2010 IEEE International Conference on Control Applications
Part of 2010 IEEE Multi-Conference on Systems and Control
Yokohama, Japan, September 8-10, 2010

Reinforcement Learning Based Multi-agent LFC Design Concerning the Integration of Wind Farms

H. Bevrani, *Senior Member*, *IEEE*, F. Daneshfar, *Student Member*, *IEEE*,
P. R. Daneshmand, *Student Member*, *IEEE*, and T. Hiyama, *Senior Member*, *IEEE*

Abstract—Frequency regulation in interconnected networks is one of the main challenges posed by wind turbines in modern power systems. The wind power fluctuation negatively contributes to the power imbalance and frequency deviation. This paper presents an intelligent agent based load frequency control (LFC) for a multi-area power system in the presence of a high penetration of wind farms, using multi-agent reinforcement learning (MARL). Nonlinear time-domain simulations on a 39-bus test power system are used to demonstrate the capability of the proposed control scheme.

Keywords—Load-frequency control, Reinforcement learning, Multi-agent systems, Wind power generator.

I. Introduction

USING conventional linear control methodologies for the LFC design in a modern power system is not more efficient, because they are only suitable for a specific operating point in a traditional structure. If the dynamic/structure of system varies; they may not perform as expected. Most of conventional control strategies provide model based controllers that are highly dependent to the specific models, and are not useable for large-scale power systems concerning the integration of RES units with nonlinearities, undefined parameters and uncertain models. If the dimensions of the power system increase, then these control design may become more difficult as the number of the state variables also increases, significantly.

Therefore, design of intelligent controllers that are more adaptive and flexible than conventional controllers is become an appealing approach. Intelligent control has been already used for the frequency regulation issue in the power systems [1-5]; however there are just few reports on the intelligent frequency control design in the presence of RES units [6].

One of the adaptive and nonlinear intelligent control techniques that can be effectively applicable in the frequency control design is reinforcement learning (RL). Some efforts are addressed in [7-9]. The RL based controllers learn and are adjusted to keep the area control error small enough in each sampling time of a LFC cycle. Since, these controllers are based on learning methods; they are independent of environment conditions and can learn a wide range of operating conditions. The RL based frequency control design is a model-free design and can easily scalable for large scale systems and suitable for frequency variation caused by wind turbine fluctuation.

The present paper addresses the LFC design using an agent based reinforcement learning for a large interconnected power system concerning the integration of wind power units. In this paper, a multi-agent RL based control structure is proposed. Each control area includes an agent that communicates with each other to control the frequency among whole interconnected system. Each agent (controller agent) provides an appropriate control action according to the area control error (*ACE)* signal, using reinforcement learning. In a multi-area power system, the learning process is considered as a multi-agent RL process and agents of all areas learn together (not individually).

The above technique has been applied to the LFC problem in a network with the same topology as IEEE 10 generators 39-bus test system integrated with wind power units, as a case study.

II. Test System

As mentioned, the wind power generation could affect the dynamic behavior of the power system. The frequency response characteristic of a power system with a high penetration of wind power may be different from that of the conventional system. The impact of wind power generation on the power system frequency response is discussed in [6, 10, 11].

Here, to illustrate the effectiveness of the proposed control strategy, and to compare the results with well-tuned PI controllers, the IEEE 10 generators, 39-bus system is considered as a test case study.

This test system is widely used as a standard system for testing of new power system analysis and control synthesis methodologies. A single-line diagram of the system is given in Fig. 1. This system has 10 generators, 19 loads, 34 transmission lines, and 12 transformers. Here, the test system is updated by two wind farms in areas 1 and 3. The 39 buses system is organized into 3 areas. Total system installed capacity are 841.2 MW of conventional generation and 45.34 MW of wind power generation. There are 198.96 MW of conventional generation, 22.67 MW of wind power generation and 265.25 MW load in Area 1. In Area 2, there are 232.83

H. Bevrani, F. Daneshfar and P. R. Daneshmand are with the Dept. of Electrical and Computer Eng., University of Kurdistan, Sanandaj, Iran. Currently, H. Bevrani is a visiting professor at Kumamoto University, Japan (Corresponding author e-mail: bevrani@ieee.org).

T. Hiyama is with Dept. of Electrical Eng., Kumamoto University, Kumamoto, Japan (e-mail: hiyama@cs.kumamoto-u.ac.jp).

MW of conventional generation, and 232.83 MW load. In Area 3, there are 160.05 MW of conventional generation, 22.67 MW of wind power generation and 124.78 MW of load.

The simulation parameters for the generators, loads, lines, and transformers of the test system are given in [11]. All power plants in the power system are equipped with speed governor and power system stabilizer (PSS). However, only one generator in each area is responsible for the LFC task; G1 in Area 1, G9 in Area 2, and G4 in Area 3. For the sake of simulation, random variations of wind velocity have been considered. Dynamics of WTGs including the pitch angle control of the blades are also considered. The start up and rated wind velocity for the wind farms are specified as about 8.16 (m/s) and 14 (m/s), respectively. Furthermore, the pitch angle controls for the wind blades are activated only beyond the rated wind velocity. The pitch angles are fixed to zero degree at the lower wind velocity below the rated one.

III. The Proposed Intelligent LFC Strategy

In practice, the LFC system is traditionally using a proportional-integral (PI) type controller [6, 14]. In this section, an intelligent control design algorithm for such a controller using MARL technique is presented. The design objective is to regulate the frequency in power system concerning the integration of wind power units with various load disturbances.

Fig. 2 shows the overall diagram of the proposed multi-agent control structure for the 3-control area power system example. Each control area includes an intelligent controller. The controller is responsible to produce an appropriate control action (ΔP_{Ci}) according to the measured area control error (*ACE*) signal and tie-line power changes ($\Delta P_{tie\text{-}i}$) using RL.

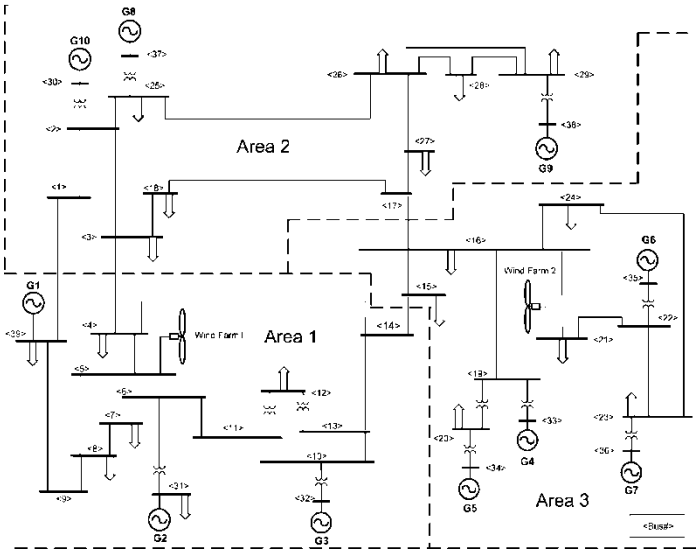

Fig. 1 Single-line diagram of 39-bus test system

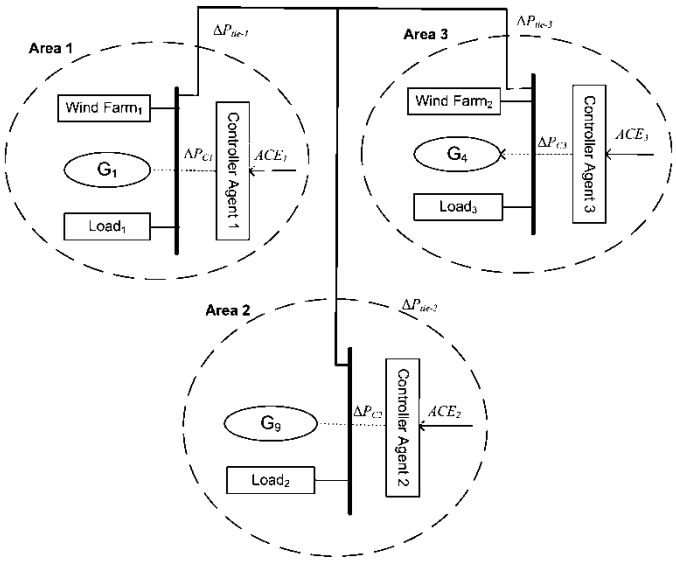

Fig. 2 The overall control framework for the 3- control area power system example

A. Controller Agent

The intelligent controller system (controller agent) functions as follows: At each instant (on a discrete time scale *k*) *k=1*, *2*, …, the controller agent observes the current state of the system, x_k , and takes an action, a_k . The state vector consists of some quantities, which are normally available to the controller agent. Here, the average of *ACE* signal over the time interval *k-1* to *k* as the state vector at the instant *k* is used. For the algorithm presented in this paper, it is assumed that the set of all possible states *X*, is finite. Therefore the values of various quantities that constitute the state information should be quantized.

The possible actions of the controller agent are the various values of ΔP_C , that can be demanded in the generation level within an LFC interval. ΔP_C is also discretised to some finite number of levels. Now, since both X and A are finite sets, a model for this dynamic system can be specified through a set of probabilities.

B. Multi-agent RL

In most RL methods, instead of calculating the state value, another term known as the action value is calculated (1), which is defined as the expected discounted reward while starting at state x_t and taking action a_t .

$$Q^{\pi}(x,a) = E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | x_t = x, a_t = a\right\} \quad (1)$$

Bellman's equation [12], as shown below, is used to find the optimal action value. In general, an optimal policy is one that maximizes the Q-function defined in the following relation [9]:

$$Q^*(x,a) = \max_{\pi} E_{\pi}\left\{r_{t+1} + \gamma \max_{\acute{a}} Q^*(x_{t+1},\acute{a}) \mid x_t = x, a_t = a\right\} \quad (2)$$

C. Developed Algorithm

Different RL methods have been proposed to solve the above equations. Here, an RL algorithm is used for estimating Q^* and the optimal policy. It is similar to the introduced algorithm in [8]. Suppose we have a sequence of samples (x_k, x_{k+1}, a_k, r), $k = 1,2,\cdots$. Each sample is such that x_{k+1} is the (random) state that resulted when action a_k is performed in state x_k and $r_k = g(x_k, x_{k+1}, a_k)$ is the consequent immediate reinforcement. Such a sequence of samples can be obtained either through a simulation model of the system or observing the actual system in operation. This sequence of samples (called training set) can be used to estimate Q^*, using a specific algorithm. Suppose Q^k is the estimate of Q^* at kth iteration. Let the next sample be (x_k, x_{k+1}, a_k, r) then we obtain Q^{k+1} as:

$$\begin{aligned} Q^{k+1}(x_k,a_k) = {} & Q^k(x_k,a_k) \\ & + \alpha\left[g(x_k,x_{k+1},a_k)\right. \\ & \left. + \gamma \max_{a\in A} Q^k(x_{k+1},\acute{a}) - Q^k(x_k,a_k)\right] \end{aligned} \quad (3)$$

where $0 < a < 1$ is a constant called the step size of learning algorithm.

At each time step (as determined by the sampling time for the LFC action) the state input vector x, to the LFC is determined, then an action in that state is selected and applied to the model, the model is integrated for a time interval equal to the sampling time of LFC to obtain the state vector $\acute{x}$ at the next time step.

Here, the exploration policy for choosing actions in different states is used. It is based on a Learning automata algorithm called pursuit algorithm [13]. This is a stochastic policy where, for each state x, actions are chosen based on a probability distribution over the action space. Let P_x^k denote the probability distribution over the action set for state vector x at the kth iteration of learning. That is, $P_x^k(a)$ is the probability of choosing action a in state x at iteration k. A uniform probability distribution is considered at $k=0$, that is

$$P_x^0(a) = \frac{1}{|A|} \quad \forall a \in A \ \forall x \in X \quad (4)$$

At the kth iteration, let the state x_k be equal to x. An action a_k, at random based on $P_x^k(.)$ is chosen. That is, $Prob(a_k = a) = P_x^k(a)$. Using the performed simulation model, the system is gone to the next state x_{k+1} by applying action a in the state x and is integrated for the next time interval. Then, Q^k is updated to Q^{k+1} using (3) and the probabilities is updated as follows.

$$\begin{aligned} P_x^{k+1}(a_g) &= P_x^k(a_g) + \beta\left(1 - P_x^k(a_g)\right) \\ P_x^{k+1}(a) &= P_x^k(a)(1-\beta) \quad \forall a \in A, a \neq a_g \\ P_y^{k+1}(a) &= P_y^k(a) \quad \forall a \in A, \forall y \in X, y \neq x \end{aligned} \quad (5)$$

where $0 < \beta < 1$ is a constant. Thus, at iteration k the probability of choosing the greedy action a_g in state x is slightly increased and the probabilities of choosing all other actions in state x are proportionally decreased.

In the present algorithm, the aim is to achieve the well-known LFC objective and to keep the *ACE* within a small band around zero. This choice is motivated by the fact that all the existing LFC implementations use this as main control objective and hence, it will be possible for us to compare the proposed RL approach with the designed linear PI based LFC approaches.

As mentioned above, in this formulation, each state vector consists of the average value of *ACE* as state variable. The control action of the LFC is to change the generation set point, ΔP_C. According to the RL algorithms application, usually a finite number of states are assumed. In this direction, state variable and action variable should be discretised to finite levels, too.

The next step is to choose an immediate reinforcement function by defining the function g. The reward matrix initially is full of zero, at each time step we get the average value of *ACE* signal, then according to its discritised values, determine the state of the system, whenever the state is desirable (i.e. $|ACE|$ is less than ε) then reward function $g(x_k, x_{k+1}, a_k)$ is assigned at zero value. When it is undesirable (i.e. $|ACE| > \varepsilon$), then $g(x_k, x_{k+1}, a_k)$ is assigned a value -$|ACE|$ (we penalized all actions which cause to go to an undesirable state with a negative value).

IV. Application to the 3-Control Area Test System

To illustrate the effectiveness of the proposed control strategy, the proposed intelligent control design is applied to the described 39-bus system (Fig. 1).

Here, the purpose is essentially to clearly show the various steps in implementation and illustrate the method. After design choices are made, the controller is trained by running the

simulation in the learning mode as explained in Section 3. After completing the learning phase, the control actions at various states have converged to their optimal values.

The simulation is run as follows: At each LFC instant *k*, controller agents of each area, average all corresponding *ACE* signal instances gained every 0.1 seconds. Three average values of *ACE* signal instances (each related to one area) form the current state vector, x_k, that is obtained according to the quantized states. When all area's state vectors are ready, then the controller agents choose the action signal a_k that consists of three ΔP_C values for three areas (action signal is gained according to the quantized actions and the exploration policy mentioned above) to change the set points of the governors using the values given by a_k.

In the performed simulation studies, the input variable is obtained as follows. As the LFC decision cycle time chosen, three values of *ACE* are calculated over a decision cycle. The averages of these values for three areas are the state variable $(x^1_{avg1}, x^1_{avg2}, x^1_{avg3})$.

Since, we use the multi-agent reinforcement learning process and agents of all areas are learning together, the state vector is also consisted of all state vectors of three areas, the action vector is consisted of all action vectors of three areas as shown in term $<(X_1, X_2, X_3),(A_1, A_2, A_3), p, (r_1, r_2, r_3)>$ or $<X, A, p, r>$.

Here $X_i = x^1_{avgi}$ is the discrete set of each area states, *X* is the joint state, A_i is the discrete set of each area actions available to the area *i*, and *A* is the joint action. In each instant time after averaging of ACE_i for each area (over three instances), depending on the current joint state (X_1, X_2, X_3) the joint action $(\Delta P_{C1}, \Delta P_{C2}, \Delta P_{C3})$ is chosen according to the exploration policy.

Consequently, the reward *r* is also dependent on the joint action which whenever the next state (*X*) is desirable (i.e. all $|ACE_i|$ are less than ε), then reward function *r* is assigned a zero value. When the next state is undesirable (i.e. $\exists ACE_i|$, $|ACE_i| > \varepsilon$) then *r* is assigned average value of -$|ACE_i|$ In this algorithm, since all agents learn together, parallel computation causes to speed up the learning process. Also this reinforcement learning algorithm is more scalable than single-agent RL.

V. Simulation Results

To demonstrate the effectiveness of the proposed control design, some simulations were carried out. In these simulations, the proposed controllers were applied to the 3-control areas model described in Fig. 1. In this section, the performance of the closed-loop system using the well-tuned conventional PI controllers is compared to the designed MARL controllers for the various possible load disturbances.

As a serious test scenario, the following load disturbances (step increase in demand) are applied to three areas: In Area 1, 3.8% of total area load at bus 8, 4.3% of total area load at bus 3 in Area 2, and 6.4% of total area load at bus 16 in Area 3 have been simultaneously increased in a step form.

The applied step load disturbances ΔP_{Li}(pu), the output power of wind farms P_{WT} (MW), and the wind velocity V_W (m/s) are shown in Fig. 3. The frequency deviation (Δf), and area control error (*ACE*) signals Area 2 and Area 3 are shown in Fig. 4, and Fig. 5. The produced mechanical power by the LFC participant unit in Area 2, corresponding electrical power, and also the overall tie line power for the same area are shown in Fig. 6.

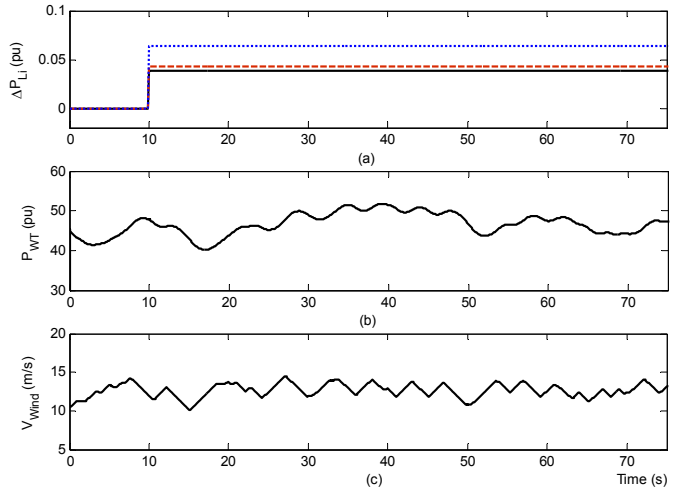

Fig. 3. a) Load step disturbances in three areas, b) Total Wind power, and c) The wind velocity pattern in Area 1

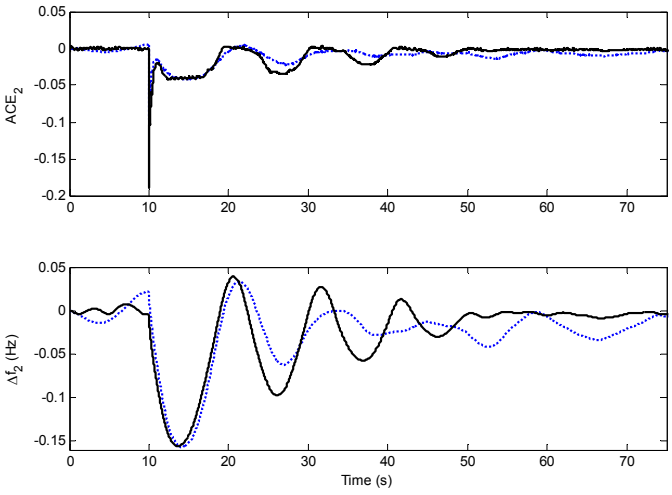

Fig. 4 Area-2 responses; proposed intelligent method (solid), linear PI control (dotted)

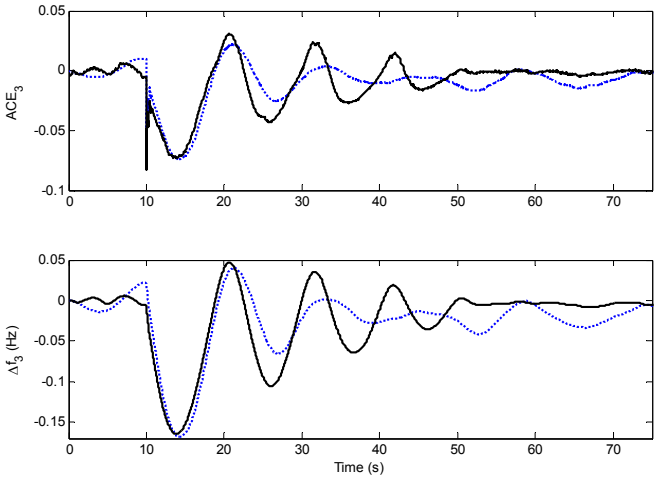

Fig. 5 Area-3 responses; proposed intelligent method (solid), linear PI control (dotted)

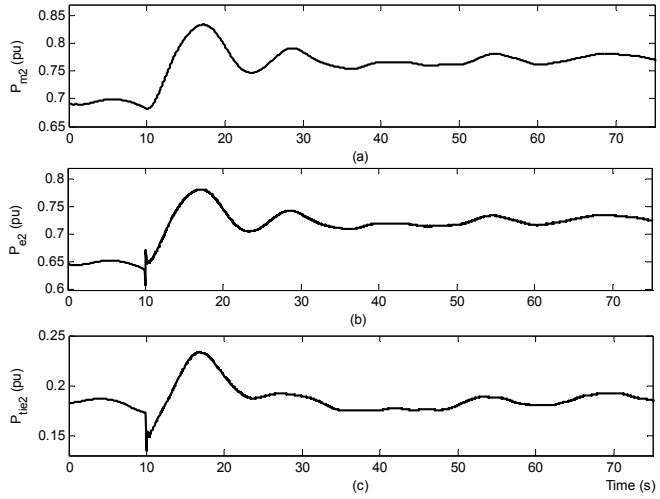

Fig. 6. Area-2 power response using the proposed MARL method

The wind penetration in this system is considered as two individual wind farms each with capacity equivalent to about half of the total penetration. However in the present simulation, the detailed dynamic nonlinear models of 39-bus power system and wind turbines are used without applying an aggregation model for generators or wind turbine units. That is why in the simulation results, in addition to the long term fluctuations, the faster dynamics on a time scale of 10 seconds are also observable [6, 11].

As shown in the simulation results, using the proposed method, the area control error and frequency deviation of all areas are properly driven close to zero. Furthermore, regarding that the proposed algorithm is an adaptive algorithm and it is based on the learning methods - in each state it finds the local optimum solution to gain the system objectives (*ACE* signal near zero) - therefore the intelligent controllers provide smoother control action signals and areas frequency deviation is less than the frequency deviation in the system with PI controllers.

VI. Conclusion

A new method for frequency regulation concerning the integration of wind power units, using MARL has been proposed. The proposed method was applied to a network with the same topology, known as New England 10-generators 39-bus system. The results show that the new algorithm performs well, in comparison of the performance of a PI control design. Two important features of new approach, model independence and flexibility in specifying the control objective; make it very attractive for application in power system operation and control. However, the scalability of MARL to realistic problem sizes is one of the great reasons to use it. In addition to scalability and benefits owing to the distributed nature of the multi-agent solution, such as parallel computation, multiple RL agents may utilize new benefits from sharing experience, e.g., by communication, teaching, or imitation.

References

[1] H. Bevrani, *et. al*, "Load-frequency regulation under a bilateral LFC scheme using flexible neural networks," *Engineering Intelligent Systems*, vol. 14, no. 2, pp. 109-117, 2006.

[2] A. Demiroren, H. L. Zeynelgil, and N. S. Sengor, "Automatic generation control for power system with SMES by using neural network controller," *Electr. Power Comp System*, vol. 31, pp. 1-25, 2003.

[3] Y. L. Karnavas, D. P. Papadopoulos, "AGC for autonomous power system using combined intelligent techniques," *Electric Power Systems Research*, vol. 62, pp. 225-239, 2002.

[4] T. Hiyama, D. Zuo, T. Funabashi, "Multi-agent based automatic generation control of isolated stand alone power system," In Proceeding of International Conference on Power System Technology, 2002.

[5] T. Hiyama, D. Zuo, T. Funabashi, "Multi-agent based control and operation of distribution system with dispersed power sources," In Proceeding of IEEE/PES Transmission and Distribution Conference and Exhibition-Asia Pacific, 2002.

[6] H. Bevrani, *Robust power system frequency control*. Springer Press, NY, 2009.

[7] F. Daneshfar, H. Bevrani, "Load-frequency control: a GA-based multi-agent reinforcement learning," *IET Gener. Transm. Distrib.*, vol. 4, no. 1, pp. 13-26, 2010.

[8] T. P. I Ahamed, P. S. N. Rao, P. S. Sastry, "Reinforcement learning controllers for automatic generation control in power systems having reheat units with GRC and dead-band," International journal of power and energy systems, vol. 26, pp. 137-146, 2006.

[9] S. Eftekharnejad, A. Feliachi, "Stability enhancement through reinforcement learning: load frequency control case study," *Bulk Power System Dynamics and Control* VII, pp. 1-8, 2007.

[10] H. Bevrani, *et al.*, "On the use of df/dt in power system emergency control" In *Proc. of IEEE Power Systems Conference & Exposition*, Seattle, Washington, USA, 2009.

[11] H. Bevrani, F. Daneshfar, P. R. Daneshmand, "Intelligent power system frequency regulation concerning the integration of wind power units", Book chapter of *Wind Power Systems: Applications of Computational Intelligence*, L. F. Wang, C. Singh, and A. Kusiak (Eds), Springer Book Series on Green Energy and Technology, Springer-Verlag, Heidelberg, expected in 2010.

[12] R. S. Sutton, A. G. Barto, *Reinforcement learning: an introduction*. MA: MIT Press, Cambridge, 1998.

[13] M. A. L. Thathachar, B. R. Harita BR, "An estimator algorithm for learning automata with changing number of actions," *International Journal of General Systems*, vol. 14, pp. 169 – 184, 1998.

[14] H. Bevrani, T. Hiyama, "On Load-Frequency Regulation with Time Delays: Design and Real Time Implementation," *IEEE Transactions on Energy Conversion*, vol. 24, no. 1. pp. 292-300, 2009.