(Deep) Reinforcement learning for electric power system control and related problems: A short review and perspectives

Mevludin Glavic[a,*]

[a]*Independent Researcher/Consultant, Maka Dizdara 66, 75000 Tuzla, Bosnia and Herzegovina*

Abstract

This paper reviews existing works on (deep) reinforcement learning considerations in electric power system control. The works are reviewed as they relate to electric power system operating states (normal, preventive, emergency, restorative) and control levels (local, household, microgrid, subsystem, wide-area). Due attention is paid to the control-related problems considerations (cyber-security, big data analysis, short-term load forecast, and composite load modelling). Observations from reviewed literature are drawn and perspectives discussed. In order to make the text compact and as easy as possible to read, the focus is only on the works published (or "in press") in journals and books while conference publications are not included. Exceptions are several work available in open repositories likely to become journal publications in near future. Hopefully this paper could serve as a good source of information for all those interested in solving similar problems.

Keywords: Electric power system, reinforcement learning, deep reinforcement learning, control, control-related problems.

1. Introduction

Power system is a vital infrastructure of modern societies. How important is best seen from the fact that complete (known as blackouts) or partial (known as brownouts) disruptions of power system result in huge economic

*Corresponding author

Email address: mevludin.glavic@gmail.com (Mevludin Glavic)

and societal costs. An example is US-Canada power system outage of August 14, 2003 [1] with estimated costs of 10 billion US dollars. In addition, more and more services are expected to rely on electricity in the future (an example is transportation systems' increased reliance on electricity due to development and deployment of electrical vehicles) and it is reasonable to expect the costs of an outage such as [1] would be much higher if happens in some future.

Complexity of present and expected future power systems is/will be increasing due to deployment of electricity generation from so-called renewable energy sources (RES), such as wind and solar, which are naturally uncertain and interfaced with power system through power electronics converters thus reducing the system inertia resulting in faster dynamics. This type of electricity generation ranges from small to large and are connected across all levels of power systems (high-voltage transmission, medium-voltage and low-voltage distribution and microgrids). New types of loads (usually interfaced with a system through power electronics converters) such as for example electrical vehicles and increased use of High-Voltage Direct Current (HVDC) to connect (usually a country geographical coverage) individual sub-systems also add to the complexity of present and expected future power systems.

Advanced control techniques are needed to ensure reliable electricity delivery from generation sources to end-users and prevent (or decrease probability) of system's blackouts/brownouts avoiding their huge economic and societal consequences. Implementation of advanced communications infrastructure in power systems together with the availability of powerful computation architectures, and power electronics devices open up the possibilities to implement advanced control schemes. All these complemented with achievements in control theory, control engineering, computer science, operational research, and applied mathematics offer a number of advanced algorithms to be used in control systems' design (see [2–4] for discussions and vision from systems (in general, including power systems) and control perspectives).

The use of recent breakthrough algorithms from machine learning opens possibilities to design power system controls with the capability to learn and update their control actions. This paper reviews considerations of Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL) to design advanced controls in electric power systems.

Research efforts in RL and DRL resulted in a number of useful methods allowing power system controllers to learn a goal-oriented control law from interactions with a system or its simulation model [5–7]. In RL and DRL set-

ting a controller observes the system state, take control actions, and observe the effects of these actions. and in this way progressively learn an algorithm (a control law) associating control actions to the observations in order to fulfil a pre-specified objective [5, 6, 8].

Some of considerations are already reviewed in [9]. In this paper the focus is on power system control while decisions (like scheduling and market decisions) and energy management are not included. RL and DRL-based power system controls are reviewed as they relate to operating states of power systems (control in normal state, preventive, emergency, and restorative control) and control levels (local, household, microgrid, subsystem, wide-area) complemented with RL and DRL considerations to control-related problem: cyber-security, short-term load forecasting, big data analysis, and component/subsystem equivalent modelling.

The paper is organized as follows. Section 2 describes ongoing changes in present and future power system structure together with presentation of power system operating states. RL and DRL are shortly introduced in Section 3 accompanied with some very recent connections between these methods and control in general. Section 4 reviews RL and DRL considerations for power system control while Section 5 discusses possible future research directions (perspectives) and Section 6 concludes.

2. Power system structure and operating states

Modern power systems undergo considerable transformation in their structure expected to be more pronounced in the future. Present and future power system structures are illustrated in Fig. 1.

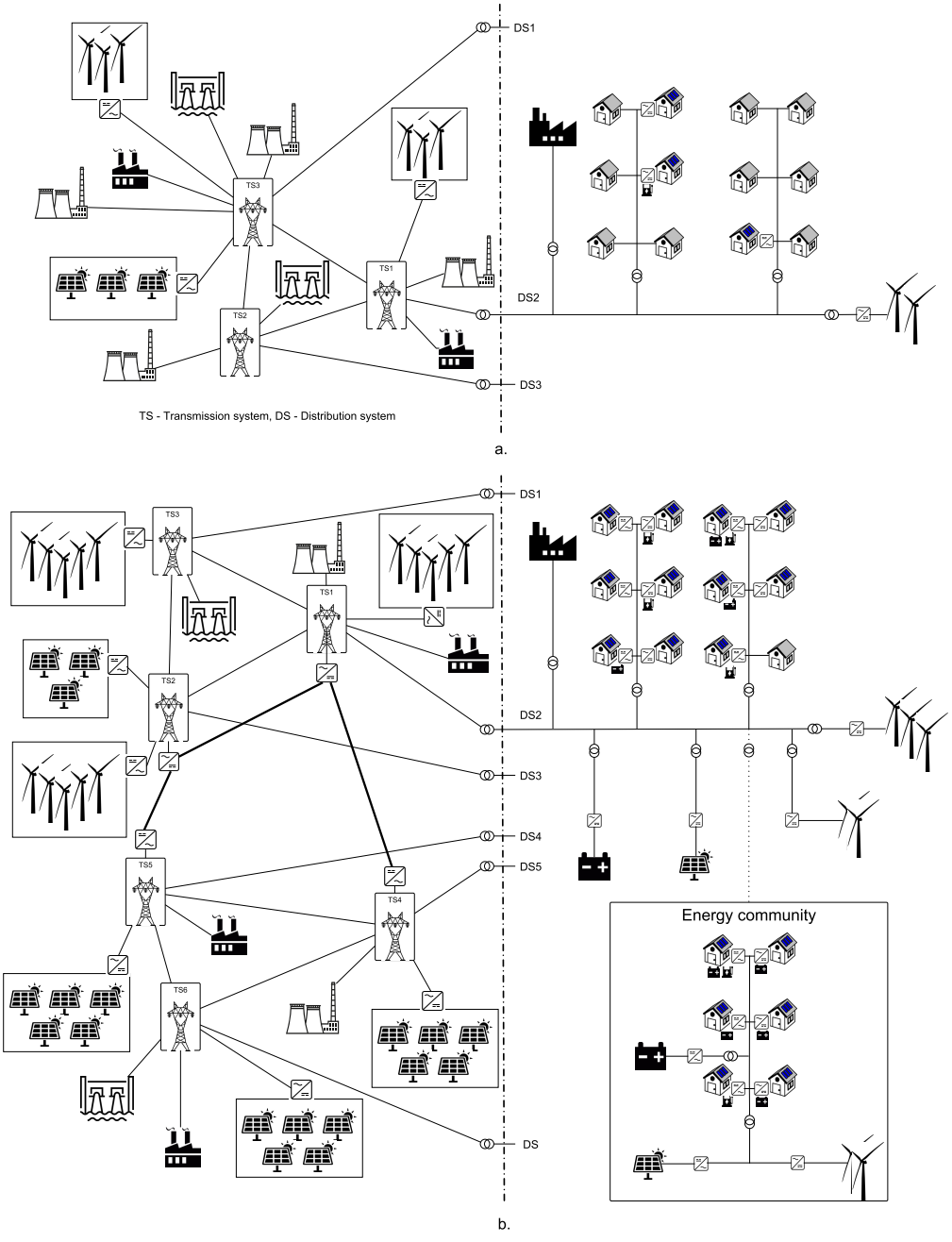

Figure 1: Structure of power system: present (a) and future (b)).

The trends driving the transformation are summarized as follows [10]:

- Changes in electricity generation sources (the mix and characteristics). These changes shifts electricity generation from large power plants to smaller generation from RES through their progressive deployment across all levels of the system (transmission and distribution) together with decommissioning of large thermal power plants (coal-fired and nuclear). This brings uncertainty in electricity generation and decrease of the system inertia since RES generators are usually interfaced with the system through power electronics converters (this makes frequency regulation and control more challenging).

- New electricity load types and changes in load profiles. An example of new type of the load is electrical vehicle with charging stations installed across the system including individual homes. Changes in load profiles are induced by possibility to generate electricity at the load side (this is termed as "prosumers" since they both generate and use electric energy), the use of electronics and controls in homes, offices and industrial sites, and growing participation of the loads in electricity markets and power system control.

- Smart grid technologies reflected in terms of advanced communications infrastructure, new instrumentation/measurement technologies (like phasor measurement units (PMU) for transmission systems and μPMU and advanced measurement infrastructure (smart meters) for distribution systems) and increase of available data.

- The progressive emergence of microgrids and energy communities as entities in power systems. These entities are similar and essentially include a group of interconnected loads and RES-based electricity generation within a geographical area that acts as a single controllable entity with respect to the grid. They could operate in grid connected mode but also could be disconnected from the system (actually this is main operation mode of energy communities) and operate as autonomous entity (this bring some flexibility in control of power systems, in particular restorative).

- Emergence of the electricity storage technologies across all levels of power systems (ranging from large storage devices connected to trans-

mission system to small ones connected to distribution system and microgrids/energy communities but also at individual load sites). These technologies revealed to be main enablers of future power system operation (possibility to smooth generation-load imbalances in uncertain operation conditions) but also could serve as important control devices across power system operation states.

- Increase in the deployment of HVDC lines to connect subsystems and electricity generation from off-shore wind-based RES and increase in deployment of so-called FACTS devices (Flexible Alternating Current Transmission System). The former transforms pure AC to hybrid AC/DC system (in transmission but also distribution systems). The later opens possibilities to control power systems more efficiently.

The transformation further led to the consideration of the concept of Internet of Things in electric power systems [11] where it often comes under term Energy Internet [12]. The widely accepted classification of electric power system operating states is the one introduced in [13]. Figure 2 illustrates five operating states as defined in [13] and adapted in [14].

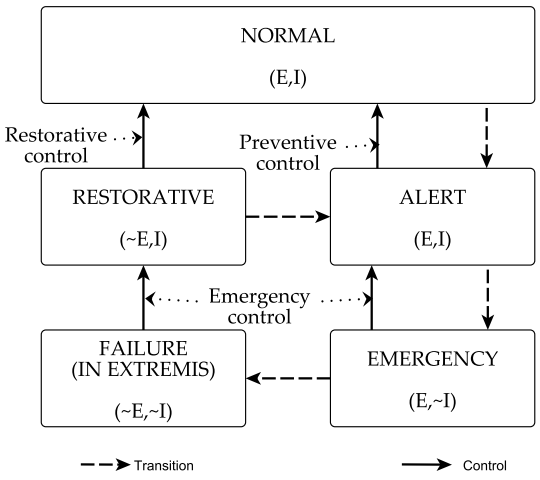

Figure 2: Power system operating states (adopted from [14])

The states are defined in terms of the status of equality (E) and inequality (I) constraints of the system (violated (indicated with "˜" in Fig. 2) or not

violated). The equality constraints express the generation-load demand balance while inequality constraints express physical limitations of power system components (usually defined in terms of current and voltage magnitudes, active, reactive and apparent powers that a system component can withstand without any damage).

Figure 2 also illustrates controls used in electric power systems. In addition to preventive, emergency and restorative controls (shown in Fig. 2) there is a need for the system to be controlled in normal operating state since continuous small variations of generations and loads are present in this state.

3. (Deep) Reinforcement learning: a short introduction and considerations for control problems in general

Many power system controls are designed as the solution of multi-stage decision optimal control problems. Dynamic programming [8] is natural framework to solve these problems. Dynamic programming, reinforcement learning and deep reinforcement learning are only briefly presented in this section (to support discussions in later sections of this paper). More details on these subjects can be found in [5–8, 15]

3.1. Dynamic programming and optimal control

Reference [16] considers dynamic programming as one of four canonical models to solve multi-stage decision optimal control problems in power systems. This section largely follows presentation of [17], where dynamic programming is formulated in the framework of discounted infinite time-horizon optimal control for a short description of dynamic programming followed by introduction to (deep) reinforcement learning. For this control, the objective is to define, for every possible initial state x_0, an optimal control sequence $u^*_{\{t\}}(x_0)$ (control policy). In order to determine this policy the value function is defined as,

$$V(x) = \max_{u_{\{t\}}} R(x, u_{\{t\}}), \tag{1}$$

$R(x, u_{\{t\}})$ is the discounted return defined as,

$$R(x_0, u_{\{t\}}) = \sum_{t=0}^{\infty} \gamma^t r(x_t, u_t). \tag{2}$$

where $r(x,u) \leq B$ is a *reward* function, $\gamma \in]0,1[$ a discount factor, and $u_{\{t\}} = (u_0, u_1, u_2, \cdots)$ a sequence of control actions applied to the system.

The value function is the solution of the Bellman equation [8],

$$V(x) = \max_{u \in U}[r(x,u) + \gamma V(f(x,u))], \tag{3}$$

The optimal control policy is deduced from the above equation as,

$$u^*(x) = \arg\max_{u \in U}[r(x,u) + \gamma V(f(x,u))]. \tag{4}$$

The value function can be re-expressed by defining the so-called Q-function,

$$Q(x,u) = r(x,u) + \gamma V(f(x,u)), \tag{5}$$

as,

$$V(x) = \max_{u \in U} Q(x,u), \tag{6}$$

while the optimal control policy is re-expressed by,

$$u^*(x) = \arg\max_{u \in U} Q(x,u). \tag{7}$$

Equation (7) provides a straightforward way to determine the optimal control law from the knowledge of Q.

3.2. Reinforcement learning

In most of the electric power system control problems state space is infinite and the Q-function must be approximated [5, 6, 8]. Prevailing approach is a state space discretization technique that divides the state space into a finite number of regions. On each region the Q-function depends only on u and, in the RL algorithms, the notion of state used is not the real state of the system x but rather the region of the state space to which x belongs denoted by s (sometimes termed as pseudo-state). In general, the knowledge of the region $s(x_t)$ at some time instant t together with u is not sufficient to predict with certainty the region to which the system will move at time $t+1$. To model this uncertainty it is assumed that the sequence of discretized states followed by a system under a certain control sequence is a Markov chain characterized by time-invariant transition probabilities $p(s'|s,u)$, which define the probability to go to a state $s_{t+1} = s'$ given that $s_t = s$ and $u_t = u$.

Using transition probabilities and a discretized reward signal ($r(s,u)$), the control problem can be reformulated as a Markov Decision Process (MDP) and search for a control policy defined over the set of discrete states S, that maximizes the *expected* return. The Q-function is now characterized by the following Bellman equation,

$$Q(s,u) = r(s,u) + \gamma \sum_{s' \in S} p(s'|s,u) \max_{u \in U} Q(s',u), . \tag{8}$$

A classical dynamic programming algorithm like the value iteration or the policy iteration algorithm [5, 6, 8] can be used to estimate solution of this problem. Optimal control policy is now defined by,

$$\hat{u}^*(x) = u^*(s(x)) = \arg\max_{u \in U} Q(s(x),u). \tag{9}$$

RL is an approach to approximately solve above problem through estimation of the Q-function by interacting with the system or its simulation model (by trial and error). The interaction works as follows:

1. at time t, the algorithm observes the state s_t, sends a control signal u_t, and receives information back from the system in terms of the successor state s_{t+1} and reward $r_t = r(s_t, u_t)$;
2. above four values are used either to estimate the transition probabilities and the associated rewards (model based) and then compute the Q-function, or learn directly the Q-function without learning any model (model-free);

A RL algorithm at each time-step selects a control signal, by using the so-called *ϵ-greedy* policy (a control signal is chosen at random in U chooses, with a probability of ϵ). The smaller the value of ϵ, the better the RL algorithms exploit the control law they have learned and the less they explore their environment (this is known as "exploration-exploitation" trade-off in RL algorithms).

Among many, most popular (at least in electric power systems community, as will be clear in later sections of this paper) RL methods are Q-learning, fitted Q-iteration, SARSA, TD, and their variants (for full details see [5, 6], including relations among mentioned RL methods since they are often not clear from electric power system literature).

3.3. Deep reinforcement learning

The rise and development of DRL is strongly connected to advances and breakthroughs in deep learning [18] and in particular deep learning for neural networks [19] (these neural networks are also known as Deep Neural Networks (DNNs)). In principle, DNNs include more (hidden) layers in between input and output layers of neural networks and enable RL to scale to decision-making problems with high-dimensional state and action spaces owning to their generalization capabilities. This is usually achieved by training DNNs to approximate (parameterized by the weights of a DNN): value function, control policy or model (in terms of transition probabilities and rewards) [15].

A general DRL framework is illustrated in Figure 3. Note that not all elements of the framework are present in every DRL method (an example is replay memory element used to store the experience that it can be reprocessed at a later time [7].

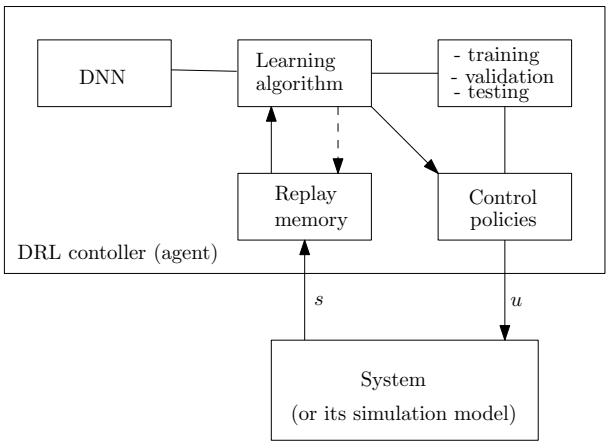

Figure 3: A general framework of DRL (adopted from [7] and slightly modified)

Each layer of a DNN consists in a non-linear transformation and the sequence of these transformations leads to learning different levels of abstraction. An arbitrarily large number of hidden layers is possible within a DNN. Two types of layers are of particular interest in DRL [7]:

- Convolutional layers: parameters of these layers consist of a set of learnable filters (or kernels), which have a small receptive field and

which apply a convolution operation to the input, passing the result to the next layer. As a result, the network learns filters that activate when it detects some specific features.

- Recurrent layers and their variants: particularly well suited for sequential data. Most important variants relevant for DRL include the long short-term memory network and neural Turing machines (able to encode information from long sequences).

DRL approaches are usually classified as: value-function based, control-policy-based and model-based. Most popular of these approaches are just noted in this section (interested readers are referred to [7, 15] for full details). Value-function-based DRL include Deep Q-networks (DQN) that combine Q-learning with a neural representation and Extensions of DQN (to avoid instability and divergence): Double DQN, multi-task learning, and rapid learning. Most popular control-policy-based DRL is the asynchronous advantage actor-critic (A3C) algorithm able to efficiently learn tasks with continuous action spaces.

3.4. (D)RL for control problems in general: a short review of recent works

A recent paper [20] comprehensively reviewed considerations of RL and DRL for solving control problems. This reference includes a long list of relevant works and is a good starting point for all those interested in the field.

Worth mentioning here are two very recent monographs [21, 22], a recent survey paper [23], and a vision paper [24] (not listed in [20] since not available at the time when [20] was prepared). Monograph [22], written by one of leading experts in the field, offers a comprehensive material on RL and DRL for solving optimal control problems and will certainly become a classic material to read whenever either RL or DRL are considered to solve optimal control problem in any engineering (and non-engineering) problems. An interesting observation from [22] is "...no methods that are guaranteed to work for all or even most problems, but there are enough methods to try on a given challenging problem with a reasonable chance that one or more of them will be successful in the end...". This indeed holds true when RL and DRL are considered for solving power system control problems. Material of [21] focused on Lyapunov-based approach in using RL in feedback control and establishing stability during the learning phase and the execution

(therefore offering a sort of safe RL). Reference [23] surveyed existing RL-based feedback control solutions to optimal regulation and tracking of single and multi-agent systems with the focus on Q-learning (as core algorithm for discrete-time) and the integral RL (as core algorithms for continuous-time systems). A vision for the field of systems and control in light of the advances in machine learning and artificial intelligence (including RL and DRL) was presented in [24]. A conclusion that can be drawn from [24] is the future systems and control developments need to fully engage with the two fields and trigger new research directions.

4. A short review of (D)RL considerations for electric power system control and control-related problems

The considerations are reviewed in terms of power system operating state for which they are designed to work. Table 1 summarizes these considerations.

Table 1: Summary of (D)RL considerations for electric power system control

Control	Reference(s)
Normal	[25–56]
Preventive	[57]
Emergency	[17, 58–76]
Restorative	[77–79]
Control-related considerations	[80–90]

4.1. Control in normal operating state

These considerations are summarized, in terms of a control problem, control level (local, subsystem, microgrid, household, wide-area), (D)RL method used, and corresponding references. The summary is displayed in Table 2.

Reference [25], through consideration of protection relays as on/off control devices, proposed DRL to set up the relay control logic able to differentiate heavy load and faulty operating conditions of a distribution system with high proliferation of electricity generation from RES. A problem was cast as a multi-agent one and term nested is used because the method exploits nested structure of electric distribution system.

Works presented in [26–29] dealt with Maximum Power Point Tracking (MPPT) control for wind energy [26, 28] and photo-voltaic electricity generation [27, 29]. They are local controls acting on individual wind or photo-voltaic sources so that maximum electricity is generated. Q-learning was used in [26] for variable-speed while [28] suggested the use of neural networks in conjunction with Q-learning for permanent magnet synchronous generator wind energy conversions systems. Photo-voltaic system was considered in [29] for MPPT through memetic computing incorporated in RL (Q-learning). Reference [27] proposed variable leaky least mean square algorithm to generate photo-voltaic inverter reference, RL algorithm (Q-learning) for MPPT and a sliding mode approach to generate switching signals. The MPPT is designed with Q-learning algorithm for extraction of maximum power from photo-voltaic panels during varied solar insolation. The work presented in [30] is also a sort of control for maximizing gathered energy (in this work from ocean waves) with Q-learning used to design controller to maximize energy absorption in each sea state optimal resistive control of a wave point absorber. Similar problem was considered in [31] using least-squares policy iteration RL method with function approximator (radial basis function) and comparisons with Q-learning and SARSA RL method suggesting better performances of the proposed approach.

References [32, 33] proposed RL to control electricity generation from electronically-interfaced electricity generators based on RES in order to support frequency regulation in the system. Actor-critic neural networks were considered in [32]. The on-line controller based on a policy iteration reinforcement learning paradigm along with an adaptive actor-critic technique was considered in [33] for wind turbines with doubly fed induction generators.

Q-learning was proposed in [34] for optimal tap setting of on-load tap changer of step-down transformers (connecting electric distribution systems with the rest of the system) in order to control distribution system side voltage under uncertain load dynamics. A sequential learning algorithm was used to learn an control-value function for each transformer based on which the optimal tap positions is determined.

Automatic Generation Control (AGC) and Load-Frequency Control (LFC) were considered in [35, 37–39, 41–46] (the objective is to keep frequency in a narrow range around nominal value, for example in Europe $[49.8 - 50.2]Hz$). AGC and LFC differs in that AGC includes LFC together with generation dispatch function for control of so called area control error that is a parameterized sum of frequency deviation and active power flows over so-called

tie-lines (the lines connecting subsystems within a larger interconnection). Most of considerations suggest the use of Q-learning (or its variant $Q(\lambda)$). Single AGC controller was defined in [35] and designed using Q-learning. The same problem as in [35] was investigated in [36] with the difference that continuous state and control spaces were considered (this was achieved through the use of radial basis function neural network trained by the RL method). Q-learning method was also used in [37] with genetic algorithms to tune controller parameters. Work presented in [38] suggested single AGC controller based on multi-step $Q(\lambda)$ method while [39] suggested the use of correlated equilibrium $Q(\lambda)$ within a multi-agent setting (similar approach was proposed in [40] with the difference that correlated equilibrium Q-learning was proposed within a multi-agent framework for AGC). A multi-objective Q-learning was used to activate rules of AGC (consisting of dynamic allocation of the AGC regulating commands among various AGC units, and activation of the secondary control reserve of those units was considered in [41]. Reference [42] proposed a lifelong learning control scheme for AGC where the wind farms, photovoltaic stations, and electric vehicles are aggregated as a wide-area virtual power plant participating in AGC with other generation plants. Q-learning was adopted, together with imitation learning and knowledge transfer, to this purpose. Reference [43] proposed a combination of $R(\lambda)$ with an imitation pre-learning process and tested it for AGC based on control performance standards. $R(\lambda)$ is an average reward RL method [5] similar to Q-learning. Work presented in [44] proposed DRL for LFC with continuous control. Off-line control policy is suggested with a DRL method and on-line control where features are extracted by stacked denoising auto-encoders. An adapted DQN for continuous spaces was use as DRL. Integral RL [23] for load frequency regulation in multi-area electric power systems was suggested in [45]. Emotional RL was proposed in [46] for AGC where the controller integrates two parts: RL and artificial emotion (this part is a function of the elements of RL (control, learning rate, reward) and essentially allows embedding domain knowledge).

A subsystem level voltage controls based on RL were considered in [47, 48] and DRL in [49] and [50]. References [47, 48] considered voltage control through Q-learning used to learn the optimal control law for reactive power control. The objective is to keep substations' voltage magnitudes within the normal range around nominal voltages ($[0.9 - 1.1]$) for distribution and ($[0.95 - 1.05]$) for transmission system. Q-learning was used to learn how to adjust a closed-loop control rule by mapping states (power flow solutions)

Table 2: Summary of (D)RL considerations for control in normal operating state

Level	Control problem	(D)RL method	Reference(s)
Device (local)	Protection relays	DRL (DQN)	[25]
	MPPT	Q-learning	[26–30]
		Least-squares policy iteration	[31]
	Frequency regulation	Q-learning	[32, 33]
	Voltage	Q-learning	[34]
Subsystem	AGC/LFC	Q-learning	[35–37, 41, 42]
		$Q(\lambda)$	[38]
		Correlated $Q(\lambda)$	[39]
		Correlated Q	[40]
		$R(\lambda)$	[43]
		DRL (DQN)	[44]
		Integral RL	[45]
		Emotional RL	[46]
	Voltage	Q-learning	[47, 48]
		DRL (DQN)	[49]
		DRL (DQN/DDPG)	[50]
	Load control	DRL (DQN)	[51, 52]
Microgrid	Transient energy storage performances	Tailored	[53]
	Power quality control	Tailored	[54]
	Active/Reactive generation power	Tailored	[55]
Household	Parameter tuning	$Q(\lambda)$	[56]

to controls (computed off-line). This is achieved though the formulation of constrained power flow problem as a multi-stage decision problem [47] and the use of an average consensus algorithm within a multi-agent RL-based framework [48]. Reference [49] proposed a two time-scale voltage regulation

scheme for distribution systems with radial grid. The optimal set-points of smart inverters are obtained, at fast scale (every second) by minimizing bus voltage deviations from their nominal values using a power flow model (exact or linear approximation). A DQN algorithm is deployed at the slower time scale (every hour) to configure a set of shunt capacitors to minimize the long-term discounted voltage deviations. Work presented in [50] considered two DRL methods (DQN and deep deterministic policy gradient (DDPG)) for subsystem level voltage control with observation that DDPG method offered much better performances after a sufficient number of training scenarios.

Load as a control mean to balance electricity generation and consumption was considered in [51, 52] using DRL to this purposes. Reference [51] considered an approach to find a near-optimal sequence of decisions based on sparse observations. This reference investigated the capabilities of different deep learning techniques, such as convolutional neural networks and recurrent neural networks, to extract relevant features for finding near-optimal policies for a residential heating system and electric water heaters, with conclusion that LSTM network offers a higher performance than stacking these time-series in the input of a convolutional neural network. Reference [52] suggested the use of a convolutional neural network to extract hidden state-time features to mitigate partial observability. A convolutional neural network is used as a function approximator to estimate the Q-function in the supervised learning step of fitted Q-iteration.

Considerations of RL for microgrid level control were presented in [53–55]. In [53] a hybrid energy storage (consisting of Lithium-Ion battery and ultra-capacitor) is controlled in order to improve transient performances in a microgird involving photo-voltaic system and diesel generators. Two neural networks are used to this purpose: one to estimate system dynamics on-line and another to to calculate the optimal control input for the storage system through on-line learning based on the estimated system dynamics. This approach is specific, it somewhat resembles DRL but does not belong to any known of DRL algorithms. In [54] a DSTATCOM (Distribution STATic COMpensator) was proposed to compensate power quality issues (the reactive power, harmonics, and unbalanced load current) in a microgrid. Voltage controller minimizes the voltage profile at point of microgrid coupling with the rest of the system, whereas the current based controller compensate the unbalanced load current in distributed generation sources. RL method proposed in [54] is also specific: for each pair of input/output signal, three different control signals are considered and in each state the adaptation unit

is used to selects one control. Reference [55] proposed a critic-based adaptive control system that include a neuro-fuzzy and a fuzzy critic controllers for the control of active and reactive powers generation in a microgrid. The fuzzy critic controller is based on a neuro-dynamic programming RL algorithm. On-line tuning of the output layer weights of the neuro-fuzzy controller is realized through reinforcement signal produced by the critic controller together with the back-propagation of error.

RL considerations to control a household were presented in [56]. $Q(\lambda)$ was proposed to learn the optimal values of only one parameter of a fuzzy controller that include a set of fuzzy rules generated by off-line optimization. A value of parameter corresponds to one set of fuzzy rules.

4.2. Preventive control

Q-learning was suggested in [57] to determine optimal control of active power generations for preventing cascading failure and blackout in smart grids. This approach belongs to subsystem level controls and considers single line outages (termed N-1 contingency in power system literature) and two consecutive line outages (termed N-1-1). The control is designed to work in normal operating state of the system and applies control action in this state in order to avoid cascading failures and possible blackouts/brownouts. Proposed approach was tested in experimental set-up in addition to tests in simulation environments (as in many other considerations reviewed in this paper).

4.3. Emergency control

Table 3 summarizes these consideration also in terms of a control problem, control level (local, subsystem, microgrid, wide-area), (D)RL method used and, corresponding references.

Two problems for which (D)RL was considered in existing works are instability (transient, oscillatory angle and voltage instabilities) and cascading failure problem (cascading outages of transmission lines and electricity generation plants usually initiated by outage of a transmission line suffering lasting overload).

Transient angle instability appears in electric power systems after large disturbances (usually outage of large generators or important transmission lines as well as three-phase short-circuit in transmission system). Imbalance in generation and demand causes fast increase/decrease of angular velocities

Table 3: Summary of (D)RL considerations for emergency control

Level	Control problem	(D)RL method	Reference(s)
Device (local)	Transient angle instability	Q-learning	[17, 60–62]
		Fitted Q-iteration	[58, 59]
		DRL (DQN)	[73]
	Oscillatory angle instability	Q-learning	[17, 60, 65–68]
		Fitted Q-iteration	[58, 59, 69]
Subsystem	Voltage (FIDVR)	DRL (DQN)	[73]
	Cascading failure	Q-learning	[76]
Wide-area	Oscillatory angle instability	Q-learning	[70, 71]
		$TD(\lambda)$	[72]
	Transient angle instability	actor-critic	[63, 64]
	Frequency instability	$TD(\lambda)$	[74]

of synchronous generators (this instability is also termed as first-swing instability). The aim of controls is to keep the system in synchronism (with angular velocities equal or very close to the nimnal value defined by the nominal system frequency). References [17, 58–64, 73] dealt with transient instability control by controlling individual electric power system components such as thyristor-controlled series capacitor [17, 58, 59] and a dynamic brake (a resistor usually located near electricity generation plant) to absorb excess of electricity generation [60, 61]. Q-learning was used in [17, 60–62] while [58, 59] suggested fitted Q-iteration. Reference [60] suggested limiting controls to stabilizing ones (derived from the concept of control Lyapunov functions) in order to ensure safety during exploration in RL. Inclusion of state history to recover Markov property in partially observable problems was considered in [62]. A dynamic brake wa also considered in [73] for emergency control using DRL (DQN was an approach of the choice in [73]) largely following implementation details presented in [17, 62]. The problem of transient angle instability was also considered within wide-area control systems

[63, 64]. Work of [63] presented an optimal wide-area system-centric controller and observer based on a hybrid RL (adaptive critic design) and TD with eligibility traces framework. A similar approach (in terms of the use of RL) was presented in [64] with an extension consisting in a value priority scheme to prioritize local and proposed global control so damping of both local and inter-area oscillations is achieved. The prioritizing scheme is designed using a derived Lyapunov energy function.

Oscillatory angle instability relates to the problem of low-frequency oscillations in the system (local modes in the range $[0.7 - 2.0]$ Hz and inter-area modes in the range $[0.1 - 0.8]$ Hz). This type of instability was considered in the context of controlling individual system components [17, 58–60, 65–69] and wide-area controls [70–72]. Some of the works [17, 58–60] are already discussed in the context of transient angle instability and some comments are valid for oscillatory angle instability consideration. A backstepping control was designed in [65] using Q-learning. Reference [66] proposed a decentralized synergetic controllers with varying parameters. Particle swarm optimization is first used to optimize parameters of the controllers followed by RL (Q-learning) to vary some of controller parameters to improve its performances. Work of [67] suggested Q-learning to design a controller for interline power controller to damp low-frequency oscillations (inter-area oscillations often present between two subsystems in a larger interconnection). In [68] a set of quadrature boosters devices, controlled by fuzzy controllers, are coordinated by Q-learning for oscillations damping. Fitted Q-iteration RL method was considered in [69] where a trajectory-based approach was designed as supplementary to existing controllers. In [70] a wide-area decentralized power system stabilizer was designed using Q-learning for damping both local and inter-area low frequency oscillations. Work presented in [71] also used Q-learning RL method and both physical and communication infrastructure brought uncertainties were addressed. Wide area control for oscillations damping presented in [72] used $TD(\lambda)$.

Reference [73]) dealt with Fault Induced Delayed Voltage Recovery (FIDVR) problem through DRL (DQN) where under-voltage load shedding was used as an emergency control. FIDVR problem is caused by a fault in transmission system resulting in slow voltage recovery in distribution system (the problem is connected to presence of reactive power loads in distribution and increased demand for this power due to reduced voltage causing slow recovery).

Frequency instability (a long-term instability caused by imbalance in generation and load demand after system survives faser transient processes) was

considered in [74] with an adaptive under-frequency load shedding as emergency control realized using $TD(\lambda)$ RL method. The approach is considered as a set of load controlling agents envisioned to be employed with strategic power infrastructure defense system presented in [75].

An approach for emergency control of cascading failures was presented in [76] where Q-learning was used to learn a control law modifying active power generation in order to control flows over transmission lines. The control is applied once te system experiences considered outages.

4.4. Restorative control

A multi-agent framework with Q-learning was suggested in [77] for restoration of power grid systems after being subjected to disturbances involving outages of lines and loss of generators. The controls included are generation, load and line's switches. Q-learning was also considered in [78] to develop optimal sequence for system restoration. This approach is based on a power flow-based model of cascading failure (consecutive outages of the system lines) and works in sequential fashion (power system components are bought back one by one through a sequence od controls). A multi-agent framework with Q-learning was also suggested in [79] for fault location, isolation, and restoration in electric power distribution systems. Q-learning was modified to capture interactions among RL-based agents through so-called Q-matrix. From control level point of view these controls belong to subsystem level.

4.5. Power system control-related considerations

As already emphasized, integration of new instrumentation technology, advanced communication infrastructures and powerful computation architectures allowed design of advanced controls in power system. However, this integration transformed modern power systems into cyber-physical systems and control-related aspects of these systems have to be fully considered. An important aspect is cyber-security since cyber attacks can make best designed controls to malfunction or degrade their performances. Several works dealt with this problem [80–88]. Reference [80] considered cyber-physical systems security from systems and control perspective in general, and shortly discussed possibilities to use RL and DRL to this purpose. Q-learning was proposed in [81] to analyze the transmission grid vulnerability under sequential topology attacks and identify critical attack sequences with consideration of physical system behaviors. A modified Q-learning (termed nearest sequence memory Q-learning) was adopted in [83] to evaluate threat imposed

by false data injection attack on voltage control of a power system. Test results revealed if even a few substations are attacked a voltage collapse with its consequences can happen in the system. Power system state estimation under cyber attacks was considered in [82, 84, 85]. Secure state estimation with assumption that measurements are sent over a wireless networks under jamming attacks was dealt in [82] and the antijamming game framework for secure state estimation using multi-agent reinforcement learning to determine optimal path against an intelligent attacker. Reference [85] considered secure state estimation with risk-averse transmission path selection method that is based on RL idea and demonstrated how proposed approach can improve secure state estimation robustness. DRL method (DQN) was proposed in [84] to defend against data integrity attacks in power systems state estimation. These types of cyber attacks are able to bypass the bad data detection mechanism in state estimation and make the system operator and controllers obtain the misleading states of system. In [86] the on-line attack detection problem was formulated as a partially observable Markov decision process (Markov property recovered through the use of a window of state history)and on-line detection algorithm using SARSA method was proposed for early cyber attacks detection. Recent work presented in [87] discusses te use of RL in a general framework of cognitive risk control for cyber attacks in smart grids. RL was proposed in [88] to evaluate false data injection attacks on automatic voltage control of power systems (in normal operating states). A Q-learning algorithm with nearest sequence memory is adopted for on-line learning of attacking strategy and optimal attack strategy is modelled as a partially observable Markov decision process. Based on kernel density estimation, a bad data detection and correction method were presented to mitigate the disruptive impacts of the attacks.

Another important aspect of modern power systems is that huge amount of data (termed big data) are available and analysis of these data can help improve performances of power system controls. Work presented in [89] suggested to integrate fuzzy cluster based analytical method, game theory and reinforcement learning to perform the security situational analysis for the smart grid.

Short-term load forecasting is of importance in any predictive control in electric power systems and a number of methods have been proposed so far. Work presented in [90] suggested RL (Q-learning) as an approach to choose most appropriate short-term load forecast method among those available. A Q-learning learns the optimal policy of selecting the best forecasting model

for the next time step, based on the model performance.

Reference [91] proposed the use of Q-learning with imitation and knowledge transfer for improved modelling of composite loads in electric power systems. In [92] a modification of Q-learning method (termed as enhanced RL, different from Q-learning since it records value function only for controls, not for state-control pair) was considered for determination of the equivalent of electric distribution system with due account of uncertainties of electricity generation from RES. This is related to the control since many controllers are designed using simulation models and improvement in component modelling yields more accurate control designs.

4.6. Observations

Based of the review of existing considerations of (D)RL for power system control and related problems the following observations are drawn:

- (D)RL was considered as a solution for electric power system control across all operating states and control levels (from local (device) to wide-area level). The considerations confirm potentials of (D)RL to solve these problems. However, all the considerations are research works and no practical implementation was reported.

- All considerations used simulation models of the system do design the controllers This is expected since hard to envision direct interaction of (D)RL with real-life electric power system due to exploration issue on such a vital infrastructure. This will be prevailing approach as long as some safety guarantees are not included in control design.

- Most of considerations are for controls in normal and emergency operating state and control-related problems while comparatively less considerations exists for preventive and restorative controls. Surprisingly low number of considerations exist for preventive control. Likely reason for this is the most controls of this type are formulated as static optimization problem.

- Prevailing RL method used is Q-learning (and its variant $Q(\lambda)$) followed by Fitted Q-iteration. A likely reason for this is success of these RL methods in other domains.

- DRL started being considered for electric power system control and related problems only recently. A likely reason is matured DRL methods

emerge also rather recently. This interest is increasing rapidly (as confirmed by several papers available on open repositories and "in press", reviewed in this paper). DQN is most used DRL method (again, its success in other domains is the main reason for it). An exception is work presented in [50] where DDPG offered better performances with respect to DQN.

- (D)RL was considered as single controller or in the context of multi-agent systems.

- Most of control-related considerations dealt with the problem of cyber-physical security of the system. This is not surprising since the problem is very important and its importance will increase in the context of Energy Internet (Internet of Things).

- All emergency control considerations relate to the emergency controls bringing a system from emergency to alert state. No considerations reported on the emergency control bringing a system from failure to restorative state.

- In general, there is a lack of efficient fusion of (D)RL models with control theory and practice in electric power systems. Few exceptions exist [59, 62, 68, 69] where RL was fused with the concept of control Lyapunov functions [62], model predictive control [59, 69] and fuzzy logic based control [68].

- In cases when the system is partially observable Markov decision problems, usual approach is to use history of states/controls to recover Markov property (see [62] where communication delays were handled through the use of states-controls history).

- Some considerations [53, 54] do not belong to any well-known RL method but are rather motivated by the spirit of RL and marked in this review as "Tailored".

- Embedding domain specific knowledge (in defining state space, control and reward) is crucial for a problem dimensionality reduction and accelerating learning.

5. Perspectives

(D)RL is a vibrant research field and new or improved existing methods emerge fast. It is reasonable to expect increased interests (from both research community and electric power system practitioners) to further consider (D)RL to solve control problems in future. The following are some future directions:

- Uisng (D)RL to control electric power system devices, in particular emerging ones, not considered previously. An example are energy storage devices allowing rapid and frequent charges/discharges such as supercapacitor, superconductive magnetic energy storage and flywheels [93]. In principle, these devices could be controlled in a similar way as dynamic braking resistor [61, 62].

- Revisiting existing RL considerations for electric power system control in the context of DRL, in particular for cases where the problem boils down to be partially observable Markov decision problem (see [62, 86] where history of states/controls were used to handle communication delays and recover Markov property at expense of increased dimensionality reasonably expected to be better handled by DRL).

- Considerations of (D)RL methods offering safe exploration. An approach in [62] proposed limiting admissible controls to stable ones and used the concept of control Lyapunov functions to this purpose. Other possibilities, in this respect, include the use of safe (D)RL (see [94, 95] and especially [96, 97] discussing safe exploration for controls, [98] for synthesis of RL and robust control with stability guarantees, and [99] for robust adversarial RL where a controller was trained in the presence of a destabilizing adversary applying disturbance to the system).

- Fusion with advanced methods coming from control theory and engineering (some example exists, see [55, 59, 62, 68, 69], for some future on control see [2–4]). Recent work presented in [100] is another example of combining RL and a model predictive control and is worth of considerations in electric power systems. This work, in line with the suggestions on combining RL and model predictive control [59, 69] shown, from control theoretic perspective, how RL methods can be used to tune parameters of economic model predictive controller and

how economic model predictive controller could be used as a new type of function approximator within RL. (D)RL could be used to coordinate existing controllers that ensure baseline properties (this is an interesting possibility since in existing electric power systems, especially large interconnections, the controls are not designed in a coordinated way and cause permanent oscillations in the system). Reference [101] offers some viable insights on this possibility. Moreover, expected increase in future electric power system operations uncertainties necessitate more deployment of robust control (see [2]). These controllers are designed on the worst-case basis but operate most of the time in non-worst-case situations and thus wasting control efforts. (D)RL could prove to be an appropriate approach to be used with robust controllers to tune their parameters for better overall performances. Another option would be to use robust RL [98] and robust adversarial RL [99].

- Many electric power system controls are designed through extensive simulations and (D)RL fits well to this kind of problems (an example is work presented in [102] where several parameters are computed through simulations for under-voltage load shedding (considered to be an expensive emergency control in electric power systems)) and the use of (D)RL could offer a viable solution for improvements and reduction of economic losses through fine tuning of the controller parameters. Similar observation holds true for so-called system integrity protection schemes design [103].

- As argued in [104], preventive control problems would be better formulated as multi-stage decision problem (particularly in presence of increased uncertainties) and it is reasonable to expect more (D)RL considerations in the future.

- The use of (D)RL methods to trade-off between preventive (open-loop) and corrective (closed-loop) controls in electric power systems [104]. Preventive controls are expensive and (D)RL could help decrease associated costs through learning the trade-off (an example potentially useful to consider was presented in [105]) for transient instability problem).

- The approach from [90] for short-term load forecasting (RL used to chose among a number of available forecasting methods at each step)

could be easily extended to equally important problem of electricity generation from RES forecasting (particularly solar and wind-based electricity generation). This is related to possible increased use of predictive controls in electric power systems in the future.

- Some methods coming from RL research sub-fields like hierarchical RL [106], preference-based RL [107] and imitative RL [108] are worth of considerations in electric power system controls. Some of them allow embedding preferences (somewhat related to embedding a domain knowledge) and accelerate RL (imitative learning particularly well-suited for multi-agent RL-based control) while hierarchical RL naturally fits many control problems. Three existing works considered imitative learning in electric power system control [42, 43, 91] where [42, 91] also considered knowledge transfer (a point to be considered more in the future). Bayesian RL permits embedding prior knowledge about controlled system and is worth considerations in electric power system controls (see [109] for a promising Bayesian RL approach).

- More considerations of other DRL methods (other than DQN since it cannot solve the problems with large continuous control space and where te optimal policy is stochastic) is expected. Reference [50] is a good example showing better performances of DDPG with respect to DQN for particular problem. Bayesian DRL is particularly interesting for future considerations. Reference [110] offers a good starting point. In addition, experience replay option in some DRL methods, if used with care, considerably improves performances of the methods (a good source on this subject, dealing with systems control, is reference [111]). DRL methods are not without the issues (particularly related to the convergence and sensitivity to involved parameters) [20], but huge undergoing research efforts will certainly offer solutions for the issues and increase interest in te use of DRL (interesting new results, in this respect, were presented in [112] with considerations of so-called "deadly triad" in RL: function approximation, bootstrapping, and off-policy learning).

- Integral RL is a popular RL algorithm among control theorists and engineers [23, 113] and is worth of more consideration for electric power system control (only one work considered the use of integral RL to this purpose [45]) in the future (integral RL offers some advantages, with

respect to other RL algorithms, such as scalability, higher efficiency and less open parameters).

- Further use of (D)RL in determining dynamic equivalent of electric distribution systems or external subsystems with high penetration of electricity generation from RES. References [91, 92] considered RL for these purposes (identification of composite load [91] and an equivalent of active distribution network [92]). Some successes in using deep learning for this purposes (an example is reference [114]) suggest that DRL could offer viable solutions to this problem.

- (D)RL considerations to design fault-tolerant controls since future uncertainties and increased complexity in electric power system structure are expected to experience inevitable failures in measurements, control actuators, etc. A good starting point is the work presented in [115].

- Energy Internet (Internet of Things) opens a number of possibilities for (D)RL considerations in this context (holds true also for cyber-physical systems since no clear distinction in the literature on these terms). Reference [116] discussed applications and challenges of DRL in this context and revealed opportunities to use DRL in all three layers of Internet of Things: perception layer (control of the physical system or its components), network layer (control of communications resources) and application layer (control of computation resources). Future considerations should take into account the use of blockchain technology in this context [117].

- Bringing (D)RL considerations to the attention of electric power system practitioners. A good starting point is embedding domain specific knowledge where the system experts could bring useful information for better use of the methods together with the use of interpretable (D)RL methods.

6. Conclusion

(D)RL considerations for electric power system control and related problems are reviewed in this paper focusing on journals publications and books. The considerations are presented as they relate to electric power system operating states and control level together with control-relevant ones. This review reveals:

- (D)RL offers viable solutions for many electric power system control problems across all its operating states. The considerations include different level of controls ranging from local to wide-area.

- Going back to important observation of [22] "...no methods that are guaranteed to work for all or even most problems, but there are enough methods to try on a given challenging problem with a reasonable chance that one or more of them will be successful in the end..." a suggestion is to try several (D)RL methods for an electric power system problem to be solved and chose the one showing best performances.

- Proliferation of smart grid technologies make electric power systems to become cyber-physical ones and due considerations, based on (D)RL, were already given to some issues these technologies bring in electricity sector.

In general, this review shows (D)RL offers a panel of promising methods to be considered in design of electric power system controllers. It is reasonable to expect more considerations due to expected future changes electric power systems (increased uncertainties and complexity). Further research is strongly encouraged together with due consideration of bringing it to the attention of electric power system practitioners. This review focused only on the works published in the journals and books. Conference papers and (D)RL considerations for electric power system decision problems (scheduling, market decisions and energy management of microgrids and buildings) are not included and they are left for a possible future extension of this review. Approaches known as approximate and adaptive dynamic programming were also considered in electric power system controls. Only some of these approaches belong to RL but not reviewed in this paper to avoid confusions (these approaches and RL are often used interchangeably in the literature) and left for a possible future extensions.

References

[1] US-DoE, Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations, Tech. rep., US Department of Energy, US-Canada Power System Outage Task Force (2004).

[2] F. Lamnabhi-Lagarrigue, A. Annaswamy, S. Engell, A. Isaksson, P. Khargonekar, R. M. Murray, H. Nijmeijer, T. Samad, D. Tilbury, P. Van den Hof, Systems and control for the future of humanity, research agenda: Current and future roles, impact and grand challenges, Annual Reviews in Control 43 (2017) 1–64.

[3] A. Annaswamy, (Ed.), Vision for smart grid control: 2030 and beyond, Tech. rep., IEEE Standards Association (2013).

[4] T. Samad, A. M. Annaswamy, Controls for smart grids: Architectures and applications, Proceedings of the IEEE 105 (2017) 2244–2261.

[5] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, 1998.

[6] L. Busoniu, R. Babuska, B. D. Schutter, D. Ernst, Reinforcement learning and dynamic programming using function approximators, CRC Press, Boca Raton, 2010.

[7] V. Fancois-Lavet, P. Henderson, R. Islam, M. G. Bellemare, J. Pineau, An introduction to deep reinforcement learning, Foundations and Trends in Machine Learning 11 (2019) 219–354.

[8] D. P. Bertsekas, Dynamic programming and optimal control, Vols I and II, Athena Scientific, Boston, 1995.

[9] M. Glavic, R. Fonteneau, D. Ernst, Reinforcement learning for electric power system decision and control: Past considerations and perspectives, IFAC PapersOnLine 50-1 (2017) 6918–6927.

[10] US-DoE, An assessment of energy technologies and research opportunities, Tech. rep., US Department of Energy (2015).

[11] G. Bedi, G. K. Venayagamoorthy, R. Singh, R. R. Brooks, K. C. Wang, Review of internet of things (IoT) in electric power and energy systems, IEEE Internet of Things Journal 5 (2018) 847–870.

[12] K. Wang, J. Yu, Y. Yu, Y. Qian, D. Zeng, S. Guo, Y. Xiang, J. Wu, A survey on energy internet: Architecture, approach, and emerging technologies, IEEE Systems Journal 12 (2018) 2403–2416.

[13] T. E. DyLiacco, Real-time computer control of power systems, Proceedings of the IEEE 62 (1974) 884–891.

[14] K. R. Padiyar (Ed.), Power System Dynamics, Stability and Control, BS Publications, Bangalore, 2008.

[15] K. Arulkumaran, M. P. Desienroth, M. Brundage, A. A. Bharath, A brief survey of deep reinforcement learning, IEEE Signal Processing Magazine 34 (2017) 26–38.

[16] W. B. Powell, S. Meisel, Tutorial on stochastic optimization in energyPart I: Modeling and policies, IEEE Transactions on Power Systems 31 (2016) 1459–1467.

[17] D. Ernst, M. Glavic, L. Wehenkel, Power systems stability control: Reinforcement learning framework, IEEE Transactions on Power Systems 19 (2004) 427–435.

[18] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 426–444.

[19] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks 61 (2015) 85–117.

[20] L. Busoniu, T. de Bruin, D. Tolic, J. Kober, I. Palunko, Reinforcement learning for control: Performance, stability, and deep approximators, Annual Reviews in Control 46 (2018) 8–28.

[21] R. Kamalapurkar, P. Walters, J. Rosenfeld, W. Dixon, Reinforcement Learning for Optimal Feedback Control: A Lyapunov-Based Approach, Springer, Cham, Switzerland, 2018.

[22] D. P. Bertsekas, Reinforcement Learning and Optimal Control, Athena Scientific, Nashua, NH, USA, 2019.

[23] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, F. L. Lewis, Optimal and autonomous control using reinforcement learning: A survey, IEEE Transactions on Neural Networks and Learning Systems 29 (2018) 2042–2062.

[24] P. P. Khargonekar, M. A. Dahleh, Advancing systems and control research in the era of ML and AI, Annual Reviews in Control 45 (2019) 1–4.

[25] D. Wu, X. Zheng, D. Kalathil, L. Xie, Nested reinforcement learning based control for protective relays in power distribution systems, arXiv:1906.10815v1, Accessed August, 2019 (2019) 1–8.

[26] C. Wei, Z. Zhang, W. Qiao, L. Qu, Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems, IEEE Transactions on Industrial Electronics 62 (2015) 6360–6370.

[27] A. Bag, B. Subudhi, P. Ray, An adaptive variable leaky least mean square control scheme for grid integration of a PV system, IEEE Transactions on Sustainable Energy In Press.

[28] C. Wei, Z. Zhang, W. Qiao, L. Qu, An adaptive network-based reinforcement learning method for MPPT control of PMSG wind energy conversion systems, IEEE Transactions on Power Electronics 31 (2016) 7837–7848.

[29] X. Zhang, S. Li, T. He, B. Yang, T. Yu, H. Li, L. Jiang, L. Sun, Memetic reinforcement learning based maximum power point tracking design for PV systems under partial shading condition, Energy 174 (2019) 1079–1090.

[30] E. Anderlini, D. I. M. Forehand, P. Stansell, Q. Xiao, M. Abusara, Control of a point absorber using reinforcement learning, IEEE Transactions on Sustainable Energy 7 (2016) 1681–1690.

[31] E. Anderlini, D. I. M. Forehand, E. Bannon, M. Abusara, Control of a realistic wave energy converter model using least-squares policy iteration, IEEE Transactions on Sustainable Energy 8 (2017) 1618–1628.

[32] J. Sun, Z. Zhu, H. Li, Y. Chai, G. Qi, H. Wang, Y. H. Hu, An integrated critic-actor neural network for reinforcement learning with application of DERs control in grid frequency regulation, International Journal of Electrical Power and Energy Systems 111 (2019) 286–299.

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

[34] H. Xi, A. D. Dominguez-Garcia, P. W. Sauer, Optimal tap setting of voltage regulation transformers using batch reinforcement learnings, arXiv:1807.10997v2, Accessed August, 2019 (2018) 1–8.

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

[36] T. P. I. Ahamed, P. S. N. Rao, P. S. Sastry, Ha neural network based automatic generation controller design through reinforcement learning, International Journal of Emerging Electric Power Systems 6 (2006) 1–31.

[37] F. Daneshfar, H. Bevrani, Load-frequency control: a GA-based multi-agent reinforcement learning, IET Generation, Transmission, Distribution 4 (2010) 13–26.

[38] T. Yu, B. Zhou, K. W. Chan, L. Chen, B. Yang, Stochastic optimal relaxed automatic generation control in non-markov environment based on multi-step $Q(\lambda)$ learning, IEEE Transactions on Power Systems 26 (2011) 1272–1282.

[39] T. Yu, H. Z. Wang, B. Zhou, K. W. Chen, J. Tang, Multi-agent correlated equilibrium $Q(\lambda)$ learning for coordinated smart generation control of interconnected power grids, IEEE Transactions on Power Systems 27 (2012) 373–380.

[40] T. Yu, X. S. Zhang, B. Zhou, K. V. Chan, Hierarchical correlated q-learning for multi-layer optimal generation command dispatch, International Journal of Electric Power and Energy Systems 78 (2016) 1–12.

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

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

[43] T. Yu, B. Zhou, K. W. Chan, Y. Yuan, B. Yang, Q. Wu, $R(\lambda)$ imitation learning for automatic generation control of interconnected power grids, Automatica 48 (2012) 2130–2136.

[44] Q. Huang, R. Huang, W. Hao, J. Tan, F. R, Z. Huang, Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action searchning, IEEE Transactions on Power Systems 34 (2019) 1653–1656.

[45] M. Abouheaf, W. Gueaieb, A. Sharaf, Load frequency regulation for multi-area power system using integral reinforcement learning, IET Generation, Transmission, Distribution In Press.

[46] L. Yin, T. Yu, L. Zhou, L. Huang, X. Zhang, B. Zheng, Artificial emotional reinforcement learning for automatic generation control of large-scale interconnected power grids, IET Generation, Transmission, Distribution 11 (2017) 2305–2313.

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

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

[49] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis, Real-time voltage control using deep reinforcement learnings, arXiv:1904.09374v1, Accessed August, 2019 (2019) 1–9.

[50] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, Z. Yi, Deep-reinforcement-learning-based autonomous voltage control for power grid operations, IEEE Transactions on Power Systems In Press.

[51] F. Ruelens, B. J. Claessens, P. Vrancx, F. Spiessens, G. Deconinck, Direct load control of thermostatically controlled loads based on sparse observations using deep reinforcement learning, arXiv:1707.08553.v1, accessed August, 2019 (2017) 1–8.

[52] B. J. Claessens, P. Vrancxs, F. Ruelens, Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control, IEEE Transactions on Smart Grid 9 (2018) 3259–3269.

[53] J. Duan, Z. Li, D. Shi, C. Lin, Z. Wang, Reinforcement-learning-based optimal control for hybrid energy storage systems in hybrid AC/DC microgrids, IEEE Transactions on Industrial Informatics In Press.

[54] M. Bagheri, V. Nurmanova, O. Abedinia, M. S. Naderi, Enhancing power quality in microgrids with a new online control strategy for DSTATCOM using reinforcement learning algorithm, IEEE Access 6 (2018) 38986–38996.

[55] S. S. Khorramabadi, A. Bakhshai, Intelligent control of grid-connected microgrids: An adaptive critic-based approach, IEEE Journal of Emerging and Selected Topics in Power Electronics 3 (2015) 493–504.

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

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

[58] D. Ernst, M. Glavic, P. Geurst, L. Wehenkel, Approximate value iteration in the reinforcement learning context. Application to electrical power system control, International Journal of Emerging Electrical Power Systems 3 (2005) 1–37.

[59] D. Ernst, M. Glavic, F. Capitanescu, L. Wehenkel, Reinforcement learning versus model predictive control: A comparison on a power system problem, IEEE Transactions on Systems, Man, and Cybernetic: Part B 39 (2009) 517–529.

[60] M. Glavic, D. Ernst, L. Wehenkel, Combining a stability and a performance-oriented control in power systems, IEEE Transactions on Power Systems 20 (2005) 525–526.

[61] M. Glavic, D. Ernst, L. Wehenkel, A reinforcement learning based discrete supplementary control for power system transient stability enhancement, Engineering Intelligent Systems for Electrical Engineering and Communications 13 (2005) 81–88.

[62] M. Glavic, Design of a resistive brake controller for power system stability enhancement using reinforcement learning, IEEE Transactions on Control Systems Technology 13 (2005) 743–751.

[63] R. Yousefian, R. Bhattarai, S. Kamalasadan, Transient stability enhancement of power grid with integrated wide area control of wind farms and synchronous generators, IEEE Transactions on Power Systems 32 (2017) 4818–4831.

[64] R. Yousefian, S. Kamalasadan, Energy function inspired value priority based global wide-area control of power grid, IEEE Transactions on Smart Grid 9 (2018) 552–563.

[65] A. Karimi, S. Eftekharnejad, A. Feliachi, Reinforcement learning based backstepping control of power system oscillations, Electric Power Systems Research 79 (2009) 1511–1520.

[66] T. Ademoye, A. Feliachi, Reinforcement learning tuned decentralized synergetic control of power systems, Electric Power Systems Research 86 (2012) 34–40.

[67] A. Younesi, H. Shayeghi, M. Moradzadeh, Application of reinforcement learning for generating optimal control signal to the IPFC for damping of lowfrequency oscillations, International Transactions on Electric Energy Systems 28 (2018) 1–23.

[68] B. H. Li, Q. H. Wu, Learning coordinated fuzzy logic control of dynamic quadrature boosters in multimachine power systems, IEE Generation, Transmission, Distribution 146 (1999) 577–585.

[69] D. Wang, M. Glavic, L. Wehenkel, Trajectory-based supplementary damping control for power system electromechanical oscillations, IEEE Transactions on Power Systems 29 (2014) 2835–2845.

[70] R. Hadidi, B. Jeyasurya, Reinforcement learning based real-time wide-area stabilizing control agents to enhance power system stability, IEEE Transactions on Smart Grid 4 (2013) 489–497.

[71] J. Duan, H. Xu, W. Liu, Q-learning-based damping control of wide-area power systems under cyber uncertainties, IEEE Transactions on Smart Grid 9 (2018) 6408–6418.

[72] R. Yousefian, S. Kamalsadan, Design and real-time implementation of optimal power system wide-area system-centric controller based on temporal difference learning, IEEE Transactions on Industry Applications 52 (2016) 395–406.

[73] Q. Huang, R. Huang, W. Hao, J. Tan, F. R, Z. Huang, Adaptive power system emergency control using deep reinforcement learning, IEEE Transactions on Smart Grid In Press.

[74] J. Jung, C. C. Liu, S. L. Tanimoto, V. Vittal, Adaptation in load shedding under vulnerable operating conditions, IEEE Transactions on Power Systems 17 (2002) 1199–1205.

[75] C. C. Liu, J. Jung, G. T. Heydt, V. Vittal, A. Phadke, The strategic power infrastructure defense (SPID) system. A conceptual design, IEEE Control Systems Magazine 20 (2000) 40–52.

[76] R. Belkacemi, A. A. Babalola, S. Zarrabian, Real-time cascading failures prevention through MAS algorithm and immune system reinforcement learning, Electric Power Components and Systems 45 (2017) 505–519.

[77] D. Ye, M. Zhang, D. Sutanto, A hybrid multiagent framework with Q-learning for power grid systems restoration, IEEE Transactions on Power Systems 26 (2011) 2434–2441.

[78] J. Wu, B. Fang, J. Fang, X. Chen, C. K. Tse, Sequential topology recovery of complex power systems based on reinforcement learning, Physica A: Statistical Mechanics and its Applications 535 (2019) 1–13.

[79] M. J. Ghorbani, M. A. Choudhry, A. Feliachi, A multiagent design for power distribution systems automation, IEEE Transactions on Smart Grid 7 (2016) 329–339.

[80] S. M. Dibaji, M. Pirani, D. B. Flamholz, M. A. A, K. H. Johansson, A. Chakrabortty, A systems and control perspective of CPS security, Annual Reviews in Control 47 (2019) 394–411.

[81] J. Yan, H. He, X. Zhong, Y. Tang, Q-learning based vulnerability analysis of smart grid against sequential topology attacks, IEEE Transactions on Information Forensics and Security 12 (2017) 200–210.

[82] J. He, C. Chen, S. Zhu, B. Yang, X. Guan, Antijamming game framework for secure state estimation in power systems, IEEE Transactions on Industrial Informatics 15 (2019) 2628–2637.

[83] Z. Wang, Y. Chen, F. Liu, Y. Xia, X. Zhang, Power system security under false data injection attacks with exploitation and exploration based on reinforcement learning, IEEE Access 6 (2018) 48785–48796.

[84] D. An, Q. Yang, W. Liu, Y. Zhang, Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach, IEEE Access 7 (2019) 110835–110845.

[85] Y. He, C. Chen, S. Zhu, B. Yang, X. Guan, Risk-averse transmission path selection for secure state estimation in power systems, IEEE Internet of Things Journal 6 (2019) 3121–3131.

[86] M. N. Kurt, O. Ogundijo, C. Li, X. Wang, Online cyber-attack detection in smart grid: A reinforcement learning approach, IEEE Transactions on Smart Grid 10 (2019) 5174–5185.

[87] M. I. Oozeer, S. Haykin, Cognitive risk control for mitigating cyber-attack in smart grid, IEEE Access In Press (2019) 1–21.

[88] Y. Chen, S. Huang, F. Liu, Z. Wang, X. Sun, Q. Yang, W. Liu, Y. Zhang, Evaluation of reinforcement learning-based false data injection attack to automatic voltage control, IEEE Transactions on Smart Grid 10 (2019) 2158–2169.

[89] J. Wu, K. Ota, M. Dong, J. Li, H. Wang, Big data analysis based security situational awareness for smart grid, IEEE Transactions on Big Data 4 (2018) 408–417.

[90] C. Feng, M. Sun, J. Zhang, Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection, IEEE Transactions on Smart Grid In Press.

[91] J. Xie, Z. Ma, S. Ma, Z. Wang, Data-driven based method for power system time-varying composite load modeling, arXiv:1905.02688v1, Accessed August, 2019 (2019) 1–8.

[92] X. Shang, Z. Li, J. Zheng, Q. H. Wu, Equivalent modeling of active distribution network considering the spatial uncertainty of renewable energy resources, International Journal of Electrical Power and Energy Systems 112 (2019) 83–91.

[93] M. Farhadi, O. Mohammed, Energy storage technologies for high-power applications, IEEE Transactions on Industry Applications 52 (2015) 1953–1961.

[94] J. Garcia, F. Fernandez, A comprehensive survey on safe reinforcement learning, Journal of Machine Learning Research 16 (2015) 1437–1480.

[95] J. Fan, W. Li, Safety-guided deep reinforcement learning via online Gaussian process estimation, arXiv:1903.02526v2, Accessed September, 2019 (2019) 1–13.

[96] T. Mannucci, E. J. van Kempen, C. de Viser, Q. Chu, Safe exploration algorithms for reinforcement learning controllers, IEEE Transactions on Neural Networks and Learning Systems 29 (2018) 1069–1081.

[97] M. Jin, J. Lavaei, Stability-certified reinforcement learning: A control-theoretic perspective, arXiv:1810.11505v1, Accessed August, 2019 (2018) 1–30.

[98] R. M. Kretchmar, P. M. Young, C. W. Anderson, D. C. Hittle, M. L. Anderson, C. C. Delnero, Robust reinforcement learning control with static and dynamic stability, International Journal of Robust and Nonlinear Control 11 (2001) 1469–1500.

[99] L. Pinto, J. Davidson, R. Sukthankar, A. Gupta, Robust adversarial reinforcement learning, arXiv:1703.02702v1, Accessed September, 2019 (2017) 1–10.

[100] S. Gros, M. Zanon, Data-driven economic NMPC using reinforcement learning, IEEE Transactions on Automatic Control In Press.

[101] E. Abramova, L. Dickens, D. Kuhn, A. Faisal, RLOC: Neurobiologically inspired hierarchical reinforcement learning algorithm for continuous control of nonlinear dynamical systems, arXiv:1903.03064v1, Accessed September, 2019 (2019) 1–33.

[102] B. Otomega, M. Glavic, T. Van Cutsem, Distributed undervoltage load shedding, IEEE Transactions on Power Systems 22 (2007) 2283–2284.

[103] V. Madani, D. Novosel, S. Horowitz, M. Adamiak, J. Amantegui, D. Karlsson, S. Imai, A. Apostolov, IEEE PSRC report on global industry experiences with system integrity protection schemes (SIPS), IEEE Transactions on Power Systems 25 (2010) 2143–2155.

[104] L. Wehenkel, M. Glavic, P. Geurts, D. Ernst, Automatic learning of sequential decision strategies for dynamic security assessment and controls, in: IEEE PES General Meeting, IEEE, 2006, pp. 1–6.

[105] D. Ruiz-Vega, M. Glavic, D. Ernst, Transient stability emergency control combining open-loop and closed-loop techniques, in: IEEE PES General Meeting, IEEE, 2003, pp. 2053–2059.

[106] A. G. Barto, S. Mahadevan, Recent advances in hierarchical reinforcement learning, Discrete Events Dynamic Systems: Theory and Applications 13 (1999) 41–77.

[107] C. Wirth, R. Akrour, G. Neumann, J. Furnkranz, A survey of preference-based reinforcement learning methods, Journal of Machine Learning Research 18 (2017) 1–46.

[108] B. Price, C. Boutilier, Accelerating reinforcement learning through implicit imitation, Journal of Artificial Intelligence Research 19 (2003) 569–629.

[109] E. D. Klenske, P. Hennig, Dual control for approximate Bayesian reinforcement learning, Journal of Machine Learning Research 17 (2016) 1–30.

[110] K. Azizzadenesheli, A. Anandkumar, Efficient exploration through Bayesian deep Q-networks, arXiv:1802.04412v4, Accessed September, 2019 (2019) 1–40.

[111] T. de Bruin, J. Kober, K. Tuyls, R. Babuska, Experience selection in deep reinforcement learning for control, Journal of Machine Learning Research 19 (2018) 1–56.

[112] K. De Asis, A. Chan, S. Pitis, R. S. Sutton, D. Graves, Fixed-horizon temporal difference methods for stable reinforcement learning, arXiv:1909.03906v1, Accessed September, 2019 (2019) 1–16.

[113] E. A. Theodorou, J. Buchli, S. Schaal, A generalized path integral control approach to reinforcement learning, Journal of Machine Learning Research 11 (2010) 3137–3181.

[114] C. Zheng, S. Wang, Y. Liu, C. Liu, W. Xie, C. Fang, S. Liu, A novel equivalent model of active distribution networks based on LSTM, IEEE Transactions on Neural Networks and Learning Systems 30 (2019) 2611–2624.

[115] Z. Wang, L. Liu, H. Zhang, G. Xiao, Fault-tolerant controller design for a class of nonlinear MIMO discrete-time systems via online reinforcement learning algorithm, IEEE Transactions on Systems, Man, and Cybernetics: Systems 46 (2016) 611–622.

[116] L. Lei, Y. Tan, S. Liu, K. Zheng, X. Shen, Deep reinforcement learning for autonomous internet of things: Model, applications and challenges, arXiv:1907.09059v1, Accessed August, 2019 (2019) 1–23.

[117] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung, M. Song, Performance optimization for blockchain-enabled industrial internet of things (IIoT) systems: A deep reinforcement learning approach, IEEE Transactions on Industrial Informatics 15 (2019) 3559–3570.