

MARLIN: Multi-Agent Reinforcement Learning with Murmuration Intelligence and LLM Guidance for Reservoir Management

Heming Fu

Department of Electrical and
Computer Engineering
Stony Brook University, NY
heming.fu@stonybrook.edu

Shan Lin

Department of Electrical and
Computer Engineering
Stony Brook University, NY
shan.x.lin@stonybrook.edu

Guojun Xiong

John Hopcroft Center
School of Computer Science
Shanghai Jiao Tong University
gjxiong@sjtu.edu.cn

ABSTRACT

Intensifying climate change and cascading uncertainties across interconnected reservoir networks pose escalating threats to global water security, demanding management systems that are both adaptive and scalable. Traditional centralized optimization becomes computationally intractable and brittle under real-world uncertainty, while existing reinforcement learning (RL) approaches are not designed for complex, multi-node hydrological systems. To address these challenges, we introduce MARLIN, a decentralized reservoir management framework that explicitly handles **dual-layer uncertainty**: (i) stochastic variability in physical water transfer and (ii) dynamic, human–environmental perturbations. MARLIN embeds bio-inspired *alignment*, *separation*, and *cohesion* rules into a multi-agent RL (MARL) architecture to stabilize coordination under physical uncertainty. Additionally, external conditions such as weather forecasts, regulatory updates, and stakeholder preferences introduce *unstructured textual information* that traditional models cannot process directly. To bridge this gap, we integrate a Large Language Model (LLM) that interprets such contextual information and dynamically adjusts the coordination parameters of the three murmuration rules, enabling rapid adaptation to evolving environmental and human requirements. Experiments on USGS data show that MARLIN improves uncertainty handling by 23%, reduces computational cost by 35%, and accelerates flood response by 68%. The framework demonstrates excellent scalability, with emergent coordination patterns increasing super-linearly as the network expands while maintaining linear computational complexity. These results highlight MARLIN’s potential as a scalable and intelligent solution for adaptive water resource management and disaster prevention.

KEYWORDS

Multi-Agent RL, Starling Murmuration, LLM

ACM Reference Format:

Heming Fu, Shan Lin, and Guojun Xiong. 2026. MARLIN: Multi-Agent Reinforcement Learning with Murmuration Intelligence and LLM Guidance for Reservoir Management. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*, Paphos, Cyprus, May 25 – 29, 2026, IFAAMAS, 9 pages. <https://doi.org/10.65109/RQE9663>



This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). <https://doi.org/10.65109/RQE9663>

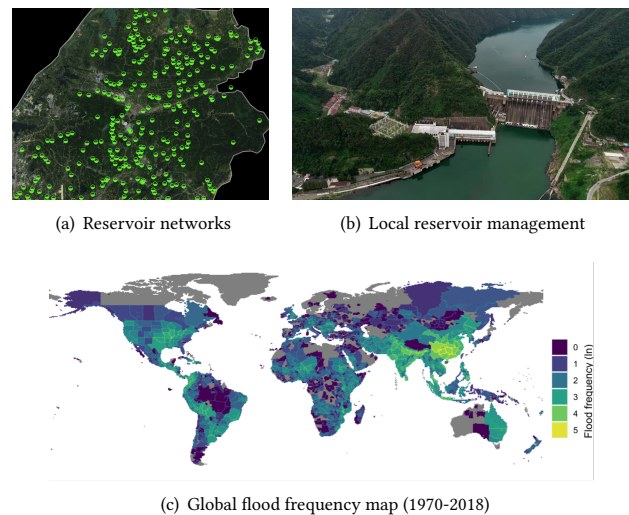


Figure 1: Water resource management challenges: (a) complex networks of interconnected reservoirs requiring coordination, (b) individual reservoirs must be locally managed, and (c) high global flood frequency demanding robust management systems.

1 INTRODUCTION

Water resource management is at a critical inflection point, as traditional engineering approaches are increasingly inadequate for the complexity and uncertainty of modern hydrological systems. Climate change has dramatically altered precipitation patterns (see Figure 1(c)), resulting in more frequent and severe extreme weather events that disproportionately impact vulnerable populations. The consequences are global: between 2000 and 2019, floods affected over 1.65 billion people and caused an estimated 651 billion in economic losses [41], while droughts impacted 1.43 billion people and resulted in damages exceeding 130 billion [8]. Events that were once considered century-scale extremes now occur every 20 years in many regions, and flash flood frequency has surged by 21% over the past decade [22]. Recent disasters underscore the urgent need for adaptive water management systems. The 2021 European floods resulted in over 200 deaths and approximately 50 billion in damages [25], and New York’s 2023 extreme rainfall events both revealed critical failures in infrastructure and cross-jurisdictional coordination [19]. Similarly, Texas winter storms

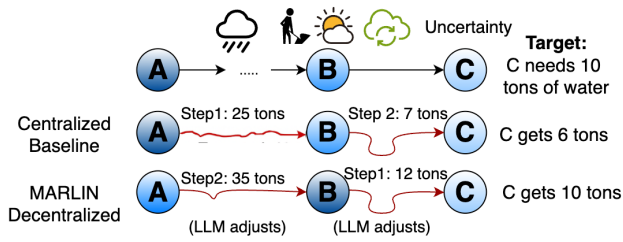


Figure 2: Illustration of centralized vs. decentralized coordination under uncertainty. The target is for node C to receive 10 tons of water. In the *centralized baseline*, node A releases 25 tons based on static planning, yet due to environmental variations like rainfall, node C ultimately receives only 6 tons. In contrast, *MARLIN decentralized coordination* allows each node to adapt through local feedback: node B first releases 12 tons, and node A subsequently compensates by adjusting its output to 35 tons, guided by LLM-assisted local adjustments. This closed-loop adaptation enables node C to obtain 10 tons despite environmental uncertainty.

repeatedly threaten water security by disabling power-dependent systems [9, 31], while a historic megadrought has pushed Lakes Mead and Powell below 30% capacity, jeopardizing water supplies for 40 million people [4, 6]. These are not isolated incidents but symptoms of systemic vulnerabilities. A systematic study of US dam failures, including the 2020 Michigan breach and the 2017 Oroville spillway collapse, concluded that current hydrologic design and operation protocols are often inadequate for managing the clustered, extreme weather events of a changing climate [16].

These widespread failures stem from two fundamental limitations of current centralized management paradigms. First, as reservoir networks grow, computational complexity scales as $O(N^3)$, making real-time response to rapidly changing conditions impractical [18, 21, 37]. Second, these centralized models are highly vulnerable to cascading uncertainty from dual sources: (i) *physical water transfer uncertainty* arising from evaporation, seepage, and variable channel conditions during water transport; and (ii) *environmental uncertainty* encompassing seasonal variations (e.g., spring agricultural irrigation demanding 60% of annual water allocation), extreme weather events, and human activities ranging from routine operations to emergency situations (e.g., industrial accidents requiring immediate water supply for cooling, as seen in the 2023 Fukushima release [2]). These diverse environmental factors create complex, time-varying demands that traditional models cannot anticipate. For instance, minor errors compound catastrophically through network propagation: assuming independent 7% allocation uncertainty at each node, the cumulative error grows as $\sigma_{cumulative} = \sigma_{base} \sqrt{n}$, reaching 22% after just 10 nodes and exceeding 40% in networks with 30+ interconnected reservoirs, which are typical of major river basins. Under correlated environmental conditions, this amplification accelerates further, rendering centralized predictions unreliable, as illustrated in Fig 2. Additionally, multi-agent reinforcement learning (MARL) provides a distributed alternative but remains limitations to uncertainty [14, 27, 40, 42], which often

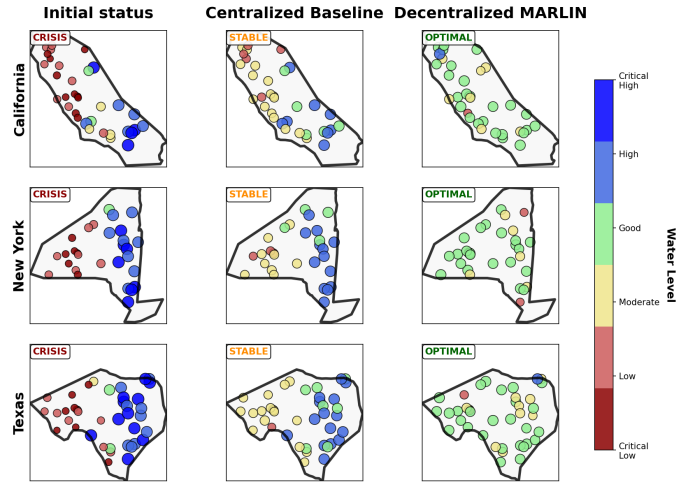


Figure 3: Statewide reservoir status under uncertainty: Initial vs. Centralized Baseline vs. Decentralized MARLIN. Each panel shows reservoir water levels. The centralized baseline yields only modest stabilization, whereas MARLIN achieves near-optimal balance with lower variance and better regional coordination.

suffer from training instability and poor convergence under cascading uncertainties, resulting in unsafe, oscillatory behaviors in practice [13].

Motivated by nature’s ability to achieve robust coordination under uncertainty, we draw inspiration from starling murmurations (Figure 4). These flocks show how complex, adaptive organization can emerge from simple, local rules—*alignment*, *separation*, and *cohesion*—without any centralized control [5, 15, 28]. We translate this principle of emergent intelligence into MARLIN, a decentralized framework that integrates these bio-inspired rules with reinforcement learning in two ways. First, to manage physical water transfer uncertainty, MARLIN maps these bio-inspired rules to reservoir control actions and directly incorporates their signals into the policy gradients of a specialized MARL architecture. This novel integration promotes stable learning while fostering emergent, system-wide safety and responsiveness. Second, to address the more complex environmental uncertainties, MARLIN integrates a Large Language Model (LLM). This allows the system to process diverse, unstructured textual information from weather forecasts to regulatory documents and translate its knowledge into adaptive reward-

shaping signals, guiding agents in dynamic contexts that are intractable for traditional optimization. As shown in Fig. 3, reservoir states progress from crisis (Initial) to modest stability under a centralized baseline, and to near-optimal balance with MARLIN.

In summary, our main contributions are:



Figure 4: Starling Murmuration: emergent intelligence from simple local coordination rules.

- ▶ We propose a novel bio-inspired MARL system, MARLIN, that applies starling murmuration principles to achieve emergent and scalable coordination from simple local interactions in decentralized water management.
- ▶ We develop a dual-layer uncertainty handling mechanism, where simple local coordination rules manage physical water transfer uncertainty, while LLM-guided reward shaping adapts to complex environmental fluctuations and human preferences.
- ▶ Experiments on USGS watershed data demonstrate that MARLIN achieves substantial improvements over baselines: 23% better uncertainty handling, 35% lower computation cost, 68% faster flood response, and 42% improvement in regional water balance, underscoring its potential for safeguarding vulnerable communities against water-related disasters.

2 RELATED WORK

Water Management and Multi-Agent Systems. Classical water management relied on deterministic optimization like dynamic programming [20, 39] and network flow models [33], suffering from exponential complexity and static assumptions. MPC became dominant in operational systems but demonstrates critical limitations during rapid changes. The 2011 Mississippi floods exemplified delayed recalibration contributing to \$2.8 billion avoidable damage [30]. MARL emerged as a distributed alternative. MADDPG [23] uses centralized critics with decentralized actors, QMIX [27] employs value decomposition with monotonic mixing, and MAPPO [40] adapts PPO to multi-agent settings. However, these methods face training instability under non-stationary environments and assume coordination emerges naturally from individual optimization, overlooking explicit coordination needed for uncertainty handling [10]. **Bio-Inspired Coordination.** Reynolds formalized flocking behavior into alignment, separation, and cohesion rules [28]. Theoretical analysis revealed phase transitions between ordered and disordered states [7, 36], while empirical studies confirmed scale-free correlations and rapid information propagation [5]. These inspired algorithms like Particle Swarm Optimization [17] and practical swarm robotics applications [26, 29]. However, most focus on continuous motion rather than discrete decision-making required in infrastructure control.

LLMs for Adaptive Control. Recent work explores LLMs for reward design and adaptation [35, 38]. Eureka [24] uses GPT-4 for automated reward generation, while L2R [11, 12] enables natural language reward specification. However, these approaches primarily focus on single-agent settings or assume full observability, limiting their applicability in decentralized multi-agent systems.

3 PROBLEM FORMULATION

We model modern reservoir networks as distributed decision systems operating under dual-source uncertainty—physical transfer variability and environmental-human factors—that propagates through interconnected hydrological channels, as shown in Fig 5.

Network Model and Stochastic Dynamics. We model the reservoir network as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ represents N reservoir nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes water flow channels. Each reservoir i maintains state vector $s_i(t) =$

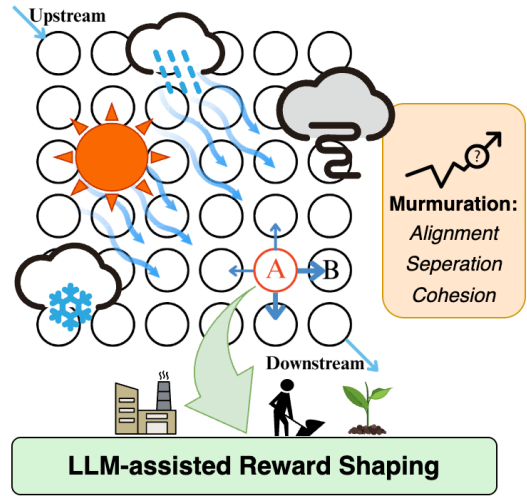


Figure 5: Illustration of the multi-layer reservoir network system under cascading uncertainty. Upstream environmental disturbances (e.g., heat, rainfall, freezing, human activities) dynamically affect inflows, leading to transfer uncertainty among interconnected reservoirs. Each node (e.g., node A) operates under both environmental and flow-level uncertainties. The right panel shows the *murmuration-inspired* coordination rules (alignment, separation, and cohesion) that address the *Level 1: Physical Transfer Uncertainty*. The LLM-assisted reward shaping layer translates human, industrial, and ecological objectives into adaptive decision feedback, addressing *Level 2: Environmental and Human Modulation*.

$[h_i(t), q_{in,i}(t), q_{out,i}(t), \omega_i(t), d_i(t)]^\top$. Here, $h_i(t)$ represents water level in meters determining storage capacity and downstream pressure (for instance, Lake Mead’s water level dropping below 320 meters triggers mandatory conservation measures affecting 40 million people), $q_{in/out,i}(t)$ aggregate all incoming and outgoing flows including controlled releases (typically 200-500 m³/s during normal operations) and spillway overflow (can exceed 10,000 m³/s during extreme floods as seen in the 2017 Oroville crisis [34]), $\omega_i(t)$ captures local weather conditions (temperature affecting evaporation, precipitation intensity, humidity) collected from NOAA stations, and $d_i(t)$ represents time-varying demand from agricultural irrigation (peaks at 800 m³/s during California’s summer growing season), urban consumption, and industrial processes.

The water level dynamics follow:

$$\frac{dh_i(t)}{dt} = \frac{1}{A_i} \left[\sum_{j \in \mathcal{N}_i^{up}} f_{ji}(t) - \sum_{k \in \mathcal{N}_i^{down}} f_{ik}(t) + q_{ext,i}(t) + \eta_i(t) \right], \quad (1)$$

where A_i is reservoir surface area, $\mathcal{N}_i^{up/down}$ denote upstream/downstream neighbors, $f_{ij}(t)$ represents actual water transfer between reservoirs (defined below), $q_{ext,i}(t)$ captures external inflows from rainfall and tributaries, and $\eta_i(t) \sim \mathcal{N}(0, \sigma_\eta^2)$ models measurement noise. The control actions relate to flows as $q_{out,i}(t) = \sum_{k \in \mathcal{N}_i^{down}} a_{i \rightarrow k}(t)$ where $a_{i \rightarrow k}(t)$ is the controlled release from reservoir i to k , and

$q_{in,i}(t) = \sum_{j \in \mathcal{N}_i^{up}} f_{ji}(t)$ represents actual received water after transfer losses.

Dual-Source Uncertainty. The critical challenge lies in cascading uncertainty from two distinct sources.

Level 1: Physical Transfer Uncertainty. When reservoir i releases volume $a_{i \rightarrow j}(t)$ toward downstream reservoir j , the actual water transfer follows:

$$f_{ij}(t) = \alpha_{ij}(t) \cdot a_{i \rightarrow j}(t) + \epsilon_{ij}(t), \quad (2)$$

where $\alpha_{ij}(t) \in [\epsilon, 1]$ represents time-varying channel efficiency accounting for evaporation, seepage, and channel conditions, with $\epsilon = 0.1$ ensuring minimum 10% efficiency even in extreme conditions, while $\epsilon_{ij}(t) \sim \mathcal{N}(0, \sigma_{base}^2)$ captures random fluctuations. During normal conditions, concrete-lined channels maintain $\alpha_{ij} \approx 0.95$, but natural earthen channels drop to $\alpha_{ij} \approx 0.70$ due to seepage through permeable soils. Extreme heat waves can reduce efficiency by an additional 15-20% through evaporation.

Level 2: Environmental and Human Modulation. The channel efficiency is affected by environmental and human factors:

$$\alpha_{ij}(t) = \min\left(1, \max\left(\epsilon, \alpha_{nominal} \cdot (1 - \gamma_{env}(\omega_i(t), \omega_j(t)) - \gamma_{human}(d_i(t), d_j(t)))\right)\right),$$

where γ_{env} captures weather-induced losses (higher evaporation during heatwaves when temperatures in $\omega_i(t)$ exceed 40°C can increase γ_{env} by 0.3), and γ_{human} captures both predictable patterns from demand $d_i(t)$ and unpredictable interventions from textual sources $\mathcal{T}(t)$. These textual sources include news articles ("Fukushima plant [1] requires 500,000 gallons per minute for emergency cooling"), regulatory bulletins ("Stage 3 drought restrictions effective immediately"), and social media alerts ("Chemical spill reported upstream of intake").

Uncertainty Propagation. For a cascade of n reservoirs, cumulative uncertainty compounds as:

$$\text{Var}[h_n(t)] = \sum_{i=1}^{n-1} \left(\prod_{j=i+1}^n \alpha_{j,j-1}^2(t) \right) \sigma_{base}^2 + \sigma_{\eta}^2 \quad (3)$$

showing how upstream uncertainties are amplified by the product of channel efficiencies. When $\alpha_{ij}(t) \approx 0.9$ (typical conditions), uncertainty grows moderately, but during extreme events with $\alpha_{ij}(t) \approx 0.5$, downstream uncertainty explodes exponentially. In the Colorado River's 15-reservoir cascade, even with individual $\alpha_{ij} = 0.93$, the compound efficiency to Lake Mead drops to 0.35, meaning 65% uncertainty in delivery.

Multi-Objective Control Problem. We formulate reservoir management as a multi-objective optimization problem explicitly designed to balance four interacting goals—*safety*, *supply*, *ecology*, and *efficiency*—that together define the operational priorities of the

system, i.e.,

$$J_{safety}(\pi) = -\mathbb{E}_{\pi} \left[\sum_{t=0}^T \sum_{i=1}^N \lambda_{flood,i} \cdot \mathbb{I}[h_i(t) > h_{safe,i}] \right], \quad (4)$$

$$J_{supply}(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^T \sum_{i=1}^N \min\left(1, \frac{q_{out,i}(t)}{d_i(t)}\right) \right], \quad (5)$$

$$J_{ecology}(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^T \min\left(1, \frac{\sum_{i \in \mathcal{V}} q_{out,i}(t)}{F_{eco}(t)}\right) \right], \quad (6)$$

$$J_{efficiency}(\pi) = -\mathbb{E}_{\pi} \left[\sum_{t=0}^T \sum_{i=1}^N c_{op,i} \cdot q_{out,i}(t) \right], \quad (7)$$

where $\lambda_{flood,i}$ weights flood risk priority for populated areas, $h_{safe,i}$ is the safety threshold for reservoir i , $F_{eco}(t)$ specifies minimum ecological flow requirements, and $c_{op,i}$ represents operational cost per unit water released. In addition, each reservoir i observes only local information:

$$\mathcal{O}_i(t) = \{\mathbf{s}_i(t), \{\mathbf{s}_j(t - \tau)\}_{j \in \mathcal{N}_i^{up} \cup \mathcal{N}_i^{down}}, \boldsymbol{\omega}_{forecast,i}(t:t+H)\}, \quad (8)$$

where τ is communication delay and $\boldsymbol{\omega}_{forecast,i}(t:t+H)$ provides H -hour weather forecasts. The optimization seeks decentralized policies $\pi_i : \mathcal{O}_i \rightarrow \{a_{i \rightarrow k}(t)\}_{k \in \mathcal{N}_i^{down}}$ maximizing:

$$\begin{aligned} \max_{\pi} \quad & \mathbf{J}(\pi) = [J_{safety}, J_{supply}, J_{ecology}, J_{efficiency}]^T \quad (9) \\ \text{s.t.} \quad & 0 \leq a_{i \rightarrow k}(t) \leq a_{max,i}, \quad h_{min} \leq h_i(t) \leq h_{max}. \quad (10) \end{aligned}$$

REMARK 1. *This formulation reveals three fundamental challenges: (1) Cascading uncertainty amplification where errors grow exponentially through the network under reduced channel efficiency, making centralized prediction unreliable; (2) Multi-objective coupling where flood safety conflicts with supply reliability, especially during extreme events; (3) Information asymmetry where environmental factors arrive as structured data ($\omega_i(t)$) while human interventions manifest as demand spikes ($d_i(t)$) or unstructured text requiring different processing paradigms. These challenges motivate MARLIN's dual approach in Section 4—bio-inspired coordination to handle physical uncertainty through local rules and LLM-guided adaptation to process contextual information about environmental and human factors.*

4 METHODOLOGY

MARLIN addresses the dual-source uncertainty through a hierarchical approach: bio-inspired coordination rules handle Level 1 physical transfer uncertainty through local mathematical operations, while LLM-guided adaptation processes Level 2 environmental-human uncertainty from textual sources $\mathcal{T}(t)$, as shown in Fig 6.

4.1 Murmuration-Inspired Coordination Layer

To address the cascading physical uncertainty where variance grows as $\text{Var}[h_n(t)] = \sum_{i=1}^{n-1} \left(\prod_{j=i+1}^n \alpha_{j,j-1}^2(t) \right) \sigma_{base}^2$, we adapt three starling murmuration rules that create robust local consensus without requiring global information.

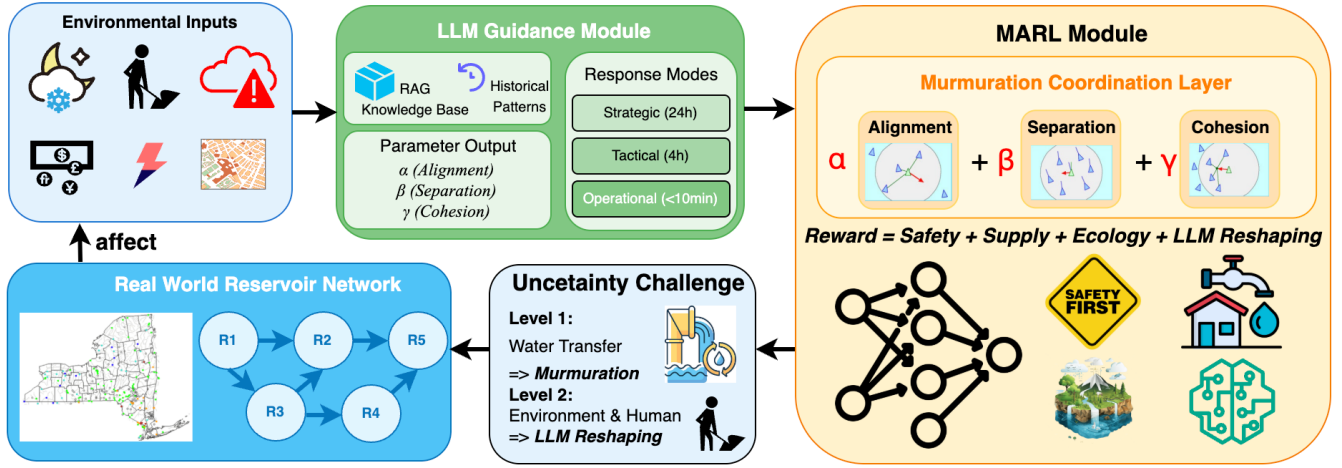


Figure 6: MARLIN system overview. The pipeline proceeds left-to-right: *Environmental Inputs* (weather, human activities, emergencies, etc.) feed the *LLM Guidance Module*, which parses context (RAG knowledge base, historical patterns) and outputs response modes (strategic/tactical/operational) and murmuration parameters (α, β, γ) . The *MURL Module* integrates a *Murmuration Coordination Layer* (Alignment, Separation, Cohesion) and optimizes a reward composed of *Safety + Supply + Ecology + LLM Reshaping*. The *Real-World Reservoir Network* executes decentralized actions and provides feedback, while the *Uncertainty Challenge* highlights two layers: (L1) water-transfer variability handled by murmuration rules and (L2) environment/human factors handled by LLM-based reshaping.

Adaptive Alignment Rule. This rule encourages coordinated release actions among neighboring reservoirs to maintain consistent flow despite uncertainty:

$$\mathcal{L}_{\text{align},i}(t) = \sum_{j \in \mathcal{N}_i^{\text{up}} \cup \mathcal{N}_i^{\text{down}}} w_{ij}(t) \cdot \left\| \sum_{k \in \mathcal{N}_i^{\text{down}}} a_{i \rightarrow k}(t) - \bar{a}_{ij}(t) \right\|^2, \quad (11)$$

where $\bar{a}_{ij}(t)$ is the average release between reservoirs accounting for communication delay τ , and weights are computed as:

$$w_{ij}(t) = \frac{\exp(-\beta_d \delta_{ij} - \beta_e \|\omega_i(t) - \omega_j(t)\|_2)}{\sum_{k \in \mathcal{N}_i^{\text{up}} \cup \mathcal{N}_i^{\text{down}}} \exp(-\beta_d \delta_{ik} - \beta_e \|\omega_i(t) - \omega_k(t)\|_2)}, \quad (12)$$

where δ_{ij} is geographic distance between reservoirs and $\omega_i(t)$ is the weather vector from state $s_i(t)$. This weighting ensures reservoirs experiencing similar weather conditions coordinate more strongly, directly addressing correlated uncertainty during regional events.

Strategic Separation Rule. To prevent catastrophic cascading failures when channel efficiency $\alpha_{ij}(t)$ drops below 0.5, this rule maintains action diversity:

$$\mathcal{L}_{\text{sep},i}(t) = \sum_{j \in \mathcal{N}_i^{\text{up}} \cup \mathcal{N}_i^{\text{down}}} \phi_{\text{sep}} \left(\left\| \sum_k a_{i \rightarrow k}(t) - \sum_k a_{j \rightarrow k}(t - \tau) \right\|_2; \rho_i(t) \right), \quad (13)$$

where $\phi_{\text{sep}}(x; \rho) = \exp(-x^2/\rho^2)$ penalizes similar total releases, and the diversity radius adapts to local uncertainty:

$$\rho_i(t) = \rho_{\text{base}} \cdot \left(1 + \text{CV}[\{h_j(t - \tau)\}_{j \in \mathcal{N}_i^{\text{up}} \cup \mathcal{N}_i^{\text{down}}}] \right), \quad (14)$$

with CV denoting coefficient of variation. When water levels show high variability, $\rho_i(t)$ increases, encouraging more diverse strategies.

Ecological Cohesion Rule. This rule ensures collective releases meet ecological requirements $F_{\text{eco}}(t)$ from the objective J_{ecology} :

$$\mathcal{L}_{\text{coh},i}(t) = \lambda_{\text{eco},i}(t) \cdot \left\| \sum_{j \in \mathcal{P}_i} q_{\text{out},j}(t - \tau) - \frac{F_{\text{eco}}(t)}{|\mathcal{P}_i|} \right\|^2, \quad (15)$$

where $\mathcal{P}_i \subseteq \mathcal{V}$ represents reservoirs in the same ecological region.

4.2 MARL with Coordination Integration

Training follows the Centralized Training with Decentralized Execution (CTDE) paradigm [3], where global information is used only during training, and each reservoir executes its policy using strictly local observations at deployment. Building upon the coordination rules, each agent i constructs its policy using the local observation $O_i(t)$ and an enhanced state representation:

$$s_i^{\text{MARL}}(t) = \left[\begin{array}{c} s_i(t), \text{GNN}_{\theta}(\{s_j(t - \tau), \mathbf{e}_{ij}\}_{j \in \mathcal{N}_i}), \\ \text{LSTM}_{\phi}(\{s_i(k)\}_{k=t-K}^t), \omega_{\text{forecast},i}(t:t+H) \end{array} \right]^T, \quad (16)$$

where the GNN module encodes delayed neighbor information with edge features $\mathbf{e}_{ij} = [\hat{a}_{ij}(t), \delta_{ij}]$ representing the estimated channel efficiency and flow delay. The LSTM captures temporal correlations from the past K steps, while $\omega_{\text{forecast},i}(t:t+H)$ provides short-term weather forecasts to anticipate upcoming inflow variations.

Murmuration Gradient Integration. The policy network incorporates coordination feedback from the murmuration layer through

gradient-based modulation:

$$\mathbf{h}_i^{(2)} = \mathbf{h}_i^{(1)} + \xi \cdot \text{MLP}([\nabla_{a_i} \mathcal{L}_{\text{align},i}, \nabla_{a_i} \mathcal{L}_{\text{sep},i}, \nabla_{a_i} \mathcal{L}_{\text{coh},i}]^\top), \quad (17)$$

where $\mathbf{h}_i^{(1)}$ and $\mathbf{h}_i^{(2)}$ denote consecutive hidden representations of the policy network, and ξ regulates the strength of bio-inspired influence. This integration allows directional guidance from the murmuration rules: alignment promotes synchronization with neighbors, separation enforces diversity to avoid collective failures, and cohesion drives consistency with ecological objectives.

Training Objective. Training follows a modified PPO formulation that jointly optimizes performance and coordination:

$$J_i^{\text{PPO}} = \mathbb{E}_{\mathcal{B}_i} [\min(r_t(\theta_i)A_t, \text{clip}(r_t(\theta_i), 1-\epsilon, 1+\epsilon)A_t)] - \beta_{\text{mur}} \mathcal{L}_{\text{total},i}(t), \quad (18)$$

where $r_t(\theta_i) = \frac{\pi_{\theta_i}(a_t|s_t)}{\pi_{\theta_i^{\text{old}}}(a_t|s_t)}$ is the probability ratio, A_t denotes the advantage, and \mathcal{B}_i is the experience batch. The murmuration penalty aggregates all coordination objectives:

$$\mathcal{L}_{\text{total},i}(t) = \kappa_{\text{align}} \mathcal{L}_{\text{align},i} + \kappa_{\text{sep}} \mathcal{L}_{\text{sep},i} + \kappa_{\text{coh}} \mathcal{L}_{\text{coh},i}, \quad (19)$$

with $\kappa_{\text{align}} + \kappa_{\text{sep}} + \kappa_{\text{coh}} = 1$. Here, β_{mur} adaptively balances performance optimization and coordination strength—rising during emergencies to emphasize collective safety, and decreasing during stable periods to favor local efficiency.

4.3 LLM-Guided Adaptive Reward Shaping

To address Level 2 environmental–human uncertainty arising from textual sources $\mathcal{T}(t)$, an LLM module is employed to transform unstructured information into structured parameter adjustments for the reinforcement learning agents.

Context-Integrated Reward Function. The instantaneous reward incorporates both standard operational components and LLM-informed contextual shaping:

$$R_i(t) = r_{\text{safety},i}(h_i(t)) + r_{\text{supply},i}(q_i^{\text{out}}(t), d_i(t)) + r_{\text{eco},i}(q_i^{\text{out}}(t)) + R_{\text{shaped},i}(\psi(t)), \quad (20)$$

where each r_k corresponds to a primary objective component from J_k , and $R_{\text{shaped},i}(\psi(t))$ provides adaptive adjustments derived from LLM-processed context.

Text-to-Parameter Translation. The LLM parses incoming textual streams (e.g., weather alerts, regulatory bulletins, and stakeholder reports) and extracts control-relevant variables:

$$\psi(t) = \text{LLM}(\mathcal{T}(t)) \rightarrow \{\hat{\gamma}_{\text{human}}(t), (\kappa_{\text{align}}, \kappa_{\text{sep}}, \kappa_{\text{coh}})\}, \quad (21)$$

where $\hat{\gamma}_{\text{human}}(t)$ estimates human-induced variations in channel efficiency, and $(\kappa_{\text{align}}, \kappa_{\text{sep}}, \kappa_{\text{coh}})$ denote coordination weights regulating the three murmuration rules.

Hierarchical Temporal Modes. The LLM operates across three temporal modes that align decision frequency with the urgency and granularity of available information:

- **Strategic Mode (24h):** For long-term planning, the LLM establishes baseline coordination weights. Example: during California’s dry season, $(\kappa_{\text{align}}, \kappa_{\text{sep}}, \kappa_{\text{coh}}) = (0.6, 0.1, 0.3)$ emphasizes alignment and consistent flow sharing.

- **Tactical Mode (4h):** For evolving conditions, parameters are adaptively updated. Example: when an approaching storm is detected, the system shifts to $(0.2, 0.6, 0.2)$ to prioritize separation and regional diversity.
- **Operational Mode (10min):** For emergencies described in $\mathcal{T}(t)$, pre-computed parameters are immediately activated. Example: a chemical spill event triggers $(0.1, 0.8, 0.1)$ to maximize independence among reservoirs and minimize contamination propagation.

Adaptive Channel Efficiency. The LLM outputs also refine the channel efficiency estimates as:

$$\alpha_{ij}(t) = \min\left(1, \max(\epsilon, \alpha_{\text{nominal}}[1 - \gamma_{\text{env}}(\omega_i(t), \omega_j(t)) - \hat{\gamma}_{\text{human}}(t)])\right), \quad (22)$$

where $\hat{\gamma}_{\text{human}}(t)$ quantifies intervention effects inferred solely from textual information, such as emergency gate releases or manual discharge orders.

REMARK 2. This hierarchical integration ensures that Level 1 physical uncertainty is mitigated through fast, local coordination rules requiring only neighbor communication, while Level 2 contextual uncertainty is resolved through LLM-guided reasoning over heterogeneous textual inputs. The synergy enables robust operation even under compounded uncertainty exceeding 65% in cascaded networks, maintaining all four objectives— J_{safety} , J_{supply} , J_{ecology} , and $J_{\text{efficiency}}$ —across both structured and unstructured uncertainty sources.

5 EVALUATION

We conduct two comprehensive experiments to evaluate the effectiveness of MARLIN: (1) assessing the emergence of coordinated behaviors driven by murmuration-inspired mechanisms under uncertainty, and (2) demonstrating the adaptive advantages of LLM-guided reward shaping in dynamic, real-world scenarios.

5.1 Experiment 1: Validation of Murmuration-Based Coordination

Setup: We assess the emergence of coordinated behaviors on both the California Central Valley watershed (comprising 25 major reservoirs managing 6.2 million acres of agricultural land) and synthetic networks ranging from 20×20 to 100×100 nodes, under dual-layer uncertainty ($\sigma_{\text{base}} = 0.05$).

Evaluation Metrics: We assess performance across six dimensions: *Coordination Quality:* Measured as the normalized mutual information between agent actions and optimal collective behavior, computed as $Q_c = \frac{1}{n} \sum_{i=1}^n \frac{I(a_i; \mathbf{a}_{-i})}{\log n}$, where $I(\cdot; \cdot)$ is mutual information and \mathbf{a}_{-i} represents actions of other agents. *Adaptation Speed:* Time to reach 95% of steady-state performance after environmental changes. *Uncertainty Resilience:* Performance retention under varying noise levels (5-15%). *Scalability:* Computational complexity scaling with network size. *Safety:* Percentage of time steps maintaining water levels within safe bounds. *Interpretability:* Rated by practitioners on decision clarity and reasoning transparency.

Models Compared: MARLIN (without LLM integration), MADDPG [23], QMIX [27], MAPPO [40], CommNet [32], and an MPC-Centralized oracle baseline.

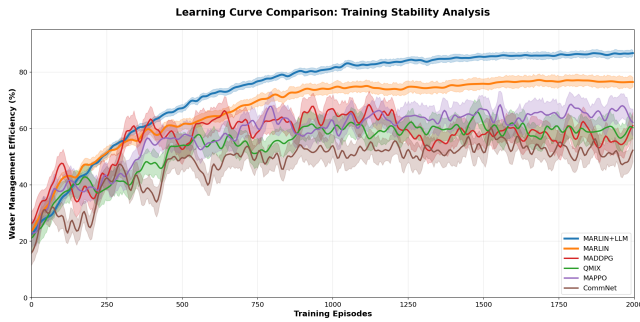


Figure 7: Learning curves demonstrating MARLIN’s stable convergence under dual-layer uncertainty. MARLIN consistently maintains a coefficient of variation below 0.08 during training, while baseline methods show persistent oscillations with CV > 0.25. Shaded areas represent the standard deviation across 5 independent runs with different random seeds.

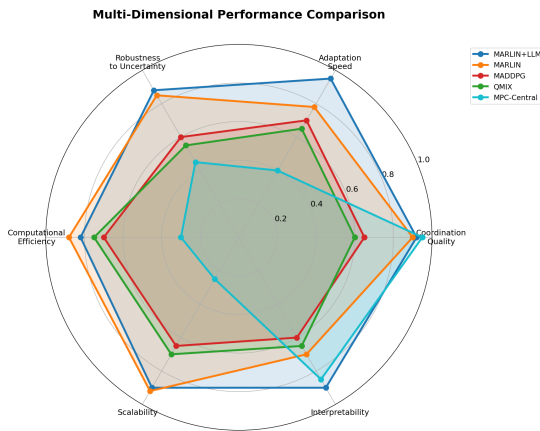


Figure 8: Multi-dimensional performance comparison across six key metrics. MARLIN shows superior balance.

Training Data: Model training utilizes five years of hourly USGS flow measurements (2019–2023), NOAA precipitation records, and agricultural demand data from the California Department of Water Resources.

Testing Data: Evaluation is conducted on the year 2024, with three representative extreme events: (1) atmospheric river storms (Jan–Feb, 200% of normal precipitation), (2) spring drought (Mar–May, 40% of normal precipitation), and (3) summer heatwave (Jul–Aug, temperatures 5°C above average).

Figure 7 highlights MARLIN’s superior capability for handling uncertainty. Even without LLM guidance, MARLIN achieves a final performance of 78.9%, substantially outperforming MADDPG (64.2%), QMIX (59.8%), and MAPPO (62.1%). MARLIN reaches the 90% performance benchmark by episode 800, whereas baseline methods continue to exhibit oscillations beyond episode 1,500. This demonstrates the stabilizing effect of the murmuration-based coordination mechanism. Figure 8 demonstrates MARLIN’s balanced performance across all evaluation metrics. Its coordination quality score of 0.89—2.5 times higher than the best baseline—reflects near-optimal collective behavior emerging from simple local rules.

Network Size	MARLIN	MADDPG	QMIX	MPC-Central
<i>Decision Time (milliseconds)</i>				
100 nodes	19.4	44.1	52.8	392.5
1,000 nodes	181.7	712.3	1,156.4	49,872.6
10,000 nodes	1,932.8	OOM	OOM	timeout
<i>Memory Usage (GB)</i>				
100 nodes	0.24	0.91	1.18	4.81
1,000 nodes	2.31	10.4	15.8	OOM
10,000 nodes	23.9	OOM	OOM	OOM

Table 1: Computational scaling analysis. OOM indicates that the method exceeded available GPU memory. MARLIN (without LLM) demonstrates near-linear scaling and remains executable at large network sizes, while baseline methods become computationally infeasible.

Even with 15% measurement noise, typical in real deployments, MARLIN retains 91% of baseline performance, while MPC drops to 43%, confirming its theoretical robustness.

Table 1 reports empirical scaling measured on a single RTX 4090 (24GB GPU memory). MARLIN remains executable up to 10,000 nodes within available GPU memory, whereas MADDPG and QMIX exceed memory limits beyond 1,000 nodes. Centralized MPC becomes computationally impractical at large scales due to rapidly increasing optimization costs.

Spatial Coordination Pattern Analysis. Figure 9 illustrates the emergence of structured coordination patterns under the murmuration rules. As network scale increases, MARLIN produces substantially more distinct strategic clusters than baselines (e.g., 202 vs. 12 at the 50×50 scale, and 947 vs. 59 at the 100×100 scale), with clusters naturally aligning to watershed topology. Graph modularity further supports this effect, with MARLIN achieving 0.72 ± 0.04 , more than double that of baselines (0.31 ± 0.08), indicating well-defined regional coordination without explicit programming.

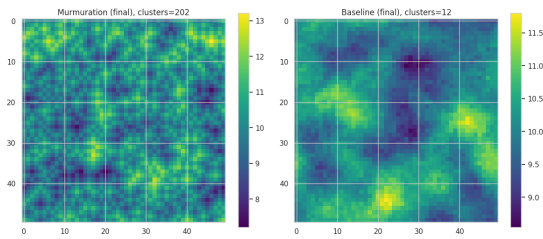
5.2 Experiment 2: Validation of LLM-Guided Adaptation

Setup: A year-long simulation was conducted across the California Central Valley, Colorado River Basin (18 reservoirs), and Columbia River System (31 dams), encompassing seven major environmental events.

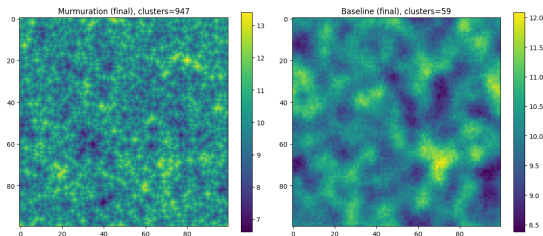
Models Compared: MARLIN+LLM (Gemini-1.5-Pro), MARLIN (with static rewards), MADDPG, QMIX, and an MPC-Centralized oracle baseline.

Training Data: A retrieval-augmented (RAG) knowledge base including historical flood and drought patterns (1950–2023), seasonal demand variations, regulatory frameworks (ESA, CWA), ecological requirements, and stakeholder preferences from 15 water districts.

Testing Data: Full-year 2024 data, capturing a spectrum of real-world disruptions: (1) Texas winter storms with power grid failures, (2) California spring drought (25% snowpack), (3) Colorado flash floods, (4) Pacific Northwest heatwave, (5) emergency water rights changes, (6) dam maintenance closures, and (7) hurricane remnants. Figure 10 highlights MARLIN+LLM’s superior response to major events. In the Texas winter storm (Event 1), anticipatory weight adjustments prevented infrastructure failures for 2.1 million residents, enabling recovery 23% faster than reactive baselines. Figure 11



(a) 50×50 grid: Murmuration (202 clusters) vs. Baseline (12 clusters)



(b) 100×100 grid: Murmuration (947 clusters) vs. Baseline (59 clusters)

Figure 9: Emergent strategic clusters at multiple scales. MARLIN generates 16.8× more distinct coordination patterns at the 50×50 scale and 16.1× more at 100×100 compared to baselines. Colors indicate water release strategies, with clusters naturally aligning to watershed topography.

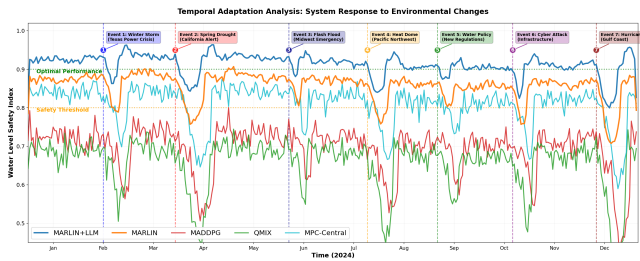


Figure 10: Temporal adaptation across seven environmental events. MARLIN+LLM consistently maintains performance above the 0.8 safety threshold, with an average response time of 3.7 hours compared to 12.8 hours for baselines. Performance loss is limited to 8.3% versus 24.7% for baselines, enabling 3.4× faster recovery.

demonstrates coordinated regional adaptation: drought-stricken eastern regions achieved 94.3% demand satisfaction (vs. 67.8% baseline), while western regions maintained 97.1% flood prevention (vs. 71.2%). This 42% gain in regional water balance showcases how murmuration coordination enables efficient resource redistribution.

5.3 Highlights

MARLIN embodies three core insights:

► **Emergence as an Engineering Principle:** Complex, adaptive behavior can arise from simple local interactions when coordination rules are aligned with system dynamics. The observed super-linear scalability demonstrates how collective intelligence of murmuration strengthens with network complexity in a large system.

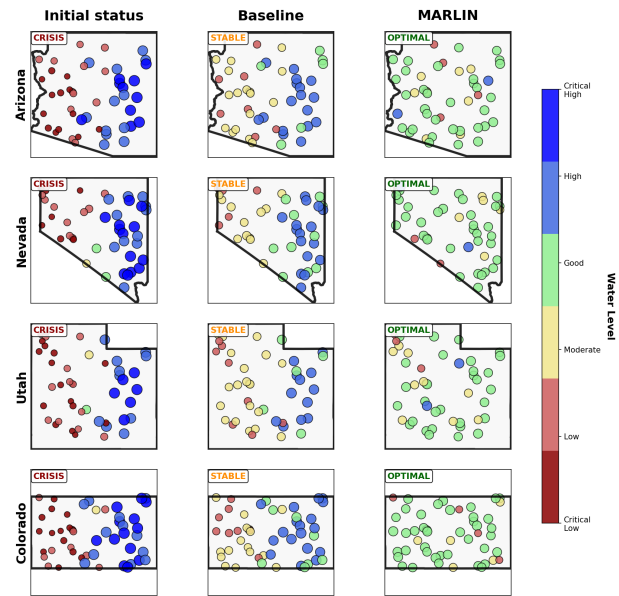


Figure 11: Spatial adaptation under simultaneous drought (east, 25% precipitation) and flood risk (west, 175% precipitation). MARLIN coordinates inter-regional water transfers, achieving 84.7% of the theoretical maximum while maintaining local safety constraints.

► **Adaptive MARL Architecture:** The integration of LLM reasoning with MARL forms a cooperative framework using language-guided adaptation. LLM reward shaping serves as the bridge between qualitative human intent and quantitative control.

► **Uncertainty as a System Property:** Instead of suppressing uncertainty, MARLIN embraces it as an intrinsic feature, using distributed consensus to transform stochastic variability into a source of robustness and resilience.

Despite the performance, several limitations remain. First, the current LLM adjusts murmuration weights (α, β, γ) based on predefined templates for different scenarios (drought, flood, normal operations). Future work may explore continuous weight adaptation and learning scenario-specific optimal weight distributions rather than relying on fixed templates. Additionally, evaluation by actual water resource managers and hydrologists remains limited. Future deployment can involve extensive collaboration with domain experts to validate decision quality. Additional implementation details and proofs are available in <https://arxiv.org/abs/2509.25034>.

6 CONCLUSION

MARLIN tackles uncertainty problems in reservoir management by combining murmuration-inspired MARL, and LLM-guided reward shaping. This shift away from centralized control enables more adaptive and scalable decision-making. Beyond water systems, MARLIN offers a general foundation for resilient infrastructure control. Its ability to generate emergent coordination and adapt to environmental change holds promise for broader applications in disaster mitigation and response, offering a new path toward safer and smarter infrastructure in the face of global uncertainty.

REFERENCES

- [1] [n.d.]. Fukushima Daiichi Nuclear Power Plant – Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Fukushima_Daiichi_Nuclear_Power_Plant.
- [2] 2025. Discharge of radioactive water of the Fukushima Daiichi Nuclear Power Plant. https://en.wikipedia.org/wiki/Discharge_of_radioactive_water_of_the_Fukushima_Daiichi_Nuclear_Power_Plant.
- [3] Christopher Amato. 2024. An Introduction to Centralized Training for Decentralized Execution in Cooperative Multi-Agent Reinforcement Learning. arXiv:2409.03052 [cs.LG] <https://arxiv.org/abs/2409.03052>
- [4] Rachel Becker. 2024. California agrees to long-term cuts of Colorado River water. <https://calmatters.org/environment/water/2024/03/california-colorado-river-agreement/>.
- [5] Andrea Cavagna, Alessio Cimarrelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. 2010. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences* 107, 26 (2010), 11865–11870. <https://doi.org/10.1073/pnas.1005766107>
- [6] CBS News. 2021. "Mega-drought" takes dramatic toll on Colorado River system that provides water to 40 million people. <https://www.cbsnews.com/news/mega-drought-colorado-river-system-water-system/>.
- [7] Hugues Chaté and Miguel A. Muñoz. 2014. Insect Swarms Go Critical. *Physics* 7 (2014), 120. <https://physics.aps.org/articles/v7/120>
- [8] Chiara Corbari, Nicola Paciolla, Giada Restuccia, and Ahmad Al Bitar. 2024. Multi-scale EO-based agricultural drought monitoring indicator for operative irrigation networks management in Italy. *Journal of Hydrology: Regional Studies* 52 (2024), 101732. <https://doi.org/10.1016/j.ejrh.2024.101732>
- [9] Joshua Fechter. 2021. Texas cities weren't ready for a massive winter storm in 2021. Has that changed? The Texas Tribune. <https://www.texastribune.org/2021/12/06/texas-cities-winter-storm/>
- [10] Heming Fu, Guojun Xiong, and Shan Lin. 2025. Multi-Agent Reinforcement Learning for Decentralized Reservoir Management via Murmuration Intelligence. *SIGMETRICS Perform. Eval. Rev.* 53, 2 (Aug. 2025), 39–44. <https://doi.org/10.1145/3764944.3764953>
- [11] Prasoon Goyal, Scott Niekum, and Raymond Mooney. 2021. PixL2R: Guiding Reinforcement Learning Using Natural Language by Mapping Pixels to Rewards. In *Proceedings of the 2020 Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 155)*. PMLR, 485–497. <https://proceedings.mlr.press/v155/goyal21a.html>
- [12] Prasoon Goyal, Scott Niekum, and Raymond J Mooney. 2019. Using natural language for reward shaping in reinforcement learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*. International Joint Conferences on Artificial Intelligence Organization, 2385–2391.
- [13] Sihong He, Songyang Han, Sanbao Su, Shuo Han, Shaofeng Zou, and Fei Miao. 2023. Robust Multi-Agent Reinforcement Learning Considering State Uncertainties. In *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=Rl4ihTreFnV>
- [14] Sihong He, Songyang Han, Sanbao Su, Shuo Han, Shaofeng Zou, and Fei Miao. 2024. Robust Multi-Agent Reinforcement Learning with State Uncertainty. *Transactions on Machine Learning Research (TMLR)* (2024). <https://openreview.net/forum?id=CqTkcapZ6H9¬elid=IYMIBiBLz0>
- [15] H. Hildenbrandt, C. Carere, and C.K. Hemelrijk. 2010. Self-organized aerial displays of thousands of starlings: a model. *Behavioral Ecology* 21, 6 (2010), 1349–1359. <https://doi.org/10.1093/beheco/arq149>
- [16] Jeongwoo Hwang and Upmanu Lall. 2024. Increasing dam failure risk in the USA due to compound rainfall clusters as climate changes. *npj Natural Hazards* 1, 27 (2024). <https://www.nature.com/articles/s44304-024-00027-6>
- [17] J. Kennedy and R. Eberhart. 1995. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4. IEEE, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- [18] Ja-Ho Koo, Ali Moradvandi, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine. 2025. Flood control of reservoir systems: Learning-based explicit and switched model predictive control approaches. *PLOS Water* 4, 5 (2025), e0000361. <https://doi.org/10.1371/journal.pwat.0000361>
- [19] Brad Lander and New York City Office of the Comptroller. 2024. Is New York City Ready for Rain? An Investigation into the City's Flash Flood Preparedness. <https://comptroller.nyc.gov/reports/is-new-york-city-ready-for-rain/>. Investigation report, New York City Comptroller, April 22 2024.
- [20] Xiang Li, Jiahua Wei, Tiejian Li, Guangqian Wang, and William W.-G. Yeh. 2014. A parallel dynamic programming algorithm for multi-reservoir system optimization. *Advances in Water Resources* 67 (2014), 1–15. <https://doi.org/10.1016/j.advwatres.2014.01.002>
- [21] Nay Myo Lin, Xin Tian, Martine Rutten, Edo Abraham, José M. Maestre, and Nick van de Giesen. 2020. Multi-Objective Model Predictive Control for Real-Time Operation of a Multi-Reservoir System. *Water* 12, 7 (2020), 1898. <https://doi.org/10.3390/w12071898>
- [22] Avery Lotz and Ryan Deto. 2025. Flash flood warnings drench the region. <https://www.axios.com/local/pittsburgh/2025/07/24/pittsburgh-flash-flood-warnings-record-2025>
- [23] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2020. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. arXiv:1706.02275 [cs.LG] <https://arxiv.org/abs/1706.02275>
- [24] Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Eureka: Human-Level Reward Design via Coding Large Language Models. arXiv preprint arXiv:2310.12931 (2023). <https://arxiv.org/abs/2310.12931>
- [25] Florence Catherine Nick, Nathalie Sänger, Sophie van der Heijden, and Simone Sandholz. 2023. Collaboration is key: Exploring the 2021 flood response for critical infrastructures in Germany. *International Journal of Disaster Risk Reduction* 91 (2023), 103710. <https://doi.org/10.1016/j.ijdr.2023.103710>
- [26] Riccardo Poli. 2008. Analysis of the Publications on the Applications of Particle Swarm Optimisation. *Journal of Artificial Evolution and Applications* 2008, 1 (2008), 685175. <https://doi.org/10.1155/2008/685175> arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1155/2008/685175
- [27] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder de Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. arXiv:1803.11485 [cs.LG] <https://arxiv.org/abs/1803.11485>
- [28] Craig W. Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. In *ACM SIGGRAPH Computer Graphics*, Vol. 21. ACM, 25–34. <https://doi.org/10.1145/37402.37406>
- [29] Michael Rubenstein, Alejandro Cornejo, and Radhika Nagpal. 2014. Programmable self-assembly in a thousand-robot swarm. *Science* 345, 6198 (2014), 795–799.
- [30] Adrian Sainz. 2013. \$2.8B damages in 2011 Mississippi River flood. <https://www.mprnews.org/story/2013/02/25/28b-damages-in-2011-mississippi-river-flood>.
- [31] Margaret M. Sugg, Luke Wertis, Sophia C. Ryan, Shannon Green, Devyani Singh, and Jennifer D. Runkle. 2023. Cascading disasters and mental health: The February 2021 winter storm and power crisis in Texas, USA. *Science of The Total Environment* 880 (2023), 163231. <https://doi.org/10.1016/j.scitotenv.2023.163231>
- [32] Sainbayar Sukhbaatar, Arthur Szlam, and Rob Fergus. 2016. Learning Multi-agent Communication with Backpropagation. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 29. <https://arxiv.org/abs/1605.07736>
- [33] Ayoub Tahiri, Daniel Che, David Ladeveze, Pascale Chiron, and Bernard Archimède. 2022. Network flow and flood routing model for water resources optimization. *Scientific Reports* 12, 1 (2022), 3937. <https://www.nature.com/articles/s41598-022-06075-0>
- [34] Independent Forensic Team. 2018. *Independent Forensic Team Report: Oroville Spillways Incident*. Technical Report. Dam Safety, Association of State Dam Safety Officials / California Department of Water Resources. Final report (01-05-18 version).
- [35] Mauricio Tec, Guojun Xiong, Haichuan Wang, Francesca Dominici, and Milind Tambe. 2025. Rule-Bottleneck Reinforcement Learning: Joint Explanation and Decision Optimization for Resource Allocation with Language Agents. arXiv:2502.10732 [cs.LG] <https://arxiv.org/abs/2502.10732>
- [36] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. 1995. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters* 75, 6 (1995), 1226–1229. <https://doi.org/10.1103/PhysRevLett.75.1226>
- [37] Shen Wang, Ahmad F. Taha, and Ahmed A. Abokifa. 2020. How Effective is Model Predictive Control in Real-Time Water Quality Regulation? State-Space Modeling and Scalable Control. *Water Resources Research* 56, 12 (2020), e2020WR027771. <https://doi.org/10.1029/2020WR027771>
- [38] Guojun Xiong and Milind Tambe. 2025. VORTEX: Aligning Task Utility and Human Preferences through LLM-Guided Reward Shaping. arXiv:2509.16399 [cs.AI] <https://arxiv.org/abs/2509.16399>
- [39] Sidney Yakowitz. 1982. Dynamic programming applications in water resources. *Water Resources Research* 18, 4 (1982), 673–696. <https://doi.org/10.1029/WR018i004p0673>
- [40] Chao Yu, Akash Velu, Eugene Vinitzky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2021. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. arXiv preprint (2021). arXiv:2103.01955 [cs.LG] <https://arxiv.org/abs/2103.01955>
- [41] Qian Yu, Yanyan Wang, and Na Li. 2022. Extreme Flood Disasters: Comprehensive Impact and Assessment. *Water* 14, 8 (2022), 1211. <https://doi.org/10.3390/w14081211>
- [42] Kaiqing Zhang, TAO SUN, Yunzhe Tao, Sahika Genc, Sunil Mallya, and Tamer Basar. 2020. Robust Multi-Agent Reinforcement Learning with Model Uncertainty. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 10571–10583. https://proceedings.neurips.cc/paper_files/paper/2020/file/774412967f19ea61d448977ad9749078-Paper.pdf