Probabilistic LMP forecasting under AC optimal power flow framework: Theory and applications

Rui Bo\textsuperscript{a, *}, Fangxing Li\textsuperscript{b}

\textsuperscript{a} Midwest Independent Transmission System Operator (MISO), St. Paul, MN 55108, USA
\textsuperscript{b} Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA

\textbf{A R T I C L E   I N F O}

Article history:
Received 18 August 2011
Received in revised form 8 November 2011
Accepted 23 January 2012
Available online 9 March 2012

Keywords:
Power market
Optimal power flow
Critical load level
Locational marginal price
Probabilistic LMP
AC optimal power flow (ACOPF)

\textbf{A B S T R A C T}

This paper presents an analytical approach for studying the impact of load forecasting uncertainty on locational marginal price (LMP), in electricity wholesale market. The random nature of load brings in uncertainty in LMP through a market clearing model, also known as an optimal power flow (OPF) problem. This work is built on alternating current OPF, a close representation of the actual problem where power losses are accurately modeled. In the context of load being a normally distributed random variable, the probabilistic LMP concept under ACOPF is firstly examined to be a mixed random variable assuming both continuous and discrete values. Then, with the LMP versus load model, the probability density function (PDF) and cumulative density function (CDF) of probabilistic LMP are formulated and shown to be differentiable almost everywhere. The derived PDF formulation reveals an interesting fact that LMP does not follow a normal distribution even when load is normally distributed. Rather, its PDF presents a piece-wise partial normal distribution pattern due to the LMP step-change phenomenon. Further, the expected value of the probabilistic LMP and its sensitivity are derived. The proposed analytical approach can be useful for power market participants in evaluating and hedging financial risks associated with LMP uncertainty. The validity and effectiveness of the proposed method will be exemplified on a modified PJM 5-bus system and the IEEE 118-bus test system.

\textsuperscript{*} Corresponding author at: 1125 Energy Park Drive, St. Paul, MN 55108, USA.
Tel.: +1 651 632 8447; fax: +1 651 632 8417.
E-mail addresses: rui.bo@ieee.org (R. Bo), fl6@utk.edu (F. Li).

1. Introduction

In North America, Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) have been created to enhance power grid reliability and achieve greater generation dispatch efficiency through operating a geographically large power grid and organizing a competitive power market in each ISO/RTO footprint [1–4]. Presently, two-thirds of the electricity consumers in the United States and more than half of the Canada’s consumers are served by ten ISOs/RTOs [5]. In the organized power market, the ISO/RTO dispatches generation in order to maximize total social benefits or minimize total generation cost while respecting all physical constraints of the power grid and generators, and complying to reliability standards set by the North American Electric Reliability Corporation (NERC) and its Regional Reliability Councils.

The generation dispatch model is also called economic dispatch (ED), and is a specific type of the optimal power flow (OPF) problem [6]. Based on the form of the power flow model, OPF models can be categorized into the direct current OPF (DCOPF) and alternating current OPF (ACOPF). DCOPF model employs a linearized and highly simplified power flow model where reactive power and power losses are often ignored. In contrast, ACOPF model is built on a more accurate power flow model where both real and reactive power flows and their losses are taken into account. Both DCOPF and ACOPF have gained real-world applications [1–4,7]. Comparisons between the ACOPF and DCOPF models in terms of market price calculation are available in [8,9].

Locational marginal price (LMP) is a very important by-product of the OPF problem. It is the Lagrange multiplier associated with the equality constraints of the nodal real power balance [10]. LMP is used for billing settlement and serves as a crucial price signal for investment decisions.

The ACOPF problem is a deterministic programming model. However, the disturbance variables, for example, the loads, are unknowns at the time of solving the dispatch problem because there is a lead time. Amount of total system load can be estimated through load forecasting tools, but there are always discrepancies between the forecasted and the actual values due to the random nature of load. The load uncertainty subsequently brings in risks in the calculated LMP.

The situation is further complicated by the step-change phenomenon identified in the LMP versus load curve. LMP may exhibit drastic pattern changes such as step changes at certain load levels while it presents steady changes at other load levels. The load levels where significant change of LMP pattern occurs are termed critical
load levels (CLLs) [11]. Methods for efficient estimation of CLLs have been proposed in [12–14] for different models such as lossless DCOPF, with-loss DCOPF, and ACOPF, respectively. To facilitate the study and present results in a concise manner, bus-level loads are assumed to change proportionally to the total system load change from the present load level. Thus, we can focus on the total system load change instead of the change at each individual bus. Certainly, the bus-level load variation pattern can be defined differently.

Ref. [15] has analytically examined the impact of load forecasting uncertainty on nodal LMP. The study is carried out for a simplified version of economic dispatch, namely, lossless DCOPF, where power losses on the transmission system are ignored. Under the DCOPF framework, the LMP versus system load curve is a staircase curve, and LMP has a finite number of values. Therefore, the nodal LMP is a discrete random variable in the context of actual load being a random variable. The probabilistic LMP concept is proposed and the probability mass function is derived.

This paper extends the previous work in [15] to the more complicated ACOPF framework to study the impact of load forecasting uncertainty on nodal LMP. The ACOPF model contains a number of nonlinear constraints, which makes it very difficult, if not impossible, to get closed form representations of ACOPF solutions and the by-products, such as LMPs.

In the lossless DCOPF framework, the nodal LMP is constant between two adjacent CLLs, while in the ACOPF framework, the LMP will steadily increase or decrease when load varies within two adjacent CLLs [16]. Fig. 1 shows a typical LMP versus system load curve for a modified PJM 5-bus system under ACOPF. It can be seen that step changes still exist at a few load levels, which are the CLLs. The LMPs between any two adjacent CLLs suggest a linear pattern, as can be seen in the inset of Fig. 1. For example, LMPs at the five busses steadily increase when system load grows from 0 MW to 600 MW, followed by a step change at 600 MW. Note that in Fig. 1 the linear pattern is not quite perceivable without proper scaling for this small system. However, it can be very significant for larger systems with high nonlinearity due to higher resistance, reactance, and reactive power. It will be illustrated with the IEEE 118-bus system.

Consequently, the random variable, LMP at hour $t$, is no longer a discrete random variable; rather, it is a mixed random variable. Specifically, it is a combination of discrete random variable and piece-wise continuous random variable which results from the step-change phenomenon. Therefore, the associated probabilistic characteristics of this random variable such as probability density function and expected value will exhibit very different patterns than those of the LMP as a discrete variable presented in [15].

Numeric study on the probabilistic LMP with respect to load uncertainty has been presented in [17], where an efficient two-point estimate method rather than Monte Carlo Simulation method is applied to get the approximated distribution of LMP. In [17], load is assumed to follow normal distribution, and the authors conclude the output variables such as LMP tend to have the same probability distribution as the input variables such as load. However, at the same time the authors do notice conflicting results where normal distribution cannot fit well the actual LMP distribution due to ‘outliers’. This puzzle was left to be answered.

Analytical approach is therefore adopted in this work, which helps to solve the puzzle. In this paper, the closed form of the LMP distribution will be derived and help explain why LMP does not actually follow a normal distribution even when load is normally distributed. Rather, LMP follows a piece-wise partial normal distribution with each piece having its own mean and standard deviation. It implies outliers will always have a natural presence in the probability distribution. LMP will appear as a normal distribution only when load distribution has sufficiently small standard deviation and its mean is sufficiently away from any CLL. The systematic and analytical study on the probabilistic characteristics of LMP with respect to load forecasting uncertainty will be useful for power market participants in formulating their bidding strategies and developing risk hedging policies.

The paper is organized as follows. Section 2 formulates the probability density function (PDF) and cumulative density function (CDF) for the probabilistic LMP. The expected value of the probabilistic LMP and its sensitivity are then derived in Section 3. The proposed analytical approach will be fully exemplified and discussed on two test systems in Section 4. Section 5 concludes the paper.

2. Probabilistic LMP and its probability density function

Analytical approach requires explicit form of the distribution of load and the function that characterizes the relationship between load and LMP. Therefore, we will first establish a mathematical model for load distribution and the LMP versus the load curve. Then we can derive, analytically, the formulations for CDF, PDF, expected value of probabilistic LMP, etc.

2.1. Load distribution

For consistency, the same assumptions adopted in [15] on actual load will be taken in this work, namely, the actual load at hour $t$ is assumed to follow a normal distribution, which has been frequently employed to model the actual load in a number of research works [17–20]. Then, we have

$$D_t \sim N(\mu_t, \sigma_t^2)$$

(1)

$$\phi(x) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-(x-\mu_t)^2 / 2\sigma_t^2}$$

(2)

$$\Phi(x) = \int_{-\infty}^{x} \phi(u)du$$

(3)

where $D_t$ is a random variable representing the actual load at hour $t$; $N$ denotes the normal distribution; $\mu_t$ denotes the mean of $D_t$; $\sigma_t^2$ denotes the variance of $D_t$; $\phi(x)$ denotes probability density function of $D_t$; and $\Phi(x)$ denotes cumulative density function of $D_t$.

Note that it is a common practice to use a single forecasted value of load ($D_t^f$) to perform a deterministic economic dispatch and to forecast the LMP, congestion, etc. [6]. It should also be noted that load $D_t$ is referred to as the real load only since the reactive load can be derived based on constant load power factors.
2.2. Models for the LMP–load curve

By definition, the LMP is the partial derivative of the total generation cost with respect to the load change, and total generation cost is a linear combination of generation due to the adoption of the linear generation cost function. In addition, marginal unit generations are demonstrated to follow a nearly perfect quadratic pattern when load varies within two adjacent CLLS in the ACOPF framework [14]. Hence, roughly speaking, the LMP should follow a linear pattern between any two adjacent CLLS, as shown in Fig. 1. Interestingly, even seemingly high nonlinear LMP–load curve can be approximated by piece-wise linear curve with good accuracy [16].

Fig. 2 shows an illustrative picture of the LMP–load curve in the ACOPF framework. The load axis is divided into \( n - 1 \) segments by a sequence of critical load levels (CLLS), \( \{D_i\}_{i=1}^{n} \). Here, \( D_1 \) represents the no-load case (i.e., \( D_1 = 0 \)), and \( D_n \) represents the maximum load that the system can supply due to the limits of total generation resources and transmission capabilities. The actual LMP in each load segment \( i \) is considered to be a straight (linear) line, with slope \( a_i \) and intercept \( b_i \).

The LMP–load curve is extended to include two extra segments in Fig. 2. One is for the load from \( D_0 \) to \( D_1 \), where \( D_0 \) denotes a negative infinite load, and the associated price is zero. The second additional segment is defined as the load range from \( D_n \) to \( D_{n+1} \), where \( D_{n+1} \) represents a positive infinite load. In this segment, the price is set as the value of lost load (VOLL) to reflect demand response to load shedding. VOLL is assumed to be a constant value for simplicity in this work.

The two extra segments are added for mathematical completeness. In fact, they have a minimal, if any, impact on the study because the typical load range under study (for instance, from 0.8 p.u. to 1.2 p.u. of an average case load) is far from these two extreme segments, and the possibility of having the forecasted load close to zero or greater than \( D_n \) (the maximum load that the system can supply), is extremely rare and numerically 0.

The LMP–load curve can be formulated as,

\[
y(D) = \begin{cases} 
  a_0 \times D + b_0, & D_0 < D \leq D_1 \\
  a_1 \times D + b_1, & D_1 < D \leq D_2 \\
  \vdots & \\
  a_{n-1} \times D + b_{n-1}, & D_{n-1} < D \leq D_n \\
  a_n \times D + b_n, & D_n < D \leq D_{n+1} 
\end{cases}
\]  

where

\[
a_0 = b_0 = a_n = 0; \quad b_n = \text{VOLL} \\
D_0 = -\infty; \quad D_1 = 0; \quad D_{n+1} = \infty.
\]

The compact representation is given as follows:

\[
y(D) = \{a_i \times D + b_i | i \in \{0, 1, \ldots, n\}, \quad D_i < D \leq D_{i+1}\}.
\]

2.3. Probabilistic LMP

Here it is assumed that the economic dispatch and LMP calculations are performed on an hourly basis. The LMP at hour \( t \), denoted by \( \text{LMP}_t \), is a function of \( D_t \), which is a random variable from the viewpoint of forecasting. This is given by

\[
\text{LMP}_t = y(D_t).
\]

Therefore, at the forecasting or planning stage, LMP should also be viewed as a random variable. This characteristic is inherited from the use of forecasted load in OPF model. Fig. 2 shows the probability density function (PDF) of \( D_t \), overlapping with the LMP–load curve.

Three types of curve segments exist in terms of the value of the price slope \( a_i \). For example, for a curve segment with a positive price slope, the corresponding price range is \( [a_i \times D_i + b_i, a_i \times D_{i+1} + b_i] \), \( i \in \{0, 1, 2, \ldots, n\} \). In theory, the actual value of LMP could be any number in this price interval, and therefore, \( \text{LMP}_t \) assumes continuous values. For a curve segment with a negative price slope, the price range is \( [a_i \times D_i + b_i, a_i \times D_{i+1} + b_i] \), \( i \in \{0, 1, 2, \ldots, n\} \). For a curve segment with a zero price slope, the price will be a constant value \( b_i \), \( i \in \{0, 1, 2, \ldots, n\} \). Throughout the load interval, and therefore \( \text{LMP}_t \), assumes discrete values(s). Due to the step change phenomenon of the LMP–load curve, there may or may not exist intersections among these price intervals. Therefore, it can be inferred that \( \text{LMP}_t \) should be a mixed random variable assuming both discrete values and piece-wise continuous values.

For an arbitrary price \( p \), we can look up the LMP–load curve to locate the corresponding load level(s). If there is no corresponding load level, for instance, \( p \) is not in any of the price intervals, the probability density value associated with \( p \) will be zero. If the corresponding load level(s) does exist, the probability density value associated with \( p \) will depend on the distance from the corresponding load level to the mean value of \( D_t \), namely, \( \mu_t \). Intuitively, the shorter the distance, the higher the probability density value is. Furthermore, the probability density function is continuous on prices within any one of the price intervals. A special case is with the zero-slope curve segment(s), where LMP holds a constant value \( b_i \). A discrete probability function (i.e., probability mass function) should be defined. To facilitate the presentation by using one single probability function, we can assign an infinite value as the probability density value for price \( b_i \) of the curve segment. This also aligns with the fact that the CDF function has a step change at \( b_i \). From PDF perspective, the infinite value at certain points can be viewed as an impulse function which is often used in signal processing. Fig. 3 shows a schematic graph of a PDF curve of the mixed random variable \( \text{LMP}_t \). The vertical arrow represents the infinite probability density value, or, an impulse.
It should be pointed out that in Fig. 3 the probability distributions of the price intervals are amplified for illustration purposes. The price intervals are typically so narrow that they will be displayed as single vertical bars when the PDF curve is drawn for the entire price range. However, if only one price interval is shown and well scaled in the graph, the corresponding PDF should manifest a continuous curve over the interval, excluding the two end points of the interval, as illustrated in Fig. 3. This characteristic of the PDF curve will be exemplified in Section 4.

Fig. 3 shows an important feature of the probabilistic LMP under ACOCP:

The probabilistic LMP at a specific (mean) load level is not a single deterministic value. Instead, it is a mixed random variable assuming both discrete and piece-wise continuous values. The associated probability density function is a piece-wise continuous function and may contain infinite/impulse values at certain discrete price(s).

Note that each of the continuous curve segments may not be a symmetric distribution. More discussion on the probabilistic LMP will be provided later when its PDF is derived.

2.4. Cumulative density function of probabilistic LMP

In order to obtain the PDF formula of LMP,
1. Use probability theory, the CDF of LMP, and can be derived as follows:
   \[ F_{\text{LMP}}(p) = \Pr(\text{LMP} \leq p) = \Pr(y(D_l) \leq p) = \Pr(D_l \in \Omega) \]
   (7)

   Define \( y(D_l) = a_i \times D_l + b_i \), \( D_l < D_{i+1} \) and \( \Omega_l = \{ x | y(x) \leq p \} \), then we have
   \[ \Omega = \bigcup_{i=0}^{n} \Omega_l \]
   (8)

   \[ \Omega_i \cap \Omega_j = \emptyset, \forall i \neq j. \]
   (9)

Therefore, the CDF function can be further derived as
   \[ F_{\text{LMP}}(p) = \Pr(D_l \in \Omega) = \sum_{i=0}^{n} \Pr(D_l \in \Omega_l) = \sum_{i=0}^{n} \Pr(D_l \in \{ x | y(x) \leq p \}) \]

   \[ = \sum_{i=0}^{n} \Pr(D_l \in \{ x | a_i x + b_i \leq p, D_l < x \leq D_{i+1} \}). \]
   (10)

In order to calculate \( \Pr(D_l \in \{ x | a_i x + b_i \leq p, D_l < x \leq D_{i+1} \}) \), three cases need to be considered respectively, i.e., (a) \( a_i > 0 \), (b) \( a_i = 0 \), and (c) \( a_i = 0 \).

1. Case I: \( a_i > 0 \)
   \[ \Pr(D_l \in \{ x | a_i x + b_i \leq p, D_l < x \leq D_{i+1} \}) = \int_{D_l}^{D_{i+1}} \psi(u)du \]
   (11)

   where \( \tilde{D}_l = p = \frac{b_i}{a_i}, p \in [y(D_l), y(D_{i+1})] \)

2. Case II: \( a_i < 0 \)
   \[ \Pr(D_l \in \{ x | a_i x + b_i \leq p, D_l < x \leq D_{i+1} \}) = \int_{D_l}^{D_{i+1}} \psi(u)du \]
   (12)

   where \( \tilde{D}_l = p = \frac{b_i}{a_i}, p \in [y(D_{i+1}), y(D_l)] \)

(3) Case III: \( a_i = 0 \)
   \[ \Pr(D_l \in \{ x | a_i x + b_i \leq p, D_l < x \leq D_{i+1} \}) = \int_{D_l}^{D_{i+1}} \psi(u)du \]

   where \( \tilde{D}_l = \begin{cases} D_{i+1} & p = b_i \\ D_l & p < b_i \\ D_{i+1} & p > b_i \end{cases} \)

Summarizing the three cases, we have
   \[ F_{\text{LMP}}(p) = \sum_{i=0}^{n} \int_{D_l}^{D_{i+1}} \psi(u)du + \sum_{i=0}^{n} \int_{D_l}^{D_{i+1}} \psi(u)du \]
   (14)

Eq. (14) implies the CDF of the probabilistic LMP is a superposition of normal distribution CDFs with different mean and standard deviation. This will also be illustrated through examples in the numeric study section.

2.5. Probability density function of probabilistic LMP

To derive the formula for the PDF of LMP, we need to know the differentiability of the CDF function first. Analysis can show \( f_{\text{LMP}}(p) \) is differentiable almost everywhere, except for prices at the boundaries of each interval, namely, \( \{ y(D_l) \}^{n+1}_{l=0} \).

As previously mentioned, for easy illustration, infinite values or impulse function will be used to represent the derivative at those non-differentiable points. Then, the probability density function of the LMP can be derived as follows
   \[ f_{\text{LMP}}(p) = F_{\text{LMP}}(p) = \sum_{i=0}^{n} \frac{1}{a_i} \ \left\{ \begin{array}{ll} \frac{p-b_i}{a_i} & p \in [y(D_l), y(D_{i+1})] \\
         & \{ y(D_l) \}^{n+1}_{l=0} \end{array} \right. \]
   \[ + \sum_{i=0}^{n} \int_{D_l}^{D_{i+1}} \psi(u)du - \frac{1}{a_i} \ \left\{ \begin{array}{ll} \frac{p-b_i}{a_i} & p \in [y(D_l), y(D_{i+1})] \\
         & \{ y(D_l) \}^{n+1}_{l=0} \end{array} \right. \]

   \[ + \sum_{i=0}^{n} \int_{D_l}^{D_{i+1}} \psi(u)du + \sum_{i=0}^{n} \int_{D_l}^{D_{i+1}} \psi(u)du \]
   (15)

Specifically, we have
   \[ f_{\text{LMP}}(p) = \infty, \quad p = y(D_l), i \in \{ 0, 1, \ldots, n \} \]

Therefore, LMP follows a piece-wise partial normal distribution with each piece having its own mean and standard deviation, namely, each of the piece-wise continuous curves takes partial form of a normal distribution, as illustrated in Fig. 3. A special case is when load distribution has sufficiently small standard deviation and its mean is far from any LMP step change or pattern change. Then only one part in (15) will be dominant and other parts will be numerically trivial, and thus LMP will appear as a normal distribution.

It should be noted that, for normal distribution, the probability decreases when the random variable assumes values farther from the mean value. It is however not the case for probabilistic LMP. In addition, each piece of the continuous curves may not be a symmetric distribution as it is defined on the interval bounded by two
consecutive CLLs, not the entire domain. These characteristics have all been illustrated in Fig. 3.

3. Expected value of probabilistic LMP

3.1. Expected value of probabilistic LMP

By the conditional expectation theory [21], the expected value of LMP, is derived as

\[ E(LMP) = \sum_{i=0}^{n} \left[ E(LMP_i | D_i < D_t \leq D_{i+1}) \times Pr(D_i < D_t \leq D_{i+1}) \right] \]

\[ = \sum_{i=0}^{n} \left[ E((D_i) | D_i < D_t \leq D_{i+1}) \times Pr(D_i < D_t \leq D_{i+1}) \right] \]

\[ = \sum_{i=0}^{n} \left[ \left( a_i \times D_i + b_i \right) | D_i < D_t \leq D_{i+1} \right) \times Pr(D_i < D_t \leq D_{i+1}) \]

\[ = \sum_{i=0}^{n} \left[ a_i \times E(D_i | D_i < D_t \leq D_{i+1}) + b_i \times Pr(D_i < D_t \leq D_{i+1}) \right] \]  \hspace{1cm} (16)

By the definition of the conditional probability density function [21], the conditional density function of \( D_i \), given any event \( D_t < D_t \leq D_{i+1} \), is

\[ f_{D_i | D_t < D_t \leq D_{i+1}}(u) = \begin{cases} \frac{f(u)}{Pr(D_t < D_t \leq D_{i+1})}, & x \in (D_i, D_{i+1}) \\ 0, & x \notin (D_i, D_{i+1}) \end{cases} \] \hspace{1cm} (17)

Therefore, the expected value of \( D_i \) given any event \( D_t < D_t \leq D_{i+1} \) is

\[ E(D_i | D_t < D_t \leq D_{i+1}) = \int_{D_i}^{D_{i+1}} u \times f_{D_i | D_t < D_t \leq D_{i+1}}(u) \times du \]

\[ = \int_{D_i}^{D_{i+1}} u \times \frac{f(u)}{Pr(D_t < D_t \leq D_{i+1})} \times du \]

\[ = \frac{1}{Pr(D_t < D_t \leq D_{i+1})} \int_{D_i}^{D_{i+1}} u \times f(u) \times du \]

\[ = \frac{1}{Pr(D_t < D_t \leq D_{i+1})} \times \left\{ \int_{D_i}^{D_{i+1}} u \times f(u) \times du - \int_{D_i}^{D_{i+1}} \Phi(D_{i+1}) - \Phi(D_i) \right\} \] \hspace{1cm} (18)

Substituting (18) into (16) will give the expected value of LMP as follows

\[ E(LMP) = \sum_{i=0}^{n} \left[ \left( a_i \times \Sigma_i \left( \Phi(D_i) - \Phi(D_{i+1}) \right) \right) + b_i \times \left( \Phi(D_{i+1}) - \Phi(D_i) \right) \right] \]

\[ + \sum_{i=0}^{n} \left[ a_i \times \left( D_i - \mu_i \right) \times \Phi(D_i) \right] - \left( D_{i+1} - \mu_{i+1} \right) \times \Phi(D_{i+1})] \] \hspace{1cm} (19)

3.2. Sensitivity of expected value of probabilistic LMP

Define \( E_{LMP} (\mu_t, \sigma_t) \frac{\partial}{\partial \mu_t} E[LMP_t] \) and take partial derivative of \( E_{LMP} (\mu_t, \sigma_t) \) with respect to \( \mu_t \), and eventually we can get the sensitivity as follows

\[ \frac{\partial E_{LMP} (\mu_t, \sigma_t)}{\partial \mu_t} = \sum_{i=0}^{n} \left[ a_i \times (D_i - \mu_i) \times \left( \Phi(D_i) - \Phi(D_{i+1}) \right) \times \left( D_i - \mu_i \right) \times \phi(D_i) \right] \]

\[ + \sum_{i=0}^{n} \left[ a_i \times (\Phi(D_{i+1}) - \Phi(D_i)) + (a_i \times (D_i - \mu_i) + b_i) \times (\Phi(D_i) - \Phi(D_{i+1}) \right] \] \hspace{1cm} (20)

The lower and upper bound can be derived from (20) and will not be elaborated in this paper due to length limitation.

4. Numeric studies

4.1. Study on a modified PJM 5-bus system

In this section, a numeric study will be performed on the modified PJM 5-bus system. The original PJM 5-Bus system has 5 buses, 6 transmission lines, 5 generators and 3 loads [22]. In [12], slight modifications have been made to the original system in order to create more congestions and facilitate illustration. The corresponding base case settings for the system, including the generator offers and generator maximum output capability, are shown in Fig. 4. Generator minimum output capabilities are all assumed to be 0 MW. The transmission line reactances and flow limits are shown in Table 1.

The original test system does not have all the necessary parameters for ACOFP runs. Therefore, in this paper, all loads are assumed to have a 0.95 lagging power factor. The generators are assumed to have a reactive power limit of 150 MVar capacitive to 150 MVar inductive. This is selected so that the system has sufficient reactive power resources and system voltage profile is not a major concern. The R/X ratio is set at 10% for each transmission line. ACOFP is performed using MATPOWER simulation package [23]. Both active and reactive power related constraints are taken into account. To align with contemporary market practices, only active power production cost is considered. A detailed analysis on reactive power LMP can be found in [24].

To calculate the LMP–load curve as shown in Fig. 1, it is assumed for simplicity that the system load change is distributed to each bus load proportional to its base case load. For illustration purpose, it is assumed that the mean \( \mu_t \) of random variable \( D_t \) is equal to the forecasted load \( D^F_t \), and the standard deviation \( \sigma_t \) is taken as 5% of the mean \( \mu_t \) unless otherwise stated. The value of lost load (VOLL) is set to $2000/MWh.

4.1.1. Approximation of LMP–load curve

A linear polynomial curve-fitting is employed to approximate the actual LMP between every two adjacent CLLs, and the coefficients are used to establish the mathematical model for the LMP–load curve. Table 2 shows the curve-fitting coefficients for the LMP curves at all buses when load is within [0, 590] MW. It implies

| Line impedance and flow limits for the PJM 5-bus system. |
| --- | --- | --- | --- | --- | --- | --- |
| X(%) | 2.81 | 3.04 | 0.64 | 1.08 | 2.97 | 2.97 |
| Limit (MW) | 400 | 999 | 999 | 999 | 999 | 240 |

Fig. 4. The base case modified from the PJM 5-bus example.
that the LMPs at all busses, except Bus E, increase slightly, while the LMP at Bus E remains 10$/MWh for the entire load interval. With the linear polynomial coefficients obtained through curve-fitting, the mathematical representation of LMP–load curve is established, which is a piece-wise linear curve with step changes at the CLLs. This mathematical model of the curve is a very good approximation to the actual LMP–load curve. In fact, the curve produced by the mathematical formula looks almost identical to the actual curve shown in Fig. 1. The largest difference is less than 0.07$/MWh, approximately 0.7% of the lowest LMP, 10$/MWh. Therefore, the approximation curve is not redrawn in this paper.

4.1.2. Cumulative density function of probabilistic LMP

Fig. 5 shows the CDF curve of the probabilistic LMP at Bus B for forecasted load at 730 MW and 900 MW, respectively. The figure suggests the staircase pattern of the CDF curve. Combined with the LMP versus Load curve as shown in Fig. 1, it can be seen that the prices at which a step change occurs coincide with the price intervals near the forecasted load level for Bus B. The corresponding PDF values for these prices are expected to be higher than the PDF values of other prices, as to be verified in Section 4.1.3.

A careful study reveals that the majority of the step changes observed in the CDF curve are not really step changes. The inset in Fig. 5 redraws the same curves in a narrow range around 24.5$/MWh, where a step change appears. It can be seen that both CDF curves move smoothly from 23.95$/MWh to 24.02$/MWh. It is consistent with the theoretical part in Section 2, where the CDF function is shown to be differentiable almost everywhere except at the price boundaries of each interval of the LMP–load curve, namely, \( \{y(D_i)\}_{i=1}^n \). Nevertheless, the change of the CDF values happens in such narrow price intervals that it looks just like a step change when plotted for a broader range of prices.

It should be noted that the probabilistic LMP at Bus B assumes only continuous values, and therefore there is no actual step change in the CDF curve. For probabilistic LMP at Bus E, it assumes discrete values such as 10$/MWh and correspondingly its CDF curve will contain real step change at 10$/MWh.

4.1.3. Probability density function of probabilistic LMP

The PDF curve of the probabilistic LMP at Bus B is shown in Fig. 6 for the same two forecasted load levels. When the forecasted load is 730 MW, the probability density function of the probabilistic LMP is mainly scattered in three price intervals: 15.19–15.23$/MWh, 22.01–22.05$/MWh, and 23.95–24.02$/MWh which is the two price intervals with a high probability density for the forecasted load at 900 MW. The probability density is numerically trivial for almost everywhere else outside these price intervals. Furthermore, these price intervals are consistent with those where the CDF values have a jump, as seen in Fig. 5.

Note that probabilistic LMP at Bus B assumes only continuous values and no discrete values, and therefore there will be no infinite value or impulse presence in the PDF curve. The vertical bars in Fig. 6 are actually smooth curves which are not legible due to scaling. A zoomed-in graph is shown as the inset in Fig. 6. It can be seen from the inset that the probability density functions are continuous, and differentiable, curves in the 23.95–24.02$/MWh range.

It should be pointed out that piece-wise smooth curves look like vertical bars because in the modified PJM 5-bus system, power loss is very small and has very limited impact on LMP. As shown in Fig. 1, LMP changes slightly with load variation except at CLLs, and therefore each of the price intervals is very narrow. PDF is then so concentrated that it looks like vertical bars. It is however not always the case. For scenarios where losses play a significant role, especially in bigger systems, LMP intervals will have a much wider spread and the corresponding piece of PDF will be more easily recognized as a smooth curve.

Similar fact that PDF may contain multiple pieces of distributions has been observed in [17]. However, the smaller pieces have been viewed as ‘outliers’ in [17] where a single normal distribution is applied to fit the actual distribution. In this work, through analytical study, the distribution has been rigorously derived. The derived PDF formulation not only characterizes the ‘outliers’ as partial normal distribution, but also indicates the natural presence of them, namely, they are not real ‘outliers’.

It should also be noted that PDF will look like a single normal distribution if the mean of random variable load is sufficiently far (e.g., $5 \sigma$) away from any CLL and probabilistic LMP assumes continuous values.

Since the probability of any single price is zero for the probabilistic LMP at Bus B, it is more useful to divide the entire price range into a few intervals and investigate the probability of an actual LMP falling into each interval. The vertical bars observed in the PDF curves in Fig. 6 can be used to help make this classification. In practice, the categorization is at the discretion of the decision maker and can vary case by case.
The probability of the LMP, at Bus B falling into the selected price intervals is calculated and shown in Table 3 for the two forecasted load levels, 730 MW and 900 MW. The results discover the fact that the deterministic LMP with respect to \( D_t \) may or may not fall into the price interval with the highest probability. For example, when the forecasted load is 900 MW, the corresponding deterministic LMP is 24.01 $/MWh and its close neighborhood 23.9–24.1 $/MWh has the highest probability of 70.87%. However, the close neighborhood 22.0–22.1 $/MWh of the deterministic LMP 22.03 $/MWh for forecasted load 730 MW has only the second highest probability of 30.88%, less than the 39.42% probability for the price interval 23.9–24.1 $/MWh. It shows that the deterministic LMP and its close vicinity do not necessarily bear the largest probability.

Table 3 reveals the likelihood of realizing the forecasted LMP and its close vicinity, and therefore, can be very useful for buyers and sellers in developing bidding strategies and making financial decisions.

### 4.1.4. Expected value of probabilistic LMP

The expected value of the probabilistic LMP for the above case is compared with the deterministic LMP, \( \bar{y}(D_t) \), which is shown in Table 4. It shows that the expected value of probabilistic LMP may differ from the deterministic LMP at a specific forecasted load. The expected value of the probabilistic LMP versus the forecasted load curve can be plotted as well. The curve is very similar to the one obtained from lossless DCOPF model presented in [15], and therefore is not shown in this paper.

### 4.1.5. Impact of load forecasting accuracy

In this section, three different levels of standard deviation of load forecasting are examined, 5%, 3%, and 1%. Fig. 7 shows the probabilities of the random variable LMP, at Bus B falling into a few price ranges for these three levels of standard deviation when forecasted system load is 730 MW. It can be seen from Fig. 7 that the probability of realizing the actual price in the range of 22.0–22.1 $/MWh where the deterministic LMP 22.03 $/MWh falls in, increases considerably with a smaller standard deviation. This is reasonable because a more accurate load forecast should lead to less deviation in the forecasted price.

The system consists of 118 busses, 54 generators and 186 branches/transfomers. System total load is 4242 MW with 9966.2 MW total generation capacity. The detailed system data are available in [25]. OPF related data are obtained from [15]. In this case study, the same assumptions for the modified PJM 5-bus system are taken on distribution pattern of load variation, standard deviation and VOLL.

For better illustration, the LMPs at only a few selected busses are shown in Fig. 8. In fact, the same set of busses as in [15] is selected for comparison purpose. In the base case, system load is 1.0 p.u., i.e., 4242 MW real load (with 1438 MVar reactive load), system total real power losses are 368.63 MW (8.7% of system load) and total reactive power losses are 2406.22 MVar.

It should be noted that the deterministic LMP–load curve in Fig. 8 appears significantly different from its counterpart obtained from lossless DCOPF model presented in [15]. This roots in the modeling difference between DCOPF and ACOPF (such as power losses and reactive power) and the resulting differences in generation dispatches and congestions. For instance, in the ACOPF case study, up to two congested branches are observed while there are a few more congested lines in the lossless DCOPF case study in [15].

In Fig. 8, LMP–load curve seems very complex, demonstrating several nonlinear patterns including step change pattern. The step change pattern can be observed from the inset of the figure.

Piece-wise linear polynomial curve-fitting approach is applied to approximate the actual curves in order to get a mathematical representation of the LMP–load curve for the analytical study on
probabilistic LMP. The presence of the nonlinear patterns seems to be challenging for a modeling approach using linear polynomial. Fortunately, CLLs have offered great assistance in the modeling by appropriately identifying the break points of the curve. The resulting piece-wise linear polynomial curve has been verified to be an exceptionally accurate model to represent the deterministic LMP–load curve. Various patterns of the curve have been captured and closely represented by the piece-wise linear approximation model. A detailed study on the validity of the piece-wise linear approximation approach has been carried out in [16]. The piece-wise linear approximation curve looks almost identical to the deterministic LMP–load curve, and therefore is not redrawn here.

Fig. 9 shows the expected value of probabilistic LMP versus forecasted load curve at the selected busses. It captures the trends of its deterministic counterpart as in Fig. 8 and rebuilds the curve in a much smoother way. The inset of Fig. 9 shows that the step changes in the deterministic LMP–load curve have been smoothed out. The expected value of probabilistic LMP skyrocket when system load grows to 7000 MW. This is because the load level is close to, actually within 3 × σl, vicinity of, the maximum load the system can serve and therefore the expected value is affected by the value of lost load (VOLL), which is set at 2000 $/MWh.

5. Discussions and conclusions

This paper extends the methodology for studying probabilistic LMP in [15] to a much more complicated framework, the ACOPF framework. First, due to the step change phenomenon of LMP versus load curve and the nonlinearity introduced by power losses and reactive power component of the ACOPF model, the probabilistic LMP, i.e., LMP at hour t, has been examined to be a mixed random variable assuming both discrete and piece-wise continuous values. Its probability distribution has piece-wise continuous part and may contain infinite/impulse values at certain price(s). In addition, based on existing work, the LMP–load curve is modeled as a piece-wise linear curve with critical load levels as breakpoints. Second, with the assumption that load follows a normal distribution, the probability density function (PDF) and cumulative density function (CDF) of probabilistic LMP have been formulated. The PDF and CDF are shown to be differentiable at almost everywhere except for the prices at the CLLs, namely, the boundary prices of each segment of the LMP–load curve. The distribution can be viewed as superposition of partial normal distributions with different mean and variance. This explains the observations in existing literature such as [17] that a single normal distribution does not fit well the actual distribution of probabilistic LMP.

Third, the formula for expected value of probabilistic LMP and its sensitivity are derived for efficient computation. The above analytics are exemplified and validated on two test systems. A few observations are listed as follows:

• The probabilistic LMP follows a piece-wise partial normal distribution with each piece having its own mean and standard deviation.
• Deterministic LMP and its close vicinity do not necessarily bear the largest probability. Meanwhile, the probability increases considerably as standard deviation becomes smaller.
• Load forecasting accuracy affects the expected value of probabilistic LMP much more significantly when forecasted load is close to a CLL.

Furthermore, when compared with the results from lossless DCOPF as shown in [15], the probabilistic LMP under ACOPF framework differs in the following aspects:

• Prices between two adjacent CLLs in the LMP–load curve can be modeled by a linear polynomial, instead of a constant value.
• The probabilistic LMP in the ACOPF framework is a mixed random variable, rather than a discrete random variable.
• The cumulative density function and probability density function are shown to be differentiable at almost everywhere except for the prices at the CLLs.
• The probability associated with a given price may not be meaningful due to the probabilistic LMP being a mixed random variable which can assume continuous values. Instead, the probability should be able to reflect the likelihood of the actual price falling into a range of prices.

Future work includes adapting the proposed analytical approach to evaluating the combined uncertainty associated with both demand and intermittent generation.

Disclaimer

The views expressed in this paper are solely of the authors and do not necessarily represent those of the MISO.

Acknowledgements

The authors would like to thank the U.S. National Science Foundation (NSF) award ECCS-1001999, the NSF/Department of Energy (DOE) Engineering Research Center program award EEC-1041877, and the CURENT Industry Partnership program for financial and facility support to complete this research.

References