Advances and Applications in Mathematical Sciences
Volume 21, Issue 5, March 2022, Pages 2781-2793

DEEP Q REINFORCEMENT LEARNING TO IMPROVE THE MPPT IN SOLAR CELL

ASHUTOSH YADAV* and ANINDITA ROY CHOWDHURY

School of Engineering and Sciences
GD Goenka University, Gurugram, India
E-mail: ashunirban@gmail.com
anindita.roychowdhury@gdgoenka.ac.in

Abstract

The architecture of an MPPT controller assures a steady flow of energy despite having different external conditions. Designing a model that can generate the maximum power irrespective of the environmental or parametric conditions is most challenging. Reinforcement learning with fractional-order is used in this method to get over these drawbacks. Allowing the model to be parametric-free, it can be readily implanted in new situations, thanks to Deep Q-learning Tracking time, peak oscillation, and overall harmonic distortion are all reduced by using fractional-order. The model has undergone rigorous testing in controlled environments with good outcomes. It is also compared to existing comparison algorithms to track time, THD and maximum power. Real-world data from the solcast is used to verify the idea, with New Delhi serving as the test site.

I. Introduction

Energy is a necessity for human survival and development, yet worldwide supply is in short supply. As a result, researchers are becoming more concerned about the environment and looking into renewable energy alternatives. Reducing environmental impact and reliance on fossil fuels can be accomplished by using a mix of renewable and non-renewable energy sources [1]. The sun, which is abundant in dry environments, is the primary source of energy [2]. It's easy to get the most out of a PV generator. A DC/DC converter is used by MPPT algorithms to achieve this goal.

They are known as Highest Power Point Tracking (MPPT) [3] because

2020 Mathematics Subject Classification: 68T07.
Keywords: MPPT, PV power, Reinforcement Learning, DQN.
*Corresponding author; E-mail: ashtoshy12@gmail.com
Received October 3, 2021; Accepted November 30, 2021

they always send the maximum amount of power, no matter what the circumstances are. In response to variations in PV power, MPPT regulates output voltage shifts. In addition, PV systems suffer large losses because to output current-voltage nonlinearity. The most common MPPT control methods have been replaced by a variety of newer ones throughout time. In PV systems, a modified fuzzy-logic (FL) MPPT control scheme and an incremental conduction maximum power point tracking (MPPT) algorithm with fuzzy controller [4] and a hybrid MPPT control that incorporates a modified P&O and an upgraded PSO [6] are among the developed methodologies. New MPPT technology is now available for any application requiring rapidly changing Artificial Neural Networks (ANN) [7]. The majority of these technologies employ models to adapt various PV systems. Several issues that emerge when PV panels are connected in various ways may be addressed if an accurate model of the PV systems and its features is obtained. The author is motivated by a model-free approach.

AI, machine learning, and robotics all rely on RL, or reinforcement learning. Using trial-and-error and self-learning, this technique can handle a wide range of real-time problems. Among other things, they're typically found in the fields of robotics, gaming, and industrial control. It integrates adaptive control theory with optimal control and nature-inspired algorithms [8]. When an agent is constantly interacting with its surroundings, real-time learning takes place. It is up to us to either reward (via positive reinforcement) or penalise such behaviour (negative feedback for the wrong action). Hence, the ideal way to ensure growth and survival is learned by iterative RL [9]. It's worth noting that RL doesn't require a model of the system or its dynamics to reproduce it. Model settings have no effect on this.

In recent years, the usage of RL to address MPPT issues has skyrocketed. A RL-based model was developed by [10] to improve the unpredictability of wind speed energy conversion. The MPPT for PV was solved in [11]. The learning agent kept an eye on the surrounding conditions and then calculated the amount of voltage variation in the PV array. This was understood as a command to perform deeds and get benefits in return. Reward-based learning takes place again and over again. In order to maximize power output from a particular solar array design using the RL MPPT agent, it may automatically adjust PV voltage. The learning process in every RL system is guided by the reward function. Either a dense or a sparse reward system is considered to exist in the brain. It is difficult to develop dense functions because the state size in real-world applications is too large.

Non-conventional MPP tracking methods have lately gained researchers' interest because of their necessity for resilience and flexibility. The fractional-order approach is one such example of this. Using fractional derivatives to increase tracking effectiveness in changing situations is common practice. Fractional calculus is used in this fractional-order control approach. It is one of the challenges of FOC to determine the best sequence in which to implement a system. Researchers recently studied fractional-order MPPT control. A FOINC-MPPT based on incremental conductance is proposed in [12]. The fractional integrator's parameters are optimized using RMO. ESC controllers use an integrator and a low pass filter instead of a high pass filter [13]. EVSS control may also be used for photovoltaic (PV) systems, which employ an incremental conductance algorithm (VFOINC) [14]. Models with a sliding mode can be used in real-time applications like wind energy conversion [15]. Real-world applications of a POFOSMC [16] perturbation observer-based fractional-order sliding-mode controller have been documented. Real-world data has been used to validate a number of Artificial Neural Network-based models [17, 18].

A. Research Gap and Contribution

The author has compiled a list of research gaps that should be addressed.

1. Reinforcement learning in MPPT reduces the model-parameter sensitivity, making the issue easier to solve. The model works well in a wide range of climatic conditions without any redesign or retuning of the parameters.

2. This means that Q-learning does not require any models. The agent is now guided by their own experiences to take the next move. Method predicts state-action value function based on goal policy. This issue is under control till the search space is limited. A multi-dimensional search space necessitates a database with millions of records. A neural network selects the action using Deep Q-Learning, which eliminates this disadvantage.

3. Non-integer-based fractional-order control systems are often preferred

over integer-based ones because they are more accurate and have a more distinct spatial domain. 3. The fractional-order approach can address this problem.

The author is prompted by the research gap and propose a novel Fractional-order MPPT control algorithm with DQN in reinforcement learning for a solar system under partial shading. In order to maintain track of MPP, an agent known as RL gets data from 7 sources, including PV power and voltage, PV generated power diversion with PV's planned capability, integral of power diversion, coupling voltage, and divergence per unit time with reference coupling DC. The agent's tracking speed is boosted by the output of the fractional-order with error and derivative error terms, making the method more efficient. Also, it is tested using real data from New Delhi, India's capital city, to confirm that it is accurate.

Following the motivation, the paper is categorized further in the proposed work discussion (section II) and the results' discussion (section III). The work is concluded in section (IV), followed by the references.

II. Proposed Deep Q Reinforcement Learning to Control MPPT

The author proposes a fractional RL order to handle the issue of maximum power point tracking (MPPT) in PV arrays. With RL, there is no need for parametric information on dynamic model parameters. The DQN (Deep Q-Network) model is used to explain the algorithm's system parallels. The figure 1 is a depiction of the proposed model, whereas figure 2 reveals the architecture of the fractional blocks.

Reward-based learning has four basic components; these are the state-space X, the reward r, the transition probability p, and the action-space U. Any change to the variable VPV in corresponds to an action in the MPPT control issue. This system moves from state $x_t \in X$ to state x_{t+1} and the agent receives feedback, which is known as a reward, that quantifies the quality of the action or step taken by the agent through every contact it has with its surroundings. Put another way, the incentive serves as a sort of "hint" to help you find the best possible solution to your problem. The goal of the RL approach is to find an optimum policy that satisfies. The expected reward J_π for the policy π is

$$J^* = \max_{\pi} J_{\pi} \max_{\pi} E_{\pi}\{r_{\pi} \mid x_t = x\} \tag{1}$$

The current, power and voltage govern the state-space of DQN in MPPT controllers.

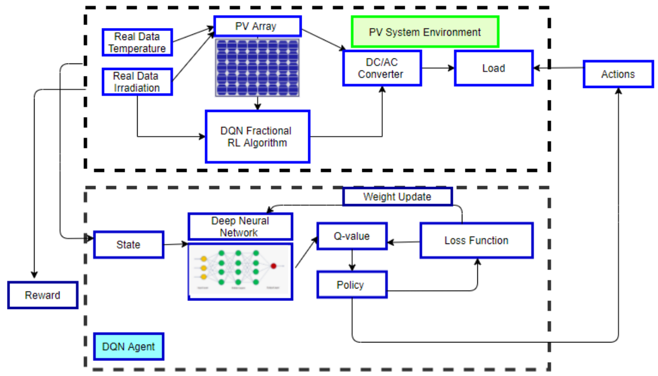

Figure 1. Proposed DQN with fractional order to improve the MPPT.

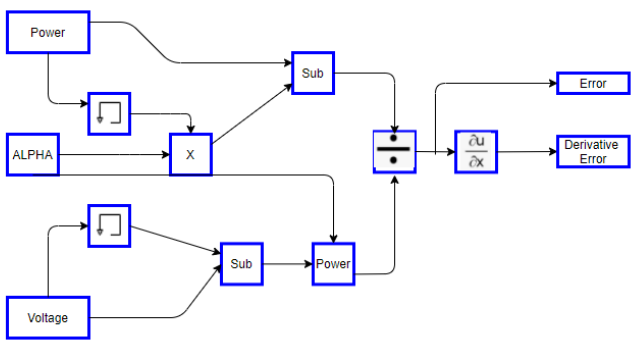

Figure 2. Schematic design of the Fractional Block.

$[V_{PV}, I_{PV}, P_{PV}, \Delta P_{PV}, \int \Delta P_{PV}, \alpha e(t)]$ and the coupling point ΔV_{DC} constitute the state space in this work. The integral of ΔV_{DC} is the duty cycle regulator in the interval of $[0, 1]$. The action space is also populated by the 100 possible discrete duty cycle values in between $[0, 1]$. The action space can also be increased to increase the complexity. However, we have set up a tradeoff between both in this work.

Reward formulation requires careful study of the problem. In the MPPT of PV array, $[\Delta V_{DC}, \Delta P_{PV}, e(t), exb, \Delta e(t)]$ are the decision elements of the reward. The violation of threshold level for these values is penalized and rewarded if the action goes along these. The pseudo code for the proposed work is listed in algorithm 1.

Algorithm 1. Pseudo code for the proposed DQN MPPT controller.

1. Set up the solar cell with open and short circuit voltage and current of 64.2 V and 5.96 A, respectively.

2. Calculate the theoretical maximum power generated from an array of 1 PV in series and 12 in parallel by $P_{npp} = (N_S \times V_{mpp}) \times (N_P \times I_{mpp})$

3. Set the initial state space, action and reward for the DQN-RL.

4. Fraction of Error and change in the error of the PV power is regulated by the parameter $0 < \alpha \leq 1$.

5. Calculate the reward function by $[\Delta V_{DC}, \Delta P_{PV}, \alpha e(t), exb, \alpha \Delta e(t)]$

6. Set the initial values of replay buffer R, learning rate β and discount factor γ.

7. Start the loop for 1 to M do

8. Initialize the state x_0

9. Form the discrete set of actions, select the action

10. Observe the new state x_0 and reward r

11. IF $|R| > N$

12. Change the Q learning policy

13. $Q^{new}(x_t, u_t) \leftarrow Q(x_t, u_t) + \beta(r_t + \gamma \max_a Q(x_{t+1}, a)) - Q(x_t, u_t)$

14. end if

15. Set $x_t = x_{t+1}$

16. End the loop

17. Save the final tuned policy for maximum MPPT

III. Results and Discussion

In order to test the hypothesis, the 330-Watt solar module is used. It is also tested using the New Delhi region's data set in normal circumstances. The Simulink model of the PV cell's performance is shown in Figure 3. This graph depicts the temperature of 25 °C under various irradiation conditions. 0.2 kW is the cell's maximum output at 200 watts/sq. m. Its efficiency rises as solar radiation increases. The RL agent's action space is scheduled to be 100×1 matrices in size over the step size (0:0.01:1). The learning rate is set at 0.01 since L2 regularisation has values of 0.0001 and a mini-batch size of $N = 64$. $M = 300$ episodes are utilized for training the model with a step size of T/T_s (total simulation time incurred and sampling time). The model considers the values of fractional order in between 0 and 1.

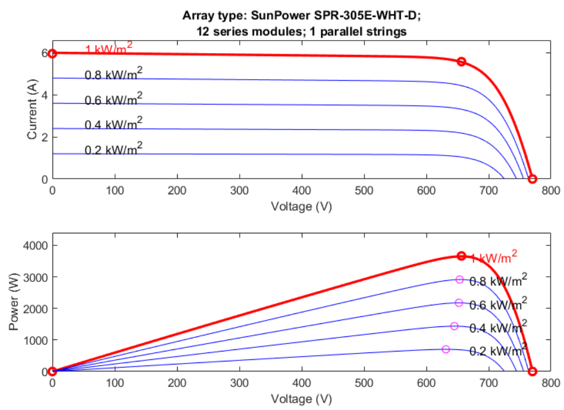

Figure 3. Current and voltage under standard environmental conditions for 330 watt solar module.

The RL training of DQN network for the proposed model is shown in figure 4 for the episodes of 300. The figure indicates the early training convergence, which is possible due to the fractional-order scheme. The proposed scheme is compared with other schemes such as perturbation and Observation (P&O) [23], a fuzzy logic controller (FLC) [24], and fractional order fuzzy logic controller (FOFLC) [25]. The harmonic distortion is the evaluation parameter here. Table 1 shows the comparative results with the state-of-the-art schemes. The New Delhi region's temperature and solar irradiance are collected from the solcast API toolkit to validate the real data.

They have temperatures ranging from 2 to 47 degrees Celsius (35.6 to 116.6 Fahrenheit), with the lowest and highest temperatures ever recorded being 2.2 and 48.4 degrees Celsius (28.0 to 119.1 degrees Fahrenheit). Summers are hotter than winters, with average temperatures ranging from 13 to 32 degrees Celsius (55 to 90 degrees Fahrenheit).

Table 1. Comparison with the state-of-the-art schemes.

Parameters	Proposed FODQN	DQNFR	FOFLC	FLC	P&O
Power (kw/m^2)	2.7	2.8	2.26	2.28	1.5
THD	-15.48	-26.1506	-6.6345	-5.250	2.23
Settling Time (s)	0.121	0.231	0.1847	0.1847	0.3225
Temperature : 0 - 50°C					
Irradiance $(\frac{W}{m^2}) = 680 - 1000$					

In order to meet the model's input criteria, the data was pre-processed. The testing is done for a day in 2019, for which the data was gathered. Figure 4 depicts the data that was gathered for analysis. Once trained and evaluated under regular settings, this model is ready to be used in real-world scenarios. Once these conditions have been replicated, the model is retested in a variety of real-world settings, including varied temperatures and levels of irradiance.

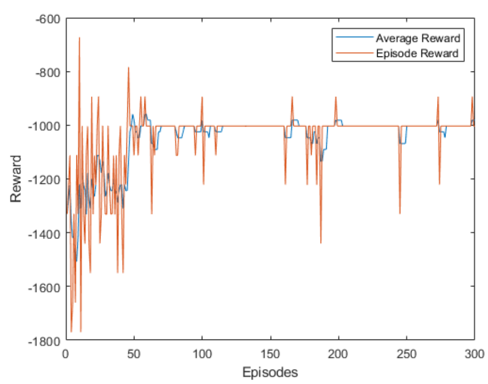

Figure 4. Episodes training for the FODQN MPPT controller for 330 watt Solar module.

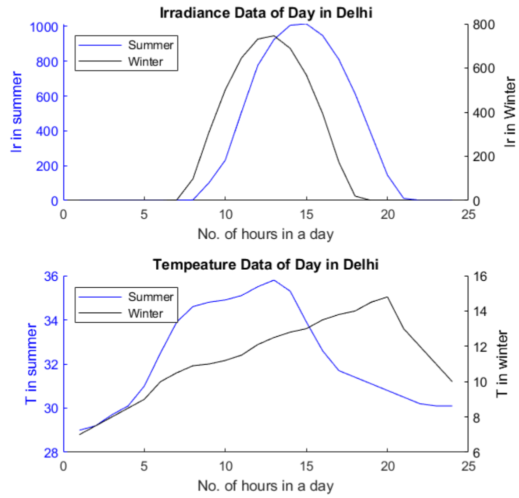

Figure 5. Testing data for a day in winter and summer.

Only the application of the pre-processed data and the observation of the outcome is required for testing. To ensure the model's validity, the author has run it through both the hot and cold seasons in New Delhi. For testing reasons, the severe weather conditions of the winter and summer months are chosen.

In light of the findings, the model may be applied in a number of settings. It is possible to utilize a field model in the same way that a computer-simulated model is used. As a further advantage, this design requires no more training after it has been implemented. Summer and winter simulation results are shown in Figures 5 and 6, respectively.

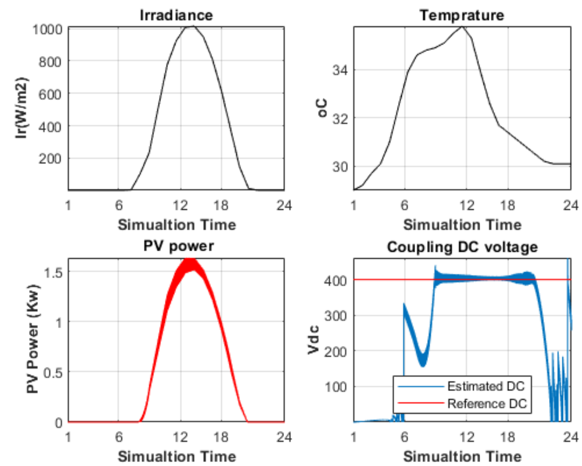

Figure 5. Photovoltaic power and coupling DC voltage for the summer data of New Delhi by FODQN MPPT.

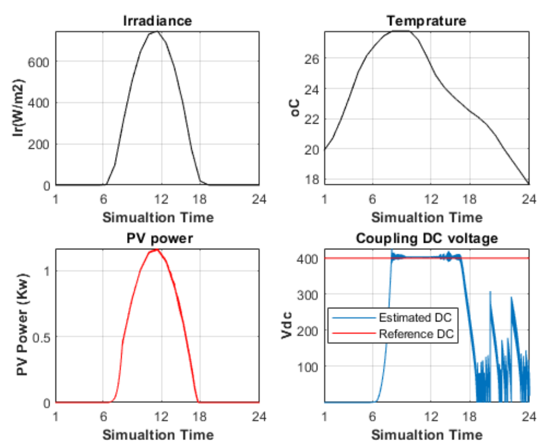

Figure 5. Photovoltaic power and coupling DC voltage for the winter data of New Delhi by FODQN MPPT.

Conclusion

Reinforcement learning and the fractional-order notion are used to develop a new model that monitors the highest power tracking point. A Deep Q-learning technique employs reinforcement learning to make the model independent of parametric design modifications for adjusting to environmental impacts. In addition, once the MPP has been taught, the network will continue to do so accurately. Tracking time for peak level, stable output, and minimal THD component are all predicted to benefit from the addition of fractional order in tracking time. This is done in two stages: first, by comparing the design to other existing algorithms, and then by evaluating it on an actual data set to see how it performs under various scenarios. FODQN has the lowest THD component, the fastest tracking time, and the biggest MPP power output when compared to other algorithms. According to the comparison study, energy harvesting may be maximized across a wide variety of radiation circumstances. As long as solar energy is at its lowest recorded level, you may still use this model in any place without adjusting its parameters and still get the most power out of it.

References

[1] Adams, Samuel and Alex O. Acheampong, Reducing carbon emissions: The role of renewable energy and democracy, Journal of Cleaner Production 240 (2019), 118245.

[2] Bouraiou, Ahmed, et al., Status of renewable energy potential and utilization in Algeria Journal of Cleaner Production 246 (2020), 119011.

[3] Moharm, State of the art in big data applications in microgrid: A review, Adv. Eng. Informatics 42 (2019).

100945,http://dx.doi.org/10.1016/j.aei.2019.100945.

[4] L. Panwar, S. C. Kaushik, S. Kothari, Role of renewable energy sources in environmental protection: A review, Renew. Sustain. EnergyRev. 15(2011), 1513-1524, http://dx.doi.org/10.1016/j.rser.2010.11.037

[5] K. Punitha, D. Devaraj and S. Sakthivel, Development and analysis of adaptive fuzzy controllers for photovoltaic system under varying atmospheric and partial shading condition, Applied Soft Computing 3(11) (2013), 4320-4332.

[6] Liu, Liqun, et al. Simulation and hardware implementation of a hill-climbing modified fuzzy-logic for maximum power point tracking with direct control method using a boost converter, Journal of Vibration and Control 21(2) (2015), 335-342.

[7] Kermadi, Mostefa, et al., An effective hybrid maximum power point tracker of photovoltaic arrays for complex partial shading conditions, IEEE Transactions on Industrial Electronics 66(9) (2018), 6990-7000.

[8] Rizzo, Santi Agatino and Giacomo Scelba, ANN-based MPPT method for rapidly variable shading conditions, Applied Energy 145 (2015), 124-132.

[9] Khan, Said G., et al. Reinforcement learning and optimal adaptive control: An overview and implementation examples, Annual reviews in control 36(1) (2012), 42-59.

[10] Zhu, Zhuangdi, Kaixiang Lin, and Jiayu Zhou, Transfer Learning in Deep Reinforcement Learning: A Survey, arXiv preprint arXiv:2009.07888 (2020).

[11] Wei, Chun, et al. Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems, IEEE Transactions on Industrial Electronics 62(10) (2015), 6360-6370.

[12] Hsu, Roy Chaoming, et al., A reinforcement learning-based maximum power point tracking method for the photovoltaic array, International Journal of Photoenergy, (2015).

[13] Al-Dhaifallah, Mujahed, et al., Optimal parameter design of fractional order control based INC-MPPT for the PV system Solar Energy 159 (2018), 650-664.

[14] Liu, Jianxin, Tiebiao Zhao, and YangQuan Chen, Maximum power point tracking with fractional order high pass filter for proton exchange membrane fuel cell, IEEE/CAA Journal of Automatica Sinica 4(1) (2017), 70-79.

[15] Yu, Kuo-Nan, Chih-Kang Liao and Her-Terng Yau, A new fractional-order based intelligent maximum power point tracking control algorithm for photovoltaic power systems, International Journal of Photoenergy, (2015).

[16] Ebrahimkhani, Sadegh, Robust fractional-order sliding mode control of doubly-fed induction generator (DFIG)-based wind turbines, ISA transactions 63 (2016), 343-354.

[17] Yang, Bo, et al., Perturbation observer based fractional-order sliding-mode controller for MPPT of grid-connected PV inverters: Design and real-time implementation, Control Engineering Practice 79 (2018), 105-125.

[18] Bouakkaz, Mohammed Salah, et al., ANN based MPPT Algorithm Design using Real Operating Climatic Condition, 2020 2nd International Conference on Mathematics and Information Technology (ICMIT), IEEE, 2020.

[19] Kumar, Pankaj, Nitai Pal and Himanshu Sharma, Performance analysis and evaluation of 10 kWp solar photovoltaic array for remote islands of Andaman and Nicobar Sustainable Energy Technologies and Assessments 42 (2020), 100889.

[20] Shongwe, Samkeliso and Moin Hanif, Comparative analysis of different single-diode PV modeling methods, IEEE Journal of Photovoltaics 5(3) (2015), 938-946.

[21] Tang, Shiqing, et al., An enhanced MPPT method combining fractional-order and fuzzy logic control, IEEE Journal of Photovoltaics 7(2) (2017), 640-650.

[22] Ma, Chengbin and Yoichi Hori, Fractional-order control: Theory and applications in motion control [past and present], IEEE Industrial Electronics Magazine 1(4) (2007), 6-16.

[23] B. Bonilla, et al., Fractional differential equations as alternative models to nonlinear differential equations, Applied Mathematics and Computation 187(1) (2007), 79-88.

[24] Ahmed, Jubaer and Zainal Salam An improved perturb and observe (P&O) maximum power point tracking (MPPT) algorithm for higher efficiency, Applied Energy 150 (2015), 97-108.

[25] Sadeq D. Al-Majidi, Maysam F. Abbod and Hamed S. Al-Raweshidy, A novel maximum power point tracking technique based on fuzzy logic for photovoltaic systems, International Journal of Hydrogen Energy 43(31) (2018), 14158-14171.

[26] Kumar, Anupam and Vijay Kumar, Performance analysis of optimal hybrid novel interval type-2 fractional-order fuzzy logic controllers for fractional-order systems, Expert Systems with Applications 93 (2018), 435-455.

[27] Babaei, S. Milad, Meisam Yahyazadeh and Hassan Fatehi Marj, Novel MPPT for Linear-Rotational Sun-Tracking System Using Fractional Fuzzy Grey-Based Sliding Mode Control, Iranian Journal of Science and Technology, Transactions of Electrical Engineering (2020), 1-23.

[28] Fathabadi, Hassan, Comparative study between two novel sensorless and sensor based dual-axis solar trackers, Solar Energy 138 (2016), 67-76.

[29] Chu, Chen-Chi and Chieh-Li Chen, Robust maximum power point tracking method for photovoltaic cells: A sliding mode control approach Solar Energy 83(8) (2009), 1370-1378.

[30] Eltamaly, Ali M., Performance of smart maximum power point tracker under partial shading conditions of photovoltaic systems, Journal of Renewable and Sustainable Energy 7(4) (2015), 043141.

[31] Guenounou, Ouahib, Boutaib Dahhou and Ferhat Chabour, Adaptive fuzzy controller based MPPT for photovoltaic systems, Energy Conversion and Management 78 (2014), 843-850.

[32] Lin, Chia-Hung, et al., Maximum photovoltaic power tracking for the PV array using the fractional-order incremental conductance method Applied Energy 88(12) (2011), 4840-4847.

[33] Yu, Kuo-Nan, Her-Terng Yau and Chi-Kang Liao, Development of a fractional-order chaos synchronization dynamic error detector for maximum power point tracking of photovoltaic power systems, applied sciences 5(4) (2015), 1117-1133.

[34] Tang, Shiqing, et al., An enhanced MPPT method combining fractional-order and fuzzy logic control, IEEE Journal of Photovoltaics 7(2) (2017), 640-650.