Received: February 5, 2023. Revised: March 23, 2023. 322

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Improvement of Perturb and Observe Based on Reinforcement Learning for Maximum Power Point Tracking Under Fast Changing Condition

Ernando Rizki Dalimunthe[1] Erwin Susanto[1,2]* Kharisma Bani Adam[3]

[1]*School of Electrical Engineering, Telkom University, Indonesia*
[2]*HUMIC Research Centre, Telkom University, Indonesia*
[3]*Department of Electrical Energy Engineering, School of Electrical Engineering, Telkom University, Indonesia*
* Corresponding author's Email: erwinelektro@telkomuniversity.ac.id

Abstract: The most frequently used maximum power point tracking (MPPT) technique is the perturb & observe (P&O) algorithm as a power tracking tool in PV system. The P&O algorithm is easy to compute and implement, but the algorithm is prone to oscillations at the maximum power point. Hence, the power becomes inaccurate due to a lot of power loss. This study utilizes the deep q- network (DQN) algorithm to improve P&O performance by correcting the output value algorithm using various step sizes by DQN. The proposed method increased the tracking speed when receiving the same value at different times by 33.3%–50%, and the oscillation rate was successfully reduced by 73.99%–83.5%. The advantages of increasing tracking speed and decreasing oscillation rate are accompanied by tracked power with averages of 95%, which is better than the P&O and DQN algorithms. It shows that the proposed method can work optimally regarding efficiency and oscillation rate and be the fastest in tracking maximum power from previous related works.

Keywords: Deep q- network (DQN), Maximum power point tracking, PV, Peturb & observe (P&O), Reinforcement learning.

1. Introduction

The role of solar power plants can be used as a reference for providers of clean electrical energy from pollution with abundant solar resources. Photovoltaic (PV) is one of the solutions for energy needs that are low in exhaust emissions compared to fossil power plants. PV generates electrical energy by converting solar radiation and temperature on the cell surface, thus PV has nonlinear characteristic because it depends on environmental conditions. Under certain environmental conditions, PV has a maximum power point that must be achieved. Power extraction by PV must be converted with a converter device to be used at specific appropriate loads.

Several algorithms have been developed as MPPT techniques grouped based on input requirements, tracker accuracy and speed, effectiveness, and data processing techniques. One of the MPPT techniques most often used methods are peturb and observe (P&O), incremental conductance (INC), and hill fractional open/ short circuit current, generally categorized as conventional methods. The methods were widely used in MPPT techniques because of their simple computation, by manually power changes calculating from input information such as current and voltage. However, the method is highly dependent on the chosen step size value, which is fixed. The larger the step size, the shorter the time required, and vice versa. It also impacts the tracking performance, which is prone to oscillations at the peak point [1]. In contrast to conventional methods, intelligent control methods and evolutionary algorithms such as fuzzy logic controllers (FLC) [2], artificial neural networks (ANN) [3], genetic algorithms (GA) [4], ant colony optimization (ACO) [5], and particle swarm optimization (PSO) [6] were reported. The algorithm requires specific information about PV to track the maximum power

point more accurately than conventional methods, but this method is more difficult to implement into the system and requires more duration time [7]. Many studies have been carried out by combining the two methods, either by combining conventional methods with intelligent control methods such as the FLC algorithm with P&O [8] that applied to the permanent magnet synchronous motor (PMSM). The results of this study show that the development of the P&O algorithm with FLC corrects the P&O deficiencies when light conditions change rapidly—however, oscillation and tracking time remain issues especially tracking time took around 50ms to reach steady state condition. To conquer each other's limitations, two vastly distinct algorithms are combined, such as the FLC algorithm with the cuckoo search algorithm [9], ANNPSO algorithm [10] and PSO-SMC [11]. The three combinations perform well in terms of efficiency and oscillation, but the tracking time is slightly longer than that of the FLC-based P&O. Improving the deficiencies of each algorithm is not only done by combining the basic algorithm with other algorithms that are considered to have advantages but these improvements can be made by modifying the algorithm itself [12].

In recent years, implementating of the reinforcement learning method as an algorithm for the MPPT technique has been extensively studied. Reinforcement learning (RL) is a machine learning method widely applied to control applications. This method has the advantage of tracking without specific system information. One of the RL methods that has been widely used for PV power control applications is the q-learning (QL) algorithm. The control is carried out without clear model information (model-free). QL application research on MPPT shows more efficient than P&O, with a lower oscillation rate and faster tracking both in constant conditions and quickly changing conditions that are partially shaded and also shows its superiority in tracking time on same input values [13-15].

The essential weakness of the QL algorithm is that the state conditions must be discretized, and the state-action pairs are stored in a table. It requires a large table composition which may cause the maximum power tracking process to be slower. One of solution in terms of state discretization is using deep reinforcement learning (DRL), which combines the concept of QL with deep learning. In addition, the implementation of reinforcement learning in energy systems and modern power has been comprehensively discussed [16]. The application of the deep q-learning network (DQN) algorithm in MPPT shows results that are as good as QL, but the process of identifying models is better [17, 18]. Compared to QL algorithm, in [13] the average power around MPP by DQN shows good result while in [17] DQN has consistently denotes improved tracking efficiency over P&O.

The combination of the RL and P&O methods was proposed in 2020 [19]. Combining the two methods is by running the QL algorithm to determine the duty cycle value, which will be used as a reference for the execution of the P&O algorithm. The proposed method reduces the detection time by 80.5% - 98.3% compared to the convergent time required by PSO in partially shaded conditions, even though at the beginning of the test, the proposed method takes longer than the PSO. Again, the RL algorithm improves the previous tracking results in less time. Furthermore, the QL algorithm has also been combined with the type 2 FLC algorithm [20].

In this paper, the MPPT control method is designed by combining conventional methods, P&O algorithm, and the DQN algorithm. In the system design, the P&O algorithm is executed earlier to obtain the duty cycle value for tracking the optimal point of PV with a fixed step size. Thus, the main contribution of this paper is to correct the oscillation level of the P&O algorithm's. Additionally, the DQN algorithm as the machine learning method, is used to reduce the tracking time when MPPT accepts repeated information. Lastly, the efficiency is also considered with the comparison of three algorithms applied to see the expected improvement results.

This paper is organized using the following sections: section II, describes the modeling of PV and boost converter and their parameters, then the fundamentals of P&O are briefly explained in the next section, also the proposed method. The fourth section presents the simulation findings and parameter tests. The analysis of the testing result is shown in the fifth part. Finally, the last section concludes the study.

2. Modelling system

As illustrated in Fig. 1, the system architecture developed in this paper includes the PV, converter, and MPPT controller.

This section presents the mathematical modeling of PV, the parameters of the chosen PV type, and the general properties of PV. Moreover, a mathematical equation for the boost converter utilized in the PV system is given. The equation can determine the values of the converter's components.

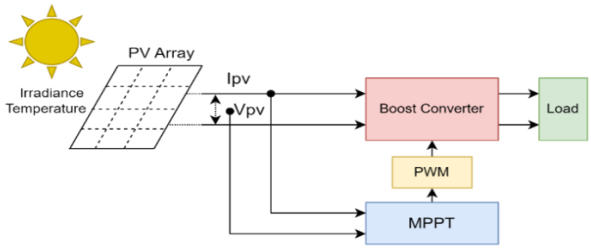

Figure. 1 PV system design

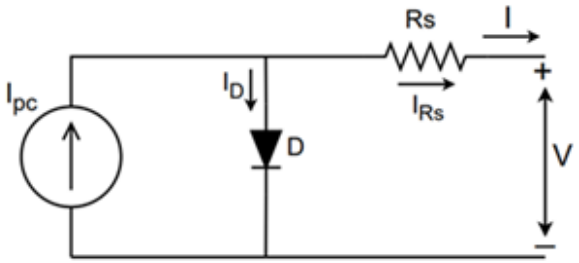

Figure. 2 PV cell equivalent circuit

2.1 PV model and characteristic

PV modeling uses equivalent electric circuit to create a model based on PV characteristics.

The equivalent PV circuit consists of the converted current from the solar cell (I_{PC}), the diode current (I_D), and the series resistance (R_S), as shown in Fig. 2. The circuit creates PV mathematical equations for modeling in software simulations [16]. I_{PC} is the current generated directly from irradiation on PV. I_{PC} is affected by the level of solar radiation and surface temperature, and PV has positive and negative poles, which are represented by a diode with current flowing in the diode (I_D). The resistance (R_S) indicates the series resistance of the solar cell. Eq. (1) shows the simple principle of the solar cell equivalent circuit above, namely the magnitude of the PV current (I) is the reduction of the photocurrent (I_{PC}) with the current through the diode (I_D), and formulated as follows:

$$I = I_{PC} - I_D \tag{1}$$

As in Eq. (1), I_{PC} and I_D can be described in Eqs. (2) and (3) as follows:

$$I_{PC} = I_{SC} + K_i(T - 298)\frac{G}{1000} \tag{2}$$

$$I_D = I_O\left[exp\left(q\frac{(V+IR_S)}{N_S AkT}\right) - 1\right] \tag{3}$$

In Eq. (2), it can be observed that the size of the output current converted from PV is determined based on temperature (T) in ^oK and solar radiation (G) in units of W/m^2. The short circuit current denoted by I_{SC} is the magnitude of the PV current at a standard temperature of 25 °C, and K_i is the temperature coefficient of the short circuit current. While refer to Eq. (3), the diode saturation current (I_O) has not been expressed, then I_O can be determined by Eq. (4) below:

$$I_O = I_{RS}\left[\frac{T}{T_r}\right]^3 exp\left[\frac{qEg}{Ak}\left\{\frac{1}{T_r} - \frac{1}{T}\right\}\right] \tag{4}$$

where I_{RS} is the reverse saturation current in units (A), T is the cell temperature in kelvin units (K), while T_r is the reference temperature, 273.15^{o}K, q is the electron charge (1.6×10^{-19}) in unit Coloumb (C), N_s is the series cell number, A is the ideality factor of semiconductor, k is The constant of Boltzmann equals to 1.3805×10^{-23} J/K and E_g is the gap energy in semiconductors which can be found in the panel specification data. Refer to Eq. (4), it can be seen that the diode saturation current is strongly influenced by the temperature level, as known that semiconductor components are susceptible to temperature increases. The PV parameters to be modeled are the type of Alta device. The Alta device was serialized with 8 cells and parallelized with 26 cells to create a PV module with the properties listed in Table 1.

2.2 Boost converter

A primary boost converter circuit is shown in Fig. 2.6 with a power MOSFET acting as the switching component, an inductor (L) acting as a filter to reduce current ripple, a diode (d) acting as the switching component, which operates when the switch is open so that current flows to the inductor, a capacitor (C) acting as a filter to reduce voltage ripple, and a resistor (R) acting as the load. The boost converter can be operated in two modes: continuous current mode (CCM) and discontinuous current mode (DCM). The difference between the two modes lies in the inductor current when switching is steady. The inductor current will flow continuously in CCM mode so that the output voltage can be adjusted by changing the pulse width and does not depend on the inductor and capacitor. In contrast, in DCM mode, the inductor current is 0, and the output voltage depends on the inductor value and the pulse width [9].

The converter's switching component makes the circuit operate in two states when the switch is open and closed. Consequently, circuit analysis is necessary whether the switch is closed or open. When the switch is closed, the input voltage is equal to the voltage across the inductor, assuming that the switch is ideal. This condition causes an increase in the current through the inductor so that the left polarity is

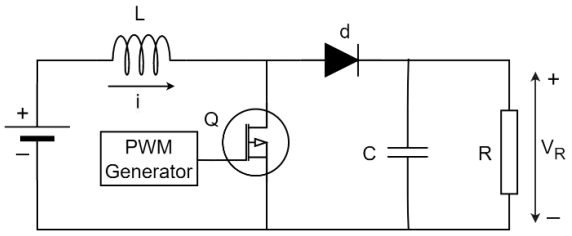

Figure. 3 Boost converter circuit

more favorable than the right side. The change in current through the inductor at a particular time is formulated as follows:

$$\frac{\Delta i_L}{\Delta t} = \frac{V_{in}}{L} \tag{5}$$

$$\Delta i_L(closed) = \frac{V_{in}DT}{L} \tag{6}$$

The reverse condition is when the switch is open, current from the DC source can flow to the capacitor, and the load, the capacitor experiences a charge.

$$V_{in} - V_o = L\frac{di_L}{dt} \tag{7}$$

$$\Delta i_L(open) = \int_{DT}^{T} \frac{(V_{in}-V_o)\,dt}{L} \tag{8}$$

Then the equation for Δi_L when the switch is open is as follows:

$$\Delta i_L(open) = \frac{(V_{in}-V_o)\,(1-D)t}{L} \tag{9}$$

In this condition, the inductor experiences discharge. The condition of the voltage on the inductor when it discharges can be formulated according to Eqs. (7), (8), and (9), where V_{in} is the input voltage, V_o is the output voltage (V), Δi_L is the change in current through the inductor (A), t is the time (s), T is periode and D is the pulse width (duty cycle). The converter operates at a steady state, so the amount of energy stored in each component must be the same at the beginning and end of the switching process. This condition causes the energy stored in the inductor to be the same between the start and end of the switch, meaning that the change in the inductor current is 0. Accordingly, two equations of Δi_L when the switch closed and open are added to one another, then the formula for boost converter output voltage is as follows:

$$V_o = V_{in}\left[\frac{1}{1-D}\right] \tag{10}$$

As is well known, the MOSFET type transistor acts as an electronic switch that opens or closes based on the switching frequency specified; the higher the value, the faster the switch opens and closes. That causes the switch to operate faster, which might raise component temperatures and lead to power losses. Temperature increases can also interfere with the diode's functioning, which conveys current (forward bias) from the inductor when the switch is open and prevents current from the opposite direction when the switch is closed (reverse bias) [21]. To achieve the best results, these two components' internal resistance, current, and maximum voltage must be considered. The values of the specified parameters are selected based on the ideal case regardless of the physical factors of the components. As a result, the boost converter is modeled on Simulink using the following equation and the component value specifications listed in Table 1.

Table 1. PV module and boost converter spesification

Spesification	Unit	Value
Maximum power (P_{max})	W	44.52
Maximum voltage (V_{max})	V	7.68
Maximum current (I_{max})	A	5.798
Open circuit voltage (V_{oc})	V	8.72
Short circuit current (I_{sc})	A	6.058
Coeficient of temperature Isc (Ki)	%/°C	0.084
Coeficient of temperature Voc (Kv)	%/°C	-0.187
Switching frequency (f)	kHz	50
Inductor (L)	μH	6.22
Capacitor (C)	μF	428.6

3. MPPT controller system

Due to the nonlinear characteristics of PV, it must be controlled with a device that tracks maximum power at every change in natural conditions. In this paper, two algortihm are to be modeled for MPPT controllers, P&O algorithm and DQN algorithm.

3.1 Perturb & observe (P&O) algorithm

As the name implies, the peturb & observe algorithm has two stages: disturbance (perturbation) and observation (observation). The process followed in this algorithm interfered by changing the reference voltage value, which impacts the output value of the duty cycle (D). The following process is observing the results of the interference in the previous stage. The way the P&O algorithm works is that if the change in power due to disturbance is positive, the next disturbance is carried out in the

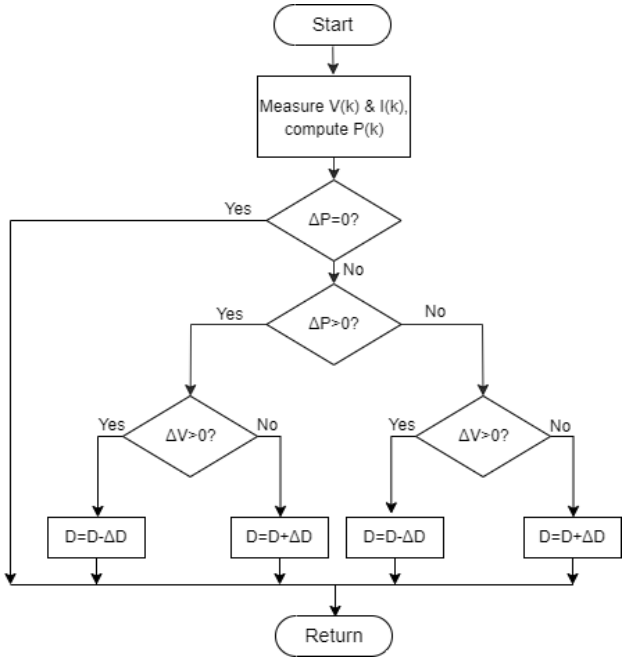

Figure. 4 P&O algorithm

same direction as before. However, if the result of the disturbance is negative, the trend of the disturbance is reversed. According to Fig. 6, the flowchart shows that algorithm works with the parameters of power, voltage, and PV output current as input.

V(k) is the reference voltage while I(k) is the reference current and multiplying the two gives the reference power (P(k)). ΔP is the subtraction of the reference power with the power after interruption, while the substraction of the reference voltage with the voltage after disturbance is called ΔV. The disturbance/ interruption itself is the step size value in the P&O algorithm, called ΔD, while the D value is the initial duty cycle. P&O's tracking technique allows it to track the maximum power but with low efficiency due to oscillations at the maximum point. In this paper, 0.01 is used as the P&O's step size which is a relatively large value. The drawback of the algorithm is fixed step size; the larger the step size value, the more the oscillation that occurs, but the tracking time becomes faster, and vice versa. In the proposed method, the P&O algorithm will be optimized with an algorithm that can apply a variable step size, and the output from the P&O will be rechallenged.

3.2 The proposed method P&O-DQN

The reinforcement learning (RL) method is a part of machine learning that works based on the cumulative rewards of the environment in making the right decisions to achieve the desired goals. The RL technique differs from other methods, such as

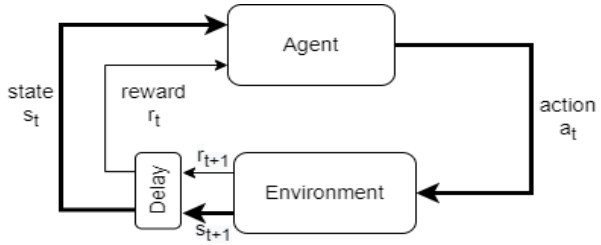

Figure. 5 Reinforcement learning general scenario

supervised learning, because RL does not require a training data set in the form of input/output pairs (labeled data) and does not require correction of sub-optimal actions. The RL technique is also different from unsupervised learning, which needs to prepare a training data set (unlabeled data) to learn structures in the form of clusters hidden in the data set. Refer to Fig. 5, RL only requires two components to solve the problem: agent and environment. The environment represents the system in which the agent operates and collects rewards from an action selection. In contrast, the agent represents a learning system that has been determined to carry out an action aimed at achieving a goal. Information about the system by RL will be represented in a process called the Markov decision process (MDP) [16]. The markov property consist of A set of environments that represents the state (s_t) and the basis of the reward signal (r_t) an agent carries out a set of actions (a_t) and take an action on the environment based on the existing state, transition of probability which from state (s_t) to next state (s_{t+1}) due to an action (a_t), and rewards function are given directly after the transition from (s_t) to (s_{t+1}) with action (a_t) and the following action (a_{t+1}) will be selected for each probability transition.

Decision-making in reinforcement learning is based on policy and value; in this paper, the topic discussed is how to track the maximum power; hence decision-making is based on value. The deep q-network DQN is a value-based algorithm with model-free characteristics; the advantage of DQN is that it combines the QL concept, which is part of reinforcement learning, with the deep neural network concept, which is part of the deep learning method. It carries out the process of finding action values based on the maximum Q value and rewards. However, using the network's learning process, the state function does not need to be discretized because the input value can be directly processed by the neural network (NN) [20]. The algorithm processes a continuous state function such as V_{pv}, I_{pv} and $D_{P\&O}$ to make the transition function available in large quantities, the concept of the proposed method

was shown in Fig. 6. The correlation between the training data will be sampled randomly with the experience replay mechanism in memory when the NN updates. This sampling uses the gradient descent method to minimize the loss function so that the maximum value can be immediately known in the following circumstances. The loss function (L (θ)) is calculated by the expected value of reward (r) after take an action with target value of Q ($Q(s_{t+1}, a_{t+1}|\theta')$) and the predicted value of Q ($Q(s_t, a_t|\theta)$ with the complete equation is expressed in Eq. (11) [13]. The next step after calculates loss function is to update the state function and the following action of the previously formed function. Formation of the following function is expressed in the Bellman equation in Eq. (12) [13].

$$\mathrm{L}(\theta) = \mathbb{E}\left[(r + \gamma \max_a Q(s_{t+1}, a_{t+1}|\theta') - Q(s_t, a_t|\theta))^2\right] \quad (11)$$

$$\theta(s_t, a_t) = \theta(s_t, a_t) + \alpha[r + \gamma max\, Q(s_{t+1}, a_{t+1}|\theta) - Q(s_t, a_t|\theta)\nabla Q(s_t, a_t|\theta)] \quad (12)$$

The parameters θ' and θ are determined as the weight of the target value and the output prediction value of the q-network. Alpha (α) is the learning rate that aims to determine how new information is obtained and replaces the old information. The discount factor has a value of $\gamma \in [0,1)$ to determine the current value of future reward and $max\, Q(s_{t+1}, a_{t+1}|\theta)$ is the Q value of the next state-action pair that is likely to be selected with the most optimal reward for the following maximum Q value.

The proposed method is denoted in Algorithm 1, The DQN as optimizer of the P&O algorithm, also receives voltage and current signals from PVs to obtain the Q value i.e., step size (ΔD). In the DQN algorithm, an agent will interact with the environment to gain experience through several parameters. In this study, the parameters of DQN are represented in Table 2. Every change in the observed state has an action taken for each change. The various actions of these changes in this study are expressed in Eq. (13).

$$D_{optimal} = D_{P\&O} \pm \Delta D_{DQN} \quad (13)$$

$$a_t = argmaxQ(s_t, a_t|\theta) \quad (14)$$

where $D_{optimal}$ is the duty cycle output optimal from the initial duty cycle obtained from the output of the P&O algorithm ($D_{P\&O}$) will be corrected by the various duty cycle step size of DQN (ΔD_{DQN}). The choice of action is adjusted to two conditions, and these conditions are based on the probability ε. The DQN algorithm initializes the parameters such as, γ, α, ε-greedy, and network information. Parameter setting information for the DQN algorithm can be observed in Table 3.

Table 2. MPPT parameters

RL parameter	Parameters in system design
Environtment	PV and converter
Agent	MPPT controller
State	Voltage, current and *duty cycle* P&O
Action	Duty cycle step (ΔD) { ±0.003, ±0.005, ±0.03, ±0.05, 0}
Reward	Power change (ΔP), duty cycle value (D)

Table 3. The parameter settings of DQN algorithm

Parameters	Value
Hidden layer	3
Relu layer	2
Full connected layer	3
Number of neurons	256
Experience replay buffer	1×10^6
γ	0.9
α	0.0001
Maximum number of episodes	500
Exploration rate (ε)	1

The epsilon value continues to decay as the training process is carried out, with the decaying exploration rate at 0.002 and the minimum epsilon value is 0.001. It is attempted that the process of obtaining future rewards is more effective by considering the experience during training. In accordance with its primary role, the epsilon affects the selection of the action value. If the probability is equal to ε then the action is chosen randomly, otherwise, the action is selected based on Eq. (14).

4. Simulation result and discussion

The goal is to test the proposed method in dynamic changing conditions of irradiance and temperature, which simulations conducted in MATLAB/Simulink with the design of the DQN algorithm based on the reinforcement learning toolbox. The results are compared with basic

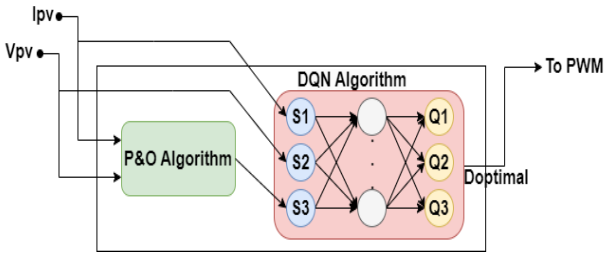

Figure. 6 MPPT controller scheme

Algorithm 1 P&O-DQN Algorithm

Connect all subsystem (PV module, MPPT & Boost Converter)

1. Measure voltage (V), current (I) of PV and calculate the Power (P);
2. Execute P&O algorithm for MPPT system and get the duty cycle value ($D_{P\&O}$);
3. Initialize DQN parameter { α, γ, maxstep, memory, ε (initial, decaying value & minimum exploration)};
4. State formation (s_t), observe the state & provide these values to the network;
5. **for** i : n do repeatly
6. Select action
7. **if** probability ε
8. Choose action randomly;
9. **else**
10. Choose available action in output layer of the network (Eq. (13));
11. **end**
12. Execute the choosen action from Eq. (14);
13. Get a new state (s_{t+1})& calculate reward;
14. Store the transition & reward function into memory;
15. **If** Data simulation > Experience replay buffer
16. Sampling the mini batch data from memory;
17. Calculate loss function from Eq. (11);
18. Do mini batch gradient descent for reducing loss function;
19. **end if**
20. Update Q value (Q') from Eq. (12);
21. Set $s_t = s_{t+1}$;
22. **end for**

algorithm of the proposed method to know the performance parameter such as settling time, oscillation range dan efficiency of power tracking among them.

Refer to Fig. 7, the simulation is carried out at varying irradiance from 500 W/m^2 – 1000 W/m^2 and temperature from 25^{o}C – 40^{o}C. The input signal consist of two times of 500 W/m^2 and 40 ^{o}C, two

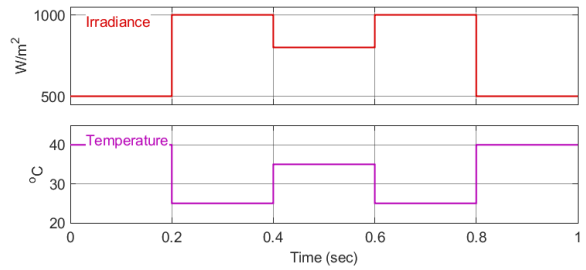

Figure. 7 Irradiance and temperature signal

times of 1000 W/m^2 and 25^{o}C and 800 W/m^2 at 35 ^{o}C. The repetitive input condition is made to evaluate the optimizer's ability for tracking the maximum power when the input is same at different times. To observe the significance, these input signals are grouped into three intervals. The first interval values with periods of 0-0.2s and 0.8-1s and the second interval values with periods of 0.2-0.4s and 0.6-0.8s and last intervals with periods of 0.4-0.6s.

4.1 Simulation result

Fig. 8 illustrates the results of the scenario; the settling time and oscillation range are look very different from each other. The DQN is the slowest for tracking the maximum power, but has small oscillation. The uncontrolled line in Fig. 8 (b) shows that PV system without MPPT controllers was unable to track the maximum power under dynamically changing condition. The comparison regarding performance of three algorithm for all scenario can be observed in Table 4.

4.2 Discussion

The evaluation of all scenario is observed involve the time settling parameter (Ts), oscillation range (Osc) and efficiency power tracked. The DQN algorithm takes longer to reach a steady state than the P&O algorithm, which only takes 5ms for each condition. This benefit is adopted by the proposed method so that the settling time on the P&O-DQN is shorter than the DQN, with a percentage increase of 91% for interval 1 and 98% for interval 2. These show better results than [19], which requires more extended tracking at the beginning of the test up to the ninth pattern, and even then, the time required is still longer than the proposed method. Moreover, in [8-10], the average settling time around MPP are 50ms, 16ms and 20ms; also, the proposed method by [11], has the slowest calculation time around 0.15s – 0.22s, when it compared with our proposed method, it can track power in less than 10ms.

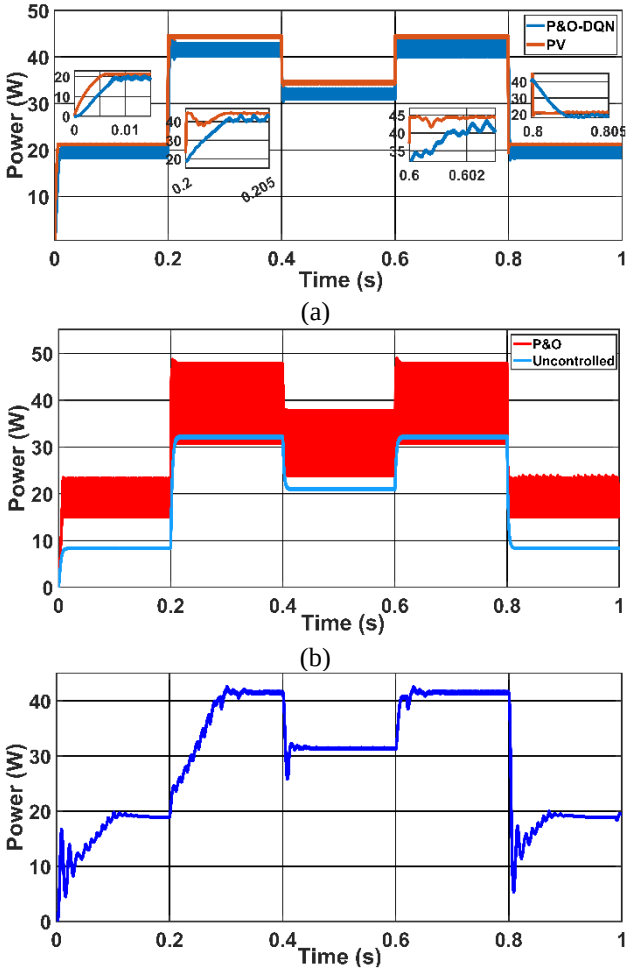

Figure. 8 The output power: (a) P&O-DQN, (b) P&O, and (c) DQN

The advantage of the proposed method is seen in repeated input conditions, the proposed method can reduce the settling time by 50% for interval 1 and 33.3% for interval 2. The same thing is consistent with the DQN, which experienced a decrease in duration under repeated input conditions. In addition, the proposed method shows its superiority compared to the P&O, which cuts the time by 60% at interval 2, when the P&O settling time does not change. The same thing is shown in [20]; even though there is no repeated input value, the tracking time consistently decreases as the input changes, but the tracking time is still relatively slow which is around 10ms-455m and also this is not shown in any other hybrid method [9-11].

Improved algorithm performance on the MPPT controller is also seen the oscillation indicator, it becomes important when the MPPT reaches its maximum point. In the hypothesis, the P&O-DQN algorithm is aimed at improving the state of the P&O algorithm at a steady state which is tracked power becomes not optimal because oscillations occur at the maximum point. Due to the perturbation, the step size is fixed so that the P&O continues to find the maximum point based on the characteristics of the P&O itself.

Table 4. Simulation result

Algorithm	Interval	Ppv (W)	Po (W)	Osc (W)	Ts (ms)
P&O-DQN	1	20.78	19.32	2.2	10 & 5
	2	43.27	41.53	3.44	4 & 2
	3	33.96	32.12	2.31	5
P&O	1	20.78	18.88	8.46	5
	2	43.27	39.43	17.18	5
	3	33.96	30.93	14	5
DQN	1	20.78	18.91	0.11	120 & 80
	2	43.27	40.79	0.44	140 & 60
	3	33.96	31.35	0.39	40

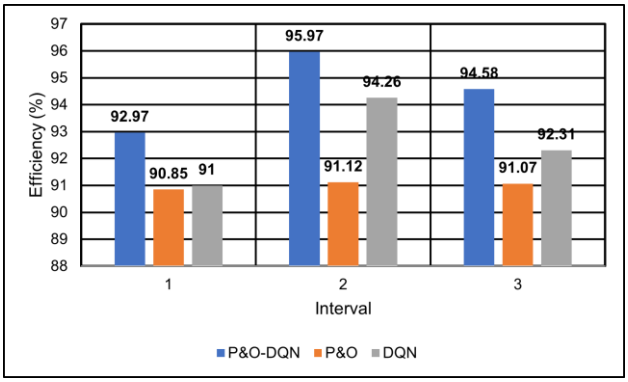

Figure. 9 Efficiency chart

Refer to Table 4, the P&O-DQN algorithm is able to reduce the oscillation level by 73.99% in interval 1, 79.97% in interval 2 and 83.5% in interval 3. It is aligned with the proposed method by [9], which can effectively reduce oscillation levels by up to 80%. In contrast, in [11, 20], oscillation occurs only in one case of five cases, and low power oscillation, respectively.

Fig. 9 illustrates that the maximum power tracking efficiency by the three algorithms that shows good results, whereas the three algorithms have an average efficiency above 90%. However, from all the tests, the proposed method managed to track the best maximum power of the other two algorithms. The P&O-DQN algorithm's average maximum power tracking efficiency is 94% - 95%, whereas, other hybrid methods can achieve 99% efficiency. It is because the value tracked at the beginning can be less optimal due to the relatively large size of the P&O step chosen, but basically, the proposed method can track power efficiently too.

5. Conclusion

The implementation of the reinforcement learning method in the perturb and observe (P&O) algorithm is a way to improve its performance. Simple computation, fast and good tracking are the advantages of the conventional P&O algorithm. It is adopted by the proposed method, a combination of the conventional one with the DQN algorithm, which has a tremendous amount of time for tracking. As a result, this combination can reduce the tracking time at the beginning and second attempt of input impacted by DQN as a machine learning method, and the proposed method has an outstanding result for it other than related works. It may be a concern for P&O; oscillation occurs at the maximum point, which makes it lose power. The P&O-DQN successfully reduces the oscillation level of the P&O; thus, compared to other methods that can obtain efficiency in 99%, the P&O-DQN can also track the maximum power efficiently.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Erwin Susanto is the paper's corresponding author. He helped develop the control strategy and coordinated the preparation of this paper. Kharisma Bani Adam assisted in the formal analysis of MPPT system performance and managed the paper organizations. Ernando Rizki Dalimunthe created the proposed control mechanism, modeled the MPPT system in MATLAB/SIMULINK, and authored the paper. All authors confirmed the final version.

References

[1] A. B. Kebede and G. B. Worku, "Comprehensive review and performance evaluation of maximum power point tracking algorithms for PV system", *Glob. Energy Interconnect.*, Vol. 3, No. 4, pp. 398–412, 2020.

[2] L. K. Narwat and J. Dhillon, "Design and Operation of Fuzzy Logic Based MPPT Controller under Uncertain Condition", *J. Phys. Conf. Ser.*, Vol. 1854, No. 1, pp. 1-12, 2021.

[3] L. Bouselham, M. Hajji, B. Hajji, and H. Bouali, "A New MPPT based ANN for PV System under Partial Shading Conditions", *Energy Procedia*, Vol. 111, pp. 924-933, 2017.

[4] Hadji, J. P. Gaubert, and F. Krim, "Real-Time Genetic Algorithms-Based MPPT: Study and Comparison (Theoretical an Experimental) with Conventional Methods", *Energies*, Vol. 11, No. 2, p. 459, 2018.

[5] S. Titri, C. Larbes, K. Y. Toumi, and K. Benatchba, "A new MPPT controller based on the Ant colony optimization algorithm for PV systems under partial shading conditions", *Applied Soft Computing*, Vol. 58, pp. 465-479, 2017.

[6] M. Alshareef, Z. Lin, M. Ma, and W. Cao, "Accelerated Particle Swarm Optimization for PV Maximum Power Point Tracking under Partial Shading Conditions", *Energies*, Vol. 12, No. 4, pp. 623, 2019.

[7] E. G. Okafor, D. Udekwe, O. C. Ubadike, E. Okafor, and P. O. Jemitola, "PV System Mppt Evaluation Using Classical , Meta -Heuristics , And *Reinforcement learning* -Based Controllers : A Comparative Study", *Journal of Southwest Jiaotong University*, Vol. 56, No. 3, pp. 1–17, 2021.

[8] M. R. Rezoug, R. Chenni, and D. Taibi, "Fuzzy logic-based perturb and observe algorithm with variable step of a reference voltage for solar permanent magnet synchronous motor drive system fed by direct-connected PV array", *Energies*, Vol. 11, No. 2, pp. 1-15, 2018.

[9] S. J. Zand, S. Mobayen, H. Z. Gul, H. Molashahi, M. Nasiri, and A. Fekih, "Optimized Fuzzy Controller Based on Cuckoo Optimization Algorithm for Maximum Power-Point Tracking of PV Systems", *IEEE Access*, Vol. 10, pp. 71699-71716, 2022.

[10] K. Khennoufi, M. Ferfra, and H. Bouzakri, "Conception and Hardware Implementation of MPPT Controller for Partially Shaded Photovoltaic Panels using Backstepping and Neural Network based Particle Swarm Optimization", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 4, pp. 545–554, 2022, doi: 10.22266/ijies2022.0831.49.

[11] K. H. Chao and M. N. Rizal, "A hybrid mppt controller based on the genetic algorithm and ant colony optimization for PV systems under partially shaded conditions", *Energies*, Vol. 14, No. 10, pp. 1-17, 2021.

[12] C. Y. Liao, R. K. Subroto, I. S. Millah, K. L. Lian, and W. T. Huang, "An Improved Bat Algorithm for More Efficient and Faster Maximum Power Point Tracking for a PV System Under Partial Shading Conditions", *IEEE Access*, Vol. 8, pp. 96378-96390, 2020.

[13] K. Y. Chou, S. T. Yang, and Y. P. Chen, "Maximum power point tracking of PV system based on reinforcement learning", *Sensors*

(Switzerland), Vol. 19, No. 22, pp. 1-17, 2019.

[14] P. Kofinas, S. Doltsinis, A. I. Dounis, and G. A. Vouros, "A reinforcement learning approach for MPPT control method of PV sources", *Renew. Energy*, Vol. 108, pp. 461–473, 2017.

[15] K. Chou and Y. Chen, "Reinforcement learning Based Maximum Power Point Tracking Control Of Partially Shaded PV System", *Journal of Marine Science and Technology*, Vol. 28, No. 5, pp. 433-443, 2020.

[16] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review", *Journal of Modern Power Systems and Clean Energy*, Vol. 8, No. 6, pp. 1029–1042, 2020.

[17] .B. C. Phan, Y. C. Lai, and C. E. Lin, "A deep reinforcement learning-based MPPT control for PV systems under partial shading condition", *Sensors*, Vol. 20, No. 11, pp. 1–22, 2020.

[18] K. Y. Chou, C. S. Yang, and Y. P. Chen, "Deep Q-Network Based Global Maximum Power Point Tracking for Partially Shaded PV System", In: *Proc. of 2020 IEEE International Conference on Consumer Electronics – Taiwan*, Taiwan, pp. 1-2, 2020.

[19] C. Kalogerakis, E. Koutroulis, and M. G. Lagoudakis, "Global MPPT based on machine-learning for PV arrays operating under partial shading conditions", *Appl. Sci.*, Vol. 10, No. 2, pp. 1-19, 2020.

[20] R. F. Iskandar, E. Leksono, and E. Joelianto, "Q-Learning Hybrid Type-2 Fuzzy Logic Control Approach for PV Maximum Power Point Tracking Under Varying Solar Irradiation Exposure", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 199–208, 2021, doi: 10.22266/ijies2021.1031.19.

[21] D. W. Hart, *Power Electronics*, McGraw-Hill, New York, N.Y. 2011.