MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Machine Learning Based State-Space Approximate Dynamic Programming Approach for Energy and Reserve Management of Power Plants

Keerthisinghe, C.; Sun, H.; Takaguchi, Y.; Nikovski, D.N.; Hashimoto, H.

TR2018-064 July 12, 2018

Abstract

This paper proposes a machine learning based state-space approximate dynamic programming (MSADP) approach to solve the self-scheduling problem faced by power plants under an integrated energy and reserve market. By extending the concept of residual demand curves (RDCs) from energy to reserve, the residual reserve curves (RRCs) is proposed to model the regulation price as a function of the power plant's reserve power. Both RRCs and RDCs are obtained using a clustering based neural network approach, which resulted in better estimates than using only a non-parametric approach. The machine learning is used to make approximations to the state space, and the dynamic programming only loops over the required states. As such, the computation effort is reduced but the solution quality does not be impacted. The value functions generated during the day-ahead optimization are used to generate optimal supply offer curves for the day-ahead market, and make real-time decisions and real-time bids to stay optimal by solving Bellman optimality condition. The effectiveness of the MSADP approach is demonstrated using empirical data obtained from the New England ISO.

IEEE PES Innovative Smart Grid Technologies Asia (ISGT Asia)

Machine Learning Based State-Space Approximate Dynamic Programming Approach for Energy and Reserve Management of Power Plants

Chanaka Keerthisinghe
Dept. of Electrical Engineering
University of Washington
Seattle, WA 98195,USA
chanakak@uw.edu

Hongbo Sun
Data Analytics Group
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139, USA
hongbo.sun@merl.com

Yusuke Takaguchi
Advanced Technology R&D Center
Mitsubishi Electric Corporation
Hyogo 661-8661, Japan
Takaguchi.Yusuke@ce.MitsubishiElectric.co.jp

Daniel Nikovski
Data Analytics Group
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139, USA
nikovski@merl.com

Hiroyuki Hashimoto
Headquarters
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139, USA
hashimoto@merl.com

***Abstract*—This paper proposes a machine learning based state-space approximate dynamic programming (MSADP) approach to solve the self-scheduling problem faced by power plants under an integrated energy and reserve market. By extending the concept of residual demand curves (RDCs) from energy to reserve, the residual reserve curves (RRCs) is proposed to model the regulation price as a function of the power plant's reserve power. Both RRCs and RDCs are obtained using a clustering based neural network approach, which resulted in better estimates than using only a non-parametric approach. The machine learning is used to make approximations to the state space, and the dynamic programming only loops over the required states. As such, the computation effort is reduced but the solution quality does not be impacted. The value functions generated during the day-ahead optimization are used to generate optimal supply offer curves for the day-ahead market, and make real-time decisions and real-time bids to stay optimal by solving Bellman optimality condition. The effectiveness of the MSADP approach is demonstrated using empirical data obtained from the New England ISO.**

Keywords—power plant self-scheduling, residual demand curves, residual reserve curves, machine learning, state-space approximate dynamic programming

I. INTRODUCTION

Under a deregulated electricity market, power plants aim to maximize their financial benefits by self-scheduling of their generation units in both the day-ahead and real-time energy and reserve markets. In this paper we focus on price makers who have the power to influence market clearing prices (MCPs). In the day-ahead electricity market, power producers aim to submit optimal stepwise supply offer curves to the independent system operator (ISO). The producer's supply offers are in pairs of price and quantity for each time-step in up to a certain number of steps that are arranged according to increasing prices [1]. Regulation offers are usually given as a single pair of price and reserve power for each hour [2]. Power producers obtain these offers by self-scheduling their generation units under uncertain MCPs, regulation clearing prices (RCPs), weather forecast and competitor offers.

To meet the needs for better coordination of energy and reserve services provided by the power producer, we has extended the concept of residual demand curves (RDCs) from energy to reserve, and use RDCs and residual reserve curves (RRCs) simultaneously to model the relationship between the powers sold and the corresponding prices in the producer's self-scheduling problem. RDCs are used to define the MCP as a monotonically non-increasing function of the producer's quota for each time-step, and obtained by subtracting the quantity offered by the competitors from the total cleared demand of the system. Similarly, we present residual reserve curves (RRCs) to model the RCP as a function of the producer's reserve power. In the literature, RDCs are usually obtained using previous days' combined competitors' offer curve and non-parametric approaches such as in [3]–[4], and reserve is assumed to be known in advance as in [5].

A clustering based artificial neural networks (ANN) approach is proposed in this paper to estimate RRCs and RDCs. The historical combined competitors' day-ahead supply offer curves are first classified into a set of clusters according to the corresponding MCP/RCP of the next day, in which the combined competitors' offer curves are obtained by adding individual competitor's supply offer curves. An ANN is then used to train a model for each cluster that maps combined competitors supply offer quotas/reserves and prices with total system cleared demand/reserve. When the system demand/reserve forecast from the ISO for next day are given, the trained non-parametric model for a chosen cluster is used to estimate the day-ahead combined supply offer curve and then the corresponding RDCs/RRCs can be generated accordingly. The cluster is chosen based on the estimated MCP/RCP. Through clustering, the proposed approach can result in better estimates than using only a non-parametric approach.

The power plant's self-scheduling and associated unit commitment problems are widely solved using mixed-integer linear programming (MILP) and dynamic programming approaches such as in [5]–[8]. Using MILP solvers, non-linear constraints, RDCs and RRCs will have to be approximated as stepwise or piece-wise linear curves, and as a consequence, the accuracy loss might not be avoidable. Dynamic programming provides a good tool for modeling non-linear constrains and all the necessary points of the RDCs and RRCs, however, it suffers from dimensionalities of state spaces.

In order to overcome the above difficulties, we model the problem as a Markov decision process (MDP) and solve using a machine learning based stat-space approximate dynamic programming (MSADP) approach. Using dynamic programming, all necessary points of RDCs and RRCs, and nonlinear constraints can be accurately modelled without linear approximation. The computational burden for dynamic programming is reduced by only looping over the desired quota and reserve states obtained using machine learning. In the day-ahead optimization, value functions defined as expected future benefit from all the possible quota, reserve and generator status combinations are generated for every time-step in the decision horizon. These value functions can help the producer to generate optimal bidding curves for the day-ahead market, and assist with real-time decisions and real-time bidding to stay optimal. At each time-step in the real-time market, the power producer can use the value functions to determine their optimal energy quota and reserve power for the next time-step given the current generation by solving Bellman optimality condition. Note that the current generation depends on the amount of energy quota that was accepted by the ISO in the day-ahead market and the amount of reserve used, and both are not predictable. Given this, the producer can adjust their prices in the real-time market and make the generation decisions to stay optimal according to Bellman's optimality condition.

The effectiveness of the proposed approach is demonstrated using empirical data obtained from the New England ISO.

II. Formulation of Self Scheduling Problem

Taken thermal plant as example, the power producer's self-scheduling problem can be formulated as a sequential decision-making problem over a given decision horizon. This problem can be described by using the input states for the producer, and the controllable states for the generation units of the plant.

For any time step k, the producer's input state variables include the MCP $s_k^{\lambda,e}$, RCP $s_k^{\lambda,r}$, forward bilateral contract price $s_k^{\lambda,f}$, required regulation reserve s_k^{r}, forward bilateral contract power s_k^{f}, and energy production quota s_k^{q} (not included forward contract power) for the power plant. RCP consists of two components, regulation capacity offer price $s_k^{\lambda,r,cap}$ and regulation service offer price $s_k^{\lambda,r,ser}$.

Among those power producer states, the forward contract power and price are given, and the MCP and RCP are functions of producer's quota and reserve powers if corresponding RDC and RRC are given, so only producer's quota and reserve powers need to be treated as independent states to be solved.

For any generation unit g, its controllable state variables at time step k include the generator up/down status $s_k^{g,s}$, power generated at the current time-step $s_k^{g,p}$, start-up command $s_k^{g,u}$ and shut-down command $s_k^{g,d}$, and generator remaining up time $s_k^{g,tu}$ and remaining down time $s_k^{g,td}$. The states of generation units are subject to its technical constraints according to its operating status and power balance required by producer's energy quota and reserve requirement.

Once a thermal generator is committed, its operation status can be categorized into four consecutive operation phases, including synchronization, soak, dispatchable, and desynchronization. The power generated during the synchronization phase is zero while the power increases linearly from synchronization/soak limit $s_k^{g,syn/soak}$ to minimum generation limit $s_k^{g,min}$ during the soak phase, and decreases linearly from $s_k^{g,min}$ to zero during desynchronization phase [9]. During the dispatchable phase, the power generation limit can range from the minimum power generation limit $s_k^{g,min}$ and maximum power generation limit $s_k^{g,max}$:

$$s_k^{g,min} \le s_k^{g,p} \le s_k^{g,max}, \tag{1}$$

according to the ramp up rate $r^{g,u}$ and ramp down rate $r^{g,d}$ as follows:

$$-r^{g,d} \le s_{k+1}^{g,p} - s_k^{g,p} \le r^{g,u}. \tag{2}$$

The network constraints for exporting the power generated to the connected power system can also be represented by modifying the power limits accordingly, if applicable. Generators can only be started when the minimum down time $\tau^{g,d}$ is satisfied:

$$s_k^{g,u}(s_k^{g,td} - \tau^{g,d}) \ge 0, \tag{3}$$

and be shut down when the minimum up time $\tau^{g,u}$ is satisfied:

$$s_k^{g,d}(s_k^{g,tu} - \tau^{g,u}) \ge 0. \tag{4}$$

The total power generated from all the generation units have to be equal to the producer's quota, forward contract power and the reserve used $\tilde{s}_k^{r}$:

$$\sum_{g=1}^{G} s_k^{g,p} = s_k^{q} + s_k^{f} + \tilde{s}_k^{r}. \tag{5}$$

where G is the total number of generation units for the plant. Meanwhile, the regulation reserve has to be met according to:

$$\sum_{g=1}^{G} \tilde{s}_k^{g} \ge s_k^{q} + s_k^{r} + s_k^{f}, \tag{6}$$

where $\tilde{s}_k^{g}$ is the maximum possible generation considering the ramp rates and the maximum power generation limit. (2)-(4) can be used to define the transition function, $\mathbf{sg}_{k+1}^{g} = \mathbf{s}^{M}(\mathbf{sg}_{k}^{g})$ when solving the self-scheduling problem using dynamic programming. The transition function describes the evolution of generator g's states from time step k to time step k+1. $\mathbf{sg}_k^{g} = \begin{bmatrix} s_k^{g,p} & s_k^{g,s} & s_k^{g,u} & s_k^{g,d} & s_k^{g,tu} & s_k^{g,td} \end{bmatrix}$. $\mathbf{s}^{M}(\cdot)$ is the system model that consists of generators' operational constraints.

Among all the generator's states, the start-up and shut-down commands and remaining up and down times can be determined if generator statuses at corresponding time-step or consecutive time steps are given. For a committed generator operated at non-dispatchable phases, its generated power is solely determined based on its corresponding up/down statuses. The power generated for generators at dispatchable phase can also be determined by simply solving an operational cost minimization problem with power balance and power generation limits when the producer's quota and reserve, generator status, and initial/terminal states are given. Therefore, only the generator up/down status combination needs to be treated as independent state variable among all generator related states.

The self-scheduling problem can be treated as a dynamic programming problem with three-dimensional state space on producer's quota, producer's reserve power and generator status combination. Its objective function is given by:

$$\mathrm{F}^{\boldsymbol{\pi}^*} = \max_{\boldsymbol{\pi}^*} \mathbb{E}\left\{\sum_{k=1}^{K} \mathrm{C}_k(\mathbf{s}_k, \boldsymbol{\pi}(\mathbf{s}_k))\right\} \quad (7)$$

where $\boldsymbol{\pi}^*$ is the optimal policy, a choice of action for for each state $\boldsymbol{\pi}: \mathbf{S} \rightarrow \mathbf{A}$ that maximizes the expected sum of future benefits over the decision horizon, $\boldsymbol{\pi}(\mathbf{s}_k)$ is an action from state $\mathbf{s}_k$, $\mathbf{S}$ and $\mathbf{A}$ are the state and action spaces, K is the total number of time steps. $\mathrm{C}_k(\mathbf{s}_k, \boldsymbol{\pi}(\mathbf{s}_k))$ is the contribution at a given time-step k, which is given by:

$$\mathrm{C}_k(\mathbf{s}_k, \boldsymbol{\pi}(\mathbf{s}_k)) = s_{k+1}^{\lambda,e} s_{k+1}^{q} + s_{k+1}^{\lambda,f} s_{k+1}^{f} + s_{k+1}^{\lambda,r,cap} s_{k+1}^{r} + \tilde{s}_{k+1}^{r} s_{k+1}^{cur} s_{k+1}^{\lambda,r,ser} - \sum_{g=1}^{G} \left[c^{fuel}\left(s_{k+1}^{g,p}\right) + c^{fixed} s_{k+1}^{g,s} + c^{start}\left(s_{k+1}^{g,td}\right) s_{k+1}^{g,u} + c^{shut}\left(s_{k+1}^{g,tu}\right) s_{k+1}^{g,d} \right] \quad (8)$$

where $\mathbf{s}_k$ consists of input states for power producer $\left[s_k^q \quad s_k^{\lambda,e} \quad s_k^r \quad s_k^{\lambda,r} \quad s_k^f \quad s_k^{\lambda,f}\right]$, and controllable states for generators $\left[\mathbf{sg}_k^1 \quad \dots \quad \mathbf{sg}_k^g \quad \dots \quad \mathbf{sg}_k^G\right]^T$. $\mathbf{s}_{k+1}$ is the state at time step k+1 resulted by following action $\boldsymbol{\pi}(\mathbf{s}_k)$ from $\mathbf{s}_k$. c^{fixed} is the generator's non-load cost, $c^{fuel}(\cdot)$, $c^{start}(\cdot)$ and $c^{shut}(\cdot)$ are the generator's fuel cost, start cost and shut-down cost respectively. s_k^{cur} is the capacity utilization ratio defined as the ratio of service offer time over length of time step. The fixed cost is a constant, and the fuel cost and start and shut-down costs are non-linear functions of generation level, remaining down time and remaining up time respectively.

III. Machine Learning Based State-Space Approximate Dynamic Programming

The producer's self-scheduling problem is solved using a machine learning based state-space approximate dynamic programming (MSADP) approach proposed in this paper.

A. Dynamic programming procedure

The problem in (7) can be cast as a Markov decision process (MDP) due to the separable objective function and Markov property of the transition functions. Given this, dynamic programming solves the MDP form of (7) by computing a value function $\mathrm{V}^{\boldsymbol{\pi}}(\mathbf{s}_k)$.

The value function is defined as the expected future benefit of following a policy, $\boldsymbol{\pi}$, starting in state, $\boldsymbol{s}_k$, and given by:

$$V^{\pi}(\boldsymbol{s}_k) = \sum_{\boldsymbol{s}' \in \boldsymbol{S}} \mathbb{P}\left(\boldsymbol{s}' | \boldsymbol{s}_k, \boldsymbol{\pi}(\boldsymbol{s}_k)\right)\left[C_k(\boldsymbol{s}_k, \boldsymbol{\pi}(\boldsymbol{s}_k), \boldsymbol{s}') + V^{\pi}(\boldsymbol{s}')\right] \quad (9)$$

where $\boldsymbol{s}'$ is the state transited from $\boldsymbol{s}_k$ through action $\boldsymbol{\pi}(\mathbf{s}_k)$, $\mathbb{P}(\boldsymbol{s}' | \boldsymbol{s}_k, \boldsymbol{\pi}(\boldsymbol{s}_k))$ is the transition probability. When the probability data is not available, or difficult to obtain, the expected value can be replaced with the maximum value, that is the right side of (9) is replaced with the maximal value of $C_k(\boldsymbol{s}_k, \boldsymbol{\pi}(\boldsymbol{s}_k), \boldsymbol{s}') + V^{\pi}(\boldsymbol{s}')$ over $\boldsymbol{s}' \in \boldsymbol{S}$.

An optimal policy, $\boldsymbol{\pi}^*$ is one that maximizes (7), and which also satisfies Bellman's optimality condition:

$$V_k^{\boldsymbol{\pi}^*}(\boldsymbol{s}_k) = \max_{\pi^*}\left(C_k(\boldsymbol{s}_k, \boldsymbol{\pi}(\boldsymbol{s}_k)) + \mathbb{E}\left\{V_{k+1}^{\boldsymbol{\pi}^*}(\boldsymbol{s}') | \boldsymbol{s}_k\right\}\right) \quad (10)$$

The expression in (10) is typically computed using backward induction, a procedure called value iteration, and then an optimal policy is extracted from the value function by selecting a maximum value action for each state.

B. Implementation of proposed approach

The proposed MSADP approach is given in Algorithm 1.

Algorithm 1: MSADP Algorithm

```
1: Load RDCs and RRCs for all time-steps.
2:  Approximate end-of-day states.
3: Initialize V_k(s_k),∀k.
4: for k = K,…, 1 do
5:    for b = 1,…, B do
6:       for y = 1,…, Y do
7:          for r = 1,…, R do
8:             Find the optimal power generation levels by solving a
               optimization problem to minimize fuel costs.
9:             Calculate fixed, start and shut-down costs.
10:            Find the instantaneous contribution using (8).
11:         end for
12:         Solve the Bellman equation (10) to calculate the expected
            future value from the state combination of current quota,
            reserve and generator statuses.
13:      end for
14:   end for
15: end for
```

The lower-case letters, *b*, *y* and *r* represent a particular quota and reserve power combination, generator status combination, and future (next time step) quota, reserve and generator status combination, respectively. Meanwhile, the upper-case letters, *B*, *Y* and *R* represent the total numbers of quota and reserve combinations, generator status combinations, and future state combinations of quota, reserve and generator statuses, respectively.

To avoid looping over all the undesired states (backward in time) that does not have a path to the desired end-of-day states, the proposed approach first approximates the end-of-day states $s_{K+1}^{g,s}$ and $s_{K+1}^{g,p}$ using either the historical data or by solving a deterministic MILP optimization over two or more days. The value functions are initialized as zeros for all time steps except the last time-step (k=K+1) where we need to penalize the undesired states by setting their values as negative infinite or provide a benefit for the desired states.

Each possible combinations of quota and reserve power in step 5 for a current time step specified in step 4 should be above the minimum power generation limit out of all the committed generators and below the total of the maximum possible power of all the committed generators. Each possible combinations of generator up/down statuses in step 6 for a current time step and quota and reserve combination specified at steps 4 and 5 should satisfy the regulation reserve, current quota and the committed forward bilateral contracts without violating any constraints.

Each possible state combination at future time step in step 7 should satisfy the minimum up and down times, and ramp up and down rates for each generator for a current time step, quota and reserve combination, and generator status specified at steps 4, 5, and 6. If no such state combination exists then penalize the current state combination. Otherwise loop over all the possible outcomes R. Given the future state *r*, the power generated from each of the generator units will be determined in step 8 to match the total power requirement for quota, reserve and forward bilateral contract power. This is a simple linear program with the objective to minimize fuel cost if the fuel cost is approximated as linear function of generated power. The power generation limits of each generator are modified to satisfy the ramp rates. Step 9 calculates the fixed cost and the start and shut-down costs using the statuses of the generators at the future state combination at next time step. Step 10 calculates the instantaneous contribution using (8).

In step 12, the expected future value from the current state combination is determined using Bellman equation (10). This is the maximum of the combined instantaneous contribution in step 10 and the expected future value for the outcome r out of all the possible outcome states R. Once we have the value functions for all the time-steps, we can move forward in time from the initial states at k = 1 to find the optimal policy. Optimal policies from different states can be used to generate the offer curves that are submitted for the day-ahead market. If the estimations are accurate enough the resulting optimal quota obtained from the value functions would be most likely get accepted by the ISO. Moreover, in real-time when the power generated at a certain time-step changes due to unforeseen reasons, the generation plant can use the value functions to stay optimal from the next time-step by making real-time bids and real-time decisions. The unforeseen reasons could be when the expected quota wasn't accepted by the ISO or when the ISO require regulation reserves.

C. Machine learning based state-space approximations

It is noted that looping over all the states in a three-dimensional state-space is computationally difficult. In order to overcome this computational burden, we use machine learning to make approximations to the state-space. The aim is to only loop over the required or desired states so the solution quality stays the same. Only the state space reduction on quota and reserve is considered here.

To approximate the quota and reserve states for the producer for a given time step, we first use historical data to learn a non-parametric model using ANN that maps MCP before producer's quota, total system cleared demand, RCP before the producer's reserve, and total system reserve requirement by the ISO (ANN inputs) with producer's quota and reserve (ANN outputs) as shown Fig. 1. If the typical values for reserve used ratio ($\text{i.e.}\,\tilde{s}_k^r/s_k^r$) that needed for determining reserve used in (5) and (8) are not available, it can also be estimated along with producer's quota and reserve by adding it to the list of ANN outputs. This non-parametric model is then used to estimate possible day-ahead quota and reserve based on the estimated next-day MCP and RCP, the reserve requirement and the demand forecast provided by ISO.

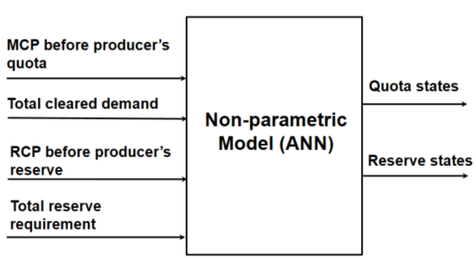

Fig. 1. Non-parametric model with inputs and outputs used to approximate the state space.

If the estimation accuracy is good enough, the estimated producer's quota and reserve values can closely match with the actual quota and reserve for the next day. The estimated quota and reserve power states are used to identify the interested state space of the self-scheduling problem, where we have a finer discretization. In this paper, ANN is a feed-forward neural network.

IV. Residual Demand Curves and Residual Reserve Curves

Conventionally, a RDC at a given time-step is obtained by subtracting the total combined competitors' offer curve from the total system cleared demand. Similarly, a RRC at a given time step is generated by subtracting the total combined reserve offer curves from the total system reserve required by the ISO.

A clustering based non-parametric approach is used in this paper to model the RDCs and RRCs. The approach is based on and tested with historical day-ahead producers' energy offer [1] and reserve offer [9] data, and their corresponding market clearing demand [10] and reserve requirements [2] from the New England ISO. Taken RDCs as example, the combined competitors' offer curves are first generated for each hour over a long period (i.e. 6 months) by adding the individual producer offers together. The maximum limit of the combined supply offer curves is set to the day-ahead cleared demand forecast from the ISO or from the demand bids. This means for a given combined supply offer curve, the maximum price is the MCP of the current day. If necessary, the combined competitors' offer curves can be separated according to the four seasons. The k-means algorithm is then used to cluster hourly combined competitors' offer curves according to the corresponding MCP. Each cluster represents a different level or range of MCPs. In the day-ahead optimization, the estimated MCP is used to identify the cluster with corresponding combined competitors' offer curves. After that, a ANN is used to train a model that maps supply offer quotas and the corresponding prices (i.e. ANN outputs) with total cleared demand (i.e. ANN inputs) in the chosen cluster. The trained model can then be used to estimate the combined supply offer curve given the day-ahead market clearing demand. RDCs are then obtained by subtracting the estimated combined competitors' offer curve from the day-ahead market clearing demand estimate. The RRCs can be estimated using the same approach as described above.

Fig. 2 gives exemplar results for modeling RDCs using the proposed approach. Fig. 2 (a) gives the distribution of historical supply offer curves at the given time-step k=20. Each point represents a historical supply offer curve, the x-axis represents its original market cleared demand, and the y-axis represents its MCP for the next day's k = 20 time-step. Fig. 2(b) shows the combined supply offers for k = 20 in the chosen cluster. Fig. 2(c) shows the errors between the actual and the estimated supply offer curves using different approaches. Fig. 2(d) shows the estimated RDCs for k = 20.

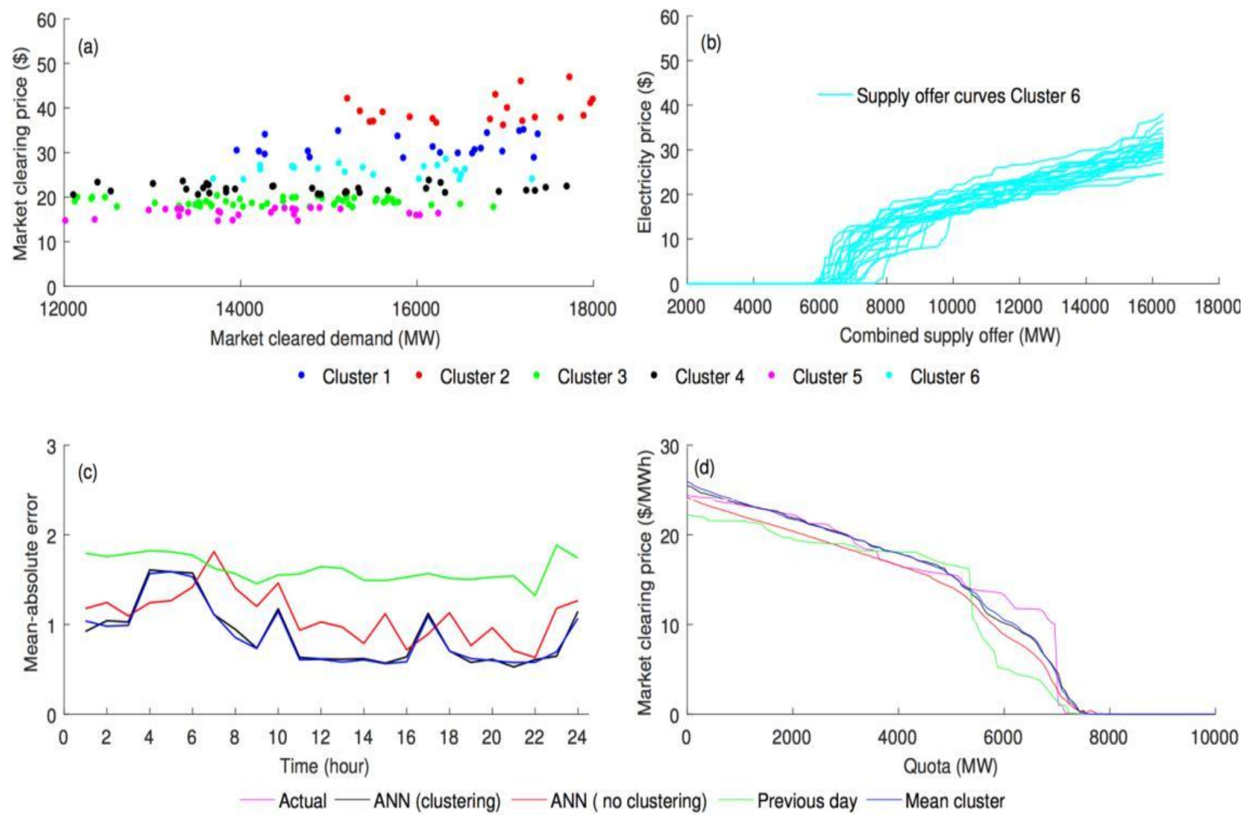

Fig. 2. Modeling RDCs: (a) day-ahead market clearing price (y-axis) from historical supply offer curves (represented by its market cleared demand in the x-axis); (b) day-ahead supply offer curves from the chosen Cluster 6; (c) mean-absolute error for the estimated combined supply offer curve over a day; (d) RDCs for k = 20.

TABLE I. THERMEL UNITS DATA

Unit	Min/ Max Power Limit [MW]	Synchro-nization/ soak limit [MW]	No load cost [$/h]	Fuel cost [$/ MWh]	Min up/ Down time [h]	Cold/ hot/warm start cost [$]	Cold/ warm start time[h]	Soak/ Synchroniza-tion/Desynch-ronization time [h]	Ramp up/ down [MW/ min]	Shut down cost [$]	Initial Power [MW]	End power [MW]
1	160/274	120	894	14	2	1350/1150/1250	4/3	1	4	0	274	170
2	180/342	140	344	15	2	400/200/285	4/3	1	4	0	313	0
3	100/378	70	167	25	1	210/100/150	3/2	1	24	0	0	0
4	150/476	100	839	10	1	350/250/300	3/2	1	24	0	476	470
5	63/152	40	965	10	2	20/10/16	4/3	1	8	0	152	0

As depicted in Fig. 2(c), the mean-absolute errors (MAE) between the actual and the estimated supply offer curves are lower for the non-parametric approach with clustering compared to the non-parametric approach without clustering. It is noted that previous day's supply offer curves determined using data from months ago has the highest MAE, and the reason for this is that historical data are only available after a certain delay (i.e. 4 months delay in New England ISO) so assuming a previous available day will result in lower quality estimates.

V. SIMULATION RESULTS

The proposed MSADP approach has been tested using empirical data obtained from the New England ISO. The techno-economic data of the thermal units are given in Table I. We use 174 days of data from January 1 to June 30, 2016 for learning and 30 days in October, 2016 for testing. The test results on an exemplar scenario on October 1, 2016 are given in this section. For the scenario, the self-scheduling problem is solved using the dynamic programming technique with proposed state-space approximation strategy. Our preliminary simulation results showed that the proposed approach can get same optimal results as MILP when the same simulation parameters are used, that is all constraints and cost functions are assumed to be linear.

The simulation results for the exemplar scenario are given in Fig. 3. Fig. 3(a) and (c) show the MCPs and RCPs before the producer's quota, and the MCPs and RCPs after the producer's quota. Fig. 3(b) and (d) show producer's quota and reserve, and system cleared demand and reserve requirement. Fig. 3(e) and (f) show RDCs and RRCs for some sample time-steps. The results show that the producer's quota/reserve is closely related to the MCPs/RCPs, slopes of the RDCs/RRCs, and the system cleared demand/reserve.

Power plant's optimal quota/reserve obtained from the day-ahead optimization as shown in Fig. 3(b) and (d) are used to create the supply offers. It is recommended for the power plant to submit their marginal cost of production for a range of

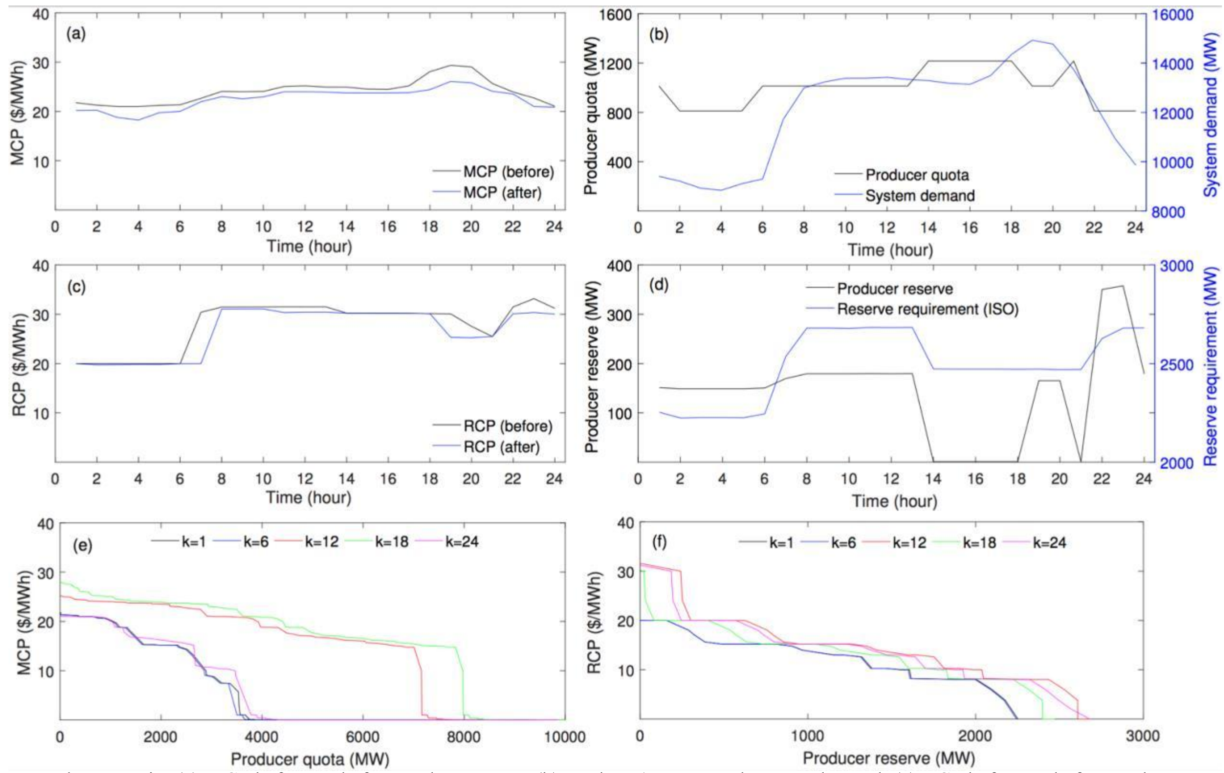

Fig. 3. Exemplar scenario: (a) MCP before and after producer quota; (b) producer's quota and system demand; (c) RCP before and after producer reserve; (d) producer's reserve and reserve requirement by the ISO; (e) RDCs; and (f) RRCs

supply offers. Given the estimations of RDCs and RRCs are accurate enough and benefits of self-scheduling, the marginal cost offer is expected to be accepted by the ISO. With the value function based decision making process, power plant can update real-time offers for upward and downward regulation reserves. For example, if the generation company is currently producing more power than their optimal then they can submit an acceptable downward reserve offer for the next hour by solving Bellman equation.

VI. Conclusions

This paper has proposed a machine learning based state-space approximate dynamic programming approach to solve the self-scheduling problem faced by power producers under an integrated energy and reserve market. Residual demand curves and residual reserve curves are estimated using a clustering based neural network approach, which resulted in better estimates than using only a non-parametric technique. The machine learning is used to make approximations to the state space, and the dynamic programming only loops over the required states. As such, the computation effort is reduced but the solution quality does not be impacted.

The value functions generated during the day-ahead optimization can be used to make optimal real-time decisions and real-time offers under market condition variations by solving Bellman's optimality condition. Producer's day-ahead offers obtained from the value functions are expected to be accepted by the independent service provider as long as all the estimations are accurate enough.

References

[1] New England ISO. (2017, March) Day-ahead energy offer data. [Online].Available: https://www.iso-ne.com/isoexpress/web/reports/pricing/tree/day-ahead-energy-offer-data

[2] New England ISO.. (2017, March) Hourly reserve requirements. [Online].Available: https://www.iso-ne.com/isoexpress/web/reports/pricing/tree/ancillary-hourly-rr

[3] J. P. Gonzlez, A. M. S. Roque, E. F. Snchez-beda, J. Garca-Gonzlez, and R. G. Hombrados, "Residual demand curves for modeling the effect of complex offering conditions on day-ahead electricity markets," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 50–61, Jan 2017.

[4] G. Aneiros, J. M. Vilar, R. Cao, and A. M. S. Roque, "Functional prediction for the residual demand in electricity spot markets," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4201–4208, Nov 2013.

[5] A. J. Conejo, F. J. Nogales, and J. M. Arroyo, "Price-taker bidding strategy under price uncertainty," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1081–1088, Nov 2002.

[6] C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis, "Optimal self-scheduling of a thermal producer in short-term electricity markets by MILP," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1965–1977, Nov 2010.

[7] H. M. I. Pousinho, J. Contreras, A. G. Bakirtzis, and J. P. S. Catalo, "Risk-constrained scheduling and offering strategies of a price-maker hydro producer under uncertainty," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1879–1887, May 2013.

[8] Walter L. Snyder H. David Powell ; John C. Rayburn, "Dynamic Programming Approach to Unit Commitment," *IEEE Transactions on Power Systems*,vol.2, no.2, pp.339-348, May 1987.

[9] New England ISO.. (2017, March) Regulation offer data. [Online]. Available: https://www.iso-ne.com/isoexpress/web/reports/pricing/tree/dayahead-regulation-offer-data

[10] New England ISO. (2017, March) Energy load and demand reports. [Online].Available: https://www.iso-ne.com/isoexpress/web/reports/loadanddemand