Forecasting Real-Time Locational Marginal Price: A State Space Approach

Yuting Ji, Jinsub Kim, Robert J. Thomas and Lang Tong
School of Electrical and Computer Engineering Cornell University, Ithaca, NY, USA.
Email: {yj246, jk752, rjt1, lt235}@cornell.edu

Abstract—The problem of forecasting the real-time locational marginal price (LMP) by a system operator is considered. A new probabilistic forecasting framework is developed based on a time in-homogeneous Markov chain representation of the real-time LMP calculation. By incorporating real-time measurements and forecasts, the proposed forecasting algorithm generates the posterior probability distribution of future locational marginal prices with forecast horizons of 6-8 hours. Such a short-term forecast provides actionable information for market participants and system operators. A Monte Carlo technique is used to estimate the posterior transition probabilities of the Markov chain, and the real-time LMP forecast is computed by the product of the estimated transition matrices. The proposed forecasting algorithm is tested on the PJM 5-bus system. Simulations show marked improvements over benchmark techniques.

Index Terms—Locational marginal price (LMP), electricity price forecasting, probabilistic forecasting, Monte Carlo techniques, Incremental optimal power flow.

I. INTRODUCTION

The problem of forecasting locational marginal price (LMP) in a deregulated electricity market is important to both system operators and market participants. Accurate LMP forecast produced in real-time are essential for demand response, revenue and risk management, and an efficient operation of a smart grid. To this end, probabilistic forecast techniques that provide the posterior distribution of the future prices are most desirable.

The forecast of electricity price is challenging. Typically, only day-ahead price forecasts are produced using historical data; probabilistic forecasts are not available in general. Applying conventional techniques such as artificial neural networks (ANNs) or support vector machines (SVMs) to forecast the day-ahead (and in real time) LMP typically generates the mean absolute prediction error (MAPE) around around 10-20% [1]. Such inaccuracy is inadequate for making critical operation decisions. As a comparison, when these techniques are applied to load forecast, the MAPEs are almost an order of magnitude lower at approximately 1-2%.

The main challenge of price forecast arises from the fact that the price of electricity is location dependent, and the determination of LMP involves the optimization of the power flow subject to various generation and transmission constraints. As a result, LMP can be highly volatile with seemingly unpredictable price spikes.

Recently, the problem of real-time LMP forecast has been considered by major system operators, motivated by the fact that short-term forecasts of the real-time LMP are potentially beneficial to large consumers for making effective demand response decisions. Such forecasts can also play an important role in congestion relief, thus reducing the overall system operating cost. An example of such an operator provided real-time price forecast is that by the Electric Reliability Council of Texas (ERCOT) [2]. Specifically, ERCOT provides the one-hour ahead real-time forecasting of electricity prices with five-minute interval updates.

The system operator has access to a wealth of information critical to determining LMPs in real-time. This includes forecasts of loads and generators, line conditions, and most importantly, real-time state estimates that are used in the LMP calculation. Thus a significant improvement of forecasting accuracy can potentially be realized. Unfortunately, there has not been a formal development of forecasting methodologies that take advantage such inside information.

A. Summary of Contributions

The main contribution of this paper is a first attempt to develop a formal LMP forecasting methodology used by a system operator with the goal of facilitating actionable information for the system operator and market participants. To this end, we adopt a probabilistic forecasting framework aimed at providing the posterior distribution of future LMPs over prediction horizons up to 6-8 hours. Such prediction horizons are sufficiently ahead for making operating decisions and sufficiently short to take advantage of real-time measurements and forecasts available internally at the operator’s energy management system.

The key of any forecast technique is a model that captures dependencies of future variables on available information. To this end, we propose a time in-homogeneous Markov chain model for the LMP calculation, which is derived based on a stylized implementation of the PJM real-time market operation [3]. In particular, we define the LMP state as the (ex-ante) congestion pattern and (ex-post) eligible generators. By modeling the LMP states as a time in-homogeneous Markov chain, the dependencies of future LMP states (thus future LMPs) on current system operating conditions are captured by a sequence of transition probabilities that can be estimated using Monte Carlo techniques and, in practice, updated using real-time measurements.

This work is supported in part by the National Science Foundation under Grant CNS-1135844 and a Grant from DoE Consortium for Electric Reliability Technology Solutions (CERTS).
The use of Markov chain is crucial in providing probabilistic forecasts. While real-time LMPs, strictly speaking, are not Markovian, the LMP state is Markovian under certain idealistic conditions. In our simulation, the Markov model seems to be sufficiently accurate to provide substantial improvement of forecast accuracy.

B. Related Work

Electricity price forecasting has been a long standing and challenging problem. A variety of forecasting techniques have been considered, including the applications of time series models, machine learning algorithms, simulation tools, and game theory strategies (refer to [1] and the reference therein).

The main difference between existing approaches and that presented here is twofold. First, we consider the LMP forecast problem from the perspective of an operator. This allows us to exploit an explicit connection between the real-time LMP and the power system state space. Specifically, it is shown in [4] that the real-time LMP calculation partitions the power system state space into price regions. The system operating point can be seen as a random walk driven by demands and supplies. When the operating point passes across boundaries of these price regions, the real-time LMP transits from one price to another. This geometric characterization of the real-time LMP has led to the Markov chain model for the LMP calculation. We note that the use of Markov chain for the short-term price forecasting has been considered by Halilcevic and Gubina [5] in a very different setting.

Second, our method falls in the category of probabilistic forecasting, i.e., it provides an estimate of the posterior price distribution in addition to the point price forecast. Most existing works on probabilistic price forecasting focus on providing prediction intervals [6]–[8]. Although prediction intervals indicate the accuracy of price forecasts, price distributions are much more useful for planning purposes. A probabilistic forecasting method for the day-ahead LMP is proposed in [9], which estimates the future price distribution based on the forecasted load distribution. However, it is not directly applicable to the real-time LMP forecasting because the real-time LMP depends not only on the load profile but also on the system state trajectory.

II. REAL-TIME LMP MODEL

The PJM ex-post real-time LMP model [3] is adopted in this paper. As illustrated in Fig. 1, the model consists of four main components: State Estimation, Security Constrained Economic Dispatch, LMP Preprocessor, and LMP Algorithm. In the following, we describe the key role of each component and the real-time LMP calculation process.

1) State Estimation: The state estimator uses real-time measurements to infer the current power system operating condition. Specifically, the state estimator receives the accurate model of power grid and the real-time meter measurement $z_t$ as inputs and calculates the estimate $\hat{x}_t$ of current system state $x_t$, which is defined as the vector of bus voltage magnitudes and phase angles at time $t$.

2) Security Constrained Economic Dispatch (SCED): Given the resulting state estimate $\hat{x}_t$ from the state estimator, SCED also receives the load forecast $d_{t+1|t}$ for the next five-minute interval, and the list of online available generators with the full set of offer data from the system operator. With these inputs, SCED solves a DC optimal power flow (OPF) problem to find an optimal economic generation adjustment that meets the demand forecast $d_{t+1|t}$ and satisfies generation, transmission, and ramping constraints.

In the context of the real-time LMP calculation, SCED produces two important outputs: the ex-ante dispatch signal $g_{t+1|t}$, which is the optimal generation level of all generators for the next five-minute interval (i.e., the solution of the DC OPF), and the predicted congestion pattern $C_{t+1|t}$, which is the set of congested lines during the next five-minute interval when all generators follow the dispatch instruction.

3) LMP Preprocessor (LMPP): The role of LMPP is to determine the set of generators that are eligible to participate in the real-time LMP calculation after the delivery of energy at time $t+1$, denoted by $G_{t+1}$. Generators are considered eligible if and only if they are marginal generators in the ex-ante dispatch signal $g_{t+1|t}$ and follow the instruction during the last five-minute interval. The criterion of eligibility is that the ex-post estimated generation $\hat{g}_{t+1}$, which is computed based on the state estimate $\hat{x}_{t+1}$ at time $t+1$, does not exceed $110\%$ of the ex-ante dispatch $g_{t+1|t}$. Such a criterion is intended to reward the generator that follows the dispatch signal in the sense that the eligible generator is paid at actual output level while the ineligible generator is only paid the portion in the ex-ante dispatch.

4) LMP Algorithm (LMPA): LMPA is the last component of the real-time LMP calculation process. Given the set of eligible generators $G_{t+1}$ from LMPP and the ex-ante congestion pattern $C_{t+1|t}$ from SCED, LMPA solves the following incremental DC OPF:
minimize \[ \sum_{i \in G_{t+1}} c_i \Delta g_i, \]
subject to \[ \sum_{i} \Delta g_i = 0, \quad (\lambda) \]
\[ \Delta g_{\text{min}} \leq \Delta g_i \leq \Delta g_{\text{max}}, \quad i \in G_{t+1} \]
\[ \sum_{i \in G_{t+1}} A_{ik} \Delta g_i \leq 0, \quad k \in C_{t+1|t} \quad (\mu_k) \]

where
\[ \Delta g_i \] change in power output for generator \( i \);
\( c_i \) real-time offer for generator \( i \);
\( \Delta g_{\text{max}} \) upper bound of change for generator \( i \);
\( \Delta g_{\text{min}} \) lower bound of change for generator \( i \);
\( A_{ik} \) matrix of shift factors for generation bus \( i \) on the binding transmission constraint \( k \);
\( \lambda \) Lagrange multiplier of the power balance constraint;
\( \mu_k \) Lagrange multiplier of transmission constraint \( k \).

It should be noted that variables of (1) are hypothetical adjustments of eligible generators since this ex-post pricing happens after the delivery of energy. The Lagrange multiplier of the power balance constraint in (1) gives the marginal energy price \( \lambda \) for generation at the reference bus and that of each congestion constraint corresponds to the shadow price \( \mu_k \) for each congested line \( k \in C_{t+1|t} \). Based on the envelop theorem, the real-time LMP at a bus \( i \) is defined as the sum of the marginal energy price and shadow prices associated with all congested lines\(^1\):

\[ \pi_{t+1}(i) = \lambda + \sum_{k \in C_{t+1|t}} A_{ik} \mu_k. \quad (2) \]

Therefore, the real-time LMP vector \( \pi_{t+1} \) can be written as

\[ \pi_{t+1} = \lambda \cdot 1 + A\mu, \quad (3) \]

where \( 1 \) is the vector of ones, and \( \mu \) is the vector of all shadow prices associated with the lines in \( C_{t+1|t} \).

### III. Real-time LMP Forecasting Algorithm

In this section, we first introduce the real-time LMP state and model the LMP state process as a discrete Markov chain. Then, we present the forecasting algorithm based on the Markov model, the LMP pricing mechanism, and a Monte Carlo method.

**A. Real-time LMP state and Markov chain model**

According to (1), the LMP \( \pi_t \) at time \( t \) is fully determined by two real-time inputs\(^2\): the set of eligible generators \( G_t \) and the set of ex-ante congested lines \( C_{t-1|t} \).

We define the pair \( S_t \triangleq (G_t, C_{t-1|t}) \) as the real-time LMP state at time \( t \). By inherent discreteness of eligible generators and congestion patterns, the LMP state is also discrete. The collection of all possible pairs of an eligible generator set and a congestion pattern forms the LMP state space, denoted as \( \Omega \).

\(^1\)In this paper, we ignore losses and hence the cost of marginal loss in LMP is set to zero.

\(^2\)We assume that the system topology, generation adjustment bounds and the cost are fixed in the five-minute interval.

In our approach, we predict the real-time LMP \( \pi_t \) via forecasting the state \( S_t \). There are important advantages of working with \( S_t \) rather than \( \pi_t \): the correlation of \( S_t \) across time can be well estimated with the help of load forecast, the knowledge of previous system operating conditions, and network parameters. In contrast, the correlation structure of price series is generally arbitrary and much harder to estimate.

With the concept of LMP states, modeling the real-time LMP state process \( (S_t) \) as a discrete Markov chain becomes natural. The Markov chain model is reasonable because eligible generators and congestion pattern at time \( t \) tend to depend on their recent values. In particular, we model \( (S_t : t = 0, \ldots, T) \), where \( T \) is the prediction horizon, as a first-order time-in-homogeneous Markov chain. The time in-homogeneity is meant to reflect the fact that the state transition depends on time-varying exogenous variables (e.g., load profiles, system operating conditions). Our approach can be generalized for a higher-order Markov chain. The use of a first-order model is purely motivated by the computational benefit. The Markov chain is defined by following elements:

- the state space: \( \Omega \);
- the initial state probability (at time 0): \( \delta_{s_0} \in \mathbb{R}^{\left| \Omega \right|} \);
- the set of transition probability matrices:

\[ Q \triangleq \{Q_{t,t+1}, t = 0, \ldots, T-1\}, \quad (4) \]

where \( T \) is the prediction horizon, and

\[ Q_{t,t+1}(i,j) = \Pr(S_{t+1} = j|S_t = i). \quad (5) \]

At time 0, our forecasting algorithm begins by observing the realization \( s_0 \) of \( S_0 \), and thus the initial state probability is set as \( \delta_{s_0} \) with the probability mass function \( \delta_{s_0}(s_0) = 1 \).

**B. Real-time LMP Forecasting Algorithm**

In order to forecast the future LMP distribution, it is necessary to estimate transition matrices of the Markov model. We employ a Monte Carlo method to infer transition matrices for each time interval in the prediction horizon.

1) **Stochastic model for sample path generation:** The first step of the Monte Carlo method is to define a stochastic model to generate multiple sample paths of the Markov chain. Given a power system, generation and transmission constraints, and the real-time LMP model, we only need to generate an actual load profile and its forecast. All other sample paths can be computed by injecting the generated load data into the LMP model.

We use a first order auto-regressive noise model and a given load profile to generate sample paths of the actual load profile and the five-minute ahead load forecast. Specifically, at each time \( t > 1 \), the actual load \( d_t \), and the five-minute ahead load forecast \( \hat{d}_{t+1|t} \) are generated from the following model:

\[ y_t = \theta y_{t-1} + \epsilon_t, \quad \theta \in (0, 1), \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \]

\[ d_t = d_t + y_t, \]

\[ \hat{d}_{t+1|t} = \hat{d}_{t+1} + \theta y_t. \quad (6) \]
This model is based on a given load profile \((d_t : t = 0, \cdots, T)\) which is considered as the mean value of the actual load profile. In practice, we can use the day-ahead load forecast. The actual load profile is generated by adding the mean value with a first order auto-regressive noise series \((y_t : t = 0, \cdots, T)\). The initial values of \(d_0\) and \(y_0\) are defined as: \(d_0 \equiv d_0 + \varepsilon_0\) and \(y_0 \equiv \varepsilon_0\). The five-minute ahead load forecast \(\hat{d}_{t+1}\) to be used by SCED is generated from \(\hat{d}_{t+1}\) and the estimated noise \(\hat{y}_{t+1} = \theta y_t\).

Given a load sample path \((d_t : t = 0, \cdots, T)\) and its up-to-date forecast \((\hat{d}_{t+1})_{t=0}^{T-1}\), the corresponding sample paths of actual power system \((x_t : t = 0, \cdots, T)\), ex-ante power system state \((x_{t+1}^t : t = 0, \cdots, T - 1)\) and ex-post state estimate \((\hat{x}_t : t = 0, \cdots, T)\) can be generated as follows. For each time \(t\), we simulate market operations and energy deliveries by using the actual load \(d_t\) to solve the DC OPF. The optimization result is treated as the actual power system state \(x_t\). To model the measurement noise \(z_t = x_t + \varepsilon_t\), we use Gaussian model \(\varepsilon_t \sim \mathcal{N}(0, \sigma^2)\) as the measurement noise. Given the measurement \(z_t\), the sample path of the state estimate \(\hat{x}_t\) is simply by applying the traditional state estimation\(^3\). With these two sample paths, the ex-ante state \(x_{t+1}^t\) is generated by SCED which uses the estimated current state condition \(\hat{x}_t\) as the starting point and finds the optimal dispatch to meet the five-minute ahead load forecast \(d_{t+1}\) subject to constraints of generation, transmission and ramping.

Once \((x_{t+1}^t : t = 0, \cdots, T - 1)\) and \((\hat{x}_t : t = 0, \cdots, T)\) are available, generating sample paths of the LMP state is trivial. The sample path of \((C_{t+1} : t = 0, \cdots, T - 1)\) is determined by the ex-ante dispatch which can be obtained from \((x_{t+1}^t : t = 0, \cdots, T - 1)\), while the sample path of eligible generators\(^4\) \((G_i : t = 0, \cdots, T)\) is generated by comparing the difference between the ex-ante dispatch and the ex-post generation estimate based on \((\hat{x}_t : t = 0, \cdots, T)\). These two sample paths form a sample path of the price state \((S_t : t = 0, \cdots, T)\).

The forecasting algorithm repeats the above steps \(N\) times (Monte Carlo runs) to collect \(N\) sample paths of \((S_t : t = 0, \cdots, T)\).

2) State transition matrix estimation and probabilistic forecasting: Given all \(N\) sample paths of \((S_t : t = 0, \cdots, T)\), we can redefine \(\Omega\) as the set of real-time LMP states that ever appear in any of sample paths at any time. This step reduces the cardinality of \(\Omega\) significantly because there are only a few lines that are likely to congest. With the price state space and all sample paths, we can obtain estimates of state transition probability matrices, \(\{Q_{i,t+1} : t = 0, \cdots, T - 1\}\). Given estimates of state transition matrices, the probability mass function \(P_T\) of \(\pi_T\) is given by

\[
P_T = \delta_{\pi_0} \cdot Q_0 \cdot Q_1 \cdot \cdots \cdot Q_{T-1} \cdot \pi_T.
\]

\(^3\)The traditional state estimation (which estimates the system state only based on the measurement and underlying mathematical equations) should not be confused with the state estimation module in the real-time LMP model.

\(^4\)Initial eligible generators \(G_0\) are assumed to be the marginal generators at time 0 in the actual system state \(x_0\).

IV. Numerical Results

We present some simulation results on the modified PJM 5-bus system illustrated in Fig. 2 to compare the proposed forecasting algorithm with benchmark techniques.

A. Simulation system settings

1) System configuration: As shown in Fig. 2, generation capacities and their marginal costs (assumed to be constant) are posted in the box beside each generator. Line 1-2 and line 4-5 are assumed to have limits, which are 400MW and 240MW, respectively. Other detailed system parameters can be found in [10].

2) Data: We use ISO New England [11] historical five-minute load profile data from three independent days as the mean value of 24-hour load profiles on bus 2, 3 and 4, respectively. To make the load profile feasible in the system, we re-scale them into range [250MW, 400MW]. We only show results of the most representative case where the load base is [325MW, 375MW] in the following. As short term load prediction algorithms with mean absolute percentage value (MAPE) below 3% available [12], we set AR parameters \(\theta\) and \(\sigma\) to achieve a reasonable load prediction error level (0.56% in MAPE for the one-hour ahead forecast).

3) Toolbox: MATPOWER [13] is used to solve ex-ante dispatch and incremental DC-OPF problems in the simulation.

B. Probabilistic price forecasting and benchmarks

We implement the proposed probabilistic LMP forecasting algorithm described in Section III with 6 hours ahead prediction horizon. Since the forecasting performance is dependent on locations, we provide detailed prediction results for each bus. To evaluate the performance of the proposed algorithm, we use the following two point price predictions at a specific bus \(i\):

- Maximum a posteriori probability (MAP) prediction:
  \[
  \hat{\pi}^{\text{MAP}}(i) = \pi^k(i),
  \]
  where \(k = \arg\max_{1 \leq k \leq |\Omega|} P_T^k(i)\).

- Minimum Absolute Error (MAE) prediction:
  \[
  \hat{\pi}^{\text{MAE}}(i) = \pi^k(i),
  \]
  where \(k = \arg\min_{1 \leq k \leq |\Omega|} \mathbb{E}[|\pi(i) - \pi^k(i)|]\).
TABLE I
RESULTS OF DIFFERENT PRICE FORECASTING ALGORITHMS

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( \pi_E )</th>
<th>( \hat{\pi}_{MAE} )</th>
<th>( \hat{\pi}_{MAE} )</th>
<th>( \pi_{NN} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus1</td>
<td>24.50</td>
<td>22.97</td>
<td>16.30</td>
<td>33.69</td>
</tr>
<tr>
<td>Bus2</td>
<td>6.42</td>
<td>6.68</td>
<td>8.87</td>
<td>8.21</td>
</tr>
<tr>
<td>Bus3</td>
<td>3.19</td>
<td>3.68</td>
<td>5.06</td>
<td>2.51</td>
</tr>
<tr>
<td>Bus4</td>
<td>24.80</td>
<td>23.15</td>
<td>17.67</td>
<td>37.43</td>
</tr>
<tr>
<td>Bus5</td>
<td>14.12</td>
<td>14.37</td>
<td>10.88</td>
<td>21.31</td>
</tr>
<tr>
<td>Average</td>
<td>14.61</td>
<td>14.25</td>
<td>11.75</td>
<td>20.63</td>
</tr>
</tbody>
</table>

assuming that \( \Omega = \{ \pi^1, \ldots, \pi^{[Q]} \} \) is the reduced price state space which is defined in Section III-B2.

A deterministic baseline benchmark is implemented on the base of the real-time LMP model without Markov chain construction in the prediction horizon. In particular, this approach uses the deterministic mean value of load profile as the one-step ahead load prediction, i.e., \( \hat{d}_{t+1}|_t = d_{t+1} \) to compute the price trajectory as the LMP forecast.

ANN is implemented as a representative benchmark of the state-of-the-art technique to be compared with the proposed forecasting algorithm and the basic deterministic prediction approach. The architecture of the neural network consists of an input layer, two hidden layers and an output layer. We train a separate neural network at each bus using all available historical data in the past 6 hours at this particular bus. For example, input data for training on bus 1 include generation profiles of both generators, and the LMP trajectory; input data on bus 2 include the historical load and price profiles.

C. Simulation results

The measurement of forecasting accuracy at bus \( i \) is accomplished by MAPE:

\[
\text{MAPE} = \frac{\mathbb{E}[\hat{\pi}(i) - \pi(i)]}{\mathbb{E}[\pi(i)]}
\]

As shown in the TABLE. I, \( \hat{\pi}_{MAE} \) has the best performance on average while \( \hat{\pi}_{NN} \) is the worst. However, the result is opposite at some specific location, e.g., bus 3, because the price is the location dependent. Such differences lead to the gap of MAPEs between different buses which is primarily due to the presence of price spikes. Since ANN is a data driven technique, it is very sensitive to unpredictable price spikes. With the knowledge of the pricing mechanism, on the other hand, the proposed algorithm is capable of predicting price spikes to some extent and hence the accuracy is considerably improved.

V. CONCLUSION

In this paper, we presented a novel probabilistic real-time LMP forecasting algorithm that exploits the knowledge of the pricing mechanism. With the comprehensive study of the PJM ex-post real-time pricing model, we provided insights of the discrete price state abstraction. Such a discrete concept motivated the modeling of the real-time price as a time inhomogeneous Markov chain. We used a Monte Carlo method to estimate price state transition probability matrices, based on which the distribution of a future price can be computed. The simulation results demonstrated that the proposed forecasting algorithm outperforms the well accepted ANN technique.

Despite its superior performance, the proposed approach suffers from the computational burden imposed by the Monte Carlo estimation of price transition probability matrices. Therefore, a more efficient and scalable computation method is desired. Incorporating a generator behavior model and the possibility of contingency events is also an interesting research direction that may improve the proposed approach.

REFERENCES