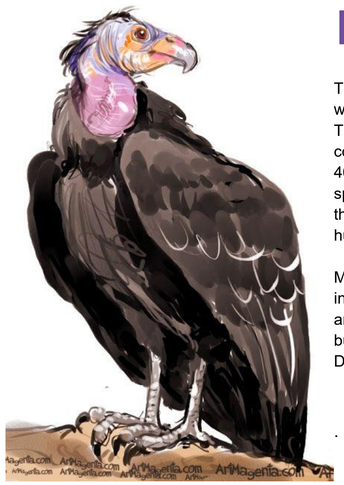
Reinforcement Learning for Long-Term Management of the California Condor Population

Team: Yuan Li, Jérémy Kim

Université de Montréal



ABSTRACT

The California condor is one of the world's most endangered birds. There are just over 500 California condors worldwide, fewer than 400 of which live in the wild. The species survives today only through intensive, long-term human conservation efforts.

Major threats to California condors include lead poisoning from ammunition, habitat loss, and rare but severe environmental Disturbances.

.

Problem: California condor recovery requires long-term management under uncertainty, balancing population sustainability and intervention costs.

Limitation of existing approaches: Traditional conservation policies are heuristic, static, and poorly adapted to stochastic disturbances.

Method: We formulate condor management as a sequential decision-making problem and apply offline reinforcement learning using a calibrated, age-structured population model.

Algorithm: Policies are learned via Fitted Q-Iteration (FQI) with Random Forest function approximation on simulated data.

Results: Learned policies achieve stable long-term populations, respond robustly to rare disasters, and minimize unnecessary interventions.

Conclusion: Extensive validation indicates that RL-derived policies are biologically plausible and suitable for decision support.

INTRODUCTION

- California condors remain critically endangered and depend on intensive human management.
- Conservation decisions are long-term, nonlinear, and subject to environmental uncertainty.
- Existing policies rely on fixed rules and expert judgment.
- We apply offline reinforcement learning to learn adaptive, cost-efficient population management strategies.

METHODOLOGY

2.1 Population Model

We model the condor population using an age-structured dynamical system with the state vector:

$$s_t = \left[N_j^t, N_s^t, N_j^t, C^t, H^t\right]$$

State includes age-structured wild population, captive stock, and habitat quality.
Nj: Juveniles (0–2 years), **Ns**: Subadults (2–6 years), - **Na**: Adults (≥6 years, breeding)
C: Captive population. **H**: Habitat quality index (baseline = 1.0)

The population dynamics are governed by an **age-structured** stochastic system incorporating survival, reproduction, lead-poisoning risk, habitat dynamics, and rare disaster events, as formalized below.

(1) Stochastic Disaster

$$\delta_t = \begin{cases} 1 - \rho_{dis}, & D_t = 1 \\ 1, & D_t = 0 \end{cases}$$

Rare disasters occur probabilistically and reduce survival of all wild birds.

(2)Effective Survival (Age-Structured)

$$s_x^t = s_x^{base} \cdot \ell_x\left(N_t, u_t^{mit}, H^t\right) \cdot \delta_t, x \in \{j, s, a\}$$

(Lead risk $\ell_x$decreases with mitigation and habitat quality.)
Survival depends on age, lead exposure, mitigation effort, habitat quality, and disasters.

(3) Reproduction (Habitat-Dependent)

$$B^t = f_{adult} N_a^t \exp\left(\kappa_{rep}(H^t - 1)\right)$$

Adult reproduction increases with habitat quality.

(4)Age Transition and Release

$$\begin{aligned} N_j^{t+1} &= B^t \\ N_s^{t+1} &= s_j^t N_j^t + s_{rel}^t u_t^{rel} \\ N_a^{t+1} &= s_s^t N_s^t + s_a^t N_a^t \end{aligned}$$

Surviving birds age forward, and released captive birds enter the subadult class.

(5)Captive Population (Logistic Growth)

$$C^{t+1} = C^t + r_C C^t\left(1 - \frac{C^t}{C_{\max}}\right) - u_t^{rel}$$

Captive population follows logistic growth and supplies releases.

(6)Habitat Recovery and Management

$$H^{t+1} = H^t + \alpha(H_0 - H^t) + \beta u_t^{mit}, H \in [H_{\min}, H_{\max}]$$

Habitat naturally recovers and is improved by mitigation efforts.

Model parameters are calibrated using a combination of published ecological studies on California condors and historical population trajectories from the real recovery program (1980–2010), by matching survival, reproduction, and long-term population trends.

2.2 Management Actions and Objective

a = [release, mitigation]

Release: number of captive-bred birds released annually
Mitigation: continuous effort level reducing lead exposure risk

$$N_{tot}^t = N_j^t + N_s^t + N_a^t$$

The objective is to maintain the total wild population within a target range while penalizing excessive intervention costs and disaster impacts. The reward function combines quadratic penalties for population deviation with linear intervention costs.

2.3 Baseline Simulations & Sanity Checks

Baseline policies: Zero-intervention, literature-inspired heuristics, and calibrated policy.
Purpose:
Verify that the environment behaves correctly under known conditions.
Establish reference points against which reinforcement learning policies could be evaluated.
Validation:
Zero-intervention policies lead to population collapse, confirming the need for active management.
Literature-based policies stabilize the population.

Baselines provide meaningful reference points for evaluating RL policies.

2.4 Model Predictive Control (MPC) Baseline.

MPC is used as a model-based baseline for comparison with RL.
Same Environment, Same Actions (Release and mitigation are chosen to minimize population deviation and intervention costs.)

Planning: At each step, MPC solves a short-horizon deterministic optimization problem.
Execution: Actions are applied in a receding-horizon manner and evaluated under stochastic dynamics.

MPC provides short-term stabilization, contrasting with RL's long-term, cost-sensitive strategies.

2.3 Reinforcement Learning Framework

Offline dataset: An offline dataset is generated by simulating random management policies.

Data coverage: Simulations span long horizons and capture both typical dynamics and rare disaster events.

Dataset size: Approximately 30,000 state–action transitions over 100-year trajectories.

Algorithm: Fitted Q-Iteration (FQI) with Random Forest function approximation.

Motivation: FQI is robust to stochasticity, suitable for long-horizon problems, and interpretable.

Policy extraction: The final policy is obtained by greedy maximization of the learned Q-function over a discretized action space.

RESULTS

The results demonstrate that the RL-derived policy achieves stable long-term population dynamics under uncertainty, while avoiding excessive intervention. Across repeated simulations, the total wild population increases smoothly over time, with occasional sharp declines caused by stochastic disaster events. These declines typically occur on the order of once every few decades, consistent with the calibrated disaster model, and are followed by gradual recovery driven by intrinsic reproduction and habitat dynamics.

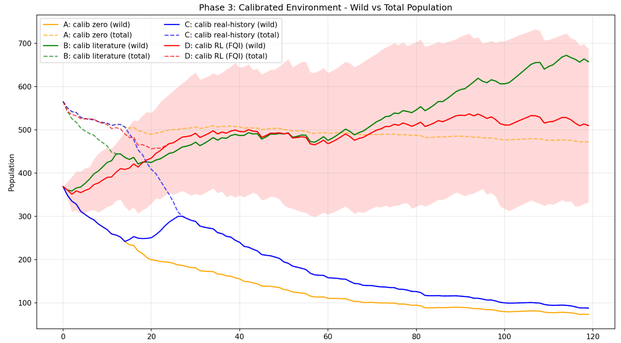


Figure 1

Figure 1. *In the calibrated environment, zero-intervention policies fail to sustain the wild population, while literature-based heuristics maintain higher population levels. The RL policy achieves long-term population stabilization and robustness under stochastic dynamics, with substantial outcome variability due to rare disaster events.Shaded region indicates stochastic outcome variability.*

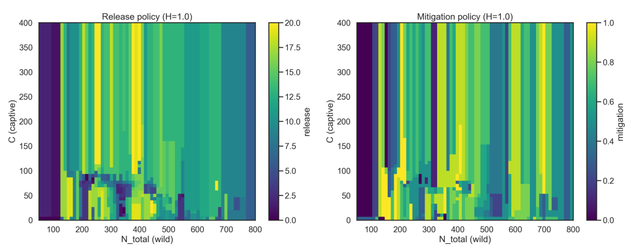


Figure 2

Figure 2. *Policy heatmaps reveal a multi-modal and cost-sensitive strategy: interventions depend jointly on wild and captive population states, rather than increasing monotonically as population declines.*

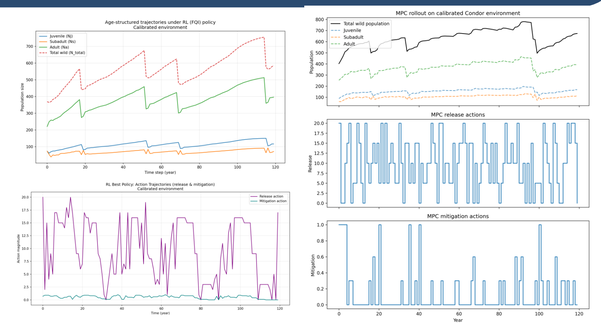


Figure 3

Figure 4

Figure 3. Under the RL policy, age-structured dynamics remain stable over long horizons, with adult populations supporting gradual recovery following rare disaster-induced shocks. The learned policy alternates between periods of active release and minimal intervention, while maintaining consistently low mitigation effort, indicating a cost-aware and adaptive control strategy.

Figure 4 shows a representative MPC rollout on the calibrated age-structured environment. MPC successfully stabilizes the wild population and prevents extinction under stochastic disturbances. However, the controller relies on frequent release actions and sparse mitigation, reflecting its short-horizon, reactive nature.

DISCUSSION.

Weak state dependence is expected: High baseline survival, reproduction, and habitat recovery place the system in a net-growth regime where aggressive intervention is rarely optimal.

Role of stochastic disasters: Rare, uncontrollable disaster events dominate large population fluctuations, encouraging cost-efficient stabilization rather than reactive over-intervention.

Policy structure: The learned policy exhibits multi-modal actions, reflecting trade-offs between intervention cost and marginal population benefit.

Limitations: Simplified lead exposure modeling, absence of spatial structure, and discontinuous policies induced by Random Forest approximation.

RL vs MPC:MPC applies more frequent short-term interventions, while RL exploits long-term ecological recovery to learn cost-sensitive, temporally extended strategies.

CONCLUSION

- Offline RL enables adaptive, long-term ecological management.
- Learned policies are stable, cost-sensitive, and biologically plausible.
- RL outperforms short-horizon MPC in exploiting long-term recovery

REFERENCES

U.S. Fish & Wildlife Service (https://www.fws.gov/)

Bruce G. Marcot , Nathan H. Schumaker , Jesse D'Elia (2024). Response of California condor populations to reintroductions, reinforcements, and reductions in spent lead ammunition pollution

Victoria J. Bakker,Myra E. Finkelstein , Daniel F. Doak , Steve Kirkland , Joseph Brandt , Alacia Welch, Rachel Wolstenholme , Joe Burnett , Arianna Punzalan , Peter Sanzenbacher (2024) Practical models to guide the transition of California condors from a conservation-reliant to a self-sustaining species

Jesse D'Elia, Nathan H. Schumaker, Bruce G. Marcot, Thomas Miewald, Sydney Watkins & Alan D. Yanahan (2022) Condors in space: an individual-based population model for California condor reintroduction planning