Electricity Price Curve Modeling and Forecasting by Manifold Learning

Jie Chen, *Student Member, IEEE*, Shi-Jie Deng, *Senior Member, IEEE*, and Xiaoming Huo, *Senior Member, IEEE*

***Abstract*—This paper proposes a novel nonparametric approach for the modeling and analysis of electricity price curves by applying the manifold learning methodology—locally linear embedding (LLE). The prediction method based on manifold learning and reconstruction is employed to make short-term and medium-term price forecasts. Our method not only performs accurately in forecasting one-day-ahead prices, but also has a great advantage in predicting one-week-ahead and one-month-ahead prices over other methods. The forecast accuracy is demonstrated by numerical results using historical price data taken from the Eastern U.S. electric power markets.**

***Index Terms*—Electricity forward curve, electricity spot price, forecasting, locational marginal price, manifold learning.**

I. Introduction

IN the competitive electricity wholesale markets, market participants, including power generators and merchants alike, strive to maximize their profits through prudent trading and effective risk management against adverse price movements. A key to the success of market participants is to model electricity price dynamics well and capture their characteristics realistically. One strand of research on modeling electricity price processes focuses on the aspect of derivative pricing and asset valuation which investigates electricity spot and forward price models in a risk-neutral world (e.g., [1]–[3]). Another research strand concerns the modeling of electricity prices in the physical world, which offers price forecasts for assisting with physical trading and operational decision-making. An accurate short-term price forecast over a time horizon of hours helps market participants to devise their bidding strategies in the auction-based pool-type markets and to allocate generation capacity optimally among different markets. The medium-term forecast with a time horizon spanning days to months is useful for balance sheet calculations and risk management applications [4].

In the second research strand of power price modeling, there is an abundant literature on forecasting spot or short-term electricity prices, especially the day-ahead prices [5]–[12]. Typically, electricity prices are treated as hourly univariate time series and then modeled by parametric models, including ARIMA processes and their variants [6]–[8], regime-switching or hidden Markov processes [9], [12], Levy processes [11], hybrid price models combining statistical modeling with fundamental supply-demand modeling [5], or nonparametric models such as the artificial neural networks [13]–[15]. While spot price modeling is important, successful trading and risk management operations in electricity markets also require knowledge on an electricity price curve consisting of prices of electricity delivered at a sequence of future times instead of only at the spot. For instance, in order to maximize the market value of generation assets, power generators would need to base their physical trading decisions over how much power to sell in the next day and in the long-term contract markets on both the short-term price forecast for electricity delivered in the next 24 h and the electricity forward price with maturity ranging from weeks to years. The non-storable nature of electricity makes the energy delivered at different time points essentially different commodities. The current market price (or spot price) of electricity may have little correlation with that of electricity delivered a few months in the future. Thus, it is imperative to be able to model the electricity price curve as a whole.

There is not much literature on modeling electricity price curves. Paper [16] proposes a parametric forward price curve model for the Nordic market, which does not model the movements of the expected future level of a forward curve. A recent paper [17] employs a weighted average of nearest neighbors approach to model and forecast the day-ahead price curve. These works offer little insight on understanding the main drivers of the price curve dynamics. Our paper contributes to this strand of research by proposing a novel nonparametric approach for modeling electricity price curves. Analysis on the intrinsic dimension of an electricity price curve is offered, which sheds light on identifying major factors governing the price curve dynamics. The forecast accuracy of our model compares favorably against that of the ARX and ARIMA model in one-day-ahead price predictions. In addition, our model has a great advantage on the predictions in a longer horizon from days to weeks over other models.

In general, the task of analytically modeling the dynamics of such a price curve is daunting, because the curve is a high-dimensional subject. Each price point on the curve essentially represents one dimension of uncertainty. To reduce the dimension of modeling a price curve and identify the major random factors influencing the curve dynamics, principle component analysis (PCA) is proposed and has been widely applied in the real-world data analysis for industrial practices. As PCA is mainly suited for extracting the linear factors of a data set, it does not appear to perform well in fitting electricity price curves with a linear factor

Manuscript received March 14, 2007; revised April 24, 2008. This work was supported in part by the National Science Foundation under Grants 0134210, 0604736, and 0700152. Paper no. TPWRS-00186-2007.

The authors are with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 USA (e-mail: jchen@isye.gatech.edu; deng@isye.gatech.edu; xiaoming@isye.gatech.edu).

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.2008.926091

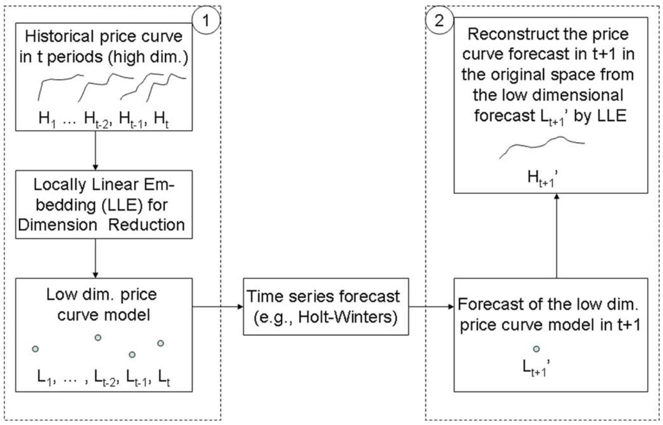

Fig. 1. Conceptual flowchart of the model.

model in a low-dimensional space. However, the following intuition suggests that there shall exist a low-dimensional structure capturing the majority of randomness in electricity price curve dynamics. Take the day-ahead electricity price curve as an example. While electricity delivered in the next 24 h are different commodities, the corresponding prices all result from equilibrating the fundamental supply and demand for electricity. The common set of demand and supply conditions in all 24 h hints a possible nonlinear representation of the 24-dimensional price curve in a space of lower dimension.

A natural extension to the PCA approach is to consider manifold learning methods, which are designed to analyze intrinsic nonlinear structures and features of high-dimensional price curves in the low-dimensional space. After obtaining the low-dimensional manifold representation of price curves, price forecasts are made by first predicting each dimension coordinate of the manifold and then utilizing a reconstruction method to map the forecasts back to the original price space. The conceptual flowchart of our modeling approach is illustrated by Fig. 1.

Our major contribution is to establish an effective approach for modeling energy forward price curves, and set up the entire framework in Fig. 1. The other major contribution is to identify the nonlinear intrinsic low-dimensional structure of price curves. The resulting analysis reveals the primary drivers of the price curve dynamics and facilitates accurate price forecasts. This work also enables the application of standard times series models such as Holt–Winters in the forecast step from box 1 to box 2.

In this paper, locally linear embedding (LLE) and LLE reconstruction are adopted for manifold learning and reconstruction. The study of the intrinsic dimension and embedded manifold indicates that there does exist a low-dimensional manifold with the intrinsic dimension around four for day-ahead electricity price curves in the New York Independent System Operator's markets (known as NYISO).

The rest of the paper is organized as follows. Section II describes a manifold based method LLE and the corresponding reconstruction method. In Section III, LLE and LLE reconstruction are applied to model and analyze the day-ahead electricity price curves in NYISO. Section IV presents the results of electricity price curve predictions based on manifold learning. Section V discusses about the extensions and restrictions of our modeling and prediction. Section VI concludes.

II. Manifold Learning Algorithm

A. Introduction to Manifold Learning

Manifold learning is a new and promising nonparametric dimension reduction approach. Many high-dimensional data sets that are encountered in real-world applications can be modeled as sets of points lying close to a low-dimensional manifold. Given a set of data points $x_1, x_2, \cdots, x_N \in \mathbb{R}^D$, we can assume that they are sampled from a manifold with noise, i.e.,

$$x_i = f(y_i) + \varepsilon_i, \quad i = 1, \cdots, N \tag{2.1}$$

where $y_i \in \mathbb{R}^d$, $d \ll D$ and ε_i is noise. Integer d is also called the intrinsic dimension. The manifold based methodology offers a way to find the embedded low-dimensional feature vectors y_i from the high-dimensional data points x_i.

Many nonparametric methods are created for nonlinear manifold learning, including multidimensional scaling (MDS) [18], [19], locally linear embedding (LLE) [20], [21], Isomap [22], Laplacian eigenmaps [23], Hessian eigenmaps [24], local tangent space alignment (LTSA) [25], and diffusion maps [26]. Survey [27] gives a review on the above methods.

Among various manifold based methods, we find that LLE works well in modeling electricity price curves. Our purpose is two-fold: to analyze the features of electricity price curves and predict the price curve at a future time. The reconstruction of high-dimensional forecasted price curves from low-dimensional predictions is a significant step for forecasting. Through extensive computational experiments, we conclude that LLE reconstruction is more efficient relative to other reconstruction methods for our purpose. Moreover, LLE and LLE reconstruction are fast and easy to implement. In next two subsections, we introduce the algorithms of LLE and LLE reconstruction, respectively.

B. Locally Linear Embedding (LLE)

Given a set of data points $x_1, x_2, \cdots, x_N \in \mathbb{R}^D$ in the high-dimensional space, we are looking for the embedded low-dimensional feature vectors $y_1, y_2, \cdots, y_N \in \mathbb{R}^d$. LLE is a nonparametric method that works as follows [20], [21].

1) Identify k nearest neighbors based on Euclidean distance for each data point x_i, $1 \le i \le N$. Let N_i denote the set of indices of the k nearest neighbors of x_i.
2) Find the optimal local convex combination of the k nearest neighbors to represent each data point x_i. That is, the following objective function (2.2) is minimized with weight coefficients w_{ij} being decision variables:

$$E(w) = \sum_{i=1}^{N} \|x_i - \sum_{j \in N_i} w_{ij} x_j\|^2 \tag{2.2}$$

where $\|\cdot\|$ is the l_2 norm and $\sum_{j \in N_i} w_{ij} = 1$.

The weight w_{ij} indicates the contribution of the jth data point to the representation of the ith data point. The optimal weights can be solved as a constrained least square problem, which is finally converted into a problem of solving a linear system of equation.

3) Find the low-dimensional feature vectors y_i, $1 \leq i \leq N$, which have the optimal local convex representations with weights w_{ij} obtained from the last step. That is, y_i's are computed by minimizing the following objective function:

$$\Phi(y) = \sum_{i=1}^{N} \left\| y_i - \sum_{j \in N_i} w_{ij} y_j \right\|^2 . \tag{2.3}$$

The problem (2.3) can be rewritten in a quadratic form

$$\Phi(y) = \sum_{ij} M_{ij}(y_i \cdot y_j) \tag{2.4}$$

involving inner products of y_i. The square $N \times N$ matrix M is given by $M_{ij} = \delta_{ij} - w_{ij} - w_{ji} + \sum_k w_{ki} w_{kj}$, where δ_{ij} is 1 if $i = j$ and 0 otherwise. To uniquely determine y_i's, we will impose the constraints $\sum_i y_i = 0$ and $(1/N) \sum_i y_i y_i^T = I$ to remove the translational and rotational degree of freedom, respectively. The quadratic optimization (2.4) with additional constraint can be solved by using the Langrange multiplier. The problem is finally converted into finding the d eigenvectors associated with the d smallest nonzero eigenvalues of matrix M, which comprise the d-dimensional coordinates of y_i's. Thus, the coordinates of y_i's are orthogonal. We refer interested readers to [20] for the details of this formulation.

LLE does not impose any probabilistic model on the data; However, it implicitly assumes the convexity of the manifold. It can be seen later that this assumption is satisfied by the electricity price data.

C. LLE Reconstruction

Given a new feature vector in the embedded low-dimensional space, the reconstruction method is used to find its counterpart in the high-dimensional space based on the calibration data set. Reconstruction accuracy is critical for the application of manifold learning in the prediction. There are a limited number of reconstruction methods in the literature. For a specific linear manifold, the reconstruction can be easily made by PCA. For a nonlinear manifold, LLE reconstruction, which is derived in the similar manner as LLE, is introduced in [20]. LTSA reconstruction and nonparametric regression reconstruction are introduced in [25]. Among all these reconstruction methods, LLE reconstruction has the best performance for electricity price data. This is an important reason for us to choose LLE and LLE reconstruction in this paper.

Suppose low-dimensional feature vectors $y_1, y_2, \cdots, y_N$ have been obtained through LLE in the previous subsection. Denote the new low-dimensional feature vector as y_0. LLE reconstruction is applied to find the approximation $\hat{x}_0$ of the original data point x_0 in the high-dimensional space based on $x_1 \cdots, x_N$ and $y_1, \cdots, y_N$. There are three steps for LLE reconstruction.

1) Identify the k nearest neighbors of the new feature vector y_0 among $y_1, \cdots, y_N$. Let N_0 denote the set of the indices of the k nearest neighbors of y_0.
2) The weights of the local optimal convex combination w_j are calculated by minimizing

$$E(w) = \left\| y_0 - \sum_{j \in N_0} w_j y_j \right\|^2 \tag{2.5}$$

subject to the sum-to-one constraint, $\sum_{j \in N_0} w_j = 1$.

3) Date point $\hat{x}_0$ is reconstructed by $\hat{x}_0 = \sum_{j \in N_0} w_j x_j$.

Remark: Solving optimization problems (2.2) and (2.5) is equivalent to solving a linear system of equations. When there are more neighbors than the high dimension or the low dimension, i.e., $k > D$ or $k > d$, the coefficient matrix associated with the system of linear equations is singular, which means that the solution is not unique. This issue is solved by adding an identity matrix multiplied with a small constant to the coefficient matrix [20]. We adopt this approach here.

Suppose $x_0^{(j)}$, $1 \leq j \leq D$, is the jth component of vector x_0. The *reconstruction error* (RE) of x_0 is defined as

$$RE(x_0) = \frac{1}{D} \sum_{j=1}^{D} \frac{\left| x_0^{(j)} - \hat{x}_0^{(j)} \right|}{x_0^{(j)}}. \tag{2.6}$$

The *reconstruction error of the entire calibration data set* (TRE)[1] is defined as

$$TRE = \frac{1}{N \times D} \sum_{i=1}^{N} \sum_{j=1}^{D} \frac{\left| x_i^{(j)} - \hat{x}_i^{(j)} \right|}{x_i^{(j)}} \tag{2.7}$$

by regarding each y_i as a new feature vector y_0.

III. Modeling of Electricity Price Curves With Manifold Learning

The data of the day-ahead market locational based marginal prices (LBMPs) and integrated real-time actual load of electricity in the Capital Zone of NYISO are collected and predicted in this paper. The data are available online (www.nyiso.com/public/market_data/pricing_data.jsp). In this section, two years (731 days) of price data from February 6, 2003 to February 5, 2005 are used as an illustration of modeling electricity price curves by manifold based methodology. Fig. 2(a) plots the hourly day-ahead LBMPs during this period, where electricity prices are treated as a univariate time series with 24×731 hourly prices. Fig. 2(b)–(d) illustrates the mean, standard deviation and skewness of 24 hourly log prices in each day after outlier processing.

The section is organized as follows. First, the data are preprocessed with log transform, outlier processing and LLP

[1]When the TRE is calculated, y_i itself is not included in its k nearest neighbors.

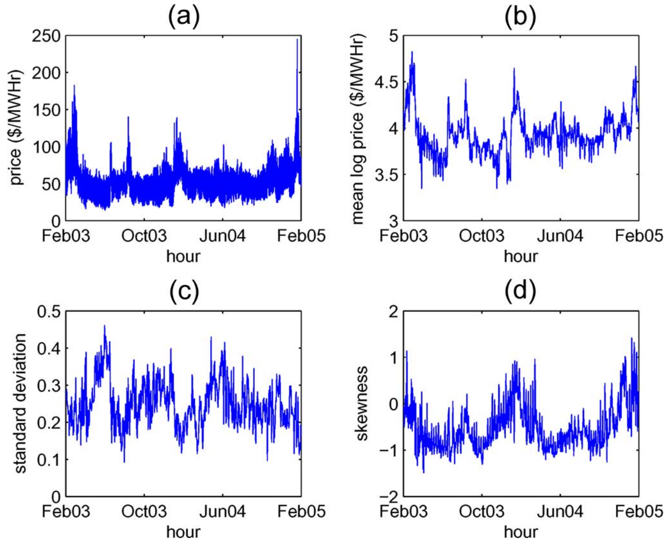

Fig. 2. Day-ahead LBMPs from February 6, 2003 to February 5, 2005 in the Capital Zone of NYISO. (a) Hourly prices. (b) Mean of log prices in each day. (c) Standard deviation of log prices in each day. (d) Skewness of log prices in each day.

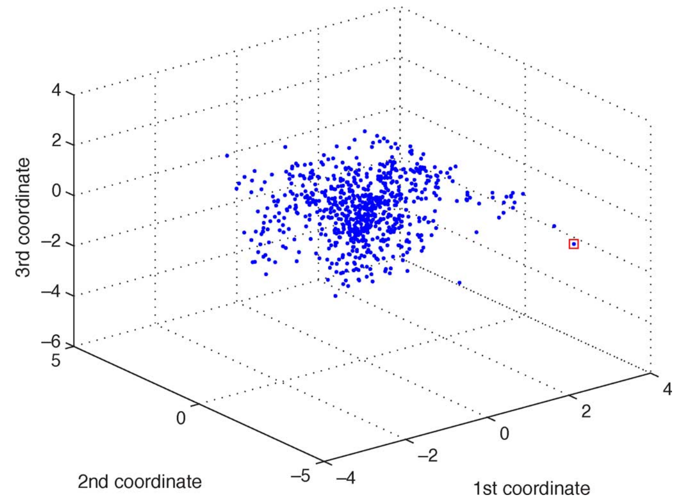

Fig. 3. Embedded three-dimensional manifold without any outlier preprocessing (but with log transform and LLP smoothing). The square indicates the day with outliers—January 24, 2005.

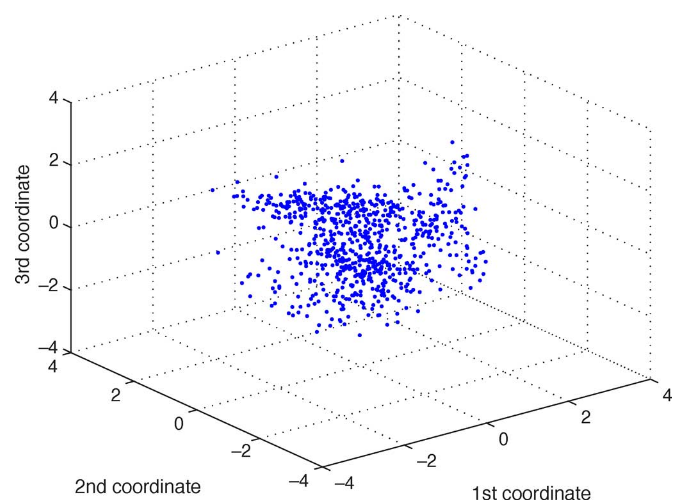

Fig. 4. Embedded three-dimensional manifold after log transform, outlier preprocessing, and LLP smoothing.

smoothing, and then the results of the manifold learning and reconstruction are illustrated. Next, the major factors of electricity price curve dynamics are analyzed with low-dimensional feature vectors. Finally, the parameter selections and the sensitivity of reconstruction error to those parameters are analyzed.

A. Preprocessing

1) Log Transform: The logarithmic (log) transforms of electricity prices are taken before the manifold learning. There are several advantages to deal with the log prices. First, electricity prices are well known to have the nonconstant variance, and log transform can make the prices less volatile. The log transform also enhances the efficiency of manifolding learning, by making the embedded manifold more uniformly distributed in the low-dimensional space and the *reconstruction error of the entire calibration data set* (TRE) reduced. Moreover, the log transform has the interpretation of the returns to someone holding the asset.

2) Outlier Processing: Outliers in this paper are defined as the electricity price spikes that are extremely different from the prices in the neighborhood. To deal with the outliers, we replace the prices in the day with outliers by the average of the prices in the days right before and right after. We remove the outliers because the embedded low-dimensional manifold is supposed to extract the primary features of the entire data set, rather than the individual and local features such as extreme price spikes. The efficiency of manifold learning is improved after outlier processing. Moreover, outliers, which represent rarely occurring phenomena in the past, often have very small probability to occur in the near future, so the processing of outliers does not severely affect the prediction of the near-term regular prices.

In the illustrated data set, only one extreme spike is identified on the right of Fig. 2(a), which belongs to January 24, 2005. In the low-dimensional manifold, the days of outliers can also be detected by the points that stand far away from the other points. Fig. 3 shows that the point corresponding to January 24, 2005 lies out of the main cloud of the points on the embedded three-dimensional manifold. Thus, we regard January 24, 2005 as a day with outliers. Fig. 4 shows that the low-dimensional manifold after removing the outliers is more uniformly distributed.

3) LLP Smoothing: The noise in (2.1) can contaminate the learning of the embedded manifold and the estimation of the intrinsic dimension. Therefore, locally linear projection (LLP) [27]–[29] is recommended to smooth the manifold and reduce the noise. The description of the algorithm is given as follows

---

ALGORITHM: LLP

For each observation x_i, $i = 1, 2, \cdots, N$,

1) Find the k-nearest neighbors of x_i. The neighbors are denoted by $\widetilde{x}_1, \widetilde{x}_2, \cdots, \widetilde{x}_k$.
2) Use PCA or SVD to identify the linear subspace that contains most of the information in the vectors $\widetilde{x}_1, \widetilde{x}_2, \cdots, \widetilde{x}_k$. Suppose the linear subspace is A_i. Let k_0

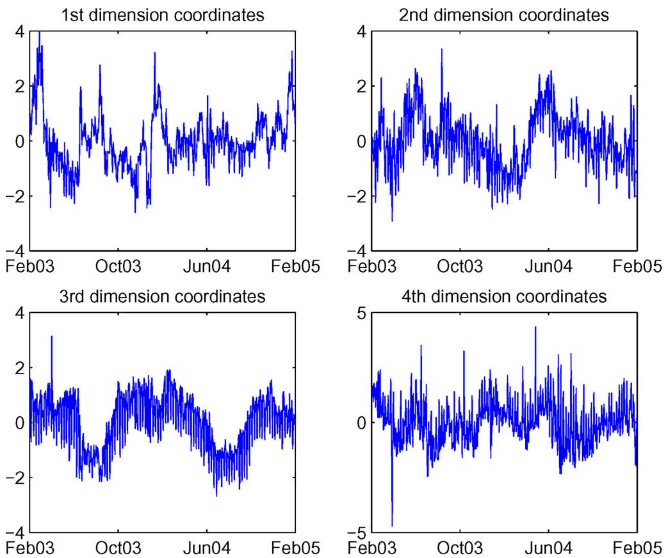

Fig. 5. Coordinates of the embedded four-dimensional manifold.

denote the assumed dimension of the embedded manifold. Then subspace A_i can be viewed as a linear subspace spanned by the singular vectors associated with the largest k_0 singular values.

3) Project x_i into the linear subspace A_i and let $\check{x}_i$, $i = 1, \cdots, N$, denote the projected points.

After denoising, the efficiency of manifold learning is enhanced, and the TRE is reduced. For the illustrated data set with the intrinsic dimension being four, the TRE is 3.89% after LLP smoothing, compared to 4.41% without LLP smoothing. The choice of the two parameters in LLP, the dimension of the linear space and the number of the nearest neighbors, will be discussed in detail in Section III-D.

B. Manifold Learning by LLE

Each price curve with 24 hourly prices in a day is considered as an observation, so the dimension of the high-dimensional space D is 24. The intrinsic dimension d is set to be four. The number of the nearest neighbors k for LLP smoothing, LLE, and LLE reconstruction is selected to be a common number 23 for all the numerical studies. The details of the parameter selections are discussed in Section III-D. Due to the ease of visualization in a three-dimensional space, all the low-dimensional manifolds are plotted with the intrinsic dimension being three. We apply LLE to the denoised data $\check{x}_i$, $i = 1, \cdots, N$, which are obtained after LLP smoothing. Fig. 4 provides the plot of the embedded three-dimensional manifold. As the low-dimensional manifold is nearly convex and uniformly distributed, LLE is an appropriate manifold based method. Fig. 5 plots the time series of each coordinates of the feature vectors in the embedded four-dimensional manifold.

Table I shows the TRE of different reconstruction methods. LLE reconstruction has the minimum reconstruction error among all the methods. LTSA reconstruction has a very large TRE, because it is an extrapolation-like method, and the reconstruction of some of the price curves has very large errors.

TABLE I
TRE OF DIFFERENT RECONSTRUCTION METHODS

Reconstruction method	TRE(%)
LLE and LLE reconstruction	3.89
PCA and PCA reconstruction	4.26
LTSA and LTSA reconstruction	4.55×10^6
LLE and nonparametric regression reconstruction	4.77

Therefore, LLE and LLE reconstruction are selected to model electricity price dynamics.

C. Analysis of Major Factors of Electricity Price Curve Dynamics With Low-Dimensional Feature Vectors

The interpretation of each dimension in the low-dimensional space and the cluster analysis to the low-dimensional feature vectors reveal the major drivers of the price curve dynamics, which suggests that our prediction methods in the next section based on the modeling of price curves with manifold learning are reasonable.

1) Interpretation of Each Dimension in the Low-Dimensional Space: There are some interesting interpretations for the first three coordinates of the feature vectors in the low-dimensional space. For each price curve, we can calculate the mean, standard deviation, range, skewness and kurtosis of the 24 hourly log prices. The sequence of each coordinates of the low-dimensional feature vectors comprises a time series. The correlation between each time series and mean log prices (standard deviation, range, skewness and kurtosis) is calculated. Table II shows which one of the four-dimensional coordinates has the maximum absolute correlation with the statistics of log prices (mean, standard deviation, range, skewness and kurtosis), and the corresponding correlation coefficients. The comparison between Figs. 2 and 5 gives more intuition about the correlations. It is found that the first coordinates have a very high correlation coefficient 0.9964 with the mean log prices within each day, and the second coordinates are highly correlated with the standard deviation of the log prices in a day with a correlation coefficient 0.7073. This also means that the second coordinates contain some other information besides standard deviation, and Table II demonstrates that the second coordinates are also correlated, but not significantly, with range and skewness. The third coordinates show both weekly and yearly seasonality in Fig. 5. Weekly seasonality is well known for electricity prices. Yearly seasonality may be caused by the shape change of the price curves over the year. The shape of price curves is often unimodal in the summer and bimodal in the winter.

2) Cluster Analysis: The yearly seasonality of electricity price curves can be clearly demonstrated by the cluster analysis of low-dimensional feature vectors.

Cluster analysis [30] (also known as data segmentation) groups or segments a collection of objects into subsets (i.e., clusters), such that those within each cluster are more closely related to each other than those assigned to different clusters.

The K-means clustering algorithm is one of the mostly used iterative clustering methods. Assume that there are K clusters. The algorithm begins with a guess of the K cluster centers.

TABLE II
ONE OF THE FOUR-DIMENSIONAL COORDINATES WHICH HAS THE MAXIMUM ABSOLUTE CORRELATION COEFFICIENT WITH THE MEAN (STANDARD DEVIATION, RANGE, SKEWNESS, AND KURTOSIS) OF LOG PRICES IN A DAY IN EMBEDDED FOUR-DIMENSIONAL SPACE

	Mean	Standard Deviation	Range	Skewness	Kurtosis
Coordinate	1st	2nd	2nd	2nd	3rd
Correlation Coefficient	0.9964	0.7073	0.5141	-0.5646	0.2611

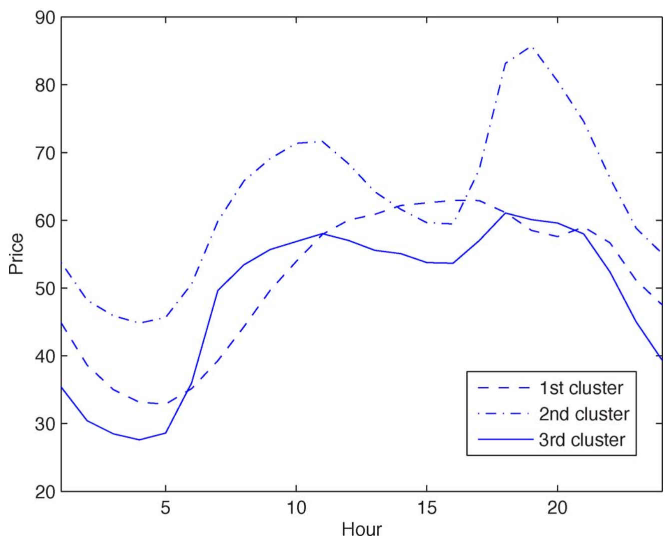

Fig. 6. Coordinate-wise average of the actual price curves in each cluster, where clustering is based on low-dimensional feature vectors.

Then, the algorithm iterates between the following two steps until convergence. The first step is to identify the closest cluster center for each data point based on some distance metric. The second step is to replace each cluster center with the coordinate-wise average of all the data points that are the closest to it.

For electricity price data, we apply K-means clustering with Euclidean distance to the low-dimensional feature vectors of the embedded four-dimensional manifold. The number of clusters is set to be three, as the yearly seasonality can be clearly illustrated with three clusters. The coordinate-wise average of price curves in each cluster is plotted in Fig. 6. The distribution of clusters is illustrated in the first graph of Fig. 7, where x axis is the date of the price curves, and y axis is the corresponding clusters. The two graphs show that the first cluster represents the price curves from the summer, which are featured with unimodal shape, and the second cluster represents the ones from the winter, which are characterized with bimodal shape. The price curves in the third cluster reveal the transition from unimodal shape to bimodal shape. The average price curves in the three clusters closely resemble the typical load shapes observed in summer, winter, and rest-of-year, respectively.

The second graph of Fig. 7 shows the distribution of clusters by applying K-means clustering with correlation distance to the high-dimensional price curves. The two graphs in Fig. 7 have the similar patterns, which gives a good illustration that low-dimensional feature vectors capture the major factors of the price curve dynamics.

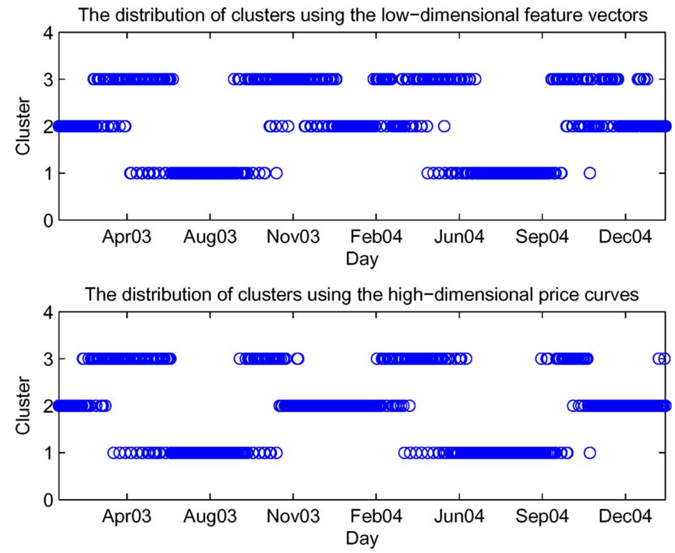

Fig. 7. Distribution of clusters.

D. Parameter Setting and Sensitivity Analysis

The selections of several parameters, including the number of intrinsic dimensions, the number of the nearest neighbors and the length of the calibration data, are discussed in this subsection.

1) Intrinsic Dimension: Intrinsic dimension d is an important parameter of manifold learning. Papers [31] and [32] provide several approaches of estimating the intrinsic dimension. In [31], the maximum likelihood estimator of the intrinsic dimension is established. In [32], the intrinsic dimension is estimated based on a nearest neighbor algorithm. Without LLP smoothing, the two methods show that the intrinsic dimension is some value between 4 and 5. Thus, it is reasonable to set the dimension of the linear space as 4 in LLP smoothing. After LLP smoothing, the intrinsic dimension is reduced to a value between 3 and 4. The numerical experiments indicate that LLP smoothing can not only denoise, but also improve the efficiency of estimating the intrinsic dimension.

Another empirical way of estimating the intrinsic dimension is to analyze the sensitivity of the TRE to the different values of the intrinsic dimension. Fig. 8 shows that the TRE is a decreasing function of the intrinsic dimension with a increasing slope. The slope of the curve in the figure has a dramatic change when the intrinsic dimension is around four. Therefore, we choose the intrinsic dimension as four in the paper.

2) Number of the Nearest Neighbors: The plot of the TRE against the number of the nearest neighbors is used to select the appropriate number of the nearest neighbors. Fig. 9 indicates the TRE first falls steeply when the number of the nearest neighbors is small, and then remains steady when the number of the nearest neighbors is greater than 22. We set the number of the nearest neighbors to be 23 for all the numerical studies. This is only one of the many choices as the reconstruction error is not sensitive to the number of the nearest neighbors within a range.

3) Length of the Calibration Data: The plot of the TRE against the length of the calibration data in Fig. 10 illustrates that the TRE is not very sensitive to the data length. Two years

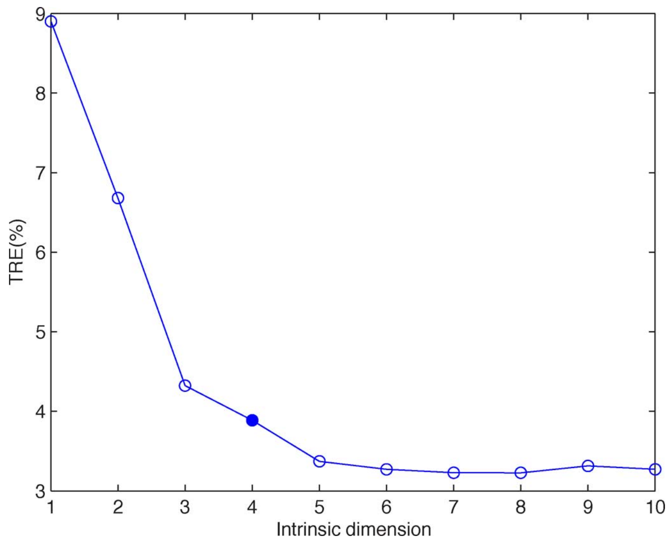

Fig. 8. Sensitivity of TRE to the intrinsic dimension (data length $= 731$ days, number of the nearest neighbors $= 23$).

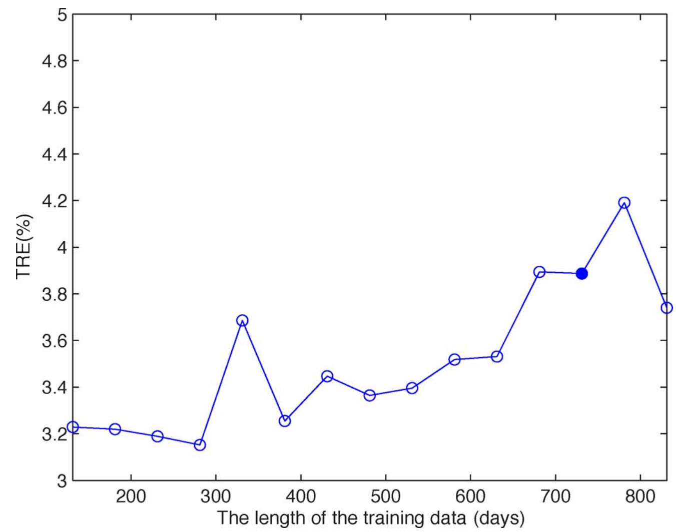

Fig. 10. Sensitivity of TRE to the length of the calibration data (intrinsic dimension $= 4$, number of the nearest neighbors $= 23$).

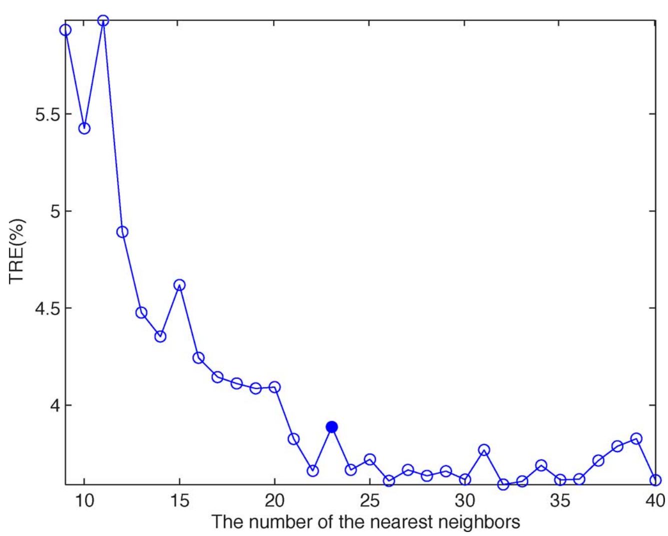

Fig. 9. Sensitivity of TRE to the number of the nearest neighbors (data length $= 731$ days, intrinsic dimension $= 4$).

of data are applied to the manifold learning, and it helps to study whether there is yearly seasonality.

IV. Prediction of Electricity Price Curves

The prediction of future electricity price curves is an important issue in electricity market related research, because accurate predictions enable market participants to earn stable profits by trading energy and hedging undesirable risks successfully. However, it is difficult to make accurate predictions for electricity prices due to their multiple seasonalities—daily and weekly seasonality. The speciality of the electricity price data often results in complicated models to forecast future electricity prices, which are often overfitting and fail to make accurate predictions in a longer horizon. Our method converts the hourly electricity price time series with multiple seasonalities into several time series with only weekly seasonality by manifold learning. After conversion, each data point in the new time series represents a day rather than an hour. The simplification of the new time series makes the longer horizon prediction easier and more accurate. Therefore, our method has an advantage in the longer horizon prediction over many other prediction methods.

A large amount of existing forecasting methods focus on one-day-ahead price predictions, i.e., the horizon of prediction is one day (24 h). Two articles [4] and [33] give a good review on many prediction methods, and make a comparison on their performance. In our paper, we compare our prediction methods with three models—ARIMA, ARX and the naive method. The ARIMA model [7] and the naive method are pure time series methods. The ARX model (also called dynamic regression model) includes the explanatory variable, load, and is suggested to be the best model in [33] and one of the best models in [4].

The longer horizon prediction has not drawn much attention so far. However, it also plays an important role in biding strategy and risk management. Our numerical results show that our prediction methods not only generate competent results in forecasting one-day-ahead price curves, but also produce more accurate predictions for one-week-ahead and one-month-ahead price curves, compared to ARX, ARIMA and the naive method. Moreover, as the new time series generated by manifold learning are simple, it is very easy to identify the time series models or utilize some nonparametric forecasting techniques. Our prediction methods also allow larger size of data for model calibration and incorporate more past information, but the size of the calibration data for ARIMA and ARX is often restricted to be several months.

A. Prediction Method

In our prediction method, we first make the prediction in the low-dimensional space, and then reconstruct the predicted price curves in the high-dimensional space from the low-dimensional prediction. There are three steps in detail.

1) Learn the low-dimensional manifold of electricity price curves with LLE. The sequence of each coordinates of the low-dimensional feature vectors comprises a time series.

2) Predict each time series in the low-dimensional space via univariate time series forecasting. Three prediction methods are applied: the Holt–Winters algorithm (HW) [34], the structural model (STR) [34] and the seasonal decomposition of time series by loess (STL) [35]. Each data point in the time series represents one day, so for the one-week-ahead (one-day-ahead or one-month-ahead) price curve predictions, seven (one or 28) data points are forecasted for each time series.
3) Reconstruct the predicted price curves in the high-dimensional space from the predictions in low-dimensional space with LLE reconstruction.

The first and third step have been described in the previous sections. In the second step, we make the univariate time series forecasting for each coordinates of the feature vectors rather than making the multivariate time series forecasting for all the time series in the low-dimensional space, because the coordinates are orthogonal to each other.

There are a variety of methods of univariate time series forecasting, among which Holt–Winters algorithm, structural model and STL are selected. Both the Holt–Winters algorithm and structural model are pure time series prediction methods (models), and do not require any model identification as in ARIMA. The STL method can involve the explanatory variable in the prediction. All the prediction methods can be easily and fast implemented in statistical software R. The following is some brief description of the three prediction methods.

1) Holt–Winters Algorithm (HW): In Holt–Winters filtering, seasonals and trends are computed by exponentially weighted moving averages. In our numerical experiments, Holt–Winters algorithm is executed with starting period equal to seven days and 14 days, respectively. This choice is due to the weekly effect of electricity prices.

2) Structural Models (STR): Structural time series model is a (linear Gaussian) state-space model for (univariate) time series based on a decomposition of the series into a number of components—trend, seasonal and noise.

3) Seasonal Decomposition of Time Series by Loess (STL): The STL method can involve explanatory variables in the prediction. As the effect of temperature is usually embodied in electricity loads, only load is utilized as an exploratory variable. We first learn the manifold with the intrinsic dimension four for both prices and loads, and then decompose each time series in the low-dimensional space of price and load curves into seasonal, trend and irregular components using loess. Let $P_{i,t}$ and $Z_{i,t}$ denote the trend[2] of the ith coordinates of the feature vectors for prices and loads at time t. Then, we regress $P_{i,t}$ on $Z_{i,t}$ and the lagged $P_{i,t}$ with the lag three. As the relationship between prices and loads are dynamic, the history data we applied to train the model are 70 days. The model is written as

$$P_{i,t} = \beta_0 + \beta_1 Z_{i,t} + \beta_2 P_{i,t-1} + \beta_3 P_{i,t-2} + \beta_4 P_{i,t-3} + \varepsilon_t.$$

B. Definition of Weekly Average Prediction Error

To assess the predictive accuracy of our methodology, three weekly average prediction errors are defined for one-day-ahead, one-week-ahead and one-month-ahead price predictions, respectively. First, we give a general definition, the weekly average p-day-ahead prediction error.

For the ith day of a certain week, $i = 1, \cdots, 7$, the calibration data are set to be the two-year data right before this day, and then p-day-ahead predictions are made, i.e., the horizon of the prediction is p days. The jth-day-ahead predictions are denoted as a 24-dimensional vector $\hat{x}_{(i,j)}$, $j = 1, \cdots, p$. The p-day-ahead prediction error for the ith day is defined as

$$\mathrm{WPE}^{(i)} = \frac{1}{p \times 24} \sum_{j=1}^{p} \frac{\left\| x_{(i,j)} - \hat{x}_{(i,j)} \right\|_1}{\bar{x}^v_{(i)}}$$

where $x_{(i,j)}$ is the actual electricity price curve corresponding to $\hat{x}_{(i,j)}$, and $\bar{x}^v_{(i)}$ is the average of the corresponding actual electricity prices of the p-day-ahead predictions for the ith day. $\|\cdot\|_1$ is the L_1 norm of a vector, which is the sum of the absolute values of all the components in the vector.

The weekly average p-day-ahead prediction error is defined as

$$\mathrm{WPE} = \frac{1}{7} \sum_{i=1}^{7} \mathrm{WPE}^{(i)}.$$

The weekly average one-day-ahead prediction error WPE_d, weekly average one-week-ahead prediction error WPE_w and weekly average one-month-ahead prediction error WPE_m are defined for $p = 1, 7, 28$, respectively.

We define σ_d, σ_w and σ_m as the standard deviations of $\mathrm{WPE}^{(i)}$, $i = 1, \cdots, 7$, for $p = 1, 7, 28$, respectively.

C. Prediction of Electricity Price Curves

Our numerical experiments are based on 12 weeks from February 2005 to January 2006, which consist of the second week of each month. Three weekly average prediction errors as defined above are calculated for each week, respectively. For each data set, the same parameter values taken from the previous section are used. The number of the nearest neighbors and the intrinsic dimension are set to be 23 and 4, respectively. Only one day, January 24, 2005, is identified with outliers. As we only have the forecasts of loads for six future days from the NYISO website, the weekly average one-week-ahead prediction error for STL and ARX is actually the weekly average six-days-ahead prediction error.

Tables III and IV provides the weekly average one-day-ahead prediction errors for the 12 weeks and their standard deviations. Our prediction methods—Holt–Winters, structural model and STL—are compared with ARX, ARIMA and the naive method. The details of the ARIMA and ARX model are in Appendices A and B. The naive predictions of a certain week are given by the actual prices of the previous week. Holt–Winters and structural model outperform all the other methods. It seems that involving the exploratory variable does not necessarily improve the prediction accuracy. STL performs slightly worse than Holt–Winters and structural model, and ARX also has less accuracy than ARIMA. This is not consistent with the results in [33] and [4], where ARX has better performance than ARIMA. A potential

[2]trend window $= 5$

TABLE III
COMPARISON OF $\mathrm{WPE}_d(\%)$ OF ONE-DAY-AHEAD PREDICTIONS FOR 12 WEEKS

date	HW7*	HW14	STR	STL	ARIMA	ARX	Naive
02/06/05 ~ 02/12/05	7.14	6.97	7.34	7.07	7.94	6.58	15.99
03/06/05 ~ 03/12/05	6.29	5.82	5.48	6.03	5.66	8.02	9.81
04/03/05 ~ 04/09/05	6.54	7.11	6.59	7.73	6.10	6.25	12.38
05/08/05 ~ 05/14/05	6.45	6.00	6.22	7.39	7.47	6.17	5.89
06/05/05 ~ 06/11/05	9.87	9.38	9.86	9.01	9.85	11.95	31.46
07/03/05 ~ 07/09/05	7.79	8.46	7.55	7.07	7.38	6.06	17.90
08/07/05 ~ 08/13/05	5.16	5.17	5.42	8.56	6.05	7.03	13.09
09/04/05 ~ 09/10/05	6.98	8.14	7.55	7.75	7.20	5.58	14.55
10/02/05 ~ 10/08/05	6.15	6.08	6.45	6.63	6.37	6.36	9.64
11/06/05 ~ 11/12/05	6.71	6.65	6.11	6.30	5.91	5.41	18.78
12/04/05 ~ 12/10/05	8.66	8.84	8.96	7.95	8.47	12.75	26.95
01/08/06 ~ 01/14/06	8.63	8.72	8.26	9.28	10.49	8.17	15.61
mean	7.20	7.28	7.15	7.56	7.41	7.53	16.00

*HW7 and HW14 stand for Holt–Winter algorithm with starting period equal to seven days and 14 days, respectively.

TABLE IV
COMPARISON OF $\sigma_d(\%)$ OF ONE-DAY-AHEAD PREDICTIONS FOR 12 WEEKS

date	HW7	HW14	STR	STL	ARIMA	ARX	Naive
02/06/05 ~ 02/12/05	3.03	3.34	3.10	2.21	4.58	2.28	4.66
03/06/05 ~ 03/12/05	1.89	1.96	1.81	2.21	2.16	3.15	5.67
04/03/05 ~ 04/09/05	2.24	2.88	2.34	2.76	1.83	3.13	3.33
05/08/05 ~ 05/14/05	3.86	3.36	3.47	3.34	4.07	2.54	1.40
06/05/05 ~ 06/11/05	3.31	3.36	4.03	2.53	4.31	6.13	5.62
07/03/05 ~ 07/09/05	3.62	4.68	3.73	4.01	3.37	2.15	8.79
08/07/05 ~ 08/13/05	1.65	1.53	2.28	3.59	3.01	2.63	4.37
09/04/05 ~ 09/10/05	3.68	3.59	3.09	3.83	4.48	2.16	8.38
10/02/05 ~ 10/08/05	2.27	2.02	2.26	3.36	2.18	3.06	4.45
11/06/05 ~ 11/12/05	2.57	2.41	2.60	2.57	2.36	3.08	7.06
12/04/05 ~ 12/10/05	4.14	4.27	3.69	2.36	2.80	6.62	10.55
01/08/06 ~ 01/14/06	4.61	4.42	3.73	3.93	5.79	4.49	4.71
mean	3.07	3.15	3.01	3.06	3.41	3.45	5.75

TABLE V
COMPARISON OF $\mathrm{WPE}_w(\%)$ OF ONE-WEEK-AHEAD PREDICTIONS FOR 12 WEEKS

date	HW7	HW14	STR	STL	ARIMA	ARX	Naive
02/06/05 ~ 02/12/05	8.31	8.33	7.65	8.28	14.57	9.00	12.19
03/06/05 ~ 03/12/05	10.85	10.59	10.19	12.57	13.28	15.30	12.33
04/03/05 ~ 04/09/05	11.03	10.88	10.53	11.74	10.68	11.46	14.59
05/08/05 ~ 05/14/05	7.55	7.15	7.36	6.89	14.21	6.16	6.71
06/05/05 ~ 06/11/05	15.58	15.23	13.88	11.39	21.64	19.49	26.68
07/03/05 ~ 07/09/05	10.85	10.04	10.41	9.60	14.63	7.03	14.44
08/07/05 ~ 08/13/05	6.82	7.09	6.42	12.87	9.49	9.14	10.28
09/04/05 ~ 09/10/05	7.46	7.95	7.31	8.52	10.36	6.86	13.17
10/02/05 ~ 10/08/05	9.78	9.60	10.11	8.97	11.84	8.30	11.57
11/06/05 ~ 11/12/05	9.45	9.44	8.99	9.42	11.24	9.00	15.18
12/04/05 ~ 12/10/05	12.94	13.09	13.30	11.55	21.78	22.92	23.94
01/08/06 ~ 01/14/06	13.95	14.08	13.28	15.00	26.01	16.37	11.52
mean	10.38	10.29	9.95	10.57	14.98	11.75	14.38

TABLE VI
COMPARISON OF $\sigma_w(\%)$ OF ONE-WEEK-AHEAD PREDICTIONS FOR 12 WEEKS

date	HW7	HW14	STR	STL	ARIMA	ARX	Naive
02/06/05 ~ 02/12/05	1.32	1.66	1.14	1.85	7.06	2.62	2.35
03/06/05 ~ 03/12/05	3.80	3.73	3.86	3.88	7.06	2.22	1.18
04/03/05 ~ 04/09/05	3.12	3.48	3.41	2.90	3.50	4.18	2.47
05/08/05 ~ 05/14/05	3.45	3.13	3.50	1.99	6.28	1.82	0.88
06/05/05 ~ 06/11/05	5.91	6.10	6.46	4.64	10.30	8.18	4.75
07/03/05 ~ 07/09/05	3.10	3.44	3.43	1.94	5.51	0.55	1.50
08/07/05 ~ 08/13/05	1.02	0.95	1.40	3.10	3.77	2.60	1.48
09/04/05 ~ 09/10/05	3.19	2.96	2.33	2.75	4.88	1.17	2.28
10/02/05 ~ 10/08/05	2.06	1.73	2.01	4.32	5.50	3.11	1.34
11/06/05 ~ 11/12/05	2.17	2.42	2.30	2.40	3.92	2.14	3.05
12/04/05 ~ 12/10/05	4.88	4.41	5.17	4.53	13.77	4.47	4.65
01/08/06 ~ 01/14/06	3.25	3.37	2.15	6.15	11.60	3.70	2.66
mean	3.11	3.12	3.10	3.37	6.93	3.06	2.38

cause is that the predictions of loads are not precise, or the correlation between loads and prices is not high enough in NYISO.

In Tables V and VI, the weekly average one-week-ahead prediction errors for the 12 weeks and their standard deviations are presented. All of our prediction methods outperform ARX, ARIMA, and the naive method. The ARIMA model acts even worse than the naive method for one-week-ahead predictions. Since the ARIMA model is a very complicated model with multiple seasonalities, it is often overfitting and makes the longer horizon predictions less accurate. The ARX model is a little simpler and given more information by the load forecasts, so it performs better than ARIMA. However, both ARX and ARIMA need to predict 168 data points for one-week-ahead predictions, while our prediction methods only need to predict seven data points for each time series. Therefore, our prediction methods have a great advantage in the longer horizon predictions. Among Holt–Winters, structural model and STL, STL has slightly worse performance than other two, and structural model is the most accurate.

The proposed method can be applied to forecast prices in a longer horizon than one week, e.g., two weeks or even one month. As there are only a few methods associated with one-month-ahead price predictions, we apply three naive methods to compare with. The first naive method takes the last month prices in the calibration data set as the predictions. The second method repeats the last week prices four times, and the third one replicates the prices of last two weeks twice, respectively, as the predictions. Tables VII and VIII provide the weekly average prediction errors of the one-month-ahead price predictions for the 12 weeks and their standard deviations. The notations—naive1, naive2 and naive3—stand for the three naive methods. From the comparison, the proposed methods outperform all the naive methods. We notice that the total stand deviation of the structural model is larger than that of the naive methods, and it is mainly due to an inaccurate prediction for one day in week five. Thus, Holt–Winters algorithm has the best performance among all the methods for one-month-ahead price predictions.

In summary, our prediction methods without a exploratory variable—Holt–Winters and structural model—outperform all

TABLE VII
COMPARISON OF $\mathrm{WPE}_m(\%)$ OF ONE-MONTH-AHEAD PREDICTIONS FOR 12 WEEKS

date	HW7	HW14	STR	Naive1	Naive2	Naive3
02/06/05 ~ 02/12/05	8.87	9.32	9.42	29.63	12.07	25.72
03/06/05 ~ 03/12/05	12.52	12.07	12.16	17.17	13.71	14.39
04/03/05 ~ 04/09/05	17.62	17.28	16.51	14.96	15.96	14.20
05/08/05 ~ 05/14/05	12.00	11.81	11.93	13.62	11.52	12.04
06/05/05 ~ 06/11/05	16.67	16.71	25.48	24.37	23.93	27.55
07/03/05 ~ 07/09/05	15.88	15.58	16.72	16.99	13.63	13.91
08/07/05 ~ 08/13/05	10.70	10.69	11.36	12.96	12.84	15.21
09/04/05 ~ 09/10/05	14.53	14.21	13.74	19.48	15.91	18.95
10/02/05 ~ 10/08/05	18.22	18.40	19.96	19.70	17.10	19.67
11/06/05 ~ 11/12/05	13.98	13.79	14.58	31.93	18.35	28.49
12/04/05 ~ 12/10/05	18.99	18.87	19.48	30.00	26.17	27.64
01/08/06 ~ 01/14/06	13.08	13.11	12.04	31.80	14.27	14.99
mean	14.42	14.32	15.28	21.88	16.29	19.40

TABLE VIII
COMPARISON OF $\sigma_m(\%)$ OF ONE-MONTH-AHEAD PREDICTIONS FOR 12 WEEKS

date	HW7	HW14	STR	Naive1	Naive2	Naive3
02/06/05 ~ 02/12/05	1.13	1.55	2.63	0.18	2.79	6.82
03/06/05 ~ 03/12/05	3.64	3.33	4.11	0.47	0.46	1.01
04/03/05 ~ 04/09/05	3.02	2.06	3.49	0.46	2.25	1.08
05/08/05 ~ 05/14/05	1.22	1.18	1.22	0.40	1.38	0.65
06/05/05 ~ 06/11/05	4.59	4.28	25.41	0.36	5.96	0.62
07/03/05 ~ 07/09/05	3.95	4.50	3.58	1.15	2.90	0.89
08/07/05 ~ 08/13/05	0.35	0.36	1.03	0.62	1.24	0.58
09/04/05 ~ 09/10/05	3.22	3.82	2.82	0.92	0.50	0.46
10/02/05 ~ 10/08/05	4.20	4.01	6.80	0.80	2.77	1.95
11/06/05 ~ 11/12/05	1.09	1.21	1.62	1.51	3.47	3.77
12/04/05 ~ 12/10/05	2.40	2.21	3.29	0.54	4.33	1.82
01/08/06 ~ 01/14/06	6.04	6.25	3.44	4.16	1.83	1.76
mean	2.90	2.90	4.95	0.96	2.49	1.78

of ARX, ARIMA and the naive method in both one-day-ahead and one-week-ahead predictions. STL is competent with ARX and ARIMA in one-day-ahead predictions, and performs better in one-week-ahead predictions. Our prediction methods have a great advantage in the longer horizon predictions spanning days to weeks.

While the numerical experiments reported in the paper are conducted with the day-ahead electricity prices in the Capital Zone in NYISO, numerical studies have been done with price signals in other zones in NYISO and demonstrate the same kinds of comparative advantages of our method.

V. DISCUSSION OF MODELING AND PREDICTION

In this section, we discuss some extensions of our modeling and prediction of electricity price curves. The restriction of our method is also discussed.

A. Modeling and Prediction With New Historical Price Curves

It is not necessary to build a new model whenever new historical price curves are coming. Denote the new historical price curve as x_0. The procedure of computing the low-dimensional feature vector of x_0 is as follows.

1) Identify the k nearest neighbors of the new data point x_0 among $x_1, \cdots, x_N$. Let N_0 denote the set of the indices of the k nearest neighbors of x_0.
2) Compute the linear weights w_j which best reconstruct x_0 from its neighbors, i.e., minimize the following objective function:

$$E(w) = \left\| x_0 - \sum_{j \in N_0} w_j x_j \right\|^2 \tag{5.8}$$

subject to the sum-to-one constraint, $\sum_{j \in N_0} w_j = 1$.
3) The low-dimensional feature vector y_0 is computed by $y_0 = \sum_{j \in N_0} w_j y_j$.

For the prediction of each dimension in the low-dimensional space, the original prediction models can still be employed. Therefore, our modeling and prediction of electricity price curves can be utilized online for real forecasting.

B. Weekday and Weekend Effect

Electricity price has different daily profiles, in particular, weekdays verses weekend. To detect the significance of the weekday and weekend effect, we can add the dummy variables, e.g., Saturday and Sunday, into our prediction method in the same fashion as electricity loads. We did the numerical experiments, but the prediction results are almost the same as those without the dummy variables. The reason may be that the effect of weekdays and weekend is mostly captured by the weekly seasonal and the effect of electricity loads.

C. Effects of Other Factors, e.g., Hurricane Events and Higher Prices for Natural Gas

Our prediction method can be extended to incorporate other factors which affect the price curve dynamics. For the irregularly occurring event, e.g., hurricane events, the invention analysis can be considered in the prediction in low-dimensional space. For the effect of natural gas price becoming higher, our prediction method can incorporate natural gas price as a modeling factor in the same manner as it does with electricity load. Exploration along these directions is left for future research.

D. Restriction of Our Method

Our model captures the price spike aspect of electricity price curves but does not focus on the prediction of the extreme spikes that are likely caused by one-of-a-kind events. For instance, the historical data set used for calibrating the forward price curve model in the New York area from February 2003 to January 2006 includes all price spikes but one outlier. This implies that the calibrated forward price curve model is capable of predicting price spikes that are of certain stationarity nature. As for the extreme spikes resulting from one-of-a-kind events, they shall not be viewed as being sampled from an embedded low-dimensional intrinsic manifold structure, thus they can be removed from the calibration data set. However, if such extreme price spikes were caused by changes to the fundamental structure of aggregate supply and demand, then the intrinsic dimension of the low-dimensional manifold would change accordingly, and yield a different set of major factors of the price dynamics in a low-dimensional space.

VI. Conclusion

We apply manifold-based dimension reduction to electricity price curve modeling. LLE is demonstrated to be an efficient method for extracting the intrinsic low-dimensional structure of electricity price curves. Using price data taken from the NYISO, we find that there exists a low-dimensional manifold representation of the day-ahead price curve in NYISO, and specifically, the dimension of the manifold is around 4. The interpretation of each dimension and the cluster analysis in the low-dimensional space are given to analyze the main factors of the price curve dynamics. Numerical experiments show that our prediction performs well for the short-term prediction, and it also facilitates medium-term prediction, which is difficult, even infeasible for other methods.

Appendix A

The procedure of identifying ARIMA model follows paper [7]. The history data we applied to train the model are 90 days. The model is as follows:[3]

$$
\begin{aligned}
&(1-\phi_1B^1-\phi_2B^2)\\
&\quad\times(1-\phi_{23}B^{23}-\phi_{24}B^{24}-\phi_{47}B^{47}-\phi_{48}B^{48}\\
&\qquad-\phi_{72}B^{72}-\phi_{96}B^{96}-\phi_{120}B^{120}-\phi_{144}B^{144}\\
&\quad\times(1-\phi_{168}B^{168}-\phi_{336}B^{336}-\phi_{504}B^{504})(1-B)\\
&\quad\times(1-B^{24})(1-B^{168})log(price_t)\\
&\qquad=c+(1-\theta_1B^1-\theta_2B^2-\theta_3B^3-\theta_4B^4-\theta_5B^5)\\
&\qquad\quad\times(1-\theta_{24}B^{24}-\theta_{48}B^{48})\\
&\qquad\quad\times(1-\theta_{168}B^{168}-\theta_{336}B^{336}-\theta_{504}B^{504})\varepsilon_t.
\end{aligned}
$$

The model estimation and prediction is implemented through the SCA system.

Appendix B

The ARX model with explanatory variable load follows paper [33]. The history data we applied to train the model are 45 days, as the relationship between prices and loads is dynamic. The model is as follows:

$$
\begin{aligned}
log(price_t)=c&+(u_1B^1+u_2B^2+u_3B^3+u_{24}B^{24}\\
&+u_{25}B^{25}+u_{48}B^{48}+u_{49}B^{49}+u_{72}B^{72}\\
&+u_{73}B^{73}+u_{96}B^{96}+u_{97}B^{97}\\
&+u_{120}B^{120}+u_{121}B^{121}+u_{144}B^{144}\\
&+u_{145}B^{145}+u_{168}B^{168}+u_{169}B^{169}\\
&+u_{192}B^{192}+u_{193}B^{193})log(price_t)\\
&+(v_1B^1+v_2B^2+v_3B^3+v_{24}B^{24}+v_{25}B^{25}\\
&+v_{48}B^{48}+v_{49}B^{49}+v_{72}B^{72}+v_{73}B^{73}\\
&+v_{96}B^{96}+v_{97}B^{97}+v_{120}B^{120}\\
&+v_{121}B^{121}+v_{144}B^{144}+v_{145}B^{145}\\
&+v_{168}B^{168}+v_{169}B^{169}+v_{192}B^{192}\\
&+v_{193}B^{193})log(load_t)+\varepsilon_t.
\end{aligned}
$$

The model estimation and prediction are implemented in MATLAB.

[3]Occasionally, we slightly change the model when it does not converge.

References

[1] B. Johnson and G. Barz, "Selecting stochastic processes for modeling electricity prices," in *Energy Modelling and the Management of Uncertainty*. London, U.K.: Risk, 1999.

[2] S. J. Deng, "Stochastic Models of Energy Commodity Prices and Their Applications: Mean-Reversion With Jumps and Spikes," UCEI POWER Working Paper P-073, 2000.

[3] J. J. Lucia and E. S. Schwartz, "Electricity prices and power derivatives: Evidence from the Nordic power exchange," *Rev. Deriv. Res.*, vol. 5, no. 1, pp. 5–50, 2002.

[4] A. Misiorek, S. Trueck, and R. Weron, "Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models," *Stud. Nonlin. Dynam. Econometr.*, vol. 10, no. 3, 2006, article 2.

[5] M. Davison, L. Anderson, B. Marcus, and K. Anderson, "Development of a hybrid model for electricity spot prices," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 257–264, May 2002.

[6] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espínola, "Forecast next-day electricity prices by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, May 2002.

[7] J. Contreras, R. Espínola, 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.

[8] A. J. Conejo, M. A. Plazas, R. Espínola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1035–1042, May 2005.

[9] A. M. Gonzalez, A. M. S. Roque, and J. G. Gonzalez, "Modeling and forecasting electricity prices with input/output hidden Markov models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 13–24, May 2005.

[10] C. Knittel and M. Roberts, "Empirical examination of deregulated electricity prices," *Energy Econ.*, vol. 27, no. 5, pp. 791–817, 2005.

[11] S. J. Deng and W. J. Jiang, "Levy process driven mean-reverting electricity price model: A marginal distribution analysis," *Decision Support Syst.*, vol. 40, no. 3–4, pp. 483–494, 2005.

[12] T. D. Mount, Y. Ning, and X. Cai, "Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters," *Energy Econ.*, vol. 28, no. 1, pp. 62–80, 2006.

[13] B. Ramsay and A. J. Wang, "A neural network based estimator for electricity spot-pricing with particular reference to weekend and public holidays," *Neurocomputing*, no. 47–57, 1998.

[14] B. R. Szkuta, L. A. Sanabria, and T. S. Dillon, "Electricity price short-term forecasting using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 851–857, Aug. 1999.

[15] L. Zhang, P. B. Luh, and K. Kasiviswanathan, "Energy clearing price prediction and confidence interval estimation with cascaded neural networks," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 99–105, Feb. 2003.

[16] N. Audet, P. Heiskanen, J. Keppo, and I. Vehvilainen, "Modeling Electricity Forward Curve Dynamics in the Nordic Market," in *Modelling Prices in Competitive Electricity Markets*. London, U.K.: Wiley, 2004.

[17] A. T. Lora, J. M. R. Santos, A. G. Expósito, J. L. M. Ramos, and J. C. R. Santos, Electricity Market Price Forecasting Based on Weighted Nearest Neighbors Techniques, Working Paper, Univ. Sevilla. Sevilla, Spain, 2006.

[18] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer-Verlag, 1997.

[19] J. B. Kruskal, "Multidimensal scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 1–27, 1964.

[20] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, 2003.

[21] L. K. Saul and S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[22] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[23] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[24] D. L. Donoho and C. Grimes, "Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 5591–5596, 2003.

[25] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2004.
[26] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems," *Appl. Comput. Harmon. Anal.* (Special Issue on Diffusion Maps and Wavelets), vol. 21, pp. 113–127, Jul. 2006.
[27] X. Huo, X. Ni, and A. K. Smith, "A survey of manifold-based learning methods, invited book chapter," in *Mining of Enterprise Data*. New York: Springer, 2005. [Online]. Available: http://www2.isye.gatech.edu/statistics/papers/06-10.pdf.
[28] X. Huo and J. Chen, "Local linear projection (LLP)," in *Proc. 1st IEEE Workshop Genomic Signal Processing and Statistics (GENSIPS)*, Raleigh, NC, Oct. 2002. [Online]. Available: http://www.gensips.gatech.edu/proceedings/.
[29] X. Huo, "A geodesic distance and local smoothing based clustering algorithm to utilize embedded geometric structures in high dimensional noisy data," in *Proc. SIAM Int. Conf. Data Mining, Workshop on Clustering High Dimensional Data and Its Applications*, San Francisco, CA, May 2003.
[30] T. Hastie, R. Tibshirani, and J. Friedman*, The Elements of Statistical Learning*. New York: Springer, 2001.
[31] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Advances in Neural Information Processing Systems 17 (NIPS2004)*. Cambridge, MA: MIT Press, 2005.
[32] P. Verveer and R. Duin, "An evaluation of intrinsic dimensionality estimators," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 81–86, Jan. 1995.
[33] A. J. Conejo, J. Contreras, R. Espínola, and M. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *Int. J. Forecast.*, vol. 21, pp. 435–462, 2005.
[34] P. J. Brockwell*, Introduction to Time Series and Forecasting*, 2nd ed. New York: Springer, 2003.
[35] R. B. Cleveland, W. S. Cleveland, J. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess," *J. Offic. Statist.*, vol. 6, pp. 3–73, 1990.

Jie Chen (S'05) received the B.S. degree in computational math from Nanjing University, Nanjing, China, in 2003. She is currently pursuing the Ph.D. degree at the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.

Her research interests are applied statistics and data mining.

Shi-Jie Deng (SM'06) received the B.Sc. degree in applied Mmthematics from Peking University, Beijing, China, the M.Sc. degree in mathematics from the University of Minnesota at Twin Cities, and the M.S. and Ph.D. degrees in industrial engineering and operations research (IEOR) from the University of California at Berkeley.

He is an Associate Professor of industrial and systems engineering at Georgia Institute of Technology, Atlanta. His research interests include financial asset pricing and real options valuation, financial engineering applications in energy commodity markets, transmission pricing in electric power systems, stochastic modeling, and simulation. He has served as a consultant to several private and public organizations on issues of risk management and asset valuation in the restructured electricity industry.

Dr. Deng received the CAREER Award from the National Science Foundation in 2002.

Xiaoming Huo (S'96-M'99-SM'04) received the B.S. degree in mathematics from the University of Science and Technology, Heifei, China, in 1993 and the M.S. degree in electrical engineering and the Ph.D. degree in statistics from Stanford University, Stanford, CA, in 1997 and 1999, respectively.

He is an Associate Professor with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His research interests include dimension reduction and nonlinear methods.

Dr. Huo received first prize in the Thirtieth International Mathematical Olympiad (IMO), which was held in Braunschweig, Germany.