A New Recursive Dynamic Factor Analysis for Point and Interval Forecast of Electricity Price

H. C. Wu, *Student Member, IEEE*, S. C. Chan, *Member, IEEE*, K. M. Tsui, and Yunhe Hou, *Member, IEEE*

***Abstract*—The functional principal component analysis (FPCA) is a recent tool in multivariate statistics and it has been shown to be effective for electricity price forecasting. However, its online implementation is expensive, which requires the computation of eigen-decomposition at each update. To reduce the arithmetic complexity, we propose a recursive dynamic factor analysis (RDFA) algorithm where the PCs are recursively tracked using efficient subspace tracking algorithm while the PC scores are further tracked and predicted recursively using Kalman filter (KF). From the latter, the covariance and hence the interval of the forecasted electricity price can be estimated. Advantages of the proposed RDFA algorithm are the low online complexity, and the availability of the prediction interval thanks to the KF framework. Furthermore, a robust extension is proposed to tackle possible non-Gaussian variation. Finally, the RDFA algorithm can be extended to predict electricity price in a longer period using a multi-factor model by capturing trends in different time horizon. Experimental results on the New England and Australian datasets show that the proposed RDFA approach is able to achieve better prediction accuracy than other conventional approaches. It thus serves as an attractive alternative to other conventional approaches to forecast electricity price and other related applications because of its low complexity, efficient recursive implementation and good performance.**

***Index Terms*—Electricity price forecasting, FPCA, interval forecast, Kalman filter, multi-factor model, OPASTr, recursive, subspace tracking.**

Nomenclature

AEMO	Australian electricity market operator.
ARMA	Autoregressive and moving average.
ARV	Autoregressive model with time varying mean.
ED	Eigen-decomposition.
FA	Factor analysis.
FPCA	Functional principal component analysis.
FV	Forward validation.
GARCH	Generalized autoregressive conditional heteroskedasticity.
KF	Kalman filter.
LMP	Locational marginal pricing.
LS	Least squares.
OPAST	Orthonormal projection approximation subspace tracking.
OPASTr	Orthonormal projection approximation subspace tracking with rank-1 modification.
MAPE	Mean absolute percentage error.
MDL	Minimum description length.
NSW	New South Wales.
OU	Ornstein-Uhlenbeck.
QRD	QR decomposition.
RDFA	Recursive dynamic factor analysis.
RRDFA	Robust recursive dynamic factor analysis.
SPE	Squared prediction error.
SVR	Support vector regression.
SSM	State space model.

Manuscript received March 17, 2012; revised April 05, 2012, August 31, 2012, October 24, 2012, and November 18, 2012; accepted November 28, 2012. Date of publication January 03, 2013; date of current version July 18, 2013. Paper no. TPWRS-00185-2012.

The authors are with the Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong (email: andrewhcwu@eee.hku.hk; scchan@eee.hku.hk; kmtsui@eee.hku.hk; yhhou@eee.hku.hk).

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.2012.2232314

I. Introduction

THE launch of the electricity market has transformed the electricity industry from a primarily technical business to one in which products are treated in much the same way as other commodities. In such a market, electricity price can be highly volatile due to the time varying nature of the demand and supply. Moreover, with increased utilization of renewable energies with variable nature, such as solar and wind power, the operation of power system will become more dynamic and stochastic in nature. In fact, renewable energies are expected to provide 20% of U.S. electricity market by 2030 [1]. To address the possible volatility of electricity price, there is an increasing interest in predicting not only the price itself but also the price interval, which can quantify the uncertainty of the forecasted price and hence to evaluate the risks of the decisions made by market participants. Finally, with the introduction of real-time pricing [2], electricity price forecasting has to be done in a much shorter interval of say 1 hour, half an hour or even shorter. These calls for efficient real-time algorithms to perform electricity price forecasting, so that decisions can be made by the consumer in a shorter time interval to provide finer control of appliances.

According to recent reviews [3], [4], common behavior of electricity prices can be broadly summarized in the following

aspects: 1) strongly seasonal nature of prices; 2) mean reversion nature; 3) price dependent volatilities and spikes. Different algorithms have been put forward for modeling these characteristics of electricity prices. A traditional approach is the autoregressive (AR) based time series models. It was shown to approximate continuous time models [5] such as the mean reverting Ornstein-Uhlenbeck (OU) processes [6], [7]. The AR model can be explicitly solved using the ordinary least squares (LS). Other variants such as the autoregressive and moving average (ARMA) model, AR with time varying mean (ARV), generalized AR conditional heteroskedasticity (GARCH) are also commonly used to predict future spot price. On the other hand, production cost models were proposed to predict the electricity price/load by simulating the strategic behavior of producers and the consumers. More recently, machine learning approaches such as the support vector regression (SVR) [8] was proposed for time series prediction in [9], and was applied recently to electricity price forecasting [10]. A desirable feature of SVR is that it is able to model possible non-linear relationships between the future price and previous price using a non-linear kernel function.

Recently, studies in the Dutch, German and French day ahead hourly prices [11] suggests it is more advantageous to use different models to describe the prices for low and peak hours because prices during low demand behave distinctively from those during peak demand. For example, the price data can be divided into 24 partitions (hours) for hourly price, where each partition represents a particular hour of all days and can be treated as a separate variable. This concept can be extended to other time scales such as half-hourly data. Since the pattern of the price for adjacent partitions will be highly similar, the application of factor analysis (FA) techniques such as functional principal component analysis (FPCA) is becoming an important research area for such problems. The usefulness of FPCA for price forecasting is its ability to explore and capture the correlation between adjacent periods of interest [12]. For example, an intraweek seasonal cycle exhibits similarity of the demand from one week to the next.

One of the major challenges of the electricity price forecast is to tackle the time-varying nature of the electricity price [3], [4]. To improve adaptability, online batch processing is usually desirable, where the prediction is performed by applying the forecasting algorithm to a data block made up of consecutive electricity price samples. Whenever a new sample is available, the existing data block is appended with the new sample and the earliest sample is discarded. This procedure is repeated for each incoming sample or blocks of samples in each update. In this regard, the forecasting algorithm can better adapt to possible changes of trend. However, this may also lead to high arithmetic complexity. For example, the eigen-decomposition (ED) in the FPCA and the cross validation for parameter tuning in the SVR have to be repeatedly applied to each update.

In this paper, to cope with the high arithmetic complexity incurred by such online real-time pricing estimation, we propose a new recursive dynamic factor analysis (RDFA) algorithm. It employs efficient recursive subspace tracking and ED algorithms to compute the PCs and PC scores in the FPCA. Since only the most recent sample is used for the updating, the memory storage required is also reduced. Moreover, we consider a dynamic factor model where the PC scores are modeled as AR processes. By assuming that the innovation is Gaussian distributed [31], [32], the Kalman filter (KF) algorithm can be used to recursively tracked the PC scores. This also allows the covariance and hence the intervals of the forecasted values to be estimated.

An outline and major contributions of the proposed algorithms are summarized below:

1) It decomposes the electricity price into deterministic and stochastic components, where longer term variations such as seasonal weather, annual generation planned outage, generation investment and retirement can be captured by the deterministic component modeled as the Fourier series, and they are estimated online using the recursive least squares (RLS). This allows the longer term variations to be effectively tracked.
2) On the other hand, the stochastic component is used to model shorter term variations, such as variations of electricity load and generation dispatches, given the deterministic components. It is approximated by a linear combination of principal components (PCs) and PC scores, which are recursively computed by an efficient subspace tracking algorithm called the orthonormal projection approximation subspace tracking (OPAST) [16]. This reduces memory requirement and arithmetic complexity which simplify real-time computation. The OPAST is commonly used in communications, array and signal processing [16], [17]. However, a major problem of subspace tracking is that the PCs estimated are up to a rotation [16] and hence a new technique based on rank-1 modification [13] is utilized to effectively compute the PCs from the tracked subspace.
3) In order to perform forecasting derived from the PCs, the PCs scores are modeled as AR models. By incorporating a regularization term to the LS estimation of the AR model, the estimation variance can be further reduced. The resultant problem can be reformulated as the state estimation problem of a linear state-space model, which can be recursively solved using the KF. A major advantage of the KF is that it can provide a density estimate of the AR coefficients, which allows us to compute the interval of the forecasted electricity price, i.e., interval forecast.
4) To tackle possible non-Gaussian variation such as price spikes in some markets, which may degrade the subsequent forecasting accuracy, we propose a robust extension for the RDFA algorithm based on robust M-estimation. Firstly, price spikes are detected using certain robust detection criteria. Afterwards, the effect of these price spikes to the RDFA algorithm is suppressed or down-weighted using a modified Huber function so as to improve the robustness of the proposed algorithm in subsequent forecasting.

Moreover, the framework can be generalized to a multi-factor model, which decomposes the electricity price into different components of different time horizon, such as hourly, daily, weekly, monthly, etc., to further improve its performance for large horizon forecasting including weekly or longer correlations. Due to page limitation, interested readers are referred to the supplementary material [15] for the details of the multi-factor model.

Though subspace tracking has been reported before for communications, array and signal processing [16], [17] and recently fault detection [13], the incorporation of interval estimation, and its application to electricity price forecasting is to our best knowledge new.

Experimental results show that the proposed approach, which combines the aforementioned subspace tracking algorithms and the KF, is able to achieve better day ahead forecast accuracy than other conventional approaches for the New England dataset [18]. Moreover, similar results can also be observed for the Australian (New South Wales) half hourly dataset [37], which shows that the robust extension of the proposed RDFA algorithm outperforms other algorithms in the presence of undesired price spikes. Its efficient recursive implementation, ability on performing interval forecast and good performance make it as an attractive alternative to other conventional approaches used in electricity price forecasting and other related applications.

Finally, it should be noted that in some electricity markets such as the Australian NEM, it is a common practice for retailers to negotiate energy prices with the supplier through financial hedging contracts [39], [40], in order to safeguard the risks of exposing to the loss incurred on the extreme price spikes. For example, the retailer and the operator can be bounded under a "contract for differences", which specifies the price for the retailer to purchase a specified physical quantity of electricity from the supplier for a defined period of time. This price is usually referred to as the strike price. If the actual price on the subsequent day is higher than the strike price, the supplier will pay the difference between the actual price and the strike price to the retailer. Therefore, in the event of contingency, the retailer will receive refund from the supplier and hence it will never lose money due to the price spike. Conversely, if the retailer has negotiated a strike price that is always higher than the actual price, it will need to pay the difference to the supplier when the power plant is under normal operation and it will suffer from lost. Due to the above reasons, the main focus of this paper is on accurately predicting the general trend for the future electricity price as this will be useful for the retailer to negotiate a better strike price to minimize the risks of losing money.

Moreover, it should be noted that even for predicting the general trend, algorithms that employ the Gaussian assumption may degrade in prediction performance if the electricity price data contain price spikes. To be specific, it may affect the prediction accuracy if the price spikes are wrongly incorporated into the prediction model since they do not reflect the general trend under normal operation. In this regard, the proposed RRDFA algorithm is robust to the effect of the price spikes on predicting the general trend of the electricity price. If the values of the price spikes need to be forecasted, then relevant information has to be used, which is not currently modeled and explored by our algorithm.

The paper is organized as follows: The background of electricity forecasting algorithms are introduced in Section II. Afterwards, the proposed RDFA algorithm is discussed in Section III. Finally, real examples using the New England dataset [18] and Australian (New South Wales) [37] and comparisons with conventional algorithms are presented in Section IV. Finally, conclusions are drawn in Section V.

II. Background

A. Auto-Regression Model

A traditional approach to model electricity price is the AR based time series models. In particular, a variant called ARV model decomposes the electricity price $x(n)$, $n = 1, 2, \ldots, N$, into deterministic and stochastic components, denoted by $x_d(n)$ and $x_s(n)$, respectively, such that $x(n) = x_d(n) + x_s(n)$. The deterministic component $x_d(n)$, which is related to the seasonal trend of the electricity price, can be estimated using the Fourier series

$$x_d(n) = \gamma_0 + \sum_{i=1}^{K_0} \gamma_i \cos(i\omega_n) + \sum_{i=1}^{K_0-1} \tilde{\gamma}_i \sin(i\omega_n) \quad (1)$$

where $\omega_n = 2n\pi/K$, K is the *base period*, $K_0 = \lfloor K/2 \rfloor$ and $\lfloor . \rfloor$ is the floor operator. The coefficients γ_i, and $\tilde{\gamma}_i$ can be obtained by using the least squares (LS) method as in [19]. After $x_d(n)$ is computed, the stochastic component, which accounts for the short-term fluctuations, is computed by $x_s(n) = x(n) - x_d(n)$. Further, $x_s(n)$ is modeled as an AR process of order L

$$x_s(n) = \sum_{i=1}^{L} x_s(n-i)\alpha_i + e(n) \quad (2)$$

where $\alpha_i, i = 1, 2, \ldots L$, are the AR coefficients and $e(n)$ is the modeling error. The AR coefficient α_i in (2) can be determined by solving the following LS problem:

$$\min_{\boldsymbol{\alpha}} . \left(\|\boldsymbol{x}_n - \boldsymbol{x}_{n,d} - (\boldsymbol{X}_{n-1} - \boldsymbol{X}_{n-1,d})\boldsymbol{\alpha}\|_2^2 \right) \quad (3)$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_L]^T$, $\boldsymbol{X}_n = [\boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \ldots, \boldsymbol{x}_{n-L+1}]$, $\boldsymbol{x}_n = [x(n-L_0+1), \ldots, x(n-1), x(n)]^T$, and L_0 is the length of the sliding window. Here, $\boldsymbol{x}_{n,d}$ and $\boldsymbol{X}_{n,d}$ are defined similarly as $\boldsymbol{x}_n$ and $\boldsymbol{X}_n$, respectively. The LS solution to (3) is

$$\boldsymbol{\alpha} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T(\boldsymbol{x}_n - \boldsymbol{x}_{n,d}) \quad (4)$$

where $\boldsymbol{A} = \boldsymbol{X}_{n-1} - \boldsymbol{X}_{n-1,d}$ and the superscript T denotes matrix transposition.

B. Functional Principal Component Analysis

In FPCA [12], the electricity price is approximated by a linear combination of a set of orthogonal basis functions (vectors) with appropriate coefficients, which are referred to as the PC vectors (or simply PCs) and PC scores respectively. With an appropriate partitioning of the data, correlation between adjacent periods of interest can be more conveniently explored. Since the pattern of the price for adjacent partitions will be highly similar, FPCA can be used to explore these correlations.

More precisely, suppose we are given the electricity price of N days, $n = 1, 2, \ldots, N$. Consider J samples of the electricity price are collected at a regular interval at the nth day. The J samples can be grouped into a vector

$$\boldsymbol{z}(n) = [x((n-1)J+1), x((n-1)J+2), \ldots, x(nJ)]^T \quad (5)$$

where x is the hourly price and J is the number of partitions/variables, which is chosen as $J = 24$ for hourly price, then each vector represents the hourly price for the nth day. The value of J can also be adjusted for other time scales, such as $J = 48$ for half-hourly data for a day. Suppose that we are given the electricity price vectors of N days, i.e., $\boldsymbol{z}(n)$, $n = 1, 2, \ldots, N$. $\boldsymbol{z}(n)$ is usually "centered", i.e., with its mean removed, before the PC functions are computed. Hence, the mean of $\boldsymbol{z}(n)$, is first computed and is subtracted from each of the measurement vector to form $\overline{\boldsymbol{z}}(n)$. Let the electricity price after centering be $\boldsymbol{Z} = [\overline{\boldsymbol{z}}(1), \overline{\boldsymbol{z}}(2), \ldots, \overline{\boldsymbol{z}}(N)]^T$. In FPCA, we wish to express the centered electricity price vector $\overline{\boldsymbol{z}}(n)$ in terms of B PCs, i.e., $\overline{\boldsymbol{z}}(n) = \sum_{m=1}^{B} t_m(n)\boldsymbol{p}_m + \boldsymbol{e}(n)$, where B is an appropriately chosen number of PCs to achieve a sufficiently small approximation error $\boldsymbol{e}$, $\boldsymbol{p}_m$ is the mth PC, and $t_m(n)$ is its associated score for $\overline{\boldsymbol{z}}(n)$. Hence, $\boldsymbol{Z}$ can be written as

$$\boldsymbol{Z} = \sum_{m=1}^{B} \boldsymbol{t}_m \boldsymbol{p}_m^T + \boldsymbol{E} = \boldsymbol{\Gamma}\boldsymbol{P}^T + \boldsymbol{E} \tag{6}$$

where $\boldsymbol{t}_m = [t_m(1), \ldots, t_m(N)]^T$, $\boldsymbol{\Gamma} = [\boldsymbol{t}_1, \ldots., \boldsymbol{t}_B]$ is the score matrix, and $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots., \boldsymbol{p}_M]^T$ is the collection of PCs or loading matrix, and $\boldsymbol{E} = [\boldsymbol{e}(1), \boldsymbol{e}(2), \ldots, \boldsymbol{e}(N)]^T$ is the error matrix. A common way to determine the PCs is to compute the ED of the empirical correlation matrix: $\boldsymbol{C_{zz}} = (1/(N-1))\boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{U}\Lambda\boldsymbol{U}^T$, where the columns of $\boldsymbol{U}$ are the eigenvector and they are also the PCs and $\boldsymbol{\Lambda} = diag\{\lambda_1, \ldots, \lambda_J\}$ contains the eigenvalues in descending order of magnitude ($\lambda_1 \geq \lambda_2 \ldots \geq \lambda_J$). If the first B largest eigenvalues and their eigenvectors $\boldsymbol{U}_B$ are retained, then one gets $\boldsymbol{P} = \boldsymbol{U_B}$. The subspace spanned by the major PCs $\boldsymbol{P} = \boldsymbol{U_B}$ is usually referred to as the signal subspace. To handle possible system changes, such as the change of the trend, the data collected can be rearranged in a manner such that only the electricity price of D most recent days is retained, i.e.,

$$\boldsymbol{Z}(n) = [\boldsymbol{z}(n), \boldsymbol{z}(n-1), \ldots, \boldsymbol{z}(n-D+1)]^T \tag{7}$$

the correlation matrix can be computed similarly, i.e.,

$$\boldsymbol{C_{zz}}(n) = \frac{1}{D-1}\boldsymbol{Z}(n)^T\boldsymbol{Z}(n) = \boldsymbol{U}(n)\boldsymbol{\Lambda}(n)\boldsymbol{U}^T(n) \tag{8}$$

for each day n to update the PC.

To perform price forecasting using the updated PCs, time series models can be built for each PC score as they are highly correlated. This can be achieved for example by AR-based time series models. Consequently, if $\hat{t}_m(n+1)$ is the one step-ahead PC score forecast of the mth PC score predicted, say by the AR model in (2), then the one day-ahead forecast for the electricity price is

$$\hat{\boldsymbol{z}}(n+1) = \boldsymbol{\mu}(n) + \sum_{m=1}^{B} \hat{t}_m(n+1)\boldsymbol{p}_m(n) \tag{9}$$

where $\boldsymbol{\mu}(n) = (1/D)\sum_{i=1}^{D} \boldsymbol{z}(n-D+i)$ is the mean of $\boldsymbol{Z}(n)$. However, the online implementation requires batch eigen-decomposition (ED) of the covariance matrix in (8) at each update, which requires high arithmetic complexity.

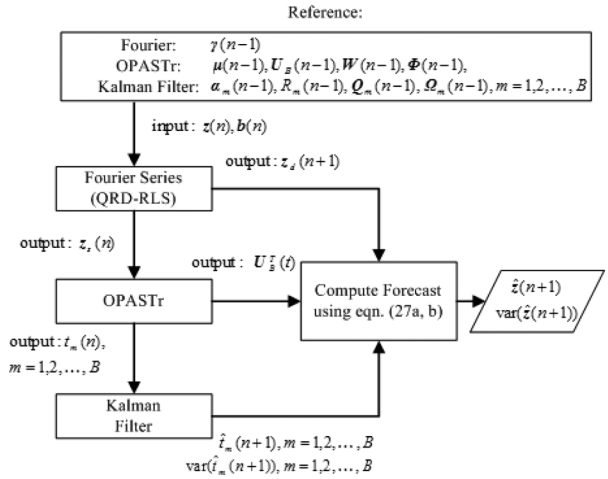


Fig. 1. Flow chart of the proposed RDFA algorithm.

III. Proposed Recursive Dynamic Factor Analysis

The proposed approach can be mainly divided into two steps:

1) The current electricity price $\boldsymbol{z}(n)$ is modeled as the sum of deterministic and stochastic components

$$\boldsymbol{z}(n) = \boldsymbol{z}_d(n) + \boldsymbol{z}_s(n) \tag{10}$$

where $\boldsymbol{z}_d(n)$ and $\boldsymbol{z}_s(n)$ are the deterministic and stochastic components, respectively. Here, the deterministic component $\boldsymbol{z}_d(n)$ is modeled as a Fourier series, which can be efficiently computed in a real-time manner using the RLS. The stochastic component $\boldsymbol{z}_s(n)$ is approximated by a linear combination of B PCs and PC scores, which are recursively computed with the low complexity subspace and PC tracking algorithms to be described later.

2) Then, the PC scores are modeled as AR model and its coefficients are assumed to be Gaussian distributed and are tracked using the KF. A distinct advantage of using the KF is that it provides the predicted AR coefficients estimate as well as its covariance. Hence the interval of the forecast values can be readily computed.

3) Furthermore, the proposed robust extension is employed to tackle possible non-Gaussian variation, such as price spikes. Firstly, price spikes are detected using certain robust detection criteria. Afterwards, the effect of these price spikes to the RDFA algorithm are suppressed or down-weighted using a modified Huber function so as to improve the robustness of the proposed algorithm in subsequent forecasting.

Fig. 1 shows a flow chart of the proposed RDFA algorithm. The rest of this section is organized as follows. First, recursive estimation of the Fourier series using the RLS is discussed in Section III-A. The subspace tracking algorithm is then discussed in Section III-B. The price and interval prediction using the KF is discussed in Section III-C. Finally, the robust extension is discussed in Section III-D.

A. Recursive Estimation of Fourier Series

As mentioned before, the deterministic component is modeled as a Fourier series, which can be solved

using the LS estimator. For efficient online computation, we employ the RLS estimator to solve for the Fourier coefficients recursively. More specifically, consider the deterministic component of the price vector $\boldsymbol{z}_d(n) = [z_{d,1}(n), z_{d,2}(n), \ldots, z_{d,J}(n)]^T$. For each variable $z_{d,j}(n), j = 1, 2, \ldots, J$, we construct a Fourier series $z_{d,j}(n) = \gamma_{0,j}(n) + \sum_{i=1}^{K_0} \gamma_{i,j}(n)\cos(i\omega_n) + \sum_{i=1}^{K_0-1} \tilde{\gamma}_{i,j}(n)\sin(i\omega_n)$ and the Fourier coefficients $\gamma_{i,j}$ and $\tilde{\gamma}_{i,j}$ can be determined by the RLS as follows:

$$\min . \left(\sum_{i=1}^{n} \beta_{LS}^{n-i} \left\| z_j(n) - \boldsymbol{b}^T(n)\boldsymbol{\gamma}_j(n) \right\|_2^2 \right) \tag{11}$$

where $\boldsymbol{z}(n) = [z_1(n), z_2(n), \ldots, z_J(n)]^T$, $\boldsymbol{b}(n) = [1, \cos(\omega_n), \ldots, \cos(K_0\omega_n), \sin(\omega_n) \ldots, \sin((K_0 - 1)\omega_n)]^T$ and $\boldsymbol{\gamma}(n) = [\gamma_{0,j}(n), \gamma_{1,j}(n), \ldots, \gamma_{K_0,j}(n), \gamma'_{1,j}(n), \gamma'_{2,j}(n), \ldots\ldots \gamma'_{K_0-1,j}(n)]^T$. Here, β_{LS} is a forgetting factor to give less weighting to data in the distant past in order to reflect changes in the process. It is well known that QR decomposition-based (QRD) method for solving RLS problem is numerically more stable than using the matrix inversion formula under finite wordlength implementation. Also, efficient hardware implementation in form of systolic arrays and Cordic processors are readily available. Interested readers are referred to [20] and [21] for detailed derivation of the QRD-RLS. Comparing with batch processing using the conventional LS, which requires a complexity of $O(K^3)$, the complexity of the QRD-RLS algorithm is only $O(K^2)$. Afterwards, the deterministic component $\boldsymbol{z}_d(n)$ is subtracted from the price vector and the resultant stochastic component $\boldsymbol{z}_s(n) = \boldsymbol{z}(n) - \boldsymbol{z}_d(n)$ is centered using a recursive mean estimator

$$\boldsymbol{\mu}(n) = \beta_\mu \boldsymbol{\mu}(n-1) + (1 - \beta_\mu)\boldsymbol{z}_s(n). \tag{12}$$

Here, β_μ is a forgetting factor for the recursive mean. The centered stochastic component, i.e.,

$$\overline{\boldsymbol{z}}_s(n) = \boldsymbol{z}_s(n) - \boldsymbol{\mu}(n) \tag{13}$$

is then used in the recursive estimation of the signal subspace.

B. Subspace Tracking

We proposed to estimate the signal subspace recursively by the OPASTr algorithm [13] summarized in Table I, which is motivated by the PAST algorithm proposed in [16]. Particularly, the signal subspace spanned by the major PCs $\boldsymbol{U}_B(n)$ is tracked recursively instead of computing the entire ED. In the PAST algorithm [16], the subspace $\boldsymbol{W}(n)$ is recursively computed by minimizing the following objective function:

$$J(\boldsymbol{W}(n)) = \sum_{i=1}^{n} \beta^{n-i} \left\| \overline{\boldsymbol{z}}_s(i) - \boldsymbol{W}(n)\overline{\boldsymbol{y}}(i) \right\|_2^2 . \tag{14}$$

Ideally, $\overline{\boldsymbol{y}}(i) = \boldsymbol{W}^T(n)\overline{\boldsymbol{z}}_s(i)$ and $J(\boldsymbol{W}(n))$ represents the energy in $\overline{\boldsymbol{z}}_s(i)$ which is outside the subspace $\boldsymbol{W}(n)$. Hence, $\boldsymbol{W}(n)$ is equal to the major PCs $\boldsymbol{U}_B(n)$ up to an orthogonal transformation or rotation, i.e., $span(\boldsymbol{W}(n)) = span(\boldsymbol{U}_B(n))$, and the outer product $\boldsymbol{W}(n)\boldsymbol{W}^T(n)$ is equal to $\boldsymbol{U}_B(n)\boldsymbol{U}_B^T(n)$. In the PAST algorithm, the projection approximation $\overline{\boldsymbol{y}}(i) \approx \boldsymbol{W}^T(i-1)\overline{\boldsymbol{z}}_s(i)$ is employed so that (14) can be relaxed to a quadratic function in $\boldsymbol{W}(n)$. Consequently, conventional RLS algorithm can be applied to solve for $\boldsymbol{W}(n)$ with very low arithmetic complexity. In the OPAST algorithm [17], an extra orthonormalization step is added to the PAST algorithm to guarantee the orthonormality of the estimated signal subspace $\boldsymbol{W}(n)$.

To apply the OPAST algorithm, an initial ED is assumed to be available either by performing an ED on an initial data block or pre-determining the eigenvalues offline. The eigenvalues so obtained can be used with the minimum description length (MDL) criterion [21] to estimate the dimension B of the signal subspace. During online application, the OPAST algorithm is invoked to update the signal subspace recursively. Due to page limitation, the detailed derivations of the PAST and the OPAST algorithms are omitted and interested readers are referred to [16], [17] for more details. However, in computing the PCs, the PAST and OPAST algorithms are not directly applicable due to the arbitrary orthogonal rotation mentioned above.

In the OPASTr algorithm [13], the PCs are further extracted from the signal subspace tracked. Given the signal subspace

TABLE I
OPASTr ALGORITHM

Initialization:

Initialize $\boldsymbol{\mu}(0) = \text{mean}(\boldsymbol{Z}_s(0))$, where the initial data block is $\boldsymbol{Z}_s(0) = [\boldsymbol{z}_s(1), \boldsymbol{z}_s(2), \ldots, \boldsymbol{z}_s(N_0)]$ and N_0 is the number of measurements in $\boldsymbol{Z}_s(0)$. Obtain $\boldsymbol{U}_B(0)$, $\boldsymbol{\Lambda}_B(0)$ from $\boldsymbol{C}_{zz}(0) = (1/L_d - 1)\boldsymbol{Z}_s^T(0)\boldsymbol{Z}_s(0)$, $\boldsymbol{W}(0) = \boldsymbol{U}_B(0)$, $\boldsymbol{\Omega}(0) = \boldsymbol{C}_{yy}^{-1}(0)$, $\boldsymbol{C}_{yy}(0) = \boldsymbol{W}^T(0)\boldsymbol{C}_{zz}(0)\boldsymbol{W}(0) = \boldsymbol{\Phi}(0)\boldsymbol{\Lambda}_B(0)\boldsymbol{\Phi}^T(0)$. β, β_μ are the forgetting factors for the OPASTr and the recursive mean estimator respectively.

Recursion:

For $n = 1,2\cdots, N$ *do*

$\boldsymbol{\mu}(n) = \beta_\mu \boldsymbol{\mu}(n-1) + (1-\beta_\mu)\boldsymbol{z}_s(n))$,

$\overline{\boldsymbol{z}}_s(n) = \boldsymbol{z}_s(n) - \boldsymbol{\mu}(n)$,

$\overline{\boldsymbol{y}}(n) = \boldsymbol{W}^T(n-1)\overline{\boldsymbol{z}}(n)$,

$\boldsymbol{g}(n) = (1/\beta)\boldsymbol{\Omega}(n-1)\overline{\boldsymbol{y}}(n)$,

$\theta(n) = (1/1 + \overline{\boldsymbol{y}}^H(n)\boldsymbol{g}(n))$,

$\boldsymbol{p}(n) = \gamma(n)\left(\overline{\boldsymbol{z}}_s(n) - \boldsymbol{W}(n-1)\overline{\boldsymbol{y}}(n)\right)$,

Orthonormalization Step for the estimated subspace

$$r(n) = \left(1/\left\|\boldsymbol{g}(n)\right\|_2^2\right)\left(\left(1/\sqrt{1 + \left\|\boldsymbol{p}(n)\right\|_2^2 \left\|\boldsymbol{g}(n)\right\|_2^2}\right) - 1\right),$$

$$\boldsymbol{p}'(n) = \tau(n)\boldsymbol{W}(n-1)\boldsymbol{g}(n) + (1 + r(n)\left\|\boldsymbol{g}(n)\right\|_2^2)\boldsymbol{p}(n),$$

Update:

$\boldsymbol{W}(n) = \boldsymbol{W}(n-1) + \boldsymbol{p}'(n)\boldsymbol{g}^T(n)$,

$\boldsymbol{\Xi}(n) = (1/\beta)\boldsymbol{\Xi}(n-1) - \theta(n)\boldsymbol{g}(n)\boldsymbol{g}^T(n)$.

Subspace eigenvector and eigenvalue computation

$\boldsymbol{y}(n) = \boldsymbol{W}^T(n)\overline{\boldsymbol{z}}(n)$,

Update $\boldsymbol{\Phi}(n)$ and $\boldsymbol{\Lambda}_B(n)$ using the rank-1 modification in Table II(B).

$[\boldsymbol{\Phi}(n), \boldsymbol{\Lambda}_B(n)] = \text{rank-1-modification}\ (\boldsymbol{\Phi}(n-1), \boldsymbol{\Lambda}_B(n-1), \boldsymbol{y}(n))$.

Compute $\boldsymbol{U}_B(n) = \boldsymbol{W}(n)\boldsymbol{\Phi}(n)$ using (16).

End for

TABLE II
RANK-1-MODIFICATION

Input: $\boldsymbol{\Phi}(n-1), \boldsymbol{\Lambda}_B(n-1), \boldsymbol{y}(n)$ **Output:** $\boldsymbol{\Phi}(n), \boldsymbol{\Lambda}_B(n)$.

$\boldsymbol{s}(n) = \boldsymbol{\Phi}^T(n-1)\boldsymbol{y}(n)$, $\tilde{\boldsymbol{s}}(n) = \boldsymbol{s}(n)/\|\boldsymbol{s}(n)\|_2$,

Solve the secular equation $1+\rho\sum_{m=1}^{B}\tilde{s}_m^2(n)/(\lambda_m(n)-\beta\lambda_m(n-1))=0$, using the Newton method reported in [26], where $\rho=(1-\beta)\|\boldsymbol{s}(n)\|_2^2$, $\tilde{\boldsymbol{s}}(n)=[\tilde{s}_1(n),\tilde{s}_2(n),\ldots,\tilde{s}_B(n)]^T$.

Recursion:

For $m=1,2\cdots,B$ *do*

$\tilde{\boldsymbol{\varphi}}_m(n)=\left((\beta\boldsymbol{\Lambda}_B(n-1)-\lambda_m(n)\boldsymbol{I})^{-1}\tilde{\boldsymbol{s}}(n)\right)\Big/\left\|(\beta\boldsymbol{\Lambda}_B(n-1)-\lambda_m(n)\boldsymbol{I})^{-1}\tilde{\boldsymbol{s}}(n)\right\|_2$,

End for

$\tilde{\boldsymbol{\Phi}}(n)=[\tilde{\boldsymbol{\varphi}}_1(n),\tilde{\boldsymbol{\varphi}}_2(n),\ldots,\tilde{\boldsymbol{\varphi}}_B(n)]$,

$\boldsymbol{\Phi}(n)=\boldsymbol{\Phi}(n-1)\tilde{\boldsymbol{\Phi}}(n)$, $\boldsymbol{\Lambda}_B(n)=diag(\lambda_1(n),\lambda_2(n),\ldots,\lambda_B(n))$.

$\boldsymbol{W}(n)$, the covariance matrix $\boldsymbol{C}_{zz}(n)=\boldsymbol{U}(n)\boldsymbol{\Lambda}(n)\boldsymbol{U}^T(n)$ is projected onto the signal subspace $\boldsymbol{W}(n)$ to obtain

$$\begin{aligned}\boldsymbol{C}_{yy}(n)&=\boldsymbol{W}^T(n)\boldsymbol{U}(n)\boldsymbol{\Lambda}(n)\boldsymbol{U}^T(n)\boldsymbol{W}(n)\\&=\boldsymbol{W}^T(n)\boldsymbol{U}_B(n)\boldsymbol{\Lambda}_B(n)\boldsymbol{U}_B^T(n)\boldsymbol{W}(n)\\&=\boldsymbol{\Phi}(n)\boldsymbol{\Lambda}_B(n)\boldsymbol{\Phi}^T(n)\end{aligned}\tag{15}$$

where $\boldsymbol{\Phi}(n)$ is a$B\times B$ orthogonal transformation satisfying $\boldsymbol{\Phi}(n)\boldsymbol{\Phi}^T(n)=\boldsymbol{I}$ and

$$\boldsymbol{U}_B(n)=\boldsymbol{W}(n)\boldsymbol{\Phi}(n).\tag{16}$$

The covariance matrix $\boldsymbol{C}_{yy}(n)=E[\boldsymbol{y}(n)\boldsymbol{y}^T(n))]$ can be recursively updated as

$$\boldsymbol{C}_{yy}(n)=\beta\boldsymbol{C}_{yy}(n-1)+(1-\beta)\boldsymbol{y}(n)\boldsymbol{y}^T(n)\tag{17}$$

where $\boldsymbol{y}(n)=\boldsymbol{W}^T(n)\overline{\boldsymbol{z}}_s(n)$. Here, we remark that $\boldsymbol{y}(n)$ is a projection of $\overline{\boldsymbol{z}}_s(n)$ on the subspace $\boldsymbol{W}(n)$ and it is different from the projection approximation $\overline{\boldsymbol{y}}(n)$ in (14). $\boldsymbol{\Phi}(n)$ can be recursively computed using the ED of $\boldsymbol{C}_{yy}(n)$. Firstly, let the ED of $\boldsymbol{C}_{yy}(n-1)$ be $\boldsymbol{\Phi}(n-1)\boldsymbol{\Lambda}_B(n-1)\boldsymbol{\Phi}^T(n-1)$. Equation (17) can be rewritten as one rank-1 modification given by

$$\boldsymbol{C}_{yy}(n)=\boldsymbol{\Phi}(n-1)\left[\beta\boldsymbol{\Lambda}_B(n-1)+(1-\beta)\boldsymbol{s}(n)\boldsymbol{s}^T(n)\right]\boldsymbol{\Phi}(n-1)^T\tag{18}$$

where $\boldsymbol{s}(n)=\boldsymbol{\Phi}^T(n-1)\boldsymbol{y}(n)$. Let the corresponding ED be

$$\beta\boldsymbol{\Lambda}_B(n-1)+(1-\beta)\boldsymbol{s}(n)\boldsymbol{s}^T(n)=\tilde{\boldsymbol{\Phi}}(n)\boldsymbol{\Lambda}_B(n)\tilde{\boldsymbol{\Phi}}^T(n).\tag{19}$$

The ED of the rank-1 update in (19) can be recursively computed using rank-1 modification [14], [22], which is summarized in Table II. Finally, the eigenvectors of $\boldsymbol{C}_{yy}(n)$ are given by

$$\boldsymbol{\Phi}(n)=\boldsymbol{\Phi}(n-1)\tilde{\boldsymbol{\Phi}}(n).\tag{20}$$

Consequently, the new PCs can be computed according to (16). Then, the PC scores $\boldsymbol{\tau}(n)=[t_1(n),t_2(n),\ldots t_B(n)]^T$ can be computed recursively as

$$\boldsymbol{\tau}(n)=\boldsymbol{U}_B^T(t)\overline{\boldsymbol{z}}_s(n).\tag{21}$$

C. Kalman Filter for Price and Interval forecasting

After PCs and PC scores are recursively updated, separate time series models such as the AR model can be built for each PC score to perform prediction because the PCs are orthogonal to each other [12], [23], [24]. Unlike the FPCA in [12], we employed the KF to recursively track the time series so that the price and interval forecast can be computed online in a real-time manner. More precisely, for each PC score $t_m(n)$ of the $m-th$ PC obtained in (21), $m=1,2,\ldots B$, an AR model of Lth order will be constructed. In contrast to the conventional LS formulation in (3), we incorporate a regularization term $\|\boldsymbol{\alpha}_m(i)-\boldsymbol{\alpha}_m(i-1)\|_2^2$ as follows:

$$\begin{aligned}\min\Bigg\{&\sum_{i=1}^{n}\left\|R_m^{-1/2}(i)\left(t_m(i)-\Sigma_{j=1}^{L}\alpha_j(i)t_m(i-j)\right)\right\|_2^2\\&+\sum_{i=1}^{n}\left\|\boldsymbol{Q}_m^{-1/2}(i)\left(\boldsymbol{\alpha}_m(i)-\boldsymbol{\alpha}_m(i-1)\right)\right\|_2^2\Bigg\}\end{aligned}\tag{22}$$

where $\boldsymbol{\alpha}_m(i)=[\alpha_{m,1}(i),\alpha_{m,2}(i),\ldots,\alpha_{m,L}(i)]^T$ are the AR coefficients, $R_m(i)$ and $\boldsymbol{Q}_m(i)$ are the covariance of the loss function $e_m(i)=t_m(i)-\Sigma_{j=1}^{L}\alpha_j(i)t_m(i-j)$ and $\boldsymbol{\varepsilon}_m(i)=\boldsymbol{\alpha}_m(i)-\boldsymbol{\alpha}_m(i-1)$ is the regularization term, respectively. The inverses $R_m^{-1/2}(i)$ and $\boldsymbol{Q}_m^{-1/2}(i)$ are used to perform scaling on each variable (whitening) in order to achieve equal variance of the transformed variables. The regularization term requires the estimate to stay close to the previous estimate and hence the variance of the estimator will be reduced. It is shown in [25] that (22) can be formulated as the following state space model (SSM):

$$\boldsymbol{\alpha}_m(n)=\boldsymbol{\alpha}_m(n-1)+\boldsymbol{\varepsilon}_m(n),\tag{23a}$$

$$t_m(n)=\boldsymbol{h}_m(n)^T\boldsymbol{\alpha}_m(n)+e_m(n).\tag{23b}$$

Equation (23a) is the state equation and it describes the evolution of the AR coefficients over time, as a function of the previous AR coefficients $\boldsymbol{\alpha}_m(n)$ and $\boldsymbol{\varepsilon}_m(n)$ represents the modeling error. Eqn. (23b) is the measurement equation which models the current PC scores $t_m(n)$ with previous scores $\boldsymbol{h}_m(n)=[t_m(n-1),t_m(n-2),\ldots,t_m(n-L)]^T$, and the AR coefficients $\boldsymbol{\alpha}_m(n)$ represent the weighting of each previous score $t_m(n-i)$, $i=1,2,\ldots L$, and $e_m(n)$ is the measurement noise. We can see that the state equation in (23a) and the measurement equations in (23b) are equivalent to the regularization term and the loss function in (22) respectively. The SSM in (23a) and (23b) can be recursively tracked using the KF as follows:

Predict

$$\begin{aligned}\boldsymbol{\alpha}_m(n|n-1)&=\boldsymbol{\alpha}_m(n-1|n-1),\\\boldsymbol{\Omega}_m(n|n-1)&=\boldsymbol{\Omega}_m(n-1|n-1)+\boldsymbol{Q}_m(n)\end{aligned}\tag{24a}$$

Update

$$\begin{aligned}e_m(n)&=t_m(n)-\boldsymbol{h}_m^T(n)\boldsymbol{\alpha}_m(n-1|n-1),\\S_m(n)&=\boldsymbol{h}_m^T(n)\boldsymbol{\Omega}_m(n|n-1)\boldsymbol{h}_m(n)+R_m(n),\\\boldsymbol{K}_m(n)&=\boldsymbol{\Omega}_m(n|n-1)\boldsymbol{h}_m(n)S_m(n)^{-1},\\\boldsymbol{\alpha}_m(n|n)&=\boldsymbol{\alpha}_m(n|n-1)+\boldsymbol{K}_m(n)e_m(n),\\\boldsymbol{\Omega}_m(n|n)&=\boldsymbol{\Omega}_m(n|n-1)-\boldsymbol{K}_m(n)\boldsymbol{h}_m^T(n)\boldsymbol{\Omega}_m(n|n-1)\end{aligned}\tag{24b}$$

where $\boldsymbol{\Omega}_m(n)$ is the covariance estimate of the AR coefficients.

A major difficulty in practical implementation of the KF is that the covariances $\boldsymbol{Q}_m(n)$ and $R_m(n)$ are often unknown.

In [26], the following recursive covariance estimators for estimating these covariances from the state error $\boldsymbol{\varepsilon}_m(n)$ and measurement error $e_m^2(n)$ for the KF were proposed:

$$\boldsymbol{Q}_m(n) = \beta_{Q\_m}\boldsymbol{Q}_m(n-1) + (1-\beta_{Q\_m})\boldsymbol{\varepsilon}_m(n)\boldsymbol{\varepsilon}_m^T(n), \quad (25a)$$

$$R_m(n) = \beta_{R\_m}R_m(n-1) + (1-\beta_{R\_m})e_m^2(n) \quad (25b)$$

where β_{Q_m} and β_{R_m} are forgetting factors for the recursive estimators of $\boldsymbol{Q}_m(n)$ and $R_m(n)$, respectively. $\boldsymbol{\varepsilon}_m(n)$ and $e_m(n)$ are the modeling error and observation noise as defined in (23a) and (23b), respectively. Hence, the one-step ahead prediction $\hat{t}_m(n+1)$ is given by

$$\hat{t}_m(n+1) = \boldsymbol{h}_m^T(n+1)\boldsymbol{\alpha}_m(n+1|n), \quad (26a)$$

$$\text{var}\left(\hat{t}_m(n+1)\right) = \boldsymbol{h}_m^T(n+1)\boldsymbol{\Omega}_m(n+1|n)\boldsymbol{h}_m(n+1) + \hat{R}_m(n+1) \quad (26b)$$

where $\hat{R}_m(n+1)$ is the predicted covariance. Since the future samples are not available, the most recent estimate $R_m(n)$ is used (i.e.. $\hat{R}_m(n+1) = R_m(n)$). From (26a) and (26b), the one day ahead prediction of the electricity price vector $\hat{\boldsymbol{z}}(n+1)$ can be determined as follows:

Point Forecast

$$\hat{\boldsymbol{z}}(n+1) = \boldsymbol{\mu}(n) + \boldsymbol{z}_d(n+1) + \Sigma_{m=1}^{B}\hat{t}_m(n+1)\boldsymbol{u}_m(n) \quad (27a)$$

Interval Forecast

$$\text{var}\left(\hat{\boldsymbol{z}}(n+1)\right) = \Sigma_{m=1}^{B}\text{var}\left(\hat{t}_m(n+1)\right)\boldsymbol{u}_m(n)\boldsymbol{u}_m(n)^T. \quad (27b)$$

The point forecast $\hat{\boldsymbol{z}}(n+1)$ is derived similarly as in (9) except that the mean $\boldsymbol{\mu}(n)$ is obtained by the recursive mean estimator in (12) and the deterministic component $\boldsymbol{z}_d(n+1)$ is obtained from the Fourier series in (11). Interested readers are referred to the supplementary material [15] for the detailed derivation of the interval forecast in (27b). To perform two days ahead prediction, one can append the prediction $\hat{t}_m(n+1)$ into $\boldsymbol{h}_m(n+2) = [\hat{t}_m(n+1), t_m(n), t_m(n-1), \ldots, t_m(n-L+2)]^T$, and then apply the KF again to predict $\hat{t}_m(n+2)$ and $\text{var}(\hat{t}_m(n+2))$ using (26a) and (26b), respectively. This procedure can be repeated for h times to compute a h days ahead prediction.

D. Robust M-Estimation

In some markets, a sudden surge in demand or forced outage may result in price spikes, which can deviate greatly from the price under normal operation. For example, the electricity price for the Australian market can goes up to several thousand Australian dollar in a short instant. The Gaussian assumption may not be valid and algorithms based on such assumption may suffer from degradation in forecasting accuracy after the occurrence of these price spikes. To address this important issue, we shall further extend our RDFA algorithm to tackle these price spikes by further incorporating the concept of robust M-estimation [33], which is a technique used to improve the robustness against the effect of the outliers such as price spikes. For the sake of presentation, we shall call it recursive robust DFA (RRDFA). In the new RRDFA algorithm, appropriate robust detection criteria are applied to the Fourier series estimation, subspace tracking and KF so as to detect the price spikes during the parameter estimation in the respective algorithms. Moreover, after these spikes are detected, we consider the following two approaches to reduce their effect on the performance of the RRDFA algorithm:

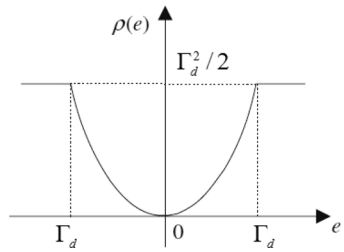


Fig. 2. Modified Huber M-estimate function.

Approach 1: If a segment over the prediction horizon is detected as price spikes, it will not be used in subsequent forecasting.

Approach 2: It is similar to approach 1 except that the contributions of the detected price spikes are down-weighed using a modified Huber function instead of removing them from subsequent forecasting, as shown in Fig. 2.

More precisely, for each daily electricity price vector $\boldsymbol{z}(n)$ defined in (5), we employ the following error measures to detect the price spikes in the Fourier series, subspace computation and KF respectively

Fourier Series:

$$e_{j,fourier}(n) = z_j(n) - \boldsymbol{b}^T(n)\boldsymbol{\gamma}_j(n), \quad j = 1, 2, \ldots, J \quad (28a)$$

Subspace:

$$\text{SPE}(n) = \|\overline{\boldsymbol{z}}_s(n) - \boldsymbol{W}(n)\overline{\boldsymbol{y}}(n)\|_2^2, \quad (28b)$$

$$T^2(n) = \left\|\overline{\boldsymbol{z}}_s^T(n)\boldsymbol{U}_B(n)\boldsymbol{\Lambda}_B^{-1}(n)\boldsymbol{U}_B^T(n)\overline{\boldsymbol{z}}_s(n)\right\|_2^2 \quad (28c)$$

Kalman filter:

$$e_m(n) = t_m(n) - \boldsymbol{h}_m^T(n)\boldsymbol{\alpha}_m(n-1|n-1), \quad m = 1, 2, \ldots B \quad (28d)$$

where $e_{j,fourier}(t)$ and $e_m(n)$ are the prediction errors for the Fourier series and KF in (11) and (24b), respectively. $SPE(t)$ and $T^2(t)$ are the squared prediction error (SPE) and T^2 score, respectively, which are detection measures commonly employed in PCA-based fault detection [34]. The following robust detection criteria are employed to detect the price spikes:

Fourier Series:

$$|\Delta_{j,fourier}(n)| > \xi\sigma_{j,fourier}(n) \quad (29a)$$

Subspace:

$$|\Delta_{\text{SPE}}(n)| > \xi\sigma_{\text{SPE}}(n), \quad (29b)$$

$$|\Delta_{T^2}(n)| > \xi\sigma_{T^2}(n) \quad (29c)$$

Kalman filter:

$$|\Delta_m(n)| > \xi\sigma_m(n) \quad (29d)$$

where $\Delta_{j,fourier}(n) = e_{j,fourier}(n)$, $\Delta_{\text{SPE}}(n) = \text{SPE}(n) - \mu_{\text{SPE}}(n)$, $\Delta_{T^2}(n) = T^2(n) - \mu_{T^2}(n)$ and $\Delta_m(n) = e_m(n)$.

ξ is a threshold quartile parameter corresponding to the upper $(1-P\{X>\xi\})$ percentile of the Gaussian distribution. Hence, the probability that the occurrence of a price spike exceeds the threshold quartile is $P\{X>\xi\}=(2/\sqrt{\pi})\int_{\xi}^{\infty}e^{-x^2}dx$. The value of ξ is chosen to achieve a certain detection rate. In many applications, the probability $P\{-\xi\le X\le\xi\}=0.99$ is often used [35], [36], which means that the probability of occurrence for a normal sample is 99%. $\mu_g(n)$ and $\sigma_g(n)$ are recursive robust location and scale estimators, respectively, and they are given by

$$\mu_{\rm d}(n)=\beta_d\mu_{\rm d}(n-1)+(1-\beta_d){\rm med}(A(d(n)))\,, \tag{30a}$$

$$\sigma^2_{\rm d}(n)=\beta_{\sigma\_d}\sigma^2_{\rm d}(n-1)+c(1-\beta_{\sigma\_d}){\rm med}\left(A\left(\Delta_{\rm d}(n)^2\right)\right) \tag{30b}$$

where $d(n)$ denotes either $SPE(n)$ or $T^2(n)$, $\Delta_{\rm d}(n)$ denotes either $\Delta_{j,fourier}(n)$, $\Delta_{\rm SPE}(n)$, $\Delta_{T^2}(n)$ and $\Delta_m(n)$, $A(d(n))=\{d(n),d(n-1),\ldots,d(n-L_d+1)\}$, L_d is the window length, med(.) is the median operator, $c=2.13$ is a correction factor for Gaussian input. β_d and β_{σ_d} are forgetting factors with values slightly smaller than 1.

If any of the error measures in (28a)to (28d) exceed the detection thresholds in (29a)to (29d), the price vector $\boldsymbol{z}(n)$ will be identified as an abnormal sample. Then, either the price vector $\boldsymbol{z}(n)$, i.e., electricity prices for the whole day, is removed from the dataset or their effect on estimating the Fourier coefficients $\boldsymbol{\gamma}_j(n)$, the subspace $\boldsymbol{W}(n)$ and the AR coefficients $\boldsymbol{\alpha}_m(n)$ are down-weighed using the modified Huber function

$$\rho_{mh}\left(d(n)\right)=\begin{cases}\frac{\Delta_d^2(n)}{2} & 0\le|\Delta_d(n)|\le\Gamma_d(n)\\ \frac{\Gamma_d^2(n)}{2} & \text{otherwise}\end{cases} \tag{31}$$

where $\Gamma_d(n)=\xi\sigma_d(n)$. From (31), we can see that the threshold $\Gamma_d(n)$ is able to change upon time. First, this function is incorporated into the objective function of the RLS for obtaining the Fourier coefficients in (11) as follows:

$$\min.\left(\sum_{i=1}^{n}\beta_{LS}^{n-i}\rho_{mh}\left(e_{j,fourier}(n)\right)\right. . \tag{32}$$

From (31) and (32), we can see that if a sample obtains an error that exceeds the threshold in (31), it will receive a constant penalty instead of the conventional squared error. In this regard, no matter how large the error of the sample is, the influence of that sample towards the optimization is bounded by the threshold in (31). This concept is regarded as robust M-estimation. The optimal solution $\boldsymbol{\gamma}_j(n)$ for minimizing this re-weighted objective function can be obtained by differentiating (32) with respect to $\boldsymbol{\gamma}_j(n)$ and setting the derivatives to zero. This yields the following M-estimate normal equation:

$$\boldsymbol{R}_j(n)\boldsymbol{\gamma}_j(n)=\boldsymbol{C}_j(n) \tag{33}$$

where

$$\begin{aligned}\boldsymbol{R}_j(n)&=\sum_{i=1}^{n}\beta_{LS}^{n-i}q\left(e_{j,fourier}(n)\right)\boldsymbol{b}(i)\boldsymbol{b}^T(i)\\&=\beta_{LS}\boldsymbol{R}_j(n-1)+q\left(e_{j,fourier}(n)\right)\boldsymbol{b}(n)\boldsymbol{b}^T(n),\end{aligned} \tag{34}$$

$$\begin{aligned}\boldsymbol{C}_j(n)&=\sum_{i=1}^{n}\beta_{LS}^{n-i}q\left(e_{j,fourier}(n)\right)z_j(i)\boldsymbol{b}(i)\\&=\beta_{LS}\boldsymbol{C}_j(n-1)+q\left(e_{j,fourier}(n)\right)z_j(n)\boldsymbol{b}(n))\end{aligned} \tag{35}$$

and $q(e_{j,fourier}(n))=1$ for $0\le|\Delta_{j,fourier}(n)|\le\Gamma_{j,fourier}(n)$ or 0 otherwise. According to [33], the following recursive least M-estimate can be used to solve for (33) by applying the matrix inversion lemma:

$$\begin{aligned}&\boldsymbol{V}_j(n)\\&\quad=\beta_{LS}^{-1}\left(\boldsymbol{I}_{2K_0-1}-\boldsymbol{K}_{RLS}(n)\boldsymbol{b}^T(n)\right)\boldsymbol{V}(n-1),\end{aligned} \tag{36a}$$

$$\begin{aligned}&\boldsymbol{K}_{j,RLS}(n)\\&\quad=\frac{q\left(e_{j,fourier}(n)\right)\boldsymbol{V}(n-1)\boldsymbol{b}(n)}{\beta_{LS}+q\left(e_{j,fourier}(n)\right)\boldsymbol{b}^T(n)\boldsymbol{V}(n-1)\boldsymbol{b}(n)},\end{aligned} \tag{36b}$$

$$\begin{aligned}&\boldsymbol{\gamma}_j(n)\\&\quad=\boldsymbol{\gamma}_j(n-1)+\left(z_j(n)-\boldsymbol{b}^T(n)\boldsymbol{\gamma}_j(n-1)\right)\boldsymbol{K}_{j,RLS}(n)\end{aligned} \tag{36c}$$

where $\boldsymbol{V}_j(n)=\boldsymbol{R}_j^{-1}(n)$. After obtaining the Fourier coefficients, the modified Huber function can be incorporated to the OPASTr algorithm similarly as in (32)–(36) as it is based on RLS. First, the recursive mean estimator in (12) is replaced by the following recursive robust location estimator:

$$\boldsymbol{\mu}(n)=\beta_\mu\boldsymbol{\mu}(n-1)+(1-\beta_\mu){\rm med}\left(A\left(z_s(n)\right)\right). \tag{37}$$

Afterwards, the subspace update formulas in Table I, which are also shown in the following

$$\boldsymbol{W}(n)=\boldsymbol{W}(n-1)+\boldsymbol{p}'(n)\boldsymbol{g}^T(n), \tag{38a}$$

$$\boldsymbol{\Xi}(n)=\left(\frac{1}{\beta}\right)\boldsymbol{\Xi}(n-1)-\theta(n)\boldsymbol{g}(n)\boldsymbol{g}^T(n) \tag{38b}$$

are modified to the following robust counterparts, respectively:

$$\begin{aligned}\boldsymbol{W}(n)=&\boldsymbol{W}(n-1)\\&+q_{T^2}\left(\Delta_{T^2}(n)\right)q_{SPE}\left(\Delta_{SPE}(n)\right)\boldsymbol{p}'(n)\boldsymbol{g}^T(n),\end{aligned} \tag{39a}$$

$$\begin{aligned}\boldsymbol{\Xi}(n)=&\left(1-q_{T^2}(n)q_{SPE}(n)\right)\boldsymbol{\Xi}(n-1)\\&+q_{T^2}(n)q_{SPE}(n)\left(\frac{1}{\beta}\right)\\&\times\left(\boldsymbol{\Xi}(n-1)-\theta(n)\boldsymbol{g}(n)\boldsymbol{g}^T(n)\right)\end{aligned} \tag{39b}$$

where $q_{T^2}(.)$ and $q_{SPE}(.)$ are robust weighing functions defined similarly as in (34) except that their thresholds are replaced by $\Gamma_{T^2}(n)$ and $\Gamma_{SPE}(n)$, respectively. For notation convenience, we have used $q_{T^2}(n)$ and $q_{SPE}(n)$ for $q_{T^2}(\Delta_{T^2}(n))$ and $q_{SPE}(\Delta_{SPE}(n))$, respectively. Moreover, the recursive covariance estimate in (17) is replaced by the following recursive robust covariance estimate:

$$\begin{aligned}\boldsymbol{C}_{yy}(n)=&\left(1-q_{T^2}(n)q_{SPE}(n)\right)\boldsymbol{C}_{yy}(n-1)\\&+q_{T^2}(n)q_{SPE}(n)\left(\beta\boldsymbol{C}_{yy}(n-1)+(1-\beta)\boldsymbol{y}(n)\boldsymbol{y}^T(n)\right).\end{aligned} \tag{40}$$

Finally, after obtaining the subspace estimate, the robust M-estimation can be incorporated by modifying the Kalman gain in (24b) as follows:

$$\boldsymbol{K}_m(n)=q_m(n)\boldsymbol{\Omega}_m(n|n-1)\boldsymbol{h}_m(n)S_m(n)^{-1} \tag{41}$$

where $q_m(n)=q_m(e_m(n))$ is the robust weighting function with thresholds $\Gamma_m(n)$ and $e_m(n)$ is the error defined in (28d).

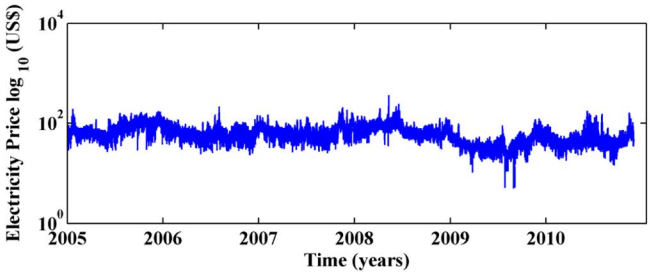


Fig. 3 Hourly electricity price of the New England dataset from January 1, 2005 to September 3, 2010.

The recursive covariance estimators for the measurement and state noise are replaced by their robust counterparts as follows:

$$\begin{aligned}\boldsymbol{Q}_m(n) &= (1-q_m(n))\,\boldsymbol{Q}_m(n-1)\\ &\quad + q_m(n)\big(\beta_{Q\_m}\boldsymbol{Q}_m(n-1)\\ &\qquad + (1-\beta_{Q\_m})\boldsymbol{\varepsilon}_m(n)\boldsymbol{\varepsilon}_m^T(n)\big), \quad (42a)\\ R_m(n) &= (1-q_m(n))\,R_m(n-1)\\ &\quad + q_m(n)\big(\beta_{R\_m}R_m(n-1)\\ &\qquad + (1-\beta_{R\_m})e_m^2(n)\big). \quad (42b)\end{aligned}$$

IV. Results

We now compare the proposed RDFA with the FPCA in [12], the ARV in [7] and the SVR in [9] on predicting the day ahead electricity prices. We consider two datasets, i.e. the New England dataset and the Australian dataset, and their results are shown in Section IV-A and Section IV-B, respectively.

A. New England Dataset

The electricity price we used is collected from New England day-ahead energy market in United States [18], which are hourly Locational Marginal Pricing (LMP) data at the Internal Hub from January 1, 2005 to September 3, 2010, as shown in Fig. 3. To investigate the average performance of various algorithms, we have performed 50 different simulations based on forward validation (FV). More precisely, the first simulation used the electricity price from January 1, 2005 to July 14, 2010 to form the training data segment, and the price on July 15, 2010 is used as the testing data segment to verify the accuracy. The next simulation will be performed similarly by shifting the data segment one day forward, and therefore the first day of the training data segment in the final simulation will be on February 18, 2005. Similar to other previous works in [7] and [27], the algorithms are invoked on the natural logarithm of the electricity price. In order to test the validity of the Gaussian assumption, we first obtain the stochastic component by subtracting the deterministic component from the log-transformed electricity price. We then employ a sliding window of 30 days long to compute the stochastic component in each calculation, since the proposed approach employs a recursive prediction model to deal with the possibly changing trend. Afterwards, we test the normality of these data blocks using the Royston's test [41], which is well-known for its high sensitivity on detecting departure from normality. In the Royston's test, if the obtained p-value is smaller than a certain significance level, e.g., 5%, then the data block considered is regarded to be non-Gaussian under the given significance level. We find that the average p-value obtained from these data blocks is 0.4001. Since these p-values are much greater than the commonly considered 5% significance level, it seems that the Gaussian assumption is deemed acceptable.

Next, the estimated electricity prices in natural logarithmic scale obtained from the algorithms are converted back to the original scale in order to assess the performance of various algorithms by means of the mean absolute percentage error (MAPE)

$$MAPE = \left(\frac{1}{K}\right)\sum_{k=1}^{K}\left|\frac{\big(x^{(k)}(n)-\hat{x}^{(k)}(n)\big)}{x^{(k)}(n)}\right| \quad (43)$$

which is calculated by averaging the prediction error obtained from those $K = 50$ simulations. In (43), $x^{(k)}(n)$ and $\hat{x}^{(k)}(n)$ denote the actual and estimated hourly electricity prices, respectively, at the kth FV iteration. The FV is usually used for parameter tuning and evaluation of accuracy for time series prediction [28], [29]. For details of FV, interested readers are referred to [28]. Here, for the proposed RDFA and FPCA, we remark that the single variable time series formed by the electricity hourly price is divided into 24 partitions/variables as in (5), where the time unit of each sample is one day. To apply (43), the predicted daily price vectors (24 variables) are converted back to the original format, i.e., single variable time series. Whereas in the ARV and SVR, (43) can be directly applied because the two algorithms are applied to the hourly price data in original format.

Furthermore, to evaluate the quality of the predicted intervals, we consider the following measures:

1) Calibration Bias [42]: It computes the deviation between the forecast nominal proportions $\hat{\alpha}$ and actual nominal proportions α (two-tailed) defined as

$$bias = \alpha - \hat{\alpha} \quad (44)$$

in which the forecast nominal $\hat{\alpha}$ can be calculated as

$$\hat{\alpha} = \frac{1}{KN}\sum_{k=1}^{K}\sum_{n=1}^{N}\delta\left(\hat{x}_{L,\alpha}^{(k)}(n) \le x^{(k)}(n) \le \hat{x}_{U,\alpha}^{(k)}(n)\right) \quad (45)$$

where N is the number of time instants of the day and hence $N = 24$ for hourly recorded electricity price. $\delta(.)$ is an indicator function returning one when the argument is true or zero otherwise, and $\hat{x}_{L,\alpha}^{(k)}(n)$ and $\hat{x}_{U,\alpha}^{(k)}(n)$ denotes, respectively, the lower and upper bound of the interval forecast at the given nominal proportion α.

2) Interval Score [43]: It favors narrow prediction interval and penalizes observations that do not lies within the predicted interval according to the nominal proportions, and is computed as follows:

$$S(\alpha) = I(\alpha) + P(\alpha) \quad (46)$$

where $I(\alpha) = (1/KN)\sum_{k=1}^{K}\sum_{n=1}^{N}(\hat{x}_{U,\alpha}^{(k)}(n) - \hat{x}_{L,\alpha}^{(k)}(n))$, $P(\alpha) = \frac{1}{KN}\sum_{k=1}^{K}\sum_{n=1}^{N}(2/(1-\alpha))\{[(\hat{x}_{L,\alpha}^{(k)}(n) - x^{(k)}(n))\,\delta(x^{(k)}(n) < \hat{x}_{L,\alpha}^{(k)}(n)] + [x^{(k)}(n) - \hat{x}_{U,\alpha}^{(k)}(n)]\delta(x^{(k)}(n) >$

$\hat{x}_{U,\alpha}^{(k)}(n))\}$. Note, a smaller interval score indicates better performance. Also, we can see that the multiplication factor $2/(1-\alpha)$ in the penalty term increases with the nominated proportion, and observations that exceed the predicted interval at higher nominal proportions receive higher penalties. To quantify the performance of the proposed approach, we shall consider the relative interval score (RIS) as follows:

$$RIS(\alpha) = \frac{S_{ref}(\alpha) - S(\alpha)}{S_{ref}(\alpha)} \times 100\% \tag{47}$$

where $S_{ref}(\alpha)$ is the RIS defined similarly as in (2) except that for a given α the lower and upper bounds of the prediction $\hat{x}_{L,\alpha}^{(k)}(n)$ and $\hat{x}_{U,\alpha}^{(k)}(n)$ are obtained from ensemble mean and standard deviation in a 30-day sliding window. Generally, the RIS with a larger positive value indicates better performance of the proposed approach.

1) Settings for the Proposed RDFA: According to [2], electricity prices during weekdays follow a similar pattern. Moreover, the electricity prices of the current day and the same day of the previous week, i.e., 7 days ago, generally exhibit similar pattern too because of similar demand. To capture the correlation of the previous 7 days, the AR model order is chosen as $L = 7$ for the RDFA. Analogously, the Fourier series base period is chosen as $K = 7$ days, which is the same as that in [7].[1] The state covariance and noise covariance for the KF are initialized as $\boldsymbol{Q}_m = 0.1\boldsymbol{I}_L$ and $R_m = 0.1$, respectively, where $\boldsymbol{I}_L$ is an $L \times L$ identity matrix. The number of chosen PCs is $B = 4$, which is determined by the MDL method in [21]. The forgetting factors β, β_{LS}, β_{μ}, β_{Q_m} and β_{R_m} are chosen as 0.995, which are commonly used in adaptive algorithms. An initial data block of length 10 days is used to initialize the RDFA algorithm. Since RDFA treats each hour of the day (or less depending on the time scale of recording) as a separate variable, the day-ahead price can be obtained in one prediction step. After the next day has elapsed and the actual price has been obtained, we update the model using the most recent actual price. This procedure is repeated for onward prediction such that the latest actual price is incorporated into the model on each day.

2) Settings for the FPCA: Similar to the proposed RDFA, the same AR model order $(L = 7)$ and the number of chosen PCs $(B = 4)$ are adopted except that batch processing is used in the FPCA. A sliding window of length 1000 days is used to provide sufficient data for batch processing.

3) Settings for ARV: For fair comparison, the AR model order, Fourier base period and the sliding window for batch processing are chosen so as to capture the trend for the same period of time. Different from the RDFA and FPCA, the ARV is directly applied to the single variable hourly electricity price, so that the time unit of each sample is one hour. Hence, the AR model order, Fourier base period and sliding window are chosen as $L = 7 \times 24 = 168$ hours, $K = 7 \times 24 = 168$ hours and $1000 \times 24 = 24\,000$ hours, respectively.

[1]In [7], the deterministic component of the two-factor ARV is modeled by a sum of two Fourier series with base periods of 7 days and 1 year, respectively. In this paper, we have considered only the former Fourier series (7 days) for modeling the deterministic component, which is sufficient for tracking short term trend (weekly) for day ahead forecast.

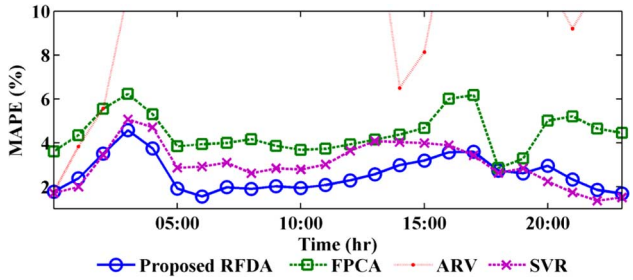


Fig. 4. MAPE of the day-ahead prediction obtained from New England dataset for different algorithms.

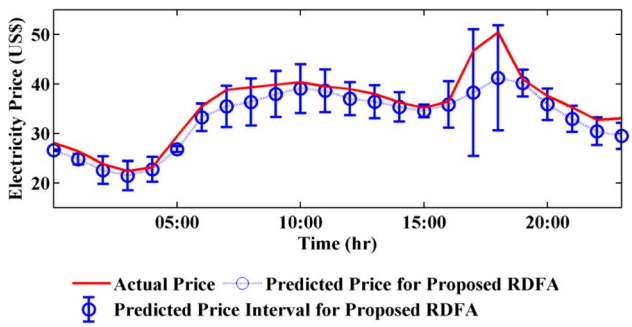


Fig. 5. Price and interval forecast obtained from New England dataset for the proposed RDFA algorithm.

4) Settings for the SVR: Settings similar to the ARV are adopted except that the Gaussian kernel is adopted for the nonlinear mapping and an additional FV is required for tuning the kernel width and other parameters, such as the epsilon and the cost parameter. This is different from the other 3 algorithms, where FV is used for evaluating the prediction accuracy only. Therefore, a double FV procedure is adopted in the SVR, where the external FV loops are used to evaluate the performance and the inner FV loops are used for parameter tuning. This avoids the optimistic bias of obtaining prediction accuracy that is higher than the actual one [30], which happen when the testing sets are re-used for both evaluation and parameter tuning.

5) Results: Fig. 4 shows the MAPE of the day ahead price prediction for the three algorithms. Lower MAPE generally implies higher prediction accuracy. Among all algorithms, we can see that the proposed RDFA algorithm outperforms other algorithms in the prediction accuracy, while it is able to reliably estimate the interval of the day ahead price as shown in Fig. 5, such that the actual price lies within the predicted interval for most of the time.

Tables III and IV show the average run-time per sample and arithmetic complexities of different algorithms. We compare the average run-time per sample of all algorithms to see whether the algorithms are able to complete the price prediction before the next sample is received. Otherwise, the price prediction of the succeeding samples will be delayed. We can see that the proposed RDFA has the lowest runtime per iteration, which is contributed by the low arithmetic complexity of the RLS, subspace tracking algorithm and the KF. On the other hand, the SVR has the longest run-time. This is mainly due to the parameter tuning

TABLE III
AVERAGE RUNTIME FOR DIFFERENT ALGORITHMS PER SAMPLE

Time (s)	RDFA	FPCA	ARV	SVR
Fourier	5.17×10^{-4}	N/A	0.23	N/A
Remaining	9.84×10^{-4}	2.70×10^{-3}	0.11	3.02×10^{3}
Total	1.50×10^{-3}	2.70×10^{-3}	0.34	3.02×10^{3}

TABLE IV
ARITHMETIC COMPLEXITY OF DIFFERENT ALGORITHMS

	Proposed RDFA	FPCA	ARV
Fourier Series	$O(JK^2)$	N/A	$O(J^3K^3)$
Subspace Computation	$4JB+O(B^3)$	$O(J^3)$	N/A
Time Series model	$O(L^{2.4})$	$O(L^3)$	$O(L^3)$

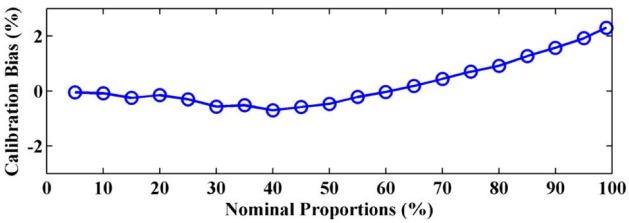


Fig. 6. Calibration bias obtained from the New England Dataset for the proposed RDFA algorithm.

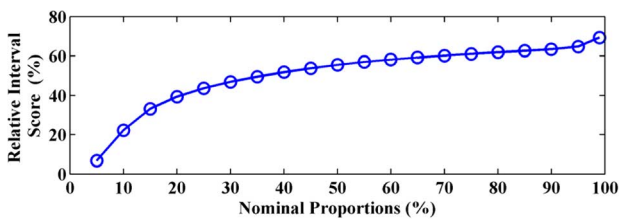


Fig. 7. Relative interval score obtained from the New England Dataset for the proposed RDFA algorithm.

process. Here, we have chosen 10 different values for each parameter, i.e., the kernel width, cost parameter and the epsilon, and SVR is re-invoked for the 1000 combinations to search for the pair of parameters that is able to achieve best average accuracy among different FV partitions. In general, the proposed RDFA algorithm outperforms other algorithms in runtime.

Figs. 6 and 7 show the calibration bias and interval score of the proposed approach respectively for the New England dataset. We find that the worse case calibration bias is around 2% and the proposed approach generally performs much better than the reference ensemble forecast in terms of the RIS.

B. Australian Dataset

We consider the half hourly electricity price of New South Wales (NSW), Australia in 2005 obtained from the Australian Electricity Market Operator (AEMO) [37]. It demonstrates strong and stochastic volatility or spikes, as shown in Fig. 8. In general, these prices spikes may be related to abnormal situations such as generation outage or a sudden surge in demand [38]. Such values are difficult to be predicted but the abnormal behavior may have substantial impact on subsequent prediction when the system returns to normal operation. Therefore, their effects should be suppressed or down-weighted.

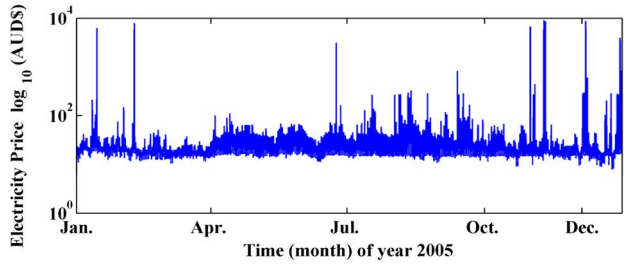


Fig. 8. Half hourly electricity price of the Australian (NSW) dataset from January 1, 2005 to December 31, 2005.

For comparison purpose, we consider the performance of the same set of algorithms. We employed the half-hourly price data from Jan.1 to December 31, 2005 for the evaluation. The same FV procedure employed in previous section (i.e., 50 simulations) is considered except that each daily sample contains a total of 48 prices because the electricity price is recorded half-hourly.

All the algorithms are assessed using the MAPE defined in (43). However, we note that in the calculation of MAPE, it is important to identify and exclude the daily samples in the testing data segment that contain the price spikes as they deviate significantly from the normal trend. If these daily samples are involved, the calculated MAPE will be dominated by these extremely large outlying values due to price spikes, and it will not be able to reflect the performance of the algorithms on predicting the normal trend, which is of our major interest. To this end, we employed the robust detection criteria of the proposed RRDFA algorithm on the whole dataset to systematically identify these daily samples and exclude them from the MAPE calculation. Therefore, the MAPE is modified as follows:

$$MAPE_{\Omega}=\left(\frac{1}{K_{\boldsymbol{\Omega}}}\right)\sum_{k=1}^{K_{\Omega}}\left|\frac{\left(x_{\Omega}^{(k)}(n)-\hat{x}_{\Omega}^{(k)}(n)\right)}{x_{\Omega}^{(k)}(n)}\right| \tag{48}$$

where $K_{\boldsymbol{\Omega}}$ denotes the set containing the FV iterations without price spikes, $K_{\boldsymbol{\Omega}}$ is the total number of FV iterations in the set $\boldsymbol{\Omega}$, and $x_{\Omega}^{(k)}(n)$ and $\hat{x}_{\Omega}^{(k)}(n)$ denote the kth actual and estimated half-hourly electricity prices, respectively, in the set $\boldsymbol{\Omega}$. Similar to (48), the calibration bias, interval score and RIS in (45)–(47) are also modified so that the general trend of the price can be properly assessed. To test the validity of the Gaussian assumption, we employ the same Royston's test as in the previous section except that an extra step is performed to either remove (approach 1) or down-weigh (approach 2) the price spikes using the robust detection criteria of the proposed RRDFA algorithm. The average p-values for the two approaches are found to be 0.2774 and 0.3175, respectively, which are much larger than the commonly considered 0.05 significance level. This suggests the Gaussian assumption is also deemed acceptable for this configuration.

1) Settings for Various Algorithms: In general, we find that the dataset is much more volatile than the New England dataset and the trend changes more quickly. In this regard, we have reduced the forgetting factors for the RDFA algorithm to 0.95 for β, β_{LS}, β_{μ}, and 0.9 for β_{Q_m} and β_{R_m}, respectively, to allow quicker adaption to the faster changing trend. The number of

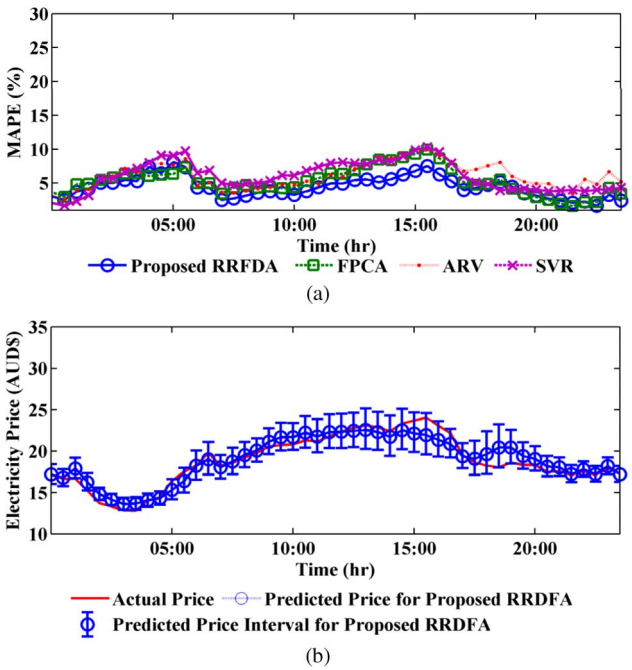


Fig. 9. (a) MAPE of the day ahead prediction obtained from Australian (NSW) dataset for the proposed RRDFA algorithm using approach 1. (b) Price and interval forecast obtained from the Australian (NSW) dataset for the proposed RRDFA algorithm using approach 1.

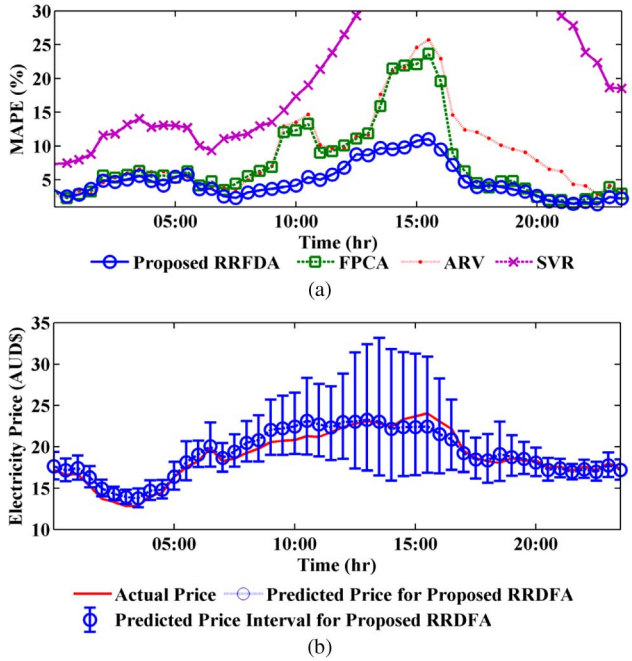


Fig. 10. (a) MAPE of the day ahead prediction obtained from Australian (NSW) dataset for the proposed RRDFA algorithm using approach 2. (b) Price and interval forecast for Australian (NSW) dataset for the proposed RRDFA algorithm using approach 2.

major PCs are chosen as $B = 9$, which is determined by the MDL method. Moreover, the state covariance and noise covariance initialization for the KF are increased to $\boldsymbol{Q}_m = \boldsymbol{I}_L$ and $R_m = 1$ as the dataset is more volatile. For the robust extension, the value of ξ in (29) is chosen as $\xi = 2.576$, which corresponds to the commonly employed 99% occurrence probability of normal sample [31], [32]. For the robust detection criteria, the forgetting factors in (29) and (30) are also chosen as 0.95 and the window length in (30) and (37) is chosen as $L_d = 20$.

For the FPCA, the number of chosen PCs is also chosen as $B = 9$. The length of the batch processing sliding window is chosen as 1 month (31 days), to allow quicker adaption to the faster changing trend. For the ARV and SVR algorithms, the same sliding window is chosen for fair comparison. For other parameters, the same values are adopted as in the previous section.

2) Results: Figs. 9(a) and 10(a) show the modified MAPE obtained from approaches 1 and 2 mentioned in Section III-D. We can see that other conventional algorithms generally exhibit much larger MAPE under the presence of price spikes. This suggests that the extreme values of price spikes generally lead to less accurate subsequent prediction of electricity price under normal operation. On the contrary, the proposed robust detection criteria are found to be very effective in improving the performance of the conventional algorithms in normal forecasting as shown in Fig. 9(a). Moreover, in view of the interval forecast as shown in Figs. 9(b) and 10(b), both approaches are generally able to reliably predict the price interval, such that the actual price lies within the predicted intervals for most of the time. In particular, we can see that approach 2 generally provides more accurate predicted prices but exhibits larger variation than

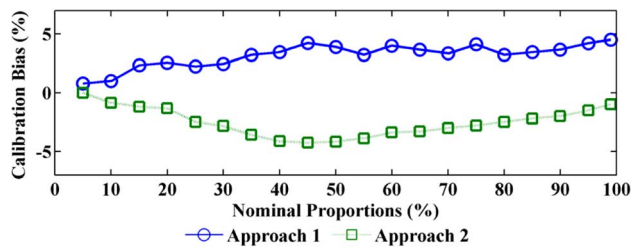


Fig. 11. Calibration bias obtained from the Australian (NSW) dataset for the proposed RRDFA algorithm.

approach 1, because in the latter, the whole daily sample will be completely removed once any one of 48 half hourly prices are identified as the spikes. Therefore, some other normal half hourly prices that capture the general trend may be sacrificed, which in turn affects the accuracy of the predicted price. On the other hand, since the effect of the price spikes is completely removed, the predicted price interval of approach 1 generally has smaller variation. Finally, we remark that both approaches have their own merits and advantages, and they can complement to each other in order to offer reliable price forecasting in the presence of price spikes.

Figs. 11 and 12 show the calibration bias and RIS of the proposed RRDFA algorithm, respectively. We can see that the worse case calibration biases are less than 5% for both approaches. Moreover, comparing with the ensemble forecast, the RRDFA algorithm has a better performance in terms of the RIS, as shown in Fig. 12. In summary, the calibration bias and interval score in all simulations are very satisfactory, which substantiates the usefulness of the proposed interval forecast.

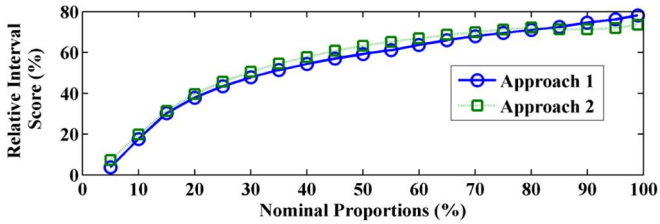


Fig. 12. Relative interval scores obtained from the Australian (NSW) dataset for the proposed RRDFA algorithm.

V. Conclusions

A new RDFA and multi-factor approach for electricity price and interval forecasting have been presented. Experimental results using the New England dataset [18] and the Australian dataset [37] shows that the proposed RDFA, which combines the OPASTr algorithm recursive ED algorithm and the KF, is able to achieve better day ahead forecast accuracy than other conventional algorithms. Moreover, a new robust extension is proposed to further improve its performance under possible non-Gaussian variation such as price spikes. Results show that the RDFA algorithm under the robust setting is able to consistently achieve better forecasting accuracy than other algorithms under the presence of price spikes. Finally, it can be extended to predict electricity price of longer term by incorporating a multi-factor model. The efficient recursive implementation, ability on performing interval forecast and good performance of the proposed RDFA algorithm make it as an attractive alternative to other conventional approaches to electricity price forecasting and other possible applications, such as demand load forecasting.

References

[1] R. Bjorgan, C. Liu, and J. Lawarrée, "Financial risk management in a competitive electricity market," *IEEE Trans. Power Syst.*, vol. 14, pp. 1285–1291, 1999.

[2] A. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 120–133, Sep. 2010.

[3] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Chichester, U.K.: Wiley, 2006.

[4] Y. Hou, C. Liu, and H. Salazar, "Electricity prices as stochastic processes," in *Advances in Electric Power and Energy*, ser. Power Engineering, M. El Hawary, Ed. Piscataway, NJ: IEEE Press/Wiley, to be published.

[5] F. Black and M. S. Scholes, "The pricing of options and corporate liabilities," *J. Polit. Econ.*, vol. 81, pp. 637–654, 1973.

[6] A. Neumaier and T. Schneider, "Multivariate autoregressive and Ornstein-Uhlenbeck processes: Estimates for order, parameters, spectral information, and confidence regions," *ACM Trans. Math. Softw.*, submitted for publication.

[7] S. Heydari and A. S. Siddiqui, "Valuing a gas-fired power plant: A comparison of ordinary linear models, regime switching approaches, and models with stochastic volatility," *Energy Econ.*, vol. 32, no. 3, pp. 709–725, 2010.

[8] C. Hsu, C. Chang, and C. Lin, A Practical Guide to Support Vector Classification, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003, tech. rep.

[9] U. Thissen, R. van Brakel, A. P. de Weijer, W. Melssen, and L. Buydens, "Using support vector machines for time series prediction," *Chemometr. Intell. Lab. Syst.*, vol. 69, no. 1–2, pp. 35–49, Nov. 2003.

[10] C. Gao, E. Bomparb, R. Napoli, and H. Zheng, "Price forecast in the competitive electricity market by support vector machine," *Physica A*, vol. 382, pp. 98–113, Aug. 2007.

[11] R. Huisman and R. Mahieu, "Regime jumps in electricity prices," *Energy Econ.*, vol. 25, pp. 425–434, 2003.

[12] H. L. Shang, Nonparametric Modeling and Forecasting Electricity Demand: An Empirical Study, Department of Econometrics and Business Statistics, Monash University, Monash Econometrics and Business Statistics Working Papers, No. 19/10, 2010.

[13] S. C. Chan, H. C. Wu, and K. M. Tsui, "Robust recursive eigen-decomposition and subspace-based algorithms with application to fault detection in wireless sensor networks," *IEEE Trans. Instrum. Meas.*, to be published.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[15] H. C. Wu, S. C. Chan, K. M. Tsui, and Y. Hou, Supplementary Material: A New Recursive Dynamic Factor Analysis for Point and Interval Forecast of Electricity Price, Mar. 2012, tech. rep. [Online]. Available: http://www.eee.hku.hk/h0895838/pess_supplementary.pdf.

[16] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[17] K. Abed-Meraim, A. Chkeif, Y. Hua, and T. Paris, "Fast orthonormal PAST algorithm," *IEEE Signal Process. Lett.*, vol. 7, no. 3, pp. 60–62, Mar. 2000.

[18] ISO New England Inc., 2011. [Online]. Available: http://www.iso-ne.com/markets/hstdata/znlinfo/hourly/index.html.

[19] K. Maribu, A. Galli, and M. Armstrong, "Valuation of spark-spread options with mean reversion and stochastic volatility," *Int. J. Electron. Bus. Manage.*, vol. 5, no. 3, pp. 173–181, 2007.

[20] S. C. Chan, Y. Zhou, and W. Y. Lau, "Approximate QR-based algorithms for recursive nonlinear least squares estimation," in *Proc. IEEE ISCAS*, 2005, pp. 4333–4336.

[21] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, Oct. 1998.

[22] J. Bunch, C. Nielsen, and D. Sorensen, "Rank-one modification of the symmetric eigenproblem," *Numer. Math.*, vol. 31, no. 1, pp. 31–48, 1978.

[23] R. J. Harris, *A Primer of Multivariate Statistics*. Mahwah, NJ: Lawrence Erlbum, 2001, 07430.

[24] T. Winter, S. Mallory, T. Allen, and D. Rosenberry, "The use of principal component analysis for interpreting ground water hydrographs," *Ground Water*, vol. 38, no. 2, pp. 234–246, Mar. 2000.

[25] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*, ser. Springer Series in Statistics, 2nd ed. New York: Springer, Apr. 2001.

[26] B. Liao, Z. G. Zhang, and S. C. Chan, "A new robust kalman filter-based subspace tracking algorithm in an impulsive noise environment," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 9, pp. 740–744, Sep. 2010.

[27] J. Janczura and R. Weron, "Modeling electricity spot prices: Regime switching models with price-capped spike distributions," in *Proc. IEEE MEPS*, , pp. 1–4.

[28] Y. Bengio and N. Chapados, "Extensions to metric-based model selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1209–1227, Mar. 2003.

[29] U. Hjorth, "Model selection and forward validation," *Scandinavian J. Statist.*, vol. 9, no. 2, pp. 95–105, 1982.

[30] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformat.*, vol. 7, no. 1, p. 91, Feb. 2006.

[31] R. Huisman, C. Huurman, and R. Mahieu, "Hourly electricity price in day-ahead markets," *Energy Econ.*, vol. 29, pp. 240–248, 2007.

[32] H. Miyaguchi, T. Ito, and T. Misawa, "Regression analysis of electric power price using meteorological informations," in *Proc. IEEE PowerCon. 2006*, Oct. 2006, pp. 1–6.

[33] Y. Zou, S. C. Chan, and T. S. Ng, "A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.

[34] J. Jackson, *A User's Guide to Principal Components*. New York: Wiley Interscience, 1991.

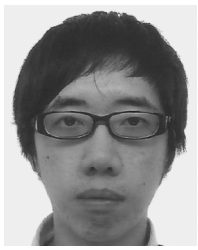
[35] D. X. Tien, K. W. Lim, and L. Jun, "Comparative study of PCA approaches in process monitoring and fault detection," in *Proc. IECON 2004*, May 2044, vol. 3, pp. 2594–2599.

[36] P. F. Odgaard, L. Bao, and S. B. Jorgensen, "Observer and data-driven-model-based fault detection in power plant coal mills," *IEEE Trans. Energy Convers.*, vol. 23, no. 2, pp. 659–668, Jun. 2008.

[37] Australian Energy Market Operator, 2012. [Online]. Available: http://www.aemo.com.au/.

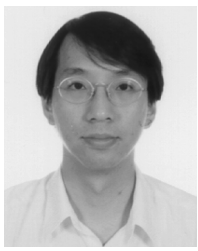
[38] C. Cuevas-Cubria, A. Schultz, R. Petchey, F. Beaini, and R. New, Energy in Australia 2011, Department of Resources, Energy and Tourism, Australian Government, Australia, 2011.

[39] S. Deng and S. Oren, "Electricity derivatives and risk management," *Energy*, vol. 31, pp. 940–953, 2006.
[40] P. Spalding, Review Into the Role of Hedging Contracts in the Existing NEM Prudential Framework, Australian Energy Market Commission, Sydney, NSW, Australia, Jun. 2010, tech. rep.
[41] J. P. Royston, "Some techniques for assessing multivarate normality based on the Shapiro-Wilk W," *J. Roy. Statist. Soc. C (Appl. Statist.)*, vol. 32, no. 2, pp. 121–133, 1983.
[42] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Statist. Assoc.*, vol. 102, no. 477, pp. 359–378, 2007.
[43] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, vol. 69, no. 2, pp. 243–268, 2007.



H. C. Wu (S'10) received the B.Eng. degree in electric engineering from the University of New South Wales, Australia, and the M.Sc. degree in electrical and electronic engineering from the University of Hong Kong, Pokfulam, in 2008 and 2009, respectively. He is currently pursuing the Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong.

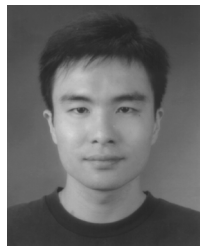
His main research interests are in power systems economics, pattern recognition, and bioinformatics.



S. C. Chan (S'87–M'92) received the B.Sc. (Eng) and Ph.D. degrees from The University of Hong Kong, Pokfulam, in 1986 and 1992, respectively.

Since 1994, he has been with the Department of Electrical and Electronic Engineering, the University of Hong Kong, where he is currently a Professor. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed A/D converter architecture, bioinformatics, smart grid and image-based rendering.

Prof. Chan is currently a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society, and an associate editor of *Journal of Signal Processing Systems* and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He was the chair of the IEEE Hong Kong Chapter of Signal Processing in 2000–2002, an organizing committee member of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, the 2010 International Conference on Image Processing, and an associate editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2008 to 2009.



K. M. Tsui received the B.Eng., M.Phil., and Ph.D. degrees in electrical and electronic engineering from The University of Hong Kong, Pokfulam, in 2001, 2004, and 2008, respectively.

He is currently working as Post-doctoral Fellow in the Department of Electrical and Electronic Engineering at the same university. His main research interests are in smart grid, high-speed A/D converter architecture, biomedical signal processing, array signal processing, digital signal processing, multirate filter bank and wavelet design, and digital filter design, realization and application.

Yunhe Hou (M'06) received the B.E. and Ph.D. degrees from Huazhong University of Science and Technology, China, in 1999 and 2005, respectively.

He worked as a Postdoctoral Research Fellow at Tsinghua University, Beijing, China, from September 2005 to August 2007. In 2008, he worked as a Postdoctoral Researcher at Iowa State University, Ames, and then as a Research Scientist at University College Dublin, National University of Ireland, Dublin, in 2009. In 2010, he was a Visiting Scientist with the Laboratory of Information and Decision Systems (LIDS) of Massachusetts Institute of Technology (MIT), Cambridge. He has been a Research Assistant Professor (since 2009) and an Assistant Professor (since 2011) with the Department of Electrical and Electronic Engineering, the University of Hong Kong.