2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Prediction Interval Estimation for Electricity Price and Demand using Support Vector Machines

Nitin Anand Shrivastava, Abbas Khosravi, Member, IEEE and Bijaya Ketan Panigrahi, Senior Memeber, IEEE

***Abstract*—Uncertainty is known to be a concomitant factor of almost all the real world commodities such as oil prices, stock prices, sales and demand of products. As a consequence, forecasting problems are becoming more and more challenging and ridden with uncertainty. Such uncertainties are generally quantified by statistical tools such as prediction intervals (PIs). PIs quantify the uncertainty related to forecasts by estimating the ranges of the targeted quantities. PIs generated by traditional neural network based approaches are limited by high computational burden and impractical assumptions about the distribution of the data. A novel technique for constructing high quality PIs using support vector machines (SVMs) is being proposed in this paper. The proposed technique directly estimates the upper and lower bounds of the PI in a short time and without any assumptions about the data distribution. The SVM parameters are tuned using particle swarm optimization technique by minimization of a modified PI-based objective function. Electricity price and demand data of the Ontario electricity market is used to validate the performance of the proposed technique. Several case studies for different months indicate the superior performance of the proposed method in terms of high quality PI generation and shorter computational times.**

***Index Terms*—Deregulation, Particle swarm optimization, Prediction interval, Support vector machines, Uncertainty.**

I. Introduction

Deregulation of power sector is a key step taken up by many countries for improving the efficiency of their power transactions. In the deregulated scenario, electricity is sold and purchased in a market in quite a similar fashion as other commodities. The prospective power generators submit their selling bids indicating the price and the quantum of power they are willing to trade with in the market. Similarly large scale consumers and retailer submit their offer bids to the market indicating their willingness and capacity of the purchase power. An independent entity known as market operator clears the submitted bids and offers and declares a uniform market clearing price which are applicable to both the buyers and the sellers. The market participants have to rely on forecasting models to estimate the future prices and demands and plan their operation and bidding strategies for maximizing their profit. Consequently, research and development of fast and accurate price and demand forecasting techniques is gradually gaining prominence in the deregulated power market scenario.

Nitin Anand is with the Department of Electrical Engineering, Indian Institute of Technology, New Delhi, India. (e-mail: anandnitin26@gmail.com)
A. Khosravi is with the Center for Intelligent Systems Research, Deakin University, Australia.(e-mail:abbas.khosravi@deakin.edu.au)
B.K.Panigrahi is with the Department of Electrical Engineering, Indian Institute of Technology, New Delhi, India. (e-mail:bijayaketan.panigrahi@gmail.com.)

Electricity is a very distinct market commodity compared to other commodities such as oil and gas owing to its non-storability and mandatory balance between demand and supply. Unlike demand, electricity prices are known for their unpredictable fluctuations and spikes and accurate forecast of electricity prices is quite a challenging problem for the market participants. Many researchers have delved into the problem of developing accurate forecasting modules encompassing a variety of market scenarios with varying ranges of forecast accuracy. Time series based methods rely on the past behavior of the prices and other exogenous variables like demand to estimate its future movement. Some of the time series based methodologies using transfer function, ARIMA and GARCH have been presented in [1], [2], [3] and reasonable accuracy levels have been achieved by them. Artificial intelligence based methods are also quite popular as they are simple and computationally efficient. The philosophy behind these methodologies is to find a mapping between the electricity prices and other explanatory variables such as historical prices, demand and weather, using historical examples. Some of the well known artificial intelligence techniques proposed in the price and load forecasting literature are neural networks [4], [5], fuzzy NNs [6], [7], [8] and support vector machines (SVM) [9].

It is impossible to eliminate forecasting errors due to the prominent uncertainty component intertwined with the electricity prices. This uncertainty amply reflects in the forecast errors generated by different models and needs to be quantified through certain measures. The commonly known statistical tools for quantifying the uncertainty in the predictions are Confidence Intervals (CIs) and Predictions Intervals (PIs). Confidence intervals are generated for an existing data based on the accuracy of the prediction of the regression i.e., of the mean of the target probability distribution. Prediction intervals take into account the accuracy with which the future targets have been encapsulated by the model [10]. PIs are more relevant to decision makers as compared to the CIs as they are more informative about the future. Wide PIs indicate higher uncertainty at that instance whereas low interval width indicate lower uncertainty. Therefore, forecasts with wide PIs should be used more cautiously by the decision makers while narrow width PIs can be used more confidently [11].

Some of the commonly known techniques for constructing PIs are Bayesian [12], delta [13], bootstrap [10], mean-variance estimation [14] and lower upper bound estimations using neural networks [15]. PIs are now being constructed for some of the real world problems and some of the contemporary applications of PIs have been reported for transportation [16], baggage handling systems [17] and financial services [18]. In [19], a hybrid method using support vector machines and a nonlinear conditional heteroskedastic forecasting model was proposed for constructing PIs of electricity prices. However, the methodology focussed on improvement of coverage probability and the width of the PIs was ignored. In [20], a decoupled extended Kalman filter-based NN method is employed for estimation of confidence intervals for electricity prices. An ARIMA-based method is developed in [21] for construction of CIs assuming a gaussian or uniform distribution for the residual errors. No discussion is made regarding PIs for a future electricity price in these papers. In [22], authors employ the delta technique for generating optimal PIs for electric load data. Delta technique is applied on the outcomes of neural network models and simulated annealing method is used for cost function minimization and optimization of network parameters. Authors reported that the proposed method could outperform traditional delta technique and generate reliable PIs.

The motivation of this work is derived from a lower-upper bound estimation technique proposed in [15] where the upper and the lower bounds of the PIs are directly generated by a two output NN model. In this work we propose a modification wherein support vector machines (SVM) are utilized for estimating the upper and lower bounds of the prediction intervals instead of NNs as proposed in the earlier work. NNs applications are limited by their possibility of getting stuck in local minima. SVM training, on the other hand, is claimed to always find a global minima [23]. The simple geometric interpretation of SVM gives a great scope for further investigations [24]. Keeping in view the advantages of SVM over NNs, SVM is considered as the learning technique in this work. Appropriate selection of the SVM parameters can tremendously boost its learning capabilities. Swarm intelligence optimization techniques are very effective in selecting optimal parameters for many real world problems. Particle swarm optimization (PSO) technique is a well known swarm intelligence method and has been used in this work for determination of optimal SVM parameters. The proposed SVM-PSO based lower-upper bound estimation method is used to construct optimal PIs for the future prices and demands of Ontario electricity markets. The performance of the proposed technique is also compared with some benchmark techniques and the ability of the proposed method to construct high quality PIs is clearly observed.

The rest of the paper is organized as follows. In Section II, review of the SVM, PSO and the various performance measures employed in the proposed method are presented. The steps of the model development are discussed in detail in Section III. The experimental studies and results are presented in Section IV. Section V highlights the conclusions of the work.

II. General Background

A. Support Vector Machines

Support vector machine algorithm originates from the Generalized Portrait algorithm developed in Russia in the late sixties and its present form was developed at the AT@T Bell Laboratories by Vapnik and associates. Initial focus of these algorithms was in the field of classification, but later on it was successfully extended to regression and time series predictions as well. In this section we give a brief introduction to support vector regression (SVR). Assume a training data $\{(x_1, y_1), ..., (x_\ell, y_\ell)\} \subset X \times \Re$ where each X denotes the input space of the sample. The objective of the algorithm is to determine a function $f(x)$ that allows at most ϵ deviation from the targets y_i for the entire training data and it should be as flat as possible. For the case of linear functions, f takes the form

$$f(x) = \langle w, x\rangle + b \quad w \in X, b \in \Re \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in X. Minimum flatness of f can be ensured by minimizing the norm $\|w\|^2$. This can be formulated as a convex optimization problem:

$$\text{minimize} \quad \frac{1}{2}\|w\|^2 \tag{2}$$
$$\text{subject to} \begin{cases} y_i - \langle w, x_i\rangle - b \leq \epsilon \\ \langle w, x_i\rangle + b - y_i \leq \epsilon \end{cases}$$

In practical situations, this constrained optimization problem may be infeasible and we have to allow for some errors. By introducing the slack variables ξ_i, ξ_i^*, the revised formulation can be stated as

$$\text{minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*) \tag{3}$$
$$\text{subject to} \begin{cases} y_i - \langle w, x_i\rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \quad i = 1, \ldots, l \end{cases}$$

The flatness of f and tolerance limits of the error beyond ϵ is regulated by the constant C. Inclusion of the constraints ensures that most of the data x_i lies in the tube $|y_i - \langle w, x_i\rangle - b| \leq \epsilon$. If x_i goes out of the tube, it results in an error ξ_i or ξ_i^* which is to be minimized in the objective function. The underfitting and overfitting of the training data is avoided by minimizing the training error $C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*)$ as well as the regularization term $\frac{1}{2}\|w\|^2$. Real world problems are primarily non linear and the above formulation can be made non-linear by mapping the data x to a higher dimensional space using a function $\Phi : x \to \Gamma$. Further details about SVMs can be found in [25], [26].

B. Particle Swarm Optimization

The PSO was developed and introduced by Kennedy and Eberhart as a stochastic optimization algorithm and is being widely used for solving a variety of complex optimization problems [27]. It is primarily a population based algorithm which employs a population of individuals to iteratively search a multi-dimensional search space for global minimum (maximum).The population in this case is referred to as *swarm* and the individuals are called as *particles*. Algorithm is initialized with a population of random particles and the particles explore the search space according to their individual velocity vector. The particles are considered to have the history of the best positions attained personally by them (personal best) and the best position attained by the entire group (group best) in every iteration. The further movement of each particle is influenced by these two factors.

Consider a PSO model with N particles which occupy positions in a D-dimensional problem space. The particles can be denoted as $x_{i1}, \ldots, x_{id}, \ldots x_{iD}$ and this population of particles represents a potential solution to the problem in hand. The velocity of each particle is denoted as $v_i = (v_{i1}, \ldots, v_{id}, \ldots v_{iD})$. The projected position and velocity of the particle i at $(k+1)^{th}$ iteration are defined by the following equations:

$$v_{ij}^{k+1} = \omega v_{ij}^k + c_1 r_{1j}^t \left(P_{ij}^k - X_{ij}^k\right) + c_2 r_{2j}^k \left(P_{gj}^k - X_{ij}^k\right) \quad (4)$$

$$X_{ij}^{k+1} = X_{ij}^k + v_{ij}^{k+1} \quad (5)$$

The constants c_1 and c_2 known as the cognitive and social coefficients respectively control the relative proportion of cognition and social interaction. r_{1j} and r_{2j} are two independent random numbers uniformly distributed in (0, 1) for the j^{th} dimension. Vector $P_g = (P_{g1}, \ldots, P_{gj}, \ldots, P_{gD})$ is the best position obtained so far by the entire population and is called *gbest*. Vector $P_i = (P_{i1}, \ldots, P_{ij}, \ldots, P_{iD})$ is the position with the personal best fitness found so far by the i^{th} particle, and it is called *pbest*. ω is known as the inertia weight parameter and it controls the velocity of particles from one iteration to the next iteration. The original version of PSO was found to lack velocity control mechanism and therefore a linearly decreasing inertia weight approach was introduced by Eberhart and Shi [28] to overcome that defect and it has been used in this work.

C. Performance measures

Quite analogous to point forecasts, the performance of the PI forecasting models need to be assessed in terms of the quality of PIs obtained by them. Coverage probability and PI width are the two commonly used PI performance measures. PI coverage probability (PICP) refers to the ability of the constructed PIs to capture the actual target variables. PICP can be mathematically stated as:

$$\text{PICP} = \frac{1}{N}\sum_{i=1}^{N} C_i \quad (6)$$

$$\text{where } C_i = \begin{cases} 1 \; t_i \in [L_i, U_i] \\ 0 \; t_i \notin [L_i, U_i] \end{cases}$$

where N is the number of samples in the test set, and L_i and U_i are lower and upper bounds of the i^{th} PI respectively.

PIs should be able to capture the target variables with a prescribed probability called the confidence level ($(1-\alpha)\%$). PIs can be considered to be reliable if $PICP \geq (1-\alpha)\%$ otherwise they should be discarded. Wide PIs have a high PICP but such PIs convey no information about the variation of the underlying target variables and are not practically useful. Therefore the width of PIs is a critical factor in evaluating their quality. PI normalized averaged width (PINAW) assesses PIs from this aspect and it can be mathematically stated as follows:

$$PINAW = \frac{1}{rN}\sum_{i=1}^{N}(U_i - L_i) \quad (7)$$

where r is the range of the underlying targets. The PICP and PINAW indices explained above assess the quality of PIs from two different aspects. A high PICP may come at the cost of wide PIs which are less informative and less useful for practical purposes. Similarly, reducing the width of intervals may reduce the possibilities of covering the desired target. Therefore evaluation of these indices independently or alone is not sufficient to give clear idea about the quality of the constructed PIs. Therefore, a comprehensive index consisting of both PICP and PINAW known as coverage width criterion (CWC) has been developed [29]:

$$\text{CWC} = PINAW * (1 + \gamma(PICP) e^{(-\eta(PICP-\mu))}) \quad (8)$$

$$\text{where } \gamma(PICP) \text{ is given by } \gamma = \begin{cases} 0 \; PICP \geq \mu \\ 1 \; PICP < \mu \end{cases}$$

where η and μ refers to the hyperparameters controlling the magnitude of CWC index. μ is the nominal confidence level for which the intervals are constructed. η value is set between 10 to 100 to penalize the invalid PIs. The CWC index is designed such that if PICP is less than the nominal confidence level, then CWC should be large regardless of the widths of intervals. If PICP is greater than the nominal confidence level, then γ becomes 0 and PINAW becomes the major criterion to be minimized. In this way, the index accommodates both requirements and gives a better indication of the quality of the PIs.

During our experiments, some irregularities were observed with respect to the present definition of CWC. It can be observed in (8) that the PINAW index is multiplied by all other terms and if PINAW is reduced to zero then the entire term (in effect, CWC) becomes zero irrespective of the PICP value. The objective of optimization technique used in our work is to minimize the CWC, therefore it is highly likely that CWC may be minimized to zero resulting in zero-width intervals. If PINAW is reduced to zero, even very low values of PICP become immaterial. Therefore, we suggest a slight improvement in the original CWC definition where PINAW is added rather than being multiplied with the rest of the terms. The modified CWC definition is stated as follows:

$$CWC = PINAW + \gamma(PICP) e^{(-\eta(PICP-\mu))} \quad (9)$$

Note that the original and modified CWC are equal if $PICP \geq (1-\alpha)\%$. Another index for evaluating the quality of PIs which is widely used in the literature is the Winkler score [30]. The score is calculated as follows:

$$\vartheta_i = U_i - L_i \quad (10)$$

$$S_i = \begin{cases} -2\alpha\vartheta_i - 4[L_i - t_i] \; if & t_i < L_i \\ -2\alpha\vartheta_i \qquad\qquad\quad if & t_i \in \vartheta_i \\ -2\alpha\vartheta_i - 4[t_i - U_i] \; if & t_i > U_i \end{cases}$$

where α is related to the confidence level ($\alpha = 0.1$ for 90% nominal confidence level). Then the overall score can be evaluated as:

$$\hat{S} = \frac{1}{N}\sum_{i=1}^{N} S_i \qquad (11)$$

III. Model development

Some of the traditional methods which are generally used for constructing PIs are the delta technique, the Bayesian technique and the Bootstrap technique. The delta technique is based on the assumption that multilayer feedforward NNs are non-linear regression models and they can be linearized using Taylor's series expansion [13]. Bayesian technique considers each parameter in a NN as a distribution and therefore the output of the network will also be in the form of distributions conditional on the observed training data [12]. Application of this method is limited by its massive computational burden and calculation of Hessian matrix. Bootstrap method [10], which is essentially a resampling method, is the most well known method for construction of PIs. However, this method requires large computational cost for large data sets. The mean-variance estimation-based method proposed by Nix and Weigand [14] uses a NN to estimate the characteristics of the conditional target distribution. The key assumption in this method is that noise is considered to be additive Gaussian with a nonconstant variance. This method requires lesser computational burden during the training and testing phase but it underestimates the variance of data, leading to a low empirical coverage probability [31]. Recently a new method for constructing the NN based PIs, called the LUBE method, was proposed in [15]. In this method, the PIs are constructed directly using a two-output NN structure without any assumption about the distribution of the sample data. The two outputs of the network directly give the lower and upper bounds of PIs. This one step process is easy, straightforward and can be easily implemented.

In this work we consider the same principle with a different approach. We use SVM as the machine learning technique in place of NNs. The motivation behind this step is to construct better quality PIs with a SVM which has some known advantages over NNs. SVM has the advantage of always finding a global minima while other approaches like NNs may be trapped in local minima [24]. In SVM, the number of parameters to be fixed are few in number and the basis functions are automatically selected. Additional advantage of SVM over NNs is that their results are stable over repeated executions. Results can be reproduced any number of times and they are also independent of the type of algorithm used to optimize the model [23]. Unlike NNs which can have multiple outputs, SVM is designed to give a single output. Therefore, in order to get two outputs corresponding to the lower and the upper bounds of the PI, we have to use two SVM models. One SVM model predicts the upper bound and the other predicts

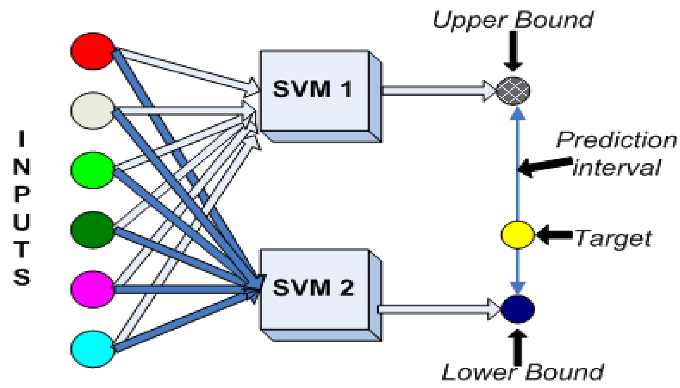

Fig. 1. Model structure for construction of PIs using two SVM models

the lower bounds subject to the same set of input features. The structure of the proposed model is depicted in Fig. 2

Radial basis function is generally considered to be a suitable kernel function for SVM. The reason is that RBF kernels are capable of analyzing higher-dimensional data with a few hyperparameters thus reducing the complexity of model selections [32]. Hence, the RBF function, $\exp\left\{-\gamma\left|x_i - x_j\right|^2\right\}$ is used in this work. C and γ are the two parameters of the RBF kernel function which influence the accuracy of forecasting [32], [33]. In order to determine the best parameter, a strategy for parameter selection needs to be devised. The most straightforward and basic technique is *grid Search* where a sequential search of the appropriate parameters is done within the possible bounds of these parameters. However, this technique is quite time consuming therefore some kind of optimization technique is required. Here we consider particle swarm optimization technique which is a powerful swarm intelligence based optimization technique. Since two SVM need to be used in this work, therefore totally four parameters (C and γ for each model) need tuning. The objective function to be minimized is the CWC index. As the PSO iterations proceed, optimal parameters which minimize the CWC index or in other words, improve the quality of the PIs, are obtained by the algorithm. Care should be taken during the training session that SVM does not over-fit the data as it may result in poor testing accuracy. The over-fitting problem during the training period can be avoided by the cross validation technique. In k-fold cross validation, the training data is divided into k equal size subsets. The model is trained using $k-1$ subsets and tested on one subset. In this way, each instance of the entire training set is predicted once and cross validation accuracy is given by the average accuracy obtained for all the subsets.

The various steps of the model are discussed as follows:

- *Step 1* Initialize the parameters of PSO algorithm such as the number of particles, maximum iterations, initial position and velocity. Each particle comprises of four dimensions i.e., the cost and the γ value for each of the two SVM's. The general range of the C should be between $[10^{-5}, 10^5]$ and that for γ should be between $[0, 10^1]$. Since the most suitable bounds of the SVM parameter for the given problem are not known apriori, therefore initially tests are conducted in the coarse range and then the search area is reduced to a finer zone with the boundaries of C and γ being $[0, 10^2]$ and $[0, 10^0]$ respectively. PIs are constructed for a confidence level of 90% and the η value lies between 10 to 100 for different case studies.

- *Step 2* The data to be studied needs to be selected from the available historical records. Since we are dealing with time series, a suitable time duration for training the model and for testing the actual performance of the model needs to be decided. A correlation analysis is done for each of the selected data set to determine prominent input features which affect the variation of electricity prices and demand.
- *Step 3* The selected data set should be splitted into two major divisions. One division corresponds to the training data which will be used to tune the model parameters and the other division corresponds to the testing data which is not seen by the model during the training session. It is very important to scale the data before applying it to the SVM model as it avoids numeric difficulties during calculation and also prevents features with large numeric ranges from dominating the features with low ranges. Therefore all the data sets are scaled in the range $[-1, +1]$.
- *Step 4* The training data is now given to the PSO optimization module. In this work, we consider a 5-fold cross validation strategy to avoid overfitting of the model to a specific set of data. The training data is further divided into 5 sets. 4 data sets are used for training at a time and PIs are constructed for the remaining data set.
- *Step 5* The PI performance measures like CWC, PICP and PINAW are evaluated for each of the cross validation step. The average CWC value of all the folds are considered as the fitness of each particle. This step is repeated for each particle of the swarm population and the fitness vector for that iteration is determined.
- *Step 6* The best fitness attained by the swarm and its corresponding particle dimensions are recorded as the group best fitness and group best position attained so far. The personal best position and the fitness of all the particles are also updated simultaneously.
- *Step 7* The termination condition for the optimization considered in this work is the maximum number of iterations. If the maximum iteration number is not reached, then the particle positions and velocities are updated and the next iteration starts with a new population. If the termination criterion is met, then the optimization is stopped and the most optimal values of the SVM parameters C and γ are recorded.
- *Step 8* The optimized values of the SVM parameters are now used to construct PIs for the testing data set. The quality of the constructed PIs is evaluated with the performance indices like PICP, PINAW, CWC and winkler score.

IV. Experimental studies and results

The data sets considered in this work are electricity prices and demands from the Ontario electricity market for the year 2010 [34]. Experiments are performed for all the months of the year 2010 with the available Hourly Ontario electricity prices (HOEP) and the total market demand. The input features required for creating the training and testing data sets for each of the month are determined using correlation analysis with the historical prices and demands respectively. The data sets are generally further scaled in the range $[-1, +1]$ in order to avoid conflicting numerical ranges of different features and also to make calculations easier.

Several experiments are now performed to evaluate the performance of the proposed methodology in different seasons. After performing the correlation analysis for historical prices for each of the data set, best six input features are selected and given as input to the model. For secure operation of the power systems and the electricity markets which affect the economy as well as the operation of entire nation, it is desirable to have forecast information with high confidence levels. Therefore in this study, we construct PIs for a nominal confidence level of $90\%(\alpha = 0.1)$, $95\%(\alpha = 0.05)$ and $99\%(\alpha = 0.01)$. For comparison purpose, we also implement two benchmark PI construction methodologies i.e., the Naïve method and NN based LUBE method. All the above studies are performed for the same set of datasets. The ANN based LUBE method is implemented using a single hidden layer network as performed in previous studies and the optimal network structure is obtained after a 5-fold cross validation. The number of neurons in the hidden layer is changed between 5 to 20.

The above experiments are implemented in the MATLAB programming environment. The training data is used to determine optimal SVM parameters which can be directly applied for the unknown test data set. During the training session, PIs are constructed for the validation data sets and the average CWC of the 5 validation sets is minimized by the PSO algorithm. The optimal parameters obtained during the training are now applied to the test data sets.

The experimental results for the test data sets for all the case studies pertaining to electricity prices are tabulated in Tables I-III for different nominal confidence levels. Table I shows the PICP, PINAW, CWC and the Winkler score obtained for the test data sets corresponding to a $90\%(\alpha = 0.1)$ nominal confidence level. From the table it can be seen that in all the months except May and December, the coverage probabilities of the PIs obtained with the proposed method are well above the nominal confidence level and therefore the PIs can be considered to be reliable. The minimum PICP is obtained for the May month (85.71%) while the maximum is for February (97.62%) followed by September (97.02%) and October (96.43%). However, their corresponding interval widths are different for each case. The highest width is obtained for the March month (70.88) which also has a high PICP. The minimum interval width is obtained for the month of July (8.13) with a corresponding PICP of (91.07%). A general observation is that with the exception of December month ($PICP = 89.88\%$ and $PINAW = 12.71$), generally higher PI widths are obtained for the fall and winter seasons and lower PI widths for the spring and summer seasons in the Ontario region.

Results corresponding to a 95.0% nominal confidence interval are presented in Table II and are quite similar to that of Table I with minor variations. The coverage probabilities of the obtained PIs are above the nominal confidence level in 8 out of 12 months. The errors for the lower PICP months are in the range 0.36% (April) to 6.31% (December). In 6 out of 12 months, a increase in the interval widths is observed with respect to 90% nominal confidence level results. The coverage probabilities show a minor increment in 6 months, decrement in 4 months and remains the same for 2 months. The winkler scores are significantly lesser for the all the seasons compared to the results of Table I.

Table III depicts the performance indices for PIs created for a nominal confidence level of 99.0%. In this case, the PICP of only one month (March) is above the nominal confidence level. The maximum and minimum error for the remaining months are 9.71% (December) and 1.38% (February). The PICP of the PIs is seen to increase in 4 months with respect to the results in Table II, remains the same for 4 months and decreases for 4 months. However, the interval widths increase for 7 months and decrease for the remaining 5 months.

TABLE I
PI PERFORMANCE INDICES FOR ONTARIO MARKET (NOMINAL CONFIDENCE = 90%)

Data Set	*PICP*	*PINAW*	*CWC*	*Winkler Score*
		Nominal confidence = 90%		
JAN	95.83	39.89	39.89	-1479.0
FEB	97.62	50.55	50.55	-1395.3
MAR	94.05	70.88	70.88	-1266.4
APRIL	94.05	34.77	34.77	-1501.6
MAY	85.71	26.88	35.40	-3099.6
JUNE	94.05	11.63	11.63	-2698.6
JULY	91.07	8.13	8.13	-5164.4
AUGUST	92.26	30.99	30.99	-1909.2
SEPTEMBER	97.02	11.02	11.02	-2422.2
OCTOBER	96.43	44.63	44.63	-1210.5
NOVEMBER	95.24	40.55	40.55	-1398.7
DECEMBER	89.88	12.71	13.77	-3506.4

TABLE II
PI PERFORMANCE INDICES FOR ONTARIO MARKET (NOMINAL CONFIDENCE = 95%)

Data Set	*PICP*	*PINAW*	*CWC*	*Winkler Score*
		Nominal confidence = 95%		
JAN	95.24	39.64	39.64	-856.20
FEB	97.62	51.98	51.98	-786.99
MAR	98.81	78.31	78.31	-635.46
APRIL	94.64	35.87	37.06	-841.76
MAY	95.24	35.43	35.43	-957.52
JUNE	95.83	11.61	11.61	-1791.3
JULY	91.07	7.84	14.97	-3889.1
AUGUST	90.48	29.18	38.78	-1325.9
SEPTEMBER	96.43	11.69	11.69	-1840.3
OCTOBER	97.02	43.75	43.75	-625.17
NOVEMBER	97.02	44.10	44.10	-735.63
DECEMBER	88.69	11.32	34.77	-2944.3

TABLE III
PI PERFORMANCE INDICES FOR ONTARIO MARKET (NOMINAL CONFIDENCE = 99%)

Data Set	*PICP*	*PINAW*	*CWC*	*Winkler Score*
		Nominal confidence = 99%		
JAN	97.02	42.02	44.71	-388.70
FEB	97.62	48.48	50.48	-400.62
MAR	99.40	73.54	73.54	-144.24
APRIL	93.45	36.10	52.12	-499.88
MAY	94.64	34.45	44.28	-425.0
JUNE	95.83	11.78	16.65	-1596.10
JULY	90.48	7.63	78.57	-4094.50
AUGUST	91.07	29.87	82.56	-800.55
SEPTEMBER	95.83	10.82	15.69	-1276.90
OCTOBER	97.02	47.50	50.19	-242.33
NOVEMBER	97.02	45.09	45.09	-314.61
DECEMBER	89.29	12.37	141.03	-2988.2

For a fair comparison, we construct and evaluate PIs for the same data sets using some benchmark techniques for all the case studies. These methods are Naïve method and recently developed NN-based LUBE method. The results of these experiments for electricity prices are presented in Table IV. The PICP index for all months of the year are below the nominal coverage level in case of results obtained with the Naïve method. The PICP for PIs generated by the LUBE ANN method are better but in two cases they are below the nominal coverage probability. High PICPs are also obtained by the proposed LUBE SVM method except in two cases where obtained PICP is below the nominal confidence level. While comparing the interval widths, we find that the Naive method generates lower width PIs in 9 months compared to LUBE NN at the cost of lower coverage probabilities. The interval widths for LUBE NN based methods are higher and the corresponding coverage probabilities are also higher. The results obtained with the LUBE SVM method are better in terms of PICP in most of the cases and the PI widths are also lower or comparative with the other methods in most of the cases.

Similar studies are performed for the electricity demand data of the Ontario market for all months of year 2010. The comparative results for PIs constructed with a nominal confidence of 90% are presented in Table V. The superior performance of the proposed method is clearly visible for all months except the May season with respect to the coverage probability. The interval width of PIs generated by the LUBE SVM method is less in all cases when compared to the Naïve and LUBE NN method.

The PIs for the month of January generated by Naïve, LUBE NN and the proposed LUBE SVM method are depicted in Figures 2, 3 and 4 respectively. The PIs generated by the Naïve method are able to capture most of the targets but they are not able to account for the cyclical patterns of the demand. LUBE-NN based PIs are able to capture the cyclical patterns but they show large variation around the true targets which is not very desirable. PIs generated by the proposed method are able to closely capture most of the targets and they are also able to follow the cyclical pattern of the demand. Therefore they convey the maximum information and are more

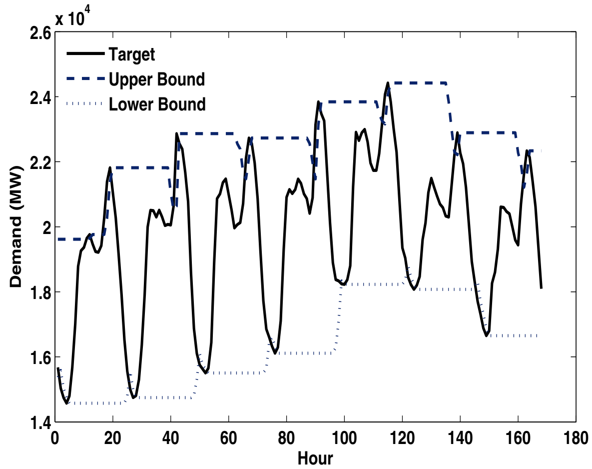

Fig. 2. PI for Ontario demand with Naïve method

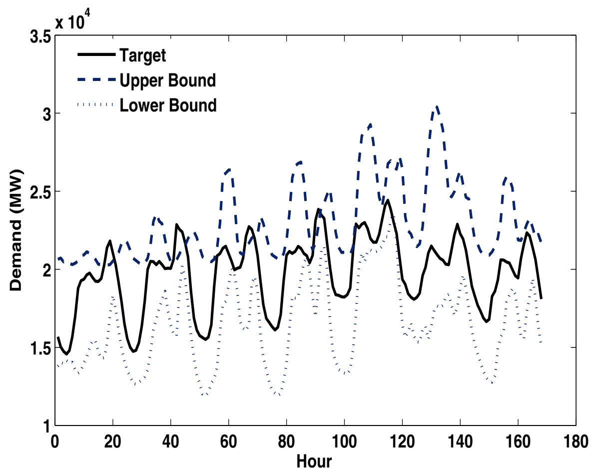

Fig. 3. PI for Ontario demand with LUBE NN method

reliable compared to PIs generated by other techniques. The PICP indices of other months are also well above the nominal confidence level except in one case. In some months, coverage probabilities close to 100% are also obtained with a minimum interval width of the generated PIs.

V. Conclusions

Prediction intervals are efficient statistical tools for quantifying the uncertainties associated with forecast models. Prediction intervals quantify the uncertainty in terms of the expected ranges within which the future targets are likely to lie. Traditionally prediction intervals are constructed using methods such as Delta, Bayesian and Bootstraps. However, their application is limited by their massive computational burden and doubtful assumptions about the data distributions. To overcome these limitations, we propose a novel technique using support vector machines where the upper and the lower bounds of the prediction intervals are directly estimated without prior assumptions about data distribution. The SVM parameters are optimized by particle swarm optimization technique which has strong capabilities of locating global

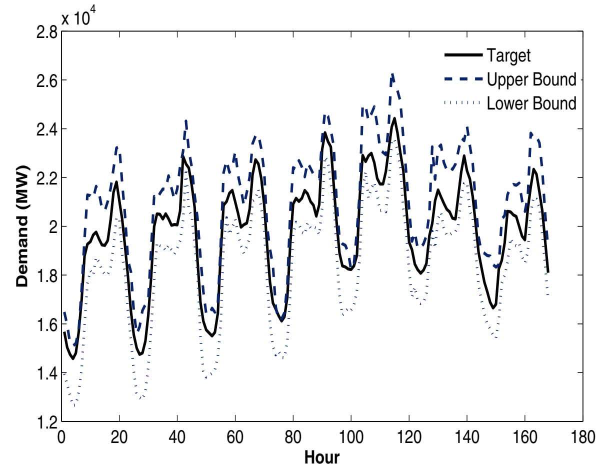

Fig. 4. PI for Ontario demand with LUBE SVM method

minima in lesser time. Model parameters are optimized by PSO through minimization of a modified PI coverage-width criterion. The performance of the proposed model is tested using data sets corresponding to the hourly electricity prices and demand of the Ontario electricity market. The performance of the proposed technique is compared with some benchmark techniques and the obtained results indicate the superiority of the method in generating high quality prediction intervals in a short time.

References

[1] F. Nogales and A. Conejo, "Electricity price forecasting through transfer function models," *Journal of the Operational Research Society*, vol. 57, pp. 350–356, 2006.

[2] J. Contreras, R. Espinola, F. Nogales, and A. Conejo, "Arima models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, 2003.

[3] R. Garcia, J. Contreras, M. van Akkeren, and J. Garcia, "A garch forecasting model to predict day-ahead electricity prices," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 867–874, 2005.

[4] Y.-Y. Hong and C.-Y. Hsiao, "Locational marginal price forecasting in deregulated electricity markets using artificial intelligence," *Proc. Gen. Transm. Dist.*, vol. 149, no. 5, pp. 621–626, 2002.

[5] C. Guan, P. Luh, L. Michel, Y. Wang, and P. Friedland, "Very short-term load forecasting: Wavelet neural networks with data pre-filtering," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 30–41, Feb 2013.

[6] N. Amjady, "Day-ahead price forecasting of electricity markets by a new fuzzy neural network," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 887–896, 2006.

[7] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Interval type-2 fuzzy logic systems for load forecasting: A comparative study," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1274–1282, Aug 2012.

[8] A. Khosravi and S. Nahavandi, "Load forecasting using interval type-2 fuzzy logic systems: Optimal type reduction," *IEEE Trans. Ind. Informat.*, vol. PP, no. 99, pp. 1–1, 2013.

[9] S. Fan, C. Mao, and L. Chen, "Next-day electricity-price forecasting using a hybrid network," *Generation, Transmission Distribution, IET*, vol. 1, no. 1, pp. 176–182, 2007.

[10] T. Heskes, "Practical confidence and prediction intervals," in *Adv. Neural Inf. Process. Syst. 9*. MIT press, 1997, pp. 176–182.

[11] A. Khosravi, S. Nahavandi, and D. Creighton, "Quantifying uncertainties of neural network-based electricity price forecasts," *Applied Energy*, vol. 112, no. 0, pp. 120 – 129, 2013.

[12] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995.

[13] J. T. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 748–757, 1997.

TABLE IV
COMPARISON OF PI QUALITY FOR ELECTRICITY PRICES WITH BENCHMARK FUNCTIONS

	NAIVE		LUBE ANN		LUBE SVM	
	PICP	***PINAW***	***PICP***	***PINAW***	***PICP***	***PINAW***
JAN	86.90	50.18	95.83	128.43	95.83	39.89
FEB	84.52	35.28	96.43	59.58	97.62	50.55
MAR	84.52	57.43	93.45	81.96	94.05	70.88
APRIL	83.93	41.97	86.90	108.08	94.05	34.77
MAY	82.14	48.30	83.33	45.19	85.71	26.88
JUNE	82.14	20.74	91.67	20.06	94.05	11.63
JULY	83.93	22.62	93.45	17.03	91.07	8.13
AUGUST	81.55	46.00	89.29	60.24	92.26	30.99
SEPTEMBER	82.74	21.46	97.52	32.30	97.02	11.02
OCTOBER	84.52	41.62	100.0	213.04	96.43	44.63
NOVEMBER	86.90	36.92	91.67	61.85	95.24	40.55
DECEMBER	87.50	26.11	92.86	34.97	89.88	12.71

TABLE V
COMPARISON OF PI QUALITY FOR ELECTRICITY DEMAND WITH BENCHMARK FUNCTIONS

	NAIVE		LUBE ANN		LUBE SVM	
	PICP	***PINAW***	***PICP***	***PINAW***	***PICP***	***PINAW***
JAN	79.17	63.50	94.64	92.20	95.83	25.61
FEB	79.17	72.09	97.62	95.64	95.83	28.83
MAR	82.74	69.68	95.24	110.94	98.81	35.72
APRIL	82.14	70.48	89.29	100.80	100.0	42.07
MAY	79.76	64.55	78.57	50.64	61.90	18.45
JUNE	83.33	65.23	89.88	84.75	94.64	26.06
JULY	79.17	68.47	96.43	68.94	96.43	23.25
AUGUST	75.00	52.84	91.07	59.19	98.21	18.28
SEPTEMBER	80.36	65.71	91.67	104.53	95.24	25.16
OCTOBER	80.95	73.29	86.90	88.03	97.02	34.70
NOVEMBER	81.55	78.85	89.88	80.78	99.40	39.87
DECEMBER	83.33	65.36	92.86	96.49	95.83	32.44

[14] D. Nix and A. Weigend, "Estimating the mean and variance of the target probability distribution," in *Proc. IEEE Int. Conf. Neural Netw. World Congr. Comput. Intell. 1994*, vol. 1, 1994, pp. 55–60 vol.1.

[15] A. Khosravi, S. Nahavandi, D. Creighton, and A. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Trans. Neural Networks*, vol. 22, no. 3, pp. 337–346, 2011.

[16] E. Mazloumi, G. Rose, G. Currie, and S. Moridpour, "Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 3, pp. 534 – 542, 2011.

[17] A. Khosravi, S. Nahavandi, and D. Creighton, "A prediction interval-based approach to determine optimal structures of neural network metamodels," *Expert Syst. Appl*, vol. 37, no. 3, pp. 2377 – 2387, 2010.

[18] D. F. Benoit and D. V. den Poel, "Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services," *Expert Syst. Appl*, vol. 36, no. 7, pp. 10 475 – 10 484, 2009.

[19] J. Zhao, Z. Dong, Z. Xu, and K. Wong, "A statistical approach for interval forecasting of the electricity price," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 267–276, 2008.

[20] L. Zhang and P. Luh, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended kalman filter method," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 59–66, 2005.

[21] M. Zhou, Z. Yan, Y. X. Ni, G. Li, and Y. Nie, "Electricity price forecasting with confidence-interval estimation through an extended arima approach," *Proc. Gen. Transm. Dist.,*, vol. 153, no. 2, pp. 187–195, 2006.

[22] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1496–1503, Aug 2010.

[23] K. P. Bennett and C. Campbell, "Support vector machines: hype or hallelujah?" *SIGKDD Explor. Newsl.*, vol. 2, no. 2, pp. 1–13, Dec. 2000.

[24] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data. Min. Knowl. Disc.*, vol. 2, pp. 121–167, 1998.

[25] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[27] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, 1995, pp. 1942–1948.

[28] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, 1998, pp. 69–73.

[29] A. Khosravi, S. Nahavandi, D. Creighton, and A. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Trans. Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.

[30] R. L. Winkler, "A decision-theoretic approach to interval estimation," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 187 – 191, 1972.

[31] R. Dybowski and S. J. Roberts, "Confidence intervals and prediction intervals for feed-forward neural networks," in *Clinical Applications of Artificial Neural Networks*. University Press, 2001, pp. 298–326.

[32] C.-C. Chang and C.-J. Lin, "A practical guide to support vector classification," pp. 1–16, 2003, available at: http://www.csie.ntu.edu.tw/ cjlin/papers/guide/guide.pdf.

[33] C.-L. Huang and C.-J. Wang, "A ga-based feature selection and parameters optimizationfor support vector machines," *Expert Syst. Appl*, vol. 31, no. 2, pp. 231 – 240, 2006.

[34] W. of Ontario electricity market. (2013, Sep.) http://www.ieso.ca @ONLINE. [Online]. Available: http://www.ieso.ca