 \text{Var}[e^{Y(t)-Y(s)}]
\end{aligned}$$

Thus,

$$\text{Var}[X(t) | X(u), 0 \leq u \leq s] = X(s)^2 e^{2\mu(t-s) + \sigma^2(t-s)} (e^{\sigma^2(t-s)} - 1) \quad (15)$$

and for conditional skewness, we have

$$\begin{aligned}
\text{skew}[X(t) | X(u), 0 \leq u \leq s] &= \text{skew}[e^{Y(t)} | Y(u), 0 \leq u \leq s] \\
&= \text{skew}[e^{Y(s)+Y(t)-Y(s)} | Y(u), 0 \leq u \leq s] \\
&= E(e^{Y(s)} (e^{Y(t)-Y(s)} - E(e^{Y(t)-Y(s)})))^3 / (\text{Var}(e^{Y(s)} (e^{Y(t)-Y(s)})))^{3/2} \\
&= (e^{\sigma^2(t-s)} + 2) \sqrt{e^{\sigma^2(t-s)} - 1} \quad (16)
\end{aligned}$$

According to (10)-(16), the conditional central moments can be computed for any length interval between the two time points.

The stationarity of the GBM process simplifies the scenario generation procedure. Rather than solving (6)-(8) at each node, the nonlinear optimization problem can be solved once for a given interval length and the results applied to each node whose branches span an interval of that length.

6.3 Division of the planning horizon and scenario tree generation

Even though the number of branches from each scenario node is controlled, the number of subperiods in the planning horizon also influences the size of scenario tree. In our 20-year case study, if every year in the planning horizon is viewed as an individual period, and just two realizations branch from each node, there would be 2^{20} leaves of the scenario tree. Thus, we wish to divide the planning horizon into fewer periods. In this paper, 20 years were divided into 10 periods with different structure by two methods. As the decision maker may mainly focus on the decisions in the immediate future, the first method is to divide the horizon into increasing length periods – short periods early in the planning horizon and long periods further in the future. Fig.2 shows one way of dividing the planning horizon. The second method is to divide the planning horizon into equal length periods. Each of the equal length periods involves two years, as Fig.3 presents.

<Fig.2 here>

Fig.2 Division of the planning horizon time into increasing length periods

<Fig.3 here>

Fig.3 Division of the planning horizon time into equal length periods

In the first method of dividing the time horizon, each year was considered as a period in the first four years; then two years were viewed as a period in the latter six years; the length of the first three periods in the last ten years is 3 years, and the final period included four years. The different planning horizon division methods resulted in different structures of the scenario sets. In the first, according to the notation defined in section 4.2, the nodal dimension D increases from 2 to 8 along the time horizon. In contrast, D is constant in the second method. The number of statistical specifications used in the moment-matching procedure varies with D . These specifications are conditional

on the realizations in the final year of the predecessor period which is defined as the reference year, and calculated according to (14)-(16). For example, given the annual electricity demand and natural gas price in the reference year, then the specifications S_{VAL_t} of those variables in target year t ranging from year $s+1$ to $s+l$ (where l is the length of the subperiod) can be found by using $t-s$ in equations (14)-(16). After identifying the specifications S_{VAL_t} and then computing y according to equation (9), the nonlinear optimization problem (6)-(8) is solved at the root node for each subperiod length. Because of the stationarity of the GBM process, the tree can then be built by applying the growth ratios corresponding to each branch from the root node to the respective branches from subsequent nodes according to the subperiod lengths.

The growth rates of annual electricity demand and annual average natural gas price are given in Tables 3 through 6. Fig.4 shows the projections of the tree onto the demand-time plane and the price-time plane.

Table 3.

The growth rates of annual electricity demand and natural gas price for periods 1-4

	Demand ratio	Price ratio	Conditional probability
Branch 1	0.9991	0.8759	0.5045
Branch 2	1.0488	1.1807	0.2871
Branch 3	1.0400	1.6067	0.2084

Table 4.

The growth rates of annual electricity demand and natural gas price for periods 5-7

	$t = s + 1$		$t = s + 2$		Conditional Probability
	Demand ratio	Price ratio	Demand ratio	Price ratio	
Branch 1	0.9992	0.8812	1.0117	0.9418	0.5084
Branch 2	1.0489	1.1422	1.0712	1.0884	0.2494
Branch 3	1.0417	1.5806	1.0850	2.0423	0.0242

Table 5.

The growth rates of annual electricity demand and natural gas price for periods 8-9

	$t = s + 1$		$t = s + 2$		$t = s + 3$		Conditional Probability
	Demand ratio	Price ratio	Demand ratio	Price ratio	Demand ratio	Price ratio	
Branch 1	0.9992	0.8812	1.0117	0.9428	1.0265	0.8414	0.5072
Branch 2	1.0489	1.1422	1.0708	1.0839	1.1196	1.5309	0.2499
Branch 3	1.0416	1.5801	1.0852	2.0414	1.0984	2.3866	0.2429

Table 6.

The growth rates of annual electricity demand and natural gas price for period 10

	$t = s + 1$		$t = s + 2$		$t = s + 3$		$t = s + 4$		Conditional Probability
	Demand ratio	Price ratio	Demand ratio	Price ratio	Demand ratio	Price ratio	Demand ratio	Price ratio	
Branch 1	0.9995	0.8729	1.0118	0.8862	1.0265	0.9381	1.0430	1.0322	0.4991
Branch 2	1.0483	1.2465	1.0811	1.3756	1.1057	1.4653	1.1329	1.5016	0.3424
Branch 3	1.0392	1.6696	1.0727	2.2109	1.1205	2.8114	1.1590	3.4928	0.1585

When generating the second set of scenarios, with equal length subperiods, we first realized possible states for even years, and applied equations (14)-(16) with $t-s$ equal to 2 to compute data values for the subsequent even years. We then applied linear interpolation to find values of annual electricity demands and natural gas prices in odd years.

According to equation (9), three branches were made from each node. However, we found two of the three annual growth ratios for annual electricity demand were very similar in the solution to (6)-(8). Substituting one of the two growth ratios with the other one for demand resulted in little change to value of the objective function. Note that creating two branches for demand and three for gas price allows (9) to be satisfied exactly as 2 (demands) + 3 (prices) + 3 (probabilities) – 1 = 7 (statistical specifications). Moreover, the stationarity and equal length subperiods allowed multiple tree nodes at the same time point to be collapsed into a single lattice node. Table 7 shows the growth rates for each branch from a lattice node. Fig.5 shows the demand and price projections of the lattice.

As each set of scenarios covers 10 time periods with three branches from each node, each contains 3^{10} scenario paths. Section 6.5 numerically compares the effects of the two methods on thinning the scenario sets to a manageable size.

Table 7.

The growth rates of annual electricity demand and natural gas price for even years in the second scenario tree

	Demand ratio	Price ratio	Conditional probability
Branch 1	1.0124	0.9015	0.5237
Branch 2	1.0803	1.3093	0.2969
Branch 3	1.0803	2.1403	0.1794

< Fig.4(a) here >

< Fig.4(b) here >

(a)

(b)

Fig.4 Scenario tree with unequal length periods: (a) demand vs. time, (b) gas price vs. time

< Fig.5(a) here >

< Fig.5(b) here >

(a)

(b)

Fig.5 Scenario lattice with equal length periods: (a) demand vs. time, (b) gas price vs. time

6.4 Scenario reduction

As a baseline for comparison, we applied fast forward selection (FFS) to select scenarios from each original scenario set. To implement the FSWC method, model (1)-(5) was solved for each scenario individually as a deterministic “wait-and-see” optimization problem by considering the set of scenarios as a singleton and setting that scenario probability to one. The optimal first-stage decisions $\{U_{gt}\}$ corresponding to each scenario were accumulated over 20

years to get the number of each generator type planned to exist in each year $\{V_{gt}\}$, where $V_{gt} = \sum_{s=1}^t U_{gs}$. Then, the key first-stage variables were defined to be the number of each type of generators in each of the last ten years, and they were formed to be a row vector, i.e. $[V_{1,11}, V_{2,11}, \dots, V_{1,20}, V_{2,20}, V_{3,20}, V_{4,20}, V_{5,20}, V_{6,20}]$. The FSWC method described

in Section 4 can be applied directly on the obtained vectors $[V_j]$, and the L_2 -norm is adopted in the k -means method in Step 3 of FSWC. As mentioned in Section 4, these quantities could partially reflect the growth trend of annual electricity demand, as higher annual electricity demand would require more generation capacity. Because the dramatic differences in annual electricity demand appear evident especially in the last ten years according to Fig.4 and Fig.5, we pay more attention to the numbers of generators added in these years. Since more generators of a given type of generator tend to be built if it has lower building cost and lower fuel price, the key first-stage variables could also partially reflect the relative differences between natural gas price and other fuel prices, which were considered as deterministic over the planning horizon. To compare the outputs of FFS and FSWC, various cardinalities of selected scenario sets were obtained. Considering the computational complexity to select from the large original scenario sets, the selected sets had modest cardinalities of 10, 20, 30, 50 and 100. Because one of the purposes of generation expansion planning is to satisfy demand over the long term, we set the penalty for unserved energy to a large value $P_c = 10^7$ \$/MWh. The scenario reduction procedures and stochastic programming solution process were implemented by using Matlab and CPLEX with the interface provided by Tomlab on computers with 3GHz CPU.

6.5. Numerical results

Scenarios selected by FFS and FSWC methods based on the generated scenario sets are presented in this subsection. Comparisons between these two methods were performed by solving (1)-(5) over a 20 year planning horizon.

6.5.1 Numerical results for the unequal-length period scenario tree

Fig. 6 and 7 illustrate the selected scenarios according to FFS and FSWC methods which are projected on the annual electricity demand vs. time and natural gas price vs. time planes, respectively. Because of the space limitation, only the selected sets with cardinality 100 are shown.

< Fig.6(a) here >
< Fig.6(b) here >
(a)
(b)

Fig.6 Scenarios selected by FFS method from the unequal-length scenario set (tree): (a) demand vs. time, (b) gas price vs. time

< Fig.7(a) here >
< Fig.7(b) here >
(a)
(b)

Fig.7 Scenarios selected by FSWC method from the unequal-length scenario set (tree): (a) demand vs. time, (b) gas price vs. time

< Fig.8(a) here >
< Fig.8(b) here >
(a)
(b)

Fig.8 Comparisons among scenarios from original set, FSWC and FFS methods based on the unequal-length scenario set (tree): (a) expected annual electricity demand, (b) expected average natural gas price

According to Fig.6 and Fig.7, the reduced scenario sets obtained by both reduction heuristics have similar contours with respect to annual electricity demands, but differ somewhat according to the natural gas prices. To make closer comparisons between the reduction methods, expected annual electricity demands and average natural gas prices over the planning horizon based on the two selected subsets as well as the original scenarios have been plotted in Fig 8. The expected demands based on each selected set are very similar to the original (Fig.8(a)). Fig.8(b) illustrates that the expected gas prices according to scenarios selected by the FFS method are closer to the original scenarios than those selected by FSWC. One explanation is that, because of the high penalty for unserved energy, the electricity demand receives more attention in the FSWC method than the natural gas price does. In contrast, the FFS method measures scenario distances by a metric that weights annual electricity demand and natural gas price equally.

To assess the performance of solutions based on the reduced scenario sets, we solved the GEP model (1)-(5) using a reduced scenario set to obtain the optimal first-stage decisions. Tables 8 and 9, which show the optimal first-stage decisions based on each reduced set, illustrate that the expansion strategies based on the two scenario reduction methods have only minor differences.

Table 8.

The cumulative number of generators added based on FFS reduction of the unequal-length period tree

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Base-load	0	0	0	7	13	18	26	31	39	45	45	45	45	45	45	45	45	45	45	45
Gas CC	0	0	0	0	0	0	0	0	0	0	0	1	9	24	37	53	55	65	65	65
Gas CT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
Nuclear	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Wind	91	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105
IGCC	13	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45

Table 9.

The cumulative number of generators added based on FSWC reduction of the unequal-length period tree

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Base-load	0	0	0	7	13	15	20	25	32	39	45	45	45	45	45	45	45	45	45	45
Gas CC	0	0	0	0	0	0	0	0	0	0	0	0	12	17	31	43	55	65	65	65
Gas CT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	23
Nuclear	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Wind	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105
IGCC	12	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45

Fig.9 shows the expected generation by generator type over the planning horizon from the solution to (1)-(5) based on the reduced scenario subsets obtained from FFS and FSWC, respectively, for the unequal-length period scenario tree. The figures show that similar generation patterns result from applying reduced scenarios obtained from either scenario reduction method. IGCC generators start to increase generation from year 2, while the Base-load (coal) generators and those generators which consume natural gas decrease their generation at the same time. Generation from Base-load (coal) then increases after the fifth year. Fig.10 illustrates the expected generation costs resulting from the FFS- and FSWC-based expansion decisions for the unequal-length period tree.

< Fig.9(a) here >

(a)

< Fig.9(b) here >

(b)

Fig.9 Illustration of average energy generation of each type of generator based on the unequal-length period tree: (a) FFS method, (b) FSWC method

< Fig.10(a) here >

(a)

< Fig.10(b) here >

(b)

Fig.10 Expected generation cost of each type of generator over selected scenarios from the unequal-length period tree: (a) FFS method, (b) FSWC method

According to Fig.10, the average generation costs of gas-fired CC and CT generators dramatically increase after the second year in each subplot, and from that year on, the generation costs of Base-load and IGCC generators are much lower than those of CC and CT generators. The potential for high natural gas prices later in the planning horizon favors energy production by coal-fueled Base-load and IGCC generators to minimize expected costs. The generation level of CC dramatically increases after the fifteenth year as the high electricity demand calls for more generators, and only CC and CT are available at that time.

Table 10 summarizes the investment cost, generation cost and penalty for unserved energy over the scenarios selected from the unequal-length period tree by the two reduction methods, respectively, and also illustrates the influences of expansion strategies on the original scenarios. Penalties occur when we solve the stochastic program based on the selected scenarios derived from FSWC method. But it does not mean FSWC performs worse than FFS method does, because FSWC selected more extreme scenarios with higher electricity demand. Comparisons in Table 10 demonstrate that there are no major differences between FFS and FSWC in any cost category. The expected investment and generation costs from FSWC with respect to the selected scenarios are both smaller than those from FFS. One possible reason is that the generation costs of CC and CT from FSWC are lower than those from FFS, and no difference exists among the other types' generation costs, as Fig.10 illustrates. The lower costs from FSWC also derive from the cost discounting in objective function (1). The discount term $(1+r)^{1-t}$ indicates that the later the

same type of generator is constructed, the lower the impact of investment cost for that type of generator. Therefore, despite the total number of Base-load and IGCC generators resulting from FFS and FSWC methods being similar, the discounted investment costs for the two types of generators are different. More CC and CT generators will be built in the strategy derived from FSWC, but because the building costs of Base-load and IGCC are much higher than those of CC and CT, the total investment cost that results from the FFS method is still higher than that derived from the FSWC method. The results of the scenario reduction heuristics are further compared by fixing the first-stage decisions according to the solution obtained with the reduced sets and then solving (1)-(5) as a continuous linear program using the original scenario tree. According to the expected penalty costs with respect to all scenarios, the expansion strategy obtained with the FSWC-reduced scenario subset can provide a more favorable outcome that not only leads to lower investment cost, but also to less unserved energy when the cardinality of the reduced scenario subset is greater than 10.

Table 10.

Expected costs and solution times based on FFS & FSWC for the unequal-length period tree

Number of selected scenarios, n		10		20		30		50		100	
		FFS	FSWC	FFS	FSWC	FFS	FSWC	FFS	FSWC	FFS	FSWC
Expected cost with respect to selected scenarios (\$Billion)	Investment cost	58.56	57.51	58.73	57.42	59.35	58.50	59.17	58.91	59.20	58.61
	Generation cost	170.93	145.32	165.86	143.84	165.06	146.91	161.15	147.74	157.83	145.39
	Penalty (\$10M/MWh)	0	0.01	0	0.43	0	0.23	0	0.14	0	0.19
	Total cost Investment	229.49	202.84	224.59	201.69	224.41	205.64	220.32	206.79	217.03	204.19
Expected cost with respect to all scenarios (\$Billion)	cost Generation	58.56	57.51	58.73	57.42	59.35	58.50	59.17	58.91	59.20	58.61
	cost	157.10	160.78	156.68	160.72	156.15	158.48	156.38	157.40	156.30	157.65
	Penalty (\$10M/MWh)	160.98	15.73	157.26	24.75	83.42	19.10	83.86	38.65	83.86	34.16
	Total cost	376.64	234.02	372.67	242.39	298.92	236.08	299.41	254.96	299.36	250.42
Scenario reduction time (CPU s)	Solve scenario problems	-	5,246	-	5,246	-	5,246	-	5,246	-	5,246
	Cluster	-	4,815	-	4,815	-	4,815	-	4,815	-	4,815
	Select	205,242	2,438	225,166	1,257	352,706	845	560,907	618	1,066,800	273
Solution time with reduced scenario set (CPU s)		1	1	3	3	5	5	14	13	56	63
Total time pruning and solution (CPU s)		20,543	12,500	225,169	11,321	352,711	10,911	560,921	10,692	1,066,856	10,397

The comparison of computational times for FFS and FSWC shown in Table 10 illustrates that FSWC may be more powerful than FFS in scenario reduction. The computational burden of FFS increases much more rapidly than the linear growth in n , the cardinality of the set of selected scenarios. Conversely, the computational times of FSWC may even decrease as n increases. One possible reason is that selecting more scenarios in FSWC leads to a smaller size of each scenario cluster, and FFS must be applied only once in each cluster. Besides, the time spent on all pairwise-distances comparison in each cluster is $O(N^2)$; thus, the smaller cluster size may save much more time.

6.5.2 Numerical results for the equal-length period lattice

The corresponding numerical results for the second scenario set are provided below. As for the first scenario set, there are minor differences between scenarios selected by the two reduction methods with respect to annual electricity demand, but larger differences in terms of the natural gas prices (Fig.11 and Fig 12). As Fig.13(a) shows, comparing to the scenarios selected by FFS, the expected annual electricity demand of the scenarios selected by FSWC is closer to the expected level based on the original scenarios. Similar to the numerical results in Section 6.5.1, the expected natural gas prices of scenarios selected by FFS are closer to the original scenarios, and the same explanation of the phenomenon applies.

< Fig.11(a) here >

(a)

< Fig.11(b) here >

(b)

Fig.11 Scenarios selected by FFS method from the equal-length period lattice: (a) demand vs. time, (b) gas price vs. time

< Fig.12(a) here >

(a)

< Fig.12(b) here >

(b)

Fig.12 Scenarios selected by FSWC method from the equal-length period lattice: (a) demand vs. time, (b) gas price vs. time

< Fig.13(a) here >

(a)

< Fig.13(b) here >

(b)

Fig. 13 Comparison of average electricity demand and natural gas price between two methods based on the equal-length period lattice: (a) expected annual electricity demand, (b) expected average natural gas price

Again, GEP model (1)-(5) was solved to obtain the optimal first-stage decisions to assess the performance of solutions based on the reduced scenario sets. Tables 11 and 12, which show the optimal first-stage decisions based on each reduced set from the equal-length period lattice, illustrate that the expansion strategies based on the two scenario reduction methods have only minor differences.

Table 11.

The cumulative number of generators added based on FFS reduction of the equal-length period lattice

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Base-load	0	0	3	8	14	20	26	32	39	45	45	45	45	45	45	45	45	45	45	45
Gas CC	0	0	0	0	0	0	0	0	0	0	0	10	22	35	48	62	65	65	65	65
Gas CT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	49
Nuclear	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Wind	100	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105
IGCC	39	40	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45

Table 12.

The cumulative number of generators added based on FSWC reduction of the equal-length period lattice

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Base-load	0	0	2	8	9	10	16	22	29	34	41	45	45	45	45	45	45	45	45	45
Gas CC	0	0	0	0	0	0	0	0	0	0	0	0	22	35	49	62	65	65	65	65
Gas CT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	49
Nuclear	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Wind	93	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105	105
IGCC	11	19	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45

Like Fig.9, Fig.14 shows the expected generation by generator type over the planning horizon obtained by solving model (1)-(5) based on the reduced scenario subsets obtained from FFS and FSWC, respectively, for the equal-length period lattice. The figures show similar generation patterns result from applying reduced scenarios obtained from either scenario reduction method. Similar to Fig.9, Base-load and IGCC generators also start to increase their generation from year 2, while the generators which consume natural gas decrease their generation at the same time in equal-length period lattice. Fig.15 illustrates the expected generation costs resulting from the FFS- and FSWC-based expansion decisions for the equal-length period lattice.

< Fig.14(a) here >

(a)

< Fig.14(b) here >

(b)

Fig.14 Illustration of average energy generation of each type of generator based on the equal-length period lattice: (a) FFS method, (b) FSWC method

< Fig.15(a) here >

(a)

< Fig.15(b) here >

(b)

Fig.15 Expected generation cost of each type of generator over selected scenarios from the equal-length period lattice: (a) FFS method, (b) FSWC method

Similar to the pattern shown in Fig.10, the average generation costs of gas-fired CC and CT generators dramatically increase after the second year in each subplot of Fig.15, and from that year on, the generation costs of Base-load and IGCC generators are much lower than those of CC and CT generators. The potential for high natural gas prices later in the planning horizon leads to more energy production by Base-load and IGCC generators to minimize expected costs.

Like Table 10, Table 13 summarizes the investment cost, generation cost and penalty for unserved energy based on scenarios selected from the equal-length period lattice by two reduction methods respectively, and also illustrates the

performance of the expansion strategies with respect to the original scenarios. For the equal-length period lattice, FSWC and FFS methods provide similar results in every cost category, especially when the reduced scenario subset cardinality is set to 20, 50 or 100. Since any unserved energy will be penalized by 10^7 \$/MWh, the penalties in expected cost with respect to all scenarios indicate that only a small amount of demand would be unsatisfied over the planning horizon.

Table 13.

Expected costs and solution times based on FFS & FSWC for the equal-length period lattice

Number of selected scenarios, n		10		20		30		50		100	
		FFS	FSWC	FFS	FSWC	FFS	FSWC	FFS	FSWC	FFS	FSWC
Expected cost with respect to selected scenarios(\$Billion)	Investment cost	59.40	56.85	59.29	57.10	59.57	57.18	59.45	57.75	61.06	57.90
	Generation cost	169.98	159.61	174.21	148.98	174.52	147.58	171.73	146.94	165.21	146.34
	Penalty(\$10M/MWh)	0	0	0	0	0	0	0	0	0	0
	Total cost Investment	229.38	216.46	233.50	206.68	234.09	204.76	231.18	204.69	226.27	204.24
Expected cost with respect to all scenarios (\$Billion)	cost Generation	59.40	56.85	59.29	57.10	59.57	57.18	59.45	57.75	61.06	57.90
	cost	158.48	161.35	158.54	163.39	158.36	163.26	158.45	161.92	156.50	161.68
	Penalty(\$10M/MWh)	0	156.13	0	0	0	0	0	0	0	0
	Total cost	217.88	374.33	217.83	220.49	217.93	220.44	217.90	219.67	217.56	219.58
Scenario reduction time (CPU s)	Solve scenario problems	-	6,618	-	6,618	-	6,618	-	6,618	-	6,618
	Cluster	-	3,448	-	3,448	-	3,448	-	3,448	-	3,448
	Select	111,657	7,316	225,158	1,947	339,612	1,844	567,544	1,796	1,433,273	1,662
Solution time with reduced scenario set (CPU s)		1	1	3	4	6	7	15	16	62	85
Total time pruning and solution (CPU s)		111,658	17,383	225,161	12,017	339,618	11,917	567,559	11,878	1,433,335	11,813

Considering the computational times summarized in Tables 10 and 13, FFS is impractical for selecting modest scenario subsets from large scenario sets. In contrast to the unsatisfactory performance of FFS for these large scenario sets, FSWC requires computational times ranging from 60% of that for FFS when selecting 10 scenarios to less than 1% when selecting 100 scenarios. But it results in similar first-stage decisions and expected costs with respect to the original scenario sets. Moreover, the time required for FSWC is approximately constant over different reduced set cardinalities.

7. Conclusions

This paper provides an approach to generating scenario trees to cover a long time horizon with variable length periods under the assumption that the random quantities follow correlated GBM processes. The cardinalities of the generated scenario sets are controlled both by combining multiple years into periods of varying lengths and by matching the number of branches from each node to the degrees of freedom in the statistical specifications. The FSWC scenario reduction heuristic applied in this paper clusters scenarios based on their wait-and-see solutions and then applies the fast forward selection heuristic within clusters. Thus, it incorporates the stochastic optimization context in addition to the probability distribution of scenarios. The numerical results obtained from a simplified generation expansion planning model indicate the usefulness and efficiency of the FSWC method. The first-stage solutions obtained from the FSWC-reduced scenarios are similar to those obtained from the FFS-reduced scenarios, as are their expected costs when implemented on the whole set of scenarios. But the computational time for FSWC is substantially smaller and does not vary significantly with the cardinality of the reduced set.

In this paper, risk associated with generation expansion decisions has been incorporated by including a penalty for unserved energy but only in terms of expectation. In ongoing research, some simplifying assumptions made in this paper are relaxed and risk is included in terms of conditional value at risk [48]. Because the success of FSWC depends on choosing appropriate key first-stage variables on which to cluster, guidance for their selection in various applications is a significant research issue.

Acknowledgement

Funding for this research was provided by the U.S. Department of Energy's Office of Electricity Delivery and Energy Reliability under the project "The Future Grid to Enable Sustainable Energy Systems."

References

- [1]. Kaut M, Wallace SW. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization* 2007;3(2):257-71.

- [2]. Lurie PM, Goldberg MS. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science* 1998;44(2):203-18.
- [3]. Smith JE. Moment methods for decision analysis. *Management Science* 1993;39(3):340-58.
- [4]. Kouwenberg R. Scenario generation and stochastic programming model for asset liability management. *European Journal of Operational Research* 2001;134(2):279-292.
- [5]. Høyland K, Wallace SW. Generating scenario trees for multistage decision problems. *Management Science* 2001;47(2):295-307.
- [6]. Høyland K, Kaut M, Wallace SW. A heuristic for moment matching scenario generation. *Computational Optimization and Applications* 2003;24:169-85.
- [7]. Dupacová J, Conigli G, Wallace SW. Scenarios for multistage stochastic programs. *Annals of Operations Research* 2000;100:25-53.
- [8]. Pflug GC. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming* 2001;89(2): 251-71.
- [9]. Ross SM. Brownian motion and stationary process. In: *Introduction to probability models*. 10th. ed. New York: Academic Press; 2010. p. 636-7.
- [10]. Kall P, Wallace SW. *Stochastic programming*. New York: John Wiley & Sons; 1994.
- [11]. Dupacová J, Growe-Kuska N, Römisch W. Scenario reduction in stochastic programming: an approach using probability metrics. *Mathematical Programming* 2003;95(3): 493-511.
- [12]. Heitsch H, Römisch W. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications* 2003;24:187-206.
- [13]. Growe-Kuska N, Heitsch H, Römisch W. Scenario reduction and scenario tree construction for power management problems. *Proceedings of IEEE Power Tech Conference; 2003 Jun 23-26. Bologna, Italy. IEEE; 2004.*
- [14]. Morales JM, Pineda S, Conejo AJ, Carrion M. Scenario reduction for futures market trading in electricity markets. *IEEE Transactions on Power Systems* 2009;24(2):878-88.
- [15]. Heitsch H, Römisch W. A note on scenario reduction for two-stage stochastic programs. *Operations Research Letters* 2007;35(6):731-8.
- [16]. Heitsch H, Römisch W. Stability and scenario trees for multistage stochastic programs. *Stochastic Programming: International Series in Operations Research & Management Science*, vol.150, pp.139-164, 2011.
- [17]. Heitsch H, Römisch W. Scenario tree reduction for multistage stochastic programs. *Computational Management Science* 2009;6(2):117-33.
- [18]. Gonen T. *Modern power system analysis*. New York: John Wiley & Sons; 1988.
- [19]. Gorenstin BG, Campodonico NM, Costa JP and Pereira MVF. Power system expansion planning under uncertainty. *IEEE Transactions on Power Systems* 1993;8(1):129-36.
- [20]. Kagiannas AG, Askounis DT, Psarras J. Power generation planning: a survey from monopoly to competition. *International Journal of Electrical Power & Energy Systems* 2004;26(6):413-21.
- [21]. Park YM, Won JR, Park JB, Kim DG. Generation expansion planning based on an advanced evolutionary programming. *IEEE Transactions on Power Systems* 1999;14(1):299-305.
- [22]. Park JB, Park YM, Won JR, Lee KY. An improved genetic algorithm for generation expansion planning. *IEEE Transactions on Power Systems* 2000;15(3):916-22.
- [23]. Pokharel S, Ponnambalam K. Investment planning for electricity generation expansion. *International Journal of Energy Research* 1997;21(2):185-94.
- [24]. Murphy FH, Sen S, Soyster AL. Electricity utility capacity expansion planning with uncertain load forecasts. *AIIE Transactions* 1982;14(1):52-9.
- [25]. Roh JH., Shahidehpour M, Fu Y. Market-based coordination of transmission and generation capacity planning. *IEEE Transactions on Power Systems* 2007;22(4):1406-19.
- [26]. Botterud A, Ilic MD, Wangensteen I. Optimal investments in power generation under centralized and decentralized decision making. *IEEE Transactions on Power Systems* 2005;20(1):254-63.
- [27]. Chuang AS, Wu F, Varaiya P. A game-theoretic model for generation expansion planning: problem formulation and numerical comparisons. *IEEE Transactions on Power Systems* 2001;16(4):885-91.
- [28]. Jin S, Ryan SM. Capacity expansion in the integrated supply network for an electricity market. *IEEE Transactions on Power Systems* 2011;26 (4): 2275 - 2284.
- [29]. Chen Q, Kang C, Xia Q and Zhong J. Power generation expansion planning model towards low-carbon economy and its application in China. *IEEE Transactions on Power Systems* 2010;25(2):1117-25.
- [30]. Jin S, Ryan SM, Watson JP, Woodruff DL. Modeling and solving a large-scale generation expansion planning problem under uncertainty. *Energy Systems* 2011: 2 (3), 209-242
- [31]. Dapklus WD, Bowe TR. Planning for new electric generation technologies-a stochastic dynamic programming approach. *IEEE Transactions on Power Apparatus and Systems* 1984;PAS-103(6):1447-53.
- [32]. Mo B, Hegge J, Wangensteen I. Stochastic generation expansion planning by means of stochastic dynamic programming. *IEEE Transactions on Power Systems* 1991;6(2):662-8.
- [33]. Bloom JA. Long-range generation planning using decomposition and probabilistic simulation. *IEEE Transactions on Power Apparatus and Systems* 1982;PAS-101(4):797-802.

- [34]. Dantzig CB, Glynn PW, Avriel M, Stone JC, Enriken R, Nakayama M (Stanford University, CA). Decomposition techniques for multi-area generation and transmission planning under uncertainty. Technical Report. Palo Alto (CA): Electric Power Research Institute; 1989. Report NO.: EL-6484.
- [35]. Fleten SE, Pettersen E. Constructing bidding curves for a price-taking retailer in the Norwegian electricity market. *IEEE Transactions on Power Systems* 2005;20(2):701-8.
- [36]. EIA [Internet]. Electric Power Industry 2009: Year in Review; c2010-10 [cited 2011 Aug 3]. Available from: http://www.eia.gov/cneaf/electricity/epa/epa_sum.html.
- [37]. EIA [Internet]. Coal Prices in selected years; c2010-10 [cited 2011 Aug 3]. Available from: http://www.eia.gov/totalenergy/data/annual/pdf/sec7_19.pdf.
- [38]. EIA [Internet]. Electric Power Annual 2009; c2010-11 [updated 2011 Apr; cited 2011 Aug 3]. Available from: <http://www.eia.gov/cneaf/electricity/epa/epa.pdf>.
- [39]. Marathe RR, Ryan SM. On the validity of the Geometric Brownian Motion assumption. *Engineering Economist* 2005;50(2):159-92.
- [40]. Wang Y. Scenario reduction heuristics for a rolling stochastic programming simulation of bulk energy flows with uncertain fuel costs [dissertation]. Ames (IA): Iowa State University; 2010.
- [41]. Mathworks [Internet]. R 2011a Statistic Tool Box: Kmeans Documentation; c2011 [cited 2011 Aug 3]. Available from: <http://www.mathworks.com/help/toolbox/stats/kmeans.html>.
- [42]. Sigman K [Internet]. Geometric Brownian motion; c2006 [cited 2011 Aug 3]. Available from: <http://www.columbia.edu/~ks20/FE-Notes/4700-07-Notes-GBM.pdf>.
- [43]. EIA [Internet]. Natural gas data; c2011-08 [cited 2011 Aug 3]. Available from: <http://www.eia.doe.gov/dnav/ng/hist/n9190us3a.htm>.
- [44]. EIA [Internet]. Annual energy review: electricity overview; c2010-08 [cited 2011 Aug 3]. Available from: <http://www.eia.doe.gov/totalenergy/data/monthly/index.cfm#electricity>.
- [45]. EIA [Internet]. Annual electric generator report; c2008-12 [cited 2011 Jun 27]. Available from: <http://www.eia.doe.gov/cneaf/electricity/page/capacity/existingunitsbs2008.xls>.
- [46]. EIA [Internet]. Annual energy outlook 2010 with projections to 2035; c2010-04 [cited 2011 Dec 1]. Available from: [ftp://ftp.eia.doe.gov/forecasting/0554\(2010\).pdf](ftp://ftp.eia.doe.gov/forecasting/0554(2010).pdf)
- [47]. Tidball R, Bluestein J, Rodriguez N, Knoke S [Internet]. Cost and performance assumptions for modeling electricity generation technologies; c2010-11 [cited 2011 Nov 30]. Available from: <http://www.nrel.gov/docs/fy11osti/48595.pdf>.
- [48]. Feng Y, Ryan S. Application of scenario reduction to LDC and risk based generation expansion planning. *Proceedings of IEEE PES General Meeting 2012*. Forthcoming.
- [49]. JCSP [Internet]. Joint coordinated system plan '08; [cited 2012 Mar. 21]. Available: https://www.midwestiso.org/Library/Repository/Study/JCSP/JCSP_Report_Volume_1.pdf.
- [50]. EIA [Internet]. AEO 2012 early release report; c2012-01 [cited 2012 Mar. 21]. Available: <http://www.eia.gov/forecasts/aeo/er/>.