# INCLUDING A DC NETWORK APPROXIMATION IN A MULTIAREA PROBABILISTIC PRODUCTION COSTING MODEL

Yuandong Ji, Member IEEE
Operations Simulation Associates
P.O. Box 1129
Ringgold, GA 30376
ji@babb.com

Benjamin F. Hobbs, Member IEEE
Dept. Geography & Environmental Engineering
The Johns Hopkins University
Baltimore, MD 21218
bhobbs@jhu.edu

**Abstract** — A multiarea power system consists of several areas (subsystems) interconnected by a transmission network. In estimating expected generation costs for such systems, transmission capacity limits of the network should be recognized. Transportation network models have generally been used because of their simplicity, but they only enforce Kirchhoff's current law. AC power flow modeling of the transmission network, which recognizes thermal, voltage and stability constraints, is theoretically best, but is too unwieldy for assessing expected costs. The so-called "DC" linearized network model is adopted here as a compromise, as it enforces both Kirchhoff's current and voltage laws while its linearity facilitates incorporation in probabilistic production costing models. In this paper, we generalize a bounding-based multiarea probabilistic production costing model to include loop flow and resistance losses based on the DC network model. This is the first multiarea model based on efficient convolution methods for production costing that also includes loop flows and resistance losses. Computational examples are presented to highlight the modeling and solution procedures.

**Keywords** — Production costing, generation, power flow, transmission, stochastic optimization, economics

## I. INTRODUCTION

Deregulated power markets require utilities to consider the impact of transmission constraints on transactions, generation costs, and long range planning. Models are needed that can estimate expected total and marginal production costs for a multiarea generation system over the entire range of possible generator outage and load scenarios. Unlike widely used single-area production costing models, multiarea models explicitly consider transmission capacity limits. In multiarea models, the busses are grouped into different areas (subsystems) that are connected by a transmission network.

Which transmission network model is adopted can be crucial in multiarea production costing. AC power-flow modeling of the transmission network, which can recognize thermal, voltage and stability constraints, is theoretically the best. But it demands tremendous computation effort and is therefore impractical for modeling the wide range of load and outage scenarios considered in probabilistic production costing. An alternative approach is to repeatedly use AC power flow and stability programs under various contingencies to determine the transfer capability of the transmission lines [1], and then apply these limits upon individual transmission link flows in a transportation network model. But in transportation models, power flow is modeled like 'apples being hauled in carts,' in F. Schweppe's phrase. This is because the transportation network only enforces Kirchhoff's current law (KCL) [2]. As a result, Kirchhoff's voltage law (KVL) will likely be violated and loop flows disregarded, causing the power transmission capability of a network to be overestimated.

A reasonable compromise between computational tractability and the need for a more realistic representation is the linearized "DC" approximation of the AC load flow equations [3,4,5]. The approximation enforces both KCL and KVL. Because the approximation is linear, the computational effort required by the resulting production costing model is acceptable. Most previous multiarea production costing models have been based on the transportation network model [2,6,7,8]. But as Lee [2] points out, there is a need for both transportation network (capacity flow) and DC network (DC flow) models; a preference towards one or the other can be based on the particular application. Lee [2] presents a multiarea model with DC load flow constraints which he is able to solve for a small system. In [9], convolution methods are used to calculate the probability distribution of flows over selected transmission lines in large systems by taking advantage of the fact that flows in linearized models can be expressed as a linear function of the generation of individual units. However, that method cannot impose limits on flows. Such limits can be imposed in DC-based chronologic models, such as GE MAPS [10], which naturally lend themselves to probabilistic analysis using Monte Carlo methods [11].

The computational intensity of the Monte Carlo approach has motivated our exploration of effective load duration curve (LDC) approaches [e.g., 12] to multiarea probabilistic production costing with DC load flow constraints. Our purpose here is to present a practical method for including DC constraints in probabilistic production costing models for large systems. In the next section, we summarize the linearized DC approximation used in our model. Then in Section III, we incorporate the derived DC model (as additional constraints) into the multiarea production costing problem formulation. A direct solution approach to the probabilistic production costing problem is discussed. In Section IV, the bounding-based solution approach to the DC flow-based production costing problem is presented. The bounding approach originally adopted a

PE-460-PWRS-0-06-1997 A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for publication in the IEEE Transactions on Power Systems. Manuscript submitted July 31, 1996; made available for printing June 9, 1997.

transportation network model [7], and has been used to estimate expected production costs for systems with as many as five areas and 132 units [13]. Incorporation of the DC flow model in the bounding approach requires only data on line impedances, additional linear constraints to enforce KVL, and minor modifications to the solution algorithms. A simple case study is presented to illustrate the mechanics of the method in Section V, along with an application to a 42 unit, four area system. Extensions of the model are summarized in Section VI, including representation of resistance losses and varying loads.

## II. DC POWER FLOW MODEL

Basic Model. A network (with $M$ nodes) is used to represent the grid connecting the areas of a $M$-area power system. Any pair of nodes $\{m,n\}$ may be connected by a transmission link $mn$ (which might represent several physical lines between these areas). The power injection, voltage, etc., of an area are assumed to be unique--which of course disregards voltage differences and transmission constraints within an area. Based on the real and reactive power injections in the areas, the real and reactive power flows can be computed based on standard AC power flow equations for each line [4,5]. However, for probabilistic production costing of large systems, this computationally intensive approach is impractical because the nonlinear AC flow equations would need to be solved numerous times.

As an alternative, the DC "linear power flow model" is sometimes used [4,5]. It consists of a set of linear equations obtained from the AC model by assuming:

(a) per unit (p.u.) voltages in all areas have magnitude 1;
(b) the resistances $r_{mn}$ of the transmission lines are relatively small compared to the reactances $x_{mn}$; and
(c) the differences in bus voltage angles $\theta_m$ across the lines are small, i.e., $\sin(\theta_m-\theta_n) \approx \theta_m-\theta_n$ and $\cos(\theta_m-\theta_n) \approx 1$.

With these assumptions, reactive power can be disregarded in calculating real power flows. As a result, real power flow $P_{mn}$ from $m$ to $n$, $n>m$, in p.u., can be calculated as:

$$P_{mn} = (\theta_m - \theta_n)/x_{mn} \qquad \forall mn \qquad (1)$$

Under the DC assumptions, node $m$'s KCL equation is:

$$P_m = -\sum_{k=1}^{m-1} P_{km} + \sum_{k=m+1}^{M} P_{mk} \qquad m=1,2,\ldots,M \qquad (2)$$

where $P_m$ is the net power injection from node $m$ into the network. Now, let $Z$ be the number of flows $P_{mn}$ (i.e., the number of branches in the network, since $P_{mn}=-P_{nm}$). Equation (2) involves only $M$-1 independent equations, since the equation for node $M$ is just the sum of the equations for nodes 1,..,$M$-1. Thus, if we arbitrarily define $\theta_m=0$, that equation plus (1),(2) gives $Z+M$ independent equations. Fixing the power injections $P_m$ and the reactances $x_{mn}$, these $Z+M$ equations can be used to solve for the $Z$ $P_{mn}$'s and the $M$ $\theta_m$'s. These equations represent the DC power flow model of our transmission network, which satisfies both KCL and KVL.

$P_m$ in turn can be expressed in terms of the generation and load variables for our multiarea costing model [7]:

$$P_m = \sum_{i\in i(m)} g_i + ue_m - L_m \qquad m=1,2,\ldots,M \qquad (3)$$

where $i(m)$ is the set of indices of generation units located in area $m$, $g_i$ is the generation of unit $i$, $L_m$ is the load (power demand) in area $m$, and $ue_m$ is the unserved energy at $m$. Thus (2) becomes:

$$\sum_{i\in i(m)} g_i + ue_m + \sum_{k=1}^{m-1} P_{km} - \sum_{k=m+1}^{M} P_{mk} = L_m \qquad m=1,2,..,M \qquad (4)$$

Equations (1),(4) and $\theta_m=0$ also constitute a DC load flow model, and can be solved for flows and angles if the generation, loads, and unserved energy are given.

Eliminating $\theta_m$. These linear equations could be included directly in the probabilistic production costing model of Sections III and IV. Constraints could then be added to limit power flows over individual lines and through interfaces. But to reduce the model's size, it is helpful to eliminate the $\theta_m$ variables. Schweppe et al. [4] present a compact way of doing so, involving matrix manipulations of (1) and (2). However, we prefer an alternative formulation that eliminates the $\theta_m$ while preserving KCL in the form (4). This fits our model because the upper bound in our solution procedure (Section IV) exploits the dual variable for that constraint.

This approach develops KVL analogues which, together with the KCL equations (4), are a DC load flow model. $Z$-($M$-1) independent KVL equations are required and can be directly written as follows. Since the network of our power system is a connected graph, a tree can always be specified for this network. Denote this tree by $\tau$. The remaining $Z$-$M$+1 links $mn \notin \tau$ define the corresponding "cotree." Each link on the cotree defines a loop with some branches from the tree $\tau$. For a loop consisting of link $mn \notin \tau$ and links $nm_1$, $m_1m_2$, ..., $m_km \in \tau$, we can express its KVL as follows:

$$(\theta_m - \theta_n) + (\theta_n - \theta_{m_1}) + (\theta_{m_1} - \theta_{m_2}) + \ldots + (\theta_{m_k} - \theta_m) = 0 \qquad (5)$$

By (1), we can rewrite (5) as follows:

$$x_{mn}P_{mn} + x_{nm_1}P_{nm_1} + x_{m_1m_2}P_{m_1m_2} + \ldots + x_{m_km}P_{m_km} = 0 \qquad (6)$$

More generally, let $rs \in B_{mn}$ be the set of links on the loop (including $mn$ itself) associated with branch $mn \notin \tau$. Assume that in each $rs$, $r$ and $s$ occur in the order in which they are encountered around the loop. Then the $Z$-($M$-1) KVL equations can be expressed as:

$$\sum_{rs\in B_{mn}} x_{rs}P_{rs} = 0 \qquad mn \notin \tau \qquad (7)$$

Equations (4),(7) constitute the $Z$ equations of the DC load flow model, and uniquely determine the $Z$ flows $P_{mn}$.

## III. PRODUCTION COSTING & DC POWER FLOW

Model Definition. In this section, we formulate the multiarea electric power system production costing problem based on the DC transmission network model of Section II.

Consider a power system with $I$ generation units located in $M$ areas which are connected by a transmission grid that can be represented by the aforementioned DC flow network. The following assumptions are made to simplify presentation of the basic method, although most are not necessary for the algorithm of the next section [13]. For planning purposes, the marginal cost for each unit is considered constant over its range of output and unit outages are independent. Also assume that unit commitment and minimum run considerations can be ignored, that transmission outages are negligible (from the point of view of expected production costs), and--for the moment--that loads in each area are constant. Define the system outage state $X=\{X_i\}$, a vector of dimension $I$, in which $X_i=$ fraction of capacity of generation unit $i$ that is available, $i=1,2,\ldots,I$. Thus $X_i=1$ indicates that unit $i$ is available, while $X_i=0$ means that $i$ is on outage. These assumptions yield the following multiarea production costing problem, which is a stochastic linear program (LP):

$$E\{C(X)\} = E\left\{ \underset{g,t,ue}{\mathrm{Min}} \left[ \sum_{i=1}^{I} C_i g_i + \sum_{m=1}^{M} CUE_m ue_m \right] \right\} \tag{8}$$

subject to:

*KCL:*

$$\sum_{i \in i(m)} g_i + ue_m + \sum_{n \neq m} (t_{nm} - t_{mn}) = L_m \qquad m=1,2,..,M \tag{9}$$

*KVL:*

$$\sum_{rs \in Bmn} x_{rs}(t_{rs} - t_{sr}) = 0 \qquad mn \notin \tau \tag{10}$$

*Lower and (random) upper bounds on generation:*

$$0 \le g_i \le X_i CAP_i \qquad i=1,2,\ldots,I \tag{11}$$

*Flow upper bounds:*

$$0 \le t_{mn} \le T_{mn} \qquad n,m = 1,2,\ldots,M \tag{12}$$

*Nonnegativity of unserved energy:*

$$ue_m \ge 0 \qquad m=1,2,\ldots,M \tag{13}$$

where:

$C(X)=$ minimum cost achievable, given outage state $X$

$C_i=$ variable generation cost (\$/MWh) of unit $i$

$CUE_m=$ social cost of unserved energy (\$/MWh), area $m$

$CAP_i=$ capacity (MW) of generation unit $i$

$E\{\ \}=$ expectation operator

$g=\{g_i\}=$ vector of power generation

$ue=\{ue_m\}=$ vector of unserved demands

$t=\{t_{mn}\}=$ vector of nonnegative (artificial) flows. The use of nonnegative flows $t_{mn}$ (instead of unrestricted $P_{mn}$) in (9), (10) and (12) is necessary because we use LP (Simplex) algorithms to solve the problem. The DC load flow model has been converted by substituting $t_{mn}$-$t_{nm}$ for $P_{mn}$ in (4),(7).

$T_{mn}=$ real power transfer capacity from area $m$ to area $n$. Since volt-ampere flows typically exceed real power flows by 5% to 15% (depending on the power angles), $T_{mn}$ should be lowered by a small amount (say, 10%) if it is based on VA rather than real power limits.

To put it another way, the expected production cost is the optimal value of the objective function of a stochastic optimization problem in which the decision variables are generation and power flows. The constraints include KCL (9), KVL (10), and bounds upon the variables. (Constraints can also be put upon selected sums of the $t_{mn}$ to represent interface flow limits.) By solving this problem, we can obtain the expected production cost, the expected generation of each unit, the marginal cost of each area, and the expected flow on each tieline.

Solution by Exhaustive Enumeration. Now consider the solution of the probabilistic production costing problem (8)-(13). Similar to the development in [7], a direct yet impractical approach is to compute *C(X)* for each possible value of state *X*, where

$$C(X) = \underset{g,t,ue}{\mathrm{Min}} \left\{ \sum_{i=1}^{I} C_i g_i + \sum_{m=1}^{M} CUE_m ue_m \right\} \tag{14}$$

subject to (9)-(13). Recognizing that this is a deterministic LP, the algorithm is conceptually straightforward. Then the expected production cost *E{C(X)}* can be calculated as

$$E\{C(X)\} = \sum P(X)C(X) \tag{15}$$

where *P(X)* is the probability that the outage state takes value *X*. Since we assume that the outages of the generation units are independent, this probability can be calculated as the product

$$P(X) = \prod_{i=1}^{M} P(X_i) = \prod_{i=1}^{M} [r_i(1-X_i) + (1-r_i)X_i] \tag{16}$$

where $0 \le r_i \le 1$ is the forced outage rate (FOR) of unit $i$, and $P(X_i=0)=r_i$, $P(X_i=1)=1$-$r_i$. Of course, this straightforward approach is infeasible for large systems, as the number of LPs that need to be solved is $2^I$. Another solution method must be sought. In the next section, we summarize a practical approach based on calculating upper and lower bounds to the true expected cost, building upon [7].

## IV. SOLUTION BY BOUNDING-BASED APPROACH

Our approach to solving (8)-(13) is to construct upper and lower bounds on the production cost, and then tighten the bounds by partitioning the state space until a satisfactory approximation is achieved.

Generally speaking, bounds upon the expected optimal cost (8) are defined below [7]:

$$\text{Lower bound:} \quad LB = C(E\{X\}) \tag{17}$$

$$\text{Upper bound:} \quad UB = \mathrm{Min}_t\, E\{C(X|t)\} \tag{18}$$

where $C(X|t)$ is the minimum cost of production and unserved load under generator state $X$, given transmission flows $t$. That is, cost is optimized for a fixed set of flows.

Lower bound computation. The lower bound (17) is the "derating cost" and can be computed by solving a LP. The derating cost *C(E{X})* is the solution of the following optimization problem:

$$C(E\{X\}) = \underset{g,t,ue}{\mathrm{Min}} \left\{ \sum_{i=1}^{I} C_i g_i + \sum_{m=1}^{M} CUD_m ue_m \right\} \tag{19}$$

subject to

$$0 \le g_i \le E\{X_i\} CAP_i = (1-r_i) CAP_i \qquad \forall i \tag{20}$$

along with constraints (9,10,12,13). This is a deterministic multiarea production costing problem that uses derated capacities for the generation units. It can be shown that this gives a lower bound by considering (a) that the optimal value of a LP is a convex function of the right hand side of the constraints and (b) Jensen's inequality [7]. Note, however, that unlike the case of the transportation network-based model [7], a minimum flow network algorithm cannot be used here because of the presence of constraint (10). Of course, a general LP code can still be used.

Upper Bound Computation. Equation (18) is the production cost, given that (a) the same transmission flows $t$ are imposed upon all states and (b) $t$ is optimized. This equation is an upper bound to $E\{C(X)\}$ since the expected cost resulting from imposing the same transmission flows $t$ on all the states can be no less than the cost that results from allowing $t$ to be tailored separately for each state.

The upper bound (18) can be calculated by generalized Benders decomposition (GBD), an iterative procedure that breaks an optimization problem into a master problem and subproblems [14]. As in [7], we define a master problem that determines a trial value of flows $t$, and subproblems, one per area, that calculate the expected production cost and unserved energy for each area, given that value of $t$. The GBD procedure iterates between the master and subproblems until a convergence criterion is satisfied.

For the DC flow problem, the master problem in iteration $J$ of the algorithm can be stated as the following LP:

$$\operatorname*{Min}_{t,c}\{c\} \tag{21}$$

subject to (10) and (12) and

$$c \ge \sum_{m=1}^{M}\left[\bar{c}_m^{(j)} + \bar{\lambda}_m^{(j)}\left[\sum_{n\neq m}\left(t_{mn} - t_{nm} - t_{mn}^{(j)} + t_{nm}^{(j)}\right)\right]\right] \quad \forall j<J \tag{22}$$

where:

- $c=$ the objective function value (expected production cost of the current trial solution).
- $\bar{c}_m^{(j)}=$ area $m$'s expected production cost, as calculated by the subproblem for each area $m$ in previous iteration $j$.
- $\bar{\lambda}_m^{(j)}=$ area $m$'s expected dual multiplier, as calculated by the subproblem for each area $m$ in iteration $j$. It equals the expected value of the dual variable to the demand constraint (24) (defined below) in that subproblem.
- $t_{mn}^{(j)}=$ transmission flow from area $m$ to area $n$ in the master problem solution for iteration $j$.

In the $J^{th}$ iteration, the master problem yields flows $t = t^{(J)}$, which are then sent to the subproblems. The subproblem for area $m$ in iteration $J$ is the familiar single area probabilistic production costing problem, just as in our transportation network-based model [7]:

$$\bar{c}_m = \operatorname*{Min}_{g,ue_m} E\left(\sum_{i\in i(m)} C_i g_i + CUD_m ue_m\right) \tag{23}$$

subject to

$$\sum_{i\in i(m)} g_i + ue_m = L_m + \sum_{n\neq m}\left(t_{mn}^{(J)} - t_{nm}^{(J)}\right) \tag{24}$$

$$0 \le g_i \le X_i CAP_i, \quad \forall i \in i(m) \tag{25}$$

$$ue_m \ge 0 \tag{26}$$

The right side of (25) is stochastic because generator availabilities $X_i$ are random. Here, the flows $t_{mn}^{(J)}$ from iteration $J$'s master problem are fixed in iteration $J$'s subproblems.

Convolution [e.g., 12] for single area probabilistic production costing can be used to solve the subproblems. Convolution is necessary because of the random upper bounds to generation (25), and Section VI's extension to include random loads. Our GBD algorithm can be proven to have finite step convergence in a manner similar to [7].

The two bounds defined above are tightened iteratively by partitioning the outage state space and calculating bounds for each subset. Three alternative partitioning methods are described in [13]; version 1 is most appropriate for smaller systems, while versions 2 and 3 have been found to be more efficient for larger systems. In each iteration, the outage space is partitioned further, and recursive formulas are used to calculate the inputs required by the upper and lower bound models. If the resulting updated bounds are not tight enough, each subset's bounds are compared to decide which subsets need to be further partitioned. This partition process continues until $LB$ and $UB$ are sufficiently close. The convergence of the bounds to the true expected production cost ("$\varepsilon$-convergence") can be proved similar to [7]. Practical convergence experience is documented in [13] for transportation-based models.

## V. COMPUTATIONAL EXAMPLES

### A. A Simple Example

To permit the reader to reproduce our method, the DC probabilistic production costing model is applied to a system having eight small generating units and three areas. The DC flow network is from [5], with $x_{12}$ = 0.2 p.u., $x_{23}$ = 0.25 p.u., $x_{13}$ = 0.4 p.u., respectively. The generation, load and transmission flow limits are drawn from [7]: the demand in area 1 is 126 MWh, in area 2 is 81 MWh, and in area 3 is 89 MWh. The unit of generation is 100 MW, and the unit voltage is 31.6 kV. The bi-directional flow limits $T_{mn}$ are: 53 MW between area 1 and area 2, 10 MW between areas 1 and 3, and 33 MW between areas 2 and 3. The unserved energy penalty $CUE_m$ is 100 \$/MWh for all m. Generating unit data are:

| Plant | Area | MW | \$/MWh | FOR |
|---|---|---|---|---|
| 1 | 1 | 100 | 10 | 0.1 |
| 2 | 1 | 50 | 30 | 0.05 |
| 3 | 1 | 50 | 30 | 0.05 |
| 4 | 2 | 75 | 15 | 0.1 |
| 5 | 2 | 50 | 20 | 0.05 |
| 6 | 3 | 50 | 15 | 0.05 |
| 7 | 3 | 25 | 40 | 0.03 |
| 8 | 3 | 25 | 40 | 0.03 |

Now we formulate a production costing problem for this system using the DC power flow network. We specify a tree with branches 12 and 13 ($\tau$ = {12,13}). Following the procedure in Section II, it is straightforward to obtain (9)-(13), with the KVL constraint (10) being:

$$0.2(t_{12}\text{-}t_{21}) + 0.25(t_{23}\text{-}t_{32}) - 0.4(t_{13}\text{-}t_{31}) = 0$$

This problem is solved below three times: the first for deterministic capacity, the second using exhaustive enumeration, and the third time with the bounding-based approach.

*Simple Problem, Deterministic Capacity.* This is a case in which $X$ is not random and is set equal to:

$$X = [1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 0\ \ 1\ \ 1]^T$$

That is, we calculate the production cost of a scenario with unit 6 out and all others available. This problem can be solved exactly by linear programming using (9)-(13) and

(14). The results are as follows, in MW:

$$g_1=100,\ g_2=20.44,\ g_3=0,\ g_4=75,$$
$$g_5=50,\ g_6=0,\ g_7=25,\ g_8=25$$
$$ue_1=0,\ ue_2=0,\ ue_3=0.56$$
$$t_{12}=0,\ t_{21}=-15.56,\ t_{13}=10,\ t_{31}=0,\ t_{23}=28.44,\ t_{32}=0$$

Total cost equals 5793.8 \$/hour. We can calculate the following power injections from the areas:

$$P_1 = 100+20.44-126 = -5.56$$
$$P_2 = 75+50-81 = 44$$
$$P_3 = 25+25+0.56-89 = -38.44$$

These results of our DC network formulation can be verified by checking to see that the original DC power flow equation (1) is satisfied. Arbitrarily let node 1 be the reference node, so that $\theta_1=0$. By (1):

$$\theta_2=\theta_1-x_{12}(t_{12}-t_{21})/100 = 0.0311$$
$$\theta_3=\theta_1-x_{13}(t_{13}-t_{31})/100 = -0.04$$

(The division by 100 MW in the above is necessary to convert the MW power flows to per unit flows, assuming a 100 MW base). Then as a check, $t_{23}-t_{32}$ (= 28.44 MW/100 MW = 0.2844 pu) should equal $(\theta_2-\theta_3)/x_{23}$ (= [0.0311-(-0.04)]/0.25 = 0.2844 pu), which it does.

Of interest is a comparison of this solution to an AC load flow. Assume that the line resistances are 10% of their reactances. Assume further that the power angles at nodes 2 and 3 are 20° and their real power injections are as given above. Finally, let node 1 be the swing bus, whose voltage magnitude is 1 p.u. and power production (real and imaginary) can vary to ensure that the power injections at the other nodes can be maintained.

The AC load flow results are similar to the DC approximation. The nonlinear AC load flow equations result in node 1 absorbing 5.22 MW of real power rather than the DC model's 5.55 MW. Comparing the AC real power flows along the three links with the above DC results, the mean absolute percentage error is 0.8%. By comparison, the transportation algorithm can potentially make errors of 100% or even more by routing all power over one path. On average, the DC power flow understates the AC volt-ampere flows by 5.8% because it omits VAR flows.

*Simple Problem, Exhaustive Enumeration.* Now let us find the expected production cost for this example by complete state enumeration: the solving of a LP for each of the $2^8=256$ possible realizations of $X$, followed by calculation of the weighted average $\sum P(X)C(X)$. The exact expected production cost thus obtained is

$$E\{C(X)\} = 5079.1 \text{ \$/hour}$$

For the same example but using the transportation network model [4], the expected production cost was instead 5025.8. In general, inserting the KVL constraints (10) of the DC model must lead to higher production costs (or at least costs that are no lower). This is because additional constraints can only shrink the feasible region, and the result cannot be a better value of the objective function.

*Simple Problem, Bounding Method.* Now let us find the expected production for this example by the bounding-based approach described in Section IV. After 6 partitions using the first partition method in [13], we have

$$UB_6 = 5082.3;\ LB_6 = 5074.1 \text{ \$/hour}$$

Thus the error $(UB-LB)/E\{C(X)\}$ is less than 0.2%. Twenty partition subsets were considered in the iterations, far less than the 256 needed in exhaustive enumeration, with less than ten used in the final partition (iteration 6). In practice, only the lower bound model using the final partition would be used in subsequent production costing runs.

B. A Larger Example

This system has 4 areas, and a total of 42 units. If we denote unit $i$ in area $m$ as having characteristics $m\_i$ = $(CAP_i,r_i,C_i)$, the data for the units are as follows:
unit 1_1=(700,0.3,14), unit 1_2=(700,0.25,16),
unit 1_3=(200,0.05,18), unit 1_4=(120,0.02,25),
unit 1_5=(100,0.03,25), unit 1_6=(50,0.025,30),
unit 1_7=(50,0.025,50), unit 1_8=(50,0.025,50),
unit 1_9=(50,0.025,50), unit 1_10=(50,0.025,50),
unit 2_11=(800,0.25,16), unit 2_12=(650,0.2,18),
unit 2_13=(150,0.035,30), unit 2_14=(100,0.02,30),
unit 2_15=(80,0.05,30), unit 2_16=(50,0.025,40),
unit 2_17=(50,0.025,40), unit 2_18=(30,0.025,40),
unit 2_19=(20,0.025,50), unit 2_20=(20,0.025,50),
unit 3_21=(900,0.2,16), unit 3_22=(600,0.1,18),
unit 3_23=(100,0.02,25), unit 3_24=(150,0.01,25),
unit 3_25=(100,0.01,30), unit 3_26=(50,0.01,35),
unit 3_27=(30,0.025,40), unit 3_28=(20,0.025,40),
unit 3_29=(20,0.025,45), unit 3_30=(20,0.025,40),
unit 3_31=(20,0.025,50), unit 3_32=(50,0.025,50),
unit 3_33=(20,0.025,50), unit 4_34=(600,0.3,16),
unit 4_35=(800,0.25,17), unit 4_36=(100,0.02,30),
unit 4_37=(100,0.01,30), unit 4_38=(80,0.02,35),
unit 4_39=(50,0.01,40), unit 4_40=(30,0.025,45),
unit 4_41=(30,0.025,45), unit 4_42=(20,0.025,50).

The bi-directional flow limits are: $T_{12}=300$ and $T_{13}=T_{14}=T_{23}=T_{24}=T_{34}=200$, in MW. The loads are $L_1=1220$, $L_2=1285$, $L_3=1188$, and $L_4=1350$, in MW. The unserved energy penalty $CUE_m=100$ \$/MWh for $m=1,2,3,4$. The reactances of the network are $x_{34}=2$ p.u., and $x_{12}=x_{13}=x_{14}=x_{23}=x_{24}=1$ p.u.. By the procedure developed in section II, the corresponding KVL constraints (15) can be:

$$(t_{24}-t_{42}) + (t_{41}-t_{14}) + (t_{12}-t_{21}) = 0$$
$$2(t_{43}-t_{34}) + (t_{31}-t_{13}) + (t_{14}-t_{41}) = 0$$
$$(t_{12}-t_{21}) + (t_{23}-t_{32}) + (t_{31}-t_{13}) = 0$$

Using the bounding-based method, the initial iteration yields the following initial bounds on the expected cost:

$$UB_0=113{,}682;\ LB_0=86{,}750 \text{ \$/hour}$$

Thus the initial gap $(UB-LB)/LB$ is 31%. After 8 partitions using the first partitioning approach in [13], about 200 subsets are created, and the gap reduces to 0.44%, with

$$UB_8=97{,}030;\ LB_8=96{,}606 \text{ \$/hour}$$

As a comparison, we also compute the production cost for this system but assuming a transportation network--i.e., where flow paths can be chosen [13]. The initial "gap" $(UB-LB)/LB$ is 31% ($UB_0=113{,}681$, $LB_0=86{,}750$, identical to the DC model). After 8 partitions, and the gap falls to 0.28% ($UB_8=95{,}224$, $LB_8=94{,}958$). Thus, the transportation model cost is about 2% less than the DC cost.

## VI. EXTENSIONS: LOSSES AND VARYING LOADS

Simple extensions of the probabilistic production costing model (8)-(13) would increase its realism. For instance, more complex representations of generators (such as must-run levels and partial outages) can be handled using standard probabilistic production costing formulations [12]. More complicated transmission formulations are also easily included, among them multiple lines between a pair of areas and interface constraints that limit the total flow over a set of lines. Basically, any transmission constraints that can be formulated as linear equalities or inequalities upon flows can be included in the model without posing difficulties for the probabilistic costing procedure.

In this section, we discuss two possible extensions in more detail: the case of varying loads and the generalization of our DC network model to include resistance losses.

Including Varying Loads. A generalized version of (8) that considers variations in loads is [13]:

$$E\{C(X,L)\} = E\left[\operatorname*{Min}_{g,t,ue}\left[\sum_{i=1}^{I} C_i g_i + \sum_{m=1}^{M} CUE_m ue_m\right]\right] \tag{27}$$

subject to (9)-(13). $L$ is the system load state, which is a vector $L=\{L_m\}$, with $L_m$ defined as in (3). Since load data are usually stored and used in hourly form, we consider $L$ to be a discrete random vector with $N$ possible values, each representing the load in one hour and having a probability of $1/N$. $N$ is the number of hours in the period to be simulated. As a result, the sample space is now the set of all possible capacity and load states $\{X,L\}$. The probability of a particular state is $P(X,L)$.

To solve this problem, one obvious but inefficient approach would be to repeat the production costing calculation for each hour in the planning period. The other extreme is to calculate costs using the average load. However this might badly underestimate the cost.

A practical approach is to group loads into a few (say, $H$) subsets of fairly similar loads using cluster analysis [13]. Then the $H$ averages of each load cluster are used in the lower bound production costing runs, while the load distribution within each cluster is used to create the single area load duration curves used in the upper bound runs. The resulting $H$ estimates of the upper and lower bounds are weighted by their corresponding hours to obtain the expected production cost. As shown in [13], this procedure preserves the property that the resulting solutions are lower and upper bounds, respectively, to the true expected cost. Those results remain applicable here because they only require that the constraint set of the models be linear, as they are for the DC flow model.

We anticipate that computational experience will be similar to that with the transportation model [13]. With that model, we found that 10 or so load clusters capture the bulk of the variation in loads, and that little additional accuracy is obtained by further subdividing the loads. We also found that load variations and capacity outages contribute roughly equally to the gap between $UB$ and $LB$, and that both load clustering and partitioning of the outage space are needed for precise estimates of expected cost.

Including Resistance Losses. Because resistance losses can significantly increase generation costs and line loadings, an important improvement in our DC multiarea probabilistic production costing model would be their explicit inclusion. Under the assumptions made by the DC load flow model (Section II), the real power loss $PL_{mn}$ due to power flow from $m$ to $n$ can be approximated as [4]:

$$PL_{mn} = |E_m|\,|E_n|\left(r_{mn}\left[r_{mn}^2 + x_{mn}^2\right]/x_{mn}^2\right)|t_{mn} - t_{nm}|^2 \tag{28}$$

where $E_m$ = complex voltage of node $m$. Consistent with the DC assumption that $|E_m| \approx |E_n| \approx 1$ p.u., we get:

$$PL_{mn} = \alpha_{mn}\,|t_{mn} - t_{nm}|^2 \tag{29}$$

where $\alpha_{mn}$ is the constant term involving $r_{mn}$ and $x_{mn}$.

As an example, consider the flow of 15.56 MW (0.1556 p.u.) on the link between 1 and 2 in the simple deterministic example at the start of Section V. If $r_{mn} = 0.1x_{mn} = 0.02$, then $\alpha_{mn} = 0.202$. Then, by (29), the loss on that line is 0.00049 p.u., or 0.049 MW. For the entire network in the deterministic problem, if resistances are one-tenth the magnitude of the reactances, then the total loss obtained by (29) (0.29 MW) is close to the loss from the AC load flow model described in that section (0.33 MW). The difference is primarily due to DC's model disregarding of the effect of VAR flows on losses.

Building upon these results, we propose the following generalized DC network production costing with resistance losses. There are two differences compared to (8)-(13): (a) deduction of the loss associated with $t_{mn}$ from the power balance at node $n$; and (b) construction of a piecewise linearization of the quadratic loss function. The model is: minimize expected cost (27), subject to:

$$\sum_{i \in i(n)} g_i + ue_m + \sum_{n \neq m}^{M} \sum_{k=1}^{L} [(1-\beta_{nm}^k)t_{nm}^k - t_{mn}^k] = L_m \quad \forall m \tag{30}$$

$$0 \le t_{mn}^k \le T_{mn}^k \qquad \forall mn,k \tag{31}$$

plus constraints (10,11,13). The new notation includes:

- $t_{mn}^k$ = MW flow for segment $k$ of the flow $t_{mn}$, $k=1,2,..,L$. This segmentation is necessary for the piecewise linearization. Note that $\Sigma_k t_{mn}^k = t_{mn}$.
- $T_{mn}^k$ = upper bound for segment $k$ of the flow $t_{mn}$ (MW), such that $\Sigma_k T_{mn}^k = T_{mn}$.
- $\beta_{mn}^k$ = resistance loss factor for segment $k$ of the flow $t_{mn}$, $0 \le \beta_{mn}^k \ll 1$. This is the slope of segment $k$ of the piecewise linear approximation of (31):

$$PL_{nm} = \sum_{k=1}^{L} \beta_{nm}^k t_{nm}^k \qquad \forall nm \tag{32}$$

This expression is the loss that occurs if $t_{nm}$ is positive and $t_{mn}$ is zero. The loss is deducted from the power flowing into node $m$, as shown in (30). Because (29) is a convex function, $\beta_{mn}^{k1} < \beta_{mn}^{k2}$ for $k1 < k2$. (Note that it will generally be suboptimal for a $t_{mn}$ and its opposite $t_{nm}$ to both be positive, since that will inflate losses. In most optimal solutions, only flows in one direction will appear in the solution; thus $|t_{mn}-t_{nm}|$ will equal whichever of the two flows is positive. Further, the segments $k=1,2,..,L$ will enter the solution in the correct order, since the optimization algorithm will choose the segments with the lowest loss rates first. Thus, (32) will be an approximation to $\alpha_{mn}|t_{mn}-t_{nm}|^2$.)

The piecewise linearization has removed the nonlinearity (quadratic losses), yielding a LP. Since the upper and lower bound models of Section IV can accommodate any linear constraints upon transmission flows, the model can be solved by our bounding approach. Thus, this is the first convolution-based multiarea probabilistic production costing model to include resistance losses; previously, losses have only been included in chronologic models [10].

Returning to the simple deterministic example in Section V, an implementation of the above method using a three piece approximation ($L$ = 3) yields a good approximation to the exact DC load flow results (losses of 0.31 MW, compared to 0.29 MW in the quadratic DC simulation and 0.33 MW in the AC simulation). As an example of the coefficients, if $\alpha_{mn}$ = 0.0202 for the line between 1 and 2, and that line's capacity $T_{12}$ of 53 MW is split into three equal segments of $T_{12}^k$ = 17.7 MW apiece, then:

$$\beta_{12}^1=\beta_{21}^1=0.0036;\ \beta_{12}^2=\beta_{21}^2=0.011;\ \beta_{12}^3=\beta_{21}^3=0.018.$$

These values were obtained by matching the vertices of the piecewise approximation with the actual quadratic function.

As noted above, our piecewise linearization assumes that the segments will come into the solution in order of increasing $\beta_{mn}^k$. This will occur in the optimal solution if, as is usually the case, the marginal cost (spot price) of power at a node $m$ is positive. But because of KVL, it is possible for the spot price at a $m$ to be *negative*; i.e., for an increase in load at $m$ to result in *lower* total generation cost (a simple example is presented in [15]). As a result, the high loss segments (high $\beta_{mn}^k$) might enter the solution before the low loss segments, which would exaggerate losses. Yet negative spot prices/marginal costs are relatively rare in practice, so for planning purposes we believe that this approximation of resistance losses will be a useful one. If this problem does occur, it can be readily identified by noting if high loss segments for $mn$ are in the solution when its low loss segments are not. In that circumstance, the easiest fix is to use a one-piece approximation for $mn$'s losses; this inflates losses somewhat, but much less so than if $mn$'s segments enter in the wrong order.

## VII. CONCLUSION

The usefulness of the bounding-based method for multiarea probabilistic production costing [7,13] has been enhanced by the modeling of loop flow phenomena by incorporation of a linearized DC load flow model. The method is demonstrated on two systems, and extensions to random loads and resistance losses are summarized. Research is now needed on the incorporation of this model into procedures for distributed planning of utilities and evaluation of power transactions.

**Acknowledgment.** This work was funded by the National Science Foundation, grant ECS 9696014, with earlier support provided by EPRI. Comments by S. McCusker and anonymous reviewers are gratefully acknowledged. The authors are responsible for any opinions or errors.

## REFERENCES

[1] R.R. Austria, X.Y. Chao, N.D. Reppen, and D.E. Welsh, "Integrated Approach to Transfer Limit Calculations," *IEEE Comp. Appl. Power*, 8, 1995, 48-52.

[2] F.Lee, *Multi-Area Power System Reliability and Production Costing*, EPRI EL-6912, Palo Alto CA, 1990.

[3] J. Mitra and C. Singh, "Incorporating the DC Load Flow Model in the Decomposition-Simulation Method of Multi-Area Reliability Evaluation," *IEEE Trans. Pow. Sys.*, 11(3), 1996, 1245-54.

[4] F.C. Schweppe, M.C. Caramanis, R.D. Tabors, and R.E. Bohn, *Spot Pricing of Electricity*, Kluwer, Boston, 1988.

[5] A.J. Wood and B.F. Wollenberg, *Power Generation, Operation, and Control*, 2nd Ed., J. Wiley, New York, 1995.

[6] Q. Ahsan, K.F. Schenk, and R.B. Misra, "Expected Energy Production Cost of Two Interconnected Systems with Correlated Demands," *IEEE Trans. Pow. Appar. Sys.*, 99(5), 1983, 2155-64.

[7] B.F. Hobbs and Y. Ji, "A Bounding Approach to Multiarea Production Costing," *IEEE Trans. Pow. Sys.*, 10(2), 1995, 853-9.

[8] L.R. Noyes, "Two-Area Probabilistic Production Costing by the Method of Bivariate Cumulants," *IEEE Trans. Pow. Appar. Sys.*, PAS-102(2), 1983, 433-43.

[9] Power Technology, Inc., "TOPS: Transmission-Oriented Production Simulation Program," Software Program-238, Schenectady, NY, 1997.

[10] N.W. Simons, L.L. Garver, and R. Mukerji, "Transmission Constrained Production Costing: A Key to Pricing Transmission Access," *Public Util. Fortnightly*, 131(3), 1993.

[11] M.V.F. Pereira et al., "Chronological Probabilistic Production Costing and Wheeling Calculations with Transmission Network Modeling," *IEEE Trans. Pow. Sys.*, Vol. 7(2), 1992, 885-9.

[12] J.P. Stremel et al., "Production Costing Using the Cumulant Method of Representing the Equivalent Load Curve," *IEEE Trans. Pow. Appar. Sys.*, 99(5), 1990, 1947-1956.

[13] B.F. Hobbs, Y. Ji, C. Chang, K.A. Loparo, J. Jober, and M. Ohman, "An Improved Bounding-Based Approach to Multiarea Probabilistic Production Costing," *IEEE Trans. Pow. Sys.*, 11(2), May 1996, 1024-30.

[14] A.M. Geoffrion, "Generalized Benders Decomposition," *J. Optim. Theory Appl.*, 10(4), 1972, 237-60.

[15] W. Hogan, "DC Optimal Power Flow Model, V.3.0,"Unpublished Memorandum, Putnam, Hayes, and Bartlett, Inc., Cambridge, MA, 1996.

Yuandong Ji is a Project Engineer with Operations Simulation Associates, Ringgold, GA. He received his B.S. degree in Instrumentation from Qinghua University, Beijing, China, a M.S. degree in Electrical Engineering from Tianjin University, Tianjin, PRC, and a Ph.D. degree in Systems Engineering from Case Western Reserve University, Cleveland, OH, in 1978, 1981 and 1987 respectively. He was an engineer at Tianjin Petro-Chemical Fibre Company, a lecturer at Tianjin University, a Research Engineer in the Motion Study Lab at Cleveland V.A. Medical Center, and a Research Associate and a Visiting Assistant Professor in the Department of Systems, Control and Industrial Engineering at Case Western Reserve University. He also was a Staff Engineer at Rigo & Rigo Associates, Berea, OH. His research interests include systems, control, environmental, electric power, and biomedical engineering problems.

Benjamin F. Hobbs is Professor of Geography and Environmental Engineering at The Johns Hopkins University, Baltimore, MD. He is also a consultant with the FERC Office of Economic Policy. He earned a Ph.D. in Environmental Systems Engineering from Cornell University in 1983. From 1977 to 1979, he was with Brookhaven National Laboratory, Upton, NY. From 1982 to 1984 he was a Wigner Fellow with Oak Ridge National Laboratory, Oak Ridge, TN. From 1984 to 1995 he was on the faculty of the departments of Systems Engineering and Civil Engineering at Case Western Reserve University, Cleveland, OH. Among his research interests are resource evaluation, simulation of deregulated power markets, and economic incentives for environmental improvement.