Short-Term Congestion Forecasting in Wholesale Power Markets

Qun Zhou, *Student Member, IEEE*, Leigh Tesfatsion, *Member, IEEE*, and Chen-Ching Liu, *Fellow, IEEE*

***Abstract*—Short-term congestion forecasting is highly important for market participants in wholesale power markets that use locational marginal prices (LMPs) to manage congestion. Accurate congestion forecasting facilitates market traders in bidding and trading activities and assists market operators in system planning. This study proposes a new short-term forecasting algorithm for congestion, LMPs, and other power system variables based on the concept of system patterns—combinations of status flags for generating units and transmission lines. The advantage of this algorithm relative to standard statistical forecasting methods is that structural aspects underlying power market operations are exploited to reduce forecast error. The advantage relative to previously proposed structural forecasting methods is that data requirements are substantially reduced. Forecasting results based on a NYISO case study demonstrate the feasibility and accuracy of the proposed algorithm.**

***Index Terms*—Congestion forecasting, convex hull algorithm, load partitioning, locational marginal price, price forecasting, system patterns, wholesale power market.**

I. Introduction

In many transmission regions, congestion in wholesale power markets is managed by locational marginal prices (LMPs), the pricing of power in accordance with the location and timing of its injection into or withdrawal from the transmission grid. Congestion and LMP forecasts are highly important for the decision-making of market participants. Accurate congestion and LMP forecasts give advantages to market traders in bidding and trading activities and to market operators for system planning.[1]

Many studies have focused on electricity price forecasting based on statistical tools [1]–[5] and structural models [6], [7], yet few studies have focused on congestion forecasting. Li [8] applies a statistical model to predict line shadow prices. EPRI [9] has developed a congestion forecasting model that uses sequential Monte Carlo simulation to produce a probabilistic load flow. The EPRI model provides congestion probabilities for transmission lines of interests, but it requires intensive data input to the load flow model.

Li and Bo [10], [11] examine LMP variation in response to load variation and they predict the next binding constraint when load is increased. However, the authors also assume that a particular system growth pattern exists and that load growth at each bus is proportional to this pattern. Most U.S. wholesale power markets operating under LMP are geographically large; hence, distributed loads do not necessarily exhibit proportional growth. Moreover, the authors' approach has not been applied in large-scale power systems where practical issues of limited data availability need to be considered.

In our previous study [12], a piecewise linear-affine mapping between distributed loads and DC-OPF system variable solutions was identified and applied to forecast congestion and LMPs under the maintained assumption that complete historical information was available regarding the marginality (or not) of generating units and the congestion (or not) of transmission lines. This method is able to give an exact prediction result since it is derived from the core structure of a wholesale power market. However, when applied to the actual forecasting of large-scale wholesale power systems, data requirements become a problem. The needed historical generation capacity data and line flow data are either publicly unavailable on market operator websites or only available with some delay. Consequently, the correct pattern of binding constraints corresponding to any possible future load point is difficult to effectively identify, which in turn prevents the accurate forecasting of system variables.

Building on [12], this study develops a new algorithm for the short-term forecasting of system variables in wholesale power systems with substantially reduced data requirements. This algorithm permits the derivation of estimated probability distributions for congestion, LMPs, and other DC-OPF system variable solutions in real-time markets and in forward markets with hour-ahead, day-ahead, and week-ahead time horizons, conditional on a given *commitment-and-line scenario* that specifies a set of generating units committed for possible dispatch and a set of transmission lines capable of supporting power flow. Moreover, given suitable availability of historical data, this scenario-conditioned forecasting algorithm can be generalized to a cross-scenario forecasting algorithm by the assignment of probabilities to different commitment-and-line scenarios.

This new forecasting algorithm makes use of two supporting techniques in order to substantially reduce the amount of required data relative to [12]. The first technique is a method developed by Bemporad *et al.* [13] and Tøndel *et al.* [14] for

Manuscript received July 13, 2010; revised July 21, 2010, October 28, 2010, and January 20, 2011; accepted February 18, 2011. Date of publication April 05, 2011; date of current version October 21, 2011. This work was supported in part by the ISU Electric Power Research Center. The project homepage is http://www.econ.iastate.edu/tesfatsi/EPRC.htm. Paper no. TPWRS-00561-2010.

Q. Zhou is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50010 USA (e-mail: qzhou@iastate.edu).

L. Tesfatsion is with the Department of Economics, Iowa State University, Ames, IA 50010 USA (e-mail: tesfatsi@iastate.edu).

C.-C. Liu is with the School of Electrical, Electronic, and Mechanical Engineering, University College Dublin, Dublin, Ireland (e-mail: liu@ucd.ie).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2011.2123118

[1]For example, during an internship at Genscape, Inc., the first author observed first-hand that the customers for Genscape's LMP forecasting services were generation companies, load-serving entities, and utilities interested in developing daily market bidding strategies and improving their over-the-counter electricity trading.

dividing the parameter space of a *quadratic-linear programming (QLP)* problem into convex subsets such that, within each convex subset, the optimal solution values can be expressed as linear-affine functions of the parameters. A similar technique is applied in this study to a QLP DC-OPF problem formulation to show that, conditional on any given commitment-and-line scenario, the load space can be divided into convex subsets within which the optimal DC-OPF system variable solutions are linear-affine functions of load. Each convex subset corresponds to a unique *system pattern*, that is, a unique array of flags reflecting a particular pattern of binding minimum or maximum capacity constraints for the committed generating units and available transmission lines specified by the commitment-and-line scenario.

The second technique concerns convex hull determination. Given any collection of points, computational geometry [15] provides algorithms to compute the corresponding *convex hull*, i.e., the smallest convex set containing these points. Convex hull algorithms have been gaining popularity in the areas of computer graphics, robotics, geographic information systems, and so forth. To date, however, they have not been applied in electricity market forecasting. A convex hull algorithm is used in this study to estimate the convex subsets of load space within which DC-OPF solutions are linear-affine functions of load when incomplete historical data prevent their exact determination.

More precisely, our new forecasting algorithm generates short-term forecasts for congestion, LMPs, and other power system variables as follows. Let L denote a vector of loads at some possible future operating point corresponding to a particular commitment-and-line scenario S. A convex hull method is first used to estimate the division of load space into convex subsets (*system pattern regions*), each corresponding to a distinct historically-observed system pattern of binding capacity constraints for the particular committed generating units and available transmission lines specified under S. A *probabilistic point inclusion test* is next used to calculate the probability that L is associated with each historical system pattern, taking into account the imprecision with which the system pattern regions in load space are estimated. The congestion conditions at L are then probabilistically forecasted using the probability-weighted historical system patterns and forecasts for LMPs and other system variables at L are calculated using the linear-affine mapping between load and DC-OPF system variable solutions that corresponds to each probability-weighted historical system pattern.

Compared to state-of-the-art forecasting techniques for power markets, our new scenario-conditioned forecasting algorithm has the following two advantages:

- First, our algorithm makes novel use of convex hull techniques to enable the short-term forecasting of congestion conditions, prices, and other system variables for large-scale wholesale power systems using only publicly available data.
- Second, our algorithm proposes the novel use of system patterns as an effective way to take generation and transmission capacity constraints into account when forecasting DC-OPF-generated system variable solutions, thus permitting more accurate forecasts to be obtained.

The remainder of this paper is organized as follows. The general formulation of our forecasting problem is presented in Section II. A detailed description of our basic scenario-conditioned forecasting algorithm is provided in Section III. In Section IV, after some practical data-availability issues are addressed, we present a probabilistic extension of our basic scenario-conditioned forecasting algorithm that is suitable for addressing these data availability issues. In Section V, we present a NYISO case study that illustrates the effectiveness of the probabilistic scenario-conditioned forecasting algorithm developed in Section IV. In Section VI, we discuss how this algorithm can be further generalized to permit cross-scenario forecasting. Concluding remarks are provided in Section VII.

II. Basic Forecasting Problem Formulation

In electricity markets, congestion occurs when the available economical electricity has to be delivered to load "out-of-merit-order" due to transmission limitations. That is, higher-cost generation needs to be dispatched in place of cheaper generation to meet this load in order to avoid overload of transmission lines. In this case, the LMP levels at different nodes separate from each other and from the unconstrained market-clearing price. Therefore, congestion is a critical factor determining the formation of LMP levels.

However, congestion patterns are difficult to anticipate since they are related to the network topology of power systems. Provided perfect information is available, such as network data, load data, and generator bidding data, a market clearing model could be utilized to obtain accurate forecasts of congestion conditions and prices. Nevertheless, two issues arise for this direct forecasting method. First, most market traders do not have direct access to the information that is needed to implement this method; they would have to depend on data published by market operators. Second, the market operators, themselves, would need a high degree of computational speed to carry out the required computations.

As a result, statistical tools have been developed that tackle these two forecasting issues by modeling the statistical correlation between prices and explanatory factors. These statistical tools lack explicit consideration for congestion, partly because no effective approach has been developed to enable these tools to capture and express the effects of congestion. Ignoring the effects of congestion makes the forecasted prices less reliable and difficult to interpret at operating points with abnormal price behaviors.

Surely it is possible to glean some useful information about future possible congestion conditions based on statistically forecasted LMPs. However, these intuitive insights, based on forecasters' experiences, cannot provide reliable congestion forecasts. From a cause-and-effect point of view, congestion is the cause while LMP is the effect. One cannot infer the cause (congestion) from the effect (LMP) since LMP is not solely driven by congestion. In particular, statistical LMP forecasting tools do not take into account the structural aspects of power markets that fundamentally drive the determination of LMPs: namely, the fact that LMPs are derived as solutions to optimal power flow problems subject to generation capacity and transmission line constraints.

TABLE I
FLAGS USED FOR SYSTEM PATTERNS

	Generating units			Transmission lines		
State	Minimum capacity	Marginal unit	Maximum capacity	Neg Congestion	No Congestion	Pos Congestion
Flag	-1	0	1	-1	0	1

As explained more carefully in Section III, the novel concept of a "system pattern" is used in this study to incorporate the structural generation capacity and transmission line aspects that drive congestion outcomes. The forecasting of congestion at a possible future operating point is thus transformed into a problem of estimating the correct system pattern at this operating point. Moreover, the forecasting of prices and other system variables at this operating point can subsequently be undertaken using the particular linear-affine mapping between load and DC-OPF system variable solutions that is associated with this system pattern.

This basic forecasting approach makes three simplifying assumptions. First, it is assumed that the forecasting of system variables at possible future operating points can be conditioned on a particular commitment-and-line scenario, that is, a particular generation commitment (designation of generating units available for dispatch) and a particular network topology (designation of available transmission lines). Second, it is assumed that a lossless DC-OPF problem formulation is used for the determination of LMPs and other system variables, implying in particular that the loss components of LMPs are neglected. Third, it is assumed that generator supply-offer behaviors are relatively static in the forecasting horizons.

III. BASIC FORECASTING ALGORITHM DESCRIPTION

A. System Patterns and System Pattern Regions

At any system operating point, the number of marginal generating units and binding transmission constraints tends to be small compared to the number of nodes, transmission lines, and generating units. For example, in the Midwest Independent System Operator (MISO) region with 36 845 network buses and 5575 generating units, the number of day-ahead binding constraints is published daily and is typically observed to be less than 20 for an hourly interval [16]. On the other hand, high-cost units such as gas and oil units are more likely to become marginal units during peak hours, the number of which is modest.

Exploiting this important characteristic of power markets, the idea of a *system pattern* is introduced consisting of a vector of flags indicating the marginal status of committed generating units and the congestion status of available transmission lines at any given system operating point; see Table I. As long as the number of marginal generating units (labeled 0) and the number of congested transmission lines (labeled -1 or 1) are relatively few in number, the number of possible system patterns can be easily handled.

As noted in Section II, the basic congestion forecasting problem can then be transformed into a problem of estimating the correct system pattern for any given possible future operating point. The congestion forecast is directly obtained once the system pattern is estimated, since the status of transmission lines is part of the system pattern. Moreover, as clarified below in Section III-D, short-term forecasts for prices and other system variables at the operating point can also be obtained making use of this estimated system pattern.

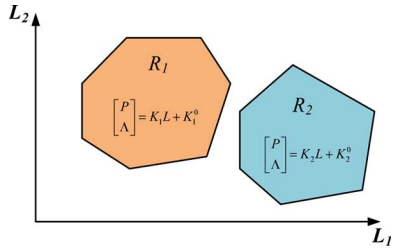

Fig. 1. Illustration of two system pattern regions (SPRs) in load space.

The proposition below provides the theoretical foundation for our proposed forecasting approach. The proposition uses the concept of a *convex polytope* for an n-dimensional Euclidean space R^n, i.e., a region in R^n determined as the intersection of finitely many half-spaces in R^n.

Proposition 1: Suppose a standard DC-OPF formulation with fixed loads and quadratic generator cost functions is used by a market operator to determine system variable solutions. Then, conditional on any given commitment-and-line scenario S, the load space can be covered by convex polytopes such that: 1) the interior of each convex polytope corresponds to a unique system pattern and 2) within the interior of each convex polytope, the system variable solutions can be expressed as linear-affine functions of the vector of distributed loads.

The proof of Proposition 1, originally derived in [17], is outlined in an appendix to this paper. The proof starts with the derivation of inequality and equality constraints constructed from the first-order KKT conditions for a DC-OPF problem conditional on a particular commitment-and-line scenario S. The inequality constraints characterize convex polytopes that cover the load space, where the interior of each convex polytope corresponds to a unique system pattern. The convex polytopes constituting the covering of the load space are referred to as *system pattern regions (SPRs)* for the fact that the interior of each convex polytope is associated with a unique system pattern.

Within each SPR, the equality constraints take the form of linear-affine equations with constant coefficients that describe fixed linear-affine relationships between DC-OPF system variable solutions and the vector of loads. The matrix of coefficients for these linear-affine functions gives the rates of change with regard to real-power dispatch levels for generating units and shadow prices for bus balance and line constraints when loads are perturbed within the region. This matrix is referred to below as the *sensitivity matrix* for this SPR.

Fig. 1 provides illustrative depictions of two SPRs, R_1 and R_2, together with their associated linear-affine mappings, when the load space is composed of two-dimensional load vectors $L = (L_1, L_2)$. The symbol P denotes the vector of unit dispatch levels and the symbol Λ denotes the vector of dual variables. The mappings are characterized by sensitivity matrices (K_1, K_2) and ordinate vectors (K_1^0, K_2^0) that are constant within each SPR, which implies that the DC-OPF solutions for P and Λ can be expressed as fixed linear-affine functions of the load vector L within each SPR.

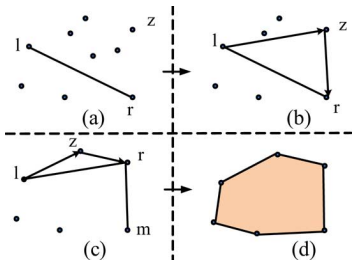

Fig. 2. Illustration of the QuickHull algorithm.

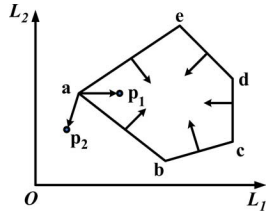

Fig. 3. Illustration of the basic point inclusion test for an SPR in a load plane.

B. Convex Hull Estimation of Historical SPRs

In practice, deriving the exact form of the SPRs is difficult due to limited access to most of the required information. This required information includes supply offer data, generating unit capacity data, and transmission limit data.

This lack of information can be overcome by applying a "convex hull algorithm" to historical load data to estimate SPRs. The *convex hull* of a point set B is the smallest convex set that contains all the points of B [18]. A *convex hull algorithm* is a computational method for computing the convex hull of a set B.

As will be concretely illustrated in Section V, each historical load point corresponding to a particular commitment-and-line scenario S can in principle be associated with a distinct system pattern based on corresponding historical data regarding the marginal status of the committed generating units and the congested status of the available transmission lines. The historical SPR corresponding to each such historically identified system pattern can then be estimated by deriving the convex hull of the collection of all historical load points that have been associated with this system pattern.

This study makes use of the "QuickHull algorithm" to estimate historical SPRs conditional on a given commitment-and-line scenario S. The *QuickHull algorithm*, developed by Barber *et al.* [19], is an iterative procedure for determining all of the points constituting the convex hull of a finite set B. At each step, points in B that are internal to the convex hull of B, and hence not viable as vertices of the convex hull, are identified and eliminated from further consideration. This process continues until no more such points can be found.

An illustrative application of the QuickHull algorithm for a finite planar set B is presented in Fig. 2. The set B is first partitioned into two subsets $B1$ and $B2$ by a line *lr* connecting a left-most upper point l to a right-most lower point r, as depicted in Fig. 2(a). More precisely, the points in B with the smallest x value are first selected and from among these points, a point with a largest y value is chosen to be the left-most upper point l; similarly for the right-most lower point r. For each subset $B1$ and $B2$, a point z in B that is furthest from lr is determined and two additional lines are constructed, $\overrightarrow{lz}$ from l to z and $\overrightarrow{zr}$ from z to r; see Fig. 2(b). By construction, points of B that lie strictly inside the resulting triangle *lzr* are strictly interior to the convex hull of B and hence can be eliminated from further consideration. The points on the triangle itself are possible vertex points for the boundary of the convex hull of B.

To continue the recursion, the above procedure is repeated for the reduced subset $BRed$ of B resulting from this elimination. Specifically, two subsets and associated triangles are formed as before for $BRed$ and the points of $BRed$ lying within the interiors of the resulting triangles are eliminated. If a triangle ever degenerates to a line, then all the points along the line lie on the boundary of the convex hull of B by construction. For example, in Fig. 2(c), the endpoints r and m of the line rm both lie on the boundary of the convex hull of B.

This process of elimination continues until no additional points to be eliminated can be found. Since B is finite, the process is guaranteed to stop in finitely many steps. All the convex hull points for B (boundary and interior) can be determined recursively in this manner. The complete convex hull for B is depicted in Fig. 2(d). By construction, this convex hull is a planar convex polytope.

The main advantage of the QuickHull algorithm relative to other such algorithms is its ability to efficiently handle high-dimensional sets B by reducing computational requirements [20]. The QuickHull algorithm has been widely used in scientific applications and appears to be the algorithm of choice for higher-dimensional convex hull computing [21].

C. Basic Point Inclusion Test

Suppose the load space has been divided up into estimated SPRs whose interiors correspond to distinct system patterns, conditional on a given commitment-and-line scenario S. Consider, now, the task of forecasting congestion conditions at some future operating point a short time into the future for which scenario S again obtains. The essence of this forecasting problem is the detection of the correct SPR for this future operating point. If the correct SPR can be detected, then congested conditions can be inferred directly from the corresponding system pattern.

This detection is undertaken in this study by means of a "point inclusion test". The *basic point inclusion test* used in this study is illustrated in Fig. 3 for an SPR in a load plane. Recall from Section III-A that each SPR takes the form of a convex polytope, i.e., a region expressable as the intersection of half-spaces; hence each SPR has flat faces with straight edges. Let the normal vectors pointing towards the interior of the SPR be constructed for each edge of the SPR. Now consider the depicted point P_1 and let $\overrightarrow{aP_1}$ denote the vector directed from the vertex a to the point P_1. The dot product between $\overrightarrow{aP_1}$ and each normal vector of each neighboring edge of a is greater than or equal to 0. If this is true for all vertices of the SPR, the point P_1 is judged to be on or inside the SPR. On the other hand, one can see that P_2 is outside the SPR since the dot product of $\overrightarrow{aP_2}$ and the normal vector for the neighboring edge connecting a to b is negative.

As will be seen in Section IV, practical data-availability issues prevent the use of the basic point inclusion test for the

exact determination of the SPR containing any possible future load point L. However, given a suitable probabilistic extension of this basic point inclusion test, the probability that any particular SPR contains L can be estimated.

D. Linear-Affine Mapping Procedure

Given sufficient generation and transmission information, each historical load point can be associated with an SPR according to the status of the generating units and transmission lines at the historical operating time. More precisely, given any commitment-and-line scenario S, consider the collection of all historically observed load points obtaining under S. Let this collection of historical load points be partitioned into subsets corresponding to distinct system patterns for scenario S. For each load subset, use the QuickHull algorithm to calculate its convex hull in load space. Each of these convex hulls then constitutes a distinct estimated SPR for scenario S. In principal, any future load point corresponding to scenario S can then be associated with one of these estimated SPRs by means of the basic point inclusion test. This association permits the prediction of congestion, prices, and other DC-OPF system variable solutions at this load point.

To see this more clearly, let Y_i^h and L_i^h denote matrices consisting of all historically observed DC-OPF system solution vectors and load vectors corresponding to a particular system pattern i for a particular commitment-and-line scenario S. Let the SPR in load space corresponding to this system pattern, denoted by R_i, be estimated by the convex hull R_{Ei} of the collection of all of the historically observed load vectors included in L_i^h.

By Proposition 1, the mapping between Y_i^h and L_i^h can be expressed in the linear-affine form

$$Y_i^h = K_i L_i^h + K_i^0 \tag{1}$$

where K_i denotes the sensitivity matrix corresponding to R_i. Normally there will be multiple historical operating points corresponding to any one SPR for a given commitment-and-line scenario S. In this case ordinary least squares (OLS) can be applied to (1) to obtain estimates $\hat{K}_i$ and $\hat{K}_i^0$ for K_i and K_i^0, as follows:

$$\begin{bmatrix} (\hat{K}_i)^T \\ (\hat{K}_i^0)^T \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T(\mathbf{Y}_i^h)^T \tag{2}$$

where $\mathbf{X} = [(L_i^h)^T \quad \mathbf{1}]$.

Now let L_i^f denote a possible load vector for a future operating time that has been found to belong to the *estimated* SPR R_{Ei}, as determined from a basic point inclusion test applied to the collection of all historically estimated SPRs corresponding to scenario S. Then the forecasted vector Y_i^f of DC-OPF system variable solutions corresponding to L_i^f can be calculated as

$$Y_i^f = \hat{K}_i L_i^f + \hat{K}_i^0. \tag{3}$$

The above linear-affine mapping procedure is modified in Section IV to accommodate some practical issues arising from data incompleteness.

IV. Extension to Probabilistic Forecasting

Practical data availability issues arise for the implementation of the basic scenario-conditioned forecasting algorithm outlined in Section III. This section discusses how these issues can be addressed by means of a probabilistic extension of this basic algorithm. Throughout this discussion, the analysis is assumed to be conditioned on a given commitment-and-line scenario S.

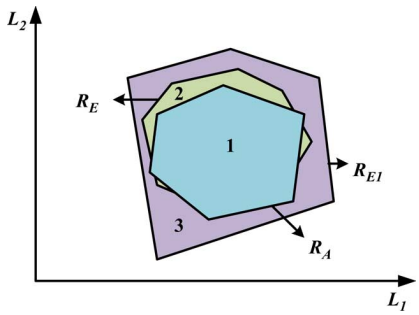

Fig. 4. Given incomplete constraint information and/or incomplete historical data, convex hull estimates for SPRs can be biased.

A. Practical Data Availability Issues

The basic scenario-conditioned forecasting algorithm proposed in Section III assumes that historical data are available regarding binding constraints for all generating units and for transmission lines on an hourly basis. In actuality, however, the marginal status of generating units is either confidential or published with limitations. Moreover, the theoretical load space cannot be fully reflected by the hourly historical load data which represent several realizations and subsets of the complete load space.

Due to these data limitations, in practice, the set $\mathcal{A}$ indexing hourly binding constraints cannot be completely determined. Consequently, estimates obtained for the SPRs could be biased.

The two basic ways in which this bias could arise are illustrated in Fig. 4 for a simple two-dimensional load space. Suppose the SPR corresponding to the true binding constraint set $\mathcal{A}$ is given by R_A (area 1) in Fig. 4.

This true SPR R_A can in principle be determined by applying the basic point inclusion test to every possible future operating point. Suppose, however, that the practically estimated binding constraint set A_{E1} is incomplete; for example, suppose A_{E1} only reflects the status of the most frequently congested lines. Given complete historical load data, the estimated convex hull R_{E1} (area 3) would then have to be larger than the true R_A (area 1) because A_{E1} is smaller (less restrictive) than the true $\mathcal{A}$. In fact, however, the actual estimated convex hull must be based on available historical load data. Since the latter is only a subset of the full load space, the result will be an actual estimated convex hull R_E (area 2) that lies within R_{E1} (area 3). In short, incompleteness of $\mathcal{A}$ and incompleteness of the practical load space each separately introduce bias in the estimate for R_A, but in opposing directions.

What are the practical implications of this bias for our basic forecasting algorithm? Two possible cases need to be handled, as illustrated in Fig. 5.

Case A: Point r in Fig. 5 lies in the interior of two different estimated SPRs, namely, R_{E1} and R_{E2} corresponding to two distinct system patterns A_1 and A_2. The true SPRs corresponding to A_1 and A_2 are denoted by the shaded regions R_{A1} and R_{A2}, respectively. The fact that the interiors of

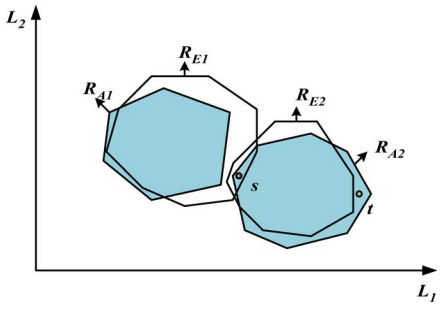

Fig. 5. Two possible types of forecast error due to biased SPR estimates.

the true SPRs do not overlap follows from Proposition 1 established in Section III-A. However, as explained above, overlap can occur for the interiors of estimated SPRs due to bias.

Case B: Point t in Fig. 5 is actually in the true SPR R_{A2}. However, point t cannot be assigned to either of the estimated SPRs because the bias in these estimates has caused point t to lie outside of both of them.

B. Probabilistic Point Inclusion Test

To mitigate the issues arising from the two types of bias discussed in Section III-A, mean and interval forecasting can be performed for the DC-OPF system variable solutions corresponding to any forecasted future load point L^f. This probabilistic forecasting can be implemented by estimating the probability of each SPR conditional on L^f, which can be characterized as a *probabilistic point inclusion test*.

More precisely, let L^f denote the forecasted load at a future operating point f and let R_i denote any particular SPR i. Let the collection of all historically identified SPRs be denoted by R^h and let CR denote the cardinality of R^h. Suppose the probability of occurrence for any SPR not in R^h is zero. Then the probability that R_i has occurred, given that L^f has been observed, can be expressed as

$$P(R_i|L^f) = \frac{P(L^f|R_i)P(R_i)}{\sum_{i\in R^h} P(L^f|R_i)P(R_i)}. \tag{4}$$

In practice, the various terms in (4) have to be estimated. In this study, it will be assumed that the prior probability $P(R_i)$ is an empirical prior estimated by the historical frequency of R_i: namely, the number of times in the past that R_i has been observed to occur divided by the total number of all past SPR observations.

The term $P(L^f|R_i)$ in (4) represents the probability of observing the load point L^f given that the true SPR is R_i. Intuitively, this probability should be a decreasing function of the distance between L^f and R_i. Therefore, this probability is estimated in this study as follows:

$$\hat{P}(L^f|R_i) = \frac{\left(\frac{1-D_i}{TD}\right)^{\gamma}}{\sum_{i\in R^h}\left(\frac{1-D_i}{TD}\right)^{\gamma}}. \tag{5}$$

In (5), the term D_i denotes the (Euclidean) distance between L^f and R_i and TD denotes the total distance calculated as the sum of the distances between L^f and each SPR in R^h. The *normalization parameter* γ in (5) can be adjusted to obtain an appropriate conditional probability measure, possibly by using historical data as training cases. A specification $\gamma = 0$ results in a uniform conditional probability (5) for L^f: namely, 1 divided by the cardinality CR of R^h. In this case, (5) is independent of the distance measures D_i. Alternatively, a specification $\gamma = 1$ implies the conditional probability (5) is derived from a linear normalization, while $\gamma = 2$ corresponds to a quadratic normalization. As will be shown below, the quadratic normalization form of the conditional probability (5) results in good forecasts for our NYISO case study.

Mean forecasts for the DC-OPF system variable solutions at the operating point f with forecasted load point L^f can then be obtained using the estimated form for the conditional probability assessments (4), denoted by P_i^f for short. Let Y_i^f denote the forecasted DC-OPF system variable solution vector corresponding to any historical SPR R_i in R^h. The mean forecast $\bar{Y}^f$ can then be calculated as

$$\bar{Y}^f = \sum_{i\in R^h} P_i^f Y_i^f. \tag{6}$$

A forecaster might also be interested in calculating upper and lower bounds for the DC-OPF system variable solutions calculated with respect to the most likely SPRs. Let nmp denote the forecaster's desired cut-off number of most probable SPRs and let MP represent the subset of R^h that contains these nmp most probable SPRs. Then the upper bound UB^f and lower bound LB^f for each forecasted DC-OPF system variable solution can be determined over the set of SPRs in MP. As a measure of dispersion, the forecaster can further consider the *coverage probability CP*, defined to be the summation of the probability assessments (4) for the nmp most probable SPRs.

Finally, another alternative might be for the forecaster to consider mean forecasts calculated using the nmp most probable SPRs, i.e., the subset MP of R^h. For example, a forecaster could choose $nmp = 1$, which would result in a point forecast for the DC-OPF system variable solutions based on a single most likely SPR R_i in R^h as determined from the estimated form of the conditional probability assessments (4).

C. Probabilistic Forecasting Algorithm

Taking into account the practical data issues addressed in Sections IV-A and B, our proposed probabilistic forecasting algorithm proceeds in four steps, as follows:

Step 1) Perform historical data processing to identify historical system patterns. Use the QuickHull algorithm to estimate historical SPRs as convex hulls of historically observed load points corresponding to distinct historical system patterns.

Step 2) For each historical SPR estimated in Step 1, a linear-affine mapping between load vectors and DC-OPF system variable solution vectors is derived using historical load and system variable data. The system variable solution vectors include real-power dispatch levels and dual variables for nodal balance and transmission line constraints. The linear-affine mapping is characterized by a sensitivity matrix and an ordinate vector.

Step 3) For any possible load point L^f in the near future for which system variable forecasts are desired, a probabilistic point inclusion test is performed. More precisely, the estimated form of the conditional probability distribution (4) is used to estimate the probability that L^f lies in each of the historical SPRs identified in Step 1.

Step 4) The results from Steps 1–3 are used to generate probabilistic forecasts at the future possible operating point L^f for generation capacity and transmission congestion conditions (system patterns) as well as for DC-OPF system variable solutions for dispatch levels and dual variables (including LMPs). For example, these probabilistic forecasts could take the form of mean and interval forecasts, or they could be point forecasts based on a most probable SPR.

V. NYISO Case Study

A. Case Study Overview

A case study using NYISO 2007 data is reported in this section for the probabilistic scenario-conditioned forecasting algorithm presented in Section IV-C. NYISO has a footprint covering 11 load zones [22]. Short-term zonal load forecasting data and binding constraints data are available at the NYISO website [23].

This forecasting algorithm is applicable for power markets using either nodal or zonal LMP pricing, since Proposition 1 in Section III-A does not rule out either form of pricing. However, NYISO's website [23] only posts daily zonal load data for its 11 load zones, which makes it impossible to forecast prices down to each node. In addition, historical NYISO price data reveal the similarity of LMPs within some of these 11 load zones, hence the negligibility of inter-zonal congestion between these zones. For this reason, to reduce our computational burden without any significant loss of information, we chose to reduce the original 11 load zones for the NYISO to 8 load zones by combining Zone Millwood with Dunwoodie and Zone West and Genesee with Central.

The top 25 most frequently congested high-voltage transmission lines during 2007 for the NYSIO day-ahead market are studied in [24]. The focus of our case study is on the five most frequently congested high-voltage transmission lines during 2007, specifically, DUNWODIE 345 SHORE RD 345 1 **(D-S)**, CENTRAL EAST-VC **(C-V)**, PLSNTVLY 345 LEEDS 345 1 **(P-L)**, WEST CENTRAL **(W-C)**, and SPRNBRK 345 EGRDNCTR 345 1 **(S-E)**. Since the marginal status of generating units is not available from the NYISO, the conditioning scenario for this empirical study is taken to be the availability of these five lines. System patterns are thus equivalent to congestion patterns for these five lines.

Regarding time period, we selected 12 test days consisting of the last day of each month in 2007. The 24 operating hours starting from 0:00 for each test day were treated as future operating points. Forecasted load data at these hours were used to identify system patterns and to generate system variable forecasts. These forecasted results were then compared with actual realizations to evaluate the performance of our algorithm. Due to space limitations, graphical illustrations are presented only for January 31 and February 28; numerical results for the last days of other months are given in tables.

TABLE II
Four Most Frequent Historical Congestion Patterns for 01/31/2007

Pattern	D-S	C-V	P-L	S-C	S-E
P1	1	0	0	0	0
P2	0	0	0	0	0
P3	1	1	0	0	0
P4	1	1	0	1	0

All calculations for this case study were implemented using Matlab 7.8 on an Intel Core 2 PC with 3.0-GHz CPU. The computational time for each daily forecast was about 2 min.

B. Implementation of Probabilistic Forecasting

Historical price and load data were first processed to identify historical system patterns and SPRs, which is Step 1 of our probabilistic forecasting algorithm. Sorted by congestion patterns, about 19 to 30 historical system patterns (hence SPRs) were found for each forecasted day. For example, the four most frequently observed congestion patterns for January 31 are shown in Table II. System patterns for other days are categorized similarly.

Step 2 of our algorithm was then carried out. Specifically, the sensitivity matrix and ordinate vector for each historical SPR were estimated by ordinary least squares, making use of the actual system operating points observed for each historical system pattern.

In Step 3, forecasted load data for the 24 operating hours of each test day were then treated as possible future load points. For each of the latter points, the probabilistic point inclusion test detailed in Section IV-B was used to assign estimated conditional probability assessments (4) giving the probability that this future load point was contained within each historical SPR. In these Step 3 calculations, we first evaluated the forecasting performance of three values (0, 1, and 2) for the normalization parameter γ in (5) on the basis of historical data. The specification $\gamma = 2$ gave the best forecast results for most historical days; hence, this value was chosen to forecast system variables for the future load points.

Finally, in Step 4, the results of Steps 1–3 above were used to generate probabilistic forecasts in the form of mean and interval forecasts. For the mean forecasts, nmp was set equal to the cardinality CR of R^h. For the interval forecasts, nmp was set equal to 4.

For the interval forecasts, the size of nmp (i.e., the cut-off number of most probable SPRs) depends on the forecaster's desired trade-off between accuracy and precision. A larger nmp tends to increase forecasting accuracy, in the sense that there is a better chance the correct SPR will be among the considered SPRs. On the other hand, the precision of any resulting mean forecast is correspondingly reduced (i.e., the variance of the forecasts across the considered SPRs is increased). In the current study, the specification $nmp = 4$ is used for interval forecasts because it results in good precision without significant loss of coverage probability.

TABLE III
FORECASTED CONGESTION PATTERNS VERSUS THE ACTUAL CONGESTION PATTERN ON 01/31/2007

Time	Forecasted	Probabilities	CP	Actual
0:00	1 0 0 0 0	0.3632	0.9541	1 0 0 0 0
	0 0 0 0 0	0.2411		
	1 1 0 0 0	0.2066		
	1 1 0 1 0	0.1432		
5:00	1 0 0 0 0	0.3451	0.9398	1 0 0 0 0
	0 0 0 0 0	0.2043		
	1 1 0 0 0	0.2418		
	1 1 0 1 0	0.1486		
10:00	1 0 0 0 0	0.4237	0.9426	1 0 0 0 0
	0 0 -1 0 0	0.0236		
	1 1 0 0 0	0.3654		
	1 1 0 1 0	0.1299		
15:00	1 0 0 0 0	0.3661	0.9452	1 1 0 0 0
	0 0 -1 0 0	0.0271		
	1 1 0 0 0	0.4243		
	1 1 0 1 0	0.1277		
20:00	1 0 0 0 0	0.4247	0.9435	1 0 0 0 0
	0 0 0 0 0	0.0244		
	1 1 0 0 0	0.3612		
	1 1 0 1 0	0.1332		

C. Congestion Pattern Forecasts

Table III reports the four most probable hourly congestion patterns, along with their associated estimated conditional probabilities and coverage probability CP (based on $nmp = 4$), for every fifth hour of January 31, 2007, starting from hour 0:00. Actual congestion patterns corresponding to each reported hour are highlighted in gray. As seen, for the reported hours, the actual congestion pattern is always included among the forecasted congestion patterns and has the highest estimated conditional probability. For future reference, note also that the first entry of the actual congestion pattern, corresponding to transmission line D-S, is always 1. This indicates that D-S is frequently congested.

The multiple forecasted congestion patterns associated with each reported hour in Table III represent several credible congestion scenarios that could occur in the future. If a forecaster desires to derive one forecast for the future congestion pattern, an intuitively reasonable option would be to select a forecasted congestion pattern that has the highest associated conditional probability (4). As observed in Table III, for the case study at hand, this approach would result in the correct prediction of the actual congestion pattern for each reported hour. In general, however, more reliable forecasts for system conditions and DC-OPF system variable solutions would be obtained by making fuller use of the conditional probability assessments (4) to form mean forecasts and interval forecasts.

D. Mean Forecasts for LMPs

One of the benefits of congestion forecasting is to enable the more precise prediction of LMPs for market operators and traders in their short-term decision making. Forecasted and actual LMPs for Zone Central on January 31 and February 28 are shown in Figs. 6 and 7. *Root mean squared error (RMSE)* and

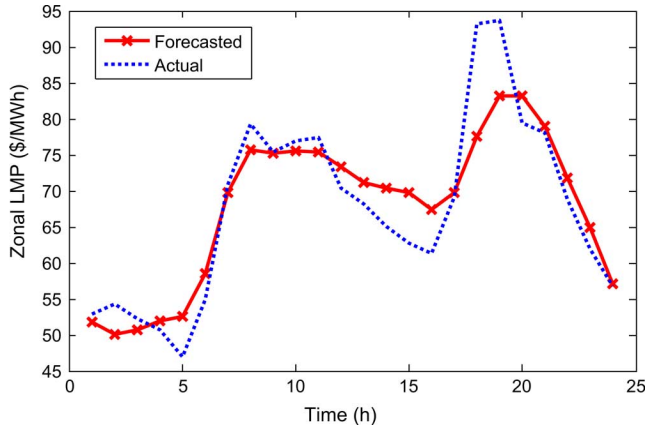

Fig. 6. Actual versus mean LMP forecasts for Zone Central on 01/31/2007.

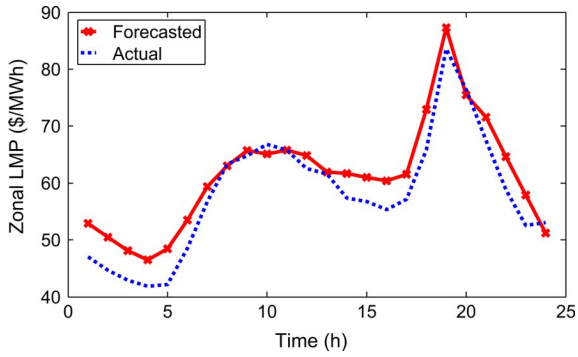

Fig. 7. Actual versus mean LMP forecasts for Zone Central on 02/28/2007.

mean absolute percentage error (MAPE) [5] are used as measures of forecast accuracy:

$$\text{RMSE} = \sqrt{\frac{1}{24}\sum_{i=1}^{24}\left(LMP_i^{actual} - LMP_i^{forecast}\right)^2} \quad (7)$$

$$\text{MAPE} = \frac{1}{24}\sum_{i=1}^{24}\frac{|LMP_i^{acutal} - LMP_i^{forecast}|}{LMP_i^{actual}}. \quad (8)$$

Table IV reports the RMSE and MAPE obtained using our probabilistic forecasting algorithm for each of our 12 test days. Corresponding forecast results obtained using a well-known statistical model—the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model [4]—are also shown for comparison. As seen, except for the slightly smaller MAPE value attained in February using GARCH, our forecasting algorithm outperforms GARCH in the sense that smaller RMSE and MAPE values are obtained.

E. Interval Forecasts for Line Shadow Prices and LMPs

Interval forecasting is recommended over mean forecasting for line shadow prices. As clarified below, interval forecasting is more informative than mean forecasting for line shadow prices because the underlying attribute of interest (negative-direction, zero, or positive-direction congestion) is measured by a discretely-valued indicator (-1, 0, or 1).

Hourly upper-bound and lower-bound interval forecasts for the line shadow prices on line D-S on January 31 and February

TABLE IV
RMSE AND MAPE VALUES FOR THE 12 TEST DAYS

Day	RMSE		MAPE	
Model	Proposed Alg	GARCH	Proposed Alg	GARCH
01/31/2007	5.026	8.689	0.0525	0.0902
02/28/2007	3.393	4.465	0.0472	0.0384
03/31/2007	4.029	7.094	0.0677	0.0727
04/30/2007	4.853	8.297	0.0535	0.1005
05/31/2007	7.401	14.741	0.0934	0.1198
06/30/2007	3.439	13.359	0.0679	0.1485
07/31/2007	3.941	11.623	0.0530	0.1082
08/31/2007	4.076	5.913	0.0671	0.0781
09/30/2007	3.249	6.636	0.0603	0.0862
10/31/2007	4.135	8.561	0.0638	0.1176
11/30/2007	6.476	7.208	0.0770	0.0855
12/31/2007	7.051	14.185	0.0903	0.1435

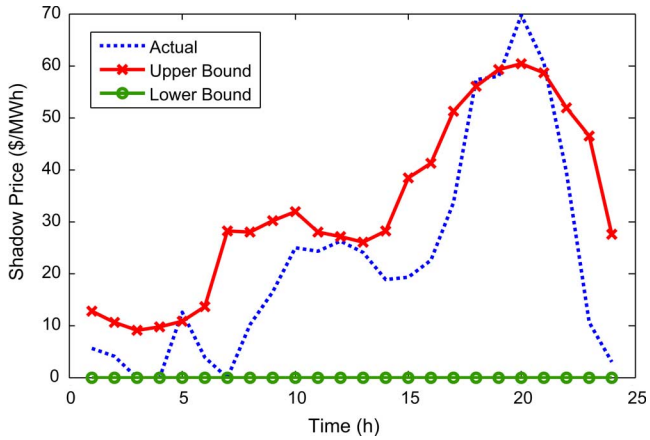

Fig. 8. Actual versus interval D-S line shadow price forecasts on 01/31/2007.

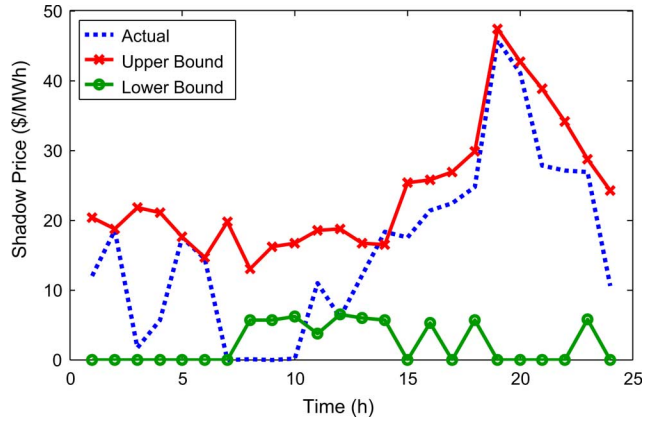

Fig. 9. Actual versus interval D-S line shadow price forecasts on 02/28/2007.

28 are shown in Figs. 8 and 9 along with actual line shadow prices for comparison. As seen, the actual line shadow prices for most hours fall within the forecasted intervals.

To better interpret these findings, consider the Table III results which forecast that line D-S (the first congestion pattern entry)

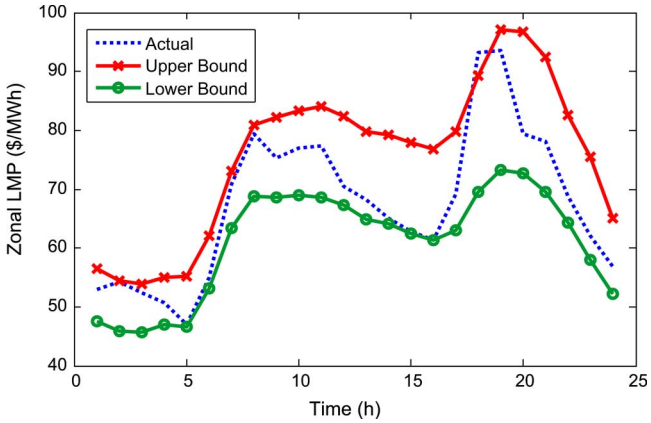

Fig. 10. Actual versus interval LMP forecasts for Zone Central on 01/31/2007.

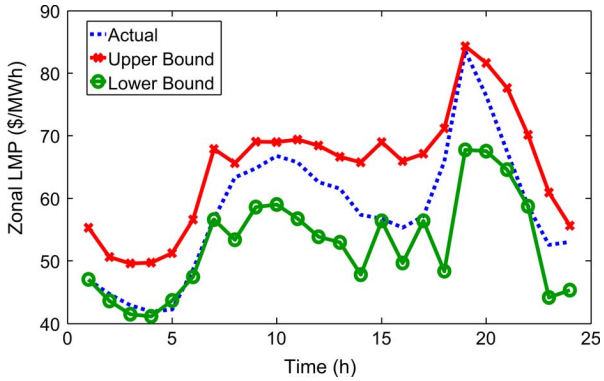

Fig. 11. Actual versus interval LMP forecasts for Zone Central on 02/28/2007.

will be either congested or not during hour 20 with varying probabilities. If congestion is forecasted, it is in the positive direction $(+1)$ and from Fig. 8, the line shadow price is estimated to be about \$60/MWh. On the other hand, if no congestion is forecasted (0), then from Fig. 8, the line shadow price is estimated to be \$0/MWh.

One final point for interval forecasts for line shadow prices is important to note. For lines for which no congestion occurs in any of the reported congestion patterns (e.g., line S-E in Table III), the corresponding upper and lower bounds for the forecasted line shadow price interval will both be zero, indicating zero congestion.

Interval forecasts for Zone Central LMPs on January 31 and February 28 are shown in Figs. 10 and 11 along with actual LMP values for comparison. For most hours, the actual LMP values fall within the upper and lower bounds of the forecasted intervals.

The interval forecasting performance for line shadow prices and zonal LMPs is measured using the accuracy-informativeness tradeoff model developed in [25]. The *statistical loss function LOSS* is defined to be

$$LOSS = \frac{|y - m|}{g} + \delta ln(g). \tag{9}$$

In (9), y denotes the actual value, m denotes the midpoint of the forecasted interval, and $\ln(g)$ denotes the natural logarithm of the width g of the forecasted interval. Also, δ determines the tradeoff between accuracy (the first term) and informativeness (the second term); in this case study, δ is set to 1. Note

TABLE V
LOSS FUNCTION VALUES AS A MEASURE OF INTERVAL FORECASTING PERFORMANCE FOR THE 12 TEST DAYS

Day	Shadow Price Forecasting		LMP Forecasting	
Model	Proposed Alg	GARCH	Proposed Alg	GARCH
01/31/2007	3.824	4.063	2.896	4.196
02/28/2007	3.729	3.977	2.835	3.649
03/31/2007	3.236	3.672	2.574	3.633
04/30/2007	3.398	3.778	3.133	4.164
05/31/2007	3.493	4.187	3.421	4.032
06/30/2007	3.838	3.897	3.365	4.425
07/31/2007	2.726	3.350	2.839	4.892
08/31/2007	2.916	3.352	2.787	3.624
09/30/2007	3.140	3.567	2.245	3.965
10/31/2007	2.825	3.335	2.725	3.799
11/30/2007	3.256	3.738	3.088	3.537
12/31/2007	3.481	3.962	3.164	3.919

that a smaller *LOSS* indicates better performance for interval forecasting.

Table V gives the *LOSS* values for the interval forecasts obtained for line shadow price and zonal LMPs using our probabilistic forecasting algorithm versus the forecasts obtained using a statistical GARCH model. As seen, our probabilistic forecasting algorithm results in uniformly lower *LOSS* values than GARCH, indicating a better forecasting performance.

A possible explanation for this performance difference is that GARCH has difficulty handling the volatility of line shadow prices, which can abruptly change from 0 to large nonzero values. In contrast, our probabilistic forecasting algorithm captures the physical meaning of these line shadow prices and this facilitates better forecasting.

In this study we observed that, in some months (January, May, November, and December), the peak-hour LMPs and line shadow prices were difficult to forecast with precision. This phenomenon could possibly be due to changes in the generating unit commitment pattern or in the transmission network topology over the forecast horizon. To enhance peak-hour forecasting results, more careful collection of historical data might be needed to ensure that these historical data correspond to the same commitment-and-line scenario as the forecasted point. Alternatively, as discussed in the following Section VI, an extended cross-scenario forecast study could be attempted.

VI. EXTENSION TO CROSS-SCENARIO FORECASTING

To this point, the forecasting algorithm developed in this study has been conditioned on a given commitment-and-line scenario S specifying a particular generating unit commitment pattern and a particular transmission network topology. One interpretation of S is that it represents anticipated conditions at a future operating point for which forecasts are desired. Another interpretation of S is that it represents a possible future system contingency (e.g., an N-1 outage scenario) under consideration in a contingency planning study.

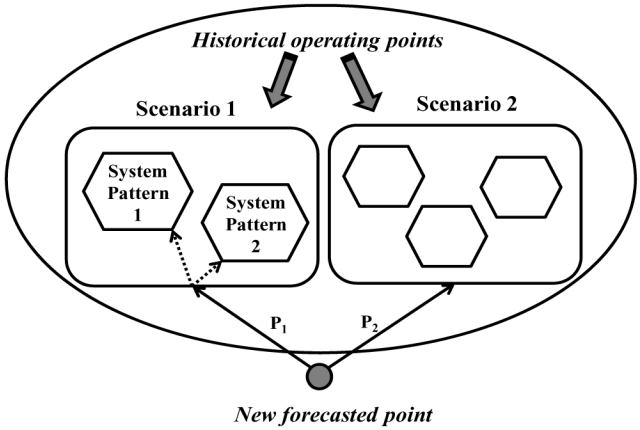

Fig. 12. Scenario-conditioned and cross-scenario forecasting.

A possibly useful extension of this algorithm would be to assign probabilities to distinct scenarios, thus permitting the probabilistic cross-scenario blending of forecasts. These scenarios could be characterized not only on the basis of system patterns, i.e., generating unit commitments and transmission network topology, but also on the basis of a variety of other types of contingencies.

As illustrated in Fig. 12, for any future operating point whose system conditions need to be forecasted, the corresponding generating unit commitment, transmission network topology, and other contingencies could be projected with some probabilities. In each of these projected scenarios, our scenario-conditioned forecasting algorithm could be applied to estimate congestion, LMPs, and other system variables. The final forecast for any system variable of interest could then be the expected value of this system variable calculated across all projected scenarios.

VII. CONCLUSION

Short-term congestion forecasting is critical for both market traders and market operators. Congestion forecasting helps to explain electricity price behaviors and facilitates the decision-making of power system participants.

This study first proposes a basic scenario-conditioned forecasting algorithm that permits the short-term forecasting of congestion, prices, and other power system variables conditional on a given generating unit commitment pattern and transmission network topology. This basic algorithm uses the novel concept of a "system pattern" to permit structural capacity constraints on generation and transmission to be properly accounted for in the forecasting procedure.

To handle practical data-availability concerns, a probabilistic extension of this basic algorithm is then proposed that can be implemented purely on the basis of publicly available information. The accuracy of this probabilistic algorithm relative to a more traditional GARCH statistical forecasting model is demonstrated for a NYISO case study.

Finally, a cross-scenario extension of this forecasting algorithm is proposed in which probabilities are assigned to different scenarios. This would permit forecasters to probabilistically average forecasts across distinct scenarios, thus potentially permitting longer forecast horizons and/or increased availability of pertinent historical data.

The proposed algorithm is targeted for energy-only markets; future work will consider the incorporation of ancillary services. Future work will also explore additional factors, such as possible strategic supply offer behaviors by generators. Moreover, alternative forms for the probabilistic point inclusion test, a key building block of our algorithm, will be systematically studied.

APPENDIX

Consider a wholesale power market operating over a transmission grid with N buses. Assume for simplicity that each bus i has one fixed load denoted by L_i and one generator with a real power level denoted by P_i. Suppose, also, that each generator i has a quadratic total cost function with coefficients a_i and b_i. Finally, suppose the objective of the market operator in each hour is to minimize the total system cost of meeting fixed load subject to an injection-equals-load balance constraint, transmission line flow limits, and generator operating capacity limits.

In particular, suppose the market operator attempts to achieve its objective in each hour by using the following standard DC-OPF formulation that assumes a lossless transmission system:

$$\min_{P} \quad \sum_{i=1}^{N} [a_i P_i + b_i P_i^2] \tag{10}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} P_i - \sum_{i=1}^{N} L_i = 0 : \lambda \tag{11}$$

$$\sum_{i=1}^{N} \beta_{ij}[P_i - L_i] \le \mathrm{F}_j^+ : \mu_j^+, \quad \text{for } j = 1:T \tag{12}$$

$$-\sum_{i=1}^{N} \beta_{ij}[P_i - L_i] \le \mathrm{F}_j^- : \mu_j^-, \quad \text{for } j = 1:T \tag{13}$$

$$P_i \le Cap_i^U : \sigma_i^U, \quad \text{for } i = 1:N \tag{14}$$

$$-P_i \le -Cap_i^L : \sigma_i^L, \quad \text{for } i = 1:N. \tag{15}$$

In these equations, β_{ij} denotes the *generation shift factor (GSF)* that measures the impact of 1-MW injection by generator i on transmission line j. Equality (11) represents the system balance constraint ensuring total generation matches total load. The transmission line flow limit constraints in two directions are expressed in (12) and (13). The last two inequalities (14) and (15) express each generator's upper and lower operating capacity limits.

Proposition 1: Consider the standard DC-OPF formulation with fixed loads and quadratic generator cost functions described in (10)–(15). Suppose this standard formulation is used by a market operator to determine system variable solutions. Then, conditional on any given commitment-and-line scenario S, the load space can be covered by convex polytopes such that: 1) the interior of each convex polytope corresponds to a unique system pattern and 2) within the interior of each convex polytope, the system variable solutions can be expressed as linear-affine functions of the vector of distributed loads.

Proof Outline [17]: First note that the DC-OPF formulation can equivalently be expressed in the following compact form:

$$\min_{P} \quad \frac{1}{2} P^T H P + \alpha^T P \tag{16}$$

$$\text{s.t.} \quad G_1 P = W_1 + S_1 L : \Lambda_1 \text{ and, for } i = 2 : (1 + 2N + 2T) \tag{17}$$

$$G_i P \le W_i + S_i L : \Lambda_i. \tag{18}$$

The notation in this general QP problem is described in [17]. The *KKT* first-order necessary conditions for (16)–(18) can then be expressed as follows:

$$HP + \alpha + G^T \Lambda = 0 \tag{19}$$

$$G_1 P - W_1 - S_1 L = 0 \text{ and, for } i = 2 : (1 + 2N + 2T) \tag{20}$$

$$\Lambda_i (G_i P - W_i - S_i L) = 0 \tag{21}$$

$$\Lambda_i \ge 0 \tag{22}$$

$$G_i P - W_i - S_i L \le 0. \tag{23}$$

Let $\mathcal{A}$ denote the set of indices corresponding to the active (binding) equality and inequality constraints for the DC-OPF problem. If the number of binding unit capacity constraints and line limit constraints are denoted by R and M, respectively, then Cardinality $(\mathcal{A}) = 1 + R + M$. Let $G^{\mathcal{A}}$, $W^{\mathcal{A}}$, and $S^{\mathcal{A}}$ represent the matrices corresponding to $\mathcal{A}$. Then, $G^{\mathcal{A}}$, $W^{\mathcal{A}}$, and $S^{\mathcal{A}}$ have row dimension $1 + R + M$ and column dimension N. Let $\Lambda^{\mathcal{A}}$ denote the multiplier vector corresponding to $\mathcal{A}$. Given $\mathcal{A}$, (19)–(21) reduce to

$$G^{\mathcal{A}} P - W^{\mathcal{A}} - S^{\mathcal{A}} L = 0 \tag{24}$$

$$HP + \alpha + (G^{\mathcal{A}})^T \Lambda^{\mathcal{A}} = 0. \tag{25}$$

Tøndel [14] defines the *linear independence constraint qualification (LICQ)* for an active set of constraints to be the assumption that these constraints are linearly independent. For the problem at hand, LICQ holds if $G^{\mathcal{A}}$ has full row rank. A generator that is at its upper capacity limit cannot at the same time be at its lower limit; hence, $[1\ 0\ \cdots\ 0]$ and $[-1\ 0\ \cdots\ 0]$ never co-exist. Moreover, the GSF matrix included in $G^{\mathcal{A}}$ has linearly independent rows. Thus, $\mathrm{rank}(G^A) = \min[1 + R + M, N]$. It follows that $G^{\mathcal{A}}$ has full row rank $1 + R + M$ if

$$1 + R + M \le N. \tag{26}$$

The regularity condition (26) requires that the number of binding constraints $[1 + R + M]$ does not exceed the number of decision variables N, a necessary condition for the existence of the DC-OPF problem solutions assumed to exist in Proposition 1. Consequently, (26) automatically holds under the assumptions of Proposition 1.

Given the LICQ (26) and the diagonal form of the matrix H, $G^{\mathcal{A}} H^{-1} (G^{\mathcal{A}})^T$ is invertible. Equations (24) and (25) can then be used to derive explicit solutions for $\Lambda^{\mathcal{A}}$ and P as shown in (27) and (28). Note that these solutions are linear-affine functions of

the load vector L. See equations (27)–(30) at the bottom of the page.

In summary, given a particular load vector L, explicit solutions have been derived for P and $\Lambda^{\mathcal{A}}$ as linear-affine functions of L. However, by construction, as long as the set $\mathcal{A}$ of active constraints remains unchanged in a neighborhood of the load vector L in the load space L, the linear-affine form of these solutions remains optimal. Such a neighborhood is given by the feasible region determined from (22) and (23). Substituting $\Lambda^{\mathcal{A}}$ and P from (27) and (28) into (22) and (23), one obtains inequalities (29) and (30). The load vectors L satisfying the latter inequalities are the intersection of a finite number of half-spaces in the load space and hence they form a convex polytope in this load space.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for constructive suggestions and comments that have greatly helped to improve the presentation of the forecasting algorithm and the case study findings.

REFERENCES

[1] G. Li, C. C. Liu, C. Mattson, and J. Lawarrée, "Day-ahead electricity price forecasting in a grid environment," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 266–274, Feb. 2007.
[2] F. Nogales, J. Contreras, A. Conejo, and R. Espinola, "Forecasting next day electricity prices by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, May 2002.
[3] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003.
[4] R. C. Garcia, J. Contreras, M. Akkeren, and J. B. C. Garcia, "A GARCH forecasting model to predict day-ahead electricity prices," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 867–874, May 2005.
[5] M. Shahidehpour, H. Yamin, and Z. Yi, *Market Operations In Electric Power Systems, Forecasting, Scheduling and Risk Management*. New York: Wiley-Interscience, 2002.
[6] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proc. IEEE*, vol. 88, no. 2, pp. 163–169, Feb. 2000.
[7] J. Bastian, J. Zhu, V. Banunarayanan, and R. Mukerji, "Forecasting energy prices in a competitive market," *IEEE Comput. Appl. Power*, vol. 12, no. 3, pp. 40–45, Jul. 1999.
[8] G. Li, C. C. Liu, and H. Salazar, "Forecasting transmission congestion using day-ahead shadow prices," in *Proc. IEEE PSCE 2006*, 2006, pp. 1705–1709.
[9] L. Min, S. T. Lee, P. Zhang, V. Rose, and J. Cole, "Short-term probabilistic transmission congestion forecasting," in *Proc. IEEE DPRT 2008*, Nanjing, China, Apr. 2008, pp. 764–770.
[10] R. Bo and F. Li, "Probabilistic LMP forecasting considering load uncertainty," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 911–922, Aug. 2009.
[11] F. Li and R. Bo, "Congestion and price prediction under load variation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 1279–1289, May 2009.
[12] Q. Zhou, L. Tesfatsion, and C. C. Liu, "Global sensitivity analysis for the short-term prediction of system variables," in *Proc. IEEE PES General Meeting 2010*, Minneapolis, MN, Jul. 29, 2010.
[13] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, pp. 3–20, 2002.
[14] P. Tøndel, T. A. Johansen, and A. Bemporad, "An algorithm for multi-parametric quadratic programming and explicit MPC solutions," *Automatica*, vol. 39, pp. 489–497, 2003.
[15] M. Berg, M. Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*. New York: Springer, 2000.
[16] MISO Fact Sheet. [Online]. Available: http://www.midwestiso.org/.
[17] Q. Zhou, L. Tesfatsion, and C. C. Liu, Short-Term Congestion Forecasting in Wholesale Power Markets, Working Paper No. 10025, Dept. Econ., Iowa State Univ., Jul. 2010. [Online]. Available: http://www.econ.iastate.edu/tesfatsi/CongestionForecasting.ZTL.pdf.
[18] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. New York: Springer-Verlag, 1985.
[19] C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa, "The Quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, Dec. 1996.
[20] Qhull. [Online]. Available: http://www.qhull.org/.
[21] S. S. Skiena, *The Algorithm Design Manual*. London, U.K.: Springer, 2008.
[22] NYISO Map. [Online]. Available: http://www.ferc.gov/images/marketoversight/mkt-electric/regmaps/2007_ny_elect_map.gif.
[23] NYISO Market Data. [Online]. Available: http://www.nyiso.com/public/marketsoperations/market/data/pricingdata/index.jsp.
[24] F. Rahimi, Market Metrics for DOE Congestion Study: Overview and Results. [Online]. Available: http://www.congestion09.anl.gov/documents/docs/techws/Rahimi.pdf.
[25] I. Yaniv and D. P. Foster, "Graininess of judgment under uncertainty: An accuracy-informativeness trade-off," *J. Experiment. Psychol., General 124*, pp. 424–432, 1995.

Qun Zhou (S'08) received the B.E. degree from Huazhong University of Science and Technology, Wuhan, China, in 2007. She is currently pursuing the Ph.D. degree at Iowa State University, Ames.

Her research interests focus on economic analysis of power systems, including short-term price forecasting and long-term transmission investment.

Leigh Tesfatsion (M'05) received the Ph.D. degree from the University of Minnesota, Minneapolis, in 1975.

She is a Professor of economics, mathematics and electrical and computer engineering at Iowa State University, Ames. Her principal research area is agent-based testbed development, with a particular focus on restructured electricity markets.

Dr. Tesfatsion is an active participant in IEEE PES working groups and task forces focusing on power economics issues and a co-organizer of the ISU Electric Energy Economics (E3) Group. She serves as an associate editor for a number of journals, including the *Journal of Energy Markets*.

Chen-Ching Liu (F'94) received the Ph.D. degree from the University of California, Berkeley.

He is a Professor at University College Dublin, Dublin, Ireland. From 2005 to 2008, he was Palmer Chair Professor of Electrical and Computer Engineering at Iowa State University, Ames. During 1983 to 2005, he was a Professor of electrical engineering at the University of Washington, Seattle.

Dr. Liu received an IEEE Third Millennium Medal in 2000 and the IEEE Power Engineering Society Outstanding Power Engineering Educator Award in 2004.

$$\Lambda^{\mathcal{A}} = -\left[G^{\mathcal{A}}H^{-1}(G^{\mathcal{A}})^T\right]^{-1}\left[G^{\mathcal{A}}H^{-1}\alpha + W^{\mathcal{A}} + S^{\mathcal{A}}L\right] \tag{27}$$

$$P = -H^{-1}\alpha + H^{-1}(G^{\mathcal{A}})^T\left[G^{\mathcal{A}}H^{-1}(G^{\mathcal{A}})^T\right]^{-1}\left[G^{\mathcal{A}}H^{-1}\alpha + W^{\mathcal{A}} + S^{\mathcal{A}}L\right] \tag{28}$$

$$0 \le \left(-\left[G^{\mathcal{A}}H^{-1}(G^{\mathcal{A}})^T\right]^{-1}\left[G^{\mathcal{A}}H^{-1}\alpha + W^{\mathcal{A}} + S^{\mathcal{A}}L\right]\right)_i,\ \forall i \in \frac{\mathcal{A}}{\{1\}} \tag{29}$$

$$W_i + S_iL \ge G_i\left[-H^{-1}\alpha + H^{-1}(G^{\mathcal{A}})^T\left[G^{\mathcal{A}}H^{-1}(G^{\mathcal{A}})^T\right]^{-1}\left[G^{\mathcal{A}}H^{-1}\alpha + W^{\mathcal{A}} + S^{\mathcal{A}}L\right]\right],\ \forall i \in \frac{\mathcal{A}}{\{1\}}. \tag{30}$$