Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale, Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Data-Driven Reinforcement Learning for Greenhouse Temperature Control

Farhat Mahmood, Sarah Namany, Rajesh Govindan, *Tareq Al-Ansari

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar
* Corresponding Author: talansari@hbku.edu.qa

ABSTRACT

Efficient temperature control in greenhouses is essential for optimal plant growth, especially in arid regions where the harsh environment poses significant challenges to maintaining a stable microclimate. Maintaining the optimum temperature range directly influences healthy plant development and overall agricultural productivity, impacting crop yields and financial outcomes. However, the greenhouse in the present case study fails to maintain the optimum temperature as it operates based on predefined settings, limiting its ability to adapt to dynamic climate conditions. To maintain an ideal temperature range within the greenhouse while dynamically adapting to fluctuating external conditions, this study introduces a control framework using Deep Deterministic Policy Gradient, a model-free deep reinforcement learning algorithm, to optimize temperature control in the closed greenhouse. A deep neural network is trained using historical data collected from the greenhouse to accurately represent the nonlinear behavior of the greenhouse system under varying conditions. The deep deterministic policy gradient algorithm learns optimal control policy by interacting with a simulated greenhouse environment, continuously adapting without needing an explicit system dynamics model. Results from the study demonstrated that, over a five-day simulation period, the deep deterministic policy gradient control system outperformed the existing greenhouse climate system in temperature regulation. It achieved a mean squared error of 0.01°C and a mean absolute error of 0.13°C. Additionally, the deep deterministic policy gradient algorithm demonstrated a significant improvement in energy efficiency, reducing total energy consumption by 6.80% compared to the current system.

Keywords: Reinforcement learning, Greenhouse temperature control, Closed environment agriculture

INTRODUCTION

By 2050, the global population is projected to rise to about 9.7 billion, significantly increasing the need for water, energy, and food. To satisfy this growing food demand, agricultural production must be increased by nearly 70% from current levels. However, the persistent effects of climate change create significant challenges for increasing crop yields via conventional open-field farming (1). More frequent extreme weather events, higher temperatures, and changing rainfall patterns compromise these traditional methods. These environmental challenges make traditional farming methods more vulnerable. Therefore, it's essential to develop innovative, high-efficiency agricultural practices that preserve land and resources while providing greater yields (2).

Controlled Environment Agriculture (CEA) represents a viable solution, as it offers a regulated environment that mitigates climate-related risks (3). Among the various CEA technologies, greenhouses effectively optimize crop production by regulating environmental variables such as temperature, humidity, and CO_2 levels. However, efficiently managing greenhouse microclimates still poses challenges, necessitating a careful balance of various factors to promote plant growth while minimizing energy and water consumption. To address these challenges, various control strategies have been developed to regulate greenhouse environments, ranging from traditional rule-based methods to advanced model-based approaches. However, the complexity and variability of greenhouse conditions require more adaptive and data-driven solutions.

Reinforcement learning (RL) is a data-driven method that can optimize greenhouse climate control without depending on explicit system models (4). Unlike traditional controllers, RL dynamically adjusts to changing conditions, making it well-suited for complex and uncertain greenhouse environments. Deep reinforcement learning (DRL) methods, such as deep deterministic policy gradient (DDPG) and proximal policy optimization (PPO), have demonstrated significant potential in managing greenhouse operations. These methods use neural networks to learn optimal control policies from data, allowing for real-time adaptation to disturbances such as solar radiation fluctuations and external temperature changes.

Several studies have explored RL-based greenhouse control strategies, demonstrating promising results in improving energy efficiency and climate stability. For instance, a DRL-based controller integrated with robust optimization reduced energy consumption by up to 57% while maintaining stable temperature regulation (5). Multi-agent reinforcement learning approaches have also been applied to coordinate multiple greenhouse subsystems—such as heating, ventilation, and cooling—leading to improved operational efficiency (6). While RL-based strategies have demonstrated significant potential, their implementation still encounters significant challenges.

Despite the advancements, multiple challenges persist in RL-based greenhouse control systems. Training RL agents requires a considerable volume of high-quality data, and inadequate exploration strategies may result in unstable learning behaviors. Furthermore, ensuring safety constraints and interpretability in RL-based controllers continues to represent a significant research area. Additionally, a significant obstacle in RL deployment is the reliance on precise greenhouse models for training and evaluation.

Many researchers use physics-based models to simulate greenhouse dynamics; however, these models often include simplifying assumptions that may not accurately reflect real-world conditions. For instance, they often oversimplify thermal behaviors and may overlook difficult-to-measure factors, such as air leakage through greenhouse walls, heat exchange with the ground, or variations in moisture levels. Such omissions can lead to less accurate predictions about temperature variations and energy consumption, limiting the effectiveness of control strategies. Furthermore, greenhouses present unique challenges compared to traditional structures because of their specific design requirements. They are designed to maximize sunlight for plant growth while also maintaining a stable internal environment. This dual objective complicates thermal and energy management, making traditional control methods less effective when environmental conditions change rapidly. Given these limitations, a data-driven approach can offer a more precise representation of greenhouse dynamics, allowing RL agents to make better-informed decisions.

To address the challenges associated with greenhouse management, this study introduces a deep deterministic policy gradient approach based on data-driven modeling. Operational data has been collected from an actual greenhouse to train a deep neural network (DNN) surrogate model, which captures the intricate relationships among various environmental variables. Utilizing the developed DNN, a DRL agent using the DDPG algorithm was developed to maintain optimal temperature levels within the greenhouse while simultaneously minimizing energy consumption. The performance of the DDPG controller was evaluated in comparison to the existing greenhouse control system under deterministic conditions.

METHODOLOGY

This study introduces a DRL approach that utilizes the DDPG algorithm to optimize greenhouse temperature control while minimizing energy consumption. The framework integrates two essential components: a DNN surrogate model and a DRL-DDPG architecture. The DNN model, trained using real-world greenhouse data, captures the dynamics of the greenhouse system, including weather conditions, internal temperature, fan operation, HVAC temperature (HT) settings, and energy usage. This surrogate model was integrated into a DRL environment to simulate greenhouse behavior, allowing the DDPG agent to learn an optimal control policy.

Greenhouse system

The greenhouse, located in Qatar, utilizes glass as its covering material to enhance the penetration of photosynthetically active radiation into the plant canopy, ensuring optimal light for crop growth. To regulate cooling, the facility is equipped with an HVAC system alongside a hydroponic setup that efficiently supplies water, nutrients, and fertilizers to the plants. Figure 1 demonstrates the closed greenhouse.

The greenhouse operates as a closed-loop system that captures hot, humid air from the plant canopy. Instead of releasing this air into the atmosphere, it is redirected to an air-to-water chiller, where its temperature is reduced below the dew point (12-13°C). This cooling mechanism enhances the condensation of water vapor present in the air, enabling the collection of the resulting water for reuse in irrigation, significantly reducing water loss. The cooled air is subsequently reheated to an optimal temperature by the HVAC system prior to its recirculation back into the greenhouse. To meet irrigation requirements, additional water is supplemented from groundwater sources and treated through a reverse osmosis process. The climate control system is calibrated to maintain optimal growth temperatures for crops, sustaining 18°C during the night and 23°C throughout the day.


Figure 1: Closed greenhouse located in Qatar.

Data collection

Data was collected every 5 min over a one-year period, from September 2023 to September 2024, utilizing a network of sensors installed at the greenhouse. These sensors continuously monitored environmental variables, with solar irradiance and external temperature recorded by devices mounted on the greenhouse roof. The internal temperature within the greenhouse was measured through a sensor box located inside the unit. During the summer months, solar irradiance reached a peak value of 1060 $W.m^{-2}$, and external temperatures to as high as 45.2 °C. A summary of the ranges for all measured variables is detailed in Table 1.

Table 1: Range of measured variables.

Parameter	Value and unit
Solar radiation	0 – 1060 $W.m^{-2}$
Outside temperature	7.5 – 45.2 °C
Fan speed	25 – 100 %
HVAC control	17 – 22 °C
Greenhouse temperature	12.1 – 25.2 °C

Deep reinforcement learning

The DDPG algorithm is a model-free, off-policy actor-critic approach in DRL. It builds upon the deterministic policy gradient method and is appropriate for dynamic environments with large, continuous action spaces (7). DDPG learns an optimal policy through the interaction of two primary components: the actor and critic networks.

The actor network, denoted as $\mu(s \mid \theta^{\mu})$, maps states to deterministic actions. Its goal is to identify the best possible action for a given state by selecting actions that maximize the Q-value predicted by the critic network. The critic network, represented as $Q(s, a \mid \theta^{Q})$, evaluates the quality of actions by estimating the expected cumulative future rewards for each state-action pair. The actor network is trained to improve its policy by maximizing the Q-value, while the critic network is updated using temporal difference learning to refine its predictions based on immediate rewards and estimated future values.

The DDPG algorithm incorporates several techniques to enhance learning stability and efficiency. A replay buffer stores past experiences, which include the current state, the action taken, the resulting reward, the next state, and whether the episode has terminated. By sampling and learning from these experiences, the agent breaks the temporal correlation between consecutive samples, improving the quality and efficiency of the training process.

Separate target networks are used for both the actor and critic. These networks provide stable reference values during training, especially when calculating the temporal difference error. To further enhance stability, DDPG employs a soft update strategy, where the target networks' parameters are gradually updated, preventing abrupt changes that could destabilize the learning process. The weight updates for the actor and critic networks are guided by mathematical equations. The actor's weights are updated to maximize the Q-value of the policy's actions(8):

$$\Delta\theta = \eta_\theta \nabla_\theta \left(\frac{1}{N} \sum_{k \in \mathcal{B}} Q\big(s(k), \pi(s(k) \mid \theta_\pi) \mid \phi_Q \big) \right) \tag{1}$$

Here, $\Delta\theta$ represents the change in the actor's parameters, $\eta_\theta$is the learning rate, ∇_θ is a minibatch sampled from the replay buffer, and N is the number of samples in the minibatch. The critic's weights are updated to minimize the difference between the predicted Q-value and the target Q-value (8):

$$\Delta\phi = \eta_\phi \nabla_\phi \left(\frac{1}{N} \sum_{k \in \mathcal{B}} \big(Q(s(k), u(k) \mid \phi_Q) - \nu(k) \big) \right) \tag{2}$$

Here, $\Delta\phi$ is the change in the critic's parameters, $u(k)$ is the action at time step k. Target Q-value is calculated as:

$$\nu(k) = r(k) + \gamma Q'\big(s(k+1), \pi'(s(k+1) \mid \theta_{\pi'}) \mid \phi_{Q'} \big) \tag{3}$$

where $r(k)$ is the immediate reward, γ is the discount factor, Q' is the target critic network, and π' is the target policy network. By iteratively refining the actor and critic networks, DDPG efficiently learns optimal control policies without requiring an explicit model of the environment, making it highly suitable for control tasks in complex, continuous systems.

System model

The surrogate model's accuracy in predicting future states is critical for an effective control strategy.

Traditional analytical models for greenhouses often rely on estimated parameters, introducing errors that compromise predictive reliability. To overcome this, a DNN was developed using PyTorch (9) to accurately represent the greenhouse system, capturing the nonlinear relationships among variables like solar irradiance, ambient temperature, HT, and fan speed.

The architecture of the DNN includes an input layer that processes 12 input features as illustrated in **Figure 2**. The model is trained using the Adam optimizer, which efficiently handles parameter updates, and the MSE loss function, which quantifies the prediction error. Training is conducted over 50 epochs to ensure the model converges to a solution that accurately captures the system's dynamics. By using this DNN-based surrogate model, the greenhouse system's behavior can be predicted with greater precision, providing a robust foundation for the control strategy.

Figure 2: DNN model for the greenhouse system.

Reward function

The reward function for the DDPG agent is designed to achieve two main goals: minimizing the difference between the greenhouse temperature and the desired set point and reducing energy consumption. By penalizing both temperature deviations and energy usage, the reward function ensures the agent prioritizes maintaining optimal conditions while being energy efficient. The function assigns specific weights to each component, allowing for flexibility in emphasizing either precise temperature control or energy savings, depending on the desired objectives. The reward function is expressed as:

$$R(k) = -\left(\alpha \cdot |T(k) - T_{\text{set}}(k)| + \beta \cdot E(k)\right) \tag{4}$$

Training

The critic network in the DDPG framework processes both state and action inputs through two fully connected layers, each containing 256 neurons. These layers use ReLU activation functions and include batch normalization to stabilize and improve the learning process. The network outputs a single Q-value, representing the expected cumulative reward for the given state-action pair. Similarly, the actor network processes the state inputs through two fully connected layers with 256 neurons each. These layers also use ReLU activation functions, along with batch normalization, to ensure stable and efficient learning. The network concludes with a Tanh activation function, which outputs the control actions within the required range. The training phase for the DRL agent spaned 500 epochs, with each epoch representing one day of operation, divided into 288 steps corresponding to 5-minute intervals. This high-resolution sampling enabled the agent to respond effectively to rapid changes in the greenhouse environment, which was crucial for maintaining optimal growing conditions.

The training process used a learning rate of $1{\times}10^{-4}$ to ensure steady weight updates. A soft update coefficient τ of 0.005 was employed for gradual updates to the target networks, enhancing stability during learning. The discount factor γ was set to 0.99, balancing the importance of immediate rewards with long-term gains.

RESULTS AND DISCUSSION

Model performance

The DNN model, developed using historical data, was evaluated for its ability to predict greenhouse temperatures over a five day period from March 27th to April 1st, 2024. The model was trained on a dataset of 24,480 samples, with predictions generated at 5 min intervals. It demonstrated high accuracy, achieving an R^2 score of 0.94 and an MSE of 0.676 °C, as illustrated in Figure 3. While the model slightly underestimated greenhouse temperatures during peak solar irradiance hours on the last two days, it successfully captured the overall diurnal trend.

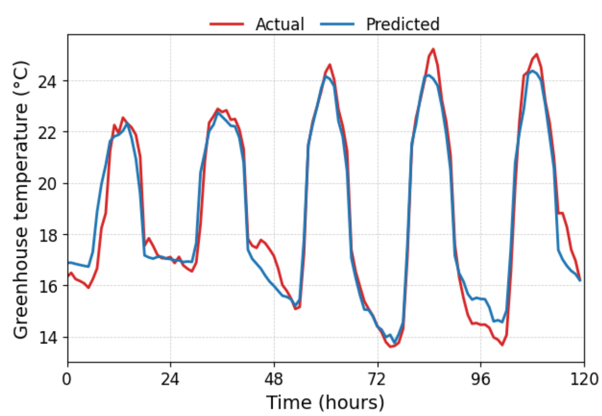

Figure 3: DNN greenhouse temperature prediction performance.

Deep deterministic policy gradient performance

The performance of the DDPG controller in regulating greenhouse temperatures was assessed over a five-day simulation, comparing its performance to the existing greenhouse control system for fixed setpoints of 23°C during the daytime and 18 °C at night. As illustrated in Figure 4, the DDPG controller outperformed the GCS in achieving temperature control. The DDPG controller (represented by the blue line) maintained temperatures within ±0.1°C of the setpoints (green line), ensuring smooth adjustments with minimal overshooting or undershooting. In contrast, the GCS experienced significant temperature variability, with actual temperatures (red line) often significantly deviating from the setpoints, reaching peaks of 24.3°C during the day and dropping to lows of 13.8°C at night. The GCS resulted in an MSE of 3.89 °C and an RMSE of 1.97 °C, while the DDPG controller achieved a lower MSE of 0.01 °C and an RMSE of 0.13 °C, demonstrating its superior performance in temperature regulation.

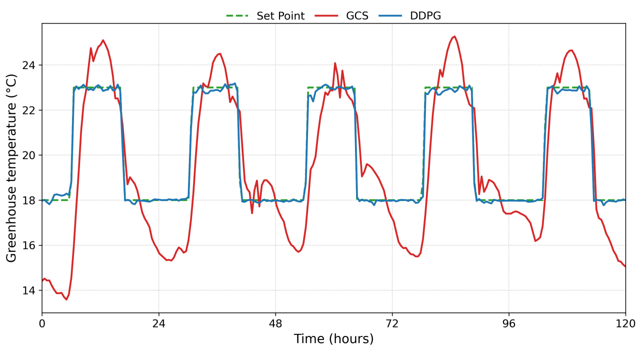

Figure 4: Temperature control performance comparison between GCS and DDPG.

The control actions of the DDPG controller were compared to the actual control actions implemented by the GCS, as demonstrated in Figure 5. The DDPG controller operated the fan at relatively lower speeds than the GCS, ensuring smoother transitions in operation. Additionally, the fan speed was always operated above 25%, maintaining consistent CO_2 circulation within the greenhouse. For HVAC control, the GCS operated on fixed temperature setpoints of 17°C at night and 22°C during the day. In contrast, the DDPG controller adopted a more dynamic approach, adjusting the temperature in response to changing environmental conditions. This adaptive behavior resulted in improved temperature regulation and overall greenhouse climate control.

Figure 6 illustrates the energy consumption over five days, demonstrating that the DDPG controller consistently resulted in lower energy consumption compared to the GCS. The DDPG controller accounted for a total energy usage of 7,789 kWh, while the GCS utilized 8,357 kWh. This led to an overall reduction of 6.80% in energy consumption with the DDPG compared to the GCS.

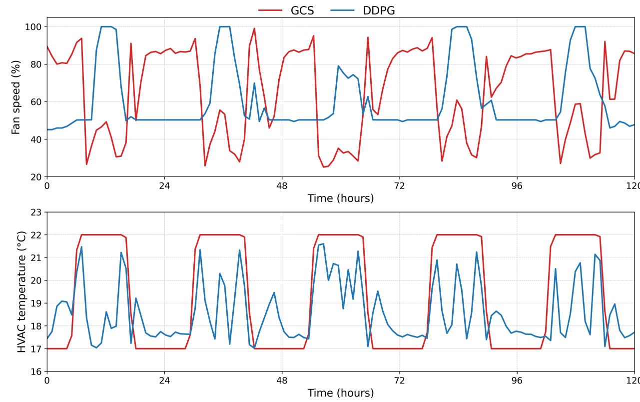

Figure 5: Control action comparison of GCS and DDPG.

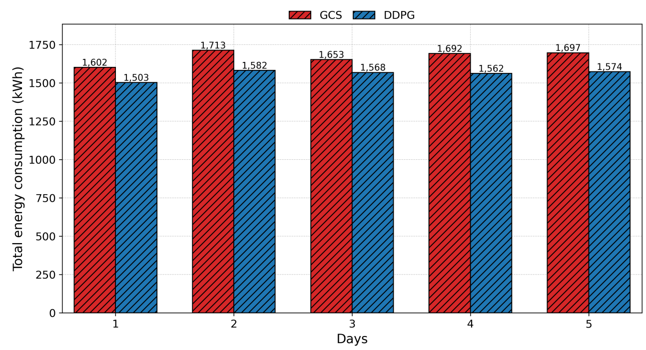

Figure 6: Energy utilization comparison of GCS and DDPG.

CONCLUSION

This study demonstrated the effectiveness of a deep reinforcement learning approach using DDPG for greenhouse temperature control. The developed control system significantly outperformed the existing greenhouse control system, achieving superior temperature regulation with a mean squared error of 0.01°C and root mean squared error of 0.13°C, compared to 3.89°C and 1.97°C respectively, for the existing greenhouse system. The DDPG controller maintained temperatures within ±0.1°C of the setpoints while implementing smoother control actions, particularly in fan speed transitions. Furthermore, the controller achieved a 6.80% reduction in energy consumption, utilizing 7,789 kWh compared to the existing greenhouse system's 8,357 kWh over the five-day testing period.

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by Hamad bin Khalifa University, Education, City Doha, Qatar, a member of the Qatar Foundation.

REFERENCES

1. Mahmood F, Govindan R, Yang D, Bermak A, Al-

Ansari T. Forecasting cooling load and water demand of a semi-closed greenhouse using a hybrid modelling approach. International Journal of Ambient Energy. 2022 Dec 31;43(1):8046–66.
2. Sajid MU, Mahmood F, Bicer Y, Al-Ansari T. Cost analysis of agricultural greenhouse yield production using spectrum selective fluids. Solar Energy. 2024 Jun;275:112607.
3. Mahmood F, Ghiat I, Hazrat B, Luqman M, Al-Ansari T. A framework based on an input-yield model for greenhouse optimisation for varying environmental conditions. Energy Conversion and Management: X. 2024 Apr;22:100555.
4. Goldenits G, Mallinger K, Raubitzek S, Neubauer T. Current applications and potential future directions of reinforcement learning-based Digital Twins in agriculture. Smart Agricultural Technology. 2024 Aug;8:100512.
5. Ajagekar A, Mattson NS, You F. Energy-efficient AI-based Control of Semi-closed Greenhouses Leveraging Robust Optimization in Deep Reinforcement Learning. Advances in Applied Energy. 2023 Feb;9:100119.
6. Ajagekar A, Decardi-Nelson B, You F. Energy management for demand response in networked greenhouses with multi-agent deep reinforcement learning. Appl Energy. 2024 Feb;355:122349.
7. Tan H. Reinforcement Learning with Deep Deterministic Policy Gradient. In: 2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA). IEEE; 2021. p. 82–5.
8. Morcego B, Yin W, Boersma S, van Henten E, Puig V, Sun C. Reinforcement Learning versus Model Predictive Control on greenhouse climate control. Comput Electron Agric. 2023 Dec;215:108372.
9. Imambi S, Prakash KB, Kanagachidambaresan GR. PyTorch. In 2021. p. 87–104.