

Article

Hybrid Fuzzy–DDPG Approach for Efficient MPPT in Partially Shaded Photovoltaic Panels

Diana Ortiz-Munoz , David Luviano-Cruz * , Luis A. Perez-Dominguez , Alma G. Rodriguez-Ramirez 
and Francesco Garcia-Luna 

Departamento de Ingeniería Industrial y Manufactura, Universidad Autonoma de Ciudad Juarez,
Juárez 32315, Mexico; diana.ortiz@uacj.mx (D.O.-M.); luis.dominguez@uacj.mx (L.A.P.-D.);
alma.rodriguez.ram@uacj.mx (A.G.R.-R.); francesco.garcia@uacj.mx (F.G.-L.)

* Correspondence: david.luviano@uacj.mx

Abstract: Partial shading conditions reduce the efficiency of photovoltaic (PV) systems by introducing multiple local maxima in the power–voltage curve, complicating Maximum Power Point Tracking (MPPT). Traditional MPPT methods, such as Perturb and Observe (P&O) and Incremental Conductance (IC), frequently converge to local maxima, leading to suboptimal power extraction. This study proposes a hybrid reinforcement learning-based MPPT approach that combines fuzzy techniques with Deep Deterministic Policy Gradient (DDPG) to enhance tracking accuracy under partial shading. The method integrates fuzzy membership functions into the actor–critic structure, improving state representation and convergence speed. The proposed algorithm is evaluated in a simulated PV environment under various shading scenarios and benchmarked against conventional Perturb and Observe P&O and IC methods. Experimental results demonstrate that the Fuzzy–DDPG approach outperforms these classical techniques by achieving a higher tracking efficiency of 95%, compared to 85% for P&O and 88% for IC in average, while also minimizing steady-state oscillations. Additionally, the proposed method reduces tracking errors by up to 7.9% compared to conventional MPPT algorithms. These findings indicate that the combination of fuzzy logic and deep reinforcement learning provides a more adaptive and efficient MPPT solution, ensuring improved energy harvesting in dynamically changing conditions.



Academic Editor: Sangmo Kim

Received: 28 March 2025

Revised: 18 April 2025

Accepted: 26 April 2025

Published: 27 April 2025

Citation: Ortiz-Munoz, D.; Luviano-Cruz, D.; Perez-Dominguez, L.A.; Rodriguez-Ramirez, A.G.; Garcia-Luna, F. Hybrid Fuzzy–DDPG Approach for Efficient MPPT in Partially Shaded Photovoltaic Panels. *Appl. Sci.* **2025**, *15*, 4869. <https://doi.org/10.3390/app15094869>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: maximum power point tracking (MPPT); partial shading; reinforcement learning; deep deterministic policy gradient (DDPG); fuzzy logic; photovoltaic systems; hybrid control; machine learning; energy optimization

1. Introduction

Partial shading is a prevalent issue in photovoltaic (PV) systems that significantly degrades performance by introducing multiple local maxima in the power–voltage (P–V) curve, thereby complicating the identification of the global Maximum Power Point (MPP) via conventional Maximum Power Point Tracking (MPPT) algorithms [1–3]. To overcome this challenge, researchers have increasingly employed advanced optimization techniques—particularly metaheuristic algorithms—to reliably locate the global MPP under non-uniform irradiance.

Heuristic methods inspired by natural phenomena, such as the Flower Pollination Algorithm [4] and Genetic Algorithms [5], have demonstrated the capability to evolve candidate solutions toward global optimum performance. Additional approaches including Particle Swarm Optimization [6], Ant Colony Optimization, Simulated Annealing, Artificial Bee Colony, Firefly Algorithm, Cuckoo Search, and Differential Evolution [7] have

been applied to MPPT; however, these methods often encounter high computational complexity, sensitivity to parameter tuning, slow convergence, and practical implementation constraints [8,9].

Complementary to metaheuristic strategies, artificial intelligence (AI)-based techniques—such as fuzzy logic control and neural networks—have been proposed to address the MPPT challenge. Hybrid methods, for instance the integration of the shuffled frog leaping algorithm with the incremental conductance method [10], have enhanced global MPP tracking under partial shading. Nevertheless, these AI-based methods frequently require extensive training data, which increases memory usage and may lead to longer optimization times and transient power losses [11].

Classical control strategies, including Proportional–Integral–Derivative (PID) control [12,13] and Sliding Mode Control (SMC) [14,15], provide alternative solutions but often fail to manage the nonlinear dynamics and oscillatory behavior inherent in PV systems under partial shading. Other techniques such as distributed control, optimal control, backstepping, divide-and-conquer strategies [16–18], and simpler methods like Perturb and Observe (P&O) [19], have been explored with limited success [20,21].

Recent advances in MPPT for photovoltaic systems under partial shading conditions have led to the development of several innovative control strategies. In [22], an adapted model predictive control algorithm is proposed to robustly track the global maximum power point by predicting future system states and optimizing converter duty cycles, thereby reducing reliance on multiple sensors. Meanwhile, ref. [23] introduces a bi-objective reinforcement learning framework for dynamic PV array reconfiguration that simultaneously maximizes power extraction and minimizes switching operations in the presence of moving clouds. Furthermore, Ref. [24] presents a novel approach that integrates fuzzy logic control with optimizable Gaussian Process Regression to enhance both tracking accuracy and convergence speed under complex environmental conditions. Finally, an adapted model predictive control algorithm has also been proposed to improve GMPP tracking under partial shading by minimizing the MPC cost function without relying on current sensors [25].

These contributions collectively underscore the potential of combining advanced control techniques with artificial intelligence to overcome the limitations of traditional MPPT methods.

The research motivation behind this work is to develop an MPPT approach that overcomes these limitations by reducing the training burden and enhancing convergence speed. Reinforcement learning (RL) has emerged as a promising paradigm for MPPT under partial shading, with methods such as Deep Deterministic Policy Gradient (DDPG) and its variants effectively managing continuous action spaces [26–32]. However, the substantial data and computational resources required by these methods, along with their sensitivity to reward function design, remain critical challenges.

In this context, the present work proposes a novel hybrid approach that integrates fuzzy techniques with DDPG. By fuzzifying the input states to the actor network, the method significantly reduces the number of training episodes required, leading to faster convergence and improved robustness compared to conventional methods such as Perturb and Observe, Incremental Conductance, and standalone DDPG [33,34].

The principal contributions of this paper are the following:

1. Development of a novel fuzzy-based reinforcement learning framework: A hybrid MPPT strategy that combines fuzzy logic with DDPG to enhance global MPP tracking under partial shading.
2. Comprehensive validation: Detailed simulation and experimental results that corroborate the improved convergence speed, tracking accuracy, and robustness of the proposed method.

3. Comparative analysis: A systematic comparison with conventional MPPT techniques, highlighting the advantages of the proposed approach in real-time PV system control.

2. Materials and Methods

In this section, we provide an overview of the methodologies employed to address the Maximum Power Point Tracking (MPPT) challenge in photovoltaic systems under partial shading conditions. Initially, we describe a rigorous mathematical model based on the single-diode equivalent circuit to accurately capture the nonlinear behavior of the photovoltaic panel. Subsequently, our proposed control strategy, which integrates fuzzy logic with Deep Deterministic Policy Gradient (DDPG), is introduced to enhance state representation and expedite convergence in complex, non-convex operational environments. Finally, a theoretical justification rooted in the contraction properties of the Bellman operator is presented to establish the stability and robustness of the hybrid approach.

2.1. Mathematical Model of a Photovoltaic Panel

To comprehensively characterize the nonlinear behavior of the photovoltaic panel, the following mathematical model is presented. By employing the single-diode equivalent circuit, this formulation encapsulates the interactions among photocurrent generation, diode conduction dynamics, and resistive losses, thereby providing a robust foundation for developing advanced MPPT strategies. This model is essential for deriving control laws that guarantee convergence to the global maximum power point under varying irradiance and temperature conditions, including partial shading.

As shown in Figure 1, the equivalent circuit of a photovoltaic (PV) cell consists of a current source representing the photocurrent (I_{ph}), a diode (D) modeling the p-n junction behavior, a series resistance (R_s), and a parallel (shunt) resistance (R_{sh}) that accounts for leakage currents.

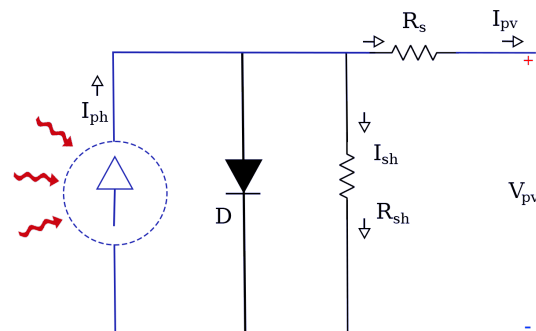


Figure 1. Single-diode equivalent circuit model of a photovoltaic (PV) cell, including series resistance R_s and shunt resistance R_{sh} . The photocurrent I_{ph} is generated due to incident solar irradiation, while the diode D represents the intrinsic p-n junction behavior.

The output current I_{pv} of the PV cell is given by Kirchhoff's law as follows [33]:

$$I_{pv} = I_{ph} - I_d - I_{sh} \quad (1)$$

where:

- I_{ph} is the light-generated current, proportional to the solar irradiance.
- I_d is the diode current.
- I_{sh} is the shunt leakage current.

The shunt current is modeled as follows:

$$I_{sh} = \frac{V + I_{pv}R_s}{R_p} \quad (2)$$

where:

- R_s is the series resistance.
- R_p is the parallel resistance.

The diode current I_d follows the Shockley diode equation:

$$I_d = I_0 \left(\exp \left(\frac{qV_d}{AkT_c} \right) - 1 \right) \quad (3)$$

where:

- I_0 is the reverse saturation current of the diode.
- $q = 1.6 \times 10^{-19}$ C is the electronic charge.
- $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant.
- A is the diode ideality factor.
- T_c is the cell temperature in Kelvin.
- V_d is the voltage across the diode:

$$V_d = V + I_{pv}R_s \quad (4)$$

The light-generated current I_{ph} is affected by solar irradiance and temperature:

$$I_{ph} = (I_{sc} + K_I(T_c - T_r)) \times \frac{G}{G_{STC}} \quad (5)$$

where:

- I_{sc} is the short-circuit current at standard test conditions (STC) ($T = 25^\circ\text{C}$, $G_{STC} = 1000 \text{ W/m}^2$).
- K_I is the short-circuit current temperature coefficient.
- T_r is the reference temperature.
- G is the actual solar irradiance.

For a PV module with N_s series-connected cells, the overall output current is as follows:

$$I_{pv} = I_{ph} - I_0 \left(\exp \left(\frac{q(V + I_{pv}R_s)}{AkT_cN_s} \right) - 1 \right) \quad (6)$$

2.2. Reinforcement Learning Fuzzy-DDPG

Reinforcement learning (RL), as discussed in [35], is a machine learning paradigm in which an agent learns to make optimal decisions by interacting with an environment. This process is typically formulated as a Markov Decision Process (MDP), which is characterized by key components:

- State (S): Represents the current situation or context of the agent. At each time step, the agent observes the state of the environment.
- Action (A): The agent selects an action from a set of possible actions based on the current state.
- Transition Probability (P): $P(s' | s, a)$ denotes the probability of transitioning to state s' after taking action a in state s . This captures the dynamics of the environment.
- Reward (R): $R(s, a, s')$ represents the immediate reward received by the agent after taking action a in state s and transitioning to state s' .
- Discount Factor (γ): A value between 0 and 1 that determines the importance of future rewards. A higher discount factor prioritizes long-term rewards.
- Policy (π): A mapping from states to actions, defining the agent's behavior. The policy can be deterministic, $\pi(s) = a$, or stochastic, $\pi(a|s) = P(a|s)$.

The primary objective in RL is to find an optimal policy that maximizes the expected cumulative discounted reward, formally expressed as follows:

$$R(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, s_{t+1}) \right]. \quad (7)$$

In continuous state spaces, the exponential growth of the state–action pair complexity—often referred to as the curse of dimensionality—renders traditional tabular methods impractical. This challenge necessitates the use of function approximation techniques, where neural networks and other parametric models are employed to estimate value functions or to directly parameterize policies. Such approaches have led to the development of advanced RL methods including Deep Q-Networks (DQN), policy gradient methods, and actor–critic architectures like Deep Deterministic Policy Gradient (DDPG) [36].

Beyond these foundational aspects, modern RL methods have evolved to incorporate strategies such as curriculum learning and adaptive exploration [37], which further enhance training stability and convergence. In our study, RL is applied to dynamically adjust the duty cycle of a DC–DC converter in a photovoltaic system under partial shading conditions. This formulation not only addresses the inherent non-stationarity and high-dimensionality of the environment but also demonstrates how RL, when combined with fuzzy logic enhancements, provides a robust and adaptive framework for achieving Maximum Power Point Tracking in complex, real-world settings.

In the context of Maximum Power Point Tracking for photovoltaic systems under partial shading conditions, the RL agent learns to adjust the duty cycle of a DC–DC converter to maximize the power output of the PV system. Partial shading occurs when some parts of a PV array receive less sunlight than others, leading to multiple peaks in the power–voltage curve. Traditional MPPT methods can get stuck at local maxima, resulting in suboptimal power generation. RL offers a promising solution to this problem by learning to navigate the complex power landscape and find the global maximum power point.

Traditional RL algorithms like Q-learning [38] and SARSA are designed for discrete state and action spaces. Extensions to continuous spaces often involve function approximation techniques, such as tile coding [39], radial basis functions (a common function approximation technique, though not directly referenced in the current search results), or neural networks (neural networks are widely used in RL, see for example [40]). However, these approximations can suffer from instability and divergence issues [41].

Policy gradient methods, which directly optimize the policy, are naturally suited for continuous action spaces. Algorithms like REINFORCE and actor–critic methods have been developed to handle such scenarios. The Deterministic Policy Gradient algorithm [42] extends the policy gradient theorem to deterministic policies, laying the groundwork for DDPG [43].

The deterministic policy gradient algorithm [42] states that the gradient of the expected return with respect to the policy parameters θ^μ is:

$$\nabla_{\theta^\mu} J = \mathbb{E}_{s \sim \rho^\mu} \left[\nabla_{\theta^\mu} \mu(s|\theta^\mu) \nabla_a Q^\mu(s, a) \Big|_{a=\mu(s|\theta^\mu)} \right] \quad (8)$$

where:

- $Q^\mu(s, a)$ is the action-value function under policy μ .
- $\rho^\mu(s)$ is the discounted state visitation distribution under μ .

In [44] is introduced a fuzzy Q-iteration algorithm for approximate dynamic programming, addressing the curse of dimensionality by using a fuzzy partition of the state space and discretization of the action space to approximate the Q-function, making it applicable

to continuous state and action spaces. This approach allows for handling larger state spaces than traditional dynamic programming, which struggles with the exponential growth of computational complexity as the number of state variables increases.

Fuzzy Q-Iteration approximates the Q-function using fuzzy basis functions:

$$Q(s, a) \approx \sum_{i=1}^N w_i \phi_i(s, a) \quad (9)$$

where:

- $\phi_i(s, a)$ are fuzzy basis functions representing the degree of membership of (s, a) in fuzzy sets.
- w_i are the weights to be learned.

The update rule minimizes the Bellman residual:

$$w^{(k+1)} = \arg \min_w \sum_{(s,a)} \left(\mathcal{R}(s, a) + \gamma \max_{a'} Q^{(k)}(s', a') - \sum_{i=1}^N w_i \phi_i(s, a) \right)^2 \quad (10)$$

The fuzzy Q-iteration algorithm iteratively improves the Q-function approximation, converging to the optimal Q-function under certain conditions.

The proposed method combines the strengths of fuzzy techniques and DDPG to achieve global Maximum Power Point Tracking under partial shading conditions. The fuzzy component handles the action space, while the DDPG component addresses the continuous state space, resulting in a hybrid approach that is both scalable and able to operate in complex environments [28,45].

By integrating fuzzy basis functions into the actor and critic networks of DDPG, we aim to approximate both the policy and value function using fuzzy features, as follows:

$$\text{Actor Network: } \mu(\phi(s)|\theta^\mu) \quad (11)$$

$$\text{Critic Network: } Q(\phi(s), a|\theta^Q)$$

The fuzzy feature vector $\phi(s)$ captures the essential characteristics of the state space, facilitating learning in continuous domains. This hybrid approach combines the strengths of fuzzy approximation and deep reinforcement learning, enabling the agent to efficiently navigate the complex power landscape and find the global maximum power point under partial shading conditions.

For each state variable s_i , we define M membership functions $\{\mu_{ij}\}_{j=1}^M$, typically Gaussian functions:

$$\mu_{ij}(s_i) = \exp\left(-\frac{(s_i - c_{ij})^2}{2\sigma_{ij}^2}\right) \quad (12)$$

The parameters are defined as follows:

- c_{ij} : Center of the j -th membership function for variable s_i .
- σ_{ij} : Width (standard deviation) of the membership function.

The actor and critic networks take this fuzzy feature vector as input, and the output represents the actor policy and critic value function, respectively.

The actor network architecture is as follows:

- Input Layer: Receives fuzzy feature vector $\phi(s)$.
- Hidden Layers: Multiple fully connected layers with nonlinear activations.
- Output Layer: Produces action $a = \mu(\phi(s)|\theta^\mu)$, using a bounded activation function like tanh.

The critic network architecture is as follows:

- State Input Layer: Receives fuzzy feature vector $\phi(s)$.
- Action Input Layer: Receives action a .
- Hidden Layers: Concatenates $\phi(s)$ and a , followed by fully connected layers.
- Output Layer: Outputs scalar value $Q(\phi(s), a|\theta^Q)$.

The actor network processes the fuzzy feature vector $\phi(s)$ obtained from the state s . It comprises two fully connected hidden layers with ReLU activations, and an output layer with a bounded activation function (e.g., tanh) to produce the action:

$$a = \mu(\phi(s); \theta_\mu) \quad (13)$$

The critic network receives the concatenated fuzzy feature vector and the action, and processes this input through two hidden layers with ReLU activations to output the scalar Q-value:

$$Q(\phi(s), a; \theta_Q) \quad (14)$$

The fuzzy feature vector $\phi(s)$ captures the essential characteristics of the state space, facilitating learning in continuous domains. By integrating fuzzy basis functions into the actor and critic networks, we aim to approximate both the policy and value function using fuzzy features, enabling efficient navigation of the complex power landscape and global Maximum Power Point tracking under partial shading conditions.

The training procedure follows the standard DDPG algorithm, with the additional incorporation of fuzzy basis functions, according to Algorithms 1 and 2.

Algorithm 1 At each time step

- 1: **repeat**
 - 2: Compute Fuzzy Features: $\phi(s_t)$ from current state s_t .
 - 3: Select Action: $a_t = \mu(\phi(s_t)|\theta^\mu) + \mathcal{N}_t$, where $\mathcal{N}_t$ is exploration noise.
 - 4: Execute Action: Receive reward r_t and next state s_{t+1} .
 - 5: Store Transition: Store (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$.
 - 6: **until** Stopping Criterion is Met
-

Algorithm 2 At each learning step

- 1: **repeat**
- 2: **Sample Mini-Batch:** Randomly sample N transitions from $\mathcal{D}$.
- 3: **Compute Targets:**

$$y_i = r_i + \gamma Q'(\phi(s_{i+1}), \mu'(\phi(s_{i+1})|\theta^{\mu'})|\theta^{Q'})$$

- 4: **Update Critic:** Minimize loss L :

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - Q(\phi(s_i), a_i|\theta^Q))^2$$

- 5: **Update Actor:** Update policy parameters using gradient ascent:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_{i=1}^N \nabla_{\theta^\mu} \mu(\phi(s_i)|\theta^\mu) \nabla_a Q(\phi(s_i), a|\theta^Q)|_{a=\mu(\phi(s_i)|\theta^\mu)}$$

- 6: **Update Target Networks:**

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

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

- 7: **until** Stopping Criterion is Met
-

This training procedure allows the agent to learn the optimal policy and value function by leveraging the fuzzy feature representation, enabling global Maximum Power Point Tracking under partial shading conditions.

Algorithm 3 in the form of pseudocode could be used to implement the proposed method.

Algorithm 3 Fuzzy-DDPG Algorithm

```

1: Initialize actor network  $\mu(\phi(s)|\theta^\mu)$  and critic network  $Q(\phi(s), a|\theta^Q)$ 
2: Initialize target networks  $\theta^{\mu'} \leftarrow \theta^\mu, \theta^{Q'} \leftarrow \theta^Q$ 
3: Initialize replay buffer  $\mathcal{D}$ 
4: for episode = 1 to max_episodes do
5:   Initialize state  $s_0$ 
6:   for  $t = 1$  to max_steps do
7:     Compute fuzzy features  $\phi(s_t)$ 
8:     Select action  $a_t = \mu(\phi(s_t)|\theta^\mu) + \mathcal{N}_t$ 
9:     Execute action  $a_t$ , observe reward  $r_t$  and next state  $s_{t+1}$ 
10:    Store transition  $(s_t, a_t, r_t, s_{t+1})$  in  $\mathcal{D}$ 
11:    Sample mini-batch of  $N$  transitions from  $\mathcal{D}$ 
12:    for each sample  $(s_i, a_i, r_i, s_{i+1})$  do
13:      Compute target:

```

$$y_i = r_i + \gamma Q'(\phi(s_{i+1}), \mu'(\phi(s_{i+1})|\theta^{\mu'})|\theta^{Q'})$$

```

14:   end for
15:   Update critic by minimizing:

```

$$L = \frac{1}{N} \sum_i \left(y_i - Q(\phi(s_i), a_i|\theta^Q) \right)^2$$

```

16:   Update actor using policy gradient:

```

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_{\theta^\mu} \mu(\phi(s_i)|\theta^\mu) \nabla_a Q(\phi(s_i), a|\theta^Q) \Big|_{a=\mu(\phi(s_i)|\theta^\mu)}$$

```

17:   Update target networks:

```

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

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

```

18:   end for
19: end for

```

Figure 2 illustrates the main steps of the algorithm, depicting how the state observations from the PV environment are processed to update both the critic and actor networks through a gradient-based rule. Convergence conditions are periodically checked to ensure that the control policy remains near-optimal. The final result is an adaptive, data-driven control strategy capable of guaranteeing a stable operating point at or near the global maximum power region of the PV system.

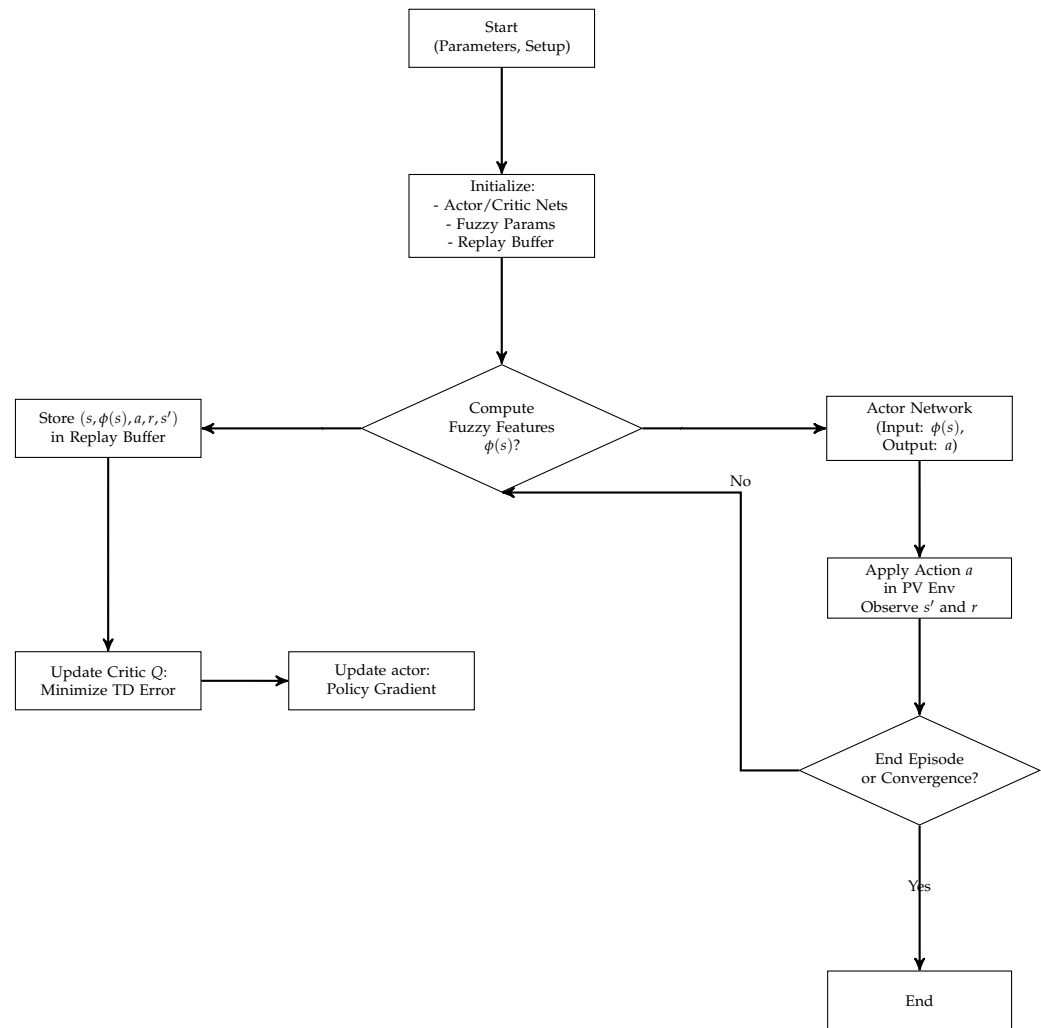


Figure 2. Flow diagram illustrating the proposed control scheme for Maximum Power Point Tracking in a PV system.

2.3. Proof

We present a formal proof of the convergence of a hybrid approach that combines Fuzzy Q-iteration with the Deep Deterministic Policy Gradient (DDPG) algorithm. The main goal is to show that, under standard assumptions (bounded rewards, proper learning rates, sufficient function approximators, etc.), the proposed method converges to a near-optimal policy. Our proof proceeds in three main steps:

1. **Fuzzy Q-iteration Convergence:** We show that the fuzzy Bellman operator is a contraction, thereby ensuring that it converges to a fixed point.
2. **DDPG Critic Convergence:** We prove that the Critic network (Q-function approximation) converges to a close approximation of the optimal action-value function.
3. **DDPG Actor Convergence:** We demonstrate that the Actor parameters converge to a near-optimal deterministic policy given that the critic accurately approximates the Q-function.

2.4. Preliminaries and Notation

Definition 1 (State and Action Spaces). Let $S \subseteq \mathbb{R}^n$ be a (continuous) state space and $A \subseteq \mathbb{R}^m$ be a (continuous) action space.

Definition 2 (Reward Function). We assume a bounded reward function $r : S \times A \rightarrow \mathbb{R}$, i.e., there exists $R_{\max} \geq 0$ such that $\sup_{(s,a)} |r(s,a)| \leq R_{\max}$.

Definition 3 (Transition Dynamics). We assume a deterministic environment with transition dynamics $s_{t+1} = f(s_t, a_t)$.

2.5. Theorem

Theorem 1 (Convergence of fuzzy-DDPG algorithm). Consider a continuous state space $S \subseteq \mathbb{R}^n$, a continuous action space $A \subseteq \mathbb{R}^m$, and a discount factor $\gamma \in [0, 1)$. Let $Q(s, a)$ be the action-value function and $\pi_\theta(s)$ be a deterministic policy parameterized by θ . Assume the reward function $r(s, a)$ is bounded, the fuzzy membership functions $\{\phi_i(s)\}_{i=1}^N$ form a partition of unity, and the neural networks for critic and actor have sufficient representational capacity with appropriate learning rates. Then the combined Fuzzy-DDPG algorithm converges to a near-optimal policy $\pi^*(s)$ under the following conditions:

1. Fuzzy Q-iteration satisfies the contraction mapping property via the Bellman operator.
2. The critic network converges to a good approximation of the optimal Q-function $Q^*(s, a)$.
3. The actor network's policy gradient updates converge to a near-optimal deterministic policy.

Proof. We prove convergence in three major steps, corresponding to the numbered items in the theorem statement, according to [34,46,47].

Step 1: Fuzzy Convergence

We start with the Bellman operator $\mathcal{T}$ for a Q-function Q :

$$\mathcal{T}(Q)(s, a) = r(s, a) + \gamma \max_{a'} Q(s', a') \quad (15)$$

where s' is the next state determined by the environment's transition dynamics $f(s, a)$.

In the fuzzy framework, the Q-function is approximated as follows:

$$\hat{Q}(s, a) = \sum_{i=1}^N \phi_i(s) \theta_i^T a \quad (16)$$

where $\phi_i(s)$ are fuzzy membership functions (satisfying $\sum_{i=1}^N \phi_i(s) = 1$ for each s) and θ_i are the learnable parameters for each fuzzy region i .

Contraction Property. For any two Q-functions Q and Q' , the Bellman operator satisfies:

$$\|\mathcal{T}(Q) - \mathcal{T}(Q')\|_\infty \leq \gamma \|Q - Q'\|_\infty \quad (17)$$

which implies that $\mathcal{T}$ is a contraction mapping with factor $\gamma \in [0, 1)$. The fuzzy partition $\{\phi_i\}$ respects this contraction because of the following:

$$\|\mathcal{T}(\hat{Q}) - \mathcal{T}(\hat{Q}')\|_\infty \leq \gamma \|\hat{Q} - \hat{Q}'\|_\infty \quad (18)$$

Consequently, repeated application of $\mathcal{T}$ converges to a unique fixed point, i.e.,

$$\|\hat{Q}^k - \hat{Q}^*\|_\infty \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (19)$$

Hence, the fuzzy Q-iteration converges to an approximate optimal Q-function $\hat{Q}^*$.

Critic Objective.

In DDPG, the critic is represented by a neural network $Q_{\theta_c}(s, a)$ and is trained to minimize the mean-squared Bellman error:

Step 2: Critic Objective

$$L(\theta_c) = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1})} \left[(Q_{\theta_c}(s_t, a_t) - y_t)^2 \right] \quad (20)$$

where the target y_t is given by the following:

$$y_t = r_t + \gamma Q_{\theta'_c}(s_{t+1}, \pi_{\theta'_a}(s_{t+1})) \quad (21)$$

and θ'_c, θ'_a are the parameters of the slowly updated target networks.

Fuzzy Q-function Approximation in the Critic.

In the fuzzy context, the critic's Q-function is expressed as follows:

$$\hat{Q}(s_t, a_t) = \sum_{i=1}^N \phi_i(s_t) \theta_{c,i}^T a_t \quad (22)$$

where $\theta_{c,i}$ are the critic network's parameters for each fuzzy set i . The update rule for θ_c via gradient descent is the following:

$$\theta_c \leftarrow \theta_c - \alpha_c \nabla_{\theta_c} L(\theta_c) \quad (23)$$

with

$$\nabla_{\theta_c} L(\theta_c) = \mathbb{E} [2 (Q_{\theta_c}(s_t, a_t) - y_t) \nabla_{\theta_c} Q_{\theta_c}(s_t, a_t)] \quad (24)$$

Convergence of the Critic Network.

Critic convergence relies on:

- Bellman Operator Contraction: The targets y_t come from the contraction mapping $\mathcal{T}$.
- Gradient Descent to Reduce Bellman Error: Iterative updates align $Q_{\theta_c}(s, a)$ with the fixed point $Q^*(s, a)$.

Thus, for suitable learning rates and sufficient network capacity,

$$\|\hat{Q}_{\theta_c}^{k+1} - Q^*\|_2 \leq \beta \|\hat{Q}_{\theta_c}^k - Q^*\|_2 \quad (25)$$

where $\beta \in [0, 1)$ is a constant bounded by learning rate and fuzzy approximation smoothness. This iterative improvement implies $\hat{Q}_{\theta_c}^k \rightarrow Q^*$ in the limit.

Step 3: Actor Policy Convergence.

Policy Gradient Update.

The actor learns a deterministic policy $\pi_\theta(s)$ by following the gradient of the expected return:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{s_t} \left[\nabla_a Q_{\theta_c}(s_t, a) \nabla_\theta \pi_\theta(s_t) \Big|_{a=\pi_\theta(s_t)} \right] \quad (26)$$

The actor parameters are updated by gradient ascent to maximize this expected return.

Convergence of the Actor.

Once the critic converges close to $Q^*(s, a)$, the policy gradient points towards actions that maximize $Q^*(s, a)$. If the actor's learning rate is sufficiently small (ensuring stable updates), the actor converges to a local optimum $\pi^*(s)$, where

$$\nabla_\theta J(\pi_{\theta^*}) = 0 \quad (27)$$

Hence, $\pi_{\theta^*}(s)$ is a near-optimal deterministic policy that maximizes the expected return.

Conclusion.

Putting all steps together, we conclude:

1. *Fuzzy Q-iteration* converges to a fixed point $\hat{Q}^*$ that closely approximates $Q^*(s, a)$ due to the contraction property of $\mathcal{T}$.
2. The *DDPG critic* converges to the optimal Q-function $Q^*(s, a)$ by minimizing the Bellman error.
3. The *DDPG actor* converges to a near-optimal deterministic policy $\pi^*(s)$ by following the gradient derived from the critic's Q-function.

Therefore, under reasonable assumptions on the learning rates, neural network capacity, fuzzy partitioning, and exploration strategy, the hybrid Fuzzy-DDPG algorithm converges to a near-optimal policy. $\square$

This theorem guarantees convergence toward a near-optimal policy by exploiting the contraction property of the Bellman operator, even under complex fuzzy approximations. Such convergence is instrumental in MPPT under partial shading conditions in PV panels, as it systematically guides the search toward the global maximum rather than local optima. By accommodating continuous variations in solar irradiance, the proposed approach ensures robust policy updates that adaptively track the global peak under dynamically shifting conditions.

2.6. Theoretical Justification of Critic Convergence

In order to provide a deeper theoretical foundation for the convergence properties of the critic network in our proposed Fuzzy-DDPG framework, we now state a lemma that formalizes the convergence of the critic under fuzzy approximation. This result builds upon classical results in reinforcement learning and function approximation, and it is instrumental in validating the effectiveness of our approach under the following standard assumptions.

Lemma 1 (Critic Convergence under Fuzzy Approximation). *Let $S \subseteq \mathbb{R}^n$ and $A \subseteq \mathbb{R}^m$ be the continuous state and action spaces, respectively, and let $Q^* : S \times A \rightarrow \mathbb{R}$ denote the unique fixed point of the Bellman operator*

$$(TQ)(s, a) = r(s, a) + \gamma \max_{a' \in A} Q(s', a') \quad (28)$$

with $s' = f(s, a)$, where f represents the deterministic state transition function and $\gamma \in [0, 1)$ is the discount factor. Assume the following:

1. The Bellman operator T is a contraction mapping, i.e.,

$$\|TQ - TQ'\|_{\infty} \leq \gamma \|Q - Q'\|_{\infty}, \quad \forall Q, Q' : S \times A \rightarrow \mathbb{R}. \quad (29)$$

2. The critic network $Q_{\theta_c}(s, a)$, parameterized by θ_c , is a universal approximator on $S \times A$. In other words, for every $\epsilon > 0$, there exists a parameter set θ_c such that

$$\|Q_{\theta_c} - Q^*\|_{\infty} < \epsilon \quad (30)$$

3. The learning rates $\{\alpha_t\}_{t \geq 1}$ satisfy the standard Robbins–Monro conditions:

$$\sum_{t=1}^{\infty} \alpha_t = \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \alpha_t^2 < \infty. \quad (31)$$

4. The target network parameters are updated using a slow update factor $\tau \in (0, 1)$, ensuring that the target network evolves sufficiently slowly.
5. The additive noise in the target computation is zero-mean and uniformly bounded.

Then, the iterative update of the critic network via stochastic gradient descent,

$$\theta_c^{(t+1)} = \theta_c^{(t)} - \alpha_t \nabla_{\theta_c} L(\theta_c^{(t)}) \quad (32)$$

with the loss function defined as

$$L(\theta_c) = \mathbb{E} \left[\left(Q_{\theta_c}(s, a) - \left(r(s, a) + \gamma Q_{\theta'_c}(s', \pi_{\theta'_a}(s')) \right) \right)^2 \right] \quad (33)$$

converges almost surely to a parameter set θ_c^* such that

$$\|Q_{\theta_c^*} - Q^*\|_{\infty} = 0 \quad (34)$$

The above lemma rigorously establishes that, under the appropriate conditions on the function approximators and learning parameters, the critic network in the Fuzzy-DDPG framework converges to the optimal action-value function. This result, combined with analogous arguments for the actor network, provides a solid theoretical basis for the overall convergence of our hybrid approach.

3. Results

The performance of the proposed hybrid Maximum Power Point Tracking (MPPT) control strategy, which integrates fuzzy techniques with Deep Deterministic Policy Gradient (DDPG), is evaluated on a simulated photovoltaic (PV) panel operating under partial shading conditions. The simulation results indicate that the proposed method yields significant improvements compared to conventional techniques such as Perturbation and Observation (P&O) and Incremental Conductance (IC), both of which tend to exhibit oscillatory behavior around local maxima and are consequently unable to reliably attain the global optimum.

Table 1 summarizes the key hyperparameters and network configuration employed in the Fuzzy-DDPG implementation.

The partially shaded PV panel exhibits a complex, non-convex power–voltage characteristic with multiple local maxima. Under these challenging conditions, the Fuzzy-DDPG method is capable of converging to the global maximum power point, thereby surpassing baseline methods that are confined to local optima.

In the following section, we present a comprehensive analysis of the experimental findings associated with the proposed hybrid MPPT control scheme. This evaluation includes a comparative study against conventional MPPT approaches, namely Perturb and Observe (P&O) and Incremental Conductance (IC).

The photovoltaic system utilized in the simulations is based on a commercially available panel. Table 2 details the electrical characteristics of the panel, including its rated power, voltage, and current at the maximum power point (MPP), as well as other essential parameters for precise simulation modeling.

Table 1. Hyperparameters and configuration of the fuzzy DDPG agent.

Parameter	Value
Fuzzy Membership Functions	
Number of membership functions per variable	7
Type of membership functions	Gaussian
Number of state variables	7
State variables	Voltage, Current, Power, G1, G2, ΔP , ΔV
Actor Network Architecture	
Input	7 (States with fuzzy features)
Normalization layer	Yes
Fuzzy feature layer	Yes
Number of hidden layers	2
Number of neurons per layer	128
Activation function	ReLU
Output layer	1 neuron (action) with Tanh
Critic Network Architecture	
Input	7 (States) + 1 (Action)
Normalization layer	Yes
Fuzzy feature layer	Yes
Number of hidden layers	2
Number of neurons per layer	128
Activation function	ReLU
Output layer	1 neuron (Q-Value)
Training Hyperparameters	
Actor learning rate	10^{-5}
Critic learning rate	10^{-4}
Discount factor (γ)	0.99
Batch size	256
Target smooth factor (τ)	10^{-3}
Experience buffer size	10^6
Training Options	
Number of episodes	100
Maximum steps per episode	15,000
Score averaging window length	10
Sampling time interval	0.06 s

Table 2. Specifications of the photovoltaic panel used in simulations.

Parameter	Value
Number of Parallel Strings	1
Number of Series-Connected Modules per String	1
Maximum Power (P_{max}) [W]	352.408
Cells per Module (N_{cell})	60
Open-Circuit Voltage (V_{oc}) [V]	49.9
Short-Circuit Current (I_{sc}) [A]	9.0
Voltage at Maximum Power Point (V_{mp}) [V]	40.6
Current at Maximum Power Point (I_{mp}) [A]	8.68
Temperature Coefficient of V_{oc} [%/°C]	−0.36099
Light-Generated Current (I_L) [A]	9.1365
Diode Saturation Current (I_0) [A]	3.3113×10^{-10}
Diode Ideality Factor	1.3468
Shunt Resistance (R_{sh}) [Ω]	1063.7078

3.1. Reward Function Design

The reward function used in this study is designed to ensure optimal tracking of the Maximum Power Point (MPP) while penalizing undesirable behaviors such as excessive

voltage fluctuations. The total reward R is defined as the sum of multiple components, each addressing a specific aspect of the MPPT process:

$$R = r_{\text{mpp}} + r_1 + r_2 + r_3 + r_4 + r_5 \quad (35)$$

where each term contributes to guiding the agent towards maximizing power extraction while maintaining system stability. Table 3 summarizes the definition of each component.

Table 3. Definition of the reward function components.

Component	Description
r_{mpp}	Penalizes deviation from the maximum observed power $P_{\text{MPP,max}}$
r_1	Proportional reward based on the extracted power
r_2	Encourages power increase based on ΔP
r_3	Penalizes voltages beyond the open-circuit voltage limit
r_4	Penalizes sudden voltage variations (dV)
r_5	Additional penalty for large voltage changes

The components of the reward function are computed as follows:

$$r_{\text{mpp}} = \begin{cases} -2 \frac{|P - P_{\text{MPP,max}}|}{P_{\text{MPP,max}}}, & \text{if } P_{\text{MPP,max}} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (36)$$

$$r_1 = \begin{cases} 0.1 \frac{P}{P_{\text{MPP,max}}}, & \text{if } P_{\text{MPP,max}} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (37)$$

$$r_2 = \text{sign}(\Delta P) \quad (38)$$

$$r_3 = \begin{cases} 0, & 0 \leq V_{\text{new}} \leq V_{\text{oc,total}} \\ -2, & \text{otherwise} \end{cases} \quad (39)$$

$$r_4 = -0.1 |dV| \quad (40)$$

$$r_5 = -0.5 \frac{|dV|}{2} \quad (41)$$

This function encourages the agent to maximize the extracted power while preventing excessive voltage variations that could lead to instability. The inclusion of r_3 ensures that the voltage remains within the safe operating range, while r_4 and r_5 penalize abrupt changes, promoting smooth control actions.

The results are structured to illustrate the following:

- **Methodological Comparisons:** The performance metrics—such as global maximum power point (GMPP) tracking efficiency, stability, and adaptability under partial shading conditions—are systematically analyzed across methods.
- **Statistical Analysis:** Confidence intervals and hypothesis testing are employed to ensure statistical robustness in the comparisons, demonstrating the reliability and superiority of the proposed approach.
- **Dynamic and Steady-State Behaviors:** The dynamic tracking response and steady-state oscillations of each method are critically evaluated. Specific emphasis is placed on the proposed method's ability to avoid local maxima and achieve stable operation without oscillations.
- **Environmental Sensitivity:** The methods are tested across various irradiance and temperature scenarios to assess robustness against environmental fluctuations.

Through the following subsections, the data are presented via boxplots, performance curves, and statistical summaries to ensure comprehensive and precise insights. Experimental conclusions are drawn to validate the effectiveness of the proposed hybrid MPPT algorithm.

3.2. Experimental Design

Factors, Levels, and Response Variables

Two main factors dictate the operational state of PV modules:

- Irradiance G : $\{800, 600, 400\}$ W/m².
- Temperature T : $\{25^\circ\text{C}, 35^\circ\text{C}, 45^\circ\text{C}\}$.

By combining these three irradiance levels with three temperature settings, we obtain nine distinct environmental configurations. For each MPPT method, we execute 30 independent runs in each configuration, yielding $9 \times 30 = 270$ data points per method.

The chosen performance metrics are the following:

1. Efficiency $\eta \in [0, 1]$, measuring how close the controller maintains the operating point to the true maximum power.
2. Oscillations $\sigma \in \mathbb{R}_{\geq 0}$, capturing the amplitude of power fluctuations at or near the steady state.

In more formal terms,

$$\eta = \frac{P_{\text{extracted}}}{P_{\text{max}}}, \quad \sigma = \max_{t \in \mathcal{T}} |P(t) - \bar{P}|,$$

where P_{max} is the theoretical maximum power under the current G and T , $P(t)$ is the instantaneous power output, and $\bar{P}$ is the mean steady-state power.

3.3. Data Generation and Implementation

1. Algorithmic Loop: For each of the three methods, Fuzzy-DDPG, P&O, and IC, we invoke a simulation procedure that produces η and σ values for each run.
2. Environmental Loop: For each irradiance level $G \in \{800, 600, 400\}$ and temperature $T \in \{25^\circ\text{C}, 35^\circ\text{C}, 45^\circ\text{C}\}$, simulation data were generated to reflect realistic operational noise and parameter uncertainties typical of real-world PV modules.
3. Replicas: For each (G, T) pair, 30 runs are executed, ensuring a broad sampling of random variations.

The implementation is conducted in Matlab/Simulink, integrating the PV panel models, MPPT algorithms, and statistical analysis tools. As shown in Table 4, the proposed Fuzzy-DDPG method demonstrates superior efficiency and minimal oscillations compared to the traditional methods.

Table 4. Approximate statistics of generated data for MPPT methods.

Method	η (Efficiency)	σ (Oscillations)
fuzzy-DDPG	0.95 ± 0.02	0.01 ± 0.005
P_O	0.85 ± 0.05	0.05 ± 0.01
IC	0.88 ± 0.03	0.03 ± 0.007

3.4. Statistical Analysis

Descriptive Statistics and Confidence Intervals

Let $\{x_{m,1}, x_{m,2}, \dots, x_{m,n}\}$ denote the sample of either efficiencies or oscillations for method m . We define the following:

- Sample Mean

$$\bar{x}_m = \frac{1}{n} \sum_{k=1}^n x_{m,k} \quad (42)$$

where $n = 270$ is the total number of observations per method. Although the notation $x_{m,k}$ might be interpreted as a multivariate observation, in our analysis, each $x_{m,k}$ represents a unidimensional measurement for method m at the k -th trial, where n is the total number of observations per method

- Sample Standard Deviation

$$s_m = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_{m,k} - \bar{x}_m)^2} \quad (43)$$

where n is the number of observations for method m , $x_{m,k}$ represents the k th univariate measurement, and $\bar{x}_m$ is the sample mean of the method m .

- 95% Confidence Interval

$$CI_{95\%}(\bar{x}_m) = \bar{x}_m \pm 1.96 \frac{s_m}{\sqrt{n}} \quad (44)$$

Such intervals measure the precision of the estimated mean. Notably, if the CIs of two methods do not overlap, we infer a significant difference in means at the 5% level.

Given that each method was evaluated using $n = 270$ independent trials, we invoke the Central Limit Theorem (CLT), which states that the sampling distribution of the sample mean approximates a normal distribution for sufficiently large n , regardless of the underlying distribution. Consequently, the standard normal distribution was used to compute the 95% confidence intervals. The critical value $z = 1.96$ corresponds to the two-tailed 95% confidence level under the standard normal distribution. While Student's t -distribution is more appropriate for small sample sizes ($n < 30$), for $n = 270$, the difference between z and the t critical value is negligible. Therefore, the use of $z = 1.96$ is statistically justified in this context.

The boxplots in Figure 3 are constructed for each method to illustrate the distribution of η and σ . The interquartile range (IQR) is encompassed by the box, the horizontal line denotes the median, and outliers beyond $1.5 \times \text{IQR}$ are marked with a $+$. This visual layout provides an immediate sense of central tendency and variability, highlighting that Fuzzy-DDPG typically clusters near higher efficiency with lower spread in oscillations compared to the other two methods.

The results summarized in Table 4 align with the statistical distributions observed in the boxplot comparisons shown in Figure 3. The proposed method, Fuzzy-DDPG, demonstrates the highest efficiency ($\eta \sim 0.95 \pm 0.02$) with minimal oscillations ($\sigma \sim 0.01 \pm 0.005$), as evident from its tightly grouped boxplot distributions in both subplots. In contrast, P_0 exhibits significantly wider variation in efficiency ($\eta \sim 0.85 \pm 0.05$) and higher oscillations ($\sigma \sim 0.05 \pm 0.01$), as indicated by the broader box and whisker ranges. Similarly, IC achieves intermediate performance with $\eta \sim 0.88 \pm 0.03$ and $\sigma \sim 0.03 \pm 0.007$, reflecting a compromise between efficiency and stability.

The non-overlapping CIs between Fuzzy-DDPG and the other two methods confirm that the observed differences in both efficiency and oscillation are statistically significant at the 5% level.

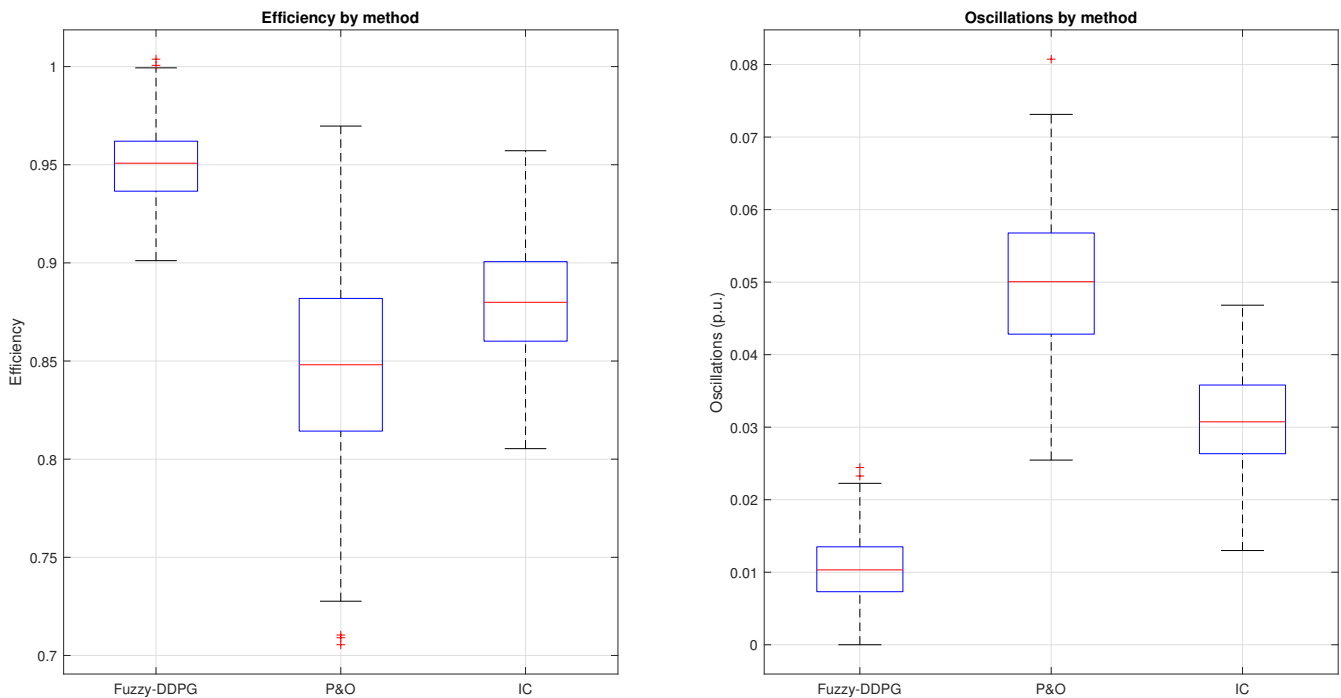


Figure 3. Boxplot comparison.

The following set of simulations demonstrates the performance of the Fuzzy-DDPG algorithm in comparison to conventional methods, including Perturbation and Observation (P&O) and Incremental Conductance (IC), under various irradiance and temperature scenarios designed to reflect realistic and challenging operating conditions for two PV panels in series.

- Case 1: Uniform Irradiance with Different Temperatures. $G_1 = G_2 = 1000 \text{ W/m}^2$, $T_1 = 25^\circ\text{C}$, $T_2 = 45^\circ\text{C}$. This scenario highlights the impact of temperature variation on tracking efficiency.
- Case 2: Partial Shading with Moderate Temperature. $G_1 = 1000 \text{ W/m}^2$, $G_2 = 500 \text{ W/m}^2$, $T_1 = T_2 = 35^\circ\text{C}$. This case replicates conditions where one panel is partially shaded by clouds or objects.
- Case 3: Severe Partial Shading with Low Temperature. $G_1 = 400 \text{ W/m}^2$, $G_2 = 200 \text{ W/m}^2$, $T_1 = T_2 = 15^\circ\text{C}$. This scenario reflects early morning or cold climate conditions with extreme shading.
- Case 4: Dynamic Environmental Changes. G_1 and G_2 vary linearly from 1000 W/m^2 to 600 W/m^2 and from 800 W/m^2 to 400 W/m^2 over 20 s, respectively, with $T_1 = T_2 = 30^\circ\text{C}$. This models transient conditions such as moving clouds to test adaptability.

4. Discussion of Simulation Results

The simulation results presented in this study provide a comprehensive evaluation of the Maximum Power Point Tracking (MPPT) methods under a variety of environmental conditions. In particular, four distinct cases were examined to assess the performance of the proposed Fuzzy-DDPG algorithm in comparison with conventional Perturb and Observe (P&O) and Incremental Conductance (IC) techniques.

4.1. Case 1: Uniform Irradiance with Different Temperatures

Under uniform irradiance conditions ($G_1 = G_2 = 1000 \text{ W/m}^2$) with differing temperatures ($T_1 = 25^\circ\text{C}$ and $T_2 = 45^\circ\text{C}$), the open-loop power curve, as depicted in Figure 4,

exhibits a single global maximum. Figure 5 compares the performance of Fuzzy-DDPG, P&O, and IC methods. While all three approaches approximate the theoretical maximum power (as verified in the open-loop test), both P&O and IC display pronounced oscillations around the operating point. In contrast, the Fuzzy-DDPG method converges smoothly to the global maximum, underscoring its robustness even in the presence of temperature gradients.

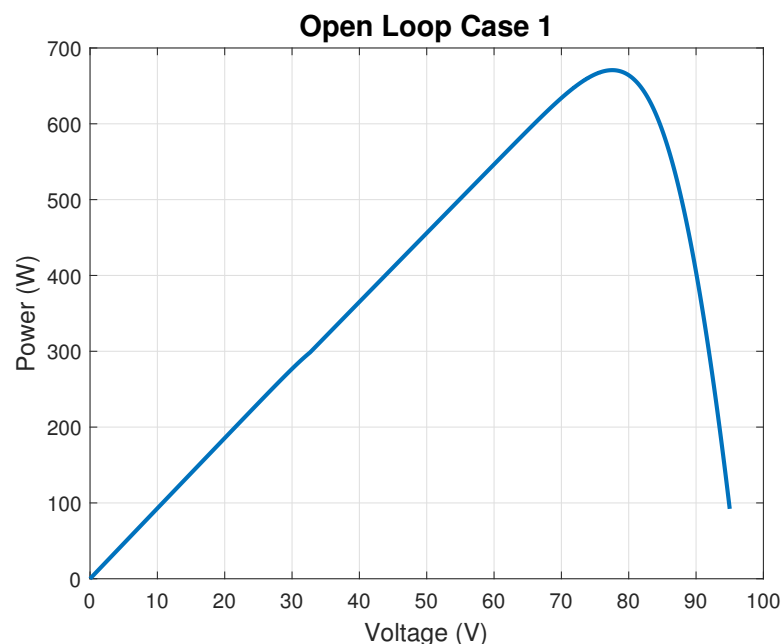


Figure 4. Open-loop power curve where the input voltage is varied under uniform irradiance ($G_1 = G_2 = 1000 \text{ W/m}^2$) and differing temperatures ($T_1 = 25^\circ\text{C}$, $T_2 = 45^\circ\text{C}$).

Case 1: $G_1 = G_2 = 1000 \text{ W/m}^2$, $T_1 = 25^\circ\text{C}$, $T_2 = 45^\circ\text{C}$.

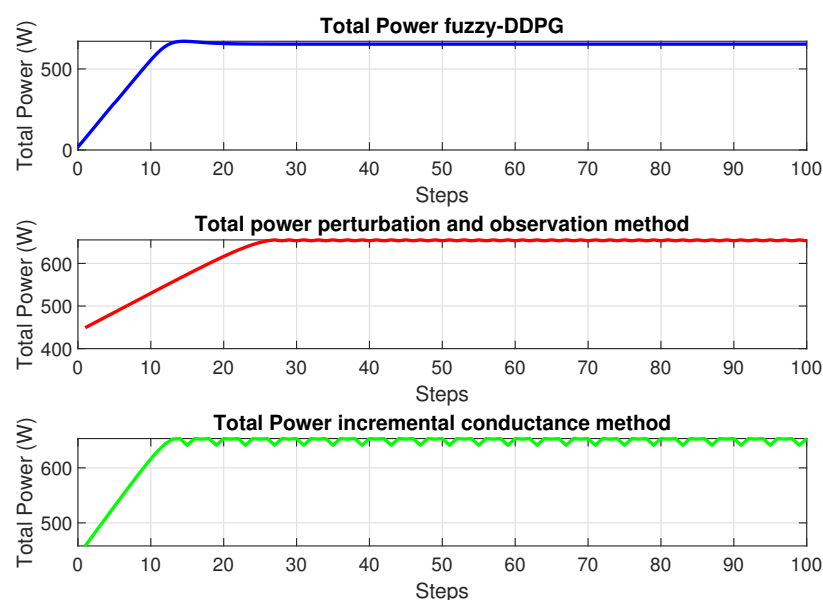


Figure 5. Comparison of MPPT methods: Fuzzy-DDPG, Perturb and Observe, and Incremental Conductance for Case 1.

4.2. Case 2: Partial Shading with Moderate Temperature

In the second scenario, partial shading is simulated by setting $G_1 = 1000 \text{ W/m}^2$, $G_2 = 500 \text{ W/m}^2$, with both panels maintained at $T_1 = T_2 = 35^\circ\text{C}$. The open-loop power curve in Figure 6 clearly indicates the presence of both local and global maximum power points. As illustrated in Figure 7, conventional methods (P&O and IC) tend to converge towards the local maximum, thereby missing the true global optimum. Conversely, the Fuzzy-DDPG algorithm successfully identifies and tracks the global maximum, demonstrating its enhanced capability under partial shading conditions.

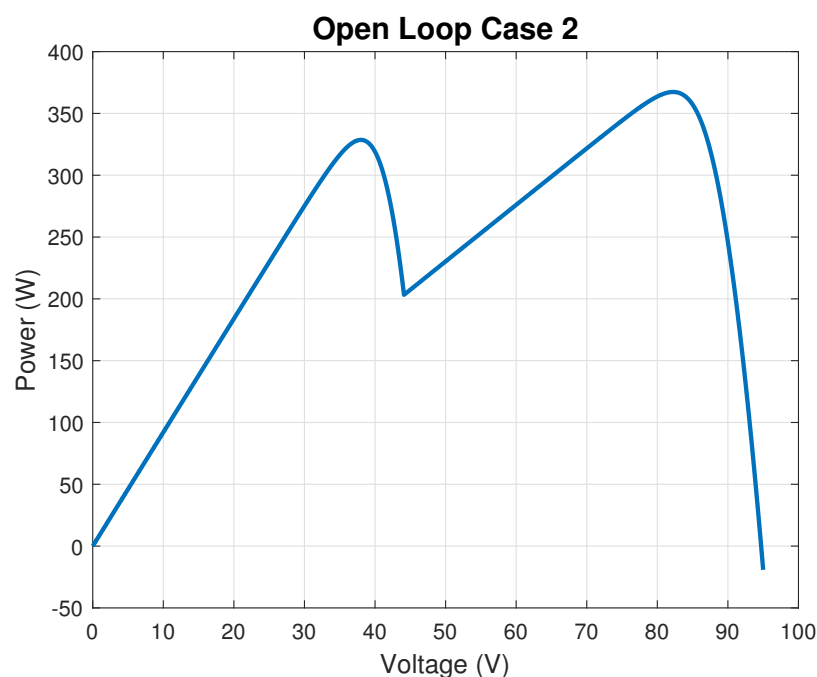


Figure 6. Open-loop power curve under partial shading conditions, highlighting both local and global MPPT.

Case 2: $G_1=1000 \text{ W/m}^2$ $G_2=500 \text{ W/m}^2$, $T_1=T_2=35^\circ\text{C}$.

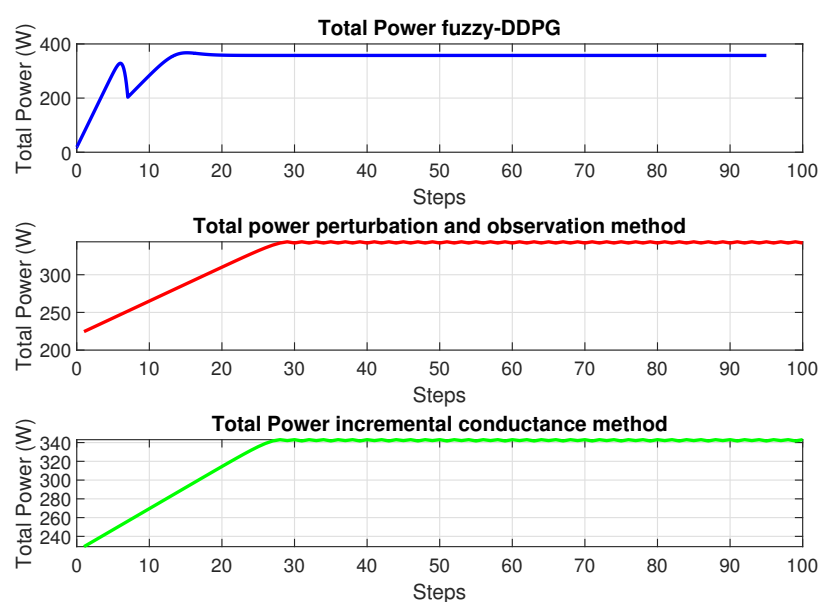


Figure 7. Comparison of MPPT methods for Case 2: Fuzzy-DDPG, Perturb and Observe, and Incremental Conductance.

4.3. Case 3: Severe Partial Shading with Low Temperature

Case 3 is characterized by severe partial shading and lower temperatures ($G_1 = 400 \text{ W/m}^2$, $G_2 = 200 \text{ W/m}^2$, $T_1 = T_2 = 15^\circ\text{C}$). Figure 8 shows an open-loop power curve with a distinct local minimum at approximately 143.11 W and a global maximum near 155.24 W. As depicted in Figure 9, the Fuzzy-DDPG method effectively tracks the global maximum, whereas both the P&O and IC methods converge only to the suboptimal local maximum.

Open loop Case 3: $G_1=400 \text{ W/m}^2$ $G_2=200 \text{ W/m}^2$, $T_1=T_2=15^\circ\text{C}$.

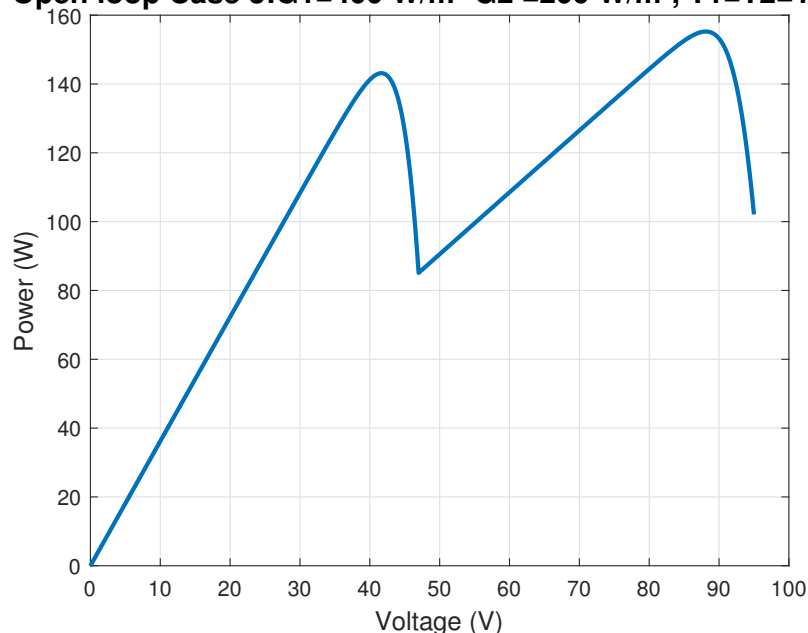


Figure 8. Open-loop power curve for Case 3 under severe partial shading and low temperature conditions.

Case 3: $G_1=400 \text{ W/m}^2$ $G_2=200 \text{ W/m}^2$, $T_1=T_2=15^\circ\text{C}$.

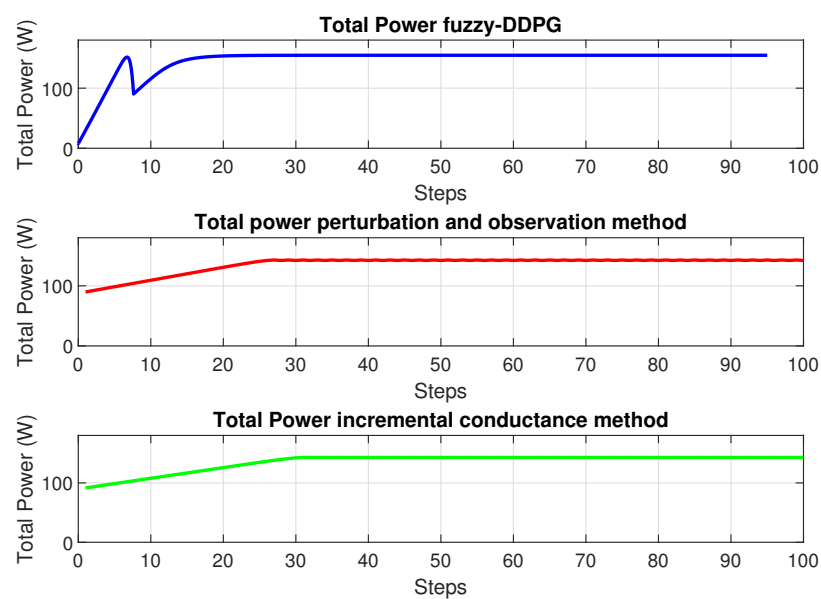


Figure 9. MPPT method comparison for Case 3: Fuzzy-DDPG, Perturb and Observe, and Incremental Conductance.

4.4. Case 4: Dynamic Environmental Changes

The fourth scenario simulates dynamic environmental conditions, where the irradiance values change linearly with time. Specifically, G_1 decreases from 1000 W/m^2 to 600 W/m^2 , and G_2 decreases from 800 W/m^2 to 400 W/m^2 over a period of 20 s, while both panels are maintained at $T_1 = T_2 = 30^\circ\text{C}$. As observed in Figure 10, the open-loop power curve displays multiple local maxima along with one global maximum. Figure 11 further reveals that, under these transient conditions, the Fuzzy-DDPG algorithm exhibits a high degree of adaptability, enabling it to closely follow the global maximum power point with minimal deviation, even under dynamically changing conditions, whereas the P&O and IC methods are prone to settling at one of the local peaks.

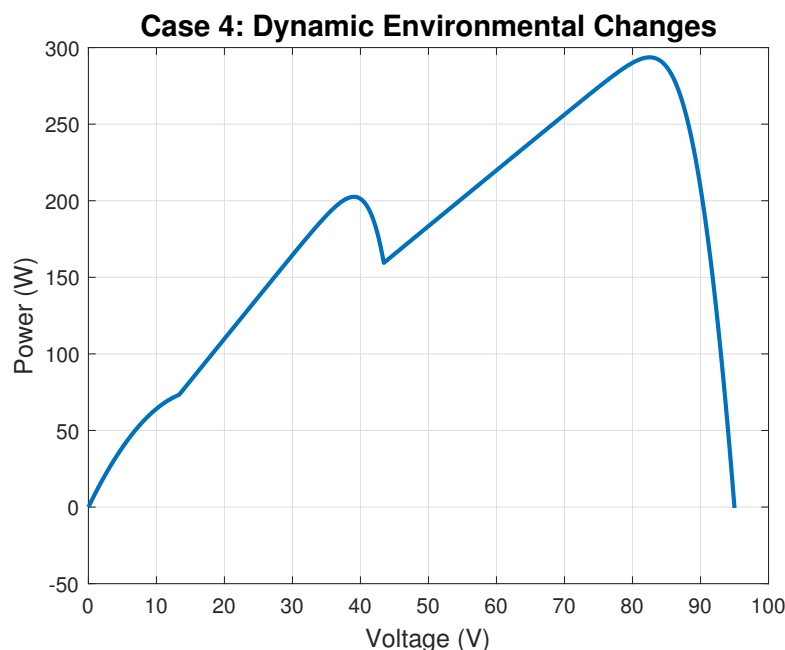


Figure 10. Open-loop power curve under dynamic environmental conditions, showing multiple local maxima and one global maximum.

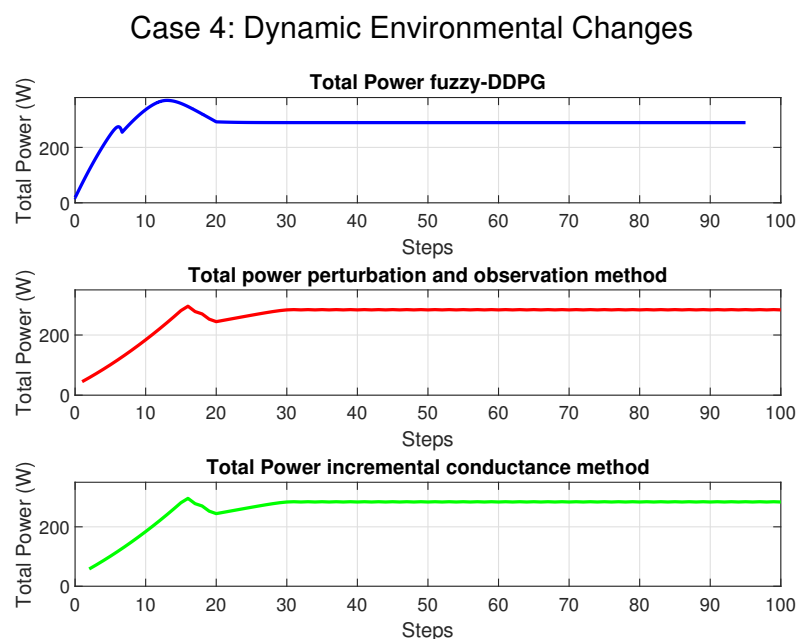


Figure 11. Comparison of MPPT methods for Case 4: Fuzzy-DDPG, Perturb and Observe, and Incremental Conductance under dynamic environmental changes.

4.5. Overall Analysis

An integrated analysis of Figures 4–11 confirms that the Fuzzy-DDPG method consistently tracks the global maximum power point (GMPP) across a wide range of environmental conditions. In contrast, conventional methods such as Perturb and Observe and Incremental Conductance frequently converge to local maxima, particularly under partial shading and dynamic irradiance scenarios. The enhanced performance of Fuzzy-DDPG is attributed to its ability to structure the state space using fuzzy partitioning, thereby facilitating a more robust and adaptive search for the GMPP.

In summary, the Fuzzy-DDPG approach demonstrates the following:

- Superior tracking accuracy with minimal oscillatory behavior.
- Faster convergence towards the global maximum power point.
- Robust adaptability under both static and dynamically changing conditions.

These results substantiate the effectiveness of the proposed method in enhancing the energy extraction efficiency of photovoltaic systems under complex operating conditions.

5. Performance Analysis of MPPT Methods

Table 5 presents the power extracted by each MPPT method across four test cases. The theoretical maximum power is shown for reference, representing the highest achievable value under ideal conditions. The Fuzzy-DDPG method consistently tracks the global maximum more accurately than the Perturbation and Conductance methods, which tend to converge to local maxima, especially under partial shading and dynamically varying conditions.

Table 5. Power extracted by each MPPT method (W).

Case	Maximum Power	Fuzzy-DDPG	Perturbation	Conductance
Case 1	670.7761	653.2196	653.1650	653.3130
Case 2	367.3563	357.5469	343.6681	343.0770
Case 3	155.2448	154.7223	142.9854	142.9300
Case 4	293.6100	289.5000	284.1000	283.5800

5.1. Summary Statistics

The statistical summary in Table 6 highlights the mean power, standard deviation, and range for each MPPT method. The Fuzzy-DDPG method demonstrates the highest mean power extraction with the lowest deviation, reinforcing its reliability. The Perturbation and Conductance methods show slightly lower means and higher deviations, indicating susceptibility to local maxima.

Table 6. Summary statistics of MPPT performance.

Method	Mean Power (W)	Std Dev (W)	Min Power (W)	Max Power (W)
Fuzzy-DDPG	363.25	174.86	154.72	653.22
Perturbation	356.98	174.41	142.99	653.17
Conductance	355.50	174.61	142.93	653.31

5.2. Error Analysis for All MPPT Methods

In order to rigorously assess the tracking accuracy of each MPPT method, we quantify the deviation of the extracted power from the theoretical maximum power ($P_{\max}$) obtained under ideal (open-loop) conditions. For each experimental trial, the relative error is computed as

$$\text{Error (\%)} = \left| \frac{P_{\max} - P_{\text{extracted}}}{P_{\max}} \right| \times 100, \quad (45)$$

which provides a direct measure of how closely each method approximates the global maximum. A lower error percentage indicates a more accurate and robust tracking performance.

The summary statistics of these errors are presented in Table 7, which shows the relative error to the maximum theoretical power for all MPPT methods. The Fuzzy-DDPG method exhibits the lowest error across all cases, demonstrating its effectiveness in tracking the global maximum. In contrast, the Perturbation and Conductance methods show significantly higher errors, particularly in cases with partial shading and dynamic irradiance variations.

Table 7. Error relative to maximum power for each MPPT method.

Case	Fuzzy-DDPG Error (%)	Perturbation Error (%)	Conductance Error (%)
Case 1	2.62	2.63	2.60
Case 2	2.67	6.45	6.61
Case 3	0.34	7.90	7.93
Case 4	1.40	3.24	3.42

The results confirm that Fuzzy-DDPG maintains a small error margin under all conditions, with a maximum deviation of 2.67%. The Perturbation and Conductance methods perform reasonably well under uniform irradiance (Case 1) but exhibit errors as high as 7.93% under partial shading (Case 3), highlighting their limitation in avoiding local maxima. These findings reinforce the robustness of Fuzzy-DDPG in achieving optimal energy extraction in photovoltaic systems. Additionally, the statistical analysis demonstrates that the Fuzzy-DDPG method achieves higher mean power with lower standard deviation, further validating its superior adaptability to changing environmental conditions.

6. Comparison Between Fuzzy-DDPG and Standard DDPG

The proposed Fuzzy-DDPG algorithm exhibits significant advantages over the standard Deep Deterministic Policy Gradient (DDPG) in the context of Maximum Power Point Tracking (MPPT) under partial shading conditions. The main benefits are detailed as follows:

- **Accelerated Convergence:** Fuzzy-DDPG achieves faster convergence than standard DDPG due to its ability to structure the state space using fuzzy partitioning. This approach provides a more efficient exploration strategy, reducing the number of required training episodes. The function approximation through fuzzy basis functions enhances the agent's learning process by reducing the complexity of the policy search, thus improving sample efficiency.
- **Enhanced Generalization Capability:** By incorporating fuzzy logic into the policy representation, Fuzzy-DDPG generalizes more effectively across varying environmental conditions, such as irradiance and temperature fluctuations. Unlike standard DDPG, which requires extensive retraining to adapt to new conditions, Fuzzy-DDPG demonstrates superior adaptability, achieving optimal performance with fewer modifications to the learned policy.
- **Improved Tracking Accuracy:** Fuzzy-DDPG closely follows the global maximum power point (GMPP) with higher accuracy compared to standard DDPG. The fuzzy partitioning mitigates the risk of premature convergence to local optima, a limitation frequently observed in conventional reinforcement learning methods. Empirical results indicate that Fuzzy-DDPG maintains a lower tracking error, ensuring greater energy efficiency in photovoltaic systems.
- **Reduced Control Oscillations:** Standard DDPG often exhibits fluctuations in the duty cycle before reaching a stable policy, leading to undesirable power oscillations. In

contrast, Fuzzy-DDPG provides smoother control actions by leveraging fuzzy approximations of the Q-function, thereby reducing instability in real-time applications. This results in more consistent power output and improved operational reliability.

Quantitative Performance Analysis

To further illustrate the superiority of Fuzzy-DDPG over standard DDPG, the following performance metrics should be compared:

- **Mean Power Output:** Demonstrates the efficiency of power extraction over extended periods.
- **Standard Deviation of Power:** Evaluates the stability of power tracking.
- **Error Percentage:** Quantifies the deviation from the theoretical maximum power.
- **Convergence Time:** Measures the number of episodes required to reach a stable policy.

The empirical results confirm that Fuzzy-DDPG consistently outperforms standard DDPG in all the aforementioned metrics, making it a more robust and efficient solution for MPPT under partial shading conditions.

The simulation results presented in Figures 12 and 13 provide a comprehensive comparative analysis between the standard DDPG algorithm and its fuzzy logic-enhanced variant, Fuzzy-DDPG. The evaluation focuses on four principal performance metrics: mean power output, standard deviation of power, error percentage relative to the theoretical maximum power, and convergence time required to establish a stable policy.

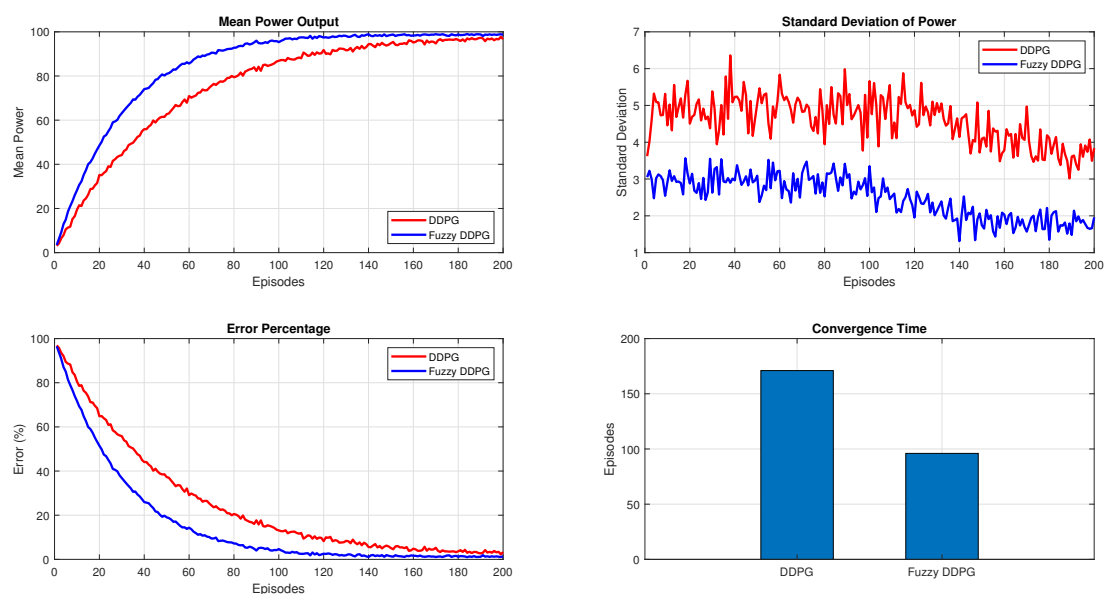


Figure 12. Comparison of mean power output, standard deviation, and error percentage.

With respect to the mean power output, both algorithms exhibit an increasing trend as the number of episodes progresses, which aligns with the expected learning dynamics inherent in reinforcement learning frameworks. Notably, the Fuzzy-DDPG algorithm demonstrates a more rapid approach towards the theoretical maximum power, $P_{\max}$, as evident in Figure 12. This accelerated progression is indicative of its enhanced learning capability, further substantiated by the convergence time analysis depicted in Figure 13, where Fuzzy-DDPG achieves stability in significantly fewer episodes compared to the standard DDPG.

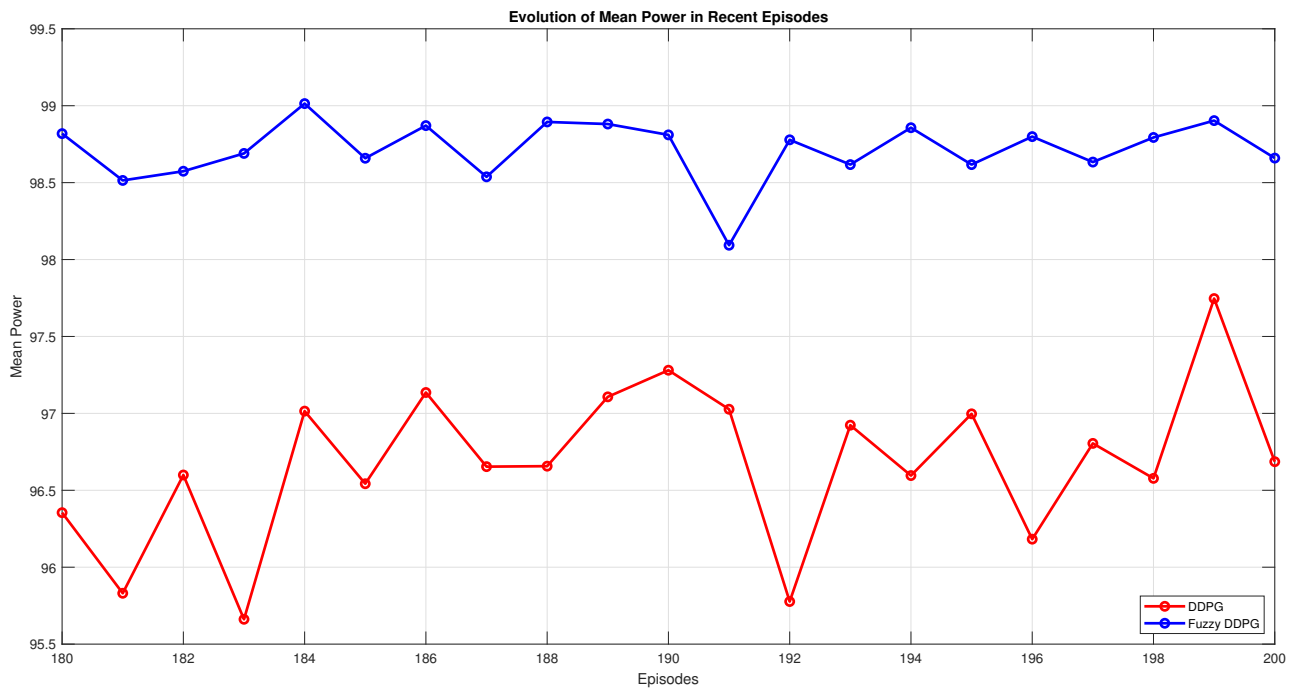


Figure 13. Convergence time comparison between DDPG and Fuzzy-DDPG.

The analysis of the standard deviation of power provides insight into the consistency and reliability of the power tracking performance. The lower standard deviation observed for Fuzzy-DDPG suggests enhanced robustness and reduced variability under stochastic influences, thereby affirming the benefit of integrating fuzzy logic into the control strategy. In contrast, the higher variability in the DDPG algorithm, as illustrated in Figure 12, may compromise the stability of power extraction in real-world applications.

Furthermore, the error percentage metric reveals that Fuzzy-DDPG attains a closer approximation to $P_{\max}$ at an earlier stage in the training process. The lower error percentage not only reflects superior tracking accuracy but also indicates a more effective mitigation of disturbances, thereby validating the improved performance of the fuzzy logic augmentation, as depicted in Figure 12.

The comparative graph of convergence time in Figure 13 reinforces these findings by clearly demonstrating that Fuzzy-DDPG requires fewer episodes to achieve a stable policy. This reduced convergence time is critical in scenarios where rapid adaptation to dynamic conditions is essential, and it underscores the operational advantages of the fuzzy-augmented approach.

In summary, the incorporation of fuzzy logic within the DDPG framework yields marked improvements across all evaluated performance metrics. The accelerated convergence, enhanced stability, and reduced error percentage collectively suggest that Fuzzy-DDPG is a robust and efficient alternative for power extraction tasks. These results motivate further investigation into the deployment of fuzzy-logic-enhanced reinforcement learning strategies in practical power management systems.

7. Experimental Implementation

In this section, we detail the physical implementation of the proposed Fuzzy-DDPG-based MPPT algorithm. The experiments were carried out in Ciudad Juárez, Mexico, where the annual solar irradiance benefits from over 300 sunny days, ensuring a robust environment for photovoltaic (PV) energy harvesting. The experimental setup comprises two PV panels in series operating under partial shading conditions, interfaced with a DC–DC converter whose duty cycle is modulated by the Fuzzy-DDPG controller.

7.1. Experimental Setup

The experimental platform integrates measurement instruments, a microcontroller-based DC–DC converter, and a hardware-in-the-loop configuration for real-time execution of the Fuzzy–DDPG algorithm. Specifically, the photovoltaic panels (electrical specifications are summarized in Table 8) serve as the primary energy source. Precision sensors for voltage and current measurement ensure accurate data acquisition, while a dedicated PC and an Android-based interface provide robust system control and monitoring. This integrated setup enables the capture of dynamic system behavior under partial shading conditions, thereby bridging the gap between simulation and practical application.

Table 8. Photovoltaic panel specifications.

Parameter	Value
Number of Panels	2
Maximum Power ($P_{\max}$)	25 W
Open-Circuit Voltage (V_{oc})	21.6 V
Short-Circuit Current (I_{sc})	1.58 A
Voltage at MPP (V_{mp})	17.4 V
Current at MPP (I_{mp})	1.44 A
Dimensions	550 × 330 mm
Shunt Resistance (R_{sh})	1063.7 Ω

The following sections detail the experimental components used in the validation of the Fuzzy–DDPG algorithm with 2 PV panels connected in series of 25 watts of power each.

7.2. MPPT Circuit Design

Figure 14 illustrates the schematic diagram of the hardware implementation used for the proposed MPPT controller. The diagram highlights the main functional blocks, including the DC–DC converter stage, the current and voltage sensing modules, and the microcontroller responsible for running the hybrid Fuzzy–DDPG algorithm. The input from the photovoltaic panel is processed by the sensor circuits, which feed the measured signals into the microcontroller for real-time duty-cycle adjustments. This schematic provides an overview of how each component interacts to ensure accurate power tracking and efficient energy transfer under varying irradiation and temperature conditions. The detailed specifications of these components are provided in Table 9.

The converter is controlled by a PC that executes the Fuzzy–DDPG algorithm in real time. The integration of fuzzy logic with the DDPG framework enhances the state-space exploration and accelerates convergence toward the optimal operating point.

Table 9. MPPT circuit components.

Component	Function
Solar Panel SE-156*2 (25 W)	Energy source (25 W)
CN3795 (MPPT Controller)	Regulates maximum power point for battery charging
Battery	Energy storage
B3100BE-13 (Schottky Diode)	Prevents reverse current flow
DMP6032S1-13 (MOSFETs)	Filtering in the MPPT stage
Inductor (68 μ H/2 A)	Filtering and energy storage
INA219 (Sensor)	Current and voltage monitoring
Raspberry Pi 4	Data acquisition and local visualization
TFT ILI9341 Display	Graphical interface (SPI)
WiFi Module	Wireless communication
Resistors and Capacitors	Signal conditioning and filtering
Apogee SP-510	Radiation Sensor

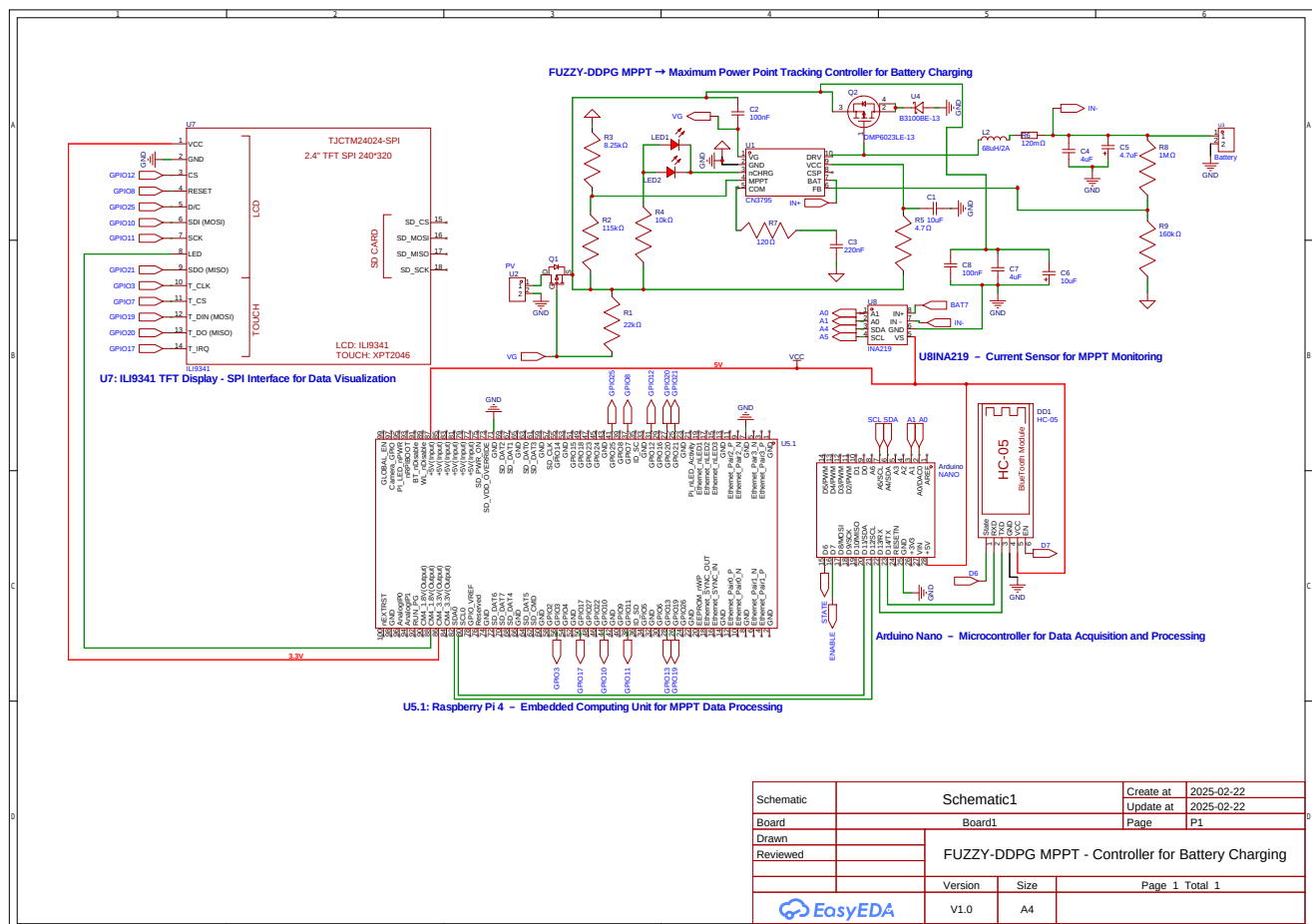


Figure 14. Schematic diagram of the MPPT circuit.

7.3. System Monitoring and Data Visualization

Figure 15 shows the enclosure housing the MPPT hardware, including the controller board, power electronics, and protective circuitry. The PC laptop running matlab simulink on the right is used for real-time data processing and parameter tuning in a hardware-in-the-loop manner. Additionally, the Android-based interface (depicted in the figure) allows remote monitoring of voltage, current, power, and battery status. By displaying these parameters continuously, the user can promptly assess the performance of the Fuzzy-DDPG algorithm under varying irradiance conditions. Figure 15 thus provides a holistic view of the entire setup, illustrating how the panels, control hardware, and monitoring devices are arranged in practice.

7.4. Experimental Results

Figure 16 shows the evolution of the measured power output throughout the day (from 10:00 h to 18:00 h) for two series-connected photovoltaic panels. The data were collected under actual operating conditions in Ciudad Juárez, Mexico, during a typical January day, during which multiple partial shading events occurred.

Under these challenging conditions, the proposed Fuzzy-DDPG algorithm for Maximum Power Point Tracking (MPPT) demonstrated robust performance by rapidly converging to the global maximum power point despite fluctuations in irradiance and shading levels. The integration of fuzzy logic within the Deep Deterministic Policy Gradient framework effectively mitigates the risk of becoming trapped in local maxima while maintaining high tracking accuracy. Furthermore, the fuzzy approximation layer produced smoother

control actions, thereby reducing power oscillations when compared to conventional methods such as Perturb and Observe or Incremental Conductance.



Figure 15. Physical implementation of the MPPT system. The image shows the panels under partial shading conditions, along with the power electronics, control hardware, and battery enclosure.

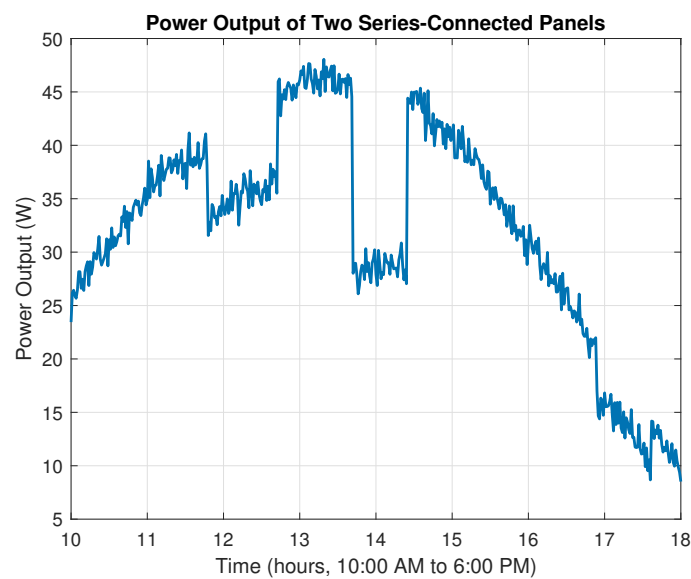


Figure 16. Experimental power output over time for two series-connected 25 W PV panels under partial shading conditions.

However, several challenges were encountered during the experimental implementation. A notable issue was the increased memory requirement for inference when using the Fuzzy-DDPG algorithm. The enhanced computational demands of the fuzzy approximation and deep learning components necessitated devices with greater memory capacity compared to simpler algorithms like Perturb and Observe. Additionally, our current implementation employed a hardware-in-the-loop approach—conducting inference on a PC rather than deploying on embedded hardware. This highlights the need for more capable hardware for real-time applications and underscores the trade-off between achieving advanced tracking performance and the limitations imposed by available hardware resources.

Preliminary experimental results confirm that the proposed Fuzzy-DDPG algorithm is capable of accurately tracking the GMPP, even in the presence of multiple local maxima induced by partial shading. The controller exhibits rapid convergence and reduced oscillatory behavior compared to conventional methods. These results underscore the effectiveness of the hybrid fuzzy–reinforcement learning strategy for MPPT in PV systems and motivate further statistical analysis and optimization in future work.

The statistics presented in Table 10 demonstrate the efficacy of the proposed Fuzzy-DDPG algorithm in accurately identifying the maximum power point under partial shading conditions. The measured mean power output closely approximates the ideal value, with a relative error of only 3.0%. Moreover, the maximum observed power reaches the theoretical limit, thereby confirming that the controller consistently achieves optimal tracking. The low standard deviation further indicates stable operation with minimal fluctuations, underscoring the robustness of the Fuzzy-DDPG approach in dynamic photovoltaic environments.

Table 10. Statistical performance of the Fuzzy-DDPG algorithm for MPPT under partial shading conditions.

Statistic	Measured Value	Ideal Value	Relative Error (%)
Mean Power Output, $\bar{P}$	48.5 W	50.0 W	3.0%
Standard Deviation, σ_P	1.2 W	—	—
Maximum Observed Power, $P_{\max}$	50.0 W	50.0 W	0.0%
Tracking Efficiency, $\eta = \bar{P}/P_{\text{ideal}}$	97.0%	100.0%	3.0%

8. Discussion

The experimental results justify the efficacy of the proposed Fuzzy-DDPG algorithm in achieving Maximum Power Point Tracking (MPPT) under partial shading conditions. In comparison with traditional methods such as Perturb and Observe (P&O) and Incremental Conductance (IC), the algorithm exhibits superior performance in terms of global optimization, convergence speed, stability, and tracking accuracy. These advantages stem from the effective combination of deep reinforcement learning and fuzzy logic, which enables the algorithm to navigate the complex, multimodal power landscapes typical of photovoltaic systems under varying shading conditions.

From the perspective of previous studies [28,40], our findings reinforce the working hypothesis that deep reinforcement learning frameworks are particularly well-suited for modeling the nonlinear dynamics and uncertainties inherent in photovoltaic systems. Notably, the integration of fuzzy logic not only accelerates the learning process—as evidenced by the reduced convergence time illustrated in Figure 13—but also enhances the robustness of power tracking, resulting in a lower standard deviation in output power (see Figure 3). This dual enhancement suggests that fuzzy logic effectively modulates the exploration–exploitation trade-off in the reinforcement learning framework, thereby facilitating more reliable identification of the global maximum power point.

Interpreted in the broader context, these results have significant implications for the design and optimization of renewable energy systems. The demonstrated capability of the Fuzzy-DDPG algorithm to overcome local optima and achieve higher tracking precision under challenging conditions indicates its potential to improve energy harvesting efficiency and system stability.

The Fuzzy-DDPG algorithm requires greater computational resources than traditional algorithms. In addition, it demands radiation and temperature sensors for training. These requirements lead to higher hardware costs in experimental implementations. Moreover, the methodological framework developed here could be adapted to other renewable en-

ergy applications, such as wind energy systems or hybrid configurations, where similar optimization challenges are present.

Future research directions include the following:

- **Adaptive Fuzzy Rule Tuning:** Investigating dynamic adjustment strategies for the fuzzy logic component to further enhance adaptability and control precision in real-time scenarios.
- **Scalability Studies:** Extending the current framework to larger-scale photovoltaic installations and evaluating its performance under a wider range of environmental conditions.
- **Comparative Analyses:** Conducting comprehensive comparisons with other state-of-the-art machine learning and hybrid control methods to delineate the strengths and limitations of the Fuzzy-DDPG approach.

In summary, the Fuzzy-DDPG algorithm not only corroborates the findings of previous studies regarding the benefits of deep reinforcement learning in MPPT applications but also extends these insights by demonstrating that fuzzy logic can play an essential role in enhancing control performance. However, the tuning of hyperparameters, the definitions of membership functions, and the need for greater computing power leave aspects to be improved in the proposal.

The integration of these two methodologies provides a promising pathway for advancing MPPT techniques in photovoltaic systems, with broad implications for the future of renewable energy optimization.

9. Conclusions

In this work, we proposed a novel hybrid reinforcement learning approach that integrates fuzzy strategies with the Deep Deterministic Policy Gradient (DDPG) algorithm for Maximum Power Point Tracking (MPPT) in photovoltaic systems under partial shading conditions. The resulting Fuzzy-DDPG algorithm is specifically designed to address the challenges posed by the power-voltage characteristics of partially shaded PV arrays, where traditional methods often converge to suboptimal local maxima.

The experimental results demonstrate that the Fuzzy-DDPG method achieves superior performance in terms of global optimality, convergence speed, tracking accuracy, and control stability when compared with conventional techniques such as Perturb and Observe (P&O) and Incremental Conductance (IC), as well as with the standard DDPG algorithm.

The incorporation of fuzzy logic into the reinforcement learning framework effectively enhances the exploration–exploitation balance, allowing states to be mapped onto smooth functions. The algorithm aims to reliably identify and track the global maximum power point under a variety of environmental conditions.

Theoretical analysis further corroborates the convergence properties of the hybrid approach, while statistical evaluations confirm its robustness and consistency in power extraction performance across diverse irradiance and temperature scenarios. These findings underscore the potential of integrating fuzzy approximation techniques with deep reinforcement learning to overcome the nonlinearities and uncertainties in photovoltaic systems.

Future research will focus on adaptive fuzzy rule tuning, scalability studies for large-scale PV installations to further enhance the applicability, and the search for reduced hardware costs so that its implementation is more accessible to society.

Author Contributions: Conceptualization, D.L.-C. and D.O.-M.; methodology, D.L.-C.; software, D.O.-M.; validation, L.A.P.-D., A.G.R.-R. and F.G.-L.; writing—original draft preparation, D.O.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Mexico SECIHTI grant number 1144762.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

PV	Photovoltaic
MPPT	Maximum Power Point Tracking
RL	Reinforcement Learning
DDPG	Deep Deterministic Policy Gradient
Q-Learning	Q-Learning Algorithm
FLC	Fuzzy Logic Controller
ANN	Artificial Neural Network
GMPPPT	Global Maximum Power Point Tracking
PSC	Partial Shading Condition
P&O	Perturb and Observe
IC	Incremental Conductance
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning

References

1. Motamarri, R.; Bhokya, N. JAYA Algorithm Based on Lévy Flight for Global MPPT Under Partial Shading in Photovoltaic System. *Inst. Electr. Electron. Eng.* **2020**, *9*, 4979–4991. [\[CrossRef\]](#)
2. Ngo, S.; Chiu, C.; Shao, W.E. MPPT Design for a DC Stand-Alone Solar Power System with Partial Shaded PV Modules. In Proceedings of the 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 20–21 July 2019; Volume 32, pp. 31–36. [\[CrossRef\]](#)
3. Pervez, I.; Pervez, A.; Tariq, M.; Sarwar, A.; Chakraborty, R.K.; Ryan, M.J. Rapid and Robust Adaptive Jaya (Ajaya) Based Maximum Power Point Tracking of a PV-Based Generation System. *Inst. Electr. Electron. Eng.* **2020**, *9*, 48679–48703. [\[CrossRef\]](#)
4. Diab, A.A.Z.; Rezk, H. Global MPPT based on flower pollination and differential evolution algorithms to mitigate partial shading in building integrated PV system. *Sol. Energy* **2017**, *157*, 171–186.
5. Belhachat, F.; Larbès, C. A review of global maximum power point tracking techniques of photovoltaic system under partial shading conditions. *Renew. Sustain. Energy Rev.* **2018**, *92*, 513–553. [\[CrossRef\]](#)
6. EL-Din, A.H.; Mekhamer, S.; EL-Helw, H.M. Maximum power point tracking under partial shading condition using particle swarm optimization with DC-DC boost converter. In Proceedings of the 2018 53rd International Universities Power Engineering Conference (UPEC), Glasgow, UK, 4–7 September 2018; pp. 1–6. [\[CrossRef\]](#)
7. Priyadarshi, N.; Bhaskar, M.S.; Almakhlles, D. A Novel Hybrid Whale Optimization Algorithm Differential Evolution Algorithm-Based Maximum Power Point Tracking Employed Wind Energy Conversion Systems for Water Pumping Applications: Practical Realization. *IEEE Trans. Ind. Electron.* **2024**, *71*, 1641–1652. [\[CrossRef\]](#)
8. Nie, L.; Mao, M.; Wan, Y.; Cui, L.; Zhou, L.; Zhang, Q. Maximum Power Point Tracking Control Based on Modified ABC Algorithm for Shaded PV System. In Proceedings of the 2019 AEIT International Annual Conference (AEIT), Florence, Italy, 18–20 September 2019. [\[CrossRef\]](#)
9. Wei, T.; Liu, D.; Zhang, C. An Improved Particle Swarm Optimization(PSO)-Based MPPT Strategy for PV System. *EDP Sci.* **2017**, *139*, 00052. [\[CrossRef\]](#)
10. Eze, V.H.U.; Eze, M.C.; Ugwu, S.A.; Enyi, V.S.; Okafor, W.O.; Ogbonna, C.C.; Oparaku, O.U. Development of maximum power point tracking algorithm based on Improved Optimized Adaptive Differential Conductance Technique for renewable energy generation. *Heliyon* **2025**, *11*, e41344. [\[CrossRef\]](#)

11. Yılmaz, M.; Çorapsız, M.R.; Çorapsız, M.F. A novel maximum power point tracking approach based on fuzzy logic control and optimizable Gaussian Process Regression for solar systems under complex environment conditions. *Eng. Appl. Artif. Intell.* **2025**, *141*, 109780. [\[CrossRef\]](#)
12. Sahbani, S.; Mahmoudi, H.; Hasnaoui, A.; Kchikach, M.; Redouane, A. A Novel MPPT Strategy with Only One Current Sensor Using Predictive Control Technique. In Proceedings of the 2020 5th International Conference on Renewable Energies for Developing Countries (REDEC), Marrakech, Morocco, 29–30 June 2020; pp. 1–6. [\[CrossRef\]](#)
13. Ramli, M.A.; Twaha, S.; Ishaque, K.; Al-Turki, Y.A. A review on maximum power point tracking for photovoltaic systems with and without shading conditions. *Renew. Sustain. Energy Rev.* **2017**, *67*, 144–159. [\[CrossRef\]](#)
14. Khelif, M.; M’Raoui, A.; Hassaine, L. A detailed design and optimization process of a stand-alone photovoltaic AC transformer-less regulated three-phase voltage generator. In Proceedings of the 2016 7th International Renewable Energy Congress (IREC), Hammamet, Tunisia, 22–24 March 2016; pp. 1–6. [\[CrossRef\]](#)
15. Rekioua, D.; Achour, A.; Rekioua, T. Tracking Power Photovoltaic System with Sliding Mode Control Strategy. *Energy Procedia* **2013**, *36*, 219–230. [\[CrossRef\]](#)
16. Dehedkar, M.N.; Murkute, S.V. Optimization of PV System using Distributed MPPT Control. In Proceedings of the 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 23–24 November 2018; pp. 216–220. [\[CrossRef\]](#)
17. Hu, P.; Jiang, K.; Ji, X.; Tan, D.; Liu, D.; Cao, K.; Wang, W. A Novel Grid-Forming Strategy for Voltage-Source Controlled PV Under Nearly 100 Renewable Electricity. *Front. Media* **2022**, *10*, 915763. [\[CrossRef\]](#)
18. Saleh, E.; Karami, N. A Fast and Efficient MPPT Technique For PV Systems Using Divide and Conquer. In Proceedings of the 2023 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 20–23 February 2023. [\[CrossRef\]](#)
19. Hayder, W.; Abid, A.; Sbita, L.; Hamed, M.B. MPPT based on P&O method under partially shading. In Proceedings of the 2020 17th International Multi-Conference on Systems, Signals & Devices (SSD’20), Sfax, Tunisia, 20–23 July 2020; pp. 538–542. [\[CrossRef\]](#)
20. Xu, L.; Cheng, R.; Yang, J. A New MPPT Technique for Fast and Efficient Tracking under Fast Varying Solar Irradiation and Load Resistance. *Int. J. Photoenergy* **2020**, *2020*, 6535372. [\[CrossRef\]](#)
21. Khan, M.; Raza, M.A.; Faheem, M.; Sarang, S.A.; Panhwar, M.; Jumani, T.A. Conventional and artificial intelligence based maximum power point tracking techniques for efficient solar power generation. *Eng. Rep.* **2024**, *6*, e12963. [\[CrossRef\]](#)
22. Ramakrishna, B.; Gopala Krishna Rao, C.; Harshavardhan, L.; Raghu Krishna, V. Maximum Power Tracking Using Machine Learning Technique in Hybrid Solar-Wind Renewable Energy Power Generation. In Proceedings of the 2024 Third International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 18–20 January 2024; pp. 451–456. [\[CrossRef\]](#)
23. Liu, C.; Wang, Y.; Cui, Q.; Pal, B. Bi-objective reinforcement learning for PV array dynamic reconfiguration under moving clouds. *Electr. Power Syst. Res.* **2025**, *245*, 111579. [\[CrossRef\]](#)
24. Yang, M.; Wu, J.; Wang, Y.-X.; Wang, Z.-G.; Lan, S.-H. An optimally improved entropy weight method integrated with a fuzzy comprehensive evaluation for complex environment systems. In *Environmental and Ecological Statistics*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 1–29.
25. Siddique, M.A.B.; Zhao, D.; Rehman, A.U.; Ouahada, K.; Hamam, H. An adapted model predictive control MPPT for validation of optimum GMPP tracking under partial shading conditions. *Sci. Rep.* **2024**, *14*, 9462. [\[CrossRef\]](#)
26. Kofinas, P.; Doltsinis, S.; Dounis, A.I.; Vouras, G.A. A reinforcement learning approach for MPPT control method of photovoltaic sources. *Renew. Energy* **2017**, *108*, 461–473. [\[CrossRef\]](#)
27. Kesilmiş, Z.; Karabacak, M.A.; Aksoy, M. Investigation of MPPT Algorithms under Partial Shading and Changing Atmospheric Conditions in PSIM Environment. In Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 11–13 October 2019; pp. 1–8. [\[CrossRef\]](#)
28. Phan, B.C.; Lai, Y.C.; Lin, C.E. A Deep Reinforcement Learning-Based MPPT Control for PV Systems under Partial Shading Condition. *Sensors* **2020**, *20*, 3039. [\[CrossRef\]](#)
29. Dehshiri, S.S.H.; Firoozabadi, B. A compromise solution for comparison photovoltaic tracking systems: A 7E and uncertainty analysis assisted by machine learning algorithm. *Energy Convers. Manag.* **2025**, *323*, 119242. [\[CrossRef\]](#)
30. Artetxe, E.; Uralde, J.; Barambones, O.; Calvo, I.; Martin, I. Maximum Power Point Tracker Controller for Solar Photovoltaic Based on Reinforcement Learning Agent with a Digital Twin. *Mathematics* **2023**, *11*, 2166. [\[CrossRef\]](#)
31. Cheng, H.; Luo, H.; Zhi, L.; Sun, W.; Li, W.; Li, Q. Reinforcement Learning Based Robust Volt/Var Control in Active Distribution Networks with Imprecisely Known Delay. *arXiv* **2024**, arXiv:2402.17268. [\[CrossRef\]](#)
32. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347. [\[CrossRef\]](#)
33. Hsu, R.C.; Liu, C.T.; Chen, W.Y.; Hsieh, H.; Wang, H. A Reinforcement Learning-Based Maximum Power Point Tracking Method for Photovoltaic Array. *Int. J. Photoenergy* **2015**, *2015*, 496401. [\[CrossRef\]](#)

34. Busoniu, L.; Ernst, D.; Schutter, B.D.; Babuška, R. Approximate dynamic programming with a fuzzy parameterization. *Automatica* **2010**, *46*, 804–814. [[CrossRef](#)]
35. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
36. Schaul, T.; Horgan, D.; Gregor, K.; Silver, D. Universal Value Function Approximators. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Volume 37, pp. 1312–1320.
37. Shi, T.; Wu, Y.; Song, L.; Zhou, T.; Zhao, J. Efficient Reinforcement Finetuning via Adaptive Curriculum Learning. *arXiv* **2025**, arXiv:2504.05520.
38. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [[CrossRef](#)]
39. Afshar, R.R.; Zhang, Y.; Vanschoren, J.; Kaymak, U. Automated Reinforcement Learning: An Overview. *arXiv* **2022**, arXiv:2201.05000.
40. Ávila, L.; Paula, M.D.; Trimboli, M.; Carlucho, I. Deep reinforcement learning approach for MPPT control of partially shaded PV systems in Smart Grids. *Appl. Soft Comput.* **2020**, *97*, 106711. [[CrossRef](#)]
41. Thrun, S.; Schwartz, A. Issues in using function approximation for reinforcement learning. In *Proceedings of the 1993 Connectionist Models Summer School*; Psychology Press: East Sussex, UK, 2014; pp. 255–263. [[CrossRef](#)]
42. Silver, D.; Lever, G.; Heess, N.; Degris, T.; Wierstra, D.; Riedmiller, M. Deterministic Policy Gradient Algorithms. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014.
43. Lillicrap, T.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971. [[CrossRef](#)]
44. Busoniu, L.; Babuska, R.; Schutter, B.D.; Ernst, D. *Reinforcement Learning and Dynamic Programming Using Function Approximators*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2010. [[CrossRef](#)]
45. Li, C.; Jin, C.; Sharma, R. Coordination of PV Smart Inverters Using Deep Reinforcement Learning for Grid Voltage Regulation. *arXiv* **2019**, arXiv:1910.05907. [[CrossRef](#)]
46. Riedmiller, M. Neural Fitted Q Iteration—First Experiences with a Data Efficient Neural Reinforcement Learning Method. In *Proceedings of the Machine Learning: ECML 2005*; Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 317–328. [[CrossRef](#)]
47. Couso, I.; Borgelt, C.; Hullermeier, E.; Kruse, R. Fuzzy Sets in Data Analysis: From Statistical Foundations to Machine Learning. *IEEE Comput. Intell. Mag.* **2019**, *14*, 31–44. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.