Integrating KPCA and Locally Weighted Support Vector Regression for Short-Term Load Forecasting

E. E. Elattar *, J. Y. Goulernas * and Q. H. Wu *

* The Department of Electrical Engineering and Electronics, The University of Liverpool
Brownlow Hill, Liverpool, L69 3GJ, U.K.
Email: {elattar, j.y.goulernas, q.h.wu}@liverpool.ac.uk

** The Department of Electrical Engineering, Minufya University
Shebin El-Kom, Egypt

Abstract—This paper proposes a new approach to solve the short term load forecasting problem that considers electricity price as one of the main characteristics of the system load. The proposed method is derived by integrating the kernel principal component analysis (KPCA) method with locally weighted support vector regression (LWSVR). LWSVR can be derived by modifying the risk function of the support vector regression algorithm with use of locally weighted regression while keeping the regularization term in its original form. In the proposed model, the first stage is using KPCA to extract features and obtain kernel principal components which used to construct the phase space of the multivariate time series of inputs. LWSVR is employed in the second stage to solve the load forecasting problem. In addition, to optimize the weighting function’s bandwidth, the weighted distance algorithm is presented. The performance of the proposed model is evaluated with the historical load, price as one of the main characteristics of the system load. The results show that the proposed method provides a relatively better forecasting performance in comparison with other published models employing the same data.

I. INTRODUCTION

Short load forecasting (STLF) has always been a very important issue in economic and reliable power systems operation such as unit commitment, reducing spinning reserve, maintenance scheduling, etc. STLF is a difficult work because the accuracy of forecasting is influenced by many unpredicted factors such as economic, temperature, etc. whose relationships are commonly complex, implicit and nonlinear.

During the last few decades, numerous investigations have been proposed to improve the accuracy of STLF. These can be classified as either traditional, or artificial intelligence (AI) based techniques. The former include linear or multiple regression [1], autoregressive moving average exogenous variable (ARMAX) [2], Kalman filtering [3], etc. These methods are based on a linear regression model and can not always represent the nonlinear characteristics of complex loads [4]. On the other hand, various artificial intelligence techniques were used for STLF, amongst these methods artificial neural networks (ANNs) have received the largest share of attention. The ANNs that have been successfully used for STLF are based on multi-layered perceptrons [5]. The neuro-fuzzy model is also used to solve STLF with success [6]. Radial basis functions (RBFs) [7], have been also used for day ahead load forecasting and giving better results than that of a back-propagation neural network.

Recently, support vector regression (SVR) [8] which proposed based on statistical learning theory (SLT), has been investigated as a promising approach to power load forecasting [9]. Its merits chiefly come from the adoption of structural risk minimization principal instead of empirical risk minimization principal, and therefore it can obtain global optimal solution by solving a quadratic problem.

With further scrutiny, one can notice the complexity of the historical load data and the uncertainty of the influencing factors such as weather, economical, and random factors. This encourages us to apply the time series reconstruction technique to the power load forecasting. The traditional time series reconstruction techniques such as the coordinate delay (CD) method have a serious problem. This problem is that there may be correlation between different features in reconstructed phase space. This will influence the quality of phase space reconstruction and modeling effect [10].

To overcome this problem, the kernel principal component analysis (KPCA) which is one type of nonlinear principal component analysis (PCA) is used recently to process the nonlinear time series. The main idea of KPCA is first to map the original inputs into a high-dimensional feature space via a kernel map, which makes data structure more linear, and then to calculate principal components in the high-dimensional feature space [11].

Our previous work [12] has shown that local prediction methods based on phase reconstruction using CD method can provide generally better results than those obtained with global methods based on phase reconstruction. The local predictor involves more than one model to utilize the local information. Therefore, the accuracy of the local predictor is better than the global predictor in which only one model is engaged for all available data that may contain irrelevant patterns to the current prediction point. Another advantage of local predictor is that the training set for each point on the reconstructed trajectory is much smaller than the global predictor which requires the use of all available training examples.

Locally weighted regression (LWR) [13] is a kind of locally weighted learning method. LWR forms a local model around
a point of interest whereby only training data that is closest to that point will be used in handling each query, instead of using all training data [14]. LWR is a method for estimating a regression surface through multivariate smoothing: the response variable is smoothed dynamically, as a function of the predictor variables [13].

In SVR, the regularization parameter is constant so that all training data contribute to the accuracy of the model to the same extent. However, in many cases, the effects of the training points are different where some training points are more important than others. In this paper, we solve the STLF problem by combining KPCA method with locally weighted support vector regression (LWSVR). LWSVR is an ameliorated SVR, which endow a weight factor to each train load datum. KPCA is adopted to reconstruct phase space of nonlinear time series, on the basis of which some kernel principal components are chosen discontinuously according to their correlative degree with model output to form final phase space of nonlinear time series. The reconstructed phase space is then used as the input of LWSVR to solve the load forecasting problem. Moreover, the weighting function’s distance is proposed to optimize this bandwidth. The proposed weighted distance algorithm which uses the Mahalanobis distance is optimized to form the input of LWSVR to solve the load forecasting problem. Finally, the weighting function’s distance algorithm which uses the Mahalanobis distance is proposed to optimize this bandwidth. The proposed weighted distance algorithm which uses the Mahalanobis distance is proposed to optimize this bandwidth. The proposed weighted distance algorithm which uses the Mahalanobis distance is proposed to optimize this bandwidth.

The paper is organized as follows: Section II outlines the time series reconstruction method. Section III describes the LWSVR algorithm. Experimental results and comparisons on load forecasting problem are presented in Section IV. Finally, Section V concludes the work.

II. PHASE SPACE RECONSTRUCTION OF TIME SERIES BASED ON KPCA

KPCA is an supervised technique that is based on performing the PCA in the feature space of a kernel. In KPCA the computations are performed in a feature space that is nonlinearly related to the input space [15].

Due to the nonlinear relationship between the input space and feature space the KPCA is nonlinear. The basic idea of KPCA is to map the data x into a high dimensional feature space \( \Phi(x) \) via a nonlinear mapping, and perform the linear PCA in that feature space as:

\[
K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)
\]

where \( x_i, x_j \) are variables in input space and \( K(x_i, x_j) \) is called kernel function.

Suppose there is a set of data \( X = \{ x_i \}_{i=1}^N \) where each \( x_i \in \mathbb{R}^n \) and the mean value \( E[X] = 0 \). By mapping \( x_i \) into \( \Phi(x_i) \), KPCA solves the eigenvalue (2) [11]:

\[
\lambda_i v_i = \tilde{C} v_i, \quad i = 1, 2, ..., N
\]

where \( \tilde{C} = \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \Phi(x_i)^T \) is the sample covariance matrix of \( \Phi(x_i) \), \( \lambda_i \) is one of the non-zero eigenvalues of \( \tilde{C} \) and \( v_i \) is the corresponding eigenvector.

Because the eigenvectors \( v_i \) in the plane which is composed of \( \Phi(x_1), \Phi(x_2), ..., \Phi(x_N) \). Therefore:

\[
\lambda_i \Phi(x_i) \cdot v_i = \Phi(x_i) \cdot \tilde{C} v_i, \quad i = 1, 2, ..., N
\]

and there exist coefficients \( \alpha_i \) (\( i = 1, 2, ..., N \)) meet:

\[
v_i = \sum_{j=1}^N \alpha_i(j) \Phi(x_j)
\]

where \( \alpha_i(j) \) are the components of \( \alpha_i \). Combining (3) with (4) and defining an \( N \times N \) matrix \( K \) which is defined by (1), the following formula can be got [11]

\[
N \lambda_i \alpha_i = K \alpha_i, \quad i = 1, 2, ..., N
\]

Assuming the eigenvectors of \( \Phi(x_i) \) is of unit length \( v_i \cdot v_i = 1 \), each \( \alpha_i \) must be normalized using the corresponding eigenvalue by \( \tilde{\alpha}_i = \frac{\alpha_i}{\sqrt{\lambda_i}}, \quad i = 1, 2, ..., N \).

Finally the principal component for \( x_i \), based on \( \tilde{\alpha}_i \), can be calculated as following:

\[
p_t(i) = v_i^T \Phi(x_i) = \sum_{j=1}^N \tilde{\alpha}_i(j) K(x_j, x_i), \quad i = 1, 2, ..., N
\]

(6)

From (6), one can notice that the maximal number of principal components that can be extracted by KPCA is \( N \). The dimension of \( p_t \) can be reduced in KPCA by considering the first several eigenvectors that is sorted in descending order of the eigenvalues.

In this paper, we employ the commonly used Gaussian kernel defined as:

\[
K(x_i, x) = \exp \left( -\frac{\|x_i - x\|^2}{2\sigma^2} \right)
\]

(7)

III. LOCALLY WEIGHTED SUPPORT VECTOR REGRESSION (LWSVR)

A. Support Vector Regression (SVR)

The SVR is a linear regression in the feature space. This space is introduced by a mapping function which maps the input space to the feature space, i.e. \( \phi : \mathbb{R}^n \rightarrow \mathbb{R}^v \). In general \( v > n \) so that the nonlinear function under approximation in the input space becomes a linear function in the future space. Given input output data \( \{ x_i, y_i \}_{i=1}^N \) where each \( x_i \in \mathbb{R}^n \) denotes the input space of the sample and has a corresponding target value \( y_i \in \mathbb{R} \) for \( i = 1, ..., N \), where \( N \) corresponds to the size of the training data. The estimating function takes the form as follows [8]:

\[
f(x) = w \cdot \phi(x) + b
\]

(8)

where \( \phi(x) \) denotes the high dimensional feature space which is nonlinearly mapped from the input space, \( w \) contains the coefficients that have to be estimated from the data and \( b \) is a real constant. The objective is to minimize the following risk function [16]:

\[
\min_{w, b, \xi, \xi^*} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi^*_i) \right\}
\]

(9)
subject to \[
\begin{align*}
\text{subject to } \quad & y_i - w \cdot \phi(x_i) + b \leq \varepsilon + \xi_i^* \quad \text{if } \|w\|^2 \leq \varepsilon \\
& y_i - w \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad \text{if } \|w\|^2 > \varepsilon \\
& \xi_i, \xi_i^* \geq 0
\end{align*}
\]

where \(x_i\) is mapped to a higher dimensional space by the function \(\phi\), the term \(\frac{1}{2}\|w\|^2\) is called the regularization term, \(C\) is the regularization constant which determines the trade-off between the flatness of \(f\) and its accuracy in capturing the training data and \(\xi_i\) is the lower training error (\(\xi_i^*\) is the upper) subject to the \(\varepsilon\)-insensitive loss function which is described as [16]:

\[
|\xi| = \begin{cases} 
0 & \text{if } |\xi| \leq \varepsilon \\
|\xi| - \varepsilon & \text{otherwise}
\end{cases} \quad (10)
\]

Introducing Lagrange multipliers \(\alpha_i\) and \(\alpha_i^*\) with \(\alpha_i, \alpha_i^* > 0\) and \(\alpha_i, \alpha_i^* \geq 0\) for \(i = 1, \ldots, N\) and according to the Karush-Kuhn-Tucker optimality conditions [8], the dual variables Lagrangian can be obtained as:

Maximize\[
L_d(\alpha, \alpha^*) = \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j)
\]

(11)

with the constraints

\[
\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \quad \text{for} \quad 0 \leq \alpha_i, \alpha_i^* \leq C \quad (12)
\]

where \(K(x_i, x_j)\) is the kernel function that is the inner product of the points \(\phi(x_i)\) and \(\phi(x_j)\) mapped into feature space. Using the kernels, all necessary computations can be undertaken directly in the input space, without calculating the explicit map \(\phi(x)\). The typical examples of kernel function are linear, polynomial, Gaussian, etc. [8]. In this paper, we employ the commonly used Gaussian kernel as defined in (7).

Finally, the regression output takes the following form:

\[
\hat{f}(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)K(x, x_i) + b \quad (13)
\]

B. Locally Weighted Regression (LWR)

Locally Weighted Regression is derived from standard linear regression. This algorithm fits a surface to "local" points using distance-weighted regression. LWR is based on the (assumption) that the neighboring values of the predictor variables are the best indicators of the response variable in that range of predictor values [13].

To estimate the value of the function \(\hat{f}(x)\) at any value of \(x\) in the \(m\)-dimensional space, the \(k\) (neighborhood size) data points whose \(x_i\) values are closest to \(x\) are used (\(1 < k \ll N\)). Each point in the neighborhood is weighted according to its distance from \(x\). The points that are close to \(x\) have large weights, and the points far from \(x\) have small weights [17].

Many weighting functions are proposed by the researchers [13]. Out of these weighting functions, Gaussian kernel, tricube kernel and quadratic kernel are the most popular [14]. The widely used weighting function is the Gaussian kernel weighting function [18] which can be defined as following:

\[
W(d_E) = e^{-\frac{d_E^2}{h^2}} \quad (14)
\]

where \(h\) is the smoothing parameter which plays an important role in local modeling. From (14), the weight of the data point \((x_i, y_i)\) then is:

\[
W_i = W(\sqrt{(x - x_i)^T(x - x_i)}) \quad (15)
\]

Thus \(W_i\) has its maximum value when \(x_i\) is closest to \(x\), and decreases as \(x_i\) increases in distance from \(x\).

C. Locally Weighted Support Vector Regression (LWSVR)

In SVR algorithm, the regularization constant \(C\) in (9) which determines the trade-off between the flatness of \(f\) and its accuracy in capturing the training data is chosen as a constant value, so all training data contribute to the accuracy of the model to the same extent [17]. In the load forecasting problem, it is common that some training points are more important than others [12]. Therefore, the model should have higher accuracy for the training input data that are closer to the new input point for prediction.

To achieve this goal, the LWSVR is proposed by modifying the risk function of the standard SVR with the use of LWR. In the proposed method, \(C\) is computed as a function of the distance between input data points, and the concept of LWR is used. Therefore, the modified risk function can be formulated as follows:

\[
\frac{1}{2} \|w\|^2 + \sum_{i=1}^{N} C_i (\xi_i + \xi_i^*) \quad (16)
\]

and

\[
C_i = W_i \times C \quad (17)
\]

where \(W_i\) is the weight function obtained from (15). Replacing the constant \(C\) in (12) using (17), the dual problem’s constraints can be written as:

\[
\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \quad \text{for} \quad 0 \leq \alpha_i, \alpha_i^* \leq C_i \quad (18)
\]

By solving this problem, the regression output can be obtained using (13).

A bandwidth parameter \(h\) in (14) defines the scale or range over which generalization is performed. This is a very important parameter which plays an important role in local modelling. If \(h\) is infinite then the local modelling becomes global. On the other hand, if \(h\) is too small, then it is possible that we will not have adequate number of data points in the neighborhood for a good prediction.

Using this parameter as a fixed value will lead to unsatisfactory performance for non-linear systems as the density and distribution of data points are unlikely to be identical at every place of the data set [19]. Therefore the weighted distance algorithm which uses the Mahalanobis distance for
optimizing the bandwidth \( (h) \) is proposed in this paper. Using this measure, the problem of scale and correlation inherent in Euclidean distance are no longer an issue.

Defining the Mahalanobis distance between the query point \( x_q \) and data point \( x \) as \( d_q = \sqrt{(x - x_q)^T S^{-1}(x - x_q)} \) where \( x \) belongs to the \( k \) nearest neighbors of the query point \( x_q \) and \( S^{-1} \) is inverse covariance matrix which computed after removing the mean from each column, the bandwidth \( h_q \) is the function of \( d_q \):

\[
h_q = \Theta(d_q) \tag{19}
\]

where \( d_{min} \leq d_q \leq d_{max} \) and \( d_{min} \) is the distance between \( x_q \) and the closest neighbor while \( d_{max} \) is the distance between \( x_q \) and the farthest neighbor.

According to the LWR method, the query point corresponding to \( d_q = d_{min} \) is most important that is \( h_{max} = \Theta(d_{min}) = 1 \) while the query point corresponding to \( d_q = d_{max} \) is the least important, that is \( h_{max} = \Theta(d_{min}) = \delta \). \( \delta \) is a real constant. This constant is a low sensitivity parameter. Therefore, after few trials, we fix it to 0.01 which gives the best results.

The bandwidth \( h_q \) can be selected as a function of \( d_q \) as following [19]:

\[
h_q = \Theta(d_q) = a \left(1 - \frac{bd_q}{d_q}\right)^2 + c \tag{20}
\]

where \( a, b \) and \( c \) are constants. By applying the boundary conditions, we can calculate these constants and get [19]:

\[
h_q = \Theta(d_q) = (1 - \delta) \left(\frac{d_{min}(d_{max} - d_q)}{d_q(d_{max} - d_{min})}\right)^2 + \delta \tag{21}
\]

Using (21), the Gaussian kernel weighting function of the query point \( x_q \) can be written as following:

\[
W(d_q) = e^{-\left(\frac{\left(1 - \delta\right)^2}{\left(\frac{d_{min}(d_{max} - d_q)}{d_q(d_{max} - d_{min})}\right)^2 + 1}\right)} \tag{22}
\]

IV. EXPERIMENTAL RESULTS

The performance of the proposed method has been evaluated using publicly available load, price and temperature data (for two years 2002 and 2003) of the Victorian electricity market, Australia. Historical loads and prices in half hourly basis was collected from National Electricity Market Management Company Limited (NEMMCO) [20], while historical weather data (temperature) was collected from the Bureau of Meteorology, VIC Climate and Consultancy Section. These data are transformed to an hourly basis by averaging two half hours.

To show the effectiveness of our proposed method, numerical simulations comparing with local SVR [12] and LWSVR based CD method are conducted. For the CD method parameters, the correlation dimension method and the mutual information method are used to selecting the embedding dimension \( (m) \) and the time delay constant \( (\tau) \), respectively. The details of how to choose the proper values of \( m \) and \( \tau \) using these two methods have been reported in [21], [12]. The optimal values of these parameters are shown in Table I.

In addition there are two important parameters in the KPCA algorithm, which are the number of principal components \( (n_c) \) and \( \sigma^2 \) in the Gaussian kernel function and the optimal values of these parameters which computed using the cross validation method are shown in Table I.

Also, choosing the neighborhood size \( (k) \) is important step in order to establish the local prediction model. Here, the value of \( k \), which gives the best performance, has been set experimentally.

There are some key parameters for SVR, which are \( C, \epsilon \) and \( \sigma \) in the Gaussian kernel function. The selection of these parameters is important to the accuracy of the forecasting. However, structural methods for confirming the selection of these parameters efficiently are lacking. Therefore, in order to get these parameters, we divided the training data into two subsets. One of them is used to train the model while the other, called the validation set, is used for select the model. According to the performance on the validation set, we infer the proper values of these parameters.

For all performed experiments we quantified the prediction performance with mean absolute percentage error (MAPE) defined as:

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|A(i) - F(i)|}{A(i)} \times 100 \tag{23}
\]

where \( A \) and \( F \) are the actual and the forecasted loads, respectively, \( N \) is the testing dataset size, and \( i \) denotes the test instance index.

We designed two cases to evaluate the performance of the proposed model.

A. Several Hours Ahead Load Forecasting

In this case, the performance of the proposed method (LWSVR based KPCA) for several hour ahead load forecasting has been tested using the hourly load, temperature and price data of Victoria dataset to forecast the hourly loads from 1 up to 6 hours ahead for several weeks in 2003. To verify the predictive ability of the proposed method, its performance has been compared with local SVR [12] method, LWSVR based CD method and two published methods. These methods are:

- Model A [22]: ANN based similar days approch.
- Model B [23]: Extended Bayesian training method.

First, we calculate the MAPE at each lead time (from 1 up to 6 hours ahead) for each method during the period from Monday, 1st September 2003 to Sunday 7th September 2003. These results are shown in Fig 1.

Fig. 1 shows that our proposed method gives the best performance amongst all methods. Moreover, as it is expected we can notice that the error is increasing proportionally with the lead time. This happens because uncertainty makes it more difficult to predict more distant events.

To further study the adaptiveness of the proposed method, we also performed simulations for the whole September month (from 1st to 30th September). The MAPE value for one hour and six hours ahead is obtained as 0.57 and 1.32, respectively.
Table I: Phase Reconstruction Parameters

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Coordination delay method parameters</th>
<th>KPCA parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Load time series</td>
<td>Temperature time series</td>
</tr>
<tr>
<td>Victoria dataset</td>
<td>$m_1$</td>
<td>$\tau_1$</td>
</tr>
</tbody>
</table>

Table II: Comparison Among Models (MAPE)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01-07/09</td>
<td>0.56</td>
<td>0.49</td>
<td>0.87</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>2</td>
<td>01-07/09</td>
<td>0.53</td>
<td>0.49</td>
<td>0.87</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>3</td>
<td>01-07/09</td>
<td>1.15</td>
<td>0.91</td>
<td>0.86</td>
<td>0.75</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>01-07/09</td>
<td>1.00</td>
<td>0.82</td>
<td>0.77</td>
<td>0.72</td>
<td>0.73</td>
</tr>
<tr>
<td>5</td>
<td>01-07/09</td>
<td>1.20</td>
<td>0.99</td>
<td>0.81</td>
<td>0.82</td>
<td>0.77</td>
</tr>
<tr>
<td>6</td>
<td>01-07/09</td>
<td>1.30</td>
<td>1.07</td>
<td>0.98</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>1</td>
<td>01-30/09</td>
<td>0.77</td>
<td>–</td>
<td>0.68</td>
<td>0.64</td>
<td>0.57</td>
</tr>
<tr>
<td>6</td>
<td>01-30/09</td>
<td>2.06</td>
<td>–</td>
<td>1.63</td>
<td>1.48</td>
<td>1.30</td>
</tr>
</tbody>
</table>

Fig. 1. MAPE plotted against lead time for each method during the period from Monday, 1st September 2003 to Sunday 7th September 2003.

Table II shows the MAPE value of each method for all the simulated cases. These results show that our proposed method outperforms the LWSVR based CD method, local SVR method and other published models. For the first week of September 2003, our proposed method improves the accuracy over LWSVR based CD method, local SVR, Model A [22] and Model B [23] by 11.11%, 14.89%, 28.57% and 18.36%, respectively for one hour ahead forecasting, while these improvements are 12.22%, 19.39%, 39.23% and 26.17%, respectively for six hours ahead forecasting. In addition, for September 2003, our proposed method improves the accuracy over LWSVR based CD method, local SVR and Model A [22] by 10.93%, 16.18% and 25.97%, respectively for one hour ahead forecasting, while these improvements are 12.16%, 20.25% and 36.89%, respectively for six hours ahead forecasting.

B. One Day Ahead Load Forecasting

In this case the hourly load for one day ahead (from 1 up to 24 hour ahead) is predicted during two test months. These months are July 2003 (Winter season in Australia) and December 2003 (Summer season in Australia). To verify the predictive ability of the proposed method, its performance has been compared with local SVR method and LWSVR based CD method. We calculate the MAPE of each day during each testing period. Then the average MAPE value of each day of the week (Monday to Sunday) during each testing period can be calculated. These results are shown in Fig. 2. Table III also summarizes the overall mean performance of each method for each testing period.
It can be seen from these results that the proposed method gives better performance than LWSVR based CD model and local SVR in all days. Moreover, the results show the superiority of the proposed method over other methods. The percentage of accuracy improvement over LWSVR based CD is ranged from 11.46% to 14.39% for the three testing months, while the percentage of accuracy improvement over local SVR is ranged from 20.93% to 25.54% for the three testing months.

Another observation from these results is that the MAPE of December is higher than July. This happens because December is unstable month in respect to load behavior. This is due to the increase in power consumption because of a gradual rise in temperature (the maximum and minimum temperatures were recorded to be nearly 40 and 12°C, respectively) and the celebration activities (Christmas and New Year).

V. CONCLUSIONS

In this paper, we have presented a locally weighted SVR based KPCA to solve STLFP problem in price environment. In the proposed model, the phase space is reconstructed based on multivariate time series using KPCA method for each scalar time series. The Euclidian distance is then used to find the neighboring points of the point under prediction. The points that are close to the point under prediction have large weights, and the points far from the point under prediction have small weights. Then the new regularization constant of SVR is calculated. Moreover, the weighting function's bandwidth is optimized using the weighted distance algorithm. According to these neighboring points and the new regularization constant, the LWSVR model is set up.

To verify the predictive ability of the proposed method, we used the historical load, temperature and price data from the Victorian electricity market in Australia and performed simulation of two cases. The proposed method has been compared with the local SVR, locally weighted SVR based CD method and some published papers employing the same datasets. The numerical results show the superiority of the proposed method over all other methods.

REFERENCES


