Day-Ahead Electricity Market Forecasting using Kernels

Vassilis Kekatos,[1] Sriharsha Veeramachaneni,[2] Marc Light,[2] and Georgios B. Giannakis[1]
[1]Digital Tech. Center and Dept. of ECE, Univ. of Minnesota, Minneapolis, MN 55455, USA
[2]Windlogics Inc., 1021 E Bandana Blvd., Suite 111, Saint Paul, MN 55108, USA
Emails: {kekatos,georgios}@umn.edu, {Marc.Light,Sriharsha.Veeramachaneni}@windlogics.com

***Abstract*—Weather and life cycles, fuel markets, reliability rules, scheduled and random outages, renewables and demand response programs, all constitute pieces of the electricity market puzzle. In such a complex environment, forecasting electricity prices is a very challenging task; nonetheless, it is of paramount importance for market participants and system operators. Day-ahead price forecasting is performed in the present paper using a kernel-based method. This machine learning approach offers unique advantages over existing alternatives, especially in systematically exploiting the spatio-temporal nature of locational marginal prices (LMPs), while nonlinear cause-effect relationships can be captured by carefully selected similarities. Beyond conventional time-series data, non-vectorial attributes (e.g., hour of the day, day of the week, balancing authority) are transparently utilized. The novel approach is tested on real data from the Midwest ISO (MISO) day-ahead electricity market over the summer of 2012, during which MISO's load peak record was observed. The resultant day-ahead LMP forecasts outperform price repetition and ordinary linear regression, thus offering a promising inference tool for the electricity market.**

***Index Terms*—Locational marginal prices, kriging filtering, machine learning, wholesale electricity market.**

I. Introduction

The smart grid vision entails advanced information technology and data analytics to enhance the efficiency, sustainability, and economics of the power grid infrastructure [16]. In this context, statistical learning tools are employed in this work to forecast the day-ahead electricity market. To appreciate the value of such predictions, consider a typical day-ahead market. An independent system operator (ISO) collects hourly supply and demand bids submitted by generator owners and utilities for the 24 hours of the following day. Compliant with network and reliability constraints, the grid is then dispatched in the most economical way. Together with power schedules, the ISO announces hourly prices for the electricity produced and/or consumed at specific nodes of the grid. Due to congestion and losses incurred by the transmission system, these locational marginal prices (LMPs) are different across the grid.

Apparently, forecasting tomorrow's LMPs constitutes a major decision making task for asset owners and market participants to plan their hedging and bidding strategies [1]. Interestingly, some ISOs recently announce pricing forecasts too in an attempt to relieve congestion [5]. These forecasts serve as signals for reliability-ensuring bidding and strategic planning in short- and long-term horizons, respectively. At a wider scale, price analytics are important to government services for identifying the so termed "national congestion corridors."

Forecasting schemes reported so far include statistical time-series analysis approaches based on auto-regressive (integrated) moving average models and their generalizations; see e.g., [4], [6], [3]. However, these models are confined to linear predictors, whereas markets involve generally nonlinear dependencies. To account for nonlinearities, artificial intelligence approaches, such as fuzzy systems and neural networks, have been also investigated [18], [8], [11], [17]. A nearest neighborhood approach has been suggested in [12]. In [19], market clearance is assumed to be solved as a quadratic program and forecasts are extracted based on the most probable outage combinations. Reviews on electricity price forecasting can be found in [1] and [14].

In this work, a kernel-based forecasting approach is considered. Contrary to existing methods, the entire day-ahead market, involving all commercial nodes and all hours, is learned by a single predictor. Using a similarity graph approach, spatio-temporal pricing correlations are systematically utilized. Based on the problem specifications, appropriate kernels are carefully designed and estimated. Furthermore, time and nodal information is transparently incorporated in the form of non-vectorial attributes. The forecasting performance of the method is corroborated by preliminary numerical tests on real data collected from the MISO market over the summer of 2012. MISO is one of the largest electricity markets with relatively high wind penetration. During this specific period, a historical load record was observed in the geographical area where MISO operates [13].

The paper is outlined as follows. Electricity market forecasting is formulated in Section II, where the novel method is also presented. The design of kernels is detailed in Section III, and an efficient algorithm for implementing kernel-based forecasting is developed in Section IV. Forecasting results on the MISO market are presented in Section V, and the paper is concluded in Section VI.

Regarding notation, lower- (upper-) case boldface letters denote column vectors (matrices); calligraphic letters stand for sets; and $(\cdot)^T$ denotes transposition. The symbol $:=$ stands for variable definition. Symbols $\odot$ and $\otimes$ denote the matrix entry-wise multiplication and the matrix Kronecker product, respectively. The operation $\text{vec}(\mathbf{A})$ transforms a matrix to a vector by stacking its columns. The following

Kronecker product properties will be needed throughout:
(P1): $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$
(P2): $(\mathbf{A} \otimes \mathbf{B})^{T} = \mathbf{A}^{T} \otimes \mathbf{B}^{T}$
(P3): $\text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = \left(\mathbf{B}^{T} \otimes \mathbf{A}\right) \text{vec}(\mathbf{X})$ and
(P4): $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C}) \otimes (\mathbf{B}\mathbf{D})$.

II. Forecasting Methods

To formulate the electricity price forecasting problem, a simple model is described first, based on which subsequent generalizations are built on. As in typical inference problems, data are partitioned into features and targets (a.k.a. regressors or predictors and responses or predictions). Data over a previous time period are provided as training feature/target pairs. Given these data, a predictor aims at finding the target values for future target variables based on the known corresponding features.

Consider the power grid as a network of N commercial pricing nodes (CPNs). The training and test periods comprise T_1 and T_2 time intervals, respectively. In the day-ahead market setup considered here, the interval is naturally an hour. The training period may potentially include all available historical price values, and $T_2 = 24$ refers to the following day.

Hourly nodal prices are stored row-wise in matrices $\mathbf{Y}_i$ of dimension $T_i \times N$ for $i = 1, 2$ corresponding to the training and the testing phase, respectively. Predictor (feature) variables are collected in $P \times 1$ vectors stored in matrices $\{\mathbf{X}_i\}_{i=1}^{2}$, with dimensions compatible with $\mathbf{Y}_i$'s. A naive forecasting method would adopt the model

$$\mathbf{Y}_i = \mathbf{X}_i \mathbf{W} + \mathbf{N}_i,\ i = 1, 2 \tag{1}$$

where $\mathbf{W}$ is the $P \times N$ matrix of prediction coefficients. Note that the n-th columns of $\mathbf{Y}_i$'s and $\mathbf{W}$, $\mathbf{y}_{i,n}$ and $\mathbf{w}_n$, correspond to the n-th CPN and its prediction coefficients, respectively. Further, each node's forecast is a linear function of the same P-dimensional feature vector.

A. Linear Prediction

An ordinary least-squares (LS) predictor yields the forecast $\hat{\mathbf{Y}}_2^{\text{LS}} := \mathbf{X}_2 \mathbf{X}_1^{\dagger} \mathbf{Y}_1$, where † denotes matrix pseudo-inverse.

To avoid overfitting, a regularized LS alternative (a.k.a. ridge regression predictor) relies on

$$\hat{\mathbf{w}}_n^{\text{R}} := \arg\min_{\mathbf{w}_n} \ \|\mathbf{y}_{1,n} - \mathbf{X}_1 \mathbf{w}_n\|_2^2 + \lambda_{\text{R}} \|\mathbf{w}_n\|_2^2 \tag{2}$$

where $n = 1, \ldots, N$, and $\lambda_{\text{R}} > 0$ denotes the regularization parameter. The prediction coefficients $\{\hat{\mathbf{w}}_n^{\text{R}}\}_{n=1}^{N}$ in (2) can be found in closed form as $\left(\mathbf{X}_1^T \mathbf{X}_1 + \lambda_{\text{R}} \mathbf{I}_P\right)^{-1} \mathbf{X}_1^T \mathbf{y}_{1,n}$, or equivalently, $\mathbf{X}_1^T \left(\mathbf{X}_1 \mathbf{X}_1^T + \lambda_{\text{R}} \mathbf{I}_{T_1}\right)^{-1} \mathbf{y}_{1,n}$ for all n. Hence, the day-ahead ridge forecast is given by

$$\hat{\mathbf{Y}}_2^{\text{R}} := \mathbf{X}_2 \mathbf{X}_1^T \left(\mathbf{X}_1 \mathbf{X}_1^T + \lambda_{\text{R}} \mathbf{I}_{T_1}\right)^{-1} \mathbf{Y}_1. \tag{3}$$

A Bayesian interpretation of ridge regression assumes:
(A1) Columns of the $\mathbf{N}_i$'s in (1) are standard Gaussian, independent across nodes and training/testing phases; and
(A2) All $\mathbf{w}_n$'s are zero-mean white Gaussian with variance λ_{R}^{-1}. Under these assumptions, $\hat{\mathbf{Y}}_2^{\text{R}}$ is the minimum mean-square error (MMSE) prediction of the day-ahead prices. Notice though that both the LS and ridge predictions are decoupled across nodes, and are restricted to be linear functions of the data. To model more complex dependencies, kernel prediction methods are pursued next.

B. Kernel Prediction

Vectorial feature data stored in $\{\mathbf{X}_i\}_{i=1}^{2}$ affect price forecasts via blocks of the so-called Gramian matrix $\mathbf{K}_v = \mathbf{X}\mathbf{X}^T$, where $\mathbf{X}^T := [\mathbf{X}_1^T \ \mathbf{X}_2^T]$. Note that the (t_1, t_2)-th entry of $\mathbf{K}_v$ is the inner product between feature vectors of two time instances; i.e., $[\mathbf{K}_v]_{t_1,t_2} = \mathbf{x}_{t_1}^T \mathbf{x}_{t_2}$. In order to allow for general nonlinear feature/target relationships, the idea is to use $[\mathbf{K}_v]_{t_1,t_2} = k(\mathbf{x}_{t_1}, \mathbf{x}_{t_2})$, where $k(\cdot,\cdot)$ is a chosen *kernel function* capturing the similarity between features.

The day-ahead electricity market forecast can be thought of as a function whose arguments are the available vectorial features along with timestamp and nodal information. To infer such a function, pairs of arguments and prices $\{\zeta_l, y_l\}_{l=1}^{L}$ are provided as historical data. The function argument is the triplet $\zeta_l := (\mathbf{x}_l, t_l, n_l) \subset \mathcal{X} \times \mathcal{T} \times \mathcal{N}$, where $\mathcal{X} \subset \mathbb{R}^P$, $\mathcal{T}$ is the space of time instants, and $\mathcal{N}$ is the finite set of CPNs. The desired prediction function is a mapping $f : \mathcal{X} \times \mathcal{T} \times \mathcal{N} \to \mathbb{R}$. Hence, different from the typical approach, where per node predictors are trained, a single model capturing the whole pricing network is pursued here.

Based on this formalism, electricity market forecasting can be posed as the regularization problem

$$\min_{f \in \mathcal{H}} \ \sum_{l=1}^{L} (y_l - f(\zeta_l))^2 + \lambda_{\text{K}} \|f\|_{\mathcal{H}}^2 \tag{4}$$

where $L := NT_1$ is the number of training data, $\mathcal{H}$ is the space in which f lies, and $\|f\|_{\mathcal{H}}$ is the induced norm in that space [2]. The sum in the cost of (4) is an LS data fitting component, while the induced norm offers a *regularizer* effectively constraining f to lie in the selected space, while at the same time enabling generalization over unseen data. These two components are balanced via $\lambda_{\text{K}} > 0$, which is typically set via cross-validation [9].

The pertinent questions in the present context of electricity price forecasting are: **(Q1)** how can the function space be selected; and **(Q2)** how the functional optimization in (4) can be practically solved. Statistical learning suggests selecting first a kernel function $k(\zeta_l, \zeta_m)$ capturing the similarity between the application-specific features. If $k(\cdot,\cdot)$ is a symmetric and positive definite function, meaning that the *kernel matrix* constructed as $[\mathbf{K}]_{l,m} := k(\zeta_l, \zeta_m)$ is positive definite for any feature collection, then it uniquely defines a reproducing kernel Hilbert space (RKHS) of functions. The latter is the family of functions expressible as $f(\zeta) = \sum_{i=1}^{\infty} k(\zeta, \zeta_i)$. The celebrated Representer's Theorem asserts that when $\mathcal{H}$ is an RKHS, the minimizer of (4) takes the form [9]

$$\hat{f}(\zeta) = \sum_{l=1}^{L} k(\zeta, \zeta_l) \hat{\alpha}_l. \tag{5}$$

In other words, the wanted prediction function can be expressed as a finite linear combination of kernel functions evaluated only at the training points.

By exploiting (5), it can be shown that the functional optimization in (4) can be equivalently solved by estimating the vector of coefficients in (5) via the quadratic program

$$\hat{\boldsymbol{\alpha}} := \arg\min_{\boldsymbol{\alpha}} \|\mathbf{y}_1 - \mathbf{K}_{11}\boldsymbol{\alpha}\|_2^2 + \lambda_{\mathrm{K}}\boldsymbol{\alpha}^T\mathbf{K}_{11}\boldsymbol{\alpha} \tag{6}$$

where $\mathbf{K}_{11}$ is the $L \times L$ kernel matrix evaluated over the training points. The vector $\boldsymbol{\alpha}$ uniquely minimizing (6) is readily found in closed form as

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{K}_{11} + \lambda_{\mathrm{K}}\mathbf{I}_{NT_1}\right)^{-1}\mathbf{y}_1. \tag{7}$$

Having acquired $\hat{\boldsymbol{\alpha}}$, the fitted function can be evaluated at any other point. Specifically, (5) dictates that at any new point ζ_{new}, the predicted function value is $\hat{f}(\zeta_{\text{new}}) = \mathbf{k}_{\text{new}}^T\hat{\boldsymbol{\alpha}}$, where $\mathbf{k}_{\text{new}}$ is an $L \times 1$ vector with entries $k(\zeta_{\text{new}}, \zeta_l)$. By stacking all function evaluations for tomorrow's market, the price forecast can be compactly expressed as

$$\hat{\mathbf{y}}_2^{\mathrm{K}} := \mathbf{K}_{21}\hat{\boldsymbol{\alpha}} \tag{8}$$

which addresses (Q2). Responding to (Q1) requires specifying the kernels, which is the topic considered next.

III. Kernels for (Non)-Vectorial Data

This section deals with constructing kernels tailored for electricity market forecasting. Recall that the kernel function $k(\zeta_l, \zeta_m)$ is a measure of similarity. Since ζ_l comprises vectorial features $\mathbf{x}_l$, time instances t_l, and nodal information n_l, kernels for each of the three components are individually defined first. With the component-wise kernels respectively denoted by $k_v(\mathbf{x}_l, \mathbf{x}_m)$, $k_t(t_l, t_m)$, and $k_s(n_l, n_m)$, the collective kernel is constructed as their product

$$k(\zeta_l, \zeta_m) := k_v(\mathbf{x}_l, \mathbf{x}_m)k_t(t_l, t_m)k_s(n_l, n_m). \tag{9}$$

If the three factors are valid kernels, i.e., symmetric and positive definite functions, then their product in (9) is certainly a valid kernel too [2].

A. Kernels for Vectorial Data

Regarding $k_v(\mathbf{x}_l, \mathbf{x}_m)$, a simple choice is the linear kernel $\mathbf{x}_l^T\mathbf{x}_m$. More complex nonlinear relationships can be captured by other vectorial kernels, such as the polynomial, the sigmoid, the Gaussian, or spline kernels [2], [9]. In this work, the Gaussian kernel

$$k_v(\mathbf{x}_l, \mathbf{x}_m) = \exp\left(-\nu\|\mathbf{x}_l - \mathbf{x}_m\|_2^2\right)$$

will be adopted with bandwidth parameter $\nu > 0$.

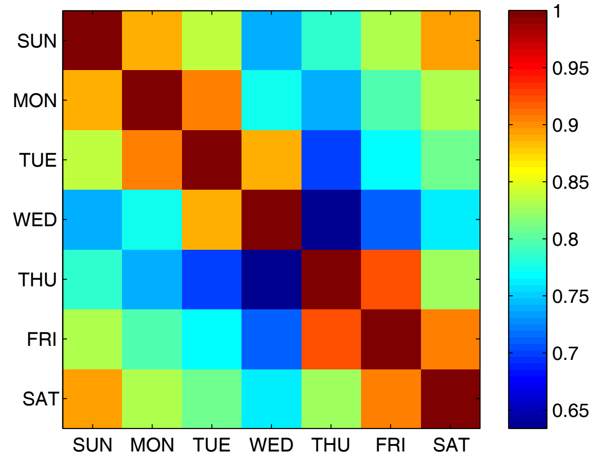

Fig. 1. A similarity matrix across the days of the week $\mathbf{K}_7$. Its entries have been empirically estimated as the correlation coefficients between daily MISO pricing signals from the summer of 2011. Observe the correlation over successive days.

B. Kernels for Timestamps

Careful kernel design for time data or timestamps can potentially capture the cyclostationarity inherent to electricity demand and consequently price. One should first recognize that the datum t_l carries two pieces of information: the day of the week d_l, and the hour of the day h_l. Moreover, the similarity between two timestamps should be decreasing with respect to their time lag. Prices corresponding to days spaced apart should be less correlated than prices between the same or consecutive days. These considerations motivate

$$k_t(t_l, t_m) = k_d(d_l, d_m)k_h(h_l, h_m)\beta^{|d_l - d_m|} \tag{10}$$

whose factors are described next.

The $k_d(\cdot,\cdot)$ factor models pairwise similarities across days of the week. Since this factor has discrete support, it is simply defined by a 7×7 matrix $\mathbf{K}_7$. It can be shown that $k_d(\cdot,\cdot)$ is a valid kernel function, provided $\mathbf{K}_7$ is positive definite [9]. The entries of $\mathbf{K}_7$ can be selected according to the designer's prior knowledge. Alternatively, they can be estimated as the correlation coefficients of historical market prices over the seven days of the week; see e.g., Fig. 1. Likewise, the $k_h(h_l, h_m)$ factor captures pairwise similarities across hours of the day. This function can be defined by a 24×24 matrix $\mathbf{K}_{24}$, as the one empirically estimated in Fig. 2.

The last factor in (10) models an exponentially decaying similarity between two time instances, and intends to incorporate the nonstationary characteristics of the electricity market. Parameter $\beta > 0$ is a decaying parameter that should be selected close to 1 yielding an effective memory of $1/(1-\beta)$ days. Finally, as with (9), the product of kernels is also a valid kernel.

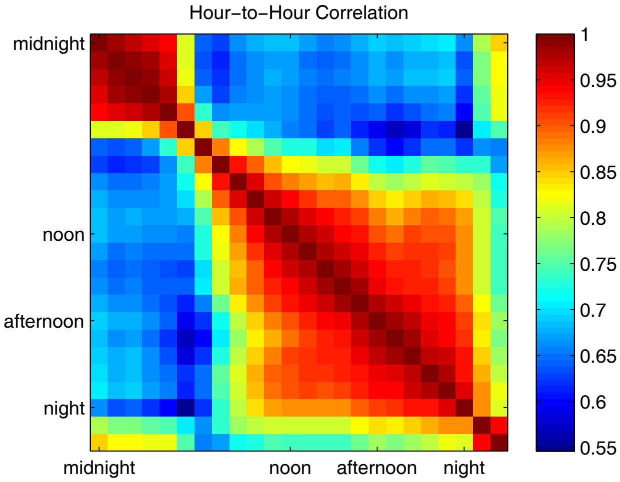

Fig. 2. A similarity matrix across hours of the day $\mathbf{K}_{24}$. Its entries have been empirically estimated as the correlation coefficients between hourly MISO pricing signals from the summer of 2011. Day and night patterns can be easily distinguished.

C. Kernels for Nodal Information

Pricing signals exhibit spatial correlations. Selecting kernels to account for not only temporal but also spatial patterns requires understanding the spatial structure of the market. Spatial correlations could be mainly attributed to the transmission infrastructure together with line and generation outages. But since the electrical topology is unavailable, other spatial similarity measures should be considered. In this paper, the information of the local balancing authority (LBA) that each CPN belongs to is exploited as a topology surrogate. The presumption here is that nodes belonging to the same LBA should experience similar prices. In addition, nodes controlled by neighboring authorities are expected to have prices correlated more than nodes controlled by non-adjacent ones.

To rigorously model spatial dependence and incorporate it in the prediction scheme, a graph-based inference approach is adopted; see also [10]. CPNs are considered as vertices of a similarity graph, connected with edges having non-negative weights. Edge weights are chosen proportional to the similarity between incident nodal prices. Edges in the same LBA are assigned unit weights; edges across nodes from different LBAs receive weight 0.5; and all other edges are zero. All weight values are stored in an $N \times N$ symmetric adjacency matrix $\mathbf{A}$. Even though edge weights are selected in a rather ad hoc manner, cross-validation could yield more meaningful values.

Having constructed the similarity graph, its *normalized Laplacian* matrix is defined as $\mathbf{L} := \mathbf{I}_N - (\mathbf{D}^{1/2})^{\dagger}\mathbf{A}(\mathbf{D}^{1/2})^{\dagger}$, where $\mathbf{D}$ is a diagonal matrix having the row sums of $\mathbf{A}$ across its main diagonal; see e.g., [10]. A kernel matrix capturing the similarity across the pricing network is the regularized Laplacian

$$\mathbf{K}_s := (\mathbf{L} + s\mathbf{I}_N)^{-1} \tag{11}$$

where $s\mathbf{I}_N$ for $s > 0$ has been added to ensure that the kernel is strictly positive definite [15].

IV. Efficient Implementation

Even though coupling price forecasts across time and nodes is beneficial from an inference perspective, the derived predictions $\hat{\mathbf{y}}_2^{\mathrm{K}}$ in (8) are computationally challenging. Indeed, the ridge predictor in (3) requires inverting a $T_1 \times T_1$ matrix, whereas the kernel one in (6)-(8) involves inverting the $NT_1 \times NT_1$ matrix $\mathbf{K}_{11} + \lambda_{\mathrm{K}}\mathbf{I}_{NT_1}$. Assuming a market of $1,000$ CPNs and a training period of three weeks (around 500 hours) leads to a $5 \cdot 10^5 \times 5 \cdot 10^5$ $\mathbf{K}_{11}$ matrix. Such a big matrix is not only hard to invert, but also non-trivial to store.

To efficiently implement (6)-(7) and facilitate cross-validation, the particular problem structure should be exploited. Note first that finding $\hat{\boldsymbol{\alpha}}$ does not necessarily require inverting $\mathbf{K}_{11} + \lambda_{\mathrm{K}}\mathbf{I}_{NT_1}$, but rather solving the linear system

$$\left(\mathbf{K}_s \otimes \mathbf{K}_{vt}^{11} + \lambda_{\mathrm{K}}\mathbf{I}_{NT_1}\right)\hat{\boldsymbol{\alpha}} = \mathbf{y}_1. \tag{12}$$

Define first the $T_1 \times N$ matrix $\hat{\mathbf{A}}$ so that $\hat{\boldsymbol{\alpha}} = \mathrm{vec}(\hat{\mathbf{A}})$. Using (P3) from Section I, the linear equations in (12) are equivalently expressed as $\mathbf{K}_{vt}^{11}\hat{\mathbf{A}}\mathbf{K}_s + \lambda_{\mathrm{K}}\hat{\mathbf{A}} = \mathbf{Y}_1$, or,

$$\mathbf{K}_{vt}^{11}\hat{\mathbf{A}} + \lambda_{\mathrm{K}}\hat{\mathbf{A}}\mathbf{K}_s^{-1} = \mathbf{Y}_1\mathbf{K}_s^{-1}. \tag{13}$$

Interestingly, (13) is an instance of the *Sylvester equation* that has been widely studied in the control literature, and can be efficiently solved using the Bartels and Stewart algorithm [7], already implemented in MATLAB's `lyap` function.

Having acquired $\hat{\mathbf{A}}$, the price forecast is provided by $\hat{\mathbf{y}}_2^{\mathrm{K}} = \mathbf{K}_{21}\hat{\boldsymbol{\alpha}}$, or more efficiently as

$$\hat{\mathbf{Y}}_2^{\mathrm{K}} = \mathbf{K}_{vt}^{21}\hat{\mathbf{A}}\mathbf{K}_s. \tag{14}$$

V. Forecasting the Midwest ISO (MISO) Market

The forecasting performance of the novel kernel-based predictors is evaluated in this section. To this end, real data from the day-ahead MISO electricity market are used. The data are related to the summer period of 2012, where a new MISO load demand peak record was observed [13]. The related commercial pricing model consists of $N = 1,732$ nodes. Day-ahead hourly LMPs are collected for the period June 1 to August 31, 2012, a total of 2,208 hours.

Regarding the feature data used, vectorial data at time instant t_l were collected in the vector $\mathbf{x}_l = \mathbf{x}_{t_l}$, and included:

- The day-ahead LMPs across MISO for the same hour during the previous operating day.
- The aggregate over the MISO region wind energy production forecast. Apart from the same hour, the previous and the next hour were also included to model volatility in wind production. Moreover, wind can be considered a rough surrogate for weather conditions too.
- The hourly regional (East, West, and Central) load forecast. Again, the loads predicted for one hour before and after the hour of interest were also included. To capture the time coupling across hours due to ramp constraints considered in unit commitment.

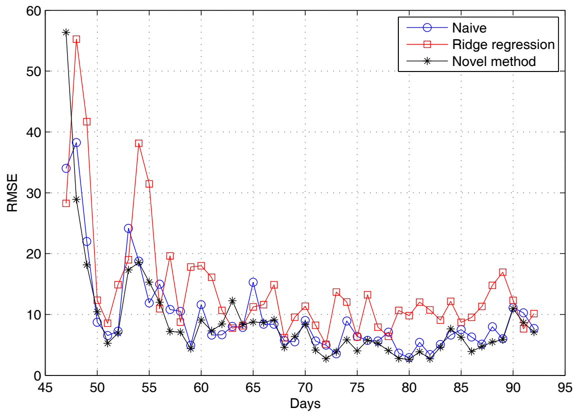

Fig. 3. The RMSE performance for the three forecasting methods tested. The evaluation period ranges from the 46th (July 17, 2012) to the 92nd (August 31, 2012) day of the summer period.

Mainly due to scheduled and random outages in transmission lines and generators, the electricity market is non-stationary; hence, to capture time variability, each day's predictor is trained independently using the feature/price data from the last three weeks. Important model parameters, namely, the Gaussian kernel bandwidth ν, the Laplacian matrix regularizer s, regularization parameters λ_{R} and λ_{K}, and the forgetting factor β need to be tuned. To tune these parameters, market data are divided in two parts. The first 46 days were used only for parameter tuning. For the remaining 46 days, the aforementioned parameters were fixed, and predictors were again trained based on their previous three weeks.

Three forecasting methods were tested: (i) a naive repetitive forecast that simply repeats yesterday's prices; (ii) the ridge regression forecast of (3); and (iii) the novel spatio-temporal kernel-based predictor of (8). The evaluation metric used was the residual mean-square error (RMSE) $\|\mathbf{y}_2 - \hat{\mathbf{y}}_2\|_2/(24N)$. Using the initial 46-day period, the parameters were set to $\lambda_{\mathrm{R}} = 1,000$, $\lambda_{\mathrm{K}} = 1$, $\nu = 1 \cdot 10^{-4}$, $\beta = 0.999$, and $s = 1$.

Fixing parameters to the aforementioned values, the forecasting results obtained for the second half of the summer period in the MISO market are depicted in Fig. 3. The initial fluctuations experienced by all methods can be attributed to high temperatures. Interestingly, the naive forecast outperforms the linear ridge regression-based predictor. The novel spatio-temporal kernel-based forecast however attains consistently the lowest RMSE.

VI. Conclusions

Undoubtedly electricity market forecasting is a challenging yet instrumental task for the smart grid operation. A novel machine learning-based approach was developed. Viewing the market as a properly defined graph, spatial correlations were indirectly modeled using information from balancing authorities. Time and calendar information can capture cyclo-stationarity as well as holiday effects on the market. Gaussian kernels on a multitude of numerical attributes model nonlinear dependencies, and avoid the curse of dimensionality. The novel price prediction algorithm operates on data publicly available, even though additional data can be readily incorporated. An efficient implementation was proposed that facilitated application of the method to the MISO 1,732-CPN network. Including additional data sources, deriving more efficient algorithms, and inferring other useful market quantities are some of the future research directions of this work.

References

[1] N. Amjady and M. Hemmati, "Energy price forecasting - problems and proposals for such predictions," *IEEE Power Energy Mag.*, vol. 4, no. 2, pp. 20–29, Mar./Apr. 2006.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[3] A. J. Conejo, M. A. Plazas, R. Espinola, 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.

[4] 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.

[5] Electric Reliability Council of Texas (ERCOT), "Ercot launches wholesale pricing forecast tool," July 11, 2012. [Online]. Available: http://www.ercot.com/news/press_releases/show/26244

[6] R. C. Garcia, J. Contreras, M. van 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.

[7] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 1996.

[8] 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. 1, pp. 13–24, Feb. 2005.

[9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2009.

[10] E. D. Kolaczyk, *Statistical Analysis of Network Data, Methods and Models*. New York, NY: Springer, 2010.

[11] G. Li, C. C. Liu, C. Mattson, and J. Lawarree, "Day-ahead electricity price forecasting in a grid environment," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 266–274, Feb. 2007.

[12] A. T. Lora, J. M. R. Santos, A. G. Exposito, J. L. M. Ramos, and J. C. R. Santos, "Electricity market price forecasting based on weighted nearest neighbors techniques," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1294–1301, Aug. 2007.

[13] Midwest Independent System Operator (MISO), "New market peak demand record set in miso region," July 25, 2012. [Online]. Available: https://www.midwestiso.org/AboutUs/MediaCenter/PressReleases/Pages/NewMarketPeakDemandRecordSetinMISORegion.aspx

[14] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. New York: IEEE-Wiley Interscience, 2002.

[15] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Proc. of the Annual Conference on Computational Learning Theory and Kernel Workshop*, ser. Lecture Notes in Computer Science, B. Schölkopf and M. Warmuth, Eds. Springer, 2003.

[16] U.S. Department of Energy, "Smart grid research and development, multi-year program plan (2010-2014)," Sep. 2011.

[17] L. Wu and M. Shahidehpour, "A hybrid model for day-ahead price forecasting," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1519–1530, Aug. 2010.

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

[19] Q. Zhou, L. Tesfatsion, and C.-C. Liu, "Short-term congestion forecasting in wholesale power markets," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2185–2196, Nov. 2011.