Annual Reviews in Control 48 (2019) 22–35

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Review article

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

Mevludin Glavic

Independent Researcher/Consultant, Maka Dizdara 66, 75000 Tuzla, Bosnia and Herzegovina

ARTICLE INFO

Article history:
Received 28 September 2019
Accepted 30 September 2019
Available online 12 October 2019

Keywords:
Electric power system
Reinforcement learning
Deep reinforcement learning
Control
Control-related problems

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.

Contents

1. Introduction . . . 22
2. Power system structure and operating states . . . 23
3. (Deep) Reinforcement learning: a short introduction and considerations for control problems in general . . . 25
3.1. Dynamic programming and optimal control . . . 25
3.2. Reinforcement learning . . . 25
3.3. Deep reinforcement learning . . . 26
3.4. (D)RL for control problems in general: a short review of recent works . . . 26
4. A short review of (D)RL considerations for electric power system control and control-related problems . . . 27
4.1. Control in normal operating state . . . 27
4.2. Preventive control . . . 29
4.3. Emergency control . . . 29
4.4. Restorative control . . . 30
4.5. Power system control-related considerations . . . 30
4.6. Observations . . . 31
5. Perspectives . . . 31
6. Conclusion . . . 33
References . . . 33

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 and societal costs. An example is US-Canada power system outage of August 14, 2003 (US-DoE, 2004) 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

E-mail address: glavic@montefiore.ulg.ac.be

https://doi.org/10.1016/j.arcontrol.2019.09.008
1367-5788/© 2019 Elsevier Ltd. All rights reserved.

electrical vehicles) and it is reasonable to expect the costs of an outage such as US-DoE (2004) 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 Annaswamy, 2013; Lamnabhi-Lagarrigue et al., 2017; Samad & Annaswamy, 2017 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 (Busoniu, Babuska, Schutter, & Ernst, 2010; Fancois-Lavet, Henderson, Islam, Bellemare, & Pineau, 2019; Sutton & Barto, 1998). In RL and DRL setting 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 (Bertsekas, 1995; Busoniu et al., 2010; Sutton & Barto, 1998).

Some of considerations are already reviewed in Glavic, Fonteneau, and Ernst (2017). 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.

The trends driving the transformation are summarized as follows (US-DoE, 2015):

- 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 transmission 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 (Bedi, Venayagamoorthy, Singh, Brooks, & Wang, 2018) where it often comes

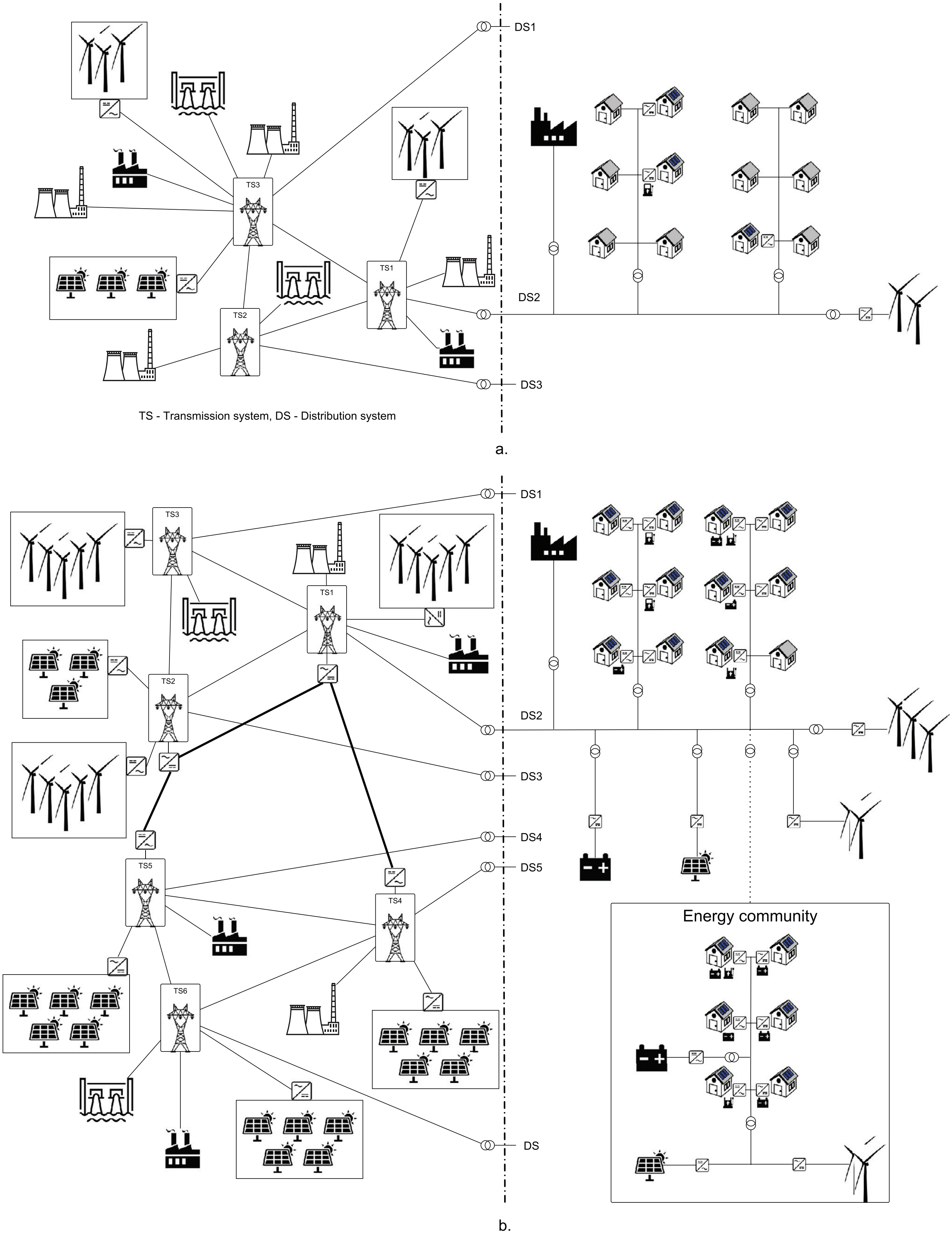

Fig. 1. Structure of power system: present (a) and future (b).

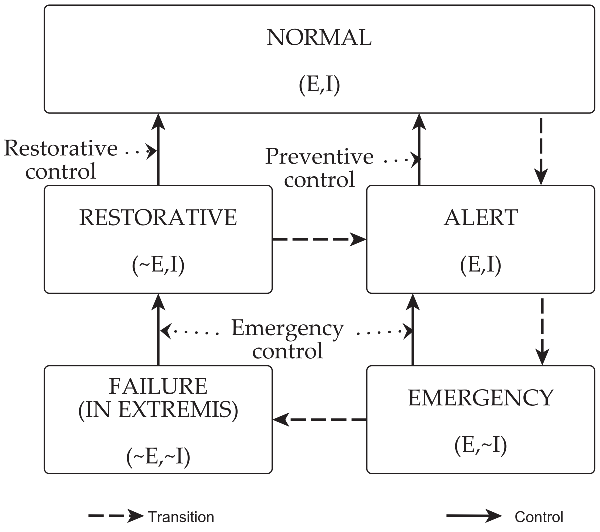

Fig. 2. Power system operating states (adopted from Padiyar, 2008).

under term Energy Internet (Wang et al., 2018). The widely accepted classification of electric power system operating states is the one introduced in DyLiacco (1974). Fig. 2 illustrates five operating states as defined in DyLiacco (1974) and adapted in Padiyar (2008).

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).

Fig. 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 (Bertsekas, 1995) 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 Bertsekas (1995), Sutton and Barto (1998), Busoniu et al. (2010), Fancois-Lavet et al. (2019) and Arulkumaran, Desienroth, Brundage, and Bharath (2017)

3.1. Dynamic programming and optimal control

Powell and Meisel (2016) 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 Ernst, Glavic, and Wehenkel (2004), 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 $r(x, u) \le B$ is a *reward* function, $\gamma \in [0, 1]$ a discount factor, and $u_{\{t\}} = (u_0, u_1, u_2, \ldots)$ a sequence of control actions applied to the system.

The value function is the solution of the Bellman equation (Bertsekas, 1995),

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

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,

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

Eq. (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 (Bertsekas, 1995; Busoniu et al., 2010; Sutton & Barto, 1998). 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 (Bertsekas, 1995; Busoniu et al., 2010; Sutton & Barto, 1998) can be used to estimate solution of this problem. Optimal control policy is now defined by,

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

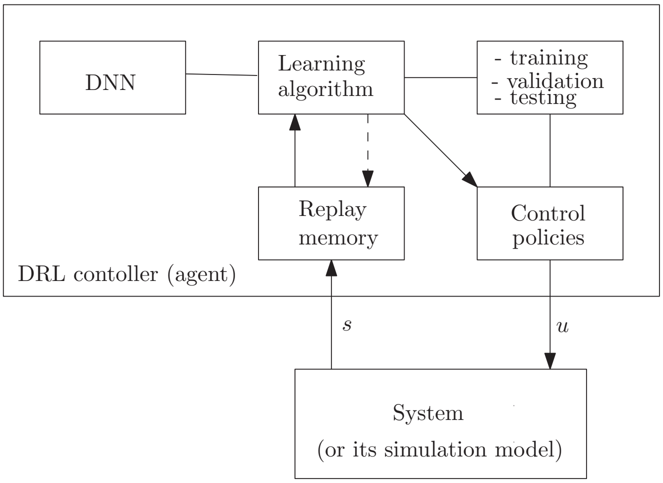

Fig. 3. A general framework of DRL (adopted from Fancois-Lavet et al., 2019 and slightly modified).

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 Busoniu et al. (2010); Sutton and Barto (1998) 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 (LeCun, Bengio, & Hinton, 2015) and in particular deep learning for neural networks (Schmidhuber, 2015) (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) (Arulkumaran et al., 2017).

A general DRL framework is illustrated in Fig. 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 (Fancois-Lavet et al., 2019).

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 (Fancois-Lavet et al., 2019):

- 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 Arulkumaran et al., 2017; Fancois-Lavet et al., 2019 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 Busoniu, de Bruin, Tolic, Kober, and Palunko (2018) 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 (Bertsekas, 2019; Kamalapurkar, Walters, Rosenfeld, & Dixon, 2018), a recent survey paper (Kiumarsi, Vamvoudakis, Modares, & Lewis, 2018), and a vision paper (Khargonekar & Dahleh, 2019) (not listed in Busoniu et al., 2018 since not available at the time when Busoniu et al., 2018 was prepared). Monograph Bertsekas (2019), 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 Bertsekas (2019) 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 Kamalapurkar et al. (2018) 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). Kiumarsi et al. (2018) 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 Khargonekar and Dahleh (2019). A conclusion that can be drawn from Khargonekar and Dahleh (2019) is the future systems and

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

Control	Reference(s)
Normal	Wu et al. (2019a), Wei et al. (2015), Bag et al. (2019), Wei et al. (2016), Zhang et al. (2019), Anderlini et al. (2016), Anderlini et al. (2017), Sun et al. (2019), Abouheaf et al. (2018), Xi et al. (2018), Ahamed et al. (2002), Ahamed et al. (2006), Daneshfar and Bevrani (2010), Yu et al. (2011), Yu, Wang, et al. (2012), Yu et al. (2016), Wang et al. (2019), Zhang et al. (2018), Yu et al. (2012), Huang et al. (2019b), Abouheaf et al. (2019), Yin et al. (2017), Vlachogiannis and Hatziargyriou (2004), Xu et al. (2012), Yang et al. (2019), Duan et al. (2019b), Ruelens et al. (2017), Claessens et al. (2018), Duan, Li, et al. (2019a), Bagheri et al. (2018), Khorramabadi and Bakhshai (2015), Wu et al. (2010)
Preventive	Zarabbian et al. (2016)
Emergency	Ernst et al. (2004), Ernst et al. (2005), Ernst et al. (2009), Glavic et al. (2005a), Glavic et al. (2005b), Glavic (2005), Yousefian et al. (2017), Yousefian and Kamalasadan (2018), Karimi et al. (2009), Ademoye and Feliachi (2012), Younesi et al. (2018), Li and Wu (1999), Wang et al. (2014), Hadidi and Jeyasurya (2013), Duan et al. (2018), Yousefian and Kamalsadan (2016), Huang et al. (2019a), Jung et al. (2002), Liu et al. (2000), Belkacemi et al. (2017)
Restorative	Ye et al. (2011), Wu et al. (2019b), Ghorbani et al. (2016)
Control-related considerations	Dibaji et al. (2019), Yan et al. (2017), He et al. (2019a), Wang, Chen, et al. (2018), An et al. (2019), He et al. (2019b), Kurt et al. (2019), Oozeer and Haykin (2019), Chen et al. (2019), Wu et al. (2018), Feng et al. (2019)

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)	Wu et al. (2019a)
	MPPT	Q-learning	Wei et al. (2015), Bag et al. (2019), Wei et al. (2016), Zhang et al. (2019), Anderlini et al. (2016)
		Least-squares policy iteration	Anderlini et al. (2017)
	Frequency regulation	Q-learning	Sun et al. (2019), Abouheaf et al. (2018)
	Voltage	Q-learning	Xi et al. (2018)
Subsystem	AGC/LFC	Q-learning	Ahamed et al. (2002), Ahamed et al. (2006), Daneshfar and Bevrani (2010), Wang et al. (2019), Zhang et al. (2018)
		$Q(\lambda)$	Yu et al. (2011)
		Correlated $Q(\lambda)$	Yu, Wang, et al. (2012)
		Correlated Q	Yu et al. (2016)
		$R(\lambda)$	Yu et al. (2012)
		DRL (DQN)	Huang et al. (2019b)
		Integral RL	Abouheaf et al. (2019)
		Emotional RL	Yin et al. (2017)
	Voltage	Q-learning	Vlachogiannis and Hatziargyriou (2004), Xu et al. (2012)
		DRL (DQN)	Yang et al. (2019)
		DRL (DQN/DDPG)	Duan et al. (2019b)
	Load control	DRL (DQN)	Ruelens et al. (2017), Claessens et al. (2018)
Microgrid	Transient energy storage performances	Tailored	Duan, Li, et al. (2019a)
	Power quality control	Tailored	Bagheri et al. (2018)
	Active/Reactive generation power	Tailored	Khorramabadi and Bakhshai (2015)
Household	Parameter tuning	$Q(\lambda)$	Wu et al. (2010)

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.

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.

Wu, Zheng, Kalathil, and Xie (2019a), 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 Wei, Zhang, Qiao, and Qu (2015, 2016), Bag, Subudhi, and Ray (2019), Zhang et al. (2019) dealt with Maximum Power Point Tracking (MPPT) control for wind energy (Wei et al., 2015; 2016) and photo-voltaic electricity generation (Bag et al., 2019; Zhang et al., 2019). They are local controls acting on individual wind or photo-voltaic sources so that maximum electricity is generated. Q-learning was used in Wei et al. (2015) for variable-speed while Wei, Zhang, Qiao, and Qu (2016) 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 Zhang et al. (2019) for MPPT through memetic computing incorporated in RL (Q-learning). Bag et al. (2019) 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 Anderlini, Forehand, Stansell, Xiao, and Abusara (2016) 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 Anderlini, Forehand, Bannon, and Abusara (2017) 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.

Sun et al. (2019) and Abouheaf, Gueaieb, and Sharaf (2018) 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 Sun et al. (2019). The on-line controller based on a policy iteration reinforcement learning paradigm along with an adaptive actor-critic technique was considered in Abouheaf et al. (2018) for wind turbines with doubly fed induction generators.

Q-learning was proposed in Xi, Dominguez-Garcia, and Sauer (2018) 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 Ahamed, Rao, and Sastry (2002), Daneshfar and Bevrani (2010), Yu, Wang, Zhou, Chen, and Tang (2012); Yu, Zhou, Chan, Chen, and Yang (2011); Yu et al. (2012), Wang, Lei, Zhang, Peng, and Jiang (2019), Zhang, Yu, Pan, Yang, and Bao (2018), Huang et al. (2019b), Abouheaf, Gueaieb, and Sharaf (2019), Yin et al. (2017) (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 Ahamed et al. (2002) and designed using Q-learning. The same problem as in Ahamed et al. (2002) was investigated in Ahamed, Rao, and Sastry (2006) 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 Daneshfar and Bevrani (2010) with genetic algorithms to tune controller parameters. Work presented in Yu et al. (2011) suggested single AGC controller based on multi-step $Q(\lambda)$ method while Yu, Wang, et al. (2012) suggested the use of correlated equilibrium $Q(\lambda)$ within a multi-agent setting (similar approach was proposed in Yu, Zhang, Zhou, and Chan (2016) 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 Wang et al. (2019). Zhang et al. (2018) 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. Yu et al. (2012) 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 (Sutton & Barto, 1998) similar to Q-learning. Work presented in Huang et al. (2019b) 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 (Kiumarsi et al., 2018) for load frequency regulation in multi-area electric power systems was suggested in Abouheaf et al. (2019). Emotional RL was proposed in Yin et al. (2017) 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 Vlachogiannis and Hatziargyriou (2004) and Xu, Zhang, Liu, and Ferrese (2012) and DRL in Yang, Wang, Sadeghi, and Giannakis (2019) and Duan et al. (2019b). Vlachogiannis and Hatziargyriou (2004) and Xu et al. (2012) 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) to controls (computed off-line). This is achieved though the formulation of constrained power flow problem as a multi-stage decision problem (Vlachogiannis & Hatziargyriou, 2004) and the use of an average consensus algorithm within a multi-agent RL-based framework (Xu et al., 2012). Yang et al. (2019) proposed a two timescale 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 Duan et al. (2019b) 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 Ruelens, Claessens, Vrancx, Spiessens, and Deconinck (2017) and Claessens, Vrancxs, and Ruelens (2018) using DRL to this purposes. Ruelens et al. (2017) considered an approach to find a near-optimal sequence of decisions

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	Ernst et al. (2004), Glavic (2005), Glavic et al. (2005a,b)
		Fitted Q-iteration	Ernst et al. (2009, 2005)
		DRL (DQN)	Huang et al. (2019a)
	Oscillatory angle instability	Q-learning	Ernst et al. (2004), Glavic et al. (2005a), Karimi et al. (2009), Ademoye and Feliachi (2012), Younesi et al. (2018), Li and Wu (1999)
		Fitted Q-iteration	Ernst et al. (2009, 2005), Wang et al. (2014)
Subsystem	Voltage (FIDVR)	DRL (DQN)	Huang et al. (2019a)
	Cascading failure	Q-learning	Belkacemi et al. (2017)
Wide-area	Oscillatory angle instability	Q-learning	Hadidi and Jeyasurya (2013), Duan et al. (2018)
		TD(λ)	Yousefian and Kamalsadan (2016)
	Transient angle instability	actor-critic	Yousefian et al. (2017), Yousefian and Kamalsadan (2018)
	Frequency instability	*TD*(λ)	Jung et al. (2002)

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. Claessens et al. (2018) 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 Duan, Li, Shi, Lin, and Wang (2019a), Bagheri, Nurmanova, Abedinia, and Naderi (2018) and Khorramabadi and Bakhshai (2015). In Duan, Li, et al. (2019a) 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 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 Bagheri et al. (2018) 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 Bagheri et al. (2018) 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. Khorramabadi and Bakhshai (2015) 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 Wu, Fang, Fang, Chen, and Tse (2010). $Q(\lambda)$ 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 Zarabbian, Belkacemi, and Babalola (2016) 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 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). Ernst et al. (2004), Ernst, Glavic, Geurst, and Wehenkel (2005), Ernst, Glavic, Capitanescu, and Wehenkel (2009), Glavic (2005), Glavic, Ernst, and Wehenkel (2005a,b), Yousefian, Bhattarai, and Kamalasadan (2017), Huang et al. (2019a), Yousefian and Kamalasadan (2018) dealt with transient instability control by controlling individual electric power system components such as thyristor-controlled series capacitor (Ernst et al., 2009; Ernst et al., 2005; Ernst et al., 2004) and a dynamic brake (a resistor usually located near electricity generation plant) to absorb excess of electricity generation (Glavic et al., 2005a; 2005b). Q-learning was used in Ernst et al. (2004), Glavic (2005), Glavic et al. (2005a,b) while Ernst et al. (2009, 2005) suggested fitted Q-iteration. Glavic et al. (2005a) 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 Glavic (2005). A dynamic brake wa also considered in Huang et al. (2019a) for emergency control using DRL (DQN was an approach of the choice in Huang et al., 2019a) largely following implementation details presented in Ernst et al. (2004) and Glavic (2005). The problem of transient angle instability was also considered within wide-area control systems (Yousefian et al., 2017; Yousefian & Kamalasadan, 2018). Work of Yousefian et al. (2017) 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 Yousefian and Kamalasadan (2018) 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 (Ademoye & Feliachi, 2012; Ernst et al., 2009; Ernst et al., 2005; Ernst et al., 2004; Glavic et al., 2005a; Karimi, Eftekharnejad, & Feliachi, 2009; Li & Wu, 1999; Wang, Glavic, & Wehenkel, 2014; Younesi, Shayeghi, & Moradzadeh, 2018) and wide-area controls (Duan, Xu, & Liu, 2018; Hadidi & Jeyasurya, 2013; Yousefian & Kamalsadan, 2016). Some of the works (Ernst et al., 2009; Ernst et al., 2005; Ernst et al., 2004; Glavic et al., 2005a) 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 Karimi et al. (2009) using Q-learning. Ademoye and Feliachi (2012) 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 Younesi et al. (2018) 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 Li and Wu (1999) 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 Wang et al. (2014) where a trajectory-based approach was designed as supplementary to existing controllers. In Hadidi and Jeyasurya (2013) 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 Duan et al. (2018) also used Q-learning RL method and both physical and communication infrastructure brought uncertainties were addressed. Wide area control for oscillations damping presented in Yousefian and Kamalsadan (2016) used $TD(\lambda)$.

Huang et al. (2019a) dealt with Fault Induced Delayed Voltage Recovery (FIDVR) problem through DRL (DQN) where undervoltage 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 Jung, Liu, Tanimoto, and Vittal (2002) 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 Liu, Jung, Heydt, Vittal, and Phadke (2000).

An approach for emergency control of cascading failures was presented in Belkacemi, Babalola, and Zarrabian (2017) 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 Ye, Zhang, and Sutanto (2011) 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 Wu, Fang, Fang, Chen, and Tse (2019b) 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 Ghorbani, Choudhry, and Feliachi (2016) 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 (An, Yang, Liu, & Zhang, 2019; Chen et al., 2019; Dibaji et al., 2019; He, Chen, Zhu, Yang, & Guan, 2019a; 2019b; Kurt, Ogundijo, Li, & Wang, 2019; Oozeer & Haykin, 2019; Wang, Chen, Liu, Xia, & Zhang, 2018; Yan, He, Zhong, & Tang, 2017). Dibaji et al. (2019) 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 Yan et al. (2017) 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 Wang, Chen, et al. (2018) 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 He et al. (2019a,b), An et al. (2019). Secure state estimation with assumption that measurements are sent over a wireless networks under jamming attacks was dealt in He et al. (2019a) and the antijamming game framework for secure state estimation using multi-agent reinforcement learning to determine optimal path against an intelligent attacker. He, Chen, Zhu, Yang, and Guan (2019b) 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 An et al. (2019) 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 Kurt et al. (2019) 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 Oozeer and Haykin (2019) discusses to use of RL in a general framework of cognitive risk control for cyber attacks in smart grids. RL was proposed in Chen et al. (2019) 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 Wu, Ota, Dong, Li, and Wang (2018) 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 Feng, Sun, and Zhang (2019) 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.

Xie, Ma, Ma, and Wang (2019) proposed the use of Q-learning with imitation and knowledge transfer for improved modelling of composite loads in electric power systems. In Shang, Li, Zheng, and Wu (2019) 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 Duan et al. (2019b) 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 (Ernst et al., 2009; Glavic, 2005; Li & Wu, 1999; Wang et al., 2014) where RL was fused with the concept of control Lyapunov functions (Glavic, 2005), model predictive control (Ernst et al., 2009; Wang et al., 2014) and fuzzy logic based control (Li & Wu, 1999).
- 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 Glavic, 2005 where communication delays were handled through the use of states-controls history).
- Some considerations (Bagheri et al., 2018; Duan, Li, et al., 2019a) 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 (Farhadi & Mohammed, 2015). In principle, these devices could be controlled in a similar way as dynamic braking resistor (Glavic, 2005; Glavic, Ernst, & Wehenkel, 2005b).
- 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 Glavic, 2005; Kurt et al., 2019 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 Glavic (2005) 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 Fan & Li, 2019; Garcia & Fernandez, 2015 and especially (Jin & Lavaei, 2018; Mannucci, van Kempen, de Viser, & Chu, 2018) discussing safe exploration for controls, Kretchmar et al. (2001) for synthesis of RL and robust control with stability guarantees, and Pinto, Davidson, Sukthankar, and Gupta (2017) 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 Ernst et al., 2009; Glavic, 2005; Khorramabadi & Bakhshai, 2015; Li & Wu, 1999; Wang et al., 2014, for some future on control see Annaswamy, 2013; Lamnabhi-Lagarrigue et al., 2017; Samad & Annaswamy, 2017). Recent work presented in Gros and Zanon (2019) 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 (Ernst et al., 2009; Wang et al., 2014) 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). Abramova, Dickens, Kuhn, and Faisal (2019) offers some viable insights on this possibility. Moreover, expected increase in future electric power system operations uncertainties necessitate more deployment of robust control (see Lamnabhi-Lagarrigue et al., 2017). 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 (Kretchmar et al., 2001) and robust adversarial RL (Pinto et al., 2017).
- 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 Otomega, Glavic, & Van Cutsem, 2007 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 (Madani et al., 2010).
- As argued in Wehenkel, Glavic, Geurts, and Ernst (2006), preventive control problems would be better formulated as multistage 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 (Wehenkel et al., 2006). 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 Ruiz-Vega, Glavic, & Ernst, 2003 for transient instability problem).
- The approach from Feng et al. (2019) 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 (Barto & Mahadevan, 1999), preference-based RL (Wirth, Akrour, Neumann, & Furnkranz, 2017) and imitative RL (Price & Boutilier, 2003) 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 (Xie et al., 2019; Yu et al., 2012; Zhang et al., 2018) where Zhang et al. (2018) and Xie et al. (2019) 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 Klenske & Hennig, 2016 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. Duan et al. (2019b) is a good example showing better performances of DDPG with respect to DQN for particular problem. Bayesian DRL is particularly interesting for future considerations. Azizzadenesheli and Anandkumar (2019) 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 de Bruin, Kober, Tuyls, & Babuska, 2018). DRL methods are not without the issues (particularly related to the convergence and sensitivity to involved parameters) (Busoniu et al., 2018), 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 De Asis, Chan, Pitis, Sutton, & Graves, 2019, 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 (Kiumarsi et al., 2018; Theodorou, Buchli, & Schaal, 2010) and is worth of more consideration for electric power system control (only one work considered the use of integral RL to this purpose Abouheaf et al., 2019) 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. Xie et al. (2019) and Shang et al. (2019) considered RL for these purposes (identification of composite load (Xie et al., 2019) and an equivalent of active distribution network Shang et al., 2019). Some successes in using deep learning for this purposes (an example is reference Zheng et al., 2019) 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 Wang, Liu, Zhang, and Xiao (2016).
- 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). Lei, Tan, Liu, Zheng, and Shen (2019) 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 (Liu, Yu, Teng, Leung, & Song, 2019).
- 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 Bertsekas (2019) "… 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

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

Abouheaf, M., Gueaieb, W., & Sharaf, A. (2019). Load frequency regulation for multi--area power system using integral reinforcement learning. *IET Generation, Transmission, Distribution, 13*, 4311–4323.

Abramova, E., Dickens, L., Kuhn, D., & Faisal, A. (2019). RLOC: Neurobiologically inspired hierarchical reinforcement learning algorithm for continuous control of nonlinear dynamical systems, arXiv:1903.03064v1, Accessed September, 2019 (pp. 1-33), https://arxiv.org/abs/1903.03064.

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

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

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

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

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

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

Annaswamy, A. (2013). Vision for smart grid control: 2030 and beyond. *Technical Report*. IEEE Standards Association.

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

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

Bag, A., Subudhi, B., & Ray, P. (2019). An adaptive variable leaky least mean square control scheme for grid integration of a PV system. *IEEE Transactions on Sustainable Energy*. Early access https://ieeexplore.ieee.org/document/8770144 .

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

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

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

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

Bertsekas, D. P. (1995). *Dynamic programming and optimal control, Vols. I and II*. Boston: Athena Scientific.

Bertsekas, D. P. (2019). *Reinforcement learning and optimal control*. Nashua, NH, USA: Athena Scientific.

Busoniu, L., Babuska, R., Schutter, B. D., & Ernst, D. (2010). *Reinforcement learning and dynamic programming using function approximators*. Boca Raton: CRC Press.

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

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

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

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

De Asis, K., Chan, A., Pitis, S., Sutton, R. S., & Graves, D. (2019). Fixed horizon temporal difference methods for stable reinforcement learning, arXiv:1909.03906v1, Accessed September, 2019 (pp. 1-16), https://arxiv.org/abs/1903.03064J.

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

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

Duan, J., Li, Z., Shi, D., Lin, C., & Wang, Z. (2019a). Reinforcement-learning-based optimal control for hybrid energy storage systems in hybrid AC/DC microgrids. *IEEE Transactions on Industrial Informatics, 15*, 5355–5364.

Duan, J., Shi, D., Diao, R., Li, H., Wang, Z., Zhang, B., Bian, D., & Yi, Z. (2019b). Deep-reinforcement-learning-based autonomous voltage control for power grid operations. *IEEE Transactions on Power Systems*. Early access https://ieeexplore.ieee.org/document/8834806 .

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

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

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

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

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

Fan, J., & Li, W. (2019). Safety-guided deep reinforcement learning via online Gaussian process estimation, arXiv:1903.02526v2, Accessed September, 2019, (pp. 1-13), https://arxiv.org/abs/1903.02526

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

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

Feng, C., Sun, M., & Zhang, J. (2019). Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection. *IEEE Transactions on Smart Grid*. Early access https://ieeexplore.ieee.org/document/8813103 .

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

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

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.

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

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

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

Gros, S., & Zanon, M. (2019). Data-driven economic NMPC using reinforcement learning. *IEEE Transactions on Automatic Control*. Early access https://ieeexplore.ieee.org/document/8701462 .

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

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

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

Huang, Q., Huang, R., Hao, W., Tan, J., R, F., & Huang, Z. (2019a). Adaptive power system emergency control using deep reinforcement learning. *IEEE Transactions on Smart Grid*. Early access https://ieeexplore.ieee.org/document/8787888 .

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

Jin, M., & Lavaei, J. (2018). Stability-certified reinforcement learning: A control theoretic perspective, arXiv:1810.11505v1, Accessed August, 2019, (pp. 1–30), https://arxiv.org/abs/1810.11505

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

Kamalapurkar, R., Walters, P., Rosenfeld, J., & Dixon, W. (2018). *Reinforcement learning for optimal feedback control: A Lyapunov-based approach*. Cham, Switzerland: Springer.

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

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

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

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

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

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

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

Lamnabhi-Lagarrigue, F., Annaswamy, A., Engell, S., Isaksson, A., Khargonekar, P., Murray, R. M., et al. (2017). Systems and control for the future of humanity, research agenda: Current and future roles, impact and grand challenges. *Annual Reviews in Control, 43*, 1–64.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 426–444.

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

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

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

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

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

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

Oozeer, M. I., Haykin, S. (2019). Cognitive risk control for mitigating cyberattack in smart grid, IEEE Access, 7, 125806–125826

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

Padiyar, K. R. (2008). *Power system dynamics, stability and control*. Bangalore: BS Publications.

Pinto, L., Davidson, J., Sukthankar, R., & Gupta, A. (2017). Robust adversarial reinforcement learning, arXiv:1703.02702v1, Accessed September, 2019, (pp. 1–10), https://arxiv.org/abs/1703.02702.

Powell, W. B., & Meisel, S. (2016). Tutorial on stochastic optimization in energy Part I: Modeling and policies. *IEEE Transactions on Power Systems, 31*, 1459–1467.

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

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

Ruiz-Vega, D., Glavic, M., & Ernst, D. (2003). Transient stability emergency control combining open-loop and closed-loop techniques. In *IEEE PES general meeting* (pp. 2053–2059). IEEE.

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

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

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

Sun, J., Zhu, Z., Li, H., Chai, Y., Qi, G., & Wang, H. (2019). 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*, 286–299.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge: MIT Press.

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

US-DoE (2004). Final Report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. *Technical Report*. US Department of Energy, US-Canada Power System Outage Task Force.

US-DoE (2015). An assessment of energy technologies and research opportunities. *Technical Report*. US Department of Energy.

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

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

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

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

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

Wang, Z., Liu, L., Zhang, H., & Xiao, G. (2016). 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*, 611–622.

Wehenkel, L., Glavic, M., Geurts, P., & Ernst, D. (2006). Automatic learning of sequential decision strategies for dynamic security assessment and controls. In *IEEE PES general meeting* (pp. 1–6). IEEE.

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

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

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

Wu, D., Zheng, X., Kalathil, D., & 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.

Wu, J., Fang, B., Fang, J., Chen, X., & 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.

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

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

Xi, H., Dominguez-Garcia, A. D., & Sauer, P. W. (2018). Optimal tap setting of voltage regulation transformers using batch reinforcement learnings, arXiv:1807.10997v2, Accessed August, 2019, (pp. 1–8), https://arxiv.org/abs/1807.10997

Xie, J., Ma, Z., Ma, S., & Wang, Z. (2019). Data-driven based method for power system time-varying composite load modeling, arXiv:1905.02688v1, Accessed August, 2019. (pp. 1–8), https://arxiv.org/abs/1905.02688

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.

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

Yang, Q., Wang, G., Sadeghi, A., & 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.

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

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

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

Yousefian, R., Bhattarai, R., & 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.

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

Yousefian, R., & Kamalsadan, S. (2016). 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*, 395–406.

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

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

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

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

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

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

Zhang, X. S., Yu, T., Pan, Z. N., Yang, B., & 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.

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