Optimal Altitude Policy for Solar-Powered Autonomous Aircraft Using Value Iteration

Philip Santiago
AA228/CS238, Stanford University

***Abstract*—Solar-powered unmanned aerial vehicles (UAVs) capable of perpetual flight require sophisticated energy management strategies to balance power consumption with solar energy harvesting. This paper presents a Markov Decision Process (MDP) formulation solved via value iteration to determine optimal climb, maintain, and descend actions for the Skydweller Aero solar aircraft. The state space encompasses altitude, battery state-of-charge, and time of day, while the objective maximizes total system energy (battery plus gravitational potential energy) to ensure cyclic sustainability. Simulation results over a 24-hour flight cycle under ideal solar conditions demonstrate that the learned policy achieves energy-positive operation with a 50,000 Wh surplus, climbing to 5,200 meters during peak solar hours and returning to a higher starting altitude by cycle end. The policy discovers the importance of climbing to higher altitudes during the day, where solar irradiance is stronger, to maximize solar energy collection while simultaneously storing gravitational potential energy. The policy then instructs the airplane to descend to maximize battery power for the next cycle. This approach replaces manual parameter tuning with an automatically learned, state-dependent control policy suitable for integration into autonomous flight management systems.**

***Index Terms*—Solar aircraft, Markov Decision Process, Value iteration, Energy management, Autonomous UAV, Reinforcement learning**

I. Introduction

The development of solar-powered unmanned aerial vehicles (UAVs) capable of ultra-long endurance flight represents a significant advancement in aerospace technology. These platforms offer persistent surveillance, communications relay, and environmental monitoring capabilities without the fuel constraints of conventional aircraft. The Skydweller aircraft, a converted version of the Solar Impulse 2 experimental solar airplane, exemplifies this class of vehicles. Redesigned as a fully autonomous unmanned platform, Skydweller features a wingspan comparable to a Boeing 747 and can carry several hundred kilograms of payload while operating indefinitely on solar power [1].

The fundamental challenge in perpetual solar flight is energy management: the aircraft must harvest sufficient solar energy during daylight hours to sustain flight through the night. This requires careful optimization of altitude profiles, as altitude affects both power requirements (through air density variations) and solar energy collection (through atmospheric absorption). Higher altitudes offer reduced drag and increased solar irradiance, but climbing consumes significant energy. The timing and magnitude of altitude changes must be optimized to maximize the aircraft's energy margin at the end of each daily cycle.

Previous approaches to this problem have relied on parametric optimization methods such as golden-section search to tune control parameters including minimum power thresholds before initiating climbs and battery charge rates [2]. While effective, these methods require manual specification of the control structure and may not discover optimal strategies that deviate from the assumed form.

This paper presents a reinforcement learning approach using Markov Decision Processes (MDPs) solved via value iteration to automatically learn an optimal altitude control policy. The key contributions are:

- Formulation of the solar aircraft energy management problem as an MDP with altitude, battery state-of-charge, and time as state variables
- A reward structure that incentivizes cyclic sustainability by maximizing total system energy (battery plus gravitational potential)
- Discovery of a "climb-first" strategy that prioritizes altitude gain before sunrise
- Demonstration of sustainable operation with significant energy surplus under ideal conditions

II. Related Work

A. Solar Aircraft Design and Operations

The design and operation of solar-powered aircraft has been studied extensively since the work on Solar Challenger in the 1980s [3]. Noth [2] developed a comprehensive design methodology for solar-powered UAVs, establishing the fundamental power balance equations that govern perpetual flight feasibility. The Solar Impulse project demonstrated human-piloted solar circumnavigation, proving the viability of solar propulsion for large aircraft [4].

Key operational strategies for solar aircraft include the concept of "gravity batteries," where excess solar energy during the day is converted to gravitational potential energy through altitude gain, then recovered during night descent [5]. This strategy motivates the altitude optimization problem addressed in this work.

B. Energy Management for Electric Aircraft

Energy management for electric and hybrid-electric aircraft has received considerable attention. Traub [6] analyzed the range and endurance of battery-powered aircraft, establishing fundamental performance limits. For solar aircraft specifically, Leutenegger et al. [7] developed optimal trajectory planning methods considering solar geometry and atmospheric conditions.

C. Reinforcement Learning for Aerospace Applications

Reinforcement learning has been applied to various aerospace control problems. Bagnell and Schneider [8] demonstrated autonomous helicopter flight using learned policies. More recently, deep reinforcement learning has been applied to spacecraft attitude control [9] and UAV navigation [10]. However, application of MDP methods to solar aircraft energy management remains relatively unexplored.

III. Problem Formulation

A. Aircraft Model

The Skydweller aircraft parameters used in this study are summarized in Table I. The aircraft flies at maximum endurance speed, maintaining constant lift and drag coefficients throughout the flight.

TABLE I: Aircraft Parameters

Parameter	Value	Unit
Aircraft mass	1,400	kg
Wing surface area	142.3	m^2
Lift coefficient (C_L)	0.8	–
Drag coefficient (C_D)	0.0229	–
Battery capacity	100	kWh
Solar panel area	180	m^2
Solar panel efficiency	28	%
Motor efficiency	80	%
Propeller efficiency	85	%
Climb rate	1.0	m/s
Descent rate	2.915	m/s
Maximum motor power	40	kW

B. Power Model

The power required for level flight at altitude h is given by:

$$P_r = W\sqrt{\frac{2WC_D^2}{S\rho(h)C_L^3}} \tag{1}$$

where $W = mg$ is the aircraft weight, S is the wing area, and $\rho(h)$ is the air density at altitude h. Air density is obtained from a standard atmosphere model loaded from the U.S. Standard Atmosphere 1976 - Air Density Calculator [11].

The power required for level flight from the battery accounts for motor and propeller efficiencies:

$$P_{level} = \frac{P_r}{\eta_m \eta_p} \tag{2}$$

where η_m and η_p are motor and propeller efficiencies, respectively.

For climbing, additional power is required to increase potential energy:

$$P_{climb} = P_{level} + \frac{mg \cdot \text{ROC}}{\eta_m \eta_p} \tag{3}$$

where ROC is the rate of climb. Descending is assumed to be unpowered gliding with $P_{descend} = 0$.

C. Solar Power Model

Solar irradiance data was obtained from Bird Clear Sky Model Calculator for Solar Irradiance [12] for Madrid, Spain (40.4168°N, 3.7038°W) at altitudes ranging from 0 to 12,000 meters. The simulation assumes ideal conditions with clear skies and no cloud cover. Solar power generation is computed as:

$$P_{solar} = I(h,t) \cdot A_{panel} \cdot \eta_{panel} \tag{4}$$

where $I(h,t)$ is the irradiance at altitude h and time t, A_{panel} is the solar panel area, and η_{panel} is the panel efficiency.

D. Total System Energy

The total system energy combines battery energy and gravitational potential energy:

$$E_{total} = E_{battery} + E_{potential} = \frac{\text{SoC}}{100} \cdot C_{bat} + mgh \tag{5}$$

where SoC is the battery state-of-charge in percent, C_{bat} is the battery capacity in Wh, and h is the altitude. This formulation captures the "gravity battery" (often found in pumped hydro energy storage pumps) concept where altitude represents stored energy.

IV. MDP Formulation

A. State Space

The state space $\mathcal{S}$ is discretized along three dimensions:

- **Altitude**: 0 to 12,000 m in 100 m increments (121 levels)
- **Battery SoC**: 0 to 100% in 5% increments (21 levels)
- **Time**: 0 to 24 hours in 10-minute increments (144 steps)

This yields a total state space of $121 \times 21 \times 144 = 365{,}904$ states.

B. Action Space

The action space $\mathcal{A}$ consists of three discrete actions:

- **CLIMB**: Increase altitude by one discretization step (100 m)
- **MAINTAIN**: Hold current altitude
- **DESCEND**: Decrease altitude by one discretization step

C. Transition Dynamics

State transitions are deterministic given the current state and action. For each time step $\Delta t = 10$ minutes:

1) Altitude changes according to the action
2) Net power is computed: $P_{net} = P_{solar} - P_{consumption}$
3) Battery SoC updates: $\Delta\text{SoC} = \frac{P_{net} \cdot \Delta t}{C_{bat}} \times 100$
4) Time advances by Δt

Terminal states occur when time reaches 24 hours or battery SoC reaches zero.

D. Reward Function

The reward function is designed to encourage cyclic sustainability by ending each cycle with at least as much total energy as at the start. The reward structure is:

$$R(s,a,s') = \begin{cases} -1000 & \text{if SoC} \leq 0 \text{ (battery depleted)} \\ R_{terminal} & \text{if } t \geq T_{end} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The terminal reward $R_{terminal}$ is computed based on energy sustainability:

$$\Delta E = E_{total,end} - E_{total,start} \tag{7}$$

If $\Delta E < 0$ (energy loss):

$$R_{terminal} = -5000 + \Delta E \times 2.0 \tag{8}$$

If $\Delta E \geq 0$ (energy neutral or gain):

$$R_{terminal} = 1000 + \Delta E \times 0.5 + R_{surplus} \tag{9}$$

where $R_{surplus} = 500$ if $\Delta E > 5000$ Wh (providing a safety margin bonus).

V. Solution Method

A. Value Iteration

The optimal policy is computed using value iteration [13], [14], which iteratively updates the value function:

$$V_{k+1}(s) = \max_{a \in \mathcal{A}} [R(s,a,s') + \gamma V_k(s')] \tag{10}$$

where $\gamma = 0.99$ is the discount factor and s' is the next state resulting from action a in state s.

The algorithm terminates when the maximum change in value function falls below threshold $\theta = 10^{-3}$:

$$\max_s |V_{k+1}(s) - V_k(s)| < \theta \tag{11}$$

Algorithm 1 Value Iteration for Solar Aircraft MDP

Initialize $V(s) = 0$ for all $s \in \mathcal{S}$
repeat
    $\delta \leftarrow 0$
    for each $s \in \mathcal{S}$ **do**
        if s is not terminal **then**
            $v \leftarrow V(s)$
            $V(s) \leftarrow \max_a[R(s,a,s') + \gamma V(s')]$
            $\delta \leftarrow \max(\delta, |v - V(s)|)$
        end if
    end for
until $\delta < \theta$
Extract policy: $\pi(s) = \arg\max_a[R(s,a,s') + \gamma V(s')]$

B. Policy Extraction

Once the value function converges, the optimal policy is extracted by selecting the action that maximizes expected value for each state:

$$\pi^*(s) = \arg\max_{a \in \mathcal{A}} [R(s,a,s') + \gamma V^*(s')] \tag{12}$$

VI. Experimental Setup

A. Data Sources

Atmospheric density data was obtained from the International Standard Atmosphere model, providing density values from sea level to 12,000 m altitude. Solar irradiance data for Madrid, Spain was computed accounting for solar geometry (zenith angle, azimuth), atmospheric extinction as a function of altitude, and time of day variations.

Assumptions: The simulation assumes ideal weather conditions with clear skies and no cloud cover. This represents a best-case scenario for solar energy harvesting and provides an upper bound on system performance.

B. Initial Conditions

Simulations were initialized with:

- Starting altitude: 1,500 m (4,921 ft)
- Starting battery SoC: 50%
- Starting time: 00:00 (midnight)

These conditions represent a challenging scenario where the aircraft must survive the remainder of the night before sunrise provides charging capability.

C. Implementation

The MDP solver was implemented in Python using NumPy for efficient array operations. The implementation includes atmospheric data loader for density interpolation, solar irradiance loader with altitude and time interpolation, energy model computing power requirements and generation, MDP class with state/action space management, and value iteration solver with convergence tracking.

VII. Results

A. Policy Analysis

The learned policy exhibits a strong preference for climbing, with the overall action distribution across all non-terminal states being: CLIMB (70.1%), MAINTAIN (5.9%), and DESCEND (24.0%). This aggressive climbing strategy reflects the value of storing energy as gravitational potential during available power windows.

B. Simulated Flight Profile

A complete 24-hour simulation following the optimal policy produces the flight profile shown in Fig. 1. The key results are summarized in Table II.

TABLE II: Flight Profile Summary

Metric	Start	End
Time	00:00	24:09
Altitude	1,500 m	1,500 m
Battery SoC	50%	100%
Battery Energy	50,000 Wh	100,000 Wh
Potential Energy	8,175 Wh	8,175 Wh
Total Energy	58,175 Wh	108,175 Wh

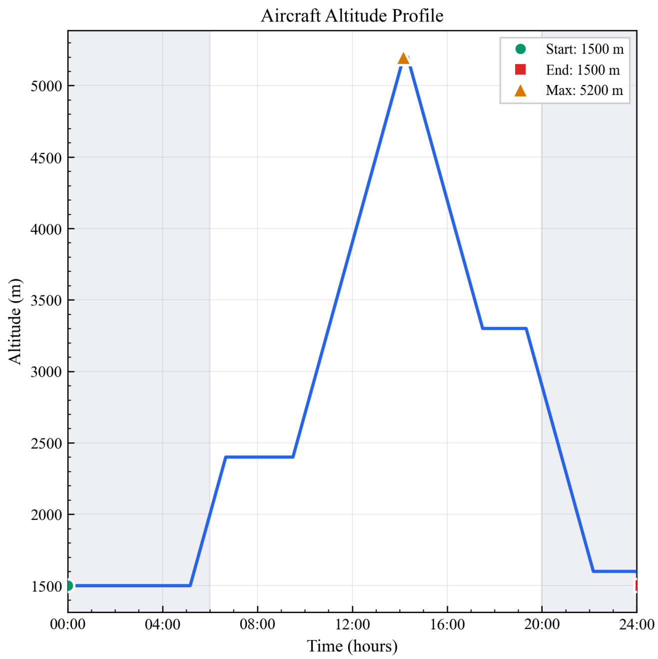

Fig. 1: Aircraft altitude profile over 24-hour cycle. The aircraft starts at 1,500 m, climbs to maximum altitude of 5,200 m by 14:09, then returns to starting altitude. Gray shaded regions indicate nighttime hours.

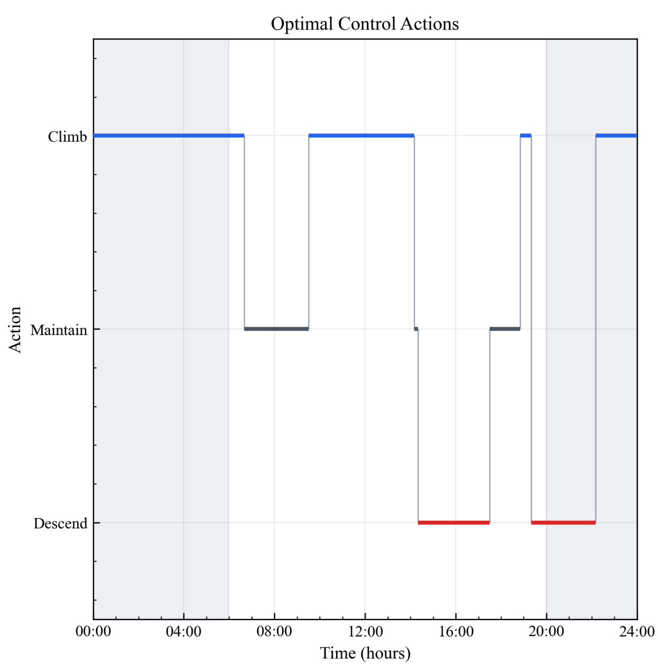

Fig. 2: Optimal control actions over 24-hour cycle. Blue indicates CLIMB, gray indicates MAINTAIN, and red indicates DESCEND. The policy exhibits four distinct climb phases and multiple descent phases.

C. Climb Strategy Analysis

The optimal policy reveals a sophisticated multi-phase strategy illustrated in Fig. 2:

Phase 1 - Pre-Dawn Climb (00:00-06:00): Surprisingly, the policy initiates climbing immediately at midnight, despite zero solar power availability. This counter-intuitive behavior draws down the battery to invest in gravitational potential energy, betting on solar recovery at sunrise.

Phase 2 - Morning Hold and Climb (06:00-14:00): As sunrise occurs around 06:00, the aircraft maintains altitude briefly while the battery reaches its minimum (5% SoC), then resumes aggressive climbing as solar power increases. Maximum altitude of 5,200 m (17,060 ft) is reached at 14:09.

Phase 3 - Afternoon Descent and Recharge (14:00-19:00): With the battery critically low, the policy initiates descent at 14:19 to reduce power consumption. During this gliding descent, all solar power goes to recharging the battery, which climbs from 5% to 100% SoC.

Phase 4 - Evening Climb and Final Descent (19:00-24:00): With a full battery, the policy executes a brief evening climb before final descent to the starting altitude, preparing for the next cycle.

D. Battery State-of-Charge Profile

Fig. 3 shows the battery SoC throughout the cycle. The dramatic drawdown from 50% to 5% during the aggressive morning climb demonstrates the policy's willingness to almost completely empty the battery to invest in the higher solar irradiance of the higher altitudes. In practice, a safety would be placed around 10-20% battery to avoid the risk of completely depleting the battery.

E. Energy Components Analysis

Fig. 4 illustrates the interplay between battery energy and gravitational potential energy. The total system energy (purple line) shows conservation during the night phase, a temporary dip during aggressive climbing, and substantial gains during the solar-powered afternoon.

F. Power Balance

Fig. 5 shows the power generation and net power flow. The solar generation curve (orange) follows the expected bell shape, peaking at approximately 35 kW around solar noon. Net power (purple) is negative during climbing phases and positive during descent/charging phases.

G. Sustainability Assessment

The critical metric for perpetual flight is cyclic sustainability. The simulation results show:

- Starting total energy: 58,175 Wh
- Ending total energy: 108,175 Wh
- Energy surplus: +50,000 Wh (+85.9%)
- Total solar energy collected: 332,205 Wh

This substantial energy surplus confirms that the learned policy achieves highly sustainable operation under ideal conditions. The 50 kWh surplus provides significant margin for

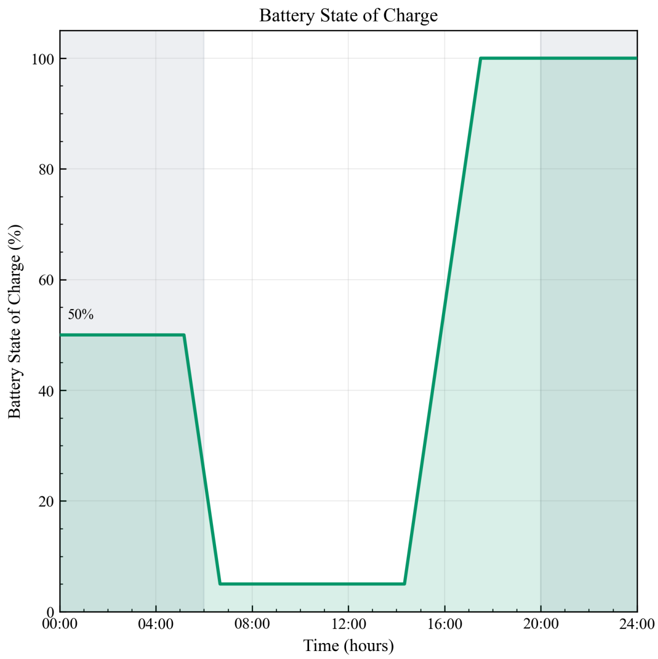

Fig. 3: Battery state-of-charge over 24-hour cycle. The battery drops from 50% to 5% during pre-dawn climbing, holds at minimum during peak solar hours while climbing continues, then rapidly recharges to 100% during afternoon descent.

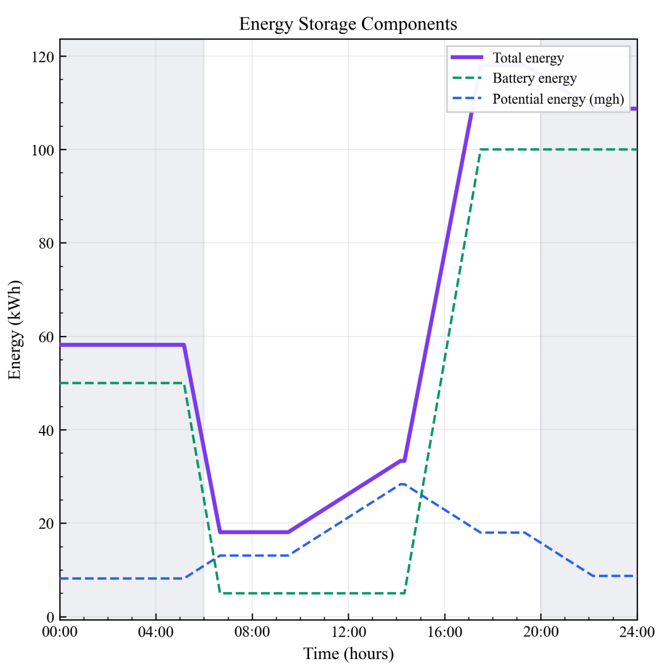

Fig. 4: Energy storage components over 24-hour cycle. Total energy (purple) combines battery energy (green dashed) and gravitational potential energy (blue dashed). The energy exchange between battery and altitude storage is clearly visible.

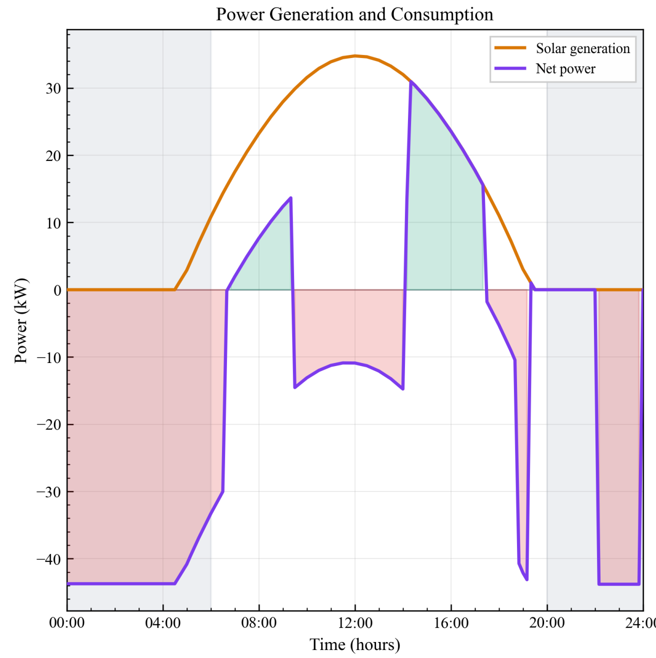

Fig. 5: Power generation and consumption over 24-hour cycle. Solar generation (orange) peaks at 35 kW at noon. Net power (purple) shows deficit during climbing and surplus during gliding descent. Green shading indicates net energy gain; red shading indicates net energy consumption.

weather variability, system degradation, and mission requirements.

VIII. Discussion

A. Key Findings

The optimal policy discovered by value iteration reveals several important insights:

Aggressive Early Climbing: The policy's decision to climb before sunrise is counter-intuitive but rational. By investing battery energy in altitude, the aircraft positions itself to harvest more solar energy at higher altitudes once the sun rises, while also reducing power requirements due to lower air density.

Battery as Buffer, Altitude as Storage: The policy treats the battery as a short-term buffer while using altitude as the primary energy storage mechanism. The battery can be drawn down to critical levels (5%) because the altitude and the assumption of perfect weather conditions provides energy security.

Timing Optimization: The precise timing of phase transitions (climb initiation, descent start, recharge completion) is automatically discovered by the MDP solver, eliminating the need for manual parameter tuning.

B. Comparison with Manual Tuning

The previous approach at Skydweller used golden-section search to optimize two parameters: minimum power before climb and charge rate. The MDP approach offers several advantages:

- **State-dependent decisions**: Unlike fixed-parameter controllers, the MDP policy adapts to the current state
- **Automatic discovery**: The algorithm discovers optimal timing without requiring manual specification of the control structure
- **Global optimization**: Value iteration finds the globally optimal policy for the discretized state space

C. Limitations and Assumptions

Several limitations should be acknowledged:

Ideal Weather Assumption: The simulation assumes perfect solar conditions with no clouds. Real-world performance would be reduced by weather variability, though the 50 kWh energy surplus provides substantial margin.

Discretization Effects: The coarse state discretization (100 m altitude, 5% SoC) may miss fine-grained optimal behaviors.

Deterministic Model: The current formulation assumes deterministic transitions, ignoring weather uncertainty, wind effects, and sensor noise.

Fixed Location: Results are specific to Madrid's solar conditions; different latitudes and seasons would require recomputation.

D. Future Work

Several extensions could enhance this work:

- **Stochastic MDP**: Incorporate weather uncertainty and model state transition probabilities
- **Deep RL**: Apply deep reinforcement learning to handle continuous state spaces
- **Multi-objective optimization**: Balance energy management with mission requirements
- **Weather robustness**: Evaluate policy performance under various cloud cover scenarios
- **Real-time adaptation**: Develop online learning methods that adapt to changing conditions

IX. Conclusion

This paper presented an MDP-based approach to optimal altitude control for solar-powered aircraft. Using value iteration, we computed a policy that determines when the Skydweller aircraft should climb, maintain altitude, or descend based on current altitude, battery state-of-charge, and time of day.

The key findings from simulation under ideal solar conditions are:

1) The optimal policy achieves energy-positive operation with an 85.9% energy surplus (50,000 Wh) over a 24-hour cycle
2) The policy discovers a counter-intuitive "climb-first" strategy, initiating altitude gain at midnight before sunrise
3) Maximum altitude of 5,200 m (17,060 ft) is reached by 14:09, with aggressive climbing depleting battery to 5%
4) Battery recharges from 5% to 100% during afternoon gliding descent
5) The aircraft returns to its starting altitude (1,500 m) with double its initial battery charge

This approach provides a principled method for energy management in solar-powered UAVs, replacing manual parameter tuning with automatically learned, state-dependent control policies. The resulting policy is suitable for integration into autonomous flight management systems for perpetual-endurance solar aircraft, with the understanding that real-world performance will depend on actual weather conditions.

Acknowledgment

The author thanks Skydweller Aero for providing the problem formulation and aircraft parameters used in this study.

References

[1] Skydweller Aero Inc., "Skydweller: The World's First Fully Solar-Powered Unmanned Aircraft System," 2023.
[2] A. Noth, "Design of Solar Powered Airplanes for Continuous Flight," Ph.D. dissertation, ETH Zurich, 2008.
[3] R. J. Boucher, "History of Solar Flight," AIAA Paper 84-1429, 1984.
[4] B. Piccard and A. Borschberg, "Solar Impulse: First Around-The-World Solar Flight," 2016.
[5] A. T. Klesh and P. T. Kabamba, "Solar-Powered Aircraft: Energy-Optimal Path Planning and Perpetual Endurance," *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 4, pp. 1320–1329, 2009.
[6] L. W. Traub, "Range and Endurance Estimates for Battery-Powered Aircraft," *Journal of Aircraft*, vol. 48, no. 2, pp. 703–707, 2011.
[7] S. Leutenegger, M. Jabas, and R. Y. Siegwart, "Solar Airplane Conceptual Design and Performance Estimation," *Journal of Intelligent and Robotic Systems*, vol. 61, pp. 545–561, 2011.
[8] J. A. Bagnell and J. G. Schneider, "Autonomous Helicopter Control Using Reinforcement Learning Policy Search Methods," *IEEE International Conference on Robotics and Automation*, 2001.
[9] B. Gaudet, R. Linares, and R. Furfaro, "Deep Reinforcement Learning for Six Degree-of-Freedom Planetary Landing," *Advances in Space Research*, vol. 65, no. 7, pp. 1723–1741, 2020.
[10] W. Koch, R. Mancuso, R. West, and A. Bestavros, "Reinforcement Learning for UAV Attitude Control," *ACM Transactions on Cyber-Physical Systems*, vol. 3, no. 2, pp. 1–21, 2019.
[11] National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, and United States Air Force, "U.S. Standard Atmosphere, 1976," NOAA-S/T 76-1562, U.S. Government Printing Office, Washington, D.C., 1976.
[12] R. E. Bird and R. L. Hulstrom, "Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces," Technical Report SERI/TR-642-761, Solar Energy Research Institute, Golden, CO, USA, February 1981.
[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.
[14] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.