IEEE Transactions on Power Systems, Vol. 3, No. 2, May 1988

THE MIXTURE OF NORMALS APPROXIMATION TECHNIQUE FOR EQUIVALENT LOAD DURATION CURVES

George Gross Nancy V. Garapic
Pacific Gas & Electric Company
San Francisco, California

Bruce McNutt
IBM Corporation
San Jose, California

Abstract - For applications such as capactiy expansion planning and screening studies of alternative expansion plans, there is a need for a fast, accurate, reliable and robust method to approximate the equivalent load duration curves (e.l.d.c.'s) for production costing. The concept of equivalent load, which is the sum of the customer load demand and the outage capacity of the generating blocks, plays a key role in probabilistic production costing. This paper describes a newly developed approximation technique based on mixtures of normal distributions. The mixtures of normals approximation (m.o.n.a.) for e.l.d.c.'s proceeds in three steps. First, the system load random variable is approximated by a mixture of normals distribution. Next, the approximation of the outage random variable of a group of one or more units by a mixture of normals distribution is obtained. Finally, these two approximations are combined to derive the m.o.n.a. of each e.l.d.c. The technique makes extensive use of the properties of mixtures of normal distributions. The construction of the m.o.n.a. for the system load random variable can be interpreted in terms of partitioning the load into various categories based on the load magnitudes. A salient feature of the m.o.n.a. technique is the simple recursive formula for convolving or "rolling in" and for deconvolving or "rolling out" the contribution of each generating block. The performance of the m.o.n.a. technique is analyzed in terms of its ability to fit the original load duration curve, its ability to fit the e.l.d.c.'s, its accuracy and robustness. Numerical results indicate that the m.o.n.a. technique is very robust, accurate and rapid. Comparison with conventional and cumulant-based techniques shows that the m.o.n.a. technique has excellent comparative computational performance on both well-behaved and pathological systems.

INTRODUCTION

Production costing is the basic tool used for assessing the variable effects of alternative resource plans. To take into account the uncertainty inherent in both the system load demand and the supply system, a probabilistic production costing framework is used [1,3]. In this framework, the planning period is divided into study periods. For each study period, the load is modeled as a random variable (r.v.) with the complement of its cumulative distribution function (c.d.f.) given by the load duration curve (l.d.c.). The loading of each generation unit of the supply system is represented by the outage capacity r.v. that incorporates the effects of forced outages. Under the reasonable assumption of mutual independence of the outage capacity r.v. of each unit and load r.v., the basic element in the simulation of the energy production process is the convolution of the distributions of r.v.'s. Central to the probabilistic production costing analysis is the equivalent load duration curve (e.l.d.c.). The e.l.d.c. is the complement of the c.d.f. of the equivalent load r.v. The equivalent load is the sum of the system load and the outage capacity of one or more generation units.

The major difference among various probabilistic production costing approaches lies in the manner in which the e.l.d.c. is represented. In the conventional approach, the e.l.d.c. is represented at each point of a uniformly spaced grid. Linear interpolation is then used to determine the value at intermediate points. In a piece-wise linear approximation of this kind, the numerical convolution process is time consuming since each successive e.l.d.c., which is the result of the convolution, must be evaluated at each grid point.

Among the most prominent alternative approaches to the approximation of e.l.d.c.'s are the cumulant-based techniques [4,5,7,8]. These are based on the use of moments of the equivalent load. An extremely attractive feature of this approach is that the convolution/deconvolution involves merely the addition/subraction of cumulants.

While in many cases cumulant-based techniques obtain good fits to the e.l.d.c.'s, they have some inherent limitations resulting in poor performance [6]. Moreover, since the approximation to the e.l.d.c.'s provided by the cumulant-based techniques is not the complement of an actual c.d.f., it is possible in certain cases to obtain negative values for the e.l.d.c. [6]. The most serious drawback of cumulant-based approaches is their poor performance in obtaining a good fit of the system load duration curve [7]. Since the inaccuracy in the approximation of the original load duration curve is propagated through the convolution process to each e.l.d.c., it is seen that this is a serious problem.

For applications such as capacity expansion planning and screening of alternative expansion plans, thers is a need for a fast, accurate, robust method to approximate the e.l.d.c.'s. We have developed a new approximation technique based on mixtures of normals distributions [9]. In this paper we describe the mixture of normals approximation (m.o.n.a.) technique and its implementation. The m.o.n.a. technique uses mixtures of normals to represent the l.d.c. and the supply system and makes extensive use of the properties of these distributions. The representation of the l.d.c. by a mixture of normals distribution is used in the development of a cumulant-based technique in [8]. We present representative numerical results. Extensive testing of the m.o.n.a. technique indicates that it is accurate, efficient and robust. Comparison with conventional and cumulant-based techniques indicates that the m.o.n.a. technique has excellent comparative computational performance on both well-behaved and pathological systems.

MIXTURE OF NORMALS

We consider a set of normally distributed r.v.'s each with mean μ_i and variance σ_i^2, $i=1,\ldots w$. A mixture of normals distribution (m.o.n.d.) is a convex combination of normal distributions. Let $\Phi(\cdot;\mu_i,\sigma_i)$ denote the c.d.f. of the ith normal r.v. $F(\cdot)$ is said to be a m.o.n.d. if

87 WM 041-7 A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE/PES 1987 Winter Meeting, New Orleans, Louisiana, February 1 - 6, 1987. Manuscript submitted January 30, 1986; made available for printing November 14, 1986.

$$F(x) = \sum_{i=1}^{w} \omega_i\, \Phi(x;\ \mu_i,\ \sigma_i) \qquad \sum_{i=1}^{w} \omega_i = 1;\ \omega_i \geq 0,\ 1 \leq i \leq w \quad (1)$$

Here ω_i is the weight of the ith normal component and w is the number of terms in the mixture. The mixture of normals density function, its mean and variance are readily evaluated from the parameters w, ω_i, μ_i, and σ_i. M.o.n.d.'s have the important property that the convolution of two m.o.n.d.'s is itself a m.o.n.d. Appendix A gives a precise statement of the m.o.n.d. properties.

THE M.O.N.A. TECHNIQUE

The m.o.n.a. technique for the representation of e.l.d.c.'s makes extensive use of m.o.n.d.'s and their properties. The first phase of the m.o.n.a. technique is the representation of the system l.d.c. Recall that the l.d.c. for a particular period such as a week or a month is obtained by rearranging the hourly loads in order of decreasing magnitude. The rearranged hourly loads are samples from distinct populations, e.g., week-day peak, shoulder, off-peak and weekend loads. Each population may appear in the system l.d.c. as a separate "mode." This is seen by the sequence of alternating flat and steep regions of the l.d.c. Because the l.d.c. is interpreted as the complement of a c.d.f. we may view the c.d.f. as a multimodal distribution. Regions of high density correspond to sharply sloping sections of the l.d.c. This observation motivates the application of a mixture of normals distribution for approximating the system l.d.c. Thus we approximate $\mathcal{L}(x)$ by

$$\mathcal{L}^a(x) = 1 - \sum_{k=1}^{K} \omega_k\, \Phi(x;\ \mu_k,\ \sigma_k) \qquad \sum_{k=1}^{K} \omega_k = 1,\ \omega_k \geq 0 \qquad (2)$$

In this approximation, the loads of the simulation period are partitioned into K classes based on load level. For each class $k = 1,2,\ldots,K$ the mean μ_k and the variance σ_k^2 are evaluated. The weight ω_k is set to be the fraction of the period that the loads belong to class k. The m.on.a. $\mathcal{L}^a(x)$ can approximate very closely the multimodal behavior of the l.d.c.'s. An example of a typical utility system l.d.c. and its m.o.n.a. is given in Fig. 1.

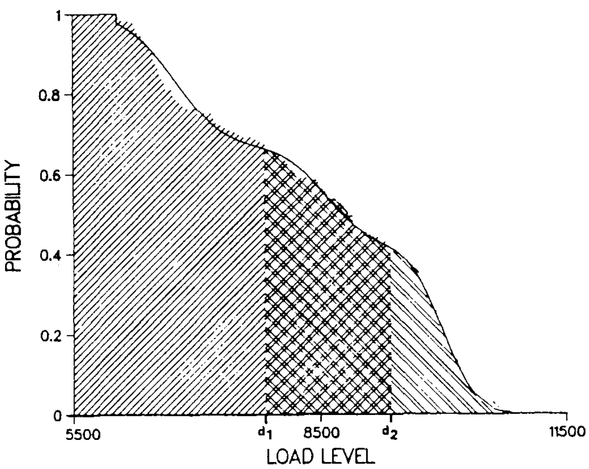

Fig. 1. System l.d.c. and its m.o.n.a. with $K = 3$. The demarcation points d_1 and d_2 are tunable parameters of the m.o.n.a.

The second phase of the m.o.n.a. technique is the approximation of the outage capacity r.v. for a group of 1 or more units by a m.o.n.d. To illustrate this phase let us consider for the sake of simplicity a system with N two-state units each loaded as a single block. Let unit i have capacity c_i and availability p_i. The probability of unit i being forced out is then $q_i = 1 - p_i$. The outage capacity r.v. $\underset{\sim}{Z}_i$ of unit i is

$$\underset{\sim}{Z}_i \triangleq \begin{cases} 0 & \text{with probability } p_i \\ c_i & \text{with probability } q_i \end{cases}$$

The c.d.f. of the outage capacity of n units is

$$F_n(x) = P\,\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \leq x\} \qquad 1 \leq n \leq N$$

The m.o.n.a. of $F_n(x)$ is obtained by partitioning $\mathcal{S}_n$, the set of all possible states of units $1,2,\ldots,n$. Let $r > 0$ be some integer. Let $\mathcal{B}_{n,j}$ be the subset of the states in which j units are on outage, for $j = 0, 1,2,\ldots,r-1$. Let $\mathcal{B}_{n,r}$ be the subset of the states in which r or more units are on outage. Then

$$\mathcal{S}_n = \mathcal{B}_{n,0} \cup \mathcal{B}_{n,1} \cup \ldots \cup \mathcal{B}_{n,r},\ \mathcal{B}_{n,i} \cap \mathcal{B}_{n,j} = \phi \quad i \neq j$$

Conditioning over the subsets of $\mathcal{S}_n$

$$F_n(x) = \sum_{j=0}^{r} P\left\{\left[\sum_{i=1}^{n} \underset{\sim}{Z}_i \leq x\right] \middle|\ \mathcal{B}_{n,j}\right\}\ P\{\mathcal{B}_{n,j}\}$$

We approximate each conditional probability density term in the summation by a normal so that we may approximate $F_n(x)$ by

$$F_n^a(x) = \sum_{j=0}^{r} \Phi(x;\ \mu_{n,j},\ \sigma_{n,j})\ \pi_{n,j} \qquad (3)$$

where,

$$\mu_{n,j} = E\left\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \middle|\ \mathcal{B}_{n,j}\right\},\ \sigma_{n,j}^2 = \mathrm{var}\left\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \middle|\ \mathcal{B}_{n,j}\right\},\ \pi_{n,j} \triangleq P\,\{\mathcal{B}_{n,j}\}$$

An important feature of this approximation is that simple recursive formulae for evaluating $\pi_{n,j}$, $\mu_{n,j}$, and $\sigma_{n,j}$ can be derived for adding (convolution) or withdrawing a unit (deconvolution). Tables I and II present the recursive relations.

The general case of multiple-state multiple-block units is treated in Appendix B. The m.o.n.a. of the outage capacity has the identical form of Eq. (3); the recursive formulae to evaluate the parameters are, however, different.

The last phase of the m.o.n.a. technique uses the convolution property of the m.o.n.d.'s to derive the m.o.n.a. of the e.l.d.c. corresponding to the loading of the group of units $1,2,\ldots,n$. We derive in Appendix C the following relation:

$$\mathcal{L}_n^a(x) = 1 - \sum_{k=1}^{K} \sum_{j=0}^{r} \omega_k\, \pi_{n,j}\, \Phi(x;\ (\mu_k + \mu_{n,j}),\ [\sigma_k^2 + \sigma_{n,j}^2]^{\frac{1}{2}}) \quad (4)$$

The application of the m.o.n.a. technique to a simple system is presented in [9]. The numerical example reproduces the m.o.n.a. computations for a 3 generating unit system.

IMPLEMENTATION AND TEST RESULTS

The m.o.n.a. technique was implemented within the framework of a test-bed developed specifically for the purpose of testing e.l.d.c. approximation techniques. The m.o.n.a., cumulant-based, and piece-wise linear approximation techniques were implemented into the test-bed. The test-bed permits the testing and the comparison of the performance of e.l.d.c. approximation techniques on a consistent basis.

Table I: Recursive evaluation of the outage capacity m.o.n.a. parameters for convolution for two-state units.

$$\pi_{n+1,0} = \pi_{n,0}\, p_{n+1}$$

$$\pi_{n+1,j} = \pi_{n,j}\, p_{n+1} + \pi_{n,j-1}\, q_{n+1} \qquad 0 < j < r$$

$$\pi_{n+1,r} = \pi_{n,r} + \pi_{n,r-1}\, q_{n+1}$$

$$\mu_{n+1,0} = 0$$

$$\mu_{n+1,j} = \frac{\pi_{n,j}\, p_{n+1}\, \mu_{n,j} + \pi_{n,j-1}\, q_{n+1}\, (\mu_{n,j-1} + c_{n+1})}{\pi_{n+1,j}} \qquad 0 < j < r$$

$$\mu_{n+1,r} = \frac{\pi_{n,r}\, (\mu_{n,r} + q_{n+1}\, c_{n+1}) + \pi_{n,r-1}\, q_{n+1}\, (\mu_{n,r-1} + c_{n+1})}{\pi_{n+1,r}}$$

$$\sigma^2_{n+1,0} = 0$$

$$\sigma^2_{n+1,j} = \frac{\pi_{n,j}\, p_{n+1}\, (\sigma^2_{n,j} + \mu^2_{n,j}) + \pi_{n,j-1}\, q_{n+1}\, [\sigma^2_{n,j-1} + (\mu_{n,j-1} + c_{n+1})^2]}{\pi_{n+1,j}} - \mu^2_{n+1,j} \qquad 0 < j < r$$

$$\sigma^2_{n+1,r} = \Big\{\pi_{n,r}\, [\sigma^2_{n,r} + p_{n+1}\, \mu^2_{n,r} + q_{n+1}\, (\mu_{n,r} + c_{n+1})^2] + \pi_{n,r-1}\, q_{n+1}\, [\sigma^2_{n,r-1} + (\mu_{n,r-1} + c_{n+1})^2]\Big\} / \pi_{n+1,r} - \mu^2_{n+1,r}$$

Table II: Recursive evaluation of the outage capacity m.o.n.a. parameters for deconvolution for two-state units.

$$\pi_{n,0} = \frac{\pi_{n+1,0}}{p_{n+1}}$$

$$\pi_{n,j} = \frac{\pi_{n+1,j} - \pi_{n,j-1}\, q_{n+1}}{p_{n+1}} \qquad 0 < j < r$$

$$\pi_{n,r} = \pi_{n+1,r} - \pi_{n,r-1}\, q_{n+1}$$

$$\mu_{n,0} = 0$$

$$\mu_{n,j} = \frac{\pi_{n+1,j}\, \mu_{n+1,j} - \pi_{n,j-1}\, q_{n+1}\, (\mu_{n,j-1} + c_{n-1})}{\pi_{n,j}\, p_{n+1}} \qquad 0 < j < r$$

$$\mu_{n,r} = \frac{\pi_{n+1,r}\, \mu_{n+1,r} - \pi_{n,r-1}\, q_{n+1}\, (\mu_{n,r-1} + c_{n+1})}{\pi_{n,r}} - q_{n+1}\, c_{n+1}$$

$$\sigma^2_{n,0} = 0$$

$$\sigma^2_{n,j} = \frac{\pi_{n+1,j}\, (\sigma^2_{n+1} + \mu^2_{n+1,j}) - \pi_{n,j-1}\, q_{n+1}\, [\sigma^2_{n,j-1} + (\mu_{n,j-1} + c_{n+1})^2]}{\pi_{n,j}\, p_{n+1}} - \mu^2_{n,j} \qquad 0 < j < r$$

$$\sigma^2_{n,r} = \frac{\pi_{n+1,r}\, (\sigma^2_{n+1,r} + \mu^2_{n+1,r}) - \pi_{n,r-1}\, q_{n+1}\, [\sigma^2_{n,r-1} + (\mu_{n,r-1} + c_{n+1})^2]}{\pi_{n,r}} - p_{n+1}\, \mu^2_{n,r} - q_{n+1}\, (\mu_{n,r} + c_{n+1})^2$$

The performance of the m.o.n.a. technique was assessed based on tests using a variety of utility systems and conditions. The test results described in this paper utilize the EPRI Synthetic Utility D, PGandE Derived, and Texas Electric Systems. Basic characteristics of these systems are given in [9]. The cases considered include both well-behaved and pathological systems. The EPRI D [2] and PGandE Derived [9] systems are large. The Texas Electric Utility System, on the other hand, is small. It was selected in order to include in the testing a system on which cumulant-based techniques have not had completely satisfactory performance [6]. The accuracy of m.o.n.a. and the other alternative e.l.d.c. approximation techniques evaluated are measured with respect to benchmark results obtained using a piece-wise linear approximation with a 1 MW grid [9].

The results of the investigation of the goodness-of-fit of the m.o.n.a. for the system l.d.c. show the strong capability of the m.o.n.a. technique to fit l.d.c.'s of any shape and form [9]. Fig. 2 displays the performance of m.o.n.a. on the Texas Electric system. The accuracy of the m.o.n.a. in Eq. (2) is virtually the same for the number of terms $2 \leq K \leq 4$. The accuracy of the m.o.n.a. fit to the l.d.c. is also not sensitive to the exact location of the chosen load demarcation points. The excellent ability of the m.o.n.a. technique to accurately approximate l.d.c.'s overcomes a major disadvantage of cumulant-based methods [6].

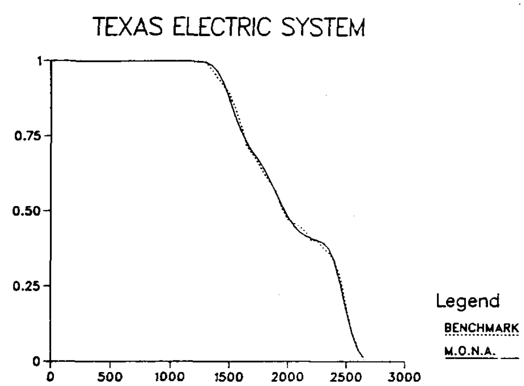

Fig. 2. Plots of the m.o.n.a. computed l.d.c. and the benchmark for the Texas Electric system.

The results of the investigation of the m.o.n.a. techniques's ability to approximate e.l.d.c.'s indicate that the m.o.n.a. for e.l.d.c.'s is accurate even for a small number of blocks, i.e., even when the large sample property of the central limit theorem does not hold. That is, the m.o.n.a. technique gives acceptable accuracy for base-loaded blocks. Test results show that the accuracy of the e.l.d.c. approximation is quite insensitive to the number of classes (r+1) in the m.o.n.a. in Eq. (3) [9]. Fig. 3 shows the pointwise error in the "mixed" e.l.d.c. using the m.o.n.a. technique for the EPRI D system. The "mixed" e.l.d.c. is obtained by plotting for each block k the (k-1)th e.l.d.c. over the interval where the block is loaded. The good fit of the m.o.n.a. technique computed e.l.d.c.'s, as evidenced by the error plot of Fig. 3, is representative of the very good accuracy of the m.o.n.a. technique. The insensitivity of the m.o.n.a. to the number of terms in the representation of the load r.v. and the precise location of the load demarcation points also hold in the approximation for e.l.d.c's [9]. The m.o.n.a. technique is very robust, as it performs very well on pathological cases. The overall experimental evidence indicates that the m.o.n.a. techniques produces excellent results for the evaluation of production costs [9].

The performance of the m.o.n.a. technique was compared to that of the following three alternative approximate techniques: homogeneous and nonhomogeneous piece-wise linear algorithms and the Type A Gram-Charlier 8 moments-based cumulant scheme. The results indicate that the comparative performance of the m.o.n.a. is

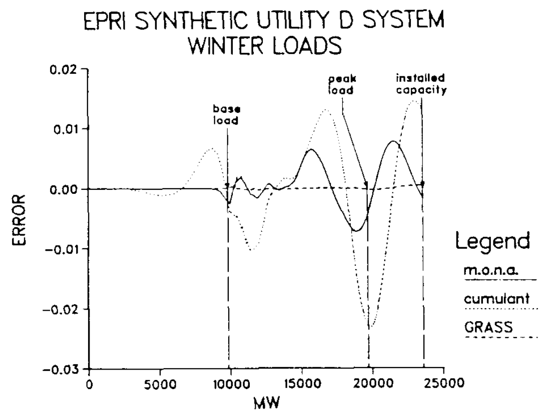

Fig. 3. The point-wise error in the "mixed" e.l.d.c. computed with three techniques for the EPRI D system. The error is measured with respect to the benchmark.

excellent in terms of accuracy and speed. Fig. 3 provides a comparison of the m.o.n.a. technique with the cumulant-based technique and the variant of the homgeneous piece-wise linear approximation algorithm of the GRASS production costing package developed at PGandE [3,9]. Table III provides a comparison of the m.o.n.a. technique with alternative approximation schemes in terms of computation times. The accuracy and efficiency of the m.o.n.a. technique coupled with its robustness lead to the conclusion that it is a superior scheme for evaluating e.l.d.c.'s.

Table III: Comparison of computation times for the approximation techniques on the PGandE derived system under winter load conditions for a 3-period simulation.

Approximation Technique	CPU Seconds
Benchmark	173.08
Cumulant-based technique	1.70
GRASS	4.48
Linear homogeneous	7.79
Linear nonhomogeneous	14.07
m.o.n.a. with 2 terms in the supply side mixture	1.89
m.o.n.a. with 6 terms in the supply side mixture	2.69

CONCLUSION

We have reported on the development and testing of an efficient and accurate approximation technique for the evaluation of e.ld.c.'s. The strong capability of the m.o.n.a. technique to approximate l.d.c.'s and e.l.d.c.'s overcome one of the principal deficiencies of other approximation techniques such as cumulant-based ones [7]. The excellent comparative performance of the m.o.n.a. technique with respect to that of commonly used approximation schemes leads to the conclusion that the m.o.n.a. technique is a superior scheme for evaluating e.l.d.c.'s. An extension of the ideas used for the m.o.n.a. technique is the development of a scheme based on a mixture of cumulants. Limited testing indicates this to be a promising approach. An extension in a different direction is the approach reported in [10].

ACKNOWLEDGEMENTS

This work was performed as part of the Electric Power Research Institute's RP1808-3 project. We are indebted to Dr. J. Delson, the EPRI Project Manager, for encouragement and support. Particular acknowledgement is made of Dr. Steve Mykytyn, who worked on the initial report on this research project; and, consequently, this paper has benefited from detailed comments from Drs. Hugh Outhred, Danny Sutanto, Brian Manhire and Messrs. Jenkins and Stremmel. We thank them for their comments and interest.

REFERENCES

[1] H. Baleriaux, et.al., "Simulation de l'exploitation d'un Parc de Machines Termiques de Production d'electricite Couple a des Stations de Pompage, "Extrait de la Revue E (Edition S.R.B.E.), Vol. 5, pp. 3-24, 1967.

[2] H.W. Zaininger, A.J. Wood, H.K. Clark, T.F. Laskowski, and J.D. Burns, "Synthetic Electric Utility Systems for Evaluating Advanced Technologies," EPRI EM-285, Final Report for Project TPS 75-615, February 1977.

[3] F.F. Wu and G. Gross, "Probabilistic Simulation of Power System Operation for Production Cost and Reliability Evaluation," IEEE International Symposium on Circuits and Systems, IEEE Pubilication No. 77 CH 1188-2 CAS, pp. 877-899, April 1977.

[4] N.S. Rau, P. Toy and K.F. Schenk, "Expected Energy Production Costs by the Method of Moments," IEEE Trans. on Pow. App. & Syst., Vol. PAS-99, pp. 1908-1917, September/October 1980.

[5] J.P. Stremel, et.al., "Production Costing Using the Cumulant Method of Representing the Equivalent Load Curve," IEEE Trans. on Pow. App. & Syst., Vol. PAS-99, pp. 1947-1955, September/October 1980.

[6] D.J. Levy and E.P. Kahn, "Accuracy of the Edgeworth Expansion of LOLP Calculations in Small Power Systems," IEEE Tran. on Pow. & Syst., Vol. PAS-101, pp. 986-996, April 1982.

[7] M. Caramanis, et.al., "Probabilistic Production Costing: An Investigation of Alternative Algorithms," Journal of Electrical Power and Energy Systems, Vol. 5, No. 2, April 1983, pp. 75-86.

[8] N.S. Rau and C. Necsulescu, "An Investigation of Two Methods for the Probabilistic Energy Production Simulation," IEEE Trans. on Pow. App. & Syst., Vol. PAS-102, pp. 2543-2551, August 1983.

[9] G. Gross, B. McNutt and N.V. Garapic, "The Mixture of Normals Approximation of Equivalent Load Duration Curves," EPRI EA/EL-4266 RP1808-3 Final Report prepared by Pacific Gas and Electric Company, October 1985.

[10] C.K. Yin and M. Mazumdar, "A Large Deviation Approach to Computation of Production Costing Indices Using Mixture of Normals Approximation," submitted to IEEE PES, 1985.

APPENDIX A: PROPERTIES OF THE M.O.N.D.

A. The mean μ of the m.o.n.d. or Eq. (1) is

$$\mu = \sum_{i=1}^{w} \omega_i \mu_i \tag{A-1}$$

B. The variance σ^2 of the m.o.n.d. of Eq. (1) is

$$\sigma^2 = \sum_{i=1}^{w} \omega_i (\sigma_i^2 + \mu_i^2) - \mu^2 \tag{A-2}$$

C. The convolution of two m.o.n.d.'s is itself a mixture of normals distribution whose parameters can be simply determined from those of the two m.o.n.d.'s.

The derivation of Eqs. (A-1) and (A-2) is a simple exercise. We prove Property C using the following Lemma [9]: Let $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ be two independent r.v.'s each with corresponding m.o.n.d.'s $F_{\underset{\sim}{X}}(x) = \sum_{i=1}^{w_x} \omega_i \Phi(x; \mu_i, \sigma_i)$ and $F_{\underset{\sim}{Y}}(y) = \sum_{j=1}^{w_y} \eta_j \Phi(y; m_j, s_j)$. Then the distribution of $\underset{\sim}{X} + \underset{\sim}{Y}$ obtained from the convolution of $F_{\underset{\sim}{X}}(\cdot)$ with $f_{\underset{\sim}{Y}}(\cdot)$, where $f(y) = \frac{dF}{dy}$, is given by

$$F_{\underset{\sim}{X} + \underset{\sim}{Y}}(z) = \sum_{k=1}^{w_z} \gamma_k \Phi(z; \mu_k', \sigma_k')$$

with

$$w_z = w_x w_y, \ \gamma_k = \omega_i \eta_j, \ \mu_k' = \mu_i + m_j, \ \sigma_k' = [\sigma_i^2 + s_j^2]^{\frac{1}{2}}$$

APPENDIX B: SUPPLY SYSTEM REPRESENTATION

Each generating unit may be represented using a multi-state model. We denote the maximum capacity of unit i by c_i. If unit i has (s_i+1) states, define for $k = 0,1,\ldots,s_i$ the quantities:

$c_{ik} \triangleq$ capacity of unit i associated with state k

$$\underset{\sim}{X}_{ik} \triangleq \begin{cases} 1 & \text{if unit i is in state k} \\ 0 & \text{otherwise} \end{cases}$$

$$p_{ik} = P\{\underset{\sim}{X}_{ik} = 1\} \qquad \sum_{k=0}^{s_i} p_{ik} = 1$$

Note that by definition $c_{i0} = 0$, the capacity when unit i is on full outage and $c_{is_i} = c_i$, the rated capacity of unit i. For simplicity, let

$$p_i \triangleq p_{i,s_i} \qquad \text{and} \qquad q_i \triangleq p_{i,0}$$

The outage capacity of unit i is

$$\underset{\sim}{Z}_i = c_i - \sum_{k=0}^{s_i} c_{ik} \underset{\sim}{X}_{ik}$$

For a group of arbitrary generating units $1,2,\ldots,n$, we define $F_n(x)$ to be the c.d.f. of the outage capacity of units $1,2,\ldots,n$:

$$F_n(x) = P\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \le x\}$$

We next provide the details for developing an approximation to $F_n(x)$ based on a m.o.n.d. The joint generation state, or simply state, of a group of n units refers to a possible realization of the r.v.'s $\underset{\sim}{X}_{i,k}$, $i = 1,2,\ldots,n$; $k = 1,2,\ldots,s_i$. A state is characterized by a vector $\underline{k}^{(n)} = [k_1,k_2,\ldots,k_n]^T$. Here k_i corresponds to that index k for which $\underset{\sim}{X}_{i,k} = 1$, i.e.

$$k_i = \sum_{k=0}^{s_i} k \underset{\sim}{X}_{ik}$$

With each state we can associate an index:

$$e_n[\underline{k}^{(n)}] = \sum_{i=1}^{n} k_i$$

The maximum value $e_{n,max}$ of $e_n[\underline{k}^{(n)}]$ is:

$$e_{n,max} = \sum_{i=1}^{n} s_i$$

which corresponds to the state of no outages in any of the n units. Let $\mathcal{S}_n$ denote the state space, i.e., the set of all possible states, of the group of n units. Let $r > 0$ be some integer value. We decompose the set $\mathcal{S}_n$ into r+1 nonintersecting subsets. For $j = 0,1,\ldots,r-1$ we define subsets $\mathcal{B}_{n,j}$ of $\mathcal{S}_n$ of states characterized by an index whose value is $e_{n,max} - j$:

$$\mathcal{B}_{n,j} = \{\underline{k}^{(n)} \in \mathcal{S}_n \mid e_n[\underline{k}^{(n)}] = e_{n,max} - j\}$$

And we define

$$\mathcal{B}_{n,r} = \{\underline{k}^{(n)} \in \mathcal{S}_n \mid e_n[\underline{k}^{(n)}] \le e_{n,max} - r\}$$

A state $\underline{k}^{(n)}$ may belong to one and only one of the r+1 subsets $\mathcal{B}_{n,0}, \mathcal{B}_{n,1}, \ldots, \mathcal{B}_{n,r}$. It follows that

$$\mathcal{S}_n = \mathcal{B}_{n,0} \cup \mathcal{B}_{n,1} \cup \ldots \cup \mathcal{B}_{n,r}, \ \mathcal{B}_{n,i} \cap \mathcal{B}_{n,j} = \phi \quad i \ne j$$

We may evaluate $F_n(x)$ by conditioning over these r+1 subsets of $\mathcal{S}_n$ [9]:

$$F_n(x) = P\left\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \le x \,\middle|\, \underline{k}^{(n)} \in \mathcal{S}_n\right\} = \sum_{j=0}^{r} F_{n,j}(x)\, \pi_{n,j}$$

where

$$\pi_{n,j} \triangleq P\{\mathcal{B}_{n,j}\}$$

$$F_{n,j} \triangleq P\left\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \le x \,\middle|\, \underline{k}^{(n)} \in \mathcal{B}_{n,j}\right\}$$

We approximate the $F_{n,j}$ by normals so that we may approximate $F_n(x)$ by $F_n^a(x)$ where:

$$F_n^a(x) = \sum_{j=0}^{r} \Phi(x; \mu_{n,j}, \sigma_{n,j})\, \pi_{n,j} \tag{B-1}$$

with

$$\mu_{n,j} = E\left\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \,\middle|\, \underline{k}^{(n)} \in \mathcal{B}_{n,j}\right\}$$

$$\sigma_{n,j}^2 = \text{var}\left\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \,\middle|\, \underline{k}^{(n)} \in \mathcal{B}_{n,j}\right\}$$

The use of the approximation in Eq. (B-1) makes numerical convolution unnecessary for evaluating e.l.d.c's.

Instead, the outage parameters $\pi_{n,j}$, $\mu_{n,j}$, and $\sigma^2_{n,j}$ must be evaluated. The evaluation is performed recursively each time the value of n is increased/decreased due to the incorporation/withdrawal of another unit into the group of units being considered.

It is convenient for the development that follows to introduce some simplifying notation. We define:

$$s \triangleq s_{n+1} \qquad t \triangleq \min\{s,j\} \qquad v \triangleq \min\{s,r\}$$

The recursive evaluation of the parameters is based on the properties of the subsets $\mathcal{B}_{n,j}$ [9].

Recursive Evaluation of the $\pi_{n,j}$. Recall that $\pi_{n,j} = P\{\mathcal{B}_{n,j}\}$ is the probability associated with the subset in which all the states have an index value of $[e_{n,max} - j]$. The $\pi_{n,j}$ can be recursively evaluated using simple probabilistic arguments [9]:

$$\pi_{n+1,0} = \pi_{n,0}\ p_{n+1} \tag{B-2}$$

$$\pi_{n+1,j} = \sum_{\ell=0}^{t} \pi_{n,j-\ell}\ p_{n+1,s-\ell} \qquad 0 < j < r \tag{B-3}$$

$$\pi_{n+1,r} = \sum_{\ell=0}^{v} \pi_{n,r-\ell}\ (\sum_{k=0}^{s-\ell} p_{n+1,k}) \tag{B-4}$$

Recursive Evaluation of the $\mu_{n,j}$. The conditional expectation $\mu_{n,j} = E\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \mid \underline{k}^{(n)} \varepsilon \mathcal{B}_{n,j}\}$ gives the expected value of the outage capacity conditioned over the subset of states with the index $[e_{n,max} - j]$. The $\mu_{n,j}$ may be recursively evaluated [9]:

$$\mu_{n+1,0} = \mu_{n,0} = 0 \tag{B-5}$$

$$\mu_{n+1,j} = \sum_{\ell=0}^{t} (\mu_{n,j-\ell} + c_{n+1} - c_{n+1,s-\ell})\ \frac{\pi_{n,j-\ell}\ p_{n+1,s-\ell}}{\pi_{n+1,j}} \tag{B-6}$$

$$\mu_{n+1,r} = \sum_{\ell=0}^{v} \left\{ (\mu_{n,r-\ell} + c_{n+1}) \sum_{k=0}^{s-\ell} p_{n+1,k} - \sum_{k=0}^{s-\ell} c_{n+1,k}\ p_{n+1,k} \right\} \cdot \frac{\pi_{n,r-\ell}}{\pi_{n+1,r}} \tag{B-7}$$

Recursive Evaluation of the $\sigma^2_{n,j}$. The conditional variance $\sigma^2_{n,j} = \mathrm{var}\{\sum_{i=1}^{n} \underset{\sim}{Z}_i \mid \underline{k}^{(n)} \varepsilon \mathcal{B}_{n,j}\}$ gives the variance of the outage capacity conditioned over the subset of states with index $[e_{n,max} - j]$. The $\sigma_{n,j}$ may be recursively evaluated using basic probabilistic arguments [9]:

$$\sigma^2_{n+1,0} = \sigma^2_{n,0} = 0 = 0 \tag{B-8}$$

$$\sigma^2_{n+1,j} = \sum_{\ell=0}^{t} [\sigma^2_{n,j-\ell} + (\mu_{n,j-\ell} + c_{n+1} - c_{n+1,s-\ell})^2] \cdot \frac{\pi_{n,j-1}\ p_{n+1,s-\ell}}{\pi_{n+1,j}} - \mu^2_{n+1,j} \tag{B-9}$$

$$\sigma^2_{n+1,r} = \sum_{\ell=0}^{v} \left[\sum_{k=0}^{s-\ell} p_{n+1,k}\ \{\sigma^2_{n,r-\ell} + [\mu_{n,r-\ell} + c_{n+1} - c_{n+1,k}]^2\} \right] \cdot \frac{\pi_{n,r-\ell}}{\pi_{n+1,r}} - \mu^2_{n+1,r} \tag{B-10}$$

The recursive evaluation of the parameters $\pi_{n,j}$, $\mu_{n,j}$, and $\sigma^2_{n,j}$ in Eqs. (B-2)-(B-10) corresponds to the convolution operation in production simulation. For every additional unit loaded to meet the demand, the values of $\pi_{n,j}$, $\mu_{n,j}$, and $\sigma_{n,j}$ are computed. When production simulation involves the loading of multiple-block units it is necessary to compute the parameters $\pi_{n,j}$, $\mu_{n,j}$, $\sigma^2_{n,j}$ for both loading and unloading. This is so because when an additional block of a previously loaded unit is loaded in the simulation, the previously loaded portion of the unit is first removed. The unit is then loaded with a capacity equal to the sum of the capacities of the previously loaded block and the additional block. Unloading of a unit (or part of a unit) corresponds to deconvolution in the production simulation. In the m.o.n.a. technique the deconvolution operation is implemented by reevaluating the parameters $\pi_{n,j}$, $\mu_{n,j}$ and $\sigma^2_{n,j}$.

For deconvolution, we assume that the parameters representing the outage distribution $F_{n+1}(x)$ to the group of n+1 units 1,2,...,n+1 are known, and that the impact of the removal of unit n+1 is to be determined. The formulae used for updating the values of $\pi_{n,j}$, $\mu_{n,j}$ and $\sigma^2_{n,j}$ are obtained using the same probabilistic arguments as in the derivation of Eqs. (B-2)-(B-10) [9].

Deconvolution: Evaluation of the $\pi_{n,j}$.

$$\pi_{n,0} = \frac{\pi_{n+1,0}}{p_{n+1}} \tag{B-11}$$

$$\pi_{n,j} = \frac{\pi_{n+1,j} - \sum_{\ell=1}^{t} \pi_{n,j-\ell}\ p_{n+1,s-\ell}}{p_{n+1,s}} \qquad 0 < j < r \tag{B-12}$$

$$\pi_{n,r} = \pi_{n+1,r} - \sum_{\ell=1}^{v} \pi_{n,r-\ell}\ (\sum_{k=0}^{s-\ell} p_{n+1,k}) \tag{B-13}$$

Deconvolution: Evaluation of the $\mu_{n,j}$.

$$\mu_{n,0} = 0 \tag{B-14}$$

$$\mu_{n,j} = \frac{\pi_{n+1,j}\ \mu_{n+1,j} - \sum_{\ell=1}^{t} (\mu_{n,j-\ell} + c_{n+1} - c_{n+1,s-\ell})\ \pi_{n,j-\ell}\ p_{n+1,s-\ell}}{\pi_{n,j}\ p_{n+1}} \qquad 0 < j < r \tag{B-15}$$

$$\mu_{n,r} = \frac{\pi_{n+1,r}\ \mu_{n+1,r} - \sum_{\ell=1}^{v} \pi_{n,r-\ell}\ \{(\mu_{n,r-\ell} + c_{n+1}) \sum_{k=0}^{s-\ell} p_{n+1,k}}{\pi_{n,r}} - \frac{\sum_{k=0}^{s-\ell} c_{n+1,k}\ p_{n+1,k}\}}{\pi_{n,r}} - (c_{n+1} - \sum_{k=0}^{s} c_{n+1,k}\ p_{n+1,k}) \tag{B-16}$$

Deconvolution: Evaluation of the $\sigma^2_{n,j}$.

$$\sigma^2_{n,0} = 0 \tag{B-17}$$

$$\sigma^2_{n,j} = \frac{\pi_{n+1,j}(\sigma^2_{n+1,j} + \mu^2_{n+1,j}) - \sum_{\ell=1}^{t} [\sigma^2_{n,j-\ell} + (\mu_{n,j-\ell} + c_{n+1} - c_{n+1,s-\ell})^2]}{\pi_{n,j}\ p_{n+1}} \cdot \frac{\pi_{n,j-\ell}\ p_{n+1,s-\ell}}{\pi_{n,j}\ p_{n+1}} - \mu^2_{n,j} \qquad 0 < j < r \tag{B-18}$$

$$\sigma^2_{n,r} = -[\sum_{k=0}^{s} p_{n+1,k} (\mu_{n,r} + c_{n+1} - c_{n+1,k})^2] + \Big\{\pi_{n+1,r} (\sigma^2_{n+1,r} + \mu^2_{n+1,r})$$

$$- \sum_{\ell=1}^{v} \sum_{k=0}^{s-\ell} [\sigma^2_{n,r-\ell} + (\mu_{n,r-\ell} + c_{n+1} - c_{n+1,k})^2] p_{n+1,k} \pi_{n,r-\ell}\Big\} / \pi_{n,r} \quad \text{(B-19)}$$

Two-State Case. The recursive equations for convolution and deconvolution in the m.o.n.a. framework are considerably simplified if it is assumed that no unit can be in a partial outage state; that is, $s_i = 1$, $1 \leq i \leq N$. We use the following notation for this case:

$$p_i \triangleq p_{i,1} \quad \text{and} \quad q_i \triangleq p_{i,0}$$

A subset $\mathcal{B}_{n,j}$ of the state space $\mathcal{S}_n$ is the collection of states with the property that j units out of a group of n units have failed for $0 \leq j < r$. $\mathcal{B}_{n,r}$ is the collection of states such that r or more units have failed out of a group of n units. In this case the recursive relations of Eqs. (B-2)-(B-10) and Eqs. (B-11)-(B-19) reduce to those given in Tables I and II, respectively.

APPENDIX C: EQUIVALENT LOAD REPRESENTATION

The e.l.d.c. $\mathcal{L}_n(x)$ associated with the equivalent load, which is the sum of tbe system load and the outages of a group of n units, is defined to be

$$\mathcal{L}_n(x) = P\{\underset{\sim}{L} + \sum_{i=1}^{n} \underset{\sim}{Z}_i > x\}$$

$\mathcal{L}_n(x)$ may be computed by convolving the original l.d.c. with the density of the sum of the outage r.v.'s of the n units:

$$\mathcal{L}_n(x) = \mathcal{L}(x) * f_n(x)$$

where $f_n(x)$ is the density function of the outage capacity

$$f_n(x) = \frac{d\, F_n(x)}{d\, x}$$

In the m.o.n.a. technique framework, we approximate $\mathcal{L}_n(x)$ by

$$\mathcal{L}^a_n(x) = \mathcal{L}^a(x) * f^a_n(x)$$

where

$$f^a_n(x) = \sum_{j=0}^{r} \pi_{n,j}\, \phi(x;\, \mu_{n,j},\, \sigma_{n,j})$$

is the derivative of $F^a_n(x)$ as defined by Eq. (B-1). Property C of Appendix A implies that $\mathcal{L}^a_n(x)$ is a m.o.n.d. It can be shown that the weights of this m.o.n.d. are $\omega_k\, \pi_{n,j}$ The means are $\mu_k + \mu_{n,j}$ and the standard deviations and $(\sigma^2_k + \sigma^2_{n,j})^{\frac{1}{2}}$, $k = 1,2,\ldots,K$; $j = 0,1,\ldots,r$ [9]. Thus

$$\mathcal{L}^a_n(x) = 1 - \sum_{k=1}^{K} \sum_{j=0}^{r} \omega_k\, \pi_{n,j}\, \Phi(x;\, (\mu_k + \mu_{n,j}),\, [\sigma^2_k + \sigma^2_{n,j}]^{\frac{1}{2}})$$

Discussion

J. K. Delson (Electric Power Research Institute, Palo Alto, CA): An important feature of EGEAS, the Electric Generation Expansion Analysis System originally distributed by EPRI in 1983, is the efficiency of its production costing subroutine. The calculation utilizes the equivalent load duration curve, but instead of the original Baleriaux method, which is based on numerical integration, an analytical expression is formed using a series expansion based on statistical cumulants. This significantly reduces the calculation effort but also introduces inaccuracies. The shortcoming is mainly apparent in analysis of smaller power systems. The tests described in this paper show that MONA is almost as fast as the cumulant method but reduces its inaccuracies. MONA first sorts load data strictly by magnitude, forming, for example, three groups of bounded loads. Bell-shaped normal distributions are fitted to each group of load data. These normal distributions are then combined into a single analytical representation of the load. A problem may arise because the goodness of fit depends on the boundary levels selected for sorting the load data. A related problem is that each normal distribution, fitted to a defined range of loads, actually covers an infinite range. The tails of the distribution that extend on either side of the load boundaries are not only superfluous but also may degrade the final analytical expression for the complete load distribution. To help overcome these limitations, an interactive but somewhat time-consuming trial and error graphics program was designed by R. Taber Jenkins to facilitate selection of the load subdivisions [1]. A further advance was made through an automatic optimum search technique, based on minimization of error through a least squares fit, that was applied to this problem by Danny Sutanto and his coworkers [2]. In Sutanto's method complete normal distributions are used from the very beginning of the subdivision process. This method not only removes the arbitrariness in the division of the loads into groups but also assures that the tails of the individual distributions will not degrade the final curve fit. In summary, this paper demonstrates a technique, MONA, that offers an attractive improvement over the basic cumulant technique, particularly for smaller power systems. Developments have continued, and as illustrated by Sutanto's technique, MONA itself may also be ready to be supplanted.

References

[1] P. S. Hill and R. T. Jenkins, "Compensation for Cumulant Load Fit Discrepancies—A Computer Graphics Approach," *Proc. on Electric Generating System Expansion Analysis*, March 5–6, 1981, The Ohio State University, Columbus, OH.

[2] D. Sutanto, H. R. Outhred, and B. Manhire, "Improvements to Probabilistic Production Cost Calculation Using Cumulant Method. Part I: Cumulant Load Fit," *Proc. Chattanooga Conference on Production Simulation*, October 22–24, 1985. EPRI Report in publication.

Manuscript received February 17, 1987.

George Gross: I wish to thank Dr. Delson for his discussion which provides the reader with his view of some of the developments in the computation of e.l.d.c.'s after our work on m.o.n.a. was completed in 1984. We are in clear agreement with his comments on the shortcomings of the cumulant technique [9]. However, I wish to make two clarifications concerning his remarks on the m.o.n.a. technique. Our extensive tests show that the goodness-of-fit of the m.o.n.a. to the l.d.c. and the e.l.d.c.'s is relatively insensitive to the location of the boundary demarcations of the K classes of loads. These results are reported in [9]. Moreover, the goodness-of-fit is also insensitive to the number of classes K in the range $2 \leqslant K \leqslant 4$. Dr. Delson points out correctly that in the m.o.n.a. each normal approximating the distribution in the class extends over an infinite range. This, however, is not a problem except at the base load point. At this point, a small correction is made to ensure that the m.o.n.a. of the l.d.c. is exactly 1 for all values less than or equal to the base load. Judging from our extensive test results, and contrary to Dr. Delson's view, this correction in no way degrades the goodness-of-fit of the approximation to l.d.c.'s and e.l.d.c.'s.

We are pleased that our work on m.o.n.a. has acted as a stimulant for additional developments in the computation of e.l.d.c.'s as cited by Dr. Delson and additional references in Appendix D of [9]. Our results with the m.o.n.a. technique indicate, indeed, that it is a very robust and accurate technique. Readers may be interested to learn that, based on these results, the m.o.n.a. approach was used in the implementation of a major commercial microcomputer-based production costing package.

Manuscript received May 4, 1987.