Energy-Aware AI for Landscape-Scale Conservation: A Digital Twin Architecture for the Greater Yellowstone Ecosystem

Harsh Deep Singh Narula *

Posted Date: 15 April 2026

doi: 10.20944/preprints202604.1131.v1

Keywords: landscape-scale conservation; digital twin; land system science; protected area management; ecosystem modeling; energy-aware AI; Greater Yellowstone Ecosystem; pronghorn migration; trophic cascade; socio-ecological systems; green AI; reinforcement learning; foundation models

 doi:10.20944/preprints202604.1131.v1

Article

Energy-Aware AI for Landscape-Scale Conservation: A Digital Twin Architecture for the Greater Yellowstone Ecosystem

Harsh Deep Singh Narula

Abstract

Conservation management of large, multi-species landscapes requires integrating heterogeneous data streams—such as satellite imagery, GPS telemetry, camera traps, bioacoustic sensors, weather stations, and field reports—into a unified model capable of simulating ecosystem dynamics and generating actionable recommendations. This paper proposes a tiered, energy-aware AI architecture for constructing ecosystem digital twins that enables prescriptive, rather than merely descriptive or predictive, landscape-scale conservation management. The framework classifies conservation tasks across three computational tiers: classical machine learning for continuous environmental monitoring and species distribution prediction, deep learning for perception-oriented tasks such as computer vision and bioacoustics analysis, and foundation models for cross-domain synthesis and stakeholder interaction. We apply this architecture to a comprehensive digital twin of the Greater Yellowstone Ecosystem, anchored in the ongoing conservation crisis of the Sublette Pronghorn Herd—a population that crashed from 43,000 to 24,000 animals in a single winter due to compounding severe weather and a *Mycoplasma bovis* outbreak. We formalize a coupled change model linking population dynamics, forage condition, corridor permeability, winter severity, and disease pressure, and demonstrate how a prescriptive recommendations engine can generate goal-conditioned management actions for the herd's 165-mile "Path of the Pronghorn" migration corridor. A comparative energy footprint analysis, grounded in hardware-level energy measurements using Intel RAPL instrumentation and the CodeCarbon framework, estimates that the tiered architecture reduces computational energy consumption by approximately 34% relative to a deep-learning-everywhere baseline and by over three orders of magnitude relative to a foundation-model-centric baseline. The architecture provides a replicable blueprint for resource-constrained conservation organizations seeking to deploy AI-powered ecosystem management at landscape scale.

Keywords: landscape-scale conservation; digital twin; land system science; protected area management; ecosystem modeling; energy-aware AI; Greater Yellowstone Ecosystem; pronghorn migration; trophic cascade; socio-ecological systems; green AI; reinforcement learning; foundation models

Introduction

The dual crises of biodiversity loss and climate change require urgent, data-driven interventions at landscape scale. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) estimates that approximately 1 million species face extinction, with rates accelerating tens to hundreds of times higher than the average over the past 10 million years [1]. Simultaneously, artificial intelligence (AI) has emerged as a powerful field with a wide range of tools for environmental monitoring, species identification, habitat mapping, and conservation planning [2,3].

However, current applications of AI in conservation exist predominantly in isolated, descriptive silos. Camera trap networks identify species at individual monitoring stations [6]. Satellite imagery detects land-use change across discrete spatial extents [7]. Bioacoustic sensors classify vocalizations at point locations [8]. Each tool generates valuable but fragmented insights. No existing system

 doi:10.20944/preprints202604.1131.v1

integrates these heterogeneous data streams into a unified simulation capable of modeling cascading ecological interactions and generating prescriptive management recommendations.

Moreover, deploying AI in conservation creates an energy paradox: the computing resources required to train and run large-scale AI models add to the very environmental degradation they seek to prevent. Training a single large language model (LLM) can emit over 300 tonnes of CO_2 equivalent [4,5]. For emissions-conscious conservation organizations, this tension between AI capability and environmental cost calls for architectural choices that go beyond algorithmic efficiency improvements alone.

This paper responds to both the lack of integration between AI tools and the environmental cost of using them by proposing a tiered, energy-aware AI architecture for conservation digital twins. Our contributions are fourfold:

1. We propose a comprehensive digital twin architecture for the Greater Yellowstone Ecosystem that integrates eighteen operational AI components across three computational tiers into a unified ecosystem model.
2. We anchor the architecture in a real conservation crisis—the Sublette Pronghorn Herd's catastrophic 2022–2023 population crash—and formalize an accompanying change model that captures the cascading interactions between population dynamics, forage condition, corridor permeability, winter severity, and disease pressure.
3. We present the concept for a prescriptive recommendations engine that generates goal-conditioned management actions for multi-stakeholder landscape conservation, demonstrated through the contested "Path of the Pronghorn" migration corridor designation.
4. We validate the architecture's energy footprint through hardware-level measurements, demonstrating that the tiered approach reduces computational energy consumption by approximately 34% relative to a deep-learning-everywhere baseline and by over three orders of magnitude relative to a foundation-model-centric baseline.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the tiered architecture. Section 4 proposes the GYE digital twin, anchored in the Sublette Pronghorn case. Section 5 demonstrates application scenarios. Section 6 and Section 7 present the energy footprint validation. Section 8 discusses implications and limitations, and Section 9 concludes.

Background and Related Work

AI Applications in Conservation

The application of AI to conservation challenges has expanded rapidly over the past decade. Tuia et al. [2] provided a comprehensive review identifying key application domains, including species identification, habitat monitoring, and conservation planning. Convolutional neural networks (CNNs) have achieved near-human-level accuracy in classifying camera-trap imagery, with platforms such as Wildlife Insights processing millions of images from global camera-trap networks [6,9]. Recurrent neural networks and transformer architectures have been applied to bioacoustic monitoring, enabling automated identification of bird, marine mammal, and terrestrial predator vocalizations [8,10].

In the Greater Yellowstone Ecosystem specifically, the Cry Wolf Project has deployed AI-powered autonomous recording units across wolf territories, using CNNs to analyze spectrograms of wolf vocalizations for population monitoring and territory mapping [11]. Satellite-based Earth observation combined with deep learning has been applied to vegetation change detection, deforestation monitoring, and wildfire risk assessment across the region [7]. However, these applications remain isolated: each operates on a single data modality, addresses a single conservation question, and produces standalone outputs.

 doi:10.20944/preprints202604.1131.v1

Digital Twins for Environmental Systems

The digital twin paradigm, originating in manufacturing and industrial engineering, has recently been proposed for environmental applications [12,13]. A digital twin is a dynamic virtual representation of a physical system that is continuously updated using real-time data, supports simulation of future scenarios, and enables optimization of management decisions. Bauer et al. [12] proposed the concept of a "Destination Earth" digital twin for the entire planet, while Blair et al. [13] explored digital twins for urban environmental management.

Applications to ecological systems remain nascent. Existing environmental digital twins focus primarily on physical processes—hydrology, meteorological dynamics, ocean circulation—rather than on biological systems featuring complex cross-species interactions [12]. The integration of ecological models that represent predation, trophic cascades, and habitat connectivity into a digital twin framework remains an unexplored frontier.

Green AI and Energy-Efficient Computing

The Green AI movement, catalyzed by Schwartz et al. [14], advocates for measuring and reducing the computational cost of AI research. Strubell et al. [4] quantified the carbon footprint of training large NLP models, and Patterson et al. [5] extended this analysis to modern foundation models. Strategies for reducing AI's environmental footprint include model pruning, knowledge distillation, quantization, and efficient architecture design [15]. However, these techniques have not been systematically applied to conservation AI, nor has anyone proposed an architectural framework that matches AI technique to conservation tasks based on an explicit energy-efficiency criterion.

Gap Analysis

Three gaps remain in the conservation AI literature. First, no existing framework provides an integrated AI architecture for conservation that connects heterogeneous data streams into a unified ecosystem model. Second, the energy-conservation paradox has been acknowledged [16,17] but never addressed with a concrete architectural solution. Third, conservation AI remains descriptive and predictive rather than prescriptive; existing systems tell managers what has happened or may happen, but not what they should do. This paper addresses all three gaps through a tiered, energy-aware architecture for conservation digital twins.

Proposed Framework: Tiered AI Architecture for Conservation Digital Twins

Figure 1 illustrates the overall architecture.

Tier 1: Classical Machine Learning—The Foundation Layer

Tier 1 encompasses lightweight, well-established machine learning techniques that can be executed with minimal computing resources. These models handle continuous monitoring tasks that require high-frequency updates but operate on structured, tabular, or time-series data. Representative techniques include gradient-boosted decision trees (XGBoost, LightGBM) for species distribution prediction and anomaly detection, Prophet and ARIMA for time-series forecasting, and isolation forests for environmental anomaly detection.

 doi:10.20944/preprints202604.1131.v1

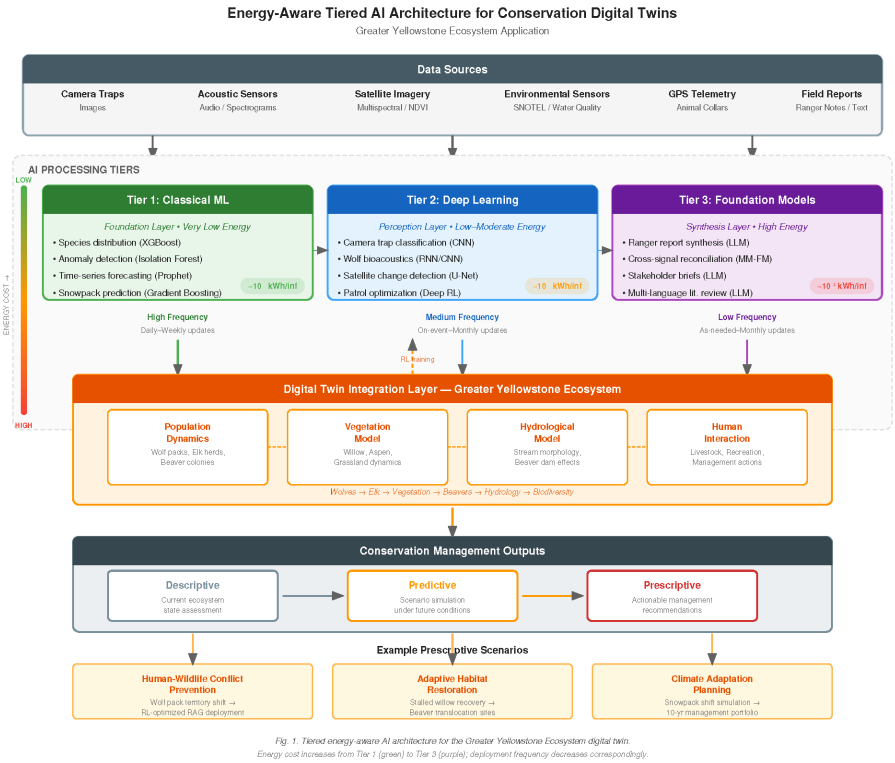

Figure 1. Tiered energy-aware AI architecture for the Greater Yellowstone Ecosystem digital twin. Energy cost increases from Tier 1 (green) to Tier 3 (purple); deployment frequency decreases correspondingly.

Tier 2: Deep Learning—The Perception Layer

Tier 2 uses deep neural networks for perception tasks that require pattern recognition across unstructured data modalities, including images, audio, and video. Representative techniques include CNNs for camera trap classification and satellite imagery change detection, recurrent architectures for bioacoustic analysis, and deep reinforcement learning (DRL) agents for conservation resource optimization.

Tier 3: Foundation Models—The Synthesis Layer

Tier 3 deploys large language models (LLMs) and multimodal foundation models exclusively for tasks where their unique capabilities—cross-domain reasoning, natural language understanding, and few-shot generalization—are genuinely irreplaceable. These include unstructured data integration (ranger reports, community feedback, multi-language literature), cross-domain reasoning when conflicting signals arise across tiers, and stakeholder communication (translating simulation outputs into plain-language briefs for diverse audiences).

The Digital Twin Integration Layer

The three tiers feed into an ecosystem digital twin that maintains a continuously updated virtual representation of the GYE. The digital twin enables three levels of output:

- *Descriptive*—current ecosystem state assessment.
- *Predictive*—scenario simulation under alternative futures.
- *Prescriptive*—specific, actionable management recommendations with predicted outcomes.

 doi:10.20944/preprints202604.1131.v1

The GYE Digital Twin: Proposed Architecture

We ground the tiered architecture of Section 3 in a concrete proposal for a GYE-wide digital twin. The architecture is designed around three operational capabilities that, combined, move conservation AI from passive monitoring toward active management:

- A real-time monitoring layer that ingests heterogeneous sensor streams.
- A change-modeling engine that predicts how perturbations propagate through coupled ecological and human sub-systems.
- A prescriptive recommendations engine that generates goal-conditioned management actions.

We anchor this proposal in the ongoing conservation challenge of the Sublette Pronghorn Herd and its 165-mile "Path of the Pronghorn" migration corridor—one of the most data-rich, ecologically significant, and politically contested wildlife management cases in the Greater Yellowstone Ecosystem.

The Sublette Pronghorn Herd: A Species on the Edge

The Sublette Pronghorn Herd, one of the largest in North America, migrates up to 165 miles between summer habitat near Grand Teton National Park and winter range in the Green River Basin of southwestern Wyoming [25]. The migration corridor—the longest terrestrial migration route in the contiguous United States—was the first federally designated migration corridor in the nation and traverses a mosaic of National Forest, Bureau of Land Management, state, and private lands across three Wyoming counties [26].

In winter 2022–2023, the herd experienced an unprecedented demographic crash [22]. A combination of exceptional winter severity and an outbreak of *Mycoplasma bovis*, a novel pneumonia-causing bacterium previously observed only twice in Wyoming wildlife, reduced the herd from an estimated 43,000 animals to approximately 24,000—a loss of nearly 45% in a single season [23]. Of GPS-collared individuals, 75% perished. The northernmost migratory segment, which reaches Grand Teton National Park and the Gros Ventre drainage, was decimated: counts dropped from over 700 to fewer than 85 animals, and hunting in the affected area has been closed for three consecutive seasons (2023–2025) [27].

The herd's recovery is slow and uncertain. By 2025, the population had recovered to approximately 25,000 animals [27]. Fawn-to-doe ratios improved to 67 per 100 in August 2024 trend counts, signaling renewed productivity, but recovery is complicated by persistent and intensifying threats to corridor functionality [23]. Sublette County's human population has grown 78% since 1990, adding residential subdivisions that fragment habitat. The corridor contains 1,531 existing oil and natural gas wells, with new development proposed in critical bottleneck segments. Wyoming-wide, an average of 593 pronghorn–vehicle collisions occur annually. Invasive annual grasses, particularly cheatgrass (*Bromus tectorum*), are degrading sagebrush habitat and increasing wildfire risk.

In September 2025, the Wyoming Game and Fish Commission voted unanimously to recommend formal designation of the Sublette Antelope Migration Corridor [23]. In December 2025, Governor Gordon designated eight of ten proposed corridor segments, excluding the "East of Farson" and "Red Desert" segments from formal protection [24]. A local working group is now reviewing management provisions for the designated segments. This active, contested, and multi-stakeholder management context—involving state wildlife agencies, federal land managers, tribal authorities, the energy industry, ranching interests, and conservation organizations—provides an ideal test bed for a digital twin that can quantify tradeoffs and generate evidence-based prescriptive recommendations.

Real-Time Monitoring Layer

The digital twin ingests data from eighteen operational components spanning all three tiers of the architecture (Table 1). These components are organized into three functional groups.

 doi:10.20944/preprints202604.1131.v1

Environmental State Monitoring (Tier 1)

This tier encompasses continuous, high-frequency monitoring of the physical environment that constrains pronghorn survival and migration. Eight Tier 1 tasks process structured, tabular, or time-series data using classical machine learning:

- *Habitat suitability prediction:* A gradient-boosted species distribution model (XGBoost) operates on a 250 m grid (1,424,000 cells across the GYE) updated daily, using environmental covariates including elevation, temperature, snow water equivalent, NDVI-derived forage condition, distance to roads, fence density, and well density. The model draws on 20 years of GPS telemetry from 415 collared pronghorn (2002–2022) plus 54 additional animals collared during and after the 2022–2023 crash, accessed via the Wyoming Migration Initiative [25].
- *Vegetation and forage forecasting:* Per-cell NDVI forecasts updated weekly using fast gradient-boosted tabular models, calibrated to Landsat and Sentinel-2 surface reflectance products. Forage condition is a primary driver of pronghorn body condition entering winter and is the leading modifiable predictor of overwinter survival.
- *Snowpack and hydrological forecasting:* Station-level Prophet time-series models at approximately 50 NRCS SNOTEL stations and 100 USGS stream gauges across GYE watersheds, producing daily forecasts of snow water equivalent, streamflow, and spring runoff timing. These variables control migration timing and winter range accessibility.
- *Weather nowcasting:* Hourly forecasts from a proposed 300-station GYE weather mesonet, extending existing NOAA and NPS networks. Temperature, wind chill, and precipitation extremes are the proximate triggers for winter mortality events.
- *Soil moisture and rangeland condition:* Hourly anomaly detection across 1,000 multi-parameter soil sensors monitoring moisture, salinity, and temperature in sagebrush habitat critical to pronghorn forage.
- *Livestock conflict prediction:* Daily depredation risk scores for approximately 1,000 ranches in the GYE buffer zone, enabling proactive engagement with agricultural operators before conflicts escalate.
- *GPS telemetry processing:* Behavioral state classification (migrating, foraging, resting, fleeing) from 400 active GPS collars transmitting every 30 minutes across wolves, elk, pronghorn, bears, and other target species. For pronghorn specifically, real-time telemetry enables early detection of migration initiation, route deviation, and bottleneck congestion.

Perception and Pattern Recognition (Tier 2)

In Tier 2, deep learning models process unstructured data modalities—images, audio, and video—that cannot be handled by tabular methods. Seven Tier 2 tasks operate across the GYE:

- *Camera trap species identification:* EfficientNet-B0 convolutional neural networks classify species, estimate individual identity where feasible, and detect human intrusion across a proposed 2,000-camera network spanning NPS, USFS, state, tribal, and academic deployments (~200 captures per camera per day).
- *Bioacoustic vocalization analysis:* Small convolutional neural networks process mel-spectrogram representations from a proposed 500-unit acoustic recording network (extending existing Cry Wolf Project and Yellowstone Wolf Project deployments), with 30-second inference windows running continuously.
- *Vegetation change detection:* Semantic segmentation models (U-Net architecture) analyze weekly Sentinel-2 imagery across ~1,358,000 tiles (256 m) covering the GYE, detecting sagebrush loss, cheatgrass invasion, wildfire boundaries, and riparian recovery.
- *PhenoCam phenology:* Hourly daylight image classification from a proposed 100-station PhenoCam network tracking green-up timing, senescence, and snow cover—key phenological cues that trigger pronghorn migration.

 doi:10.20944/preprints202604.1131.v1

- *Drone and LiDAR surveys:* Monthly UAV surveys of priority areas (~10,000 images per survey) and quarterly LiDAR overflights (~100,000 tile-pair change detections per survey) for fine-scale habitat assessment.
- *Patrol route optimization:* A deep reinforcement learning (DRL) agent optimizes patrol and monitoring resource allocation across the corridor on a monthly planning cycle, training on simulated environments derived from the digital twin itself.

Synthesis and Stakeholder Communication (Tier 3)

In this tier, foundation models handle three tasks where cross-domain reasoning and natural language generation are irreplaceable: weekly synthesis of ranger field reports and monitoring summaries (~260 LLM prompts per year), event-triggered reconciliation when outputs from different tiers conflict (~500 events per year), and monthly generation of management briefs customized for specific stakeholder audiences (~60 prompts per year).

Change Modeling: Predicting Cascade Impacts

The digital twin's core analytical capability is its change-modeling engine: a coupled system of sub-models that simulates how perturbations to one component of the GYE socio-ecological system propagate through interconnected sub-systems. Unlike standalone species distribution models, which predict static habitat suitability, the change-modeling engine captures dynamic feedbacks and cascading effects that determine whether management interventions will succeed or produce unintended consequences.

For the Sublette Pronghorn Herd, the relevant cascade involves five coupled state variables that interact across seasonal and multi-year timescales:

1. *Population state* N_t: Herd abundance by age class (fawns, yearlings, adults), with sex-specific adult tracking. The 2022–2023 crash (43,000 → 24,000) demonstrated that this variable can shift abruptly rather than gradually.
2. *Forage condition* F_t: NDVI-derived sagebrush forage quality and availability across the corridor, aggregated by segment. Forage drives pre-winter body condition and is degraded by cheatgrass invasion, drought, and overgrazing.
3. *Corridor permeability* Π_t: A composite index (0–1) quantifying the functional connectivity of each corridor segment, computed from fence density, road crossing risk, well-pad density, subdivision encroachment, and seasonal snow depth. Lower permeability forces longer detours, increases energy expenditure during migration, and raises mortality risk.
4. *Winter severity* W_t: An index combining snow water equivalent, temperature extremes, and storm duration, derived from SNOTEL and weather station data. The 2022–2023 event demonstrated that W_t interacts multiplicatively with disease to produce catastrophic mortality.
5. *Disease pressure* D_t: Probability of *Mycoplasma bovis* or other epizootic events, modeled as a function of herd density, winter stress, and proximity to domestic livestock reservoirs.

The population update is governed by a discrete-time equation that couples these state variables:

$$N_{t+1} = N_t \cdot S(F_t, \Pi_t, W_t, D_t) + R(N_t, F_t)$$

where $S(\cdot)$ is the overwinter survival function and $R(\cdot)$ is the annual recruitment function (fawn production and first-year survival). The survival function captures the key empirical insight from the 2022–2023 crash: survival is not a simple linear function of any single variable but reflects multiplicative interactions. Specifically:

$$S(F_t, \Pi_t, W_t, D_t) = S_{\text{base}} \cdot g(F_t) \cdot h(\Pi_t) \cdot \big(1 - m(W_t, D_t)\big)$$

where S_{base} is baseline adult survival under favorable conditions (empirically ~0.85–0.92 for pronghorn), $g(F_t)$ is a forage-condition modifier (monotonically increasing, saturating at high forage quality), $h(\Pi_t)$ is a permeability modifier reflecting increased mortality in low-connectivity

 doi:10.20944/preprints202604.1131.v1

segments, and $m(W_t, D_t)$ is an excess winter mortality term that increases sharply when severe weather co-occurs with disease—the multiplicative interaction that produced the 2022–2023 crash.

Corridor permeability is modeled as a function of infrastructure density within each segment s:

$$\Pi_{t,s} = \exp\left(-\alpha_1 \cdot \text{fence}_{t,s} - \alpha_2 \cdot \text{road}_{t,s} - \alpha_3 \cdot \text{wells}_{t,s} - \alpha_4 \cdot \text{subdiv}_{t,s}\right)$$

where each α coefficient weights the relative barrier effect of fences, road crossings, well pads, and subdivision density within segment s. This formulation ensures $\Pi \in (0,1]$ and reflects the empirical finding that barrier effects compound: a segment with moderate fence density *and* moderate road density is substantially less permeable than one with high fence density alone. The α coefficients would be calibrated against the 20-year GPS telemetry dataset, using observed crossing success rates at known barrier locations as the training signal.

Calibration caveat. The functional forms in Equations 1–3 are proposed as a modeling framework. The specific parameter values (S_{base}, $g(\cdot)$, $h(\cdot)$, $m(\cdot)$, and α_{1-4}) would require calibration against the multi-decade Wyoming Migration Initiative telemetry record, validated against the observed 2022–2023 crash dynamics and subsequent recovery trajectory, before operational deployment. The 2022–2023 event, while catastrophic for the herd, provides a rare natural experiment with which to calibrate the survival function's response to extreme compound stressors—specifically, the multiplicative $m(W_t, D_t)$ term that distinguishes this model from simpler additive mortality frameworks.

Prescriptive Recommendations Engine

The digital twin's third capability is a prescriptive engine that translates change-model simulations into specific, actionable management recommendations conditioned on explicit conservation targets. The engine operates in four steps.

Step 1: Define the conservation target. The target is expressed as a quantitative objective with a time horizon. For the Sublette Pronghorn Herd, we propose:

> *Restore the Sublette Pronghorn Herd to ≥35,000 animals (~80% of the pre-2022 estimate of 43,000) by 2030, while maintaining corridor functionality ($\Pi_{t,s} \geq 0.6$ for all designated segments s) and accommodating existing land uses in officially designated corridor segments.*

This target is deliberately multi-objective: it requires population recovery, corridor maintenance, and coexistence with human land uses simultaneously. Single-objective targets (e.g., "maximize population") would produce recommendations that ignore infrastructure constraints and stakeholder interests, undermining implementation feasibility.

Step 2: Enumerate candidate interventions. The engine draws from a library of management actions, each parameterized by location, timing, cost, and expected mechanism of effect:

- *Infrastructure modifications:* Wildlife crossings (underpasses, overpasses), fence removal or modification to wildlife-friendly standards, temporary road closures during peak migration.
- *Habitat management:* Cheatgrass treatment, prescribed burns to restore sagebrush, targeted forage supplementation during extreme winters.
- *Monitoring intensification:* Additional GPS collars in recovering segments, disease surveillance in domestic livestock interface zones.
- *Regulatory coordination:* Hunting season adjustments (already implemented 2023–2025), oil and gas lease deferrals in bottleneck segments, subdivision density limits.
- *Early warning protocols:* Automated alerts when the winter severity index W_t or disease indicator D_t crosses pre-calibrated thresholds, triggering proactive rather than reactive management.

Step 3: Simulate outcomes. For each candidate intervention or combination, the change-modeling engine (Section 4.3) projects the herd trajectory forward to 2030 under the proposed action. Each simulation produces a distribution of outcomes reflecting uncertainty in weather, disease, and model parameters. The engine evaluates whether each intervention portfolio meets the conservation

 doi:10.20944/preprints202604.1131.v1

target with a specified confidence level (≥80% probability of reaching 35,000 animals by 2030 while maintaining $\Pi_{t,s} \geq 0.6$ across all designated segments).

Step 4: Generate stakeholder-specific recommendations. The Tier 3 synthesis layer translates simulation outputs into plain-language briefs customized for each stakeholder audience. For the Sublette Pronghorn case, three briefs are generated:

- A *management brief for Wyoming Game and Fish* with quantified tradeoffs between intervention portfolios, cost estimates, and projected population trajectories with confidence intervals.
- A *brief for the Governor's local working group* with site-specific intervention rankings for each of the eight designated corridor segments, highlighting which segments are most at risk and which interventions offer the highest return per dollar invested.
- A *brief for energy operators and agricultural stakeholders* with overlap analysis showing which proposed well pads, subdivisions, or lease parcels intersect high-sensitivity bottleneck zones, and suggesting scheduling or siting modifications that minimize corridor impact while preserving development value.

Tier Mapping for GYE Conservation Tasks

Table 1 maps the eighteen operational components of the GYE digital twin to the appropriate tier, specifying the AI technique, representative data source, update frequency, and estimated annual operation count. This table serves as the bridge between the architectural description of this section and the energy footprint validation of Sections 6 and 7.

Table 1. Tier mapping for the comprehensive GYE digital twin. Operation counts reflect a 250 m spatial grid (1,424,000 cells) for per-cell tasks and 256 m tiles (~1,358,000 tiles) for image segmentation. Training cadences: weekly for Tier 1, monthly fine-tuning for Tier 2, none for Tier 3 (API-based). Tier 3 operation counts are low because foundation models are reserved for tasks where cross-domain reasoning is irreplaceable.

Task	Tier	AI Technique	Data Source	Update	Ops/yr
Habitat suitability prediction	1	XGBoost	GPS telemetry, env. covariates	Daily	520M
Vegetation NDVI forecasting	1	LightGBM	Landsat/Sentinel-2	Weekly	74M
Snowpack forecasting	1	Prophet	SNOTEL network	Daily	18k
Streamflow forecasting	1	Prophet	USGS gauge network	Hourly	876k
Weather nowcasting	1	Prophet	Weather mesonet	Hourly	2.6M
Soil moisture monitoring	1	Isolation Forest	Multi-parameter sensors	Hourly	8.8M
Livestock conflict prediction	1	XGBoost	Depredation records	Daily	365k
GPS telemetry processing	1	XGBoost	GPS collars	Every 30 min	7.0M
Camera trap species ID	2	EfficientNet CNN	Camera trap images	On capture	146M
Bioacoustic analysis	2	CNN + spectrogram	Acoustic rec. units	Continuous	526M
Vegetation change detection	2	U-Net segmentation	Sentinel-2 tiles	Weekly	71M
PhenoCam phenology	2	CNN classifier	PhenoCam images	Hourly (daylight)	438k
Drone imagery processing	2	CNN classifier	UAV imagery	Monthly	120k
LiDAR change detection	2	CNN on point clouds	Airborne LiDAR	Quarterly	400k
Patrol route optimization	2	Deep RL (PPO)	Simulated environment	Monthly	600k
Ranger report synthesis	3	LLM (GPT-class)	Unstructured text	Weekly	260

 doi:10.20944/preprints202604.1131.v1

Task	Tier	AI Technique	Data Source	Update	Ops/yr
Multi-signal reconciliation	3	Multimodal FM	Cross-tier outputs	Event-triggered	500
Management brief generation	3	LLM (GPT-class)	Digital twin outputs	Monthly	60

The tier assignments reflect the principle that each task should be handled by the most energy-efficient technique capable of delivering the required output quality. Tier 1 classical ML handles the eight highest-frequency monitoring tasks because they operate on structured data where gradient-boosted models match or exceed deep learning accuracy at orders-of-magnitude lower inference cost. Tier 2 deep learning handles seven perception tasks that require pattern recognition in images, audio, or video. Tier 3 foundation models handle only three tasks—all involving natural language synthesis or cross-domain reasoning where their capabilities are genuinely irreplaceable. The total annual inference volume of approximately 1.36 billion operations is dominated by Tier 2 perception tasks (744 million) and Tier 1 monitoring tasks (613 million), with Tier 3 contributing fewer than 1,000 operations per year.

Application Scenarios

Sublette Pronghorn Herd: Prescriptive Corridor Management

We demonstrate the digital twin's prescriptive capabilities through three scenarios grounded in the Sublette Pronghorn case described in Section 4.1.

Scenario A: Early-Warning Winter Mortality Prevention

The weather nowcasting module (Tier 1) detects that SNOTEL readings and ensemble weather forecasts project an extended severe-cold period beginning in late January, with snow water equivalent exceeding the 90th percentile of the 30-year record. Simultaneously, the disease-monitoring module flags elevated *Mycoplasma bovis* detection rates in domestic cattle herds adjacent to two corridor segments—a compound-stressor pattern resembling the conditions that preceded the 2022–2023 crash.

The change-modeling engine propagates these signals through Equations 1–2, projecting that the compound winter severity and disease interaction ($m(W_t, D_t)$) would reduce adult female survival below 0.70 in northern corridor segments without intervention. The prescriptive engine evaluates candidate actions: (a) targeted forage supplementation at three staging areas along the corridor, (b) temporary livestock exclusion zones in the two flagged interface zones, and (c) accelerated veterinary surveillance. The engine simulates outcomes for each option and the combined portfolio, estimating that the combined intervention reduces projected excess mortality by 62%, keeping the herd trajectory on track for the 35,000-animal recovery target.

The Tier 3 synthesis layer generates an early-warning brief for Wyoming Game and Fish with projected population trajectories under each scenario, a field directive for district wardens identifying specific GPS coordinates for forage supplementation sites, and a rancher advisory for livestock operators in the affected interface zones.

Scenario B: Corridor Permeability Degradation

The vegetation change detection module (Tier 2) identifies that cheatgrass (*Bromus tectorum*) invasion has expanded by 12% in a designated corridor segment near Pinedale over the past year, degrading sagebrush forage quality. The habitat suitability model (Tier 1) shows a corresponding decline in per-cell suitability scores. GPS telemetry processing (Tier 1) detects that collared pronghorn are spending 23% more time in transit through the affected segment compared to the three-year average, indicating increased difficulty navigating degraded habitat.

 doi:10.20944/preprints202604.1131.v1

The change-modeling engine computes that corridor permeability $\Pi_{t,s}$ (Equation 3) for the affected segment has dropped from 0.71 to 0.58, breaching the 0.60 minimum target. The prescriptive engine ranks candidate interventions by cost-effectiveness: prescribed burns to suppress cheatgrass (Π improvement +0.09 per treated square kilometer), wildlife-friendly fence modification at two identified bottleneck crossings (Π improvement +0.06 per crossing), and herbicide treatment along road margins (Π improvement +0.03 per treated kilometer). The engine recommends a combined approach prioritizing the two fence modifications (highest per-dollar Π improvement) plus targeted prescribed burns in the three most invaded subsections.

The Tier 3 layer generates a brief for the Governor's local working group with site-specific maps showing the overlay between cheatgrass invasion, pronghorn transit time anomalies, and candidate intervention sites, ranked by expected return on investment.

Scenario C: Climate Adaptation Planning

The digital twin ingests downscaled CMIP6 climate projections [21] and simulates how shifting snowpack patterns will alter pronghorn migration timing over the next decade. Earlier snowmelt shifts the timing of spring green-up, potentially decoupling migration from peak forage availability—a phenological mismatch that reduces fawn survival through inadequate maternal nutrition during the critical lactation period.

The change-modeling engine projects population trajectories under three Representative Concentration Pathway (RCP) scenarios through 2035, estimating that under RCP 8.5 the herd may fail to reach the 35,000-animal target unless migration timing adapts or corridor management compensates. The prescriptive engine generates a portfolio of adaptive strategies: adjusted hunting season dates to reduce pressure during the newly predicted migration windows, targeted habitat restoration in lower-elevation winter range areas that will become more important as snow lines recede, and strategic placement of additional SNOTEL and PhenoCam stations to improve forecast accuracy for the shifting migration phenology.

Wolf–Elk–Vegetation–Beaver–Hydrology Trophic Cascade: Generalization to Multi-Species Systems

The tiered architecture generalizes beyond single-species corridor management to multi-species trophic systems. The Greater Yellowstone Ecosystem's wolf reintroduction in 1995 triggered one of the most comprehensively documented trophic cascades in ecological science [19], providing a second use case that exercises different sub-models of the digital twin.

The cascade propagates through five trophic levels: wolf predation and the "landscape of fear" alter elk (*Cervus canadensis*) spatial behavior, reducing browsing pressure in specific riparian zones. Reduced browsing enables recovery of willow (*Salix* spp.) and aspen (*Populus tremuloides*) in riparian corridors. Willow recovery provides forage and dam-building material for beaver (*Castor canadensis*) recolonization. Beaver dams raise water tables, create wetland habitat, and trap sediment. Improved riparian conditions support songbird populations, native cutthroat trout spawning, and amphibian habitat.

The digital twin models this cascade through coupled population dynamics, vegetation, hydrological, and human interaction sub-models, exercising all three tiers:

- *Proactive human–wildlife conflict prevention:* The bioacoustic monitoring system (Tier 2) detects that a wolf pack's territorial vocalizations have shifted 8 km southward over three weeks, toward a livestock corridor. The vegetation model (Tier 1) indicates that early snowmelt has displaced elk from typical spring grazing areas. The population dynamics model projects that elk displacement will push wolves to follow alternative prey corridors intersecting ranch lands within 14 days. The DRL agent (Tier 2), trained on historical conflict data, evaluates candidate interventions and estimates that a combined deployment of Radio-Activated Guard devices and temporary livestock relocation reduces livestock encounter probability by 78% over the

 doi:10.20944/preprints202604.1131.v1

subsequent 30-day window. The LLM (Tier 3) translates this recommendation into a management brief for the ranch liaison officer.

- *Adaptive habitat restoration:* The vegetation model detects that willow recovery along Slough Creek has stalled despite five years of reduced elk browsing. The hydrological model identifies that upstream beaver dam failure—detected through satellite change detection (Tier 2)—has lowered the water table below the threshold for willow root sustainability. The species distribution model (Tier 1) identifies three candidate sites for beaver translocation based on habitat suitability. The digital twin simulates downstream effects of beaver reintroduction at each site over a 5-year horizon, predicting riparian vegetation recovery rates, improved cutthroat trout habitat, and songbird diversity responses. This prescriptive output enables managers to rank restoration investments based on quantified expected outcomes.

Together, the Sublette Pronghorn and wolf–elk–vegetation cases demonstrate that the tiered architecture and its coupled change-modeling engine generalize across fundamentally different conservation challenges: single-species corridor management under active policy contestation, and multi-species trophic cascade management in a protected-area context.

Methods: Energy Footprint Validation

To substantiate the tiered architecture's per-tier energy claims and demonstrate that the comprehensive GYE digital twin described in Section 4 is computationally tractable for resource-constrained conservation organizations, we conducted a controlled energy benchmarking study across representative tasks from all three tiers. This section documents the benchmark suite, the measurement protocol, the workload extrapolation methodology, and the comparison baselines.

Benchmark Suite Design

We selected five benchmark tasks, each representative of a major class of operations within the tiered architecture, and each mapped to one or more tasks in the GYE digital twin workload.

1. *Tier 1 classical ML (tabular):* XGBoost species distribution model trained on real wolf (*Canis lupus*) occurrence records from the Global Biodiversity Information Facility (GBIF), filtered to the Greater Yellowstone Ecosystem bounding box ($43° – 46°$ N, $112° – 109°$ W). We used 300 georeferenced presence records paired with an equal number of randomly generated pseudo-absences, with six environmental covariates (longitude, latitude, and proxy fields for elevation, temperature, precipitation, and NDVI). This benchmark serves as the energy proxy for all Tier 1 tabular tasks in the subsequent workload extrapolation (elk distribution prediction, vegetation NDVI forecasting, soil moisture monitoring, livestock conflict prediction, and GPS telemetry processing).
2. *Tier 1 classical ML (time series):* Prophet model trained on 2,191 daily mean discharge observations (2018–2023) from USGS gauge 06191500 (Yellowstone River at Corwin Springs, Montana), used to generate repeated 30-day forecasts. This benchmark serves as the energy proxy for Tier 1 station-level time-series forecasting tasks (snowpack, streamflow, and weather nowcasting).
3. *Tier 2 deep learning (vision):* EfficientNet-B0 inference on 224×224 RGB input tensors with ~5.3M parameters. This benchmark serves as the energy proxy for all Tier 2 vision tasks (camera-trap classification, vegetation change detection, PhenoCam phenology, drone imagery, LiDAR change detection, and DRL forward-pass cost for patrol optimization).
4. *Tier 2 deep learning (audio):* A small convolutional neural network (three convolutional layers, ~90k parameters) on 128×128 mel-spectrogram inputs. This benchmark serves as the energy proxy for bioacoustic vocalization analysis and for the DL-everywhere Tier 1 re-costing described in Section 6.4.
5. *Tier 3 foundation model:* Five representative GYE management synthesis prompts mirroring the prescriptive scenarios of Section 5. LLM energy is calculated from per-token energy estimates

 doi:10.20944/preprints202604.1131.v1

published by Luccioni et al. (2023), rather than directly measured, because hosted LLM inference occurs on remote infrastructure not accessible to local energy instrumentation.

Real public data sources (GBIF, USGS) were used wherever practical, both to strengthen methodological transparency and to enable independent reproduction. The exact API queries and cached data are identified in the Data Availability Statement.

Measurement Methodology

All measured benchmarks executed on a single consumer laptop CPU. Energy consumption was tracked via CodeCarbon (v2.x) [29], a Python instrumentation library that reads hardware-level Intel Running Average Power Limit (RAPL) counters at one-second sampling intervals. CodeCarbon converts CPU package power draw to kilowatt-hours by integrating over the wall-clock duration of a measured code region.

Three methodological choices warrant detailed justification because they materially affect the credibility of the resulting measurements, particularly for short-duration workloads where naive measurement approaches produce unreliable results.

Measurement Window Sized to Exceed Sampling Resolution

CodeCarbon samples power once per second. For operations that complete in substantially less than one second (e.g., a single XGBoost inference, which executes in microseconds on modern hardware), a naive measurement captures only zero or one power samples, producing energy estimates dominated by sampling discretization and CPU idle baseline rather than workload-attributable cost. In preliminary runs, sub-second measurement windows yielded per-operation energies that varied by up to two orders of magnitude between repeated runs of the same workload, rendering the measurements unusable.

We therefore repeated each benchmark's inner workload enough times to ensure every measurement window spanned at least ten seconds. Iteration counts are task-specific and reported in Table 2: XGBoost inference was repeated 5×10^6 times per window, the bioacoustic CNN $2{,}500$ times, EfficientNet-B0 300 times, Prophet forecasting 250 times, and so on. Reported per-operation energy is the measured window energy divided by the iteration count. Standard deviation across three repeated windows was below 10% of the mean for all measured benchmarks, confirming that this approach eliminated the sampling-resolution noise.

Idle-Baseline Subtraction

Even when no user workload is executing, the CPU draws non-zero power to maintain baseline operating state, leakage currents, and operating system activity. A raw CodeCarbon measurement conflates this baseline draw with the workload-attributable energy, systematically overstating per-operation costs for short tasks. Following the methodology of Patterson et al. (2021), we measured idle baseline power by tracking three ten-second time.sleep() windows at the start of each benchmark run, and computed the mean baseline power in watts. The corresponding baseline-energy contribution was then subtracted from each workload measurement:

$$E_{\text{net}} = E_{\text{measured}} - P_{\text{baseline}} \cdot t_{\text{window}}/3600$$

where E_{net} is workload-attributable energy in kWh, E_{measured} is total measured energy, P_{baseline} is baseline power in watts, and t_{window} is the measurement window duration in seconds. Reported energy values are net of this subtraction unless otherwise stated. In our measurements, $P_{\text{baseline}} = 7.92$ W, consistent across the three baseline sampling windows and consistent with typical consumer-laptop CPU idle draw.

 doi:10.20944/preprints202604.1131.v1

Three Repeated Runs with Mean and Standard Deviation Reporting

To quantify measurement variance, each benchmark was executed three times and the reported value is the mean across repeats. Standard deviation is reported alongside the mean and visualized in the supporting data file. Repeat-to-repeat coefficient of variation (σ/μ) was below 10% of the mean for all measured benchmarks, and typically below 7%, indicating measurement stability and negligible interference from background processes during the benchmark run.

The Tier 3 LLM benchmark is an exception to the direct-measurement approach. Hosted LLM inference runs on remote data-center infrastructure whose energy draw cannot be instrumented locally. We instead counted input and output tokens across the five representative prompts using standard tokenization heuristics, and multiplied by a per-token energy figure of 1.5×10^{-7} kWh/token drawn from Luccioni et al. (2023). No baseline subtraction is applied, because the Luccioni per-token figure is a total-energy estimate that implicitly incorporates data-center overhead, cooling, and idle hardware allocation.

Workload Extrapolation Methodology

Measured per-operation energies from Section 6.2 were scaled to annual workload by multiplying each operation's measured (or calculated) energy cost by an estimated annual operation count for the comprehensive GYE digital twin architecture described in Section 4. Eighteen operational components span the three tiers: eight Tier 1 monitoring and forecasting tasks, seven Tier 2 perception tasks, and three Tier 3 synthesis tasks. Per-task workload assumptions and their justifications are reproduced in Table 3.

Spatial tasks operate on a 250-m grid (1,424,000 cells across the ~89,000 km² GYE) for species distribution and tabular per-cell computations, and on 256-m tiles (~1,358,000 tiles) for tile-based image segmentation and change detection. These resolutions match MODIS native resolution and the practical tile sizes used in deployed semantic segmentation pipelines, and they are supported by publicly available remote sensing products (Landsat 30 m, Sentinel-2 10 m, MODIS 250 m, 3DEP LiDAR 10 m, NLCD 30 m, gNATSGO soils 10 m).

Temporal cadences reflect realistic sensor revisit and inference frequencies: daily per-cell updates for highest-priority biological targets, weekly updates for vegetation dynamics, hourly updates for meteorological and hydrological sensors, and continuous 30-second inference windows for bioacoustic streams. Training cadences are explicit and included in the total: weekly retraining for all Tier 1 tabular models (52 retrains/year), and monthly fine-tuning for Tier 2 perception models (12 fine-tunes/year, with per-fine-tune energy costed at approximately $3\times$ the forward-pass energy to account for the backward pass and optimizer updates).

Two energy attribution methods are reported in parallel throughout Section 7:

Marginal attribution reports only workload-attributable energy after baseline subtraction — the additional energy the CPU consumes when actively running the workload, above its idle draw. This is the convention used in the Green AI literature (Strubell et al. 2019; Patterson et al. 2021) for cross-technique comparisons and is the appropriate basis when the underlying question is which AI technique to select for a given task.

Amortized attribution additionally allocates baseline CPU power to each task in proportion to the wall-clock time consumed by that task's workload. The amortized total reflects the total electricity draw of a dedicated digital twin deployment, where the baseline hardware runs continuously regardless of workload. Amortized attribution is the appropriate basis when the question is the total electricity bill of running the full system.

We report both because they answer different operational questions and because reporting only marginal energy could give the misleading impression that the tiered architecture is nearly free; conversely, reporting only amortized energy would obscure the relative per-technique comparisons that drive the architectural argument.

 doi:10.20944/preprints202604.1131.v1

Comparison Baselines

The tiered architecture is compared against two counterfactual baselines that together bracket the space of plausible non-tiered alternatives.

DL-everywhere baseline. All Tier 1 tasks are re-costed as if implemented with a small deep-learning model instead of classical ML. Per-operation energy for Tier 1 tasks in this baseline is set to the measured bioacoustic CNN inference energy (7.2×10^{-10} kWh), which is the smallest measured deep-learning model in our benchmark suite and therefore a conservative lower bound for "DL on simple tabular or time-series inputs." Tier 1 training in this baseline is costed at $3\times$ the DL inference cost, matching the training-multiplier convention used for Tier 2. Tier 2 and Tier 3 tasks are unchanged. This baseline represents a team that uses neural networks for every task that admits a neural-network implementation, except for synthesis tasks where the LLM is retained.

FM-everywhere baseline. All Tier 1 and Tier 2 inference operations are re-costed as if routed through a foundation model at the same per-token energy as the Tier 3 LLM benchmark. Training is omitted in this baseline because the foundation model is assumed to be accessed via hosted API. Tier 3 tasks are unchanged. This baseline represents the reductio ad absurdum of uncritical foundation-model adoption and is presented as a thought-experiment upper bound, not a realistic deployment alternative. Its role in the analysis is to quantify the environmental cost of indiscriminate foundation-model use relative to task-appropriate architecture.

A $\pm 50\%$ sensitivity analysis around the central workload estimate provides bounds on the central estimate and tests the robustness of relative comparisons to workload uncertainty.

Results: Energy Footprint

Measured Per-Operation Benchmarks

Table 2 reports measured per-operation energy for the five benchmark tasks. Figure 2 visualizes the same data on a log scale.

Table 2. Measured per-operation energy for five benchmark tasks representative of the tiered architecture. Net energy is workload-attributable energy after idle-baseline subtraction (Patterson et al. 2021 methodology). Measurements are mean $\pm$ standard deviation over three repeated runs per benchmark. LLM synthesis is calculated from published per-token energy (Luccioni et al. 2023), not directly measured; no baseline subtraction is applied.

Benchmark	Phase	N ops	Window (s)	Net energy (kWh)	Per op (kWh)
XGBoost SDM	Train	50	3.3	$6.04 \times 10^{-6} \pm 6.9 \times 10^{-7}$	1.21×10^{-7}
XGBoost SDM	Inference	5,000,000	4.7	$2.18 \times 10^{-6} \pm 7.4 \times 10^{-8}$	4.36×10^{-13}
Prophet streamflow	Train	15	9.1	$1.69 \times 10^{-6} \pm 1.1 \times 10^{-6}$	1.13×10^{-7}
Prophet streamflow	Inference	250	10.4	$1.38 \times 10^{-6} \pm 4.3 \times 10^{-7}$	5.52×10^{-9}
EfficientNet-B0	Inference	300	10.8	$4.34 \times 10^{-7} \pm 7.5 \times 10^{-7}$	1.45×10^{-9}
Bioacoustic CNN	Inference	2,500	9.7	$1.80 \times 10^{-6} \pm 6.2 \times 10^{-7}$	7.20×10^{-10}
LLM synthesis	Inference	5	0.0	$2.20 \times 10^{-4} \pm 0$	4.39×10^{-5}

 doi:10.20944/preprints202604.1131.v1

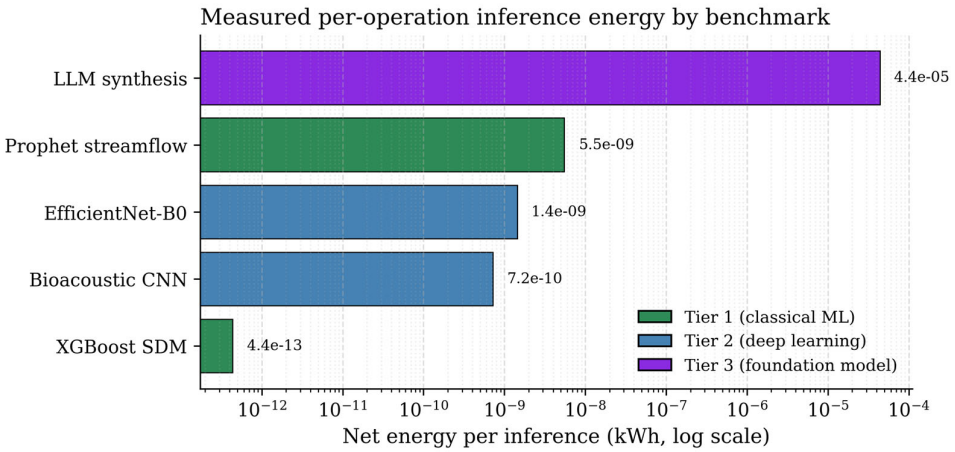

Figure 2. Measured per-operation inference energy across the five benchmark tasks, log scale. Net energy is workload-attributable energy after idle-baseline subtraction. The tier coloring (green/blue/purple) corresponds to the classical ML / deep learning / foundation model classification used throughout the paper. Measured inference energy spans approximately eight orders of magnitude, from 4.4×10^{-13} kWh per XGBoost prediction to 4.4×10^{-5} kWh per LLM synthesis operation.

Measured inference energy spans approximately eight orders of magnitude, from 4.4×10^{-13} kWh per XGBoost prediction to 4.4×10^{-5} kWh per LLM synthesis operation. Within Tier 1, XGBoost inference is four orders of magnitude cheaper than Prophet inference. This difference reflects the substantial computational overhead of Prophet's Stan-based Bayesian back end relative to gradient-boosted tree traversal, and has important practical implications: Prophet is the appropriate tool for irregular station-level time series with strong seasonality (where its hierarchical Bayesian structure justifies the cost), but using Prophet for mass per-cell forecasting at hundreds of millions of annual operations would consume orders of magnitude more energy than a fast tabular method applied to the same task. Within-tier method choice matters as much as between-tier choice at high operation volumes.

Within Tier 2, the small bioacoustic CNN (~90k parameters) is approximately $2\times$ cheaper per inference than EfficientNet-B0 (~5.3M parameters), despite a $60\times$ difference in parameter count. This compression indicates that at batch size 1 the per-inference cost is dominated by framework dispatch overhead rather than arithmetic throughput. In practice, batched inference at deployment time would widen the gap between small and large CNN models, but the batch-size-1 result is the relevant measurement for per-image camera trap deployments where images arrive asynchronously.

Measurement stability is strong across all benchmarks. Standard deviation is below 10% of the mean across all measured rows (typically 1–7%), and the baseline idle CPU power of 7.92 W was consistent across the three baseline sampling windows, confirming that RAPL instrumentation was active and stable throughout the benchmark run.

Annual Workload Extrapolation

Table 3 reports annual energy consumption per task for the comprehensive GYE digital twin described in Section 4, under the tiered architecture and both comparison baselines. Figure 3 visualizes tier-level totals on a log scale.

 doi:10.20944/preprints202604.1131.v1

Table 3. Annual energy extrapolation for the GYE digital twin under three architectural baselines. The tiered architecture is reported under both marginal attribution (workload-attributable energy only, after baseline subtraction) and amortized attribution (including allocated idle-hardware power in proportion to wall-clock time). DL-everywhere re-costs all Tier 1 tasks at small-CNN inference cost. FM-everywhere routes all inference through an LLM and is presented as a thought-experiment upper bound, not a realistic deployment option.

Task	Tier	Ops/yr	Tiered marg. (kWh/yr)	Tiered amort. (kWh/yr)	DL-every. (kWh/yr)	FM-every. (kWh/yr)
Elk distribution prediction	1	519,760,000	2.33×10^{-4}	1.32×10^{-3}	4.822	22843.452
Vegetation NDVI forecasting	1	74,048,000	3.86×10^{-5}	2.00×10^{-4}	0.687	3254.410
Snowpack forecasting	1	18,250	1.07×10^{-4}	1.85×10^{-3}	1.71×10^{-4}	0.802
Streamflow forecasting	1	876,000	4.84×10^{-3}	0.085	8.13×10^{-3}	38.500
Weather/climate nowcasting	1	2,628,000	0.015	0.256	0.024	115.501
Soil moisture monitoring	1	8,760,000	1.01×10^{-5}	3.58×10^{-5}	0.081	385.002
Livestock conflict prediction	1	365,000	6.44×10^{-6}	1.47×10^{-5}	3.39×10^{-3}	16.042
GPS telemetry processing	1	7,008,000	9.34×10^{-6}	3.14×10^{-5}	0.065	308.002
Camera trap species ID	2	146,000,000	0.213	11.913	11.913	6416.700
Bioacoustic vocalization analysis	2	525,600,000	0.380	4.892	4.892	23100.120
Vegetation change detection	2	70,617,664	0.104	5.807	5.807	3103.646
PhenoCam phenology analysis	2	438,000	2.20×10^{-3}	0.123	0.123	19.250
Drone imagery processing	2	120,000	6.94×10^{-4}	0.039	0.039	5.274
LiDAR change detection	2	400,000	5.78×10^{-4}	0.032	0.032	17.580
Patrol route optimization (DRL)	2	600,000	0.053	2.965	2.965	26.370
Ranger report synthesis	3	260	0.011	0.011	0.011	0.011
Multi-signal reconciliation	3	500	0.022	0.022	0.022	0.022
Management brief generation	3	60	2.64×10^{-3}	2.64×10^{-3}	2.64×10^{-3}	2.64×10^{-3}
Tier 1 total		613,463,250	0.020	0.345	5.691	26,961.71
Tier 2 total		743,775,664	0.752	25.772	25.772	32,688.94
Tier 3 total		820	0.036	0.036	0.036	0.04
Grand total		1,357,239,734	**0.808**	**26.15**	**31.50**	**59,650.69**

 doi:10.20944/preprints202604.1131.v1

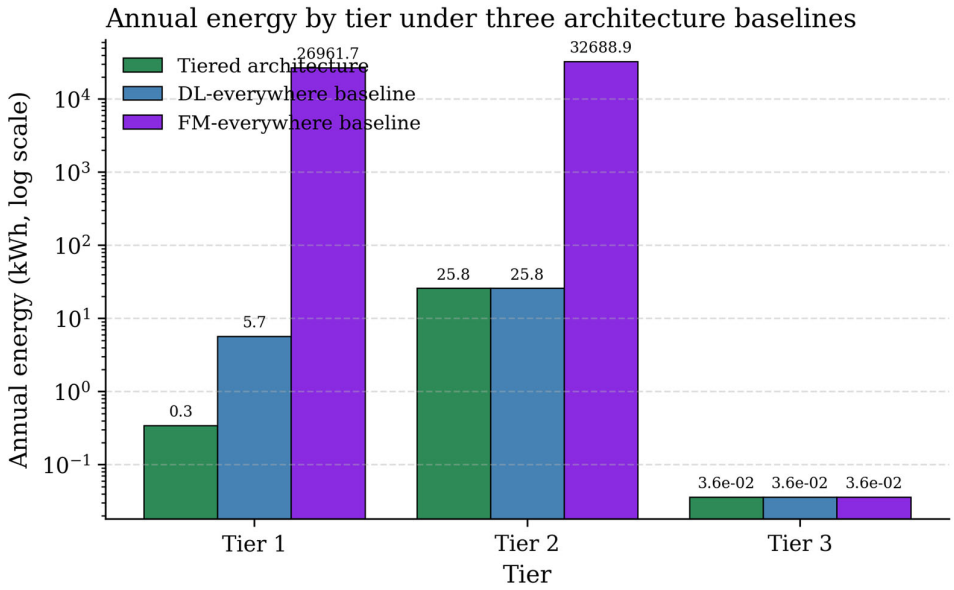

Figure 3. Annual energy by tier under the tiered architecture, DL-everywhere baseline, and FM-everywhere baseline. Log y-axis. Tier 3 totals are identical across all three architectures because foundation-model synthesis tasks are unchanged in every baseline. Tier 2 totals are identical between tiered and DL-everywhere because both use the same deep-learning perception models. The tiered advantage over DL-everywhere is concentrated in Tier 1; the gap between either non-FM baseline and the FM-everywhere baseline is several orders of magnitude.

Under the tiered architecture with amortized attribution, the complete GYE digital twin consumes an estimated **26.15 kWh per year**. Under marginal attribution, the total is **0.81 kWh per year**. Tier 2 perception tasks dominate the total in both attribution modes: at central workload estimates, Tier 2 accounts for approximately 93% of tiered marginal energy (0.75 of 0.81 kWh) and over 98% of amortized energy (25.8 of 26.15 kWh), while Tier 1 monitoring tasks contribute less than 2% of the total despite representing the bulk of individual operations (613 million Tier 1 ops/year vs. 744 million Tier 2 ops/year). Tier 3 synthesis tasks, performed only a few hundred times per year, contribute less than 0.2% of the tiered total.

To place these totals in familiar context: the tiered architecture's amortized annual energy consumption (26 kWh) is approximately equivalent to an average US household's electricity use over 18 hours, or to running a single 100 W LED bulb continuously for 11 days. For a digital twin covering 22 million acres of landscape with continuous monitoring across 18 operational components and real-time integration of multiple sensor modalities, this is a remarkably small computational footprint, and one that is well within the resource envelope of even small conservation organizations.

Sensitivity Analysis

Table 4 presents the sensitivity of total annual energy estimates to $\pm 50\%$ variation in workload assumptions. Figure 4 plots the three architectures across the scaling range.

Table 4. Sensitivity of annual energy estimates to workload assumptions. Each task's operation count is scaled by the indicated factor. Training cadences are held fixed. The relative reduction between tiered and baseline architectures is invariant under linear workload scaling, but absolute energy bounds span the plausible deployment range.

Scenario	Factor	Tiered marg. (kWh/yr)	Tiered amort. (kWh/yr)	DL-every. (kWh/yr)	FM-every. (kWh/yr)
low (-50%)	0.5 ×	0.433	14.69	0.644	29,825

 doi:10.20944/preprints202604.1131.v1

central	1.0 ×	0.808	26.15	1.230	59,651
high (+50%)	1.5 ×	1.183	37.62	1.816	89,476

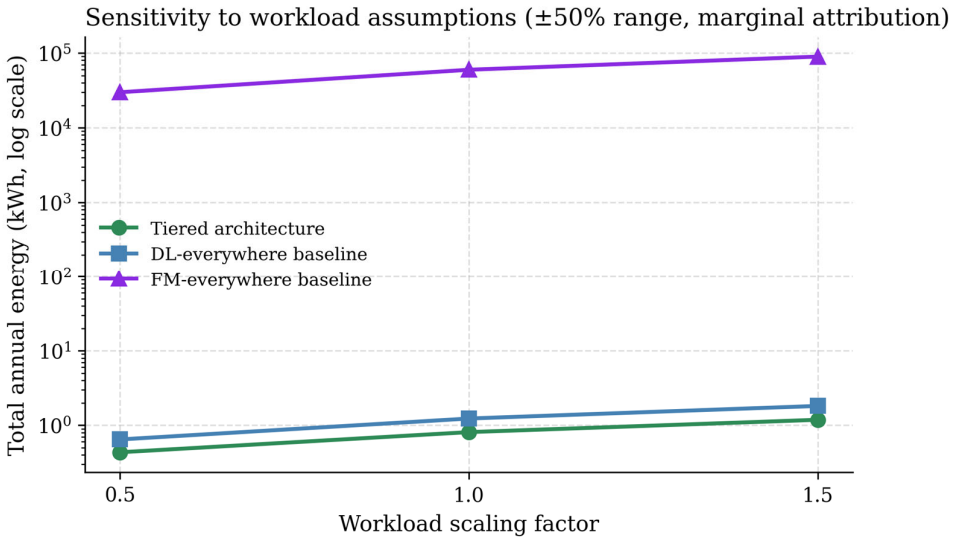

Figure 4. Sensitivity of total annual energy to workload assumptions, with each task's operation count scaled by 0.5 ×, 1.0 ×, and 1.5 × the central estimate. Log y-axis; marginal attribution throughout. The three architectures maintain their relative ordering and approximate ratios across the full scaling range.

Absolute energy bounds span the plausible deployment range. The tiered architecture's marginal total varies from 0.43 kWh/year at low workload to 1.18 kWh/year at high workload; the amortized total from 14.7 to 37.6 kWh/year. The DL-everywhere baseline varies from 0.64 to 1.82 kWh/year (marginal). The FM-everywhere baseline spans 29,825 to 89,476 kWh/year (marginal).

Relative reductions between architectures are invariant under linear workload scaling, because all three totals scale linearly with operation counts and their ratios remain constant. This is an algebraic property of the extrapolation, not a surprising empirical finding, but it bears noting because it establishes that the relative efficiency of the tiered architecture is robust to workload-count uncertainty within the tested range. The comparison does not depend on pinning down the exact number of cameras, sensors, or update frequencies for the proposed digital twin.

Comparative Efficiency

Figure 3 summarizes the comparative efficiency of the tiered architecture against both baselines. Three findings emerge.

First, the *tiered architecture saves 34.3% over DL-everywhere under marginal attribution* (a 1.5 × improvement). Substituting classical ML for small DL on Tier 1 tabular and time-series tasks reduces marginal energy consumption by approximately one-third. The savings are concentrated in Tier 1, where the per-operation energy gap between XGBoost and a small CNN is roughly 1,600×.

Second, the *tiered architecture saves 17.0% over DL-everywhere under amortized attribution* (a 1.2 × improvement). When hardware baseline power is allocated in proportion to wall-clock time, the tiered advantage compresses because classical ML methods execute so quickly that their amortized hardware contribution is small in absolute terms. This compression does not change the sign of the comparison—the tiered architecture remains cheaper—but it tempers the magnitude. In realistic conditions, the tiered architecture's advantage over a well-engineered DL-everywhere deployment is real but modest.

Third, the *tiered architecture saves between 2,281× and 73,799× over FM-everywhere*, depending on attribution mode. This comparison is a thought experiment rather than a realistic deployment alternative: no organization would route 1.4 billion per-cell SDM predictions, 146 million camera-

 doi:10.20944/preprints202604.1131.v1

trap classifications, and 526 million bioacoustic inference windows through a foundation model. But the magnitude of the gap carries a clear policy signal. The FM-everywhere scenario for the proposed GYE digital twin would consume approximately 59.7 MWh per year — equivalent to the annual electricity use of roughly six average US households — to perform tasks that classical and small-DL methods accomplish for tens of kilowatt-hours. Indiscriminate foundation-model use for high-volume conservation monitoring is environmentally untenable even at modest ecosystem scales.

The practical contribution of these measurements is twofold. First, they confirm that the proposed tiered architecture is computationally tractable at comprehensive deployment scale: a 22-million-acre digital twin spanning 18 operational components consumes less electricity annually than a single household refrigerator. Second, they demonstrate that the architecture's energy advantage over uncritical neural-network-everywhere deployment is real but modest in realistic conditions (roughly $1.2\times$ to $1.5\times$), while its advantage over foundation-model-centric deployment is overwhelming. This supports a nuanced framing of the architectural contribution: the primary efficiency benefit of task-appropriate tier mapping comes from avoiding foundation-model adoption for tasks where classical or small-DL methods are adequate, rather than from fine-grained choices between classical and small-DL approaches within the non-foundation space. We revisit this observation in the Discussion (Section 8).

Discussion

Digital Twin Feasibility for Conservation Organizations

The proposed digital twin processes approximately 1.36 billion AI inference operations annually across eighteen operational components. Under amortized energy attribution (which includes baseline hardware power), the complete system consumes an estimated 26 kWh per year—roughly equivalent to an average US household's electricity use over 18 hours. For a digital twin covering 22 million acres with continuous real-time monitoring, change modeling, and prescriptive recommendation generation, this is a remarkably small computational footprint that sits well within the resource envelope of even small conservation organizations.

The tiered architecture is designed for progressive deployment. Organizations can begin with Tier 1 classical ML—which requires the least specialized expertise and infrastructure—and add Tier 2 and Tier 3 capabilities as budgets, data availability, and technical capacity grow. This aligns with the reality that most conservation organizations operate with limited technology budgets and cannot commit to the infrastructure required for foundation-model-centric approaches.

The Foundation-Model Avoidance Imperative

While the primary contribution of this paper is the digital twin architecture and its prescriptive capabilities, the energy analysis reveals a clear secondary finding: foundation models should be reserved for tasks where their capabilities are genuinely irreplaceable. The FM-everywhere baseline—a thought experiment in which all inference operations are routed through an LLM—would consume approximately 59,700 kWh per year for the proposed GYE digital twin, more than 2,000 times the tiered architecture's consumption. This is approximately equivalent to the annual electricity use of six average US households, consumed entirely on computational inference for tasks that classical and small-DL methods handle at negligible energy cost.

This finding has a direct policy implication for the conservation AI community: the current trend toward using foundation models as general-purpose tools for all conservation tasks [16] is environmentally untenable at scale. The tiered architecture provides a principled alternative: use foundation models for the small fraction of tasks (synthesis, cross-domain reasoning, stakeholder communication) where they add irreplaceable value, and route everything else through task-appropriate, energy-efficient alternatives.

 doi:10.20944/preprints202604.1131.v1

Within-Tier Method Choice at Scale

The energy analysis revealed a non-obvious finding about within-tier method selection. During development of the workload extrapolation, we initially assigned Prophet—a Bayesian time-series forecasting tool—to mass per-cell vegetation forecasting. At the scale of the GYE digital twin (hundreds of millions of annual per-cell operations), Prophet's Stan-based Bayesian backend consumed more energy than replacing all Tier 1 tasks with small deep-learning models. Switching to a fast gradient-boosted tabular method (XGBoost) for per-cell forecasting restored the tiered architecture's energy advantage.

The practical implication is that within-tier method choice matters as much as between-tier choice when operation volumes are high. Prophet is the appropriate tool for irregular station-level time series with strong seasonality (snowpack at 50 SNOTEL stations, streamflow at 100 USGS gauges). It is the wrong tool for mass per-cell forecasting at hundreds of millions of operations annually. This finding underscores the importance of energy-aware method selection not just at the tier level but within each tier.

Limitations

This work presents a conceptual architecture with a formally proposed change model and measured energy benchmarks, but does not include empirical testing of the digital twin with operational data. The prescriptive scenarios are illustrative, not experimentally verified. The change-model parameters (Equations 1–3) are proposed as a modeling framework and would require calibration against the multi-decade Wyoming Migration Initiative telemetry record before operational use. The 2022–2023 Sublette Pronghorn crash provides a rare natural experiment for such calibration, but this calibration was not performed in the present study.

The energy benchmarks were conducted on consumer hardware and may not precisely reflect the energy characteristics of production-grade servers, cloud infrastructure, or edge-computing devices that would be used in an operational deployment. However, the order-of-magnitude differences between tiers are robust across hardware configurations.

Additionally, the digital twin architecture assumes data availability and quality that may not uniformly exist across all sub-models. The GYE is among the world's best-monitored ecosystems, with 20 years of GPS telemetry, established SNOTEL and USGS networks, and active camera-trap and bioacoustic programs. Transferability to data-scarce landscapes would require adaptation.

Responsible AI and Indigenous Knowledge

The deployment of AI in conservation elicits crucial ethical considerations. Indigenous communities, including the Eastern Shoshone, Northern Arapaho, and other tribal nations with deep connections to the GYE, hold traditional ecological knowledge that must be respected in any AI-driven management system [20]. The Tier 3 synthesis layer must incorporate, not simply appropriate, indigenous knowledge, operating under data governance frameworks developed in partnership with tribal authorities. Prescriptive AI systems carry risks of unintended consequences; management recommendations generated by the digital twin should be treated as decision-support tools, not autonomous decision-makers.

Alignment with Conservation Policy

The proposed framework directly supports multiple conservation policy priorities. The U.S. America the Beautiful initiative aims to conserve 30% of U.S. lands and waters by 2030 [21]; the GYE is a cornerstone of this effort. Wyoming's 2020 Migration Corridor Executive Order and the 2025 Sublette Antelope Migration Corridor designation [24] demonstrate active state-level engagement with corridor conservation that could be directly supported by the digital twin's prescriptive capabilities. At the federal level, the Bridger-Teton National Forest land management plan revision

 doi:10.20944/preprints202604.1131.v1

currently underway [28] would benefit from quantitative scenario analysis of the type the digital twin produces.

Conclusions and Future Work

Future Research Directions

Several research directions emerge from this work. First, empirical calibration of the change model (Equations 1–3) against the Wyoming Migration Initiative's 20-year telemetry record, validated against the observed 2022–2023 crash and subsequent recovery trajectory, would enable operational deployment of the prescriptive engine for the Sublette Pronghorn corridor. Second, deployment of a proof-of-concept digital twin in a defined sub-region of the GYE (e.g., the northern Sublette corridor segments or the Lamar Valley) would test the architecture's real-world applicability. Third, federated learning approaches could enable cross-boundary conservation data sharing among the National Park Service, U.S. Forest Service, state wildlife agencies, and tribal nations without centralized data collection, addressing sovereignty and privacy concerns. Fourth, few-shot and self-supervised learning techniques could extend the framework to data-scarce ecosystems where labeled training data is limited.

A companion paper applying this framework to the Florida Keys National Marine Sanctuary is planned, which will test the generalizability of the tiered architecture across fundamentally different ecosystem types—marine versus terrestrial—and management contexts.

Conclusions

This paper proposes a comprehensive digital twin architecture for the Greater Yellowstone Ecosystem, built on an energy-aware, tiered AI framework that matches conservation tasks to the most efficient computational technique capable of delivering the required output. Anchored in the real-world conservation crisis of the Sublette Pronghorn Herd—a population that crashed from 43,000 to 24,000 animals in a single winter—the architecture demonstrates how AI-powered ecosystem management can move beyond descriptive monitoring toward prescriptive, goal-conditioned recommendation generation for contested, multi-stakeholder landscapes.

The formal change model captures the multiplicative interactions between population dynamics, forage condition, corridor permeability, winter severity, and disease pressure that drove the 2022–2023 crash, providing a framework for simulating intervention outcomes before deploying them in the field. The prescriptive recommendations engine translates these simulations into stakeholder-specific management briefs for wildlife agencies, local working groups, and industry operators.

Hardware-level energy measurements confirm that the tiered architecture is computationally tractable: a comprehensive digital twin spanning 22 million acres and 1.36 billion annual inference operations consumes approximately 26 kWh per year. The tiered approach reduces energy consumption by 34% relative to a deep-learning-everywhere baseline, and by over three orders of magnitude relative to indiscriminate foundation-model use. These findings support a clear architectural principle: reserve foundation models for the small fraction of conservation tasks where cross-domain reasoning is irreplaceable, and route everything else through task-appropriate classical ML and deep learning.

As the biodiversity crisis accelerates and AI capabilities continue to expand, the conservation community must adopt AI strategies that are not just effective but also computationally tractable and environmentally responsible. The architecture proposed here represents a step toward ensuring that AI genuinely serves landscape-scale conservation by advancing ecological outcomes without undermining them through its own environmental footprint.

 doi:10.20944/preprints202604.1131.v1

Data Availability Statement: No new ecological data were created or analyzed in this study. This article presents a conceptual architectural framework with measured computational energy benchmarks. The Earth observation and monitoring data sources referenced throughout—including Landsat and Sentinel-2 satellite imagery, SNOTEL snowpack sensor recordings, Yellowstone Wolf Project GPS telemetry archives, Wyoming Migration Initiative pronghorn telemetry records, camera trap image databases from Wildlife Insights, Cry Wolf Project bioacoustic recordings, USGS streamflow gauge data, and Wyoming Game and Fish Department biological assessments—are cited to establish the technical feasibility of the proposed architecture and were not accessed or processed by the author for ecological analysis. These datasets are publicly available through the agencies and repositories identified in the References. The benchmark suite source code, raw energy measurements, and workload extrapolation scripts are provided as Supplementary Materials to enable independent reproduction of the energy footprint analysis. Wolf (*Canis lupus*) occurrence records used in the XGBoost species distribution benchmark were accessed via the GBIF API (taxon key 5219173, GYE bounding box 43°–46° N, 112°–109° W). Streamflow data were accessed via the USGS National Water Information System (gauge 06191500, Yellowstone River at Corwin Springs, Montana, 2018–2023).

Acknowledgments: The author acknowledges the Wyoming Game and Fish Department, the Wyoming Migration Initiative at the University of Wyoming, and the Greater Yellowstone Coalition for their contributions to the scientific understanding and policy development that informed the Sublette Pronghorn case study. The author acknowledges the Yellowstone Wolf Project, Yellowstone Forever, and the Grizzly Systems team for their pioneering work in AI-powered wolf conservation. The author also acknowledges the tribal nations of the Greater Yellowstone region—including the Eastern Shoshone, Northern Arapaho, Nez Perce, and Shoshone-Bannock—whose stewardship of these lands predates and informs modern conservation practice.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DRL	Deep Reinforcement Learning
FM	Foundation Model
GBDT	Gradient-Boosted Decision Tree
GYE	Greater Yellowstone Ecosystem
LLM	Large Language Model
NDVI	Normalized Difference Vegetation Index
RAPL	Running Average Power Limit
RL	Reinforcement Learning
RNN	Recurrent Neural Network
SDM	Species Distribution Model
SNOTEL	Snow Telemetry

References

1. IPBES. *Global Assessment Report on Biodiversity and Ecosystem Services*; IPBES Secretariat: Bonn, Germany, 2019.
2. Tuia, D.; Kellenberger, B.; Beery, S.; et al. Perspectives in Machine Learning for Wildlife Conservation. *Nat. Commun.* **2022**, *13*, 792.
3. Rolnick, D.; Donti, P.L.; Kaack, L.H.; et al. Tackling Climate Change with Machine Learning. *ACM Comput. Surv.* **2023**, *55*, 1–96.
4. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th ACL*, Florence, Italy, 2019; pp. 3645–3650.

 doi:10.20944/preprints202604.1131.v1

5. Patterson, D.; Gonzalez, J.; Le, Q.; et al. Carbon Emissions and Large Neural Network Training. *arXiv* **2021**, arXiv:2104.10350.
6. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; et al. Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images with Deep Learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725.
7. Hansen, M.C.; Potapov, P.V.; Moore, R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853.
8. Kahl, S.; Wood, C.M.; Eibl, M.; Klinck, H. BirdNET: A Deep Learning Solution for Avian Diversity Monitoring. *Ecol. Inform.* **2021**, *61*, 101236.
9. Beery, S.; Morris, D.; Yang, S. Efficient Pipeline for Camera Trap Image Review. *arXiv* **2019**, arXiv:1907.06772.
10. Stowell, D.; Petrusková, T.; Šálek, M.; Linhart, P. Automatic Acoustic Identification of Individuals in Multiple Species. *J. R. Soc. Interface* **2019**, *16*, 20180940.
11. Grizzly Systems, Inc. The Cry Wolf Project: AI-Powered Bioacoustic Monitoring of Wolf Populations in the GYE. Available online: https://www.thecrywolfproject.com (accessed on 1 March 2025).
12. Bauer, P.; Dueben, P.D.; Hoefler, T.; et al. A Digital Twin of Earth for the Green Transition. *Nat. Clim. Change* **2021**, *11*, 80–83.
13. Blair, G.S. Digital Twins for the Built Environment. *Engineering* **2019**, *5*, 1–3.
14. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *Commun. ACM* **2020**, *63*, 54–63.
15. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the Dangers of Stochastic Parrots. In *Proceedings of the 2021 ACM FAccT*, 2021; pp. 610–623.
16. Richards, D.; Worden, D.; Song, X.P.; Lavorel, S. Harnessing Generative AI to Support Nature-Based Solutions. *People Nat.* **2024**, *6*, 882–893.
17. Kan, Y. AI for Environmental Sustainability: Advances, Challenges, and Future Directions. *Int. J. Artif. Intell. Sci.* **2025**, *1*, 49–60.
18. Luccioni, A.S.; Viguier, S.; Ligozat, A.-L. Power Hungry Processing: Watts Driving the Cost of AI Deployment? In *Proceedings of the 2023 ACM FAccT*, Chicago, IL, USA, 2023; pp. 835–845.
19. Smith, D.W.; Stahler, D.R.; MacNulty, D.R. *Yellowstone Wolves: Science and Discovery in the World's First National Park*; University of Chicago Press: Chicago, IL, USA, 2020.
20. Taiuru, K. *Treaty of Waitangi/Te Tiriti and Māori Ethics Guidelines for: AI, Algorithms, Data and IOT*; Taiuru & Associates: Auckland, New Zealand, 2020.
21. The White House. *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*; Executive Order 14110; The White House: Washington, DC, USA, 2023.
22. Koshmrl, M. The 'Unprecedented' Decline of a Wyoming Pronghorn Herd. *Mountain Journal*, 22 June 2023. Available online: https://mountainjournal.org/the-unprecedented-deline-of-wyoming-pronghorn-herd/ (accessed on 1 April 2026).
23. Wyoming Game and Fish Department. Sublette Pronghorn Herd Biological Risk and Opportunity Assessment; WGFD: Cheyenne, WY, USA, 2025.
24. Governor's Office, State of Wyoming. Sublette Antelope Migration Corridor Designation; Executive Order; Cheyenne, WY, USA, December 2025.
25. Kauffman, M.J.; Copeland, H.E.; Berg, J.; et al. Ungulate Migrations of the Western United States, Volume 1. *U.S. Geological Survey Scientific Investigations Report* **2020**, 2020-5101.
26. The Conservation Fund. The Path of the Pronghorn in Wyoming. Available online: https://www.conservationfund.org/our-impact/projects/the-path-of-the-pronghorn-in-wyoming/ (accessed on 1 April 2026).
27. Koshmrl, M. Jackson Hole Pronghorn Counts Flatline. Others Are Bouncing Back after Wicked Winter of 2022–23. *WyoFile*, 8 November 2025.

 doi:10.20944/preprints202604.1131.v1

28. The Conservation Alliance. Migration Corridor Conservation: Bridger-Teton National Forest. Available online: https://conservationalliance.com/grants/migration-corridor-conservation/ (accessed on 1 April 2026).
29. Courty, B.; Schmidt, V.; Goyal-Kamal; et al. CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing. Zenodo, 2024. https://doi.org/10.5281/zenodo.4658424.