Machine Learning Based Preventive and Corrective Control to Prevent Cascading Failures Following Large Disturbances

Final Project Report

S-94G

Power Systems Engineering Research Center

Empowering Minds to Engineer
the Future Electric Energy System

Machine Learning Based Preventive and Corrective Control to Prevent Cascading Failures Following Large Disturbances

Final Project Report

Project Team
Vijay Vittal, Project Leader
Arizona State University

Graduate Students
Yuling Wang
Arizona State University

PSERC Publication 22-01

September 2022

For information about this project, contact

Vijay Vittal
Arizona State University
School of Electrical, Computer, and Energy Engineering
P.O. BOX 875706
Tempe, AZ 85287-5706
Phone: 480-965-1879
Fax: 480-727-2052
Email: Vijay.Vittal@asu.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Notice Concerning Copyright Material

Acknowledgments

This is the final report for the Power Systems Engineering Research Center (PSERC) research project titled “Explore the Efficacy of Machine Learning Based Preventive and Corrective Control to Prevent Cascading Failures Following Large Disturbances” (Project S-94G).

We would like to thank and acknowledge the technical feedback and helpful suggestions provided by the project advisor, especially: Dr. Di Shi.

Executive Summary

Power systems are becoming more complex due to the high penetration of renewable energy and integration of distributed energy resources. The introduction of electric vehicles, also increases the uncertainty due to the nature of the device characteristics and operation pattern decided by the user behavior increasing the complexity of system operation, especially emergency control. Reinforcement learning (RL) techniques now have been widely explored and applied to various electric power operation analyses under different control structures. With massive synchronized data from phasor measurement units (PMU), it is possible to explore the application of RL to ensure that electricity is delivered reliably.

Steady state stability control with RL has been studied by many researchers, while the system dynamic stability and performance after serious disturbances using RL has not been examined. However, the current control algorithms utilized in the power system dynamic control primarily utilize a discrete action space, the performance of which is not satisfactory when dealing with a parameter adjustment problem within a certain range. Hence, this project focuses on the continuous reinforcement learning methods to deal with system dynamic control problems after a disturbance.

This research applies the deep deterministic policy gradient (DDPG) approach integrated with detailed time-domain simulations to develop generator controls to prevent cascading voltage failures following large disturbances. The study in this report utilizes detailed time-domain simulations interfaced with reinforcement learning to determine appropriate generator control actions for voltage control. The DDPG controller regulates the magnitude of the voltage reference value of the generator's excitation system to achieve generator output regulation and bus voltage restoration in the vicinity of the disturbance area. In the DDPG algorithm, two hidden layers with rectified units of ReLU and Tanh are applied to the actor network and two hidden layers with ReLU are applied to the critic network to realize policy updates. The reward structure consists of accumulated voltage variations and generator regulation values in considering the system cost to realize voltage regulation. Historical data of voltage and voltage rate of change are also discussed when designing the reward function, to achieve a better voltage restoration level and dynamic performance.

The dynamic simulation and training platform is built based on the power system simulation software Siemens-PTI PSS/E and Python. The power grid operation environment is created through the API available in PSS/E to simulate dynamic changes on the power grid side. The IEEE 9-bus system and 2000-bus Texas synthetic grid system are used as the test systems, based on which time-domain simulations are conducted and interfaced with the reinforcement learning controller. The results show that the controller can provide accurate commands to the excitation system in real time based on the system operating conditions. The process of voltage recovery is stable and fast under the control of the DDPG controller, without which the system voltage would degrade slowly and enhance the risk of losing stability when only the conventional generator control is used.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Research motivation and objectives

Power system resilience and reliability are vital to the economic viability of society. The US-Canada power system outage on August 14, 2003 cost 10 billion US dollars [1]. More and more essential services, such as electrical transportation, rely on electricity, so it is of great importance to guarantee power system stability and dynamic performance. Power systems are becoming increasingly complex with renewable energy sources integration, which introduces significant uncertainty due to their natural characteristics when interfaced with power systems through power electronics converters. So advance control techniques are needed to ensure that electricity is transmitted and delivered reliably.

In the power grid, phasor measurement units (PMUs) which work as communication and measurement devices make it possible to transfer synchronized dynamic data across power systems. Based on this massive data, online stability prediction and corrective control can be applied using Reinforcement Learning (RL) techniques which have now matured and are being applied to various power system applications. Steady-state stability control utilizing reinforcement learning has been studied by many researchers, while the area of dynamic stability which needs detailed information exchange between agents and the operating power grid environment requires further development coordinating effective preventive and corrective controls combined with RL.

This study applies RL methods in conjunction with detailed time domain simulations to develop generator controls to prevent cascading voltage failures following large disturbances. The study utilizes detailed time-domain simulations interfaced with reinforcement learning to determine appropriate generator control actions. The focus of this study is on the dynamic process of power system operation, so the simulation and training platform is built based on power system simulation software Siemens-PTI PSSE 35.2 and Python with which systematic algorithmic training can be achieved during the dynamic simulation process.

1.2 Related works

Reinforcement learning (RL) has been increasingly applied in power systems to solve control related problems, which involve power system protection control (such as the utilization of DRL based methods to control the protection relay logic [2]), photo-voltaic and wind resources control (such as the combination of Maximum Power Point Tracking (MPPT) control with Q-learning to generate switching signals [3]), load-frequency control [4-9] or voltage control [1].

Voltage stability has been one of the most important control problems for power system operation. RL control was considered in [10] and [11] for subsystem voltage control, which utilized a Q-learning algorithm to learn a reactive power optimal control scheme to keep the voltage within the normal range. Q-learning was also adopted in [12] for optimal tap setting of on-load tap changer of step-down transformers (connecting electric distribution systems with the rest of the system) to control the distribution system side voltages under uncertain load dynamics. Q-learning in [13] was used to provide a control scheme of active power generations to prevent system cascading failure, and the controller operates in the system normal state and takes actions in the form of preventive control to make adjustments in case of cascading failure when the system suffers large disturbances. Reference [14] proposed a two time-scale voltage control scheme, including fast inverter control and switching of shunt capacitors at a slower time control based on the Deep Q-Network (DQN) algorithm. Reference [15] applied DQN and Deep Deterministic Policy Gradient (DDPG) for subsystem voltage control and found that DDPG performed better with sufficient training scenarios. Reference [16] adopted multi-agent deep deterministic policy gradient (MADDPG), which is a multi-agent continuous actor-critic based algorithm, to realize voltage regulation among substations based on power flow data. It, however, focused on the steady-state performance of the system.

There have been many explorations and attempts in the area of steady-state voltage control based on reinforcement learning [16], [17]. However, [18] and [19] addressed transient stability issues to keep the system in synchronism by controlling power system components, such as thyristor-controlled series capacitors and dynamic braking resistors. Reference [18] considered the historical states of the power system to recover partially observable problems that have the Markov property. References [19-20] also considered the transient angle instability problems. Reference [19] proposed a system-centric controller and observer based on hybrid RL method. Reference [20] proposed a control scheme to damp local and inter-area oscillations considering local prioritization. Reference [21] utilized dynamic braking for power system emergency control based on Deep Reinforcement Learning methods, an open simulation platform was built as well to develop, train, and benchmark RL algorithms for power system control problems.

This study aims to solve the voltage control problem based on the continuous reinforcement learning algorithm in the power system environment and will focus on the dynamic simulation process to consider the dynamic factors that may influence power system operation. The voltage control action in the power grid is achieved by adjusting the generators' excitation system under large load changes. The DDPG algorithm, which deals with the continuous action space is considered and implemented in this study to continuously control the voltage reference of the generator excitation system. Since we focus on the dynamic process, a dynamic power system training and simulation platform is built based on a commercial power system software package and Python. By utilizing the reinforcement learning algorithm, this study will finally build a controller that can help the generator flexibly change its output within the specified limits to satisfy the needs of the system and provide voltage support when

disturbances or large changes occur.

1.3 Report organization

The report is organized as follows: Chapter 1 presents the motivation and objectives of the research. Chapter 2 introduces the relevant theoretical background, which includes neural networks and reinforcement learning algorithm. In Chapter 3, the proposed reinforcement learning based voltage control method is explained in detail. The design of different reward functions is discussed, and neural network structures are introduced. The power system dynamic training and simulation platform are discussed in this chapter as well. Chapter 4 introduces the IEEE 9-bus test system, the 2000-bus Texas synthetic grid model, and the design of parameters, and the simulations are demonstrated with results analyzed. Conclusions and future work are presented in Chapter 5.

2. The reinforcement learning method

This chapter introduces the theoretical background of the basic structure of deep neural networks (DNN) and deep deterministic policy gradient (DDPG) algorithm.

2.1 Deep Neural Networks

Deep neural networks (DNN) provide the platform to incorporate large amounts of information and are applied widely in non-linear problems. A DNN typically learns the functional relationship between the inputs and outputs during the training process. A DNN includes an input layer, multiple hidden layers and an output layer.

If the output of each layer is linear, its modeling capability will be limited when dealing with complex problems. Therefore, activation functions are introduced to account for non-linear factors. There are different types of activation functions, such as sigmoid, tanh and rectified linear unit (ReLU).
The sigmoid function can output only positive values, and is described in (2-1) [22]

$$f(z) = \frac{1}{1+e^{-z}} \tag{2-1}$$

The output value range of the tanh function is (-1, 1), and when the input is 0, the output is 0 as well. The tanh function [22] is defined as

$$f(z) = \frac{2}{1+e^{-2z}} - 1 \tag{2-2}$$

The ReLU function is commonly used as it is resilient to the vanishing gradient problem. Equation (2-3) shows the ReLU activation function [23]:

$$f(z) = \max(0, z) \tag{2-3}$$

2.2 Deep Deterministic Policy Gradient Algorithm

The Deep Deterministic Policy Gradient (DDPG) algorithm adopts an actor-critic structure with a continuous actions space [24]. The Bellman equation as shown in (2-4) is used to update the critic value:

$$Q_{j+1}^{(s,a)} = Q_j^{(s,a)} + \alpha\left[R_j + \gamma \max Q_j^{(s',a')} - Q_j^{(s,a)}\right] \tag{2-4}$$

where α is the learning rate and γ is the discount rate, R_j represents the reward of each training

step *j*. Due to the limited computation capacity, Q-learning and DQN are not suitable for large action spaces or continuous action spaces since every single action is selected based on a matched Q-value in the above two algorithms [24]. In DDPG, a parameterized policy function is used to choose a deterministic action according to system states. The actor is updated by applying the chain rule to the expected reward with respect to the actor parameters as

$$\nabla_{\theta\mu} J = \frac{1}{N} \sum \nabla_a Q(s,a)\Big|_{s=s_j, a=\mu(s_j)} \nabla_{\theta\mu} \mu(s\big|\theta^{\mu})\Big|_{s=s_j} \tag{2-5}$$

where *J* is the starting distribution, which represents the initial gradient ascent of the actor network parameters to the *Q* value, $\mu(s\big|\theta^{\mu})$ is the parameterized policy function, and θ^{μ} refers to parameters of the policy network. The control action is obtained from a deep neural network, which makes it possible to deal with a continuous action space in practical large-scale systems. DDPG is an off-policy method which uses off-policy data and the Bellman equation to learn the *Q*-function and uses the *Q*-function to learn the action policy. The network used to update the *Q* value and calculate the target value will sometimes lead to instability of the process. DDPG utilizes a copied actor network $\mu'(s\big|\theta^{\mu'})$ and a critic network $Q'(s,a\big|\theta^{Q'})$ to calculate the target values, respectively, and it has a total of four networks to estimate the policy and value functions: actor, target-actor, critic, and target-critic. The main actor and critic networks will update every step that parameters update, and the target networks will update and track the learned network slowly, which are soft updates: $\theta' \leftarrow \rho\theta + (1-\rho)\theta'$, where ρ is a hyperparameter whose value lies between 0 and 1, usually close to 0 to constrain the speed of updating. The use of targets networks significantly improves the stability of learning. During the action exploration, a random decaying noise is added into the policy as shown below in (2-6)

$$\mu'(s_j) = \mu(s_j\big|\theta_j^{\mu}) + \xi_j \tag{2-6}$$

where $\xi_j + 1 = r_d * \xi_j$ and r_d is the decay rate.

The DDPG algorithm is detailed in Algorithm 1 [24] as below.

Algorithm 1: DDPG algorithm for power system voltage control

input : system voltage states

output: generator excitation system voltage reference value

Initialize the critic network Q and actor network μ with random weights θ and ϕ.;

Initialize the critic network Q' and actor network μ' with random weights $\theta' \leftarrow \theta$ and $\phi' \leftarrow \phi$.;

Initialize the experience replay buffer D.;

for *episode 1 to* M, **do**

  Initialize the power system environment and obtain initial state S_0;

  Initialize a random process N for action exploration;

  for *step 1 to* T, **do**

    Select action $a_t = \mu(s_t|\theta + N_t)$ according to the current policy and exploration noise;

    Execute action a_t and observe reward r_t, and observe next state s_{t+1} ;

    Store transition (s_t, a_t, r_t, s_{t+1}) in D;

    Sample a random minibatch of B transition (s_j, a_j, r_j, s_{j+1}) from D;

    Compute the critic target:

$$y_j = R_j + \gamma Q'(s_{j+1}, \mu'(s_{j+1}|\theta^{\mu'})|\theta^{'})$$

    Update the critic Q-function by one step of gradient descent using:

$$\text{L=1/N} \sum_j (y_j - Q(s_j, (a_j|\theta^Q))^2$$

    Update the target networks as:

$$\nabla_{\theta^\mu} J = \frac{1}{N} \sum \nabla_a Q(s,a)|_{s=s_j, a=\mu_{(s_j)}} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s=s_j}$$

    Update the network parameters:

$$\theta^{'} \leftarrow \rho\theta + (1-\rho)\theta^{'}$$

$$\phi^{'} \leftarrow \rho\phi + (1-\rho)\phi^{'}$$

3. Reinforcement learning based voltage control

This chapter introduces the system model to which the DDPG algorithm is applied. The voltage control problem is defined and formulated as a Markov decision process (MDP). The structure of the neural networks adopted is shown. The power system dynamic simulation and training platform are introduced in detail as well.

3.1 Power System Model

The dynamic process of power grid operation possesses a highly non-linear characteristic, which is essentially a process of sequential decision-making under-uncertainty. The power system model can be formulated as follows [21]:

$$\mathbf{P}: \min \int_{T_0}^{T_c} C(\mathbf{x}_t, \mathbf{y}_t, \mathbf{s}_t) dt \tag{3-1}$$

subject to

$$\dot{\mathbf{x}}_t = f(\mathbf{x}_t, \mathbf{y}_t, d_t, \mathbf{a}_t) \tag{3-2}$$

$$0 = g(\mathbf{x}_t, \mathbf{y}_t, d_t, \mathbf{a}_t) \tag{3-3}$$

$$\mathbf{x}_t^{\min} \le \mathbf{x}_t \le \mathbf{x}_t^{\max}, \forall t \in [T_0, T_c] \tag{3-4}$$

$$\mathbf{y}_t^{\min} \le \mathbf{y}_t \le \mathbf{y}_t^{\max}, \forall t \in [T_0, T_c] \tag{3-5}$$

$$\mathbf{a}_t^{\min} \le \mathbf{a}_t \le \mathbf{a}_t^{\max}, \forall t \in [T_0, T_c] \tag{3-6}$$

where $\mathbf{x}_t$ denotes dynamic state variables in the power system; $\mathbf{y}_t$ represents the algebraic states in the power system, such as voltage of the buses of the power grid; $\mathbf{a}_t$ is the control action of the power system, such as generator regulation; d_t represents the system disturbance or fault that occurs during system operation; T_0 and T_c represent the time horizon of this dynamic process.

Equation (3-1) represents minimizing the total cost of the corrective control, including the cost of control actions, the control effectiveness in terms of system states (sometimes the control effectiveness can be reflected by system states). Equation (3-2) describes the dynamic system model, such as the behavior of generators and the relevant control systems. Equation (3-3) represents the power system constraints that describe the power balance between generators, loads and transmission branches. Equation (3-4)-(3-6) are the operational constraints of the system dynamic states, algebraic states, and control actions. Equations (3-1) - (3-6) together describe the optimal decision-making model during power system operation [21] [25].

3.2 Markov Decision process Formulation for the Voltage Control Problem

3.2.1 Definition of Action, State and Observation

For voltage control under generator regulation, the control actions are defined as a vector of generator reference voltage magnitudes. By continuously adjusting this parameter of the generator excitation system, the output reactive power will be regulated so as to provide appropriate voltage support to the system under disturbance. Different measurements obtained by system measurement devices are usually used as the system states according to the problem that needs to be dealt with. Similar to [26] and [27], the reinforcement learning method used in this study only considers the bus voltage magnitude as the state in the Markov decision process. So, the observation of the controller is the bus voltage magnitude. The power system environment state transition is realized by a set of differential algebraic equations of the form (3-2) and (3-3). The limits on the controller defined in (3-6) are considered in the definition of the action space by setting appropriate control bounds.

3.2.2 Definition of Reward

- **Consider voltage magnitude deviation and regulation cost**

  The reward function r_t is designed to evaluate the control effectiveness of the actions during the implementation of reinforcement learning. To restore the voltage level under the control, the reward is designed to motivate the controller to reduce the deviation of the bus voltage magnitude from the bus reference value V_{ref} . As shown in (3-7), if the system diverges during operation under control actions, a distinguishable negative reward will be given. Otherwise, with less bus voltage deviation, the reward will become larger based on the first term of (3-7) in the case of system convergence, which will guide the controller to regulate its action to reach more desirable states. Meanwhile, the objective is to reach the desired goal with less control effort, so the second term considers the cost during all the control processes, which try to help the system restore the voltage with less regulation of the generator excitation system. Variables c_1 and c_2 are the weights of these two components and they are chosen based on expert knowledge of the system as well as trial-and-error selection [21]. The definition of $\Delta V(t)$ and $\Delta a(t)$ can be seen in (3-8) - (3-9).

$$r_t = \begin{cases} Huge\ \ penalty,\ \ system\ \ tends\ \ to\ \ diverge \\ -c_1 * \sum_i \Delta V_i(t) - c_2 * \sum_j \Delta a_j(t), \quad otherwise \end{cases} \tag{3-7}$$

$$\Delta V_i(t) = \left| V_i(t) - V_{ref} \right| \tag{3-8}$$

$$\Delta a_j(t) = \left| a_j(t) - 1 \right| \tag{3-9}$$

- **Consider voltage magnitude deviation, regulation cost and historical voltage data**

  Power systems possess significant inertia and the dynamic process is a sequential process, which means the current state of the system is affected by both the previous control actions as well as the previous states. Significant information lies in the massive historical state data within a given simulation, in our case this historical information is provided by the bus voltage magnitude. Therefore, the historical voltage magnitude data is input into the agent to help the DDPG agent learn a more accurate policy to cope with system disturbances. The reward function considering the historical data is formulated as (3-10):

$$r_t = \begin{cases} Huge \quad penalty, \quad system \quad runs \quad to \quad diverge \\ -c_1 * \sum_i \Delta V_i(t) - c_2 * \sum_j \Delta a_j(t) - c_3 * \sum_{t-c_t}^{t} \sum_i \Delta V_{history-i}(t), \quad otherwise \end{cases} \tag{3-10}$$

$$\Delta V_{history-k}(t) = \left| V_{history-k}(t) - V_{ref} \right| \tag{3-11}$$

  where $\Delta V_{history-i}$ is the historical voltage magnitude difference of bus *i* with bus reference value $V_{ref} = 1pu$, c_t is the historical time range considered in a certain past time instant during system operation, and c_3 is the weight related to the historical data in the reward function.

- **Consider voltage magnitude deviation, regulation cost, historical voltage data and voltage rate of change**

  During the system dynamic evolution and control implementation after a disturbance or large change in load, the dynamic performance after the control action is implemented is also of significant importance. The objective is to avoid system oscillations and voltage fluctuations, to facilitate the system voltage recovery in a more stable fashion. Both the rates of voltage changes and their historical values are considered in the reward function to guide the agent to generate a control policy that will aid in the recovery of the system voltage with a more desirable dynamic performance. The reward function considering both voltage historical data and voltage rate of change is shown as (3-12):

$$r_t = \begin{cases} Huge \quad penalty, \quad system \quad begins \quad to \quad diverge \\ -c_1 * \sum_i \Delta V_i(t) - c_2 * \sum_j \Delta a_j(t) - c_3 * \sum_{t-c_t}^{t} \sum_i \Delta V_{history-i}(t) - c_4 * \sum_{t-c_t}^{t-\Delta t} \sum_i \dfrac{V_{history-i}(t) - V_{history-i}(t-\Delta t)}{\Delta t}, \quad otherwise \end{cases} \tag{3-12}$$

  where c_4 is the weight related to the rate of voltage change in the reward function, Δt is the time interval of every learning step during the training process, when applied to a practical power system, Δt should be the data sampling time step of the measurement device.

3.3 Neural Network Architecture

The neural network structures adopted for the DDPG algorithm in this study are shown in Fig 3.1 and Fig 3.2. Both actor and critic neural networks have two hidden layers, which are connected with activation functions. The actor neural networks adopt Relu and Tanh activation functions and critic networks adopt Relu as the activation function.

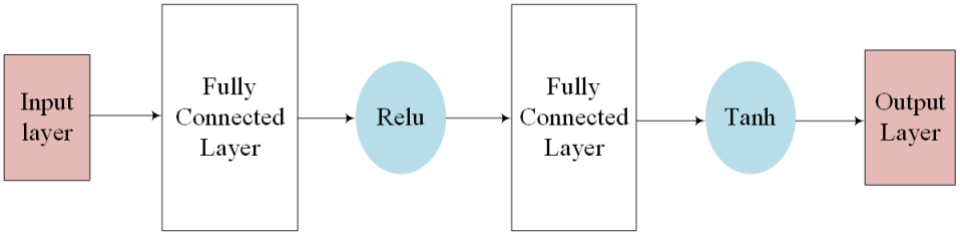

Figure 3.1 Actor neural network structure

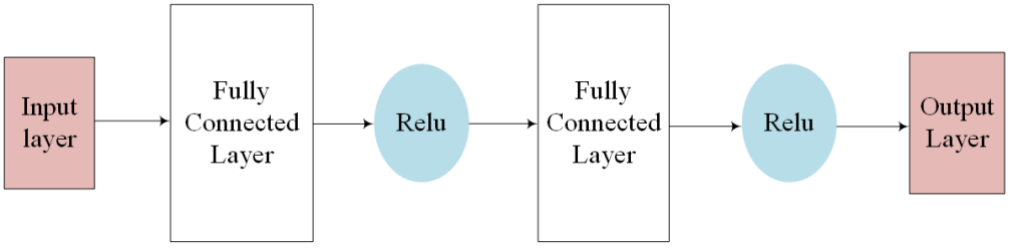

Figure 3.2 Critic neural network structure

3.4 Simulation Platform Development

A power system dynamic simulation platform has been built for the implementation of the reinforcement learning algorithm in the power system dynamic simulation environment. The structure of the whole system is shown in Fig. 3.3.

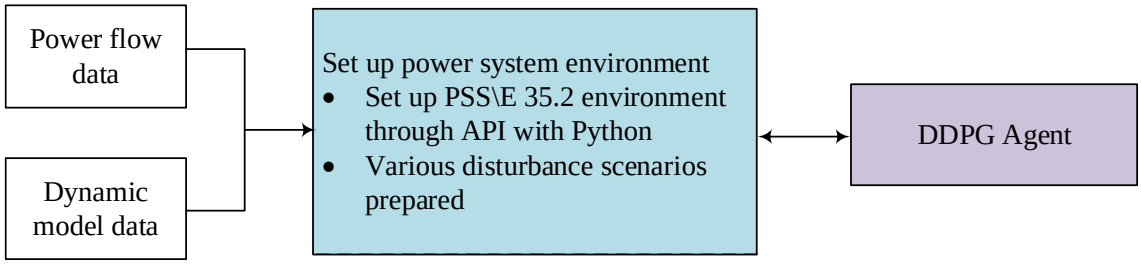

Figure 3.3 Diagram of the DDPG-based agent and power system environment

The time-domain simulation software Siemens-PTI PSS/E is used as the power system simulator to conduct power system dynamic simulations and emulate the power grid environment. PSS/E provides APIs based on an open Python technology, which can communicate with the DDPG agent in real time to exchange information during the training process.

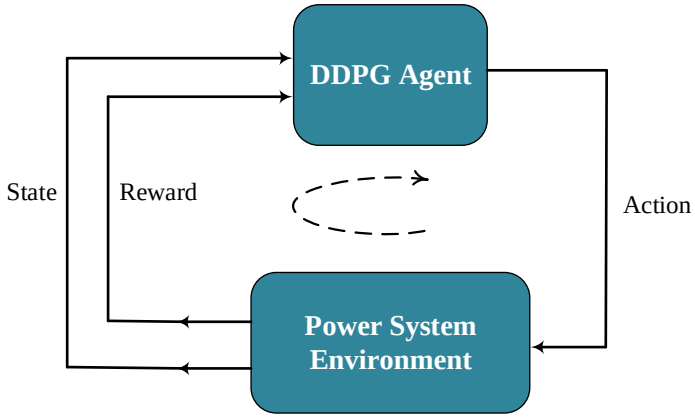

Figure 3.4 Diagram of the data interaction between DDPG agent and power system environment

The disturbance is randomly introduced into the environment during each training episode, in which one round of dynamic simulation will begin. Each episode contains certain steps, as shown in Fig. 3.4. The DDPG agent will generate actions in each step during training according to the current policy $a = \mu(s|\theta^{\mu})$ that the agent learned. In the case considered, it is a reference command that the controlled generator should follow. The action will be sent to the power system environment through the API between PSS/E and python. The simulator (PSS/E) on the power system side will receive this command and incorporate it into the simulation. The training step interval for the power system simulation is set as 1 second. After the action is executed, the dynamic simulation will run for 1 second, and the state of the power system environment will change because of this action, generating the next state at the end of simulation. The power system simulator will output the system states through the API to the DDPG agent, based on which the reward will then be calculated. The current state, next state, current action, and reward will be added into the replay buffer. After the buffer is filled, a new set of observations will be randomly selected for network updating. Then the DDPG agent will generate the new action based on the latest version of the network, another step of training starts and the dynamic simulation will move to the next second until this episode is over. Then another disturbance scenario will be generated, and another round of training begins until the final episode is examined.

This simulation platform is based on power system dynamic simulation (based on power flow data and dynamic data) and can be used for training under different scenarios based on different algorithms. The power system operation information will be exchanged with the agent at each step during system operation which can reflect system dynamic characteristics during control combined with RL algorithms.

4. Results for test systems

The IEEE 9-bus system and the 2000-bus Texas synthetic grid systems are used as the test systems, based on which time domain simulations are conducted and interfaced with the reinforcement learning controller.

4.1 Description of the test system

4.1.1 IEEE 9-bus System

Figure 4.1 shows the network structure of the IEEE 9-bus system [28], which includes 3 generators and 9 buses. The system parameters are shown in Table 4.1. Generator 1, a hydraulic turbine, is set as the generator connected to the slack bus 1. Generators 2 and 3 are steam turbines. Loads are located at buses 5, 6, and 8.

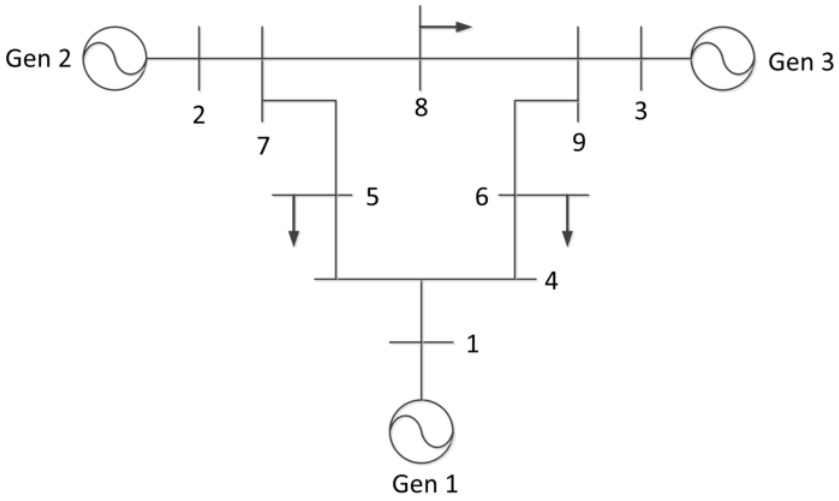

Figure 4.1 IEEE 9-bus test system

Table 4.1 Parameters of the IEEE 9-bus system

Bus Number	Voltage (kV)	Generator Output (MVA)	Load (MVA)
1	16.5	247.5	/
2	18	192	/
3	13.8	128	/
4	230	/	/
5	230	/	125+j50
6	230	/	90+j30
7	230	/	/
8	230	/	100+j35
9	230	/	/

4.1.2 Texas 2000-Bus Synthetic Power System

The Texas 2000-bus synthetic power system is a set of test data built according to public information and statistical analysis of the real power system [29] - [31]. The generation and load distributions are similar to the actual power grid. The whole 2000-bus power test system is synthetic and has four voltage levels of 500/230/161/115 kV. The total generation capacity in this system is 98GW with a load of 67GW and 19GVAr. The heavily loaded area is in southeast Texas around the Houston area, and the Northern part of the Texas grid, which can be seen from the system structure shown in Figure 4.2.

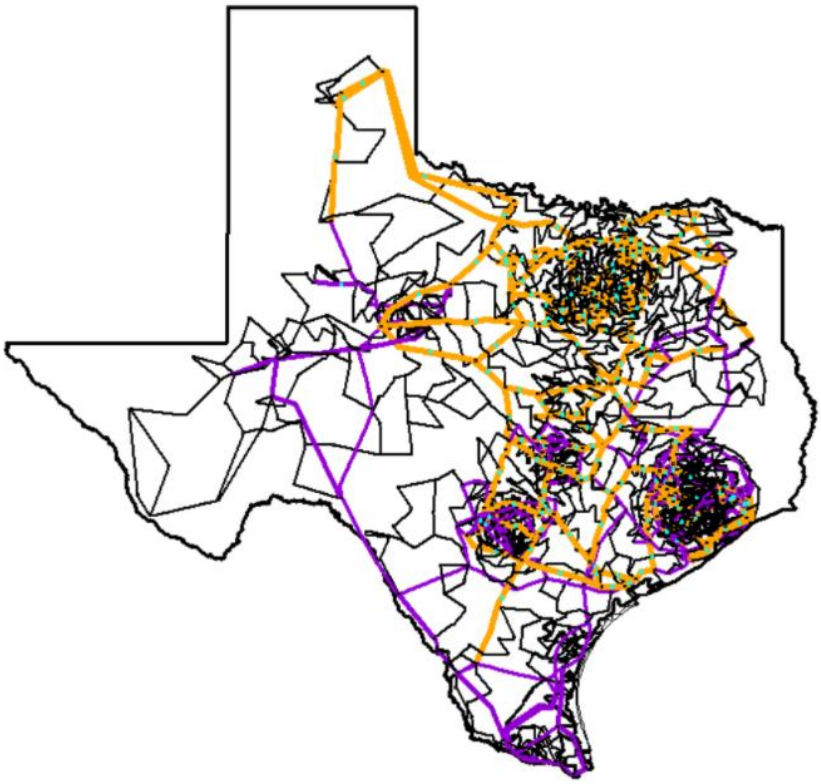

Figure 4.2 Texas 2000-bus synthetic power system

Dynamic simulations are conducted based on this data and an oscillation is found in this system when no disturbance is added, as seen in Fig 4.3. This indicates that a suitable initial condition was not determined for the time domain simulation and as a result some system parameters or control settings can be erroneous.

To remedy this problem and make preparations for the agent training, the appropriate parameter(s) of the test system is adjusted. The test case was simulated in post PSS/E and PSLF which have excellent initial condition analysis capability for dynamic initialization. According to the warning information that PSS/E and PSLF provided, we adjusted the maximum limit of the governor and the excitation

system, and this reduces the oscillation. Generator 6216 greatly influences system stability. The lead and lag time constant and "Vrmax" of generator 6216's excitation system were further adjusted to appropriate values. As a result, the initialization is successful and a flat run indicating this is obtained, as shown in Fig 4.4. The following training and testing simulations are based on this corrected data.

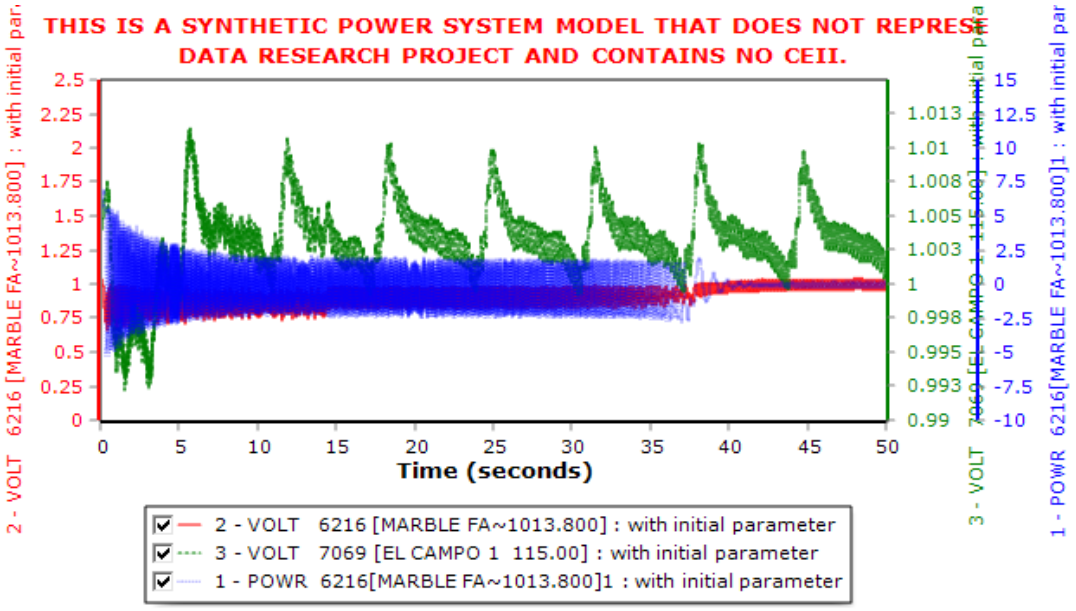

Figure 4.3 System oscillation without any disturbance

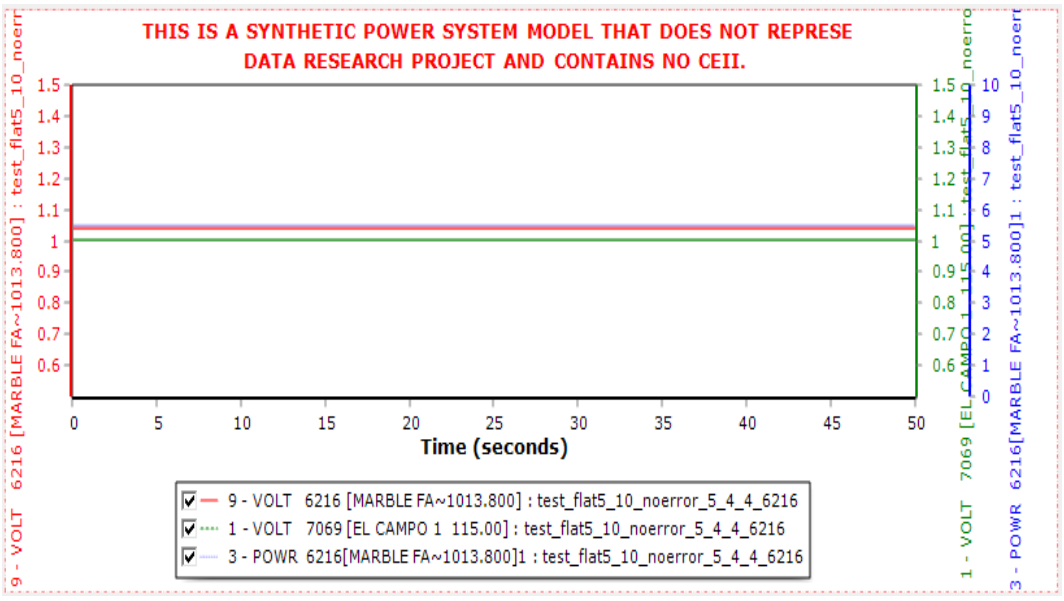

Figure 4.4 Flat voltage and power without any disturbance

4.2 Simulation Parameters

The training parameters are crucial to the convergence of the algorithm. Generally, as shown in Table 4.2, the hyperparameters include learning rate for both actor and critic networks, the discount rate γ,

batch size, memory capacity, and training step. Exploration noise, which is used to enrich the training exploration, is also an important parameter that influences the learning performance.

The learning rate determines the learning speed of the agent. The larger the learning rate is, the faster the agent will learn, but this could also lead to oscillations and result in a loss of the optimal solution. A smaller learning rate can make the learning process more precise but the drawback is the learning speed is slower. The process could be easily trapped into an overfitting situation. Generally, the learning rate is set to the range of 0.01-0.001. We used a learning rate of 0.001 due to the good performance during training.

Table 4.2 Training Parameters

Hyper parameters	Parameter values
Layer	2, 2 (actor, critic)
Activation Function	([ReLU, Tanh], [ReLU, ReLU])
Units of MLP per Layer	32
Learning Rate Actor	0.001
Learning Rate Critic	0.001
Discount rate γ	0.9
Batch Size	128
Memory Capacity	10000
Max Step	50
Exploration Noise	3

The discount rate essentially determines how much the reinforcement learning agent cares about rewards in the distant future relative to those in the immediate future. If it is set as 0, the agent will only learn about actions that produce an immediate reward, if it is set as 1, the agent will evaluate every action based on the total sum of all the future rewards. Most actions do not have long-lasting repercussions and need to be traded off to avoid irrelevant information. The discount rate is set as 0.9 in the training conducted and it provides satisfactory results.

The batch size indicates the number of training examples utilized in one iteration. Since the number of states and actions are not large in our case, the batch size is set as 128. The memory capacity is the capacity of the datasets that the agent will randomly sample from to train the network. It is used to break the correlation between different data to avoid inefficient learning. A value of 10,000 memory capacity is suitable for a 128-batch size.

The step interval is set as 1 second so the maximum step setting as 50 means every round of dynamic simulation will last 50 seconds after the disturbance is introduced. A 5 seconds initialization time is set so the total simulation will be 55 seconds for each episode.

Exploration noise is used to introduce more explorations during the training process which will enrich the data set during the information exchange with the power system environment. The exploration noise is set by trial and error and a final value of 3 is obtained.

4.3 Simulations on the IEEE 9-bus System

Case 1- Considering voltage magnitude deviation and regulation cost

The system structure can be seen in Fig 4.5. Various amounts of load changes are randomly introduced at bus 8 in the system at 5 seconds which results in around 3% - 5% voltage fluctuation.

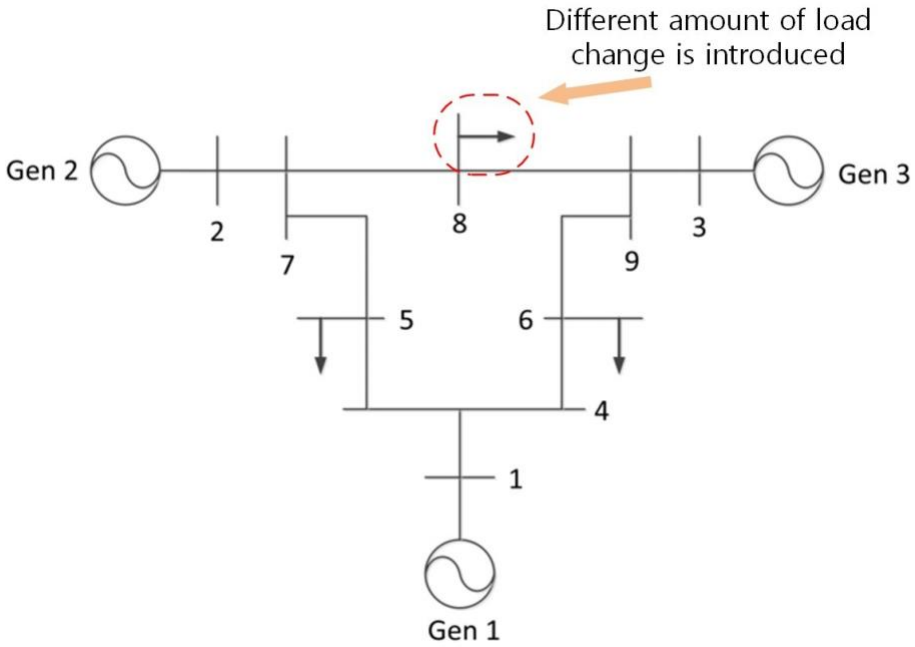

Figure 4.5 IEEE9 system structure

The power grid's state is the voltage magnitudes of the buses in the disturbed area, and two generators, generators 2 and 3, participate simultaneously in the voltage regulation as the controlling action. The action bound is set as [-1.3, 1.3]. The step interval is set as 1 second, which means the power grid environment will exchange information with the DDPG agent, send back current states and get action commands every 1 second. With the reward function of (3-7) that considers controlled bus voltage magnitude deviation and generator regulation cost, the DDPG controller learns in the training process under the above fault scenarios, the moving average reward finally reaches a satisfactory level after 2000 episodes of training, as shown in Fig 4.6.

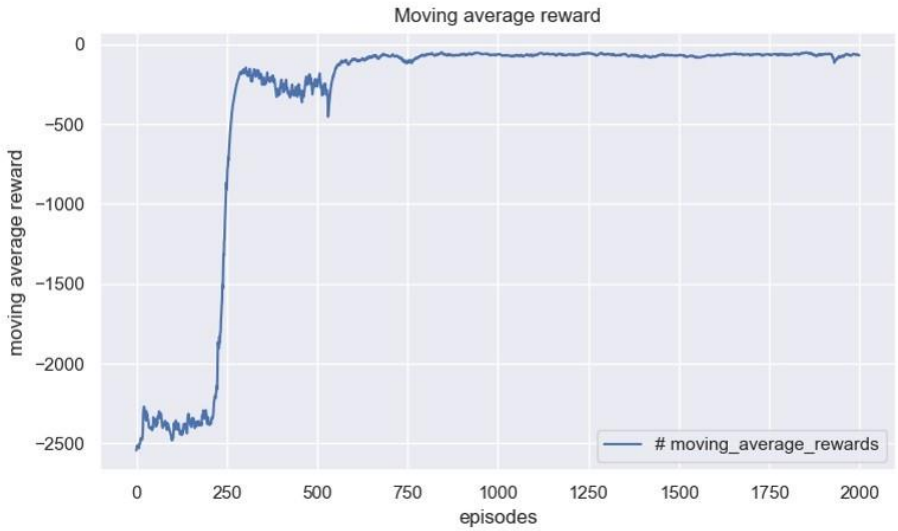

Figure 4.6 Moving average reward of IEEE9 system

After being fully trained, the DDPG agent is applied to the system to provide voltage support after system disturbance. The test result can be seen in Fig. 4.7 and Fig. 4.8. With only conventional generator control, the system bus voltage magnitudes are seriously affected and keep decreasing after the load change disturbances, which puts the system at high risk of losing stability. With the DDPG agent controlling generators 2 and 3, the system voltage can be significantly improved to the normal level. The excitation system voltage reference value of the two controlled generators can be seen in Fig. 4.9 and Fig. 4.10. Under the control of the DDPG agent, generator 3 will provide full voltage support after detecting the disturbance and the generator 2 is responsible for the voltage regulation according to the system operating condition. The two generators cooperate under the control of the DDPG agent to help the system restore voltage.

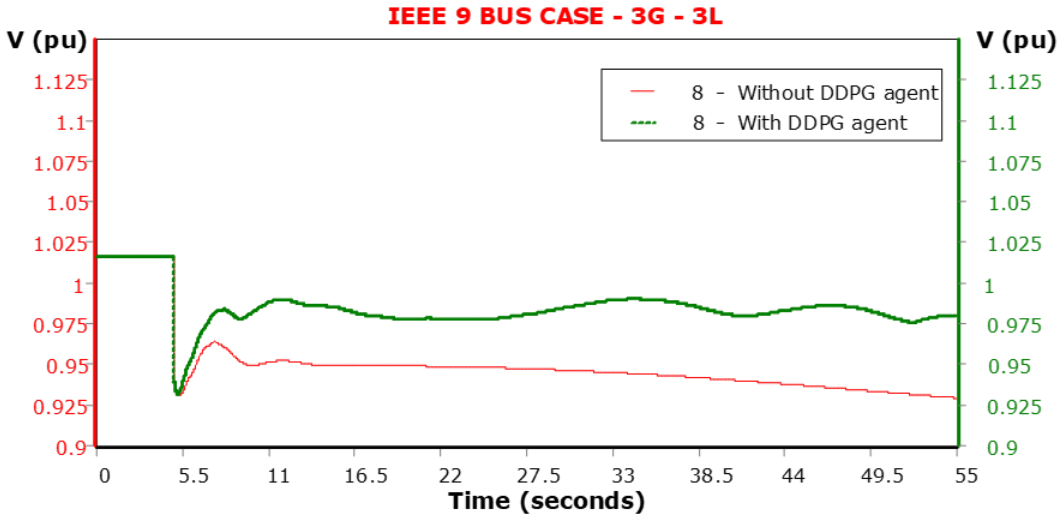

Figure 4.7 Case 1-Voltage of bus 8 with and without DDPG controller

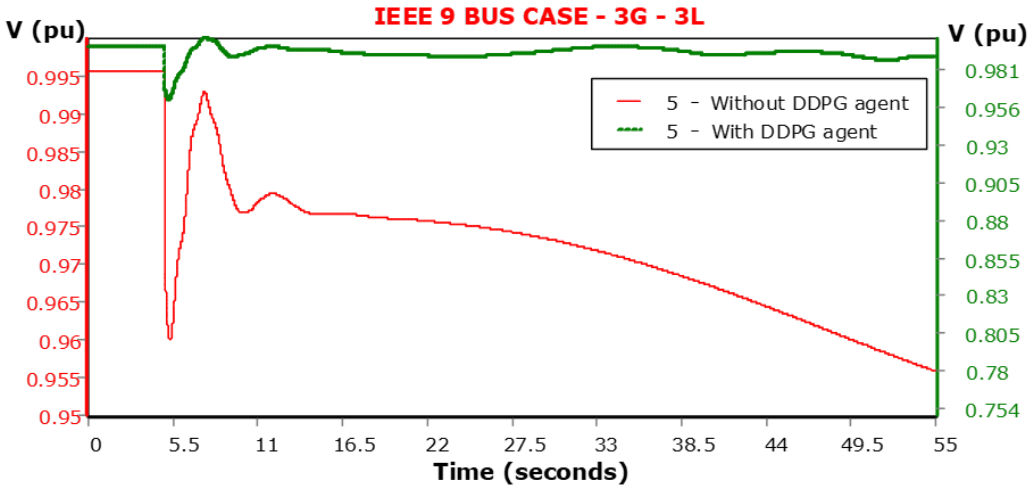

Figure 4.8 Case 1-Voltage of bus 5 with and without DDPG controller

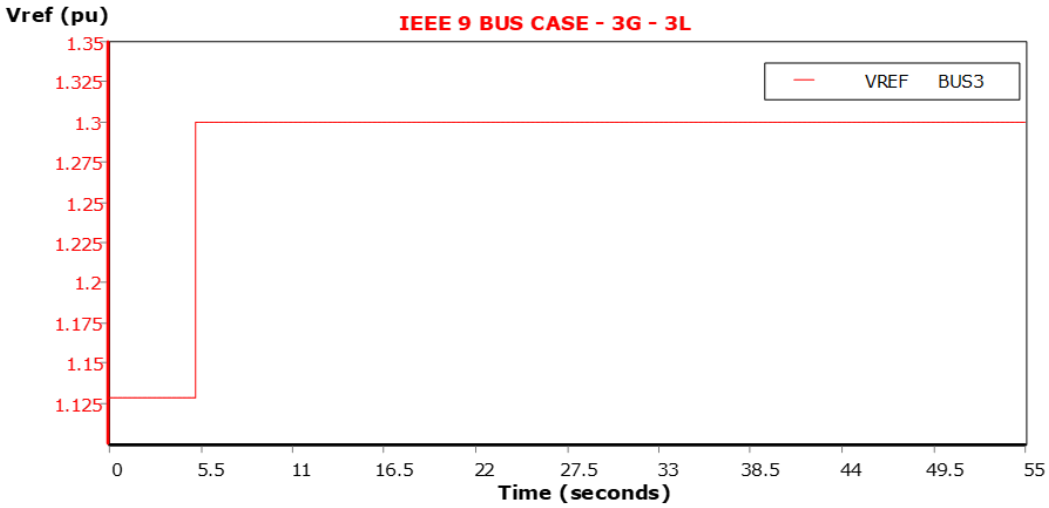

Figure 4.9 Case 1-Generator 3 voltage reference command of the excitation system

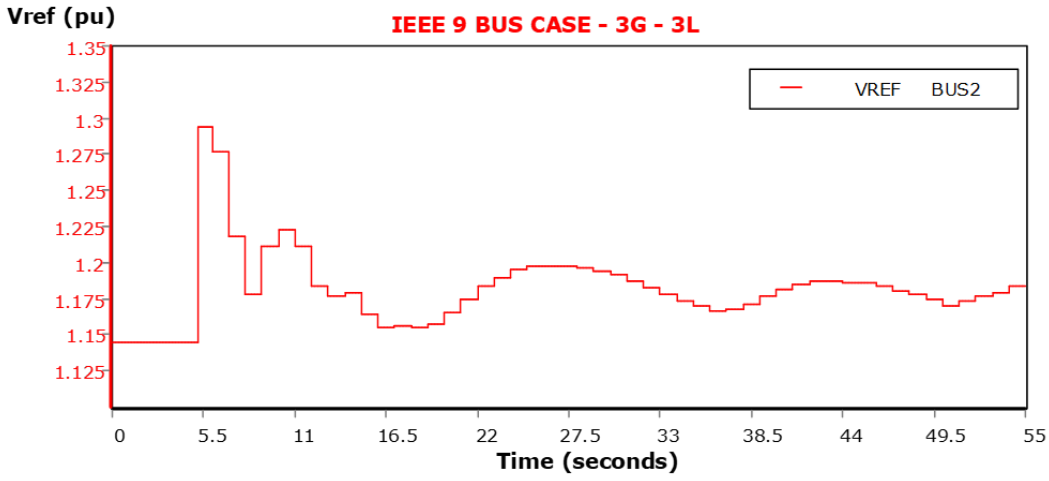

Figure 4.10 Case 1-Generator 2 voltage reference command of the excitation system

Case 2-Consider voltage deviation, regulation cost and historical voltage data

To further analyze the impact of historical data on the agent control performance, the DDPG agent is trained with the reward function of (3-10) that considers the historical voltage data besides bus voltage magnitude and generator regulation cost. The data in the past 5 seconds are considered, which means c_t is 5 in this case. The moving average reward can reach and maintain at a high level after 2000 episodes' training, which can be seen in Fig. 4.11.

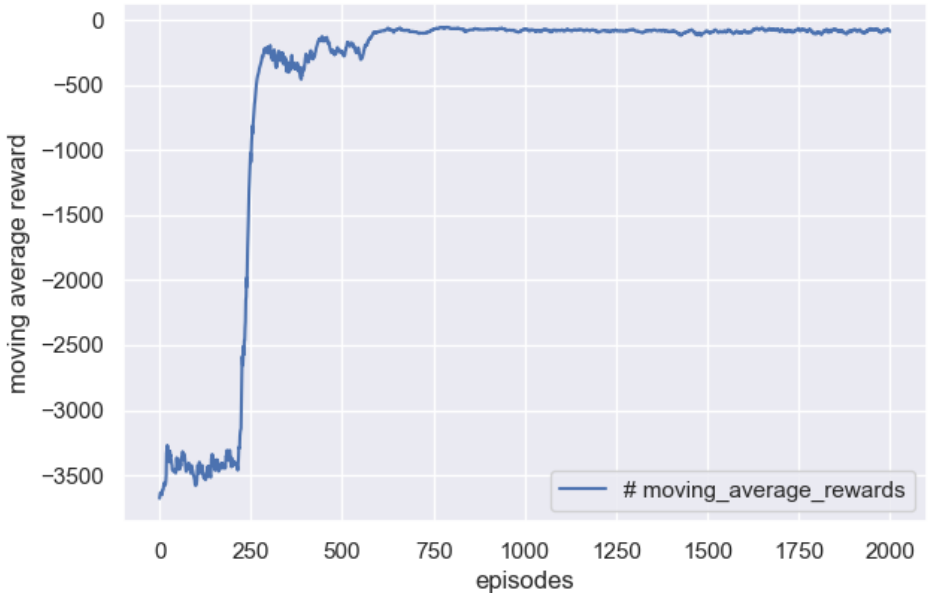

Figure 4.11 Case 2-Moving average reward

After the training has converged, the DDPG controller is integrated into the power system environment during dynamic simulation after system expereriences a load change at bus 8. The test results are shown in Fig. 4.12-4.13.

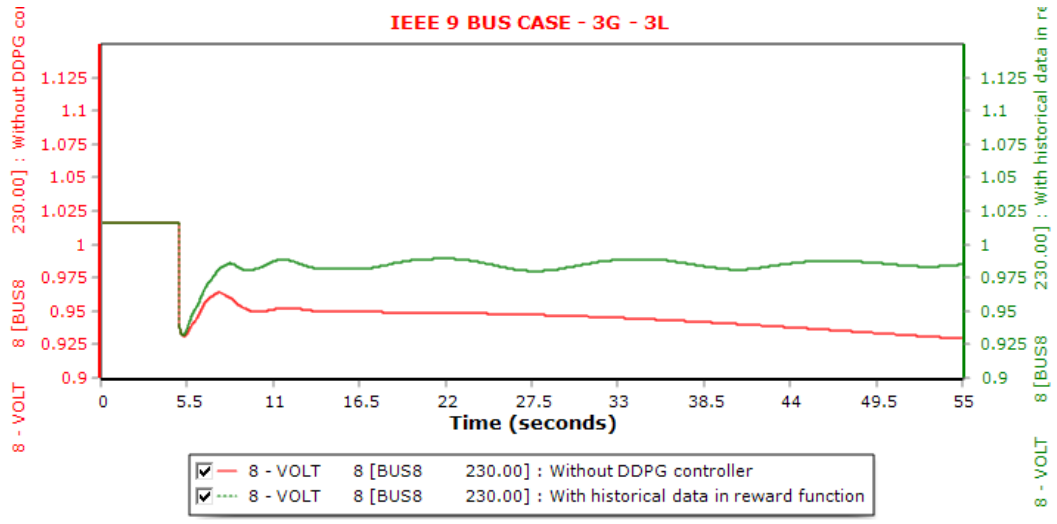

Figure 4.12 Case 2-Voltage of bus 8 with and without DDPG controller

which show that the control policy of the DDPG agent with historical voltage data considered can provide additional support to the system, and this can help the system voltage recover to a normal level.

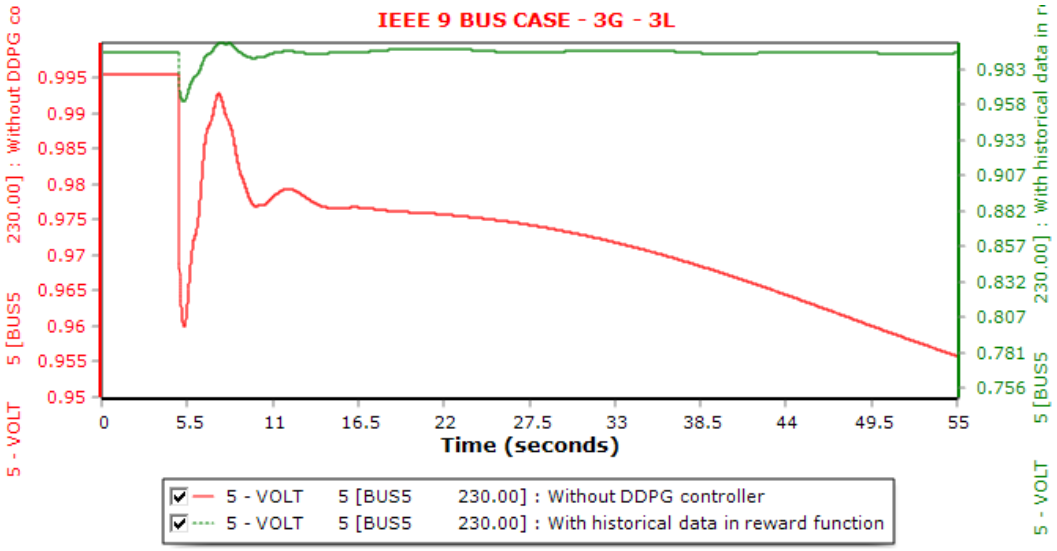

Figure 4.13 Case 2-Voltage of bus 5 with and without DDPG controller

Case 3-Consider voltage deviation, regulation cost, historical voltage data and rate of change of voltage

Further training of the DDPG agent with the reward function (3-12) that considers rate of voltage change of the historical data is also conducted. Since the rate of change is calculated using each historical voltage and its previous voltage values, there are 4 rates of voltage change data for each controlled bus when a 5 seconds interval of historical data is considered with a control interval of 1s. The moving average reward can be seen in Fig. 4.14.

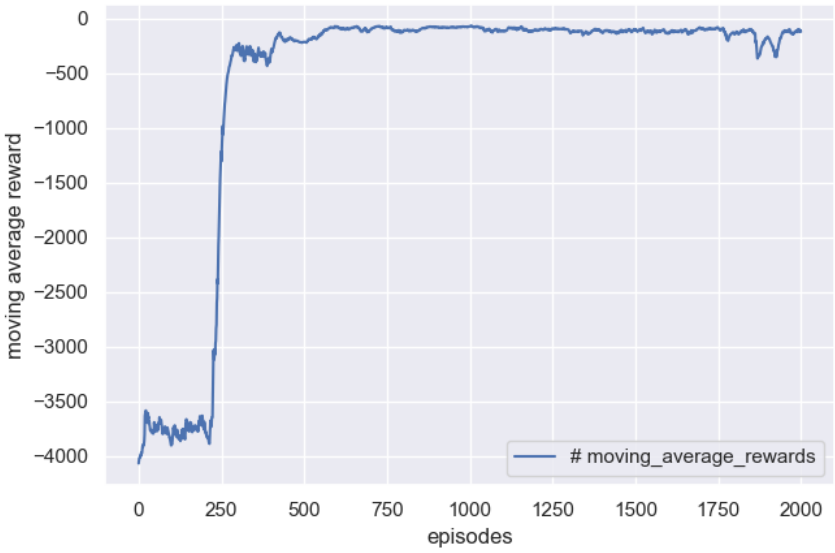

Figure 4.14 Case 3-Moving average reward

After being fully trained, the DDPG controller is tested with a random load change and the results can be seen in Fig 4.15 and Fig 4.16, which show the agent can provide voltage support and help the system voltage recover by considering the rate of voltage change.

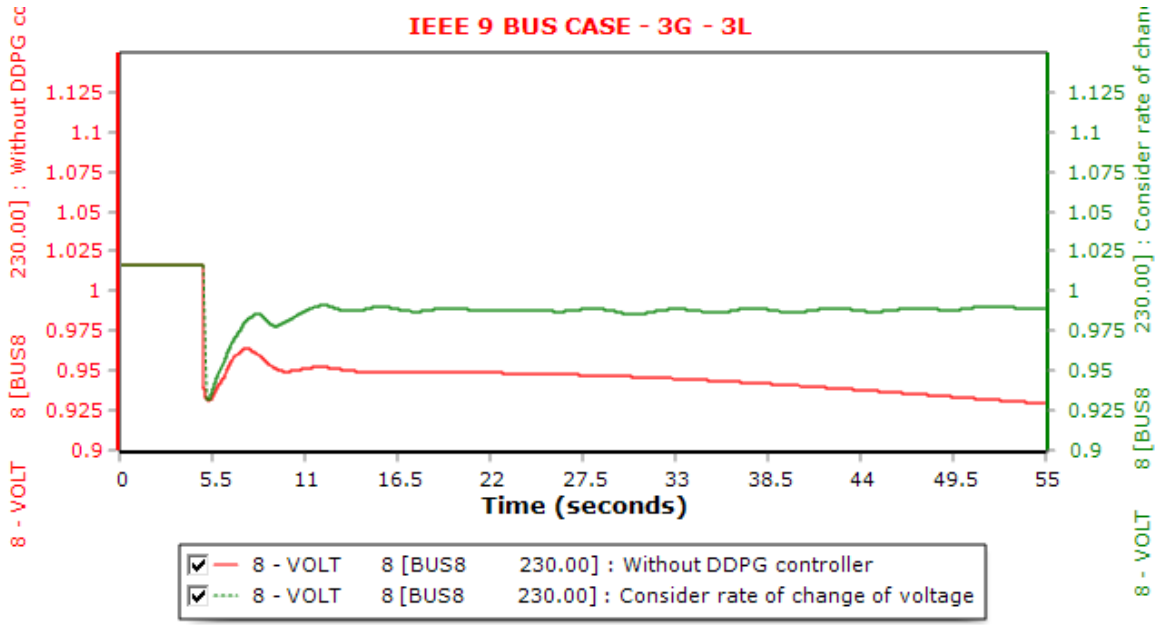

Figure 4.15 Case 3-Voltage of bus 8 with and without DDPG controller

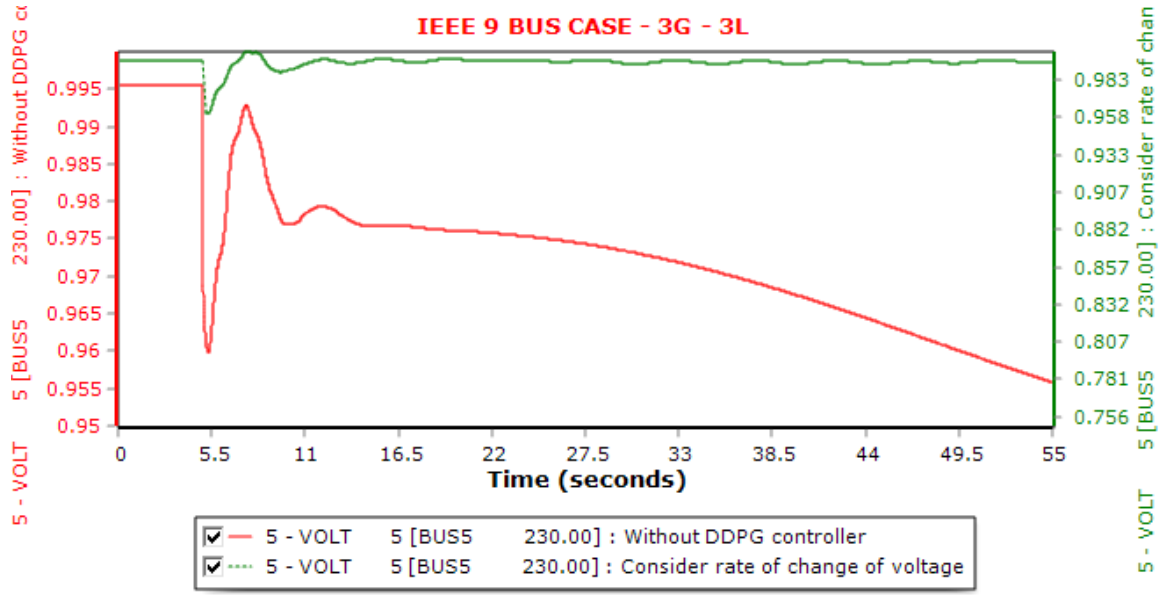

Figure 4.16 Case 3-Voltage of bus 5 with and without DDPG controller

Case 1 to Case 3 analyzed the DDPG agent control performance with different information considered. When adding the same disturbance at bus 8, the comparison of the voltage control effects can be seen in Fig 4.17. The red curve in Case 1 (also shown in Fig. 4.18 to present a clearer curve), which considered only voltage magnitude deviation, has more fluctuations in voltage magnitude during the whole dynamic simulation with the action applied at every 1 second by the DDPG agent. With historical data considered, the green curve (also shown in Fig. 4.19) of Case 2 performs a little bit

better, and the voltage magnitude is higher and has less fluctuations than Case 1. The blue curve (also shown in Fig. 4.20) of Case 3, which considers not only the historical voltage data but also the voltage rate of changes, has the most smooth voltage curve with less fluctuation with actions generated from the DDPG controller. Adding the historical voltage data and voltage rate of changes as the system states besides bus voltage magnitude can provide more information to the agent, so that the DDPG agent learns the policy to reduce the deviation and fluctuation to maximize the reward function, which improves the voltage recovery process.

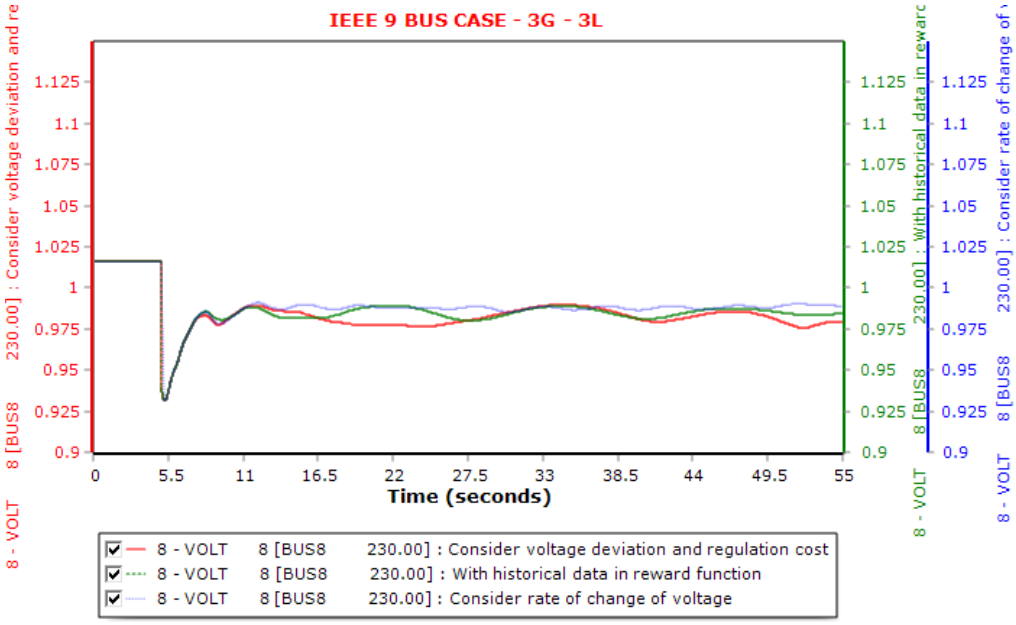

Figure 4.17 Voltage compare between Case 1 to Case 3

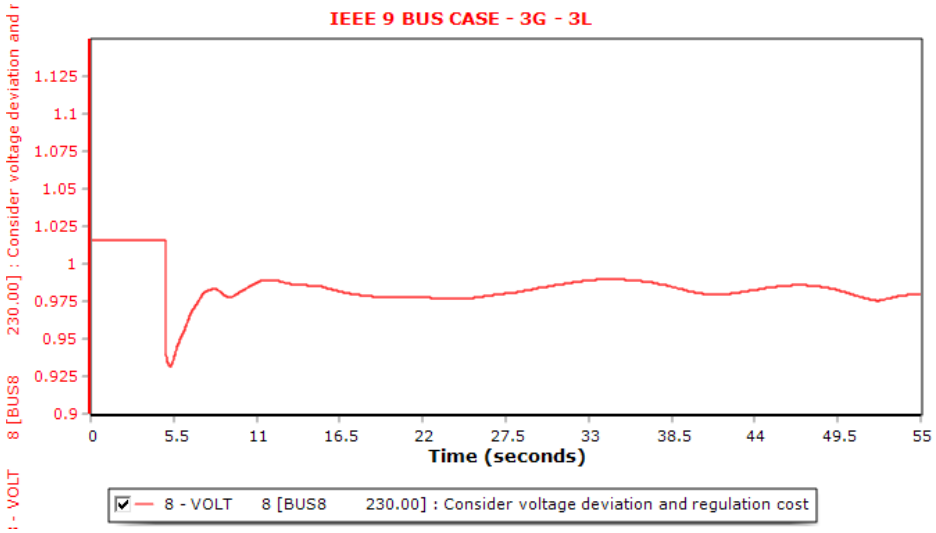

Figure 4.18 Voltage of bus 8 under Case 1

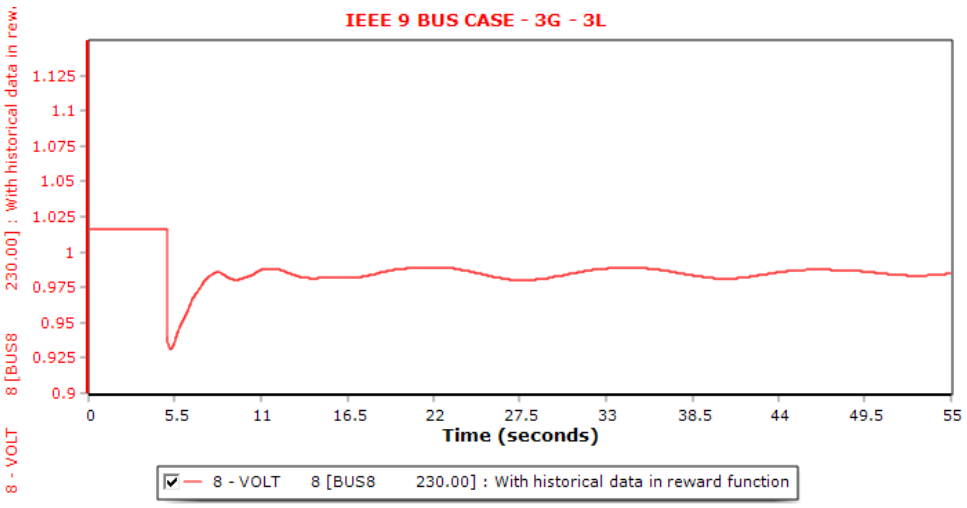

Figure 4.19 Voltage of bus 8 under Case 2

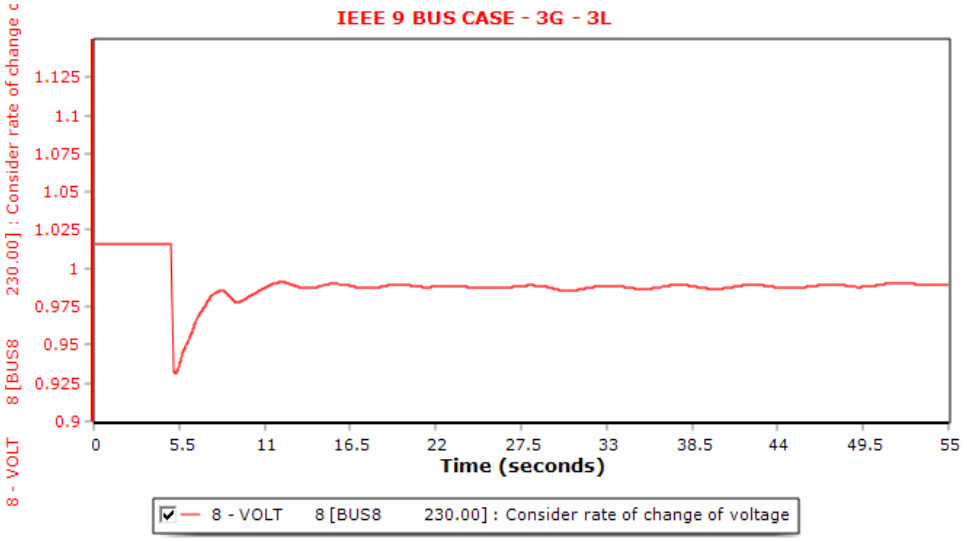

Figure 4.20 Voltage of bus 8 under Case 3

Case 4-With random selected location of the disturbance

The disturbance of Case 1 to Case 3 is located at bus 8. This section analyzes the scenario with randomly selected disturbance locations. The disturbance is randomly added to bus 5, bus 6 and bus 8 with random load change amount. Only voltage magnitude and regulation cost are considered in this scenario, which uses (3-7) as the reward function. After the agent is well trained, the test results are shown in Fig. 4.21 and Fig. 4.22. We can see from the results that the voltage can still be recovered with the random location of disturbance. This scenario is more likely to happen in the real power grid since faults always appear with significant uncertainty.

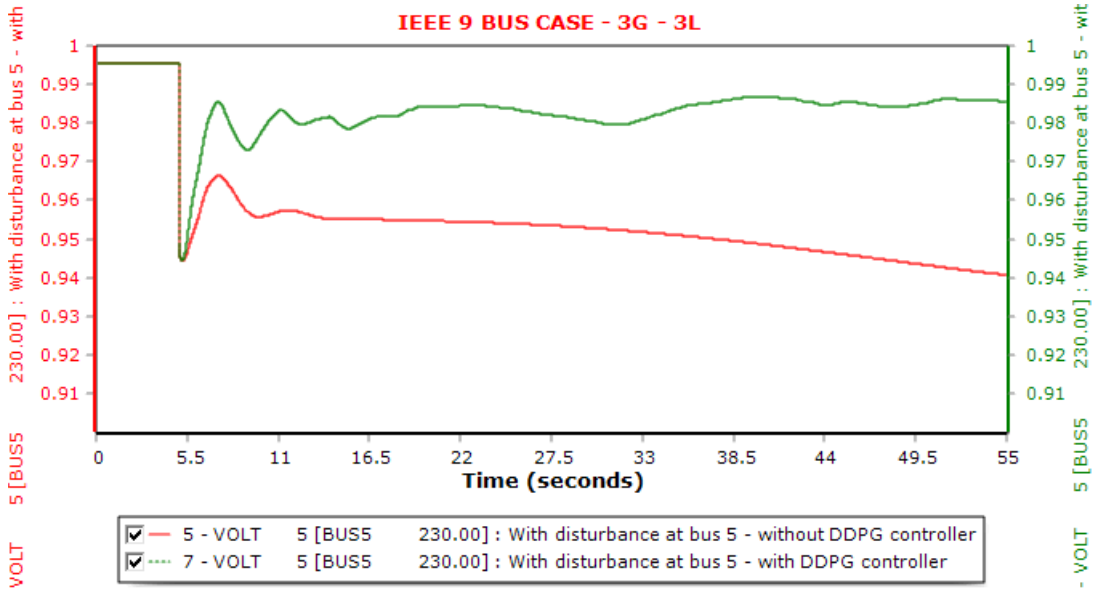

Figure 4.21 Case 4-Voltage of bus 5 when disturbance occurs at bus 5

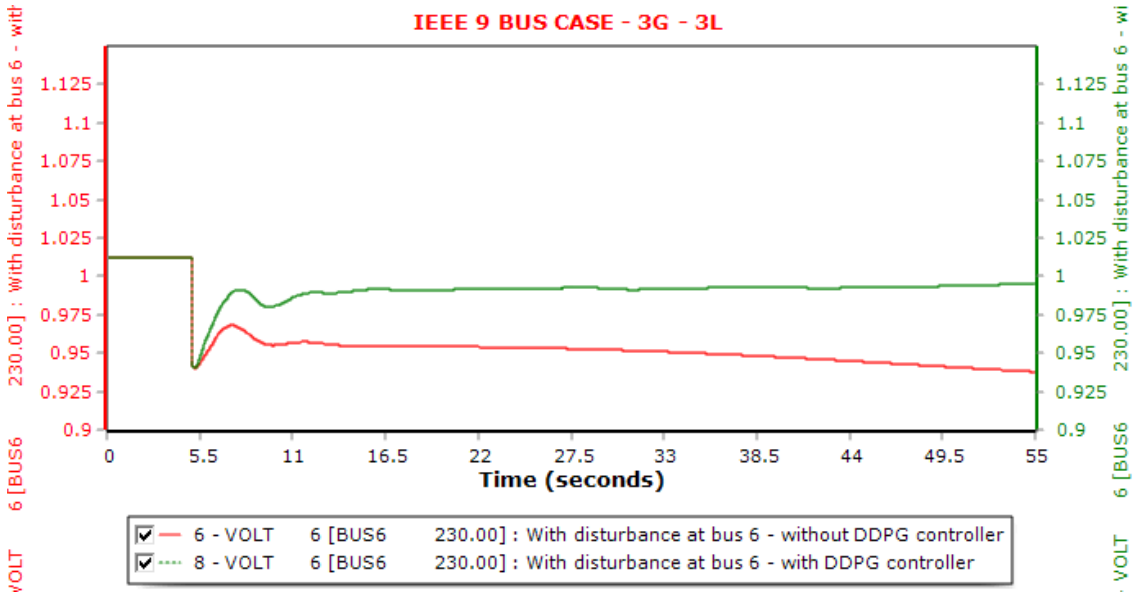

Figure 4.22 Case 4-Voltage of bus 6 when disturbance occurs at bus 6

4.4 Simulations Based on the Texas 2000-bus Power System

The real power grid is usually a large-scale system. In order to test the proposed method, simulations for the Texas 2000-bus synthetic power system are conducted to further test the efficacy of the proposed reinforcement learning based voltage control method on a more realistic system.

Various amounts of load changes are randomly introduced around bus 7068 (located near the Houston area) at 5s during the dynamic simulation to result in a 3%-5% voltage decrease. The power grid environment state is also the bus voltage magnitudes of buses in the disturbance area.

As for the controlled generators, the system structure can be seen in Fig. 4.23, generators 7307-7311 have the smallest electrical distance from the bus at which the load disturbance is placed. Generator 7098 and generator 7099 have large capacity and a significant margin in terms of reactive power capability and are also located electrically close to the disturbance area. Hence, they are good choice as the controlled generators. Since generator 7098 is connected to the swing bus of the system, generator 7099 is set as the controlled generator, together with generator 7310. As a result, generators 7099 and 7310 are set as the two controlled generators for this case.

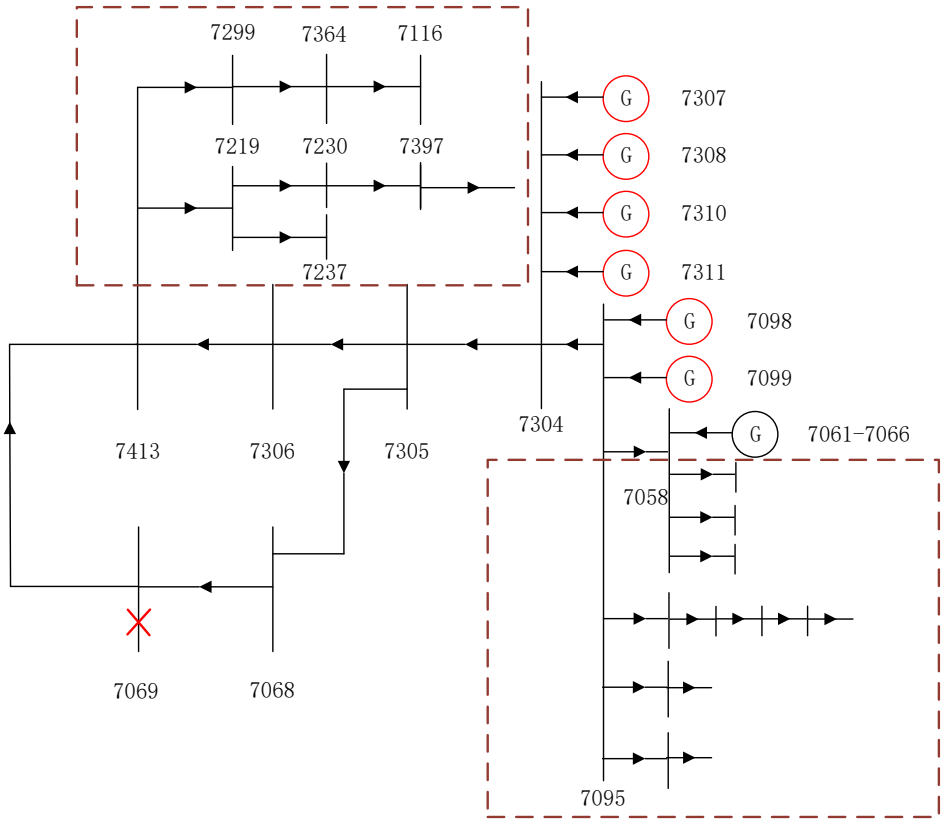

Figure 4.23 System structure near disturbance area

Since the load change amount is larger in this 2000-bus system, the action bound is set as [-2, 2]. The step interval is still set as 1 second. The DDPG controller learns during the training process under the aforementioned fault scenarios, and the moving average reward finally reaches a high level after 2000 episodes of training, as shown in Fig 4.24.

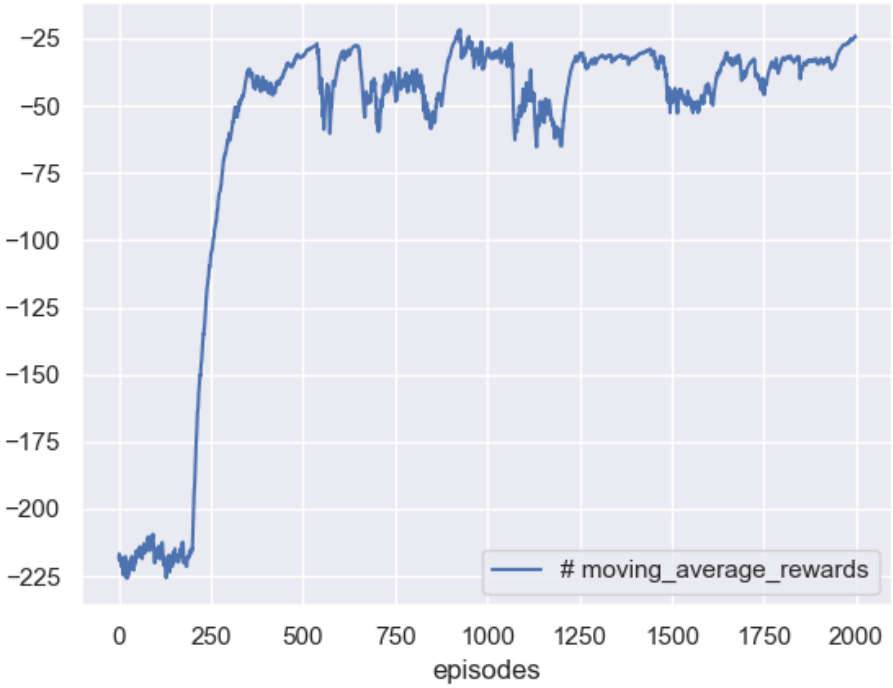

Figure 4.24 Moving average reward of 2000-bus system

Case 1- With 230MVar load change

After being fully trained, the DDPG agent is applied to the system to provide voltage support after a system disturbance. A 230MVar change of reactive power is added to bus 7069 of the 2000-bus system as shown in Fig. 4.23, with the test results shown in Fig. 4.25. The green curve is the voltage with conventional generator control and the red curve is with the DDPG controller. With only the conventional generator control, the system bus voltages decrease after the disturbance, stay at a low level and cannot recover. With the DDPG agent controlling the generators 7099 and 7310, the system voltage can be improved to normal levels.

The excitation system voltage reference value of the two controlled generators can be seen in Fig. 4.26 and Fig. 4.27. The voltage curve (green curve) is included to observe the effect of the voltage reference values in maintaining the voltage. After fully training, the DDPG agent appears to learn well and operate effectively. When the voltage is decreased from the standard normal value of 1 pu, the agent will adjust the commands and quickly provide support for the system. After the voltage roughly recovers to 1 pu, it keeps precisely regulating the command to stabilize the whole voltage level. From Fig. 4.28, it can be observed that the command for the two controlled generators have similar trends in general. The two generators cooperate well under the control of the DDPG controller to restore the system voltage.

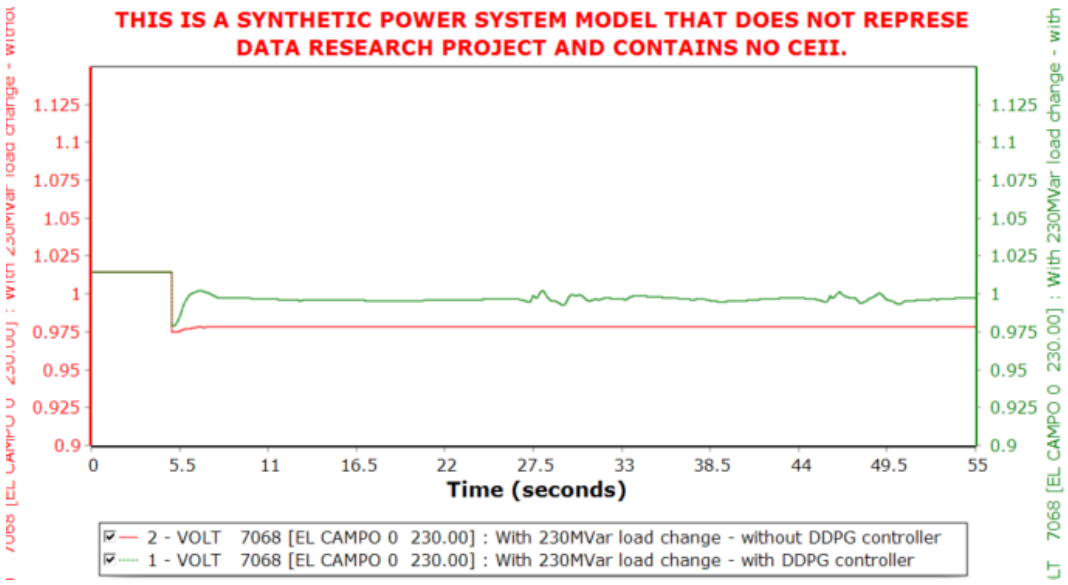

Figure 4.25 Case 1: Voltage of bus 7068 with and without DDPG controller

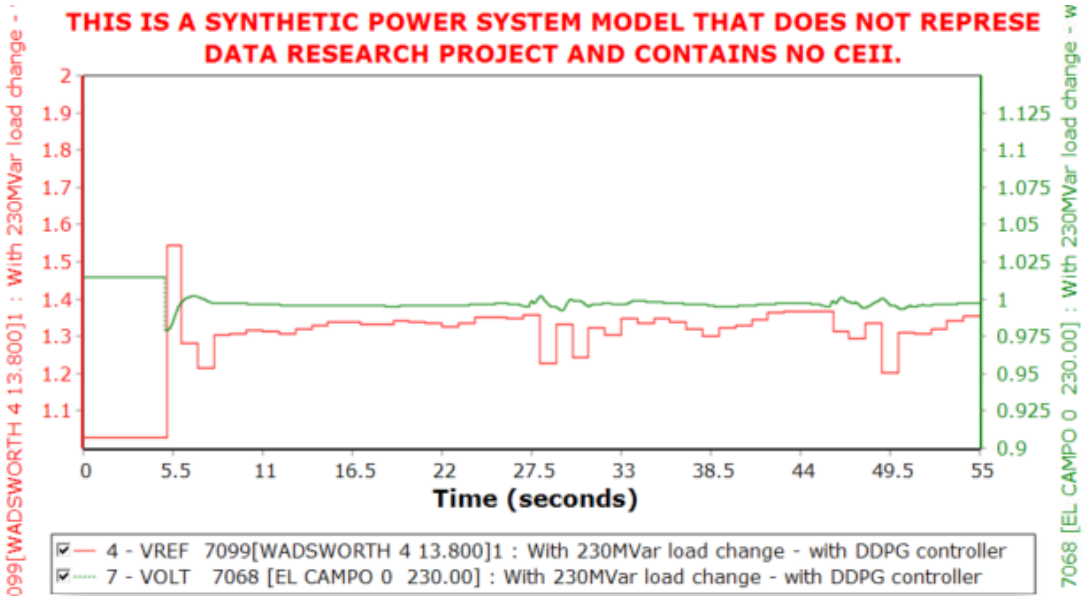

Figure 4.26 Case 1: Generator 7099 voltage reference command of the excitation system

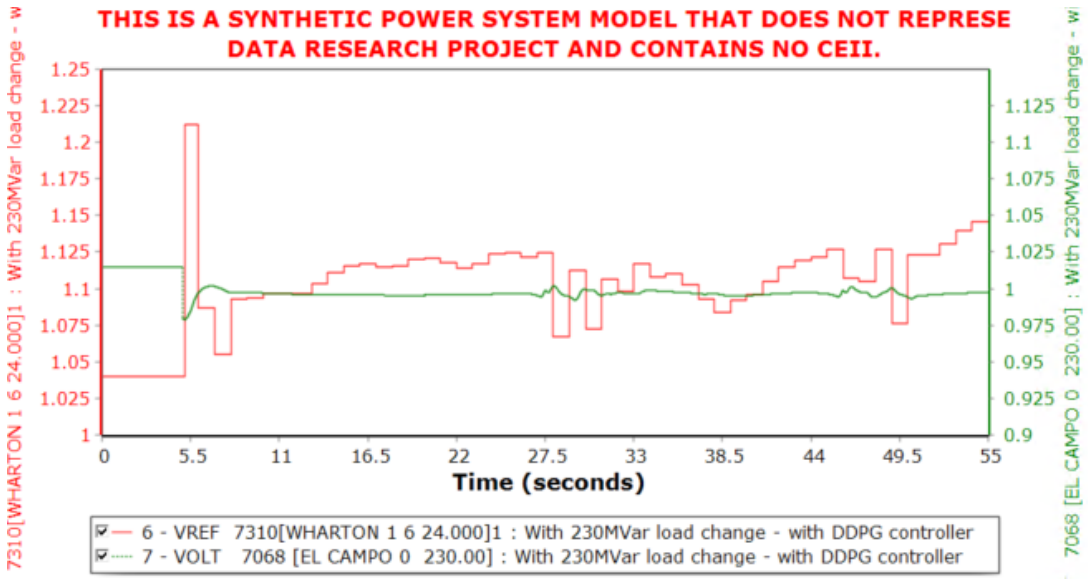

Figure 4.27 Case 1: Generator 7310 voltage reference command of the excitation system

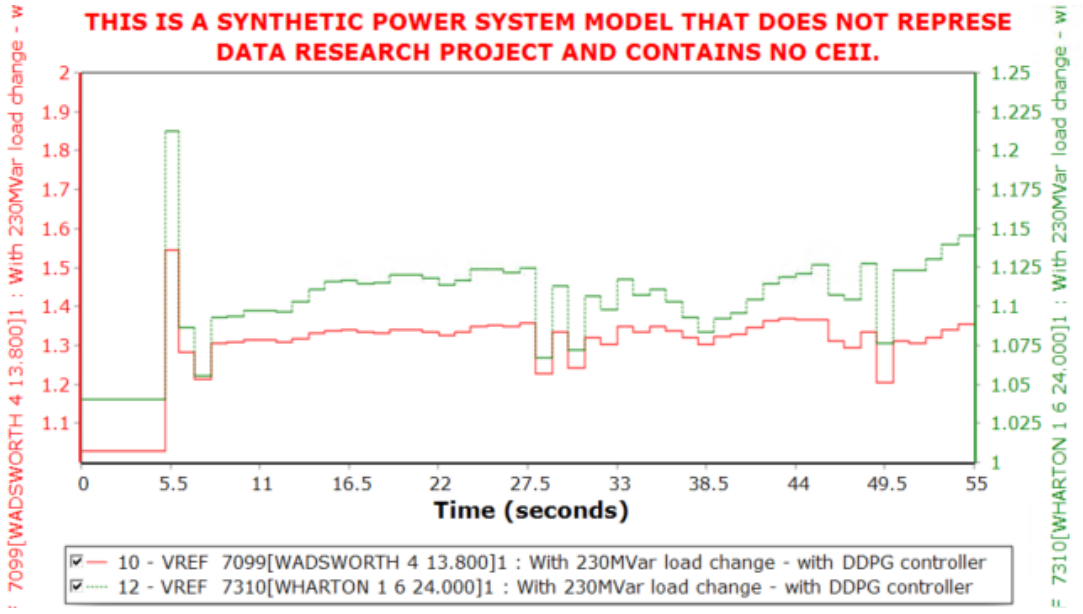

Figure 4.28 Case 1: Trend of the two generator voltage reference commands

Case 2- With 280MVar load change

A larger disturbance with 280MVar load change is introduced into the system to fully test the effectiveness of the proposed controller. The voltage curve compared with Case 1 of 230MVar load change can be seen in Fig. 4.29, and the red curve is the voltage with 280MVar load change which decreased more compared to the green curve of Case 1 after the disturbance is added into the system at 5 seconds. The excitation system voltage reference values of the two controlled generators can be seen in Fig. 4.30 and Fig. 4.31. The voltage curve (green one) is also retained in the figure to make the results more intuitive. It can be observed that after the voltage drop, the controller will output high value commands to make the two generators provide better voltage support to the system after a serious

disturbance. The voltage can be improved and can finally reach nearly 1 pu under the control of the controller. The controller will keep monitoring and regulating the system voltage during the whole dynamic simulation. The output commands from the controller are also presented in Fig. 4.32 which shows the trends of the whole control time range. From all the results presented, it is seen the controller can effectively provide support and precisely help the system restore voltage.

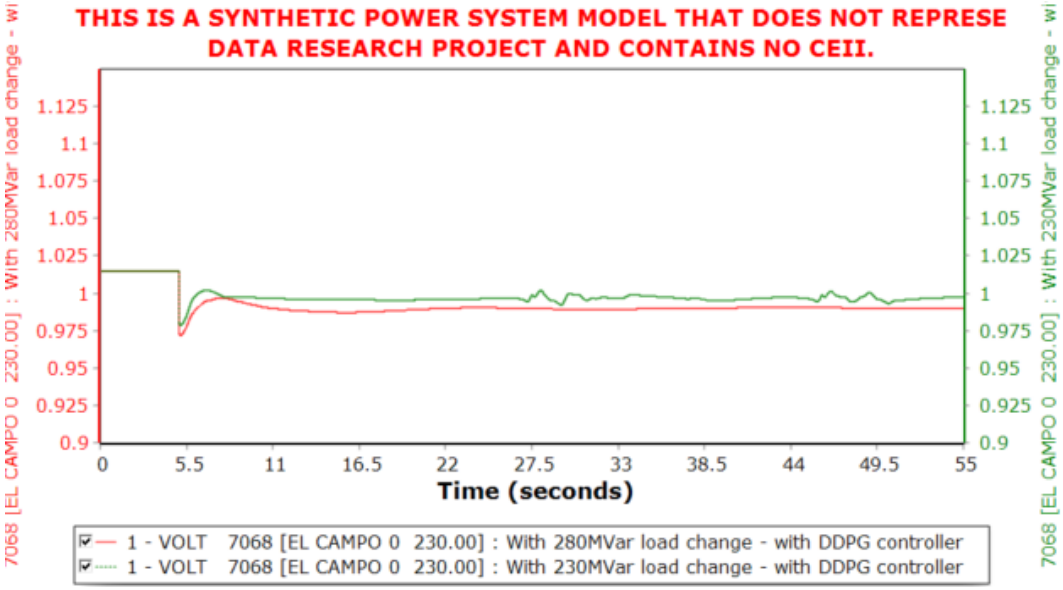

Figure 4.29 Voltage compare between Case 1 and Case 2.

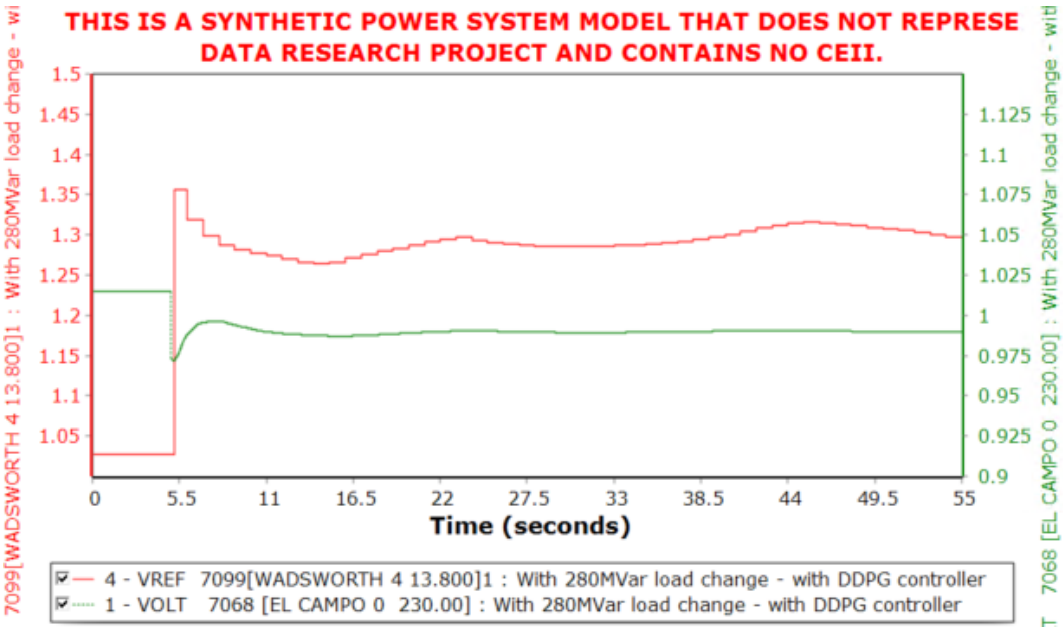

Figure 4.30 Case 2: Generator 7099 voltage reference command of the excitation system

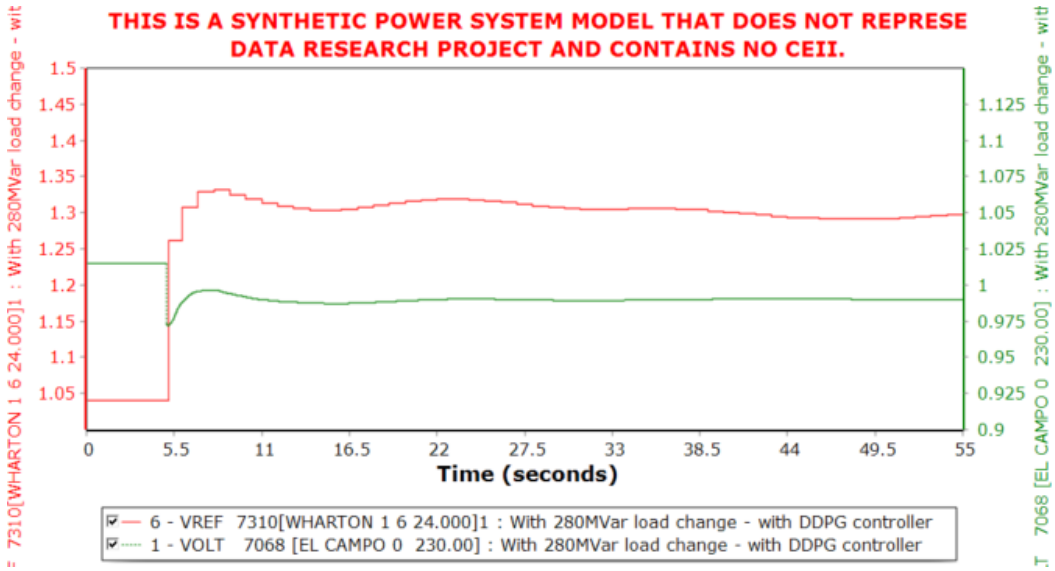

Figure 4.31 Case 2: Generator 7310 voltage reference command of the excitation system

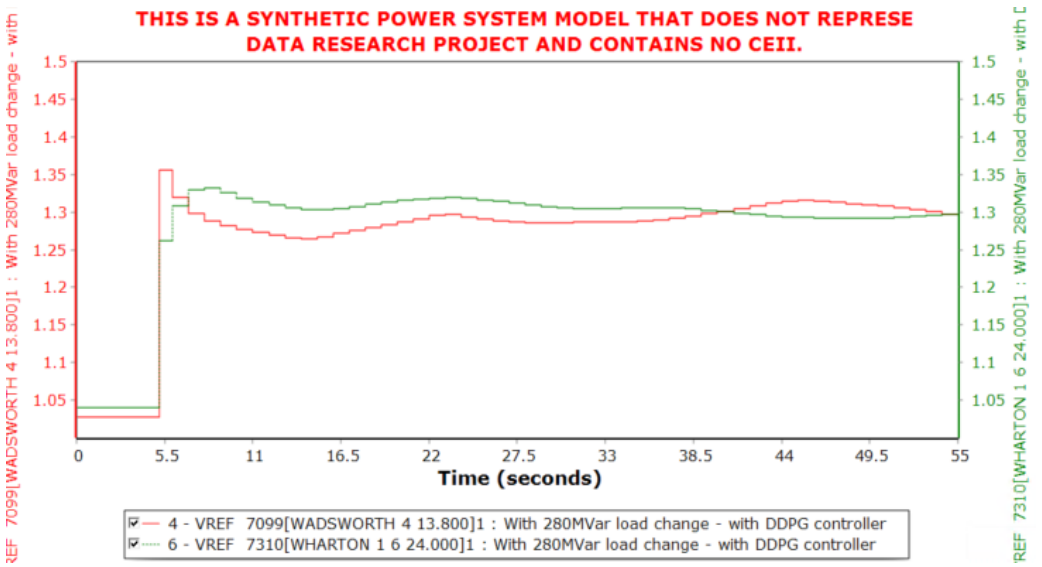

Figure 4.32 Case 2: Trend of the two generators voltage reference command

Case 3- With random selected location of disturbance

A random disturbance at buses 7219, 7306 and 7069 is introduced, respectively. Only voltage magnitude and regulation cost are considered and (3-7) is used as the reward function. After the agent is well trained, the test results are shown in Fig. 4.33 and Fig. 4.34. We can see from the results that the voltage can still be supported with a random disturbance. Since the capacity of the generator regulation amount is limited, the voltage cannot be completely restored to around 1 pu, but it still can help improve the system voltage, which is also very important to the safe operation of the system.

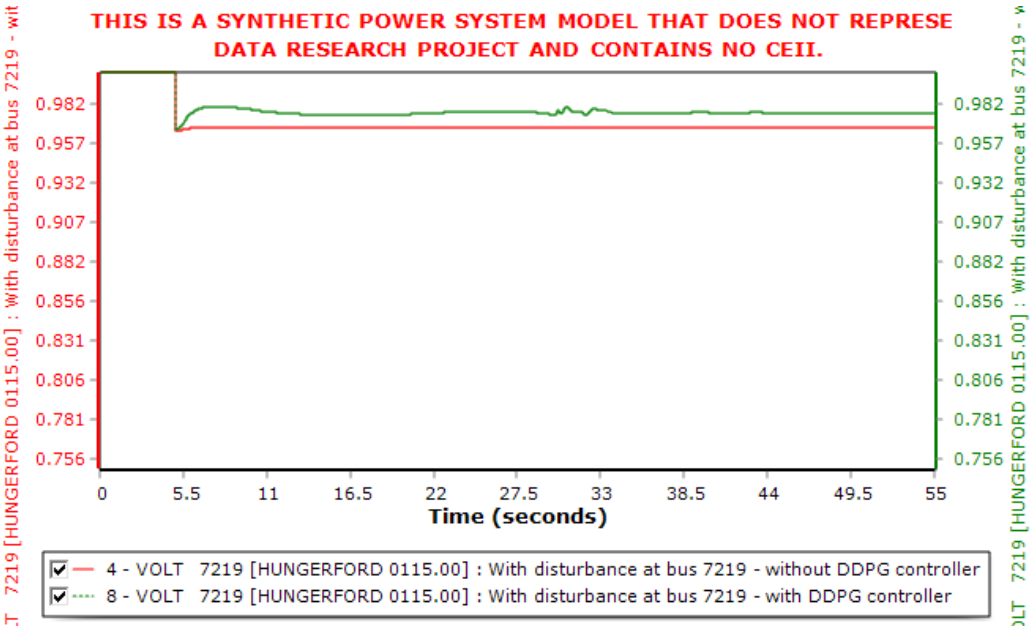

Figure 4.33 Case 3: Voltage of bus 7219 when disturbance occurs at bus 7219

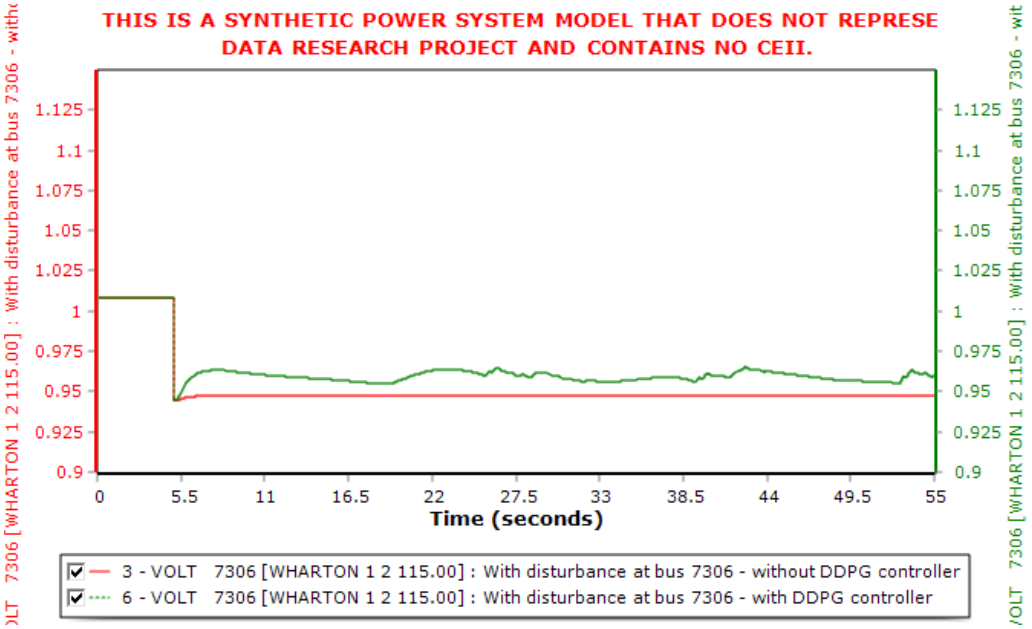

Figure 4.34 Case 3: Voltage of bus 7306 when disturbance occurs at bus 7306

Case 4- Considering voltage deviation, regulation cost, historical voltage data and rate of change of voltage

As the case in section 4.3, different reward functions are applied to the controller in the 2000-bus system, which include the base case that considers the voltage deviation and generator regulation cost, the case that considers historical voltage deviation and the case that adds rate of change of voltage in the reward function besides all the above factors. The simulation results are shown in Fig. 4.35.

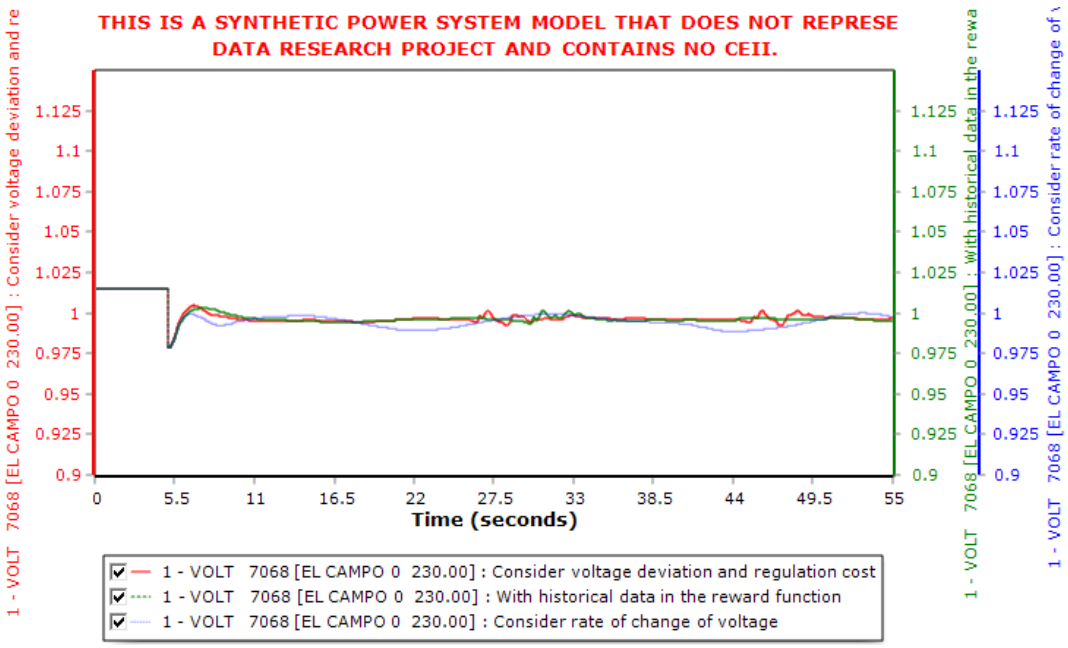

Figure 4.35 Case 4: Voltage comparison with different reward function

We can see that, with the addition of the rate of change of voltage into the reward function, the voltage recovers faster with a smoother curve (blue). The base case (red curve) and the case that considers historical voltage magnitude deviation has some oscillation. This will guide the agent to mitigate the oscillation and realize a maximum reward value, which will improve the system dynamic performance during the operation. The result shows that considering the voltage historical data and rate of change of voltage can improve the system dynamic performance when the controller restores the system voltage after disturbance.

5. Conclusions and future work

5.1 Main conclusions

The study in this report explores the efficacy of reinforcement learning in preventing voltage cascading failures following large disturbances in con-junction with detailed time-domain simulations. The results demonstrate that the controller can output accurate commands to the excitation system in real time based on the current system operating condition. The voltage recovery process is stable and fast under the control of the DDPG controller, without which the system voltage will keep decreasing slowly and have a risk of losing stability using only the conventional generator control. Different factors are considered in the reward function, the case that consider the historical voltage data and the rate of change of voltage has better dynamic performance.

This study also builds a DDPG based power system dynamic simulation platform to train the controller in the power system environment based on dynamic simulation. This platform provides a basis for training the RL algorithm during the system dynamic simulation, which makes it possible to further consider the system dynamic characteristics in future research.

5.2 Future work

The work presented observed the bus voltages of one disturbance area, control of more generators under multiple bus disturbances could be explored in the future to meet the needs of the practical power grid. Since voltage control is local problem, more control structures with multiple control devices could be considered in the future. Besides, more reinforcement learning algorithms are worth trying, such as proximal policy optimization (PPO) and soft actor-critic (SAC) [32], which have better training stability.

References

[1] M. Glavic, "(deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," Annual Reviews in Control, vol. 48, pp. 22-35, 2019.

[2] Wu, D., Zheng, X., Kalathil, D., and Xie, L. (2019a). Nested reinforcement learning based control for protective relays in power distribution systems, arXiv:1906.10815v1, Accessed August, 2019, (pp. 1–8), https://arxiv.org/abs/1906. 10815.

[3] Bag, A., Subudhi, B., and Ray, P. (2019). An adaptive variable leaky least mean square control scheme for grid integration of a PV system. IEEE Transactions on Sustainable Energy, vol. 11, no. 3, pp. 1508-1515, July 2020, doi: 10.1109/TSTE.2019.2929551.

[4] Ahamed, T. P. I., Rao, P. S. N., and Sastry, P. S. (2002). A reinforcement learning approach to automatic generation control. Electric Power Systems Research, 63, 9–26.

[5] Wu, J., Fang, B., Fang, J., Chen, X., and Tse, C. K. (2010). Online tuning of a supervisory fuzzy controller for low-energy building system using reinforcement learning. Control Engineering Practice, 18, 532–539.

[6] Wu, J., Fang, B., Fang, J., Chen, X., and Tse, C. K. (2010). Online tuning of a supervisory fuzzy controller for low-energy building system using reinforcement learning. Control Engineering Practice, 18, 532–539.

[7] Wang, H., Lei, Z., Zhang, X., Peng, J., and Jiang, H. (2019). Multiobjective reinforcement learning-based intelligent approach for optimization of activation rules in automatic generation control. IEEE Access, 7, 17480–17492.

[8] Zhang, X. S., Yu, T., Pan, Z. N., Yang, B., and Bao, T. (2018). Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and EVs. IEEE Transactions on Power Systems, 33, 4097–4110.

[9] Abouheaf, M., Gueaieb, W., and Sharaf, A. (2018). Model-free adaptive learning control scheme for wind turbines with doubly FED induction generators. IET Renewable Power Generation, 12, 1675–1686.

[10] Vlachogiannis, J. G., and Hatziargyriou, N. (2004). Reinforcement learning for reactive power control. IEEE Transactions on Power Systems, 19, 1317–1325

[11] Xu, Y., Zhang, W., Liu, W., & Ferrese, F. (2012). Multiagent-based reinforcement learning for optimal reactive power dispatch. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 42, 1742–1751.

[12] H. Xi and Dominguez-Garcia, "Optimal tap setting of voltage regulation trans-formers using batch reinforcement learning," IEEE Transactions on Power Systems, vol. 35, p. 3, 2020.

[13] Zarabbian, S., Belkacemi, R., and Babalola, A. A. (2016). Reinforcement learning approach for congestion management and cascading failure prevention with experimental application. Electric Power Systems Research, 141, 179–190.

[14] Yang, Q., Wang, G., Sadeghi, A., and Giannakis, G. B. (2019). Real-time voltage control using deep reinforcement learnings, arXiv:1904.09374v1, Accessed August, 2019, (pp. 1–9), https://arxiv.org/abs/1904.09374

[15] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, and Z. Yi, "Deep-reinforcement-learning-based autonomous voltage control for power grid operations," IEEE Transactions on Power Systems, vol. 35, no. 1, pp. 814–817, 2020.

[16] M. Glavic, "(deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," Annual Reviews in Control, vol. 48, pp. 22–35, 2019.
[17] R. Diao, Z. Wang, D. Shi, Q. Chang, J. Duan, and X. Zhang, "Autonomous volt-age control for grid operation using deep reinforcement learning," in 2019 IEEE Power Energy Society General Meeting (PESGM), 2019, pp. 1–5.
[18] Glavic, M. (2005). Design of a resistive brake controller for power system stability enhancement using reinforcement learning. IEEE Transactions on Control Systems Technology, 13, 743–751.
[19] Yousefian, R., Bhattarai, R., and Kamalasadan, S. (2017). Transient stability enhancement of power grid with integrated wide area control of wind farms and synchronous generators. IEEE Transactions on Power Systems, 32, 4818–4831.
[20] Yousefian, R., and Kamalasadan, S. (2018). Energy function inspired value priority based global wide-area control of power grid. IEEE Transactions on Smart Grid, 9, 552–563.
[21] Q. Huang and R. Huang, "Adaptive Power System Emergency Control Using Deep Reinforcement Learning," IEEE Transactions on smart grid, vol. 11, no. 2, pp. 1171–1182, Mar 2020.
[22] J. Heaton, "T81-558: Applications of deep neural networks: module 3 and module 4."
[23] M. Bernico, Deep Learning Quick Reference: Useful Hacks for Training and Optimizing Deep Neural Networks with TensorFlow and Keras. Packt Publishing Ltd, 2018.
[24] T. Lillicrap and J. Hunt, "Continuous control with deep reinforcement learning," arXiv preprint, 2015.
[25] L. Jin and R. Kumar, "Model Predictive Control-Based Real-Time Power System Protection Schemes," IEEE Transactions on power systems, vol. 25, no. 2, pp. 988–998, May 2010.
[26] A. Singhal and V. Ajjarapu, "Real-time local volt/var control under external dis-turbances with high PV penetration," IEEE Transactions on smart grid, vol. 10, no. 4, p. 3849–3859, Jul 2019.
[27] R. Lowe, "Multi-agent actor-critic formixed cooperative-competitive environ-ments," Proc. Adv. Neural Inf. Process. Syst., p. 6379–6390, 2017.
[28] J. Conto, "IEEE9_jconto", https://drive.google.com/drive/folders/0B7uS9L2Woq_7fmd4YXVxMEZKT3dJV2FleGkzS2FzVmd1RHhBNVdUTGpvdldkMnl2bXRLM1k?resourcekey=0-nuCqXu2XJ0_fxBzwHcmCGg, accessed: 2022-08-02.
[29] A. B. Birchfield; T. Xu; K. M. Gegner; K. S. Shetye; T. J. Overbye, "Grid Structural Characteristics as Validation Criteria for Synthetic Networks," in IEEE Transactions on Power Systems, vol. 32, no. 4, pp. 3258-3265, July 2017.
[30] A. B. Birchfield; K. M. Gegner; T. Xu; K. S. Shetye; T. J. Overbye, "Statistical Considerations in the Creation of Realistic Synthetic PowerGrids for Geomagnetic Disturbance Studies," in IEEE Transactions on Power Systems, vol. 32, no. 2, pp. 1502-1510, March 2017.
[31] K. M. Gegner; A. B. Birchfield; T. Xu; K. S. Shetye; T. J. Overbye, "A methodology for the creation of geographically realistic synthetic powerflow models," 2016 IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, 2016, pp. 1-6.

[32] C. Xu, R. Zhu and D. Yang, "Karting racing: A revisit to PPO and SAC algorithm," 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), 2021, pp. 310-316, doi: 10.1109/CISAI54367.2021.00066.