Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Predicting Electricity Pool Prices Using Hidden Markov Models ⋆

Ouyang Wu * **Tianbo Liu** ** **Biao Huang** *** **Fraser Forbes** ****

* *Department of Chemical and Material Engineering, University of Alberta, Edmonton, AB, T6G-2V4, Canada (e-mail: owu@ualberta.ca).*
** *Department of Chemical and Material Engineering, University of Alberta, Edmonton, AB, T6G-2V4, Canada (e-mail: tianbo2@ualberta.ca)*
*** *Department of Chemical and Material Engineering, University of Alberta, Edmonton, AB, T6G-2V4, Canada (e-mail: biao.huang@ualberta.ca)*
**** *Department of Chemical and Material Engineering, University of Alberta, Edmonton, AB, T6G-2V4, Canada (e-mail: fraser.forbes@ualberta.ca)*

Abstract: In this paper, Alberta electricity spot market or Power Pool pricing is studied and the pool price is modeled through a hidden Markov model and multiple local ARX models. By selecting and preprocessing the exogenous factors (e.g. the price forecast from Alberta Electric System Operator (AESO), demand forecast and so forth), a one-hour ahead prediction model for pool price is formulated with parameters being estimated from the real data. Validation results show that this approach can improve the price forecasting and in particular, for high pool prices.

Keywords: Hidden Markov Model, Regime-switching, Alberta's Electricity Market, Local Models, Periodic Patterns, Autoregressive Exogenous Model

1. INTRODUCTION

In recent decades, electricity market deregulation has become a world-wide trend. By introducing competition, it is expected that electricity market efficiency is improved, which provides opportunities and also presents challenges to both the generators and the consumers. As a result, electricity price prediction has become an important issue in deregulated electricity market areas.

There are a number of challenges in the prediction of electricity prices due to their high volatility and erratic nature. Based on the characteristics of electricity pricing, price regime-switching models are proposed to model switching between different states such as normal pricing and spike pricing. Ethier and Mount (1998) applied Markov regime-switching models to electricity prices in United States and Australia markets, and they confirmed the existence of two states with different means and variances. Huisman and Mahieu (2003) proposed Markov regime-switching to model price spikes in Europe electricity markets with three states, which include normal electricity price dynamics, a jump state describing sudden increases or decreases, and a state describing a recovery from a jump state to a normal state. Hidden Markov model (HMM) is a common approach to deal with electricity price model with hidden regimes, which has been applied in many financial problems (see Mamon and Elliott (2007)). In the system identification literature, a multi-linear model approach becomes popular in approximating nonlinear systems; see Jin and Huang (2010), Jin et al. (2011) and Jin and Huang (2012). Among these literature, Jin and Huang (2012) proposed an identification approach for switched Markov autoregressive exogenous model based on expectation-maximization (EM) algorithm, and simulations show good performance in solving nonlinear identification problem.

In this paper, a Markov regime-switching model based on the electricity spot price is combined with multiple autoregressive exogenous (ARX) models to predict the pool price in the Alberta electricity market.

2. BACKGROUND

In 1996, Alberta's electricity market began to evolve to a deregulated market with full deregulation established in 2001 (Market Surveillance Administrator (2010)). All wholesale electrical energy generated in Alberta which is not consumed on site, must flow through a power pool that is operated by Alberta's independent system operator called Alberta Electric System Operator (AESO). Thus, the power pool is Alberta's wholesale spot electricity market, and the hourly electricity price for power pool is called pool price.

Due to mechanics of electricity pricing, there are many characteristics specific to the Alberta Power Pool, which are summarized in Xiong (2004) and Market Surveillance

⋆ This work is supported in part by NSERC and AITF.

Administrator (2010). First, there are apparent on-peak and off-peak electricity price patterns, and the on-peak period is often from 8:00 to 21:00 during weekdays. Also, there are pronounced periodic effects for the pool price, like daily, intra-daily and weekly repeating patterns, or even the monthly repeating patterns. For example, prices vary with demands in a day, which presents an hourly pattern. Another characteristic of the Power Pool is price spikes. Since shocks in demand and supply are common, the pool prices may be quite volatile in certain periods. For example, unplanned outages along transmission lines can drive the pool price to a high level, such as 500 \$/MWh or more.

AESO provides a pool price forecast and a load forecast. The two-hour ahead pool price forecast can have large prediction errors compared to the actual pool price as the historical data showed, especially when prices spike. One reason for these errors is that generators are free to modify their supply offers two hours ahead, which could result in the dispatch level in the next two hour period being quite different from the current dispatch level. In this case, a prediction model for Alberta's pool price with better forecasting performance would benefit customers with their decisions on electricity consumption.

3. ALGORITHM

In this section, a Markov regime-switching autoregressive exogenous model is formulated. The regime-switching mechanism is based on a feature-extracted electricity pool price with hidden Markov models based on the work of Liu (2013). To identify the sub-models of the Markov states or regimes, a maximize-likelihood estimator is applied to estimate models from the complete data log likelihood function.

3.1 Feature extraction for electricity pool price

To build a hidden Markov model for electricity pool pricing, some simplifications are required. The pool price sequence is transformed into a symbol sequence with reduced representation set of features (i.e., feature extraction is used). Based on the pool price time sequence, the data are divided into three types: peak-up, peak-down and off-peak using data segmentation. The specific rules are as follows:

First, the pool prices are divided into five groups based on their absolute value with group index from low to high represented as: #1: less than 30 \$/MWh, #2: from 30 \$/MWh to 100 \$/MWh, #3: from 100 \$/MWh to 300 \$/MWh, #4: from 300 \$/MWh to 500 \$/MWh, #5: more than 500 \$/MWh. Then, a new processed sequence as discretized trend of price is developed by calculating the group index differences for every two neighboring pool price as follows:

$$S_k = \begin{cases} 1 \text{ for } d_k > 2 \\ 2 \text{ for } d_k < -2 \\ 3 \text{ for } else \end{cases} \tag{1}$$

where, S_k is the element of the pool price symbol sequence at time k; d_k is the group index differences between each two neighboring electricity prices.

The peak price or spike price always occurs during the peak hours between 9 am and 4 pm. Meanwhile, the off-peak characteristics or low prices may also appear in these peak hours. In the off-peak hours, the high prices and peak characteristics may occur. Then, it is assumed that there are some mechanisms that govern the changes of pool price that follows the Markov chain, which is defined as a discrete state process $\{I_k\}$ with three regimes as peak-up, peak-down and off peak. Therefore, a hidden Markov model for the price regimes is built based on the feature-extracted pool price sequence.

3.2 Parameter estimation and decoding for hidden Markov model

The discrete hidden Markov model (HMM) in terms of processes I_k and S_k is formulated (Elliott et al. (1994)):

$$\begin{aligned} F_X(k+1) &= A \cdot F_X(k) \\ F_Y(k) &= C \cdot F_X(k) \end{aligned} \tag{2}$$

where, $F_X(k)$ is the probability vector function for hidden state I_k at time instant k, $F_X(k) \in S_X = \{e_1, e_2, ..., e_M\}$; $F_Y(k)$ is the probability vector function for the discretized observed symbol S_k at time instant k, $F_Y(k) \in S_Y = \{f_1, f_2, ..., f_N\}$; e_i and f_i are the unit vector in S_X and S_Y respectively with unity in the i^{th} position and zeros elsewhere.

$$\begin{aligned} F_X(k) &= \sum_{i=1}^{M} P(I_k = i | I_{k-1}, \Theta_m) \cdot e_i \\ F_Y(k) &= \sum_{j=1}^{N} P(S_k = j | I_k, \Theta_m) \cdot f_j \end{aligned} \tag{3}$$

$A = [a_{ij}]^T \in R^{M \times M}$, $a_{ij} = P(I_{k+1} = j | I_k = i)$ is the probability of the state j given the previous state i that defined as the transition probability; $C = [c_{ij}]^T \in R^{N \times M}$, $c_{ij} = P(S_k = j | I_k = i)$ refers to the probability that symbol j is seen when in state i, which is defined as the emission probability; the hidden Markov model is assumed to be homogeneous so that a_{ij} and c_{ij} do not depend on time instant k, and $\sum_{j=1}^{M} a_{ij} = 1$, $\sum_{j=1}^{M} c_{ij} = 1$ are satisfied; Θ_m is the parameters for hidden Markov model denoted as $\Theta_m = \{A, C\}$.

As only the discretized observation sequence $\{S_k\}$ is known, an expectation-maximization (EM) algorithm is applied for the hidden Markov modelling to estimate the parameters of transition probabilities $\{a_{ij}\}$ and emission probabilities $\{c_{ij}\}$.

The EM algorithm is an iterative approach for maximum likelihood estimation with missing data. There are two iterative steps: the E-step calculates a lower bound of the likelihood function called Q function, which is based on the old parameter estimation; the M-step maximizes the Q function with respect to the parameters to find new estimates of parameters. Here we use Baum-Welch algorithm (Durbin (1998)) as the EM algorithm computation for hidden Markov model.

Based on the estimated model parameters Θ_m and the discretized observation sequence $\{S_k\}$, the posterior state probability for the state sequence $\{I_k\}$ can be calculated via a forward algorithm and a backward algorithm.

The forward probability is defined as:

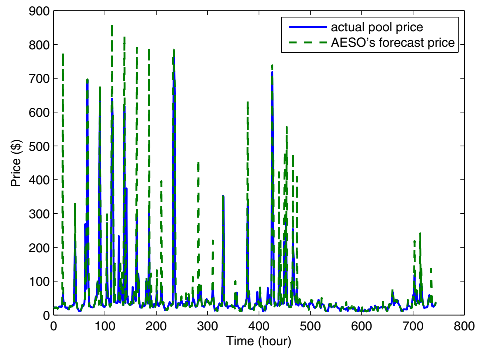

Fig. 1. Actual pool price and AESO's pool price forecast (Dec 2013)

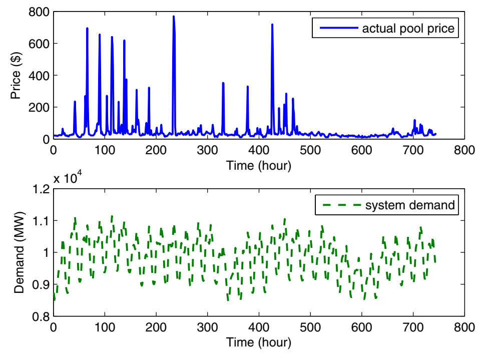

Fig. 2. Actual pool price and AESO's system demand forecast (Dec 2013)

$$f_i(k) = P(S_{1:k}, I_k = i|\Theta_m) \tag{4}$$

The backward probability is defined as:

$$b_i(k) = P(S_{k+1:L}|I_k = i, \Theta_m) \tag{5}$$

The procedures for calculating $f_i(k)$ and $b_i(k)$ can be found in Durbin (1998). Given the forward probability and backward probability, the smoothed posterior state probability can be calculated as follows:

$$P(I_k = i|S_{1:L}, \Theta_m) = \frac{f_i(k) \cdot b_i(k)}{P(S_{1:L})} \tag{6}$$

3.3 Input variables selection and data preprocessing

Some pool price characteristics are: 1) the electricity price in Alberta shows strong periodic behavior; 2) the AESO's forecast pool price reflects the fluctuation of future electricity prices; 3) the day ahead forecast demand by AESO affects the bidding results of the generators, which is related to the pool price; 4) the actual demand is correlated with the pool price at the same time instant; however, since actual demand is unavailable at the time of prediction, the historical data for the immediate past can be used for prediction. Therefore, the time sequence, the real-time forecast pool price by AESO, the history data of actual system demand and the real-time day ahead forecast demand by AESO are chosen as the input variables to predict the real time pool price.

Here we choose ARX model as the local sub-model for pool price prediction. Thus, the input variables are transformed to build a linear relationship with the pool price. From Fig. 1, the forecast pool price shows good correlation with actual price; however, the relation between actual pool price, system demands and time sequence appears to be non-linear; see Fig. 2.

To build a linear correlation with the actual pool price, the time sequence is preprocessed. First, the time sequence is transformed to be periodic with respect to a 24 hour time clock to appropriately reflect the periodic pattern of pool price. Then the weights for on-peak and off-peak hours are calculated based on the following weighting formula, and preprocessed time sequence is presented in Fig. 6.

$$\begin{aligned} &F(k) = K(k) \cdot \exp(\frac{(k-k_p)^2}{2\sigma_p^2}) \\ &k_p \sim PMF \\ &K(k) = f(P(I_k|S_{1:k-1}, \Theta_m)) \end{aligned} \tag{7}$$

where, $F(k)$ is the preprocessed time sequence, weighted by peak-price magnitude $K(k)$ and Gaussian function $\exp(\frac{(k-k_p)^2}{2\sigma_p^2})$, and σ_p^2 is a tuning parameter; k_p is the hourly time instant with the peak price in a day, which is a random variable with a probability mass function (PMF) based on historical data; peak-price magnitude $K(k)$ is function of the posterior state probability $P(I_k|S_{1:k-1}, \Theta_m)$ at time instant k to show the possibility of the price that is governed by the on-peak or off-peak price state. Here, $P(I_k|S_{1:k-1}, \Theta_m)$ works as a predictor, as opposed to the smoother $P(I_k|S_{1:L}, \Theta_m)$, which can be derived using Bayes' rule as follows:

$$\begin{aligned} P(I_k|\ S_{1:k-1}, \Theta_m) &= \sum_{I_{k-1}} P(I_k, I_{k-1}|S_{1:k-1}, \Theta_m) \\ &= \sum_{I_{k-1}} P(I_k|I_{k-1}) P(I_{k-1}|S_{1:k-1}, \Theta_m) \\ &= \sum_{I_{k-1}} P(I_k|I_{k-1}, \Theta_m) \frac{P(S_{1:k-1}, I_{k-1}|\Theta_m)}{P(S_{1:k-1})} \\ &= \sum_{i} a_{ij} \cdot \frac{f_i(k-1)}{P(S_{1:k-1})} \end{aligned} \tag{8}$$

Preprocessing of system demands is based on the electricity market mechanism. All of the generators in Alberta submit their offers to the Power Pool with their available capacity and desired prices. These offers are ranked from lowest to highest in price to meet the system demands with the high-price surplus capacities to be dispatched off. The hourly supply offer curve can be drawn as a piece-wise function as in Fig. 3 (Market Surveillance Administrator (2010)). In most cases, when system demand is more than 9000 MW, a high pool price would be triggered, which conforms to the piece-wise behavior of hourly supply offer curve. Therefore, the demand time series can be processed such that the portions over 9000 MW are emphasized, while the low demand portions are flattened. The preprocessed curves are presented in Fig. 4 and Fig. 5.

3.4 Model identification and prediction

For each hidden state or regime $\{I_k\}$, it is assumed that the input-output relationship follows the local linear model as:

$$y_k = \phi_k^T \theta_{I_k} + v_k, \quad k = 1, 2, \ldots, L \tag{9}$$

where, ϕ_k is the regressors, expressed as:

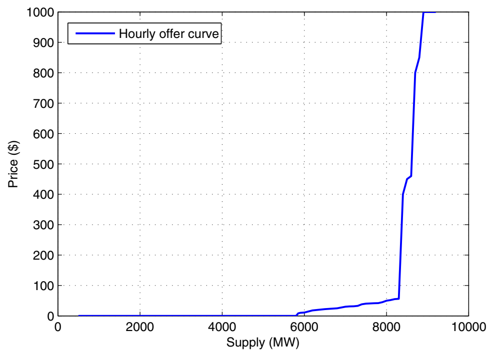

Fig. 3. Typical hourly offer curve

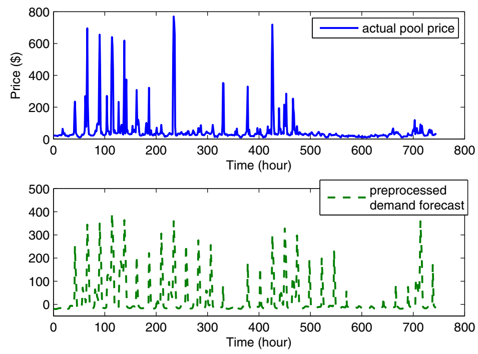

Fig. 4. Preprocessed forecast demand(Dec 2013)

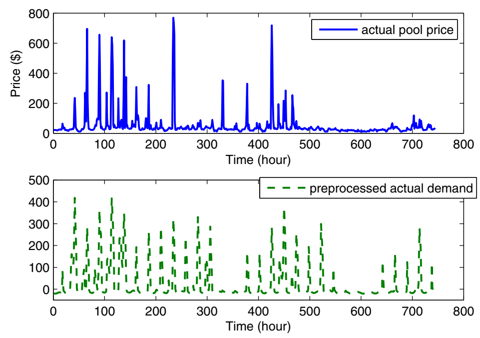

Fig. 5. Preprocessed actual demand(Dec 2013)

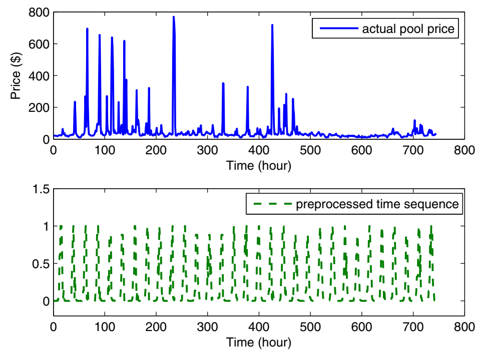

Fig. 6. Preprocessed time sequence(Dec 2013)

$$\phi_k = [y_{k-1}, y_{k-2}, \ldots, y_{k-n_a}, u_{k-1}^T, u_{k-2}^T, \ldots, u_{k-n_b}^T]^T$$

n_a and n_b are orders of denominator and numerator for ARX model; θ_{I_k} are the parameters for the local sub-model with hidden state I_k that indicates the model identity; v_k is assumed to be Gaussian noise with zero mean and variance σ^2; Here, y_k is the actual pool price at time instant k, and $\{u_k\}$ is the time series vector for inputs such as preprocessed demands, AESO's price forecast and preprocessed time sequence. To estimate the parameters for the switching model, the posterior state possibilities $P(I_k|S_{1:L}, \Theta_m)$ for hidden state $\{I_{1:L}\}$ is computed based on the discrete observed pool price symbols $\{S_{1:L}\}$ according to section 3.2, and the parameters for the local sub-models are calculated through following maximum-likelihood estimation:

$$\Theta_{ml} = \arg\max_{\Theta} \ln P(Z_{1:L}|\Theta) \tag{10}$$

The log likelihood function can be derived using Jensen's inequality ($\ln \sum_{i=1}^n \lambda_i x_i \geq \lambda_i \sum_{i=1}^n \ln x_i$):

$$\begin{aligned}
\ln P(Z_{1:L}|\Theta) &= \ln \sum_{I_{1:L}} P(Z_{1:L}, I_{1:L}|\Theta) \\
&= \ln \sum_{I_{1:L}} P(Z_{1:L}|I_{1:L}, \Theta) P(I_{1:L}|S_{1:L}, \Theta_m) \\
&\geq \sum_{I_{1:L}} P(I_{1:L}|S_{1:L}, \Theta_m) \ln P(Z_{1:L}|I_{1:L}, \Theta) \\
&= \sum_{I_{1:L}} P(I_{1:L}|S_{1:L}, \Theta_m) \ln \prod_{k=1}^{L} P(Z_k|Z_{1:k-1}, I_k, \Theta) \\
&= \sum_{k=1}^{L} \sum_{i=1}^{M} P(I_k = i|S_{1:L}, \Theta_m) \ln P(Z_k|Z_{1:k-1}, \theta_{I_k})
\end{aligned} \tag{11}$$

where Z_k is the observed data at time instant k, and $Z_k = \{y_k, u_k\}$; for each local sub-model, the noise is assumed to follow zero-mean Gaussian distribution.

To maximize the log-likelihood function over parameters Θ, derivative is taken with respect to each local sub-model parameter θ_i, and let it be zero. Then, we have:

$$\begin{aligned}
\theta_i =& [\sum_{k=1}^{L} P(I_k = i|S_{1:L}, \Theta_m) \phi_k \phi_k^T]^{-1} \cdot \\
& [\sum_{k=1}^{L} P(I_k = i|S_{1:L}, \Theta_m) \phi_k y_k]
\end{aligned} \tag{12}$$

The prediction for next hour's pool price is:

$$E(k+1) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(I_k = i|S_{1:k}, \Theta_m) \cdot a_{ij} \cdot E^j(k+1) \tag{13}$$

where, $E^j(k+1)$ is the prediction from the j^{th} local sub-model ($E^j(k+1) = \phi_{k+1}^T \theta_j$); $P(I_k = j|S_{1:k}, \Theta_m)$ is the posterior state possibility given current and history data (filter), based on forward algorithm as follows:

$$\begin{aligned}
P(I_k = i|S_{1:k}, \Theta_m) &= \sum_{i=1}^{M} \frac{P(S_{1:k}, I_k = i|\Theta_m)}{P(S_{1:k})} \\
&= \sum_{i=1}^{M} \frac{f_i(k)}{P(S_{1:k})}
\end{aligned} \tag{14}$$

4. VALIDATION STUDIES

In this section, validation studies for monthly predictions are presented using the proposed approach, and the data is from the AESO's website [1]. Hourly Predictions within a month (around 750 data points) are chosen for validation studies as it is long enough to include some price spikes and periodic behaviors. For training purposes, a batch of historical data is selected in length of a month (around 750 data points) due to consideration of a balance between computation cost and modelling accuracy.

4.1 Monthly training set selection

Parameter estimates are required to predict the pool price, and the parameters may vary seasonally. The proposed parameter estimation approach is applied to a batch of historical data (i.e., training data). A monthly training set selection rule is proposed to obtain a better monthly prediction performance with more robust parameter estimates.

To evaluate the prediction performance, the root mean squared error (RMSE), correlation coefficient and fitting rate are applied as monthly validation metrics.

$$Fitting\ rate = 1 - \frac{norm(F - Y)}{norm(Y - mean(Y))} \tag{15}$$

$$Corr = \frac{\sum_{j=1}^{n}(f_j - mean(F))(y_j - mean(Y))}{\sqrt{\sum_{i=1}^{n}(f_i - mean(F))^2(y_i - mean(Y))^2}} \tag{16}$$

where, f_i, y_i refer to the prediction and actual value in time instant i respectively, and F, Y are their matrix form over the entire validation data set; correlation coefficient measures the linear relationship between two data sets, and 1 means perfect positive linear correlation; fitting rate measures the variation of the output in percentage that is subtracted from 1.

Two case studies are presented in Table 1 and Table 2. Considering the predictions in January 2014 and June 2014, respectively, as examples, the prediction performance using monthly training sets are given in the Tables and are compared to AESO's forecast. January is a typical month for AESO to provide good forecasts where the new approach does slightly better, while June is a typical month in which the new approach may provide a better prediction than AESO's forecast. Moreover, we found that choosing the same month from the previous year as the training set produces the best predictions.

An exception is April 2014; see Table 3. Compared to January and June in 2014, April 2014 is an unusual month since no price spike occurs. The prediction performances using any month of the previous year cannot outperform the AESO's. The reason is that the spike-price characteristic depicted in the sub-models is not applicable in April 2014. On the other hand, AESO's price forecasts tend to have reasonably good performance on the low pool prices. Therefore, the proposed approach is more useful in the relatively high-price region.

By setting the threshold for the high pool price as 100$/MWh, the prediction performance for high pool price is calculated and shown in Table 4 and Table 5. The results also demonstrate that using the same month of the previous year as training set leads to the best predictions.

[1] http://ets.aeso.ca/

Table 1. Validation results for January 2014

Method	Month(Tr)	RMSE	Corr(%)	Fit(%)
AESO's	N/A	32.01	90.16	53.56
Proposed	Dec 2013	42.73	83.72	38.00
	Nov 2013	36.91	86.44	46.45
	Oct 2013	36.74	87.78	46.69
	Sept 2013	35.38	85.83	48.67
	Aug 2013	38.01	83.43	44.86
	Jul 2013	39.45	82.22	42.77
	Jun 2013	41.34	80.74	40.03
	May 2013	35.96	85.63	47.84
	Apr 2013	36.12	86.33	47.59
	Mar 2013	38.65	84.29	43.92
	Feb 2013		N/A	
	Jan 2013	31.32	89.34	54.55

Table 2. Validation results for June 2014

Method	Month(Tr)	RMSE	Corr(%)	Fit(%)
AESO's	N/A	53.11	83.99	34.84
Proposed	May 2014	37.01	89.14	54.60
	Apr 2014		N/A	
	Mar 2014	46.01	83.45	43.55
	Feb 2014	33.57	91.49	58.82
	Jan 2014	33.35	91.45	59.07
	Dec 2013	41.83	91.87	48.69
	Nov 2013	35.63	91.75	56.29
	Oct 2013	34.76	92.29	57.36
	Sept 2013	34.76	92.29	57.36
	Aug 2013	33.61	91.27	58.77
	Jul 2013	36.84	89.28	54.80
	Jun 2013	34.72	90.79	57.40

Table 3. Validation results for April 2014

Method	Month(Tr)	RMSE	Corr(%)	Fit(%)
AESO's	N/A	3.61	97.22	75.53
Proposed	Mar 2014	4.88	96.77	66.95
	Dec 2013	7.06	94.72	52.20
	Aug 2013	7.47	88.16	49.46
	Jul 2013	6.54	90.67	55.78
	Jun 2013	9.55	84.15	35.42
	May 2013	6.37	94.21	56.88
	Apr 2013	7.90	93.87	46.56

Table 4. Validation results on high-price region for January 2014

Method	Month(Tr)	RMSE	Corr(%)	Fit(%)
AESO's	N/A	139.54	81.42	37.03
Proposed (high-price)	Dec 2013	185.86	69.44	16.12
	Nov 2013	159.83	75.35	27.87
	Oct 2013	159.50	78.25	28.02
	Sept 2013	151.74	75.27	31.52
	Aug 2013	162.28	71.71	26.76
	Jul 2013	169.51	71.34	23.51
	Jun 2013	170.32	67.86	23.14
	May 2013	149.87	75.07	32.37
	Apr 2013	151.21	76.73	31.76
	Mar 2013	148.81	76.26	32.85
	Feb 2013		N/A	
	Jan 2013	134.40	80.76	39.34

4.2 Robust monthly prediction on high pool price

To test the robustness of the proposed pool price prediction, the training set selection method discussed in the previous section is tested in various monthly pool price prediction. The results are presented in Table 6. These results confirm that the proposed approach has better prediction performance for high pool prices than AESO's forecast.

4.3 Special case for monthly training set selection rule

In Table 6, September 2014 and April 2014 show no high prices so that no high-price prediction performance is given. Besides, the proposed approach for pool price prediction in February 2014 is not applicable since February 2013 has no price spikes but all off-peak price is similar to April 2014 and September 2014, which means no spikes for parameter estimation of the local sub-models and hidden Markov models. Expanding training set to the neighboring month which includes the spikes, can fix this problem, and corresponding high price prediction performance with training months including January and February in 2013 outmatches AESO's forecast; see Table 6 and Table 7.

Table 5. Validation results on high-price region for June 2014

Method	Month(Tr)	RMSE	Corr(%)	Fit(%)
AESO's	N/A	268.46	63.42	1.84
Proposed (high-price)	May 2014	182.93	78.24	33.11
	Apr 2014		N/A	
	Mar 2014	229.93	62.38	15.93
	Feb 2014	164.00	81.46	40.03
	Jan 2014	164.98	80.88	39.68
	Dec 2013	206.74	83.67	24.41
	Nov 2013	175.22	82.39	35.93
	Oct 2013	170.96	84.03	37.49
	Sept 2013	163.41	82.65	40.25
	Aug 2013	157.83	82.76	42.29
	Jul 2013	179.60	80.88	34.33
	Jun 2013	159.05	82.27	41.84

Table 6. Validation results on high-price region for different months using same month training set selection rule

Month(Val)	Method	RMSE	Corr(%)	Fit(%)
Sept 2014	N/A		N/A	
Aug 2014	AESO's	192.15	88.48	45.94
Aug 2014	Proposed	138.88	94.64	60.93
Jul 2014	AESO's	253.45	71.53	22.35
Jul 2014	Proposed	220.86	77.88	32.33
Jun 2014	AESO's	268.46	63.42	1.84
Jun 2014	Proposed	159.05	82.27	41.84
May 2014	AESO's	199.93	76.84	29.47
May 2014	Proposed	150.92	84.84	46.76
Apr 2014	N/A		N/A	
Mar 2014	AESO's	80.91	81.30	-5.72
Mar 2014	Proposed	58.46	73.52	23.61
Feb 2014	AESO's	94.32	94.31	61.92
Feb 2014	Proposed		N/A	
Jan 2014	AESO's	139.54	81.42	37.03
Jan 2014	Proposed	134.40	80.76	39.34
Dec 2013	AESO's	200.04	68.53	0.98
Dec 2013	Proposed	128.06	77.56	36.61
Nov 2013	AESO's	168.04	82.70	-24.27
Nov 2013	Proposed	84.96	80.66	37.16
Oct 2013	AESO's	200.42	72.70	5.57
Oct 2013	Proposed	129.99	79.71	38.75

Table 7. Validation results for February 2014

Method	Month(Tr)	RMSE	Corr(%)	Fit(%)
AESO's	N/A	40.89	96.44	69.79
Proposed	Mar 2013	48.38	93.94	64.25
	Feb 2013		N/A	
	Jan 2013	40.62	95.96	69.98
	Jan+Feb 2013	39.73	96.58	70.64
	Feb+Mar 2013	35.63	91.75	56.29
High-price	Jan+Feb 2013	88.99	95.05	64.07

5. SUMMARY

A pool price prediction approach is developed and applied in Alberta's electricity market. The prediction model is a combination of the Markov regime-switching mechanism and a multi-model identification approach. The system identification approach is based on the hidden Markov model and maximum likelihood estimation. To apply this approach on Alberta's pool price prediction, some heuristic strategies like feature extraction and data preprocessing are applied to the data. In validation studies, a training set selection strategy is proposed and applied in the prediction of high pool price. Validation studies illustrate good performance and the robustness of prediction, particularly in high pool price regions.

REFERENCES

Durbin, R. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids.* Cambridge university press.

Elliott, R.J., Aggoun, L., and Moore, J.B. (1994). *Hidden Markov Models: estimation and control.* Springer.

Ethier, R. and Mount, T. (1998). Estimating the volatility of spot prices in restructured electricity markets and the implications for option values. *Cornell University, Ithaca New York.*

Huisman, R. and Mahieu, R. (2003). Regime jumps in electricity prices. *Energy economics*, 25(5),425-434.

Jin, X. and Huang, B. (2010). Robust identification of piecewise/switching autoregressive exogenous process. *AIChE journal*, 56(7),1829-1844.

Jin, X. and Huang, B. and Shook, D.S. (2011). Multiple model LPV approach to nonlinear process identification with EM algorithm. *Journal of Process Control*, 21(1),182-193.

Jin, X. and Huang, B. (2012). Identification of switched Markov autoregressive eXogenous systems with hidden switching state. *Automatica*, 48(2),436-441.

Liu, T. (2013). Economic optimization of steam operationa. M.Sc Thesis, *Univerisity of Alberta.*

Mamon, R.S and Elliott, R.J (2007). *Hidden markov models in finance.* Springer.

Market Surveillance Administrator (2010). *Alberta wholesale electricity market.* [Online] September 2010. Available from: http://albertamsa.ca/uploads/pdf/-Reports/Reports/Alberta%20Wholesale%20Electricity-%20Market%20Report%20092910.pdf. [Accessed: 8th November 2014]

Xiong, L. (2004). Stochastic models for electricity prices in Alberta. M.Sc Thesis, *Univerisity of Calgary.*