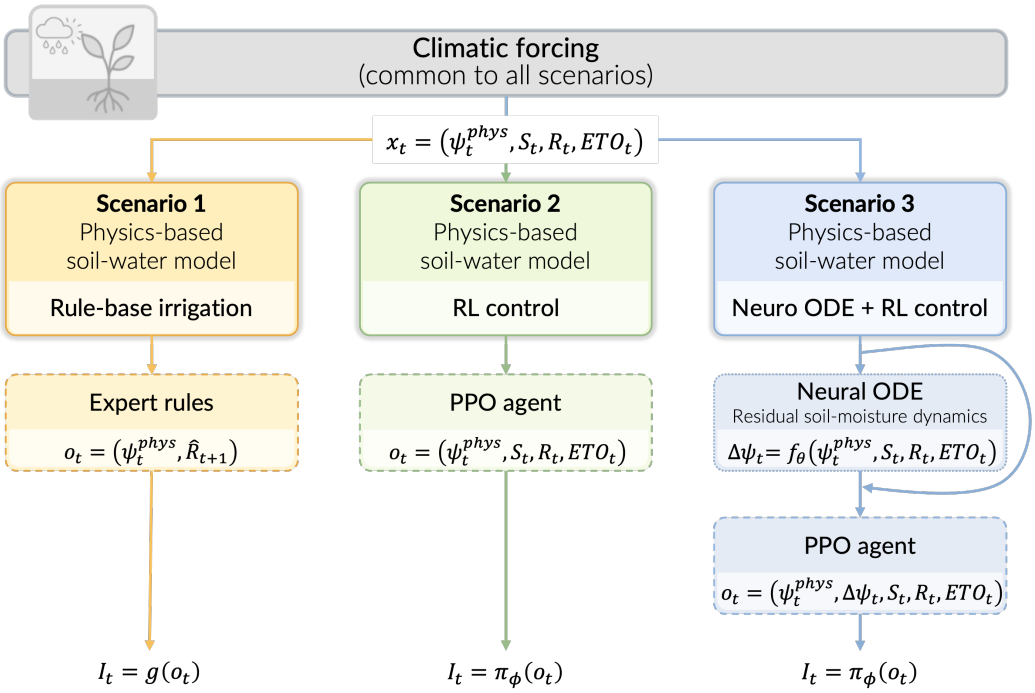
Graphical Abstract

Control-aware physics-informed reinforcement learning for adaptive irrigation under climatic uncertainty

Raymond Houé Ngouna, Philippe Berton, Fabien Dauriac



Highlights

Control-aware physics-informed reinforcement learning for adaptive irrigation under climatic uncertainty

Raymond Houé Ngouna, Philippe Berton, Fabien Dauriac

- A physics-informed reinforcement learning framework is proposed for irrigation control.
- Rule-based, reinforcement learning, and hybrid control strategies are compared under identical conditions.
- A Neural ODE-inspired residual model is integrated to correct daily prediction errors.
- Hybrid control mitigates extreme stress while preserving water-use efficiency within the tested soil and climatic parameter settings.

Control-aware physics-informed reinforcement learning for adaptive irrigation under climatic uncertainty

Raymond Houé Ngouna[a], Philippe Berton[b], Fabien Dauriac[b]

[a] *Université de Technologie de Tarbes Occitanie Pyrénées, Laboratoire Génie de Production, 47 Avenue d'Azereix, Tarbes, 65000, France*
[b] *Rives et Eaux du Sud-Ouest, BP 449 Chemin de Lalette, Tarbes, 65000, France*

Abstract

Effective irrigation management under increasingly variable climatic conditions requires control strategies that are both adaptive and grounded in physical principles. While physics-based models provide interpretability and internal consistency, they often depend on structural simplifications. In contrast, data-driven methods such as reinforcement learning address uncertainty but frequently lack physical guarantees. This study presents a physics-informed reinforcement learning framework for irrigation control under climatic uncertainty, integrating a process-based soil–water balance model with learning-based components for decision-making and system dynamics. Three irrigation strategies are systematically compared within a unified simulated environment: (i) a rule-based baseline informed by expert heuristics, (ii) a reinforcement learning controller utilizing Proximal Policy Optimization (PPO) that interacts directly with the physical model, and (iii) a hybrid approach in which the physical model is augmented by a Neural ODE–inspired residual model, implemented in discrete time to correct daily prediction errors. In the hybrid configuration, the learned residual is also provided to the controller, enabling decisions that explicitly account for model error. All strategies are evaluated under identical stochastic climatic inputs and soil properties using a controlled experimental protocol to ensure fair comparison. Results indicate that reinforcement learning improves adaptability relative to rule-based control, while the hybrid neuro-physical approach further reduces systematic stress biases and extreme events without increasing overall water consumption. The proposed framework elucidates the strengths and limitations of combining physical modeling, learning-based control, and residual dynamics correction for irrigation under uncertainty, and establishes a transparent

foundation for future extensions toward more realistic sensing and field-scale deployment. Notably, the soil–water formulation is designed to be control-aware rather than purely descriptive, prioritizing numerical robustness, interpretability, and suitability for closed-loop learning over exhaustive physical detail.

Keywords:
Intelligent irrigation, Environmental decision support systems, Control-aware modelling, Physics-informed reinforcement learning, Hybrid neuro-physical models, Adaptive control under uncertainty

---

1. Introduction

Climate change is intensifying the frequency and severity of hydro-climatic fluctuations, thereby increasing pressure on agricultural water management systems. Irrigation, which accounts for a substantial share of global freshwater withdrawals, is particularly vulnerable to rainfall variability, seasonal variability, and complex soil–plant interactions. Conventional irrigation practices, often based on fixed schedules or expert-defined thresholds, generally lack the flexibility to respond to non-stationary environmental conditions. As a result, these methods may lead to inefficient water use, greater drainage losses, and heightened crop water stress.

From a modelling perspective, irrigation management represents a dynamic environmental control problem in which decisions are made sequentially under uncertainty and must comply with physical constraints. Process-based soil–water models have traditionally been employed to represent irrigation processes and inform decision-making, providing interpretability and physical consistency through simplified frameworks such as bucket models or FAO-56-type formulations (Raes et al., 2009). However, these models depend on structural assumptions and parameterisations that may insufficiently capture context-specific dynamics, spatial heterogeneity in soils, or the cumulative effects of stochastic climatic influences.

Recent advances in reinforcement learning (RL) have enabled the development of adaptive control policies that leverage temporal dependencies and accommodate delayed system responses in complex, stochastic environments (Sutton and Barto, 2018). RL-based methods are increasingly applied to irrigation and water management, demonstrating potential improvements in water-use efficiency relative to fixed rule-based strategies. Reinforcement

learning is particularly suitable for daily, closed-loop decision-making under delayed feedback, whereas policy search and multi-objective evolutionary algorithms are more commonly used for fixed or seasonal irrigation policies. However, fully data-driven control methods continue to face challenges regarding interpretability, adherence to physical laws, and operational safety. These issues are especially significant in environmental systems, where actions may have delayed and potentially irreversible consequences.

Physics-informed learning provides a systematic framework for constraining data-driven models using established structural properties of the system, which enhances both stability and out-of-sample performance (Willard et al., 2022). Within this context, Neural Ordinary Differential Equations (Neural ODEs) have become a versatile method for modeling residual dynamics that augment existing physical models while preserving interpretability (Rackauckas et al., 2020). In environmental applications, such hybrid models are increasingly viewed not as replacements for process-based formulations, but as corrective mechanisms that address systematic model errors arising from simplifications, parameter uncertainty, or unresolved processes.

Many operational irrigation systems function on discrete daily decision cycles and depend on sensor-level observations, including soil matric potential and rainfall measurements. Therefore, effective integration of learning-based methods necessitates alignment among control decisions, temporal discretization, and physical modeling assumptions. In this study, learning is incorporated at two levels: (i) at the policy level, using reinforcement learning with continuous irrigation actions, and (ii) at the dynamics level, employing a discrete-time residual neural model inspired by Neural ODE formulations. This distinction supports adaptive decision-making while preserving a physically interpretable and computationally efficient simulation framework that is consistent with available data and operational constraints.

Although interest in learning-based irrigation control is growing, several open questions persist. These include the comparative advantages of reinforcement learning versus expert rule-based strategies when interacting with simplified physical models, the degree to which learned residual dynamics can address structural model mismatch, and the effects of closer model-controller integration on interpretability, transferability, and practical deployment. Addressing these issues necessitates controlled comparative studies that systematically examine modeling assumptions, parameter configurations, and performance trade-offs under consistent physical and climatic conditions.

The Environmental Modelling & Software community increasingly em-

phasizes transparent model design, explicit uncertainty management, and reproducible evaluation workflows (Refsgaard et al., 2007; Jakeman et al., 2006). This study adheres to these principles by employing a configuration-driven experimental framework, systematically varying soil and meteorological parameters, and distinctly separating training and evaluation phases for learning-based controllers.

Accordingly, this study investigates the following research questions:

- **RQ1:** To what extent does reinforcement learning enhance irrigation control performance relative to expert rule-based strategies when applied to a physics-based soil-water model?
- **RQ2:** To what extent does augmenting the physical model with learned residual dynamics increase the stability and safety of learning-based irrigation control under stochastic climatic forcing?
- **RQ3:** What are the effects of increased model–controller integration on interpretability, reproducibility, and transferability in environmental decision-support systems?

The main contributions of this study are as follows:

1. A physics-informed reinforcement learning framework that formulates irrigation as a finite-horizon sequential decision-making problem under climatic uncertainty, explicitly distinguishing latent physical states from sensor-level observations.
2. A structured and reproducible comparison of three irrigation control paradigms: rule-based control, reinforcement learning, and hybrid neuro-physical reinforcement learning, conducted within a unified physical simulation environment.
3. The integration of a discrete-time Neural ODE-inspired residual model to correct simplified soil-water balance dynamics, thereby preserving physical consistency and interpretability.
4. A systematic evaluation under stochastic rainfall and evapotranspiration forcing, which highlights trade-offs among water-use efficiency, stress avoidance, and drainage losses as functions of soil and meteorological parameterisation.
5. A transferable modelling and control framework intended to facilitate future integration of real sensor data and extension to other environmental systems that require adaptive control under uncertainty.

The structure of this paper is as follows. Section 2 reviews research on irrigation modelling, rule-based control, reinforcement learning, and hybrid physics-informed approaches. Section 3 details the materials and methods, including the problem formulation, simulated data sources, the physical environment, control scenarios, and the experimental design. Section 4 presents and discusses the results for each scenario, with particular attention to parameter sensitivity and performance trade-offs. Section 5 concludes the paper and suggests directions for future research.

2. Related Work

This section reviews the relevant literature on irrigation modelling and control, focusing on the transition from process- and rule-based approaches to learning-based and hybrid physics-informed methods. The aim was to position the present work within established environmental modelling paradigms and identify the methodological gaps addressed by the proposed framework.

2.1. Process-based modelling of irrigation systems

Process-based models are fundamental to irrigation modelling, offering mechanistic representations of soil-water-plant interactions grounded in mass balance principles and evapotranspiration formulations. Classical approaches include conceptual bucket models and FAO-56-based methods, as implemented in decision-support tools such as AquaCrop (Raes et al., 2009). These models provide interpretability, transparency, and relatively low data requirements, which are critical for scenario analysis and policy-oriented research.

However, process-based models depend on simplified assumptions about soil heterogeneity, root water uptake, and boundary conditions. Their performance often declines under highly variable or extreme climatic conditions, where unmodeled dynamics and parameter uncertainty become significant. Recent reviews indicate that, while physically based models are essential, their use alone may be inadequate for adaptive irrigation management in the context of climate change (Fatichi et al., 2016; Seneviratne et al., 2021).

2.2. Rule-based and heuristic irrigation control

Rule-based irrigation strategies, frequently derived from expert knowledge or agronomic guidelines, continue to be widely implemented. These methods generally use fixed thresholds for soil moisture, crop water stress indices, or accumulated evapotranspiration deficits. Their primary advantages are

simplicity, interpretability, and ease of deployment, especially in data-limited settings.

Heuristic controllers are inherently static and often fail to adapt to non-stationary climatic conditions. Recent comparative studies demonstrate that threshold-based rules may perform suboptimally when rainfall variability or delayed soil responses are pronounced, thereby motivating the investigation of adaptive control strategies (Jones et al., 2022). As a result, rule-based control is increasingly considered a baseline rather than a viable long-term solution for climate-resilient irrigation.

2.3. Reinforcement learning for irrigation and water management

Reinforcement learning (RL) provides a flexible framework for sequential decision-making in uncertain and dynamic environments (Sutton and Barto, 2018). In the context of water management, RL has been applied to reservoir operation, canal regulation, and irrigation scheduling, with several studies reporting improved performance compared to static policies (Yang et al., 2021; Giuliani et al., 2021).

Despite these advances, purely data-driven RL approaches encounter significant challenges in environmental applications. They typically require large amounts of training data, may violate physical constraints, and can generate policies that are difficult to interpret or trust. Recent reviews emphasize that insufficient physical consistency and robustness remain major obstacles to the operational adoption of RL in environmental systems (Rolnick et al., 2022; Reichstein et al., 2019).

In the environmental modelling community, adaptive control and policy search methods have been extensively studied to support decision-making under deep uncertainty and competing objectives. Policy search approaches, in particular, offer a principled framework for evaluating trade-offs among robustness, efficiency, and risk in water management systems (Giuliani et al., 2016). Additionally, many-objective evolutionary algorithms, such as the Borg framework, are widely used to identify diverse and non-dominated control strategies across complex environmental objectives (Hadka and Reed, 2013).

Within this context, reinforcement learning should be regarded not as a replacement for existing environmental decision-support methodologies, but as a complementary adaptive control paradigm. RL shares conceptual foundations with policy search and many-objective optimization, while providing

enhanced capacity to leverage temporal structure and delayed system responses.

2.4. Physics-informed and hybrid learning approaches

To overcome these limitations, physics-informed machine learning has emerged as a promising paradigm that integrates mechanistic knowledge into data-driven modelling. Comprehensive surveys indicate that constraining learning with physical structure enhances generalization, stability, and interpretability, especially in data-limited contexts (Willard et al., 2022; Karniadakis et al., 2021).

Neural Ordinary Differential Equations (Neural ODEs) offer a natural framework for hybrid modelling by enabling continuous-time representations in which neural networks parameterize unknown or residual dynamics (Rackauckas et al., 2020). In environmental and Earth system sciences, Neural ODEs and related universal differential equation frameworks are increasingly used to augment, rather than replace, physical models, thereby preserving interpretability while improving model fidelity (Rackauckas et al., 2021; Beucler et al., 2021).

2.5. Hybrid modelling for control under uncertainty

Although physics-informed learning has been extensively studied for system identification and forecasting, its integration with reinforcement learning for control remains comparatively underexplored in environmental modelling. Recent work in safe and model-based RL emphasizes the importance of incorporating physical structures to improve robustness and avoid unsafe or unrealistic control policies (Berkenkamp et al., 2017; Perkins et al., 2023).

Few studies have conducted systematic comparisons between rule-based control, reinforcement learning, and hybrid neurophysical control within a unified environmental modelling framework. In particular, the potential of learning residual dynamics via Neural ODEs to enhance irrigation control robustness under climatic variability has received limited attention. This gap motivates the present study, which investigates control strategies to increase modelling complexity in a physics-based irrigation environment.

While policy search and multi-objective evolutionary approaches have been successfully applied to irrigation management problems (Giuliani et al., 2016, 2021), the present work differs in both its control formulation and modelling focus. In particular, policy search methods typically optimise static or

season-level irrigation strategies, whereas this study addresses daily closed-loop control with explicit state feedback and delayed system responses. Moreover, the proposed framework systematically isolates the effects of control learning and dynamics correction by comparing rule-based, reinforcement learning, and hybrid neuro-physical formulations under identical physical and climatic forcing. This controlled experimental design enables a clearer assessment of how learning-based control and residual dynamics augmentation jointly influence performance, stability, and interpretability, aspects not explicitly addressed in existing policy-search-based studies.

3. Materials and Methods

This section outlines the materials and methods adopted in this study. The system dynamics, scope of the work, and data sources are first clarified. The irrigation management problem is then formulated. Subsequently, the physics-based soil–water environment used to simulate daily dynamics is presented, followed by the definition of the three control scenarios investigated: rule-based control, reinforcement learning, and hybrid neuro-physical reinforcement learning. Finally, the experimental design, training procedures, and evaluation protocols employed to ensure reproducibility and fair comparison across control strategies are detailed.

3.1. Soil–water dynamics and physical assumptions

Figure 1 illustrates the main physical processes represented in the soil–water balance model used across all scenarios. The model describes the root zone as an effective control volume subject to external water inputs from rainfall and irrigation, internal storage dynamics, and losses through evapotranspiration and drainage. Rainfall is treated as an exogenous hydro-climatic forcing, while irrigation constitutes a controllable management input. Both contribute to soil-water storage subject to physical capacity constraints.

Water losses are represented through evapotranspiration, which aggregates soil evaporation and plant transpiration into a single flux driven by atmospheric demand and regulated by soil-water availability. This formulation follows standard agro-hydrological practice, in which reference evapotranspiration is computed from meteorological variables and scaled by crop coefficients and stress reduction functions to reflect plant response under water-limited conditions (Allen et al., 1998; Monteith, 1965). Such representations are widely adopted in operational irrigation models and crop simula-

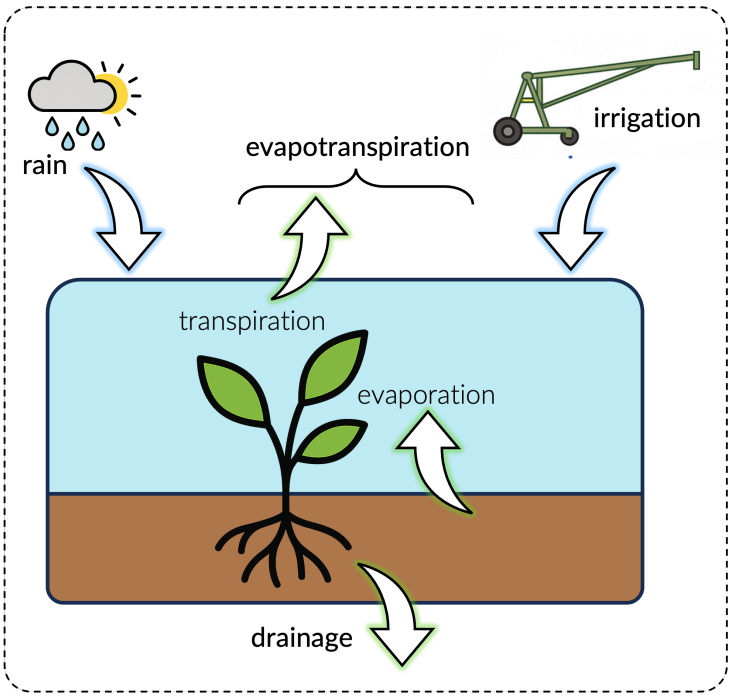


Figure 1: Schematic representation of the soil-water balance processes considered in the physical model including water inputs from rainfall and irrigation, losses due to evapotranspiration (evaporation and transpiration), and drainage from the root zone.

tion tools, as they provide a parsimonious yet physically grounded description of land–atmosphere exchanges (Raes et al., 2009).

Drainage represents gravitational percolation losses occurring when soil-water storage exceeds field capacity, leading to water leaving the effective root zone and becoming unavailable for crop uptake. In conceptual models, this process is commonly represented using threshold-based or linear drainage functions, which approximate the integrated effect of soil hydraulic conductivity and profile structure without explicitly resolving vertical flow (Rodríguez-Iturbe and Porporato, 2004).

The soil profile is modeled here as a single, homogeneous root-zone compartment characterized by an effective storage capacity and a monotonic retention relationship linking soil-water storage to matric potential. This bucket-type representation neglects vertical heterogeneity, preferential flow paths, layered hydraulic properties, and lateral redistribution. While such processes are known to play a critical role in real soils, their explicit rep-

resentation substantially increases model dimensionality, introduces strong nonlinearities, and induces delayed and history-dependent responses in soil moisture dynamics (Vereecken et al., 2007; Fatichi et al., 2016).

In particular, soil heterogeneity and hydraulic nonlinearity generate memory effects whereby past rainfall, irrigation, and evapotranspiration influence future water availability over multiple time scales. These delayed responses complicate both state estimation and control, especially under stochastic climatic forcing and sparse sensor observations (Seneviratne et al., 2010). Accurately capturing such dynamics typically requires multi-layer Richards-equation-based models (Appendix A) or detailed ecohydrological formulations (Appendix B), which are computationally demanding and difficult to integrate with learning-based controllers in a reproducible manner.

The simplifications adopted in this study are therefore intentional. By using a reduced-order, physically interpretable soil–water balance model operating at a daily time step, we isolate the effects of control strategy design and learning-based adaptation under uncertainty. This choice enables transparent coupling with reinforcement learning and hybrid neuro-physical approaches, while preserving the dominant hydrological feedbacks relevant for irrigation decision-making. The resulting physics-based environment thus constitutes a control-oriented reference model on which all control and learning strategies are evaluated, ensuring that observed performance differences arise from the controller design rather than from discrepancies in the physical model.

Appendix C presents a comparative analysis of the purposes and control compatibility of alternative soil–water modelling formulations. Appendix D demonstrates the stability and well-posedness of these formulations.

3.2. Data sources and scope of the study

The study utilized simulated data generated by a physics-based soil-water balance model driven by stochastic climatic forcing. This simulation-based approach enables controlled experimentation, systematic comparison of control strategies, and reproducibility across a wide range of hydroclimatic conditions, while avoiding confounding effects such as data scarcity, sensor noise, or site-specific calibration.

The simulated environment produces daily trajectories of rainfall, reference evapotranspiration, soil-water storage, and soil matric potential (tension), which together define the state and observation variables used by the control algorithms. In all scenarios, irrigation decisions were evaluated

against identical simulated physical dynamics and weather realizations, ensuring fair and consistent comparisons across the control strategies.

Although the present study focuses on simulated data, the proposed framework is explicitly designed to be transferable to real-world irrigation systems. In operational settings, soil-water storage is typically not directly observable; instead, the soil-water status is monitored using tensiometers that measure the soil matric potential ψ_t. Rainfall inputs can be obtained from on-site rain gauges or open-access meteorological datasets (e.g., national weather services or global reanalysis products). The observation structure adopted in this study, based on ψ_t and climatic variables, directly reflects these practical constraints.

3.3. Problem formulation

This subsection establishes a unified mathematical formulation of the irrigation control problem that underpins all scenarios considered in this study. Irrigation management is formalized as a finite-horizon sequential decision-making process, explicitly distinguishing between (i) the latent physical state of the soil-water system, (ii) the observable variables available to the controller, and (iii) the control actions constrained by agronomic and operational limits. This formulation provides a common reference framework for rule-based control, reinforcement learning, and hybrid neurophysical approaches, ensuring that performance differences arise from the control strategy rather than discrepancies in system representation.

3.3.1. System and time discretization

A single agricultural plot evolving over a growing season of length T days is considered, discretized into daily decision steps $t \in \{0, \ldots, T-1\}$. Irrigation decisions are made once per day, consistent with the temporal resolution of the meteorological forcing and the operational granularity targeted in this study.

3.3.2. Latent state, observations, and retention relationship

Let S_t (mm) denote the *latent* root-zone soil-water storage at day t, which represents the amount of plant-available water in the effective root zone of the plant. In operational settings, S_t is not generally measured directly. Instead, soil-water status is monitored using tensiometers that provide soil matric potential (tension) ψ_t (cbar), which is treated as the primary observable variable.

The link between latent storage and observed tension is defined by the soil-water retention curve:

$$\psi_t = f_{\mathrm{ret}}(S_t), \qquad S_t = f_{\mathrm{ret}}^{-1}(\psi_t), \tag{1}$$

where $f_{\mathrm{ret}}(\cdot)$ is a soil-specific monotonic mapping determined by the hydraulic properties.

Simulator-accessible state. For controlled benchmarking, the simulation environment maintains and updates the soil-water storage S_t internally. In Scenarios 2 (reinforcement learning) and Scenario 3 (hybrid physics-informed control), S_t is optionally provided to the agent as part of the observation vector in order to isolate the effect of the control strategy under identical physical dynamics and climatic forcing. This variable is not directly measurable in operational settings and is therefore treated as *simulator-accessible* rather than emph sensor-accessible.

This configuration represents an upper-bound assessment of learning-based control performance rather than a deployable solution. In practical applications, controllers rely on sensor-level observations such as soil-water tension, rainfall, and reference evapotranspiration. Extensions to partially observable settings, including belief-state estimation or recurrent policies, are identified as directions for future work.

3.3.3. Climate drivers

Daily climate forcing is represented by rainfall R_t (mm), reference evapotranspiration $ET0_t$ (mmday^{-1}), and crop coefficient Kc_t (dimensionless). We denote exogenous drivers compactly as

$$d_t := (R_t, ET0_t, Kc_t), \tag{2}$$

and treat them as stochastic disturbances sampled from a (possibly nonstationary) distribution:

$$d_t \sim \mathcal{P}_d. \tag{3}$$

3.3.4. Physics-based dynamics (mass balance)

Root zone dynamics are governed by a bucket-type soil-water mass balance:

$$S_{t+1} = \mathrm{clip}(S_t + \eta_I I_t + R_t - ET_{c,t} - D_t,\ 0,\ S_{\max})\,, \tag{4}$$

where I_t (mm) is the applied irrigation depth (control action), $\eta_I \in (0, 1]$ is the irrigation efficiency, $S_{\max}$ is the maximum admissible storage, and clip$(\cdot)$ enforces physical bounds. Drainage is computed as

$$D_t := D(S_t), \tag{5}$$

It is typically activated when the storage exceeds the field capacity. Crop evapotranspiration was computed following an FAO-inspired structure:

$$ET_{c,t} = Kc_t \, ET0_t \, f_{ET}(\psi_t), \tag{6}$$

where $f_{ET}(\psi_t) \in [0, 1]$ is the stress reduction factor driven by soil tension. The next-day tension is obtained from storage using Eq. (1).

3.3.5. Sequential decision-making and objective

At each day t, the controller selects an irrigation action

$$I_t \in [0, I_{\max}], \tag{7}$$

where $I_{\max}$ (mm) is the maximum allowable daily irrigation depth.

The controller receives an observation vector $\mathbf{o}_t$. In this study, we used a compact, physically meaningful representation:

$$\mathbf{o}_t = \begin{cases} (\psi_t,\ R_t,\ ET0_t) & \text{(sensor-level baseline)}, \\ (\psi_t,\ S_t,\ R_t,\ ET0_t) & \text{(simulator-accessible benchmarking, used in Scenario 2)}. \end{cases} \tag{8}$$

The problem is therefore partially observed in operational settings. Nonetheless, an MDP-over-observations formulation is adopted for tractability and reproducible comparisons, with partial-observability extensions explicitly discussed in Section 4.

The control objective is to limit crop water stress while reducing irrigation water usage and hydrological losses. We define a daily reward as

$$r_t = -\Big(\alpha \, \mathcal{L}_{\text{stress}}(\psi_t) + \beta \, I_t + \gamma \, D_t\Big), \tag{9}$$

where $\mathcal{L}_{\text{stress}}(\psi_t)$ penalizes excessive tension, and $\alpha, \beta, \gamma > 0$ weight the trade-offs between stress avoidance, irrigation cost, and drainage loss, respectively. A policy π is evaluated by the expected discounted return:

$$J(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} \gamma^t r_t \right], \tag{10}$$

with a discount factor $\gamma \in (0, 1]$ and expectation taken over stochastic climate forcing. This formulation is standard in reinforcement learning (Sutton and Barto, 2018) and aligns with the environmental decision-making under uncertainty.

3.4. Physics-based irrigation environment

All scenarios interact with the same physics-based environment, as implemented by Eqs. (4)–(6). Each episode corresponded to a full growing season. The environment returns daily observations (Eq. (8)), accepts a bounded irrigation action I_t, updates the latent storage S_t, computes ψ_t using the retention curve, and provides the reward defined in Eq. (9). Meteorological forcing $(R_t, ET0_t, Kc_t)$ was generated from a stochastic weather process with explicit seeds to ensure reproducibility.

3.5. Control scenarios

We consider three irrigation control scenarios representing increasing levels of learning and model integration. All scenarios share the same underlying physics-based soil-water dynamics, climatic forcing protocol, action bounds, and evaluation metrics; they differ only in the design of the irrigation controller and, in the most advanced setting, in the representation of system dynamics. In particular, the third scenario incorporates a Neural ODE-inspired residual correction, implemented here in discrete time to match the daily decision cycles.

3.5.1. Scenario 1: Rule-based irrigation control (physics + heuristics)

Scenario 1 couples the physics-based environment with a deterministic rule-based policy representative of common operational practices. At each day t, the rule maps the current observed tension and a simple one-day-ahead rainfall forecast to an irrigation decision.

We considered parameterized rule families (single-threshold, comfort-band, proportional), each defined by fixed thresholds and dose parameters. The season is simulated by iterating the physics-based update (Eq. (4)) from an initial condition at field capacity $S_0 = S_{\mathrm{fc}}$, with $\psi_0 = f_{\mathrm{ret}}(S_0)$.

This scenario provides an interpretable, low-cost baseline that is robust by construction but non-adaptive; parameters remain fixed across seasons and cannot optimize trade-offs under variability.

3.5.2. Scenario 2: Reinforcement learning with a physics-based environment (physics + PPO)

Scenario 2 replaces fixed heuristics with a policy learned through reinforcement learning via direct interaction with the physics-based environment. The agent observes $\mathbf{o}_t$ (Eq. (8)), selects a continuous irrigation depth $I_t \in [0, I_{\max}]$, and receives a reward r_t (Eq. (9)) and experience transitions induced by the process model (Eq. (4)).

Learning algorithm. We used Proximal Policy Optimization (PPO), an on-policy policy-gradient method with clipped updates and generalized advantage estimation. Both the policy and value functions were parameterized as multilayer perceptrons. Training proceeds over many simulated seasons under stochastic forcing, with explicit random seeds controlling both the climate realization and the timing of learning initiation. Scenario 2 serves as a learning-based baseline that isolates the effect of RL when the environment dynamics are purely physics-based (no learned correction).

3.5.3. Scenario 3: Hybrid neuro-physical control (physics + residual correction + PPO)

Scenario 3 augments the physics-based environment with a learned residual correction that compensates for systematic model mismatches while preserving the mass-balance structure. The hybrid update was implemented in the *tension space*, consistent with tensiometer-driven monitoring.

Hybrid transition with residual correction. First, the physical model computes the nominal next-day storage and tension as follows:

$$S_{t+1}^{\text{phys}} = \text{clip}(S_t + \eta_I I_t + R_t - ET_{c,t} - D_t,\ 0,\ S_{\max})\,, \tag{11}$$

$$\psi_{t+1}^{\text{phys}} = f_{\text{ret}}(S_{t+1}^{\text{phys}}). \tag{12}$$

The residual model then predicts an additive correction as follows:

$$\Delta\psi_t = f_\theta(\psi_t, I_t, R_t, ET0_t), \tag{13}$$

and the hybrid prediction is

$$\psi_{t+1} = \psi_{t+1}^{\text{phys}} + \Delta\psi_t, \qquad S_{t+1} = f_{\text{ret}}^{-1}(\psi_{t+1}). \tag{14}$$

This ensures consistency between the corrected tension and physically admissible storage via the inverse retention curve.

Residual model architecture. In our implementation, f_θ is a lightweight MLP with two hidden layers of 64 units and tanh activations, mapping a 4D input $[\psi_t, I_t, R_t, ET0_t]^\top$ to a scalar output $\Delta\psi_t$.

Residual training (pretraining). The residual model was pretrained using supervised regression on simulated trajectories. For each sample, the inputs are

$$\mathbf{x}_t = [\psi_t,\ I_t,\ R_t,\ ET0_t]^\top, \tag{15}$$

and targets are defined as the discrepancy between a perturbed "reference" next-day tension and the nominal physical prediction:

$$y_t = \psi_{t+1}^{\text{ref}} - \psi_{t+1}^{\text{phys}}. \tag{16}$$

In the current implementation, ψ_{t+1}^{ref} is generated by injecting stochastic perturbations into the physical update (emulating unresolved processes and observation noises). The parameters were optimized using Adam with a robust regression loss (Smooth L1). After pretraining, f_θ is fixed and used in the inference mode during the RL training.

Generation of reference trajectories and perturbation structure. The residual dynamics model is pretrained using reference trajectories generated from a perturbed version of the physics-based soil–water model. These perturbations are designed to emulate plausible sources of model mismatch rather than to represent measurement noise or an alternative high-fidelity simulator.

Specifically, three classes of perturbations are introduced:

- *Evapotranspiration stress response*: the stress reduction function $f_{ET}(\psi_t)$ is modified by a season-consistent bias term and a low-amplitude stochastic fluctuation, reflecting uncertainty in crop response and soil–plant interactions.
- *Drainage sensitivity*: the drainage function $D(S_t)$ is perturbed by varying its effective slope near field capacity, emulating unresolved soil heterogeneity and preferential flow pathways.
- *Irrigation efficiency*: the effective irrigation efficiency η_I is perturbed by a multiplicative factor that remains constant over a season, representing spatial variability and operational uncertainty.

Perturbations include both stochastic components and systematic biases that persist over each simulated season. This structure ensures that the resulting discrepancies are temporally correlated and hydrologically coherent, rather than resembling unstructured noise.

The reference next-day soil-water tension ψ_{t+1}^{ref} is then obtained by propagating the perturbed model forward in time, while the nominal physical prediction ψ_{t+1}^{phys} is computed using the unperturbed model. The residual target is defined as

$$\Delta\psi_t = \psi_{t+1}^{\text{ref}} - \psi_{t+1}^{\text{phys}}. \tag{17}$$

This residual learning strategy does not aim to approximate a specific alternative physical formulation. Instead, it trains the neural correction to compensate for persistent, control-relevant discrepancies that arise when simplified soil–water models interact with stochastic climatic forcing.

Discrete-time integration choice. Although we refer to this module as a "Neural ODE" for consistency with project terminology, the implemented residual correction is *discrete-time*: the network predicts the one-day correction $\Delta\psi_t$ directly (Eq. (13)).This choice (i) aligns with the daily forcing and decision frequency, (ii) reduces the computational overhead to a single forward pass per day, and (iii) avoids solver-induced numerical issues. Continuous-time residual formulations and higher-frequency data assimilation are considered extensions of this approach.

Reinforcement learning on the hybrid environment. A PPO agent is then trained on the hybrid environment using the same reward structure as Scenario 2 (Eq. (9)). This isolates the benefit of correcting the dynamics-level mismatch while keeping the policy-learning mechanism unchanged.

3.6. Experimental design and evaluation protocol

This subsection details the experimental design and evaluation protocol adopted to ensure fair, transparent, and reproducible comparisons among the three control scenarios. Particular attention was given to isolating the effects of the control strategy from those of the physical model and climatic forcing. All experiments shared identical environment dynamics, weather generation procedures, and evaluation metrics, differing only in the control formulation. The protocol further emphasizes configuration-driven reproducibility, controlled randomness, and consistent performance assessments across independent runs. All reported hyperparameters correspond to a stable region identified through preliminary tuning and are not optimized per scenario.

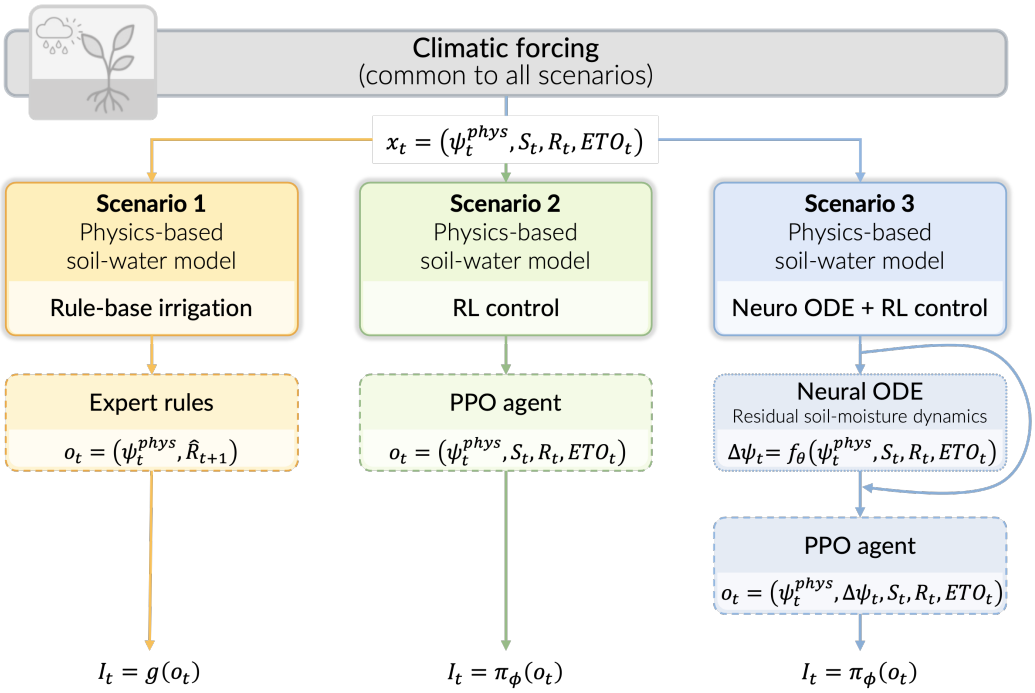


Figure 2: Overview of the three irrigation control scenarios considered in this study. Scenarios are ordered by increasing learning involvement and model–controller coupling. All scenarios share a common climatic forcing and a physics-based soil-water model, while differing in the irrigation decision mechanism and, in Scenario 3, in the representation of system dynamics.

Figure 2 summarizes the experimental design adopted in this study. All scenarios share the same underlying physics-based soil-water model and are driven by identical climatic forcing, but differ in the irrigation decision mechanism and, in Scenario 3, in the representation of system dynamics. Scenario 1 serves as a rule-based baseline, where irrigation decisions are computed using expert heuristics based on partial system information. Scenario 2 replaces heuristic rules with reinforcement learning control, using a PPO agent that interacts directly with the physical model. Scenario 3 further extends this framework by augmenting the physical model with learned residual dynamics via a Neural Ordinary Differential Equation, while reinforcement learning remains responsible for irrigation control. Unlike Scenario 2, Scenario 3 exposes the learned residual $\Delta\psi_t$ to the controller, enabling irrigation decisions that are explicitly informed by model discrepancy.

3.6.1. Configuration-driven reproducibility

To ensure transparent and reproducible experiments, parameters are centralized in a configuration module separating (i) environment parameters (season length T, $I_{\max}$, seeds), (ii) soil parameters ($S_{\max}$, S_{fc}, retention curve parameters, drainage and efficiency η_I), (iii) weather parameters (ET0 seasonality and rainfall generator), and (iv) training parameters (total interaction budget for PPO). This design supports controlled sensitivity analyses and aligns with the EMS best practices for reproducible assessments in Environmental Modelling and Software. All configuration files are available upon request / in supplementary material.

3.6.2. Training and evaluation separation

To ensure a fair, interpretable, and reproducible comparison across the control paradigms, a strict separation was enforced between the controller configuration, training (when applicable), and evaluation for all scenarios. Although only Scenarios 2 and 3 involved explicit learning, all scenarios were evaluated under identical physical, soil, and meteorological conditions using the same performance indicators.

Scenario 1: Physics-based model with rule-based irrigation control. Scenario 1 implements a baseline irrigation strategy based on expert-defined rules interacting with a physics-based soil-water bucket model. This scenario serves as a reference case, isolating the effect of heuristic control without any learning or adaptive policy optimization.

The simulation is defined over a growing season of fixed length T (days), with climatic forcing generated deterministically from a prescribed random seed. Daily weather inputs include rainfall R_t, reference evapotranspiration $\mathrm{ET0}_t$, and crop coefficient $K_c(t)$. The soil system is represented by a conceptual bucket model parameterized by a water retention curve, a drainage function, and an irrigation efficiency coefficient η_I. Unless specified otherwise, default soil parameters are used.

At the beginning of the season, soil-water storage is initialized at field capacity,

$$S_0 = S_{\mathrm{fc}}, \tag{18}$$

and converted to soil-water tension via the retention relationship $\psi_0 = S \rightarrow \psi(S_0)$.

At each day t, irrigation is determined by a predefined rule function

$$I_t = g(\psi_t, I_{\max}, \widehat{R}_{t+1}), \tag{19}$$

where ψ_t denotes the current soil-water tension, $I_{\max}$ is the maximum admissible daily irrigation depth, and $\widehat{R}_{t+1}$ is a one-day-ahead rainfall forecast (when available). The rule function may implement a single tension threshold or a comfort-band strategy, and internally clips the action to the feasible range $I_t \in [0, I_{\max}]$.

The physical soil-water dynamics are then updated deterministically. Crop evapotranspiration is computed as

$$\mathrm{ETc}_t = K_c(t)\,\mathrm{ET0}_t\,f_{\mathrm{ET}}(\psi_t), \tag{20}$$

where $f_{\mathrm{ET}}(\psi_t)$ is a stress reduction factor derived from the soil model. Drainage losses D_t occur when soil storage exceeds field capacity. The daily water balance is given by

$$S_{t+1} = \mathrm{clip}(S_t + \eta_I I_t + R_t - \mathrm{ETc}_t - D_t,\ 0,\ S_{\max})\,, \tag{21}$$

with $\mathrm{clip}(\cdot)$ enforcing physical bounds on soil-water storage. The updated soil-water tension is obtained via the inverse retention relation $\psi_{t+1} = S \to \psi(S_{t+1})$.

This procedure is repeated sequentially for $t = 0, \ldots, T-1$, producing time series of soil storage, soil tension, irrigation, evapotranspiration, and drainage. All actions are fully determined by the irrigation rule and the current system state; no policy network, learning mechanism, or optimization procedure is involved.

Scenario 1 therefore provides a transparent and interpretable benchmark that reflects common rule-based irrigation practices, against which the benefits of reinforcement learning control (Scenario 2) and hybrid neuro-physical control (Scenario 3) can be systematically assessed.

Scenario 2: Physics-based model with reinforcement learning control (PPO). In Scenario 2, irrigation control is achieved through a reinforcement learning agent interacting directly with the physics-based soil-water model described in Section 3.4. The agent is trained using Proximal Policy Optimization (PPO), as implemented in the Stable-Baselines3 library (Raffin et al., 2021), without any custom policy architecture or parameterisation.

The control policy is represented by the standard `MlpPolicy` provided by Stable-Baselines3. At each decision step t, the agent receives a continuous-valued observation vector

$$\mathbf{o}_t = [\psi_t,\ S_t,\ R_t,\ \mathrm{ET0}_t], \tag{22}$$

where ψ_t denotes soil-water tension, S_t the soil-water storage, R_t the rainfall input, and $\mathrm{ET0}_t$ the reference evapotranspiration. These variables jointly characterize the system's hydrological state and the prevailing climatic conditions.

The policy network consists of two fully connected hidden layers with 64 units each and ReLU activation functions, corresponding to the default PPO configuration in Stable-Baselines3. A shared feature extractor feeds two output heads: a policy head, which outputs the mean of a Gaussian distribution over the one-dimensional continuous action space, and a value head, which estimates the scalar state-value function $V(\mathbf{o}_t)$. In addition, PPO maintains a learnable log-standard deviation parameter for the action distribution.

The irrigation action I_t is sampled from the Gaussian policy, squashed through a hyperbolic tangent function, and rescaled to satisfy the operational constraints of the irrigation system:

$$I_t \in [0, I_{\max}], \tag{23}$$

resulting in a continuous irrigation dose expressed in millimeters. This action is then applied to the physics-based soil-water model, which updates the system state according to the water balance equations.

In this scenario, the reinforcement learning agent learns irrigation strategies solely through interaction with the fixed physical model, without any corrections or augmentations to the underlying system dynamics. Scenario 2 therefore isolates the contribution of learning-based control, providing a principled comparison with the rule-based strategy of Scenario 1 and the hybrid neuro-physical formulation introduced in Scenario 3.

Scenario 3: Hybrid environment with Neural ODE residual and PPO. Scenario 3 follows a two-stage learning protocol that explicitly separates model identification from policy optimization.

Stage 1: Pretraining of the Neural ODE residual model. Prior to reinforcement learning, the residual dynamics model was pre-trained in a supervised manner using simulated trajectories generated from a physics-based environment. The Neural ODE (implemented here as a discrete-time residual model) learns to predict a one-day correction $\Delta\psi_t$ to the soil-water tension based on inputs $(\psi_t, I_t, R_t, ET0_t)$. Training uses a fixed number of trajectories (typically 32), over 50 epochs, with a batch size of 256 and a learning rate of 10^{-3}. The objective is to minimize the discrepancy between the physical prediction

and the perturbed reference trajectory, yielding a stable residual model prior to control learning.

Once pretrained, the Neural ODE parameters are frozen and embedded within the environmental dynamics.

Stage 2: PPO training on the hybrid environment. The PPO agent is then trained in the hybrid environment (physics + Neural ODE correction) using the same algorithmic structure as in Scenario 2. The training budget is again defined by a fixed number of interaction steps (e.g., 50,000 timesteps), and the policy architecture remains a multilayer perceptron with continuous outputs. The PPO hyperparameters (learning rate, discount factor, clipping range, and GAE parameters) were kept consistent with Scenario 2 to isolate the effect of the hybrid dynamics.

Evaluation protocol. For all three scenarios, an evaluation was conducted after configuration or training using identical soil parameters, weather realizations, and initial conditions. No learning, adaptation, or parameter tuning was performed during the evaluation. Performance metrics, including soil-water tension dynamics, irrigation volumes, drainage losses, and aggregated efficiency indicators, were computed over full growing seasons.

This strict separation between configuration, training, and evaluation ensures that the observed performance differences arise from the controller design and system representation (rule-based, physical RL, or hybrid), rather than from stochastic variability, online adaptation, or unequal exposure to environmental conditions. This also reflects realistic deployment settings, where irrigation policies are typically calibrated or trained offline and then applied operationally without continuous retraining.

Importantly, the selected soil and climatic parameterisation should be interpreted as a reference operating regime chosen to enable controlled comparisons across control strategies, rather than as a representative or exhaustive characterization of agricultural conditions.

3.6.3. Performance indicators

The model performance was evaluated using both trajectory-level and aggregated indicators derived from the seasonal simulations. All indicators are consistently defined across scenarios and are directly linked to the notation summarized in Table 1.

At the trajectory level, we analyzed the temporal evolution of soil-water tension (ψ_t) and soil-water storage (S_t) to assess the occurrence, duration,

and severity of water-stress episodes, as well as the dynamics of depletion and recovery within the root zone. These trajectories provide insights into the controllers' ability to regulate soil-water status under stochastic climatic forcing.

At the aggregated level, several seasonal performance metrics were computed as follows: (i) The mean soil matric potential $\bar{\psi}$ summarizing the overall stress conditions; (ii) The fraction of days spent within an agronomically optimal tension range, denoted τ_{opt}; (iii) Total irrigation volume $I_{\text{tot}} = \sum_t I_t$; (iv) Cumulative drainage losses $D_{\text{tot}} = \sum_t D(S_t)$; (v) and Water-use efficiency metric Eff defined as the ratio between productive evapotranspiration and total water inputs.

Together, these indicators capture the key trade-offs between stress avoidance, water-use efficiency, and hydrological loss. They are used consistently across scenarios to ensure that observed performance differences can be attributed to the controller design rather than to confounding variations in physical parameters or climatic forcing.

3.6.4. Notation summary

To avoid ambiguity across the modelling, control, and learning components, Table 1 summarizes the notation used consistently throughout Section 3.

Table 1: Notation used throughout Section 3.

Symbol	**Unit**	**Description**
t	day	Discrete time index ($t = 0, \ldots, T-1$)
T	day	Length of the growing season (time horizon)
Soil-water state variables		
S_t	mm	soil-water storage in the root zone (latent physical state)
$S_{\max}$	mm	Maximum soil-water storage (soil capacity)
S_{fc}	mm	soil-water storage at field capacity
ψ_t	cbar	Soil matric potential (tension), observable via tensiometers
f_{ret}	–	soil-water retention function linking $S_t \leftrightarrow \psi_t$
Hydrological fluxes		

Continued on next page

Table 1 continued

Symbol	Unit	Description
I_t	mm	Irrigation depth applied at day t (control action)
$I_{\max}$	mm	Maximum allowable daily irrigation depth
R_t	mm	Rainfall at day t
$\widehat{R}_{t+1}$	mm	One-day-ahead rainfall forecast used by rule-based control (Scenario 1)
$ET0_t$	$\mathrm{mm\,day^{-1}}$	Reference evapotranspiration at day t
Kc_t	–	Crop coefficient at day t
ETc_t	mm	Crop evapotranspiration ($ETc_t = Kc_t \cdot ET0_t \cdot f_{ET}(\psi_t)$)
$f_{ET}(\psi_t)$	–	Water-stress reduction factor for evapotranspiration
$D(S_t)$	mm	Drainage loss as a function of soil-water storage
Mass balance and dynamics		
η_I	–	Irrigation efficiency coefficient
f_{phys}	–	Physics-based soil-water balance model
f_{res}	–	Learned residual dynamics (Neural ODE component)
Decision-making and learning		
$\mathbf{o}_t$	–	Observation vector available to the controller
$\mathbf{o}_t$	–	$(\psi_t, R_t, ET0_t)$ (default observation setting)
a_t	mm	Control action selected by the policy ($a_t = I_t$)
$g(\cdot)$	–	Rule-based irrigation function used in Scenario 1, defined as $I_t = g(\psi_t, I_{\max}, \widehat{R}_{t+1}; \boldsymbol{\kappa})$
$\boldsymbol{\kappa}$	–	Parameters of the irrigation rule (e.g., tension thresholds or comfort band limits)
$\pi(\cdot)$	–	Control policy (rule-based or learned)
π_θ	–	Parametric RL policy with parameters θ (Scenarios 2–3)
Reinforcement learning formulation		
r_t	–	Immediate reward at day t
γ	–	Discount factor for future rewards
$V(\mathbf{o}_t)$	–	State-value function approximation
$\hat{A}_t$	–	Advantage estimate (GAE)

Continued on next page

Table 1 continued

Symbol	Unit	Description
$J(\theta)$	–	Expected cumulative return optimized by PPO
Neural ODE residual model (Scenario 3)		
f_θ	–	Neural network parameterizing residual correction
$\Delta\psi_t$	cbar	Residual correction to soil-water tension
ψ_{t+1}^{phys}	cbar	Physical model prediction of soil tension
ψ_{t+1}	cbar	Hybrid prediction: $\psi_{t+1}^{\text{phys}} + \Delta\psi_t$
Performance indicators		
$\bar{\psi}$	cbar	Mean soil matric potential over the season
τ_{opt}	%	Fraction of days within optimal tension range
I_{tot}	mm	Total irrigation volume over the season
D_{tot}	mm	Total drainage loss over the season
Eff	–	Water-use efficiency metric ($ETc/(I+R)$)

4. Results and discussion

The results for the three irrigation control scenarios described in Section 3.5 are presented and discussed: (i) rule-based control, (ii) reinforcement learning in a physics-based environment, and (iii) hybrid reinforcement learning with Neural ODE–augmented dynamics. All simulations were performed over a complete growing season using identical soil and climatic conditions.

The experimental design and evaluation protocol adhered to established best practices in environmental modeling systems (EMS) to ensure transparency, reproducibility, and interpretability in model assessment. Model performance was characterized using complementary indicators that capture both system-level behavior and decision-relevant outcomes, consistent with EMS recommendations for environmental model evaluation (Bennett et al., 2013). Fixed physical configurations, controlled stochastic forcing, and consistent training and evaluation settings across scenarios ensured that observed performance differences were attributable to the control strategy rather than experimental artifacts.

4.1. Scenario 1: Rule-based control — conservative stability

Figure 3 illustrates the seasonal dynamics obtained using rule-based irrigation strategies.

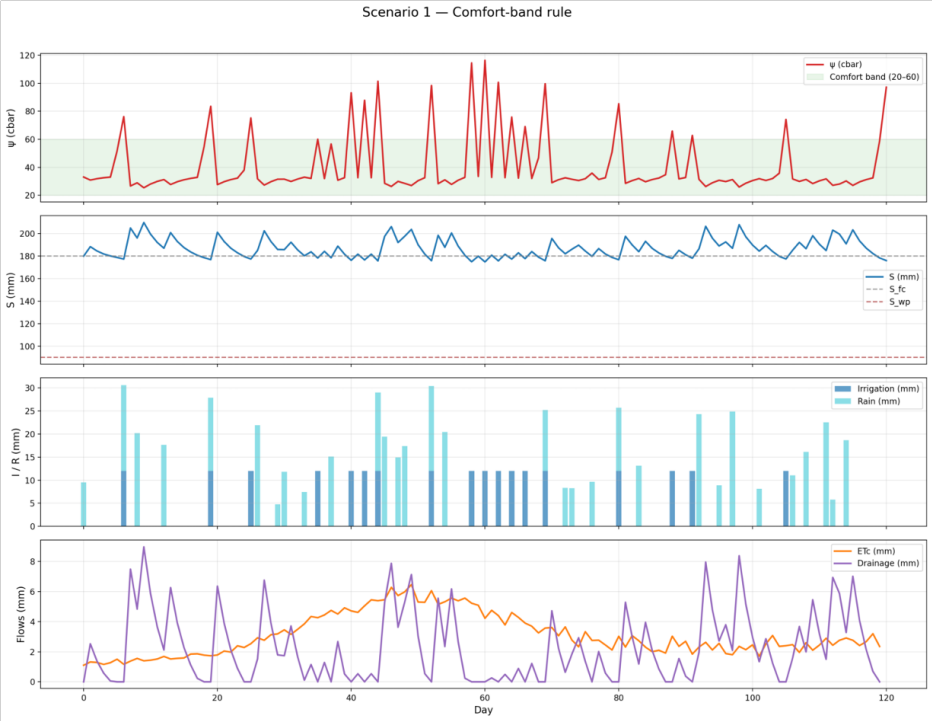


Figure 3: Seasonal dynamics of rule-based irrigation control (Scenario 1): soil-water tension, soil-water storage, irrigation and rainfall inputs, and hydrological fluxes.

The findings indicate that rule-based control maintains soil-water tension primarily within or near the agronomically optimal range for the crop. This conservative approach leads to a high proportion of days within the comfort zone, reflecting the strict enforcement of predefined thresholds or bands. Nevertheless, irrigation actions are initiated reactively and often abruptly, resulting in frequent water applications.

Soil-water storage remains near field capacity for most of the season; however, are specified as infiltration at the surface and free drainage or a age losses, especially when irrigation coincides with rainfall events. These findings suggest that while rule-based control is robust regarding stress avoidance and interpretability, it demonstrates inefficient water use under variable climatic conditions.

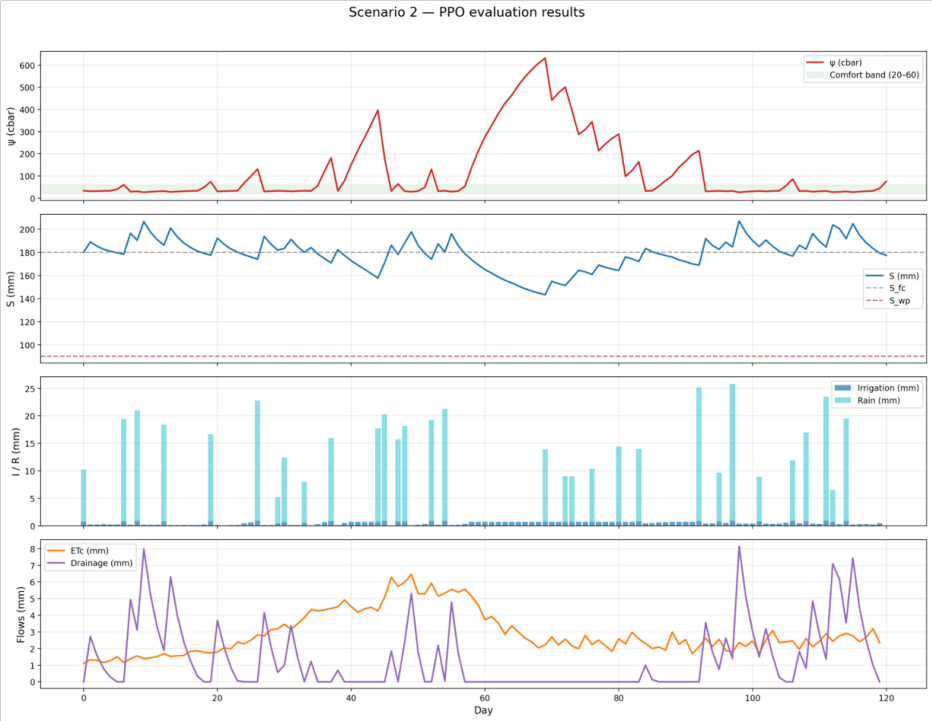


Figure 4: Seasonal dynamics of PPO-based irrigation control in the physics-based environment (Scenario 2).

4.2. Scenario 2: Physics-based reinforcement learning — efficiency with risk

Figure 4 presents the seasonal trajectories obtained using the PPO controller interacting with the physics-based environment.

In comparison to Scenario 1, the irrigation actions in Scenario 2 were smoother and significantly reduced in magnitude. This resulted in the lowest cumulative irrigation volume among the three scenarios. Scenario 2 also achieved the highest water-use efficiency. However, this improvement was accompanied by pronounced and persistent peaks in soil-water tension during dry periods, which indicates episodes of severe plant stress.

The occurrence of extreme soil-water tension values highlights the limitations of the controller's ability to anticipate cumulative water deficits when relying exclusively on a simplified physical model. Although drainage losses are reduced compared to the rule-based baseline, the presence of extreme stress episodes raises concerns about agronomic reliability.

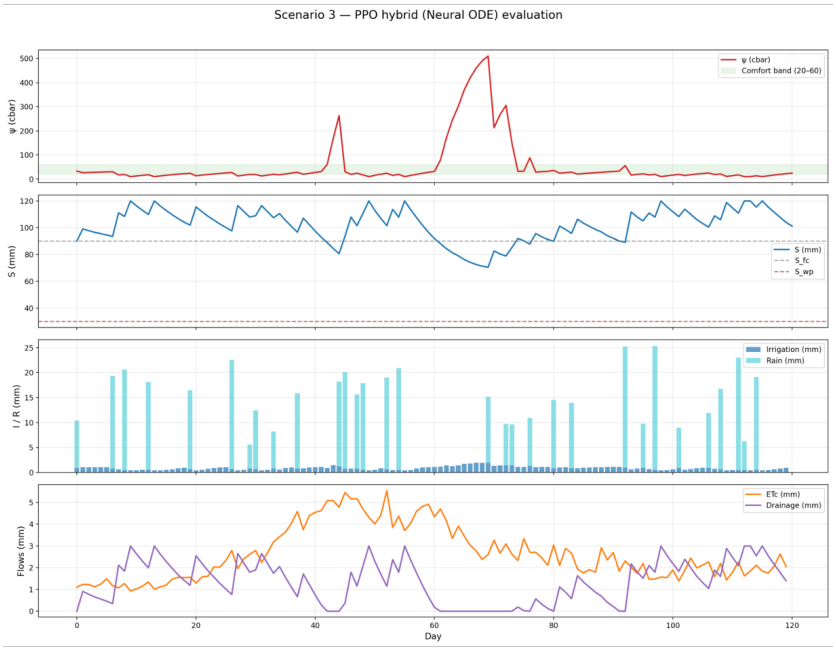


Figure 5: Seasonal dynamics of hybrid PPO control with Neural ODE–augmented dynamics (Scenario 3).

4.3. Scenario 3: Hybrid reinforcement learning with Neural ODE — moderated trade-offs

Scenario 3 addresses the limitations identified in physics-based reinforcement learning by introducing a hybrid neuro-physical formulation. This approach augments the physical soil-water model with learned components. By incorporating residual dynamics, this scenario moderates extreme stress responses while retaining the adaptive benefits of learning-based control.

Figure 5 presents the results obtained with the hybrid neuro-physical control strategy.

The hybrid approach substantially reduced the severity and duration of extreme soil-water tension peaks observed in Scenario 2. Soil-water storage trajectories were smoother and avoided significant depletion during extended dry periods. Irrigation actions remained moderate and were more closely aligned with rainfall events, resulting in reduced drainage losses.

However, Scenario 3 did not maximize the duration spent within the optimal tension range, which remained lower than that achieved by the rule-based strategy. Instead, the hybrid controller represented a compromise between

stress mitigation and water-use efficiency.

4.4. Soil and climatic parameterisation on control performance

All results were generated using a reference soil-climate configuration that represents a moderately deep agricultural soil exposed to semi-arid seasonal conditions. The specified root zone depth ($Z_r = 600$ mm), combined with *control-awareness tilting point values*, establishes a substantial buffering capacity that supports control strategies designed to anticipate delayed stress responses. In this context, excessive irrigation results mainly in drainage losses determined by the drainage coefficient ($k_d = 0.30$), which accounts for the higher penalties incurred by aggressive rule-based strategies without corresponding reductions in plant stress.

The evapotranspiration stress threshold ($?^{crit}_{?ET} = 80$ cbar) further shapes controller behavior by establishing a narrow boundary between non-stress and stress conditions. Within this range, learning-based controllers can implement precise irrigation adjustments. From a climatic standpoint, explicitly accounting for model discrepancy, thereby referencing evapotranspiration, combined with stochastic rainfall, results in intermittent rather than continuous water deficits. These circumstances favor adaptive policies that weigh short-term irrigation actions against anticipated atmospheric demand, rather than relying solely on fixed thresholds.

Consequently, these soil and climatic parameters establish the operating regime in which performance differences among control scenarios are observed and delineate the extent to which the results can be generalized. Alternative configurations, including shallower soils, increased drainage capacities, or more arid rainfall patterns, may modify the comparative advantages of rule-based, reinforcement learning, and hybrid neurophysical controllers.

Interactive configuration and reproducibility. Beyond the experimental results presented in this study, the proposed framework was implemented within an interactive interface developed using a Streamlit web application. This interface enables users to adjust soil, climatic, and control parameters in real time, allowing for rapid exploration of alternative configurations and sensitivity analyses without modifying the underlying codebase. This implementation enhances reproducibility, supports scenario-based experimentation, and serves as a practical link between the methodological contributions of this study and their application in decision support and educational settings.

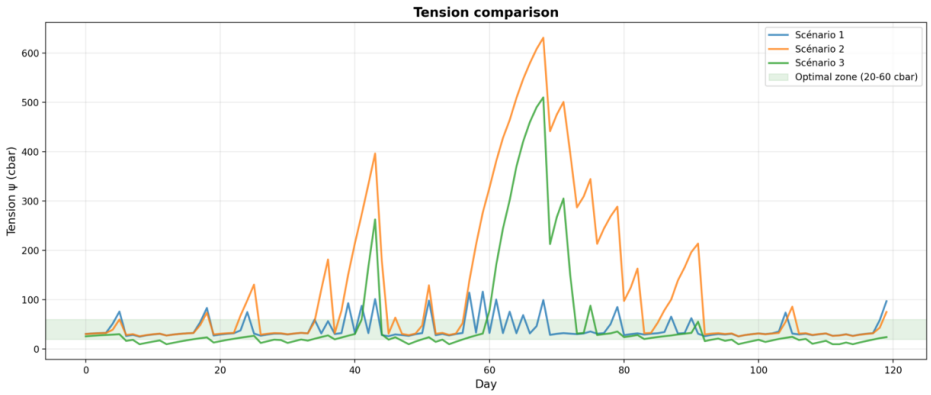


Figure 6: Comparison of soil-water tension dynamics across the three scenarios. The shaded area indicates the agronomically optimal tension range.

These effects align with the soil and climatic sensitivity discussed in Section 4.4.

4.5. Comparative analysis across scenarios

To consolidate the scenario-wise analysis, comparative indicators were evaluated based on seasonal simulations using the soil and climatic parameterisation described in Section 3.6. The results indicate a moderately deep root zone ($Z_r = 600$ mm), a relatively narrow optimal soil-water tension window ($\psi_{\text{fc}} \approx 33$ cbar, $\psi_{ET}^{\text{crit}} = 80$ cbar), and notable drainage sensitivity governed by the coefficient $k_d = 0.30$. The climatic forcing is characterized by moderate reference evapotranspiration variability and intermittent rainfall events, which are representative of semi-arid to sub-humid growing conditions.

4.5.1. Comparison of soil-water tension dynamics

Figure 6 presents a comparison of soil-water tension trajectories across the three scenarios.

Under the selected soil parameters, the rule-based controller (Scenario 1) maintained soil-water tension predominantly within the optimal range. This outcome results from conservative threshold settings relative to ψ_{ET}^{crit} and a relatively high irrigation efficiency ($\eta_I = 0.85$), which together promote frequent replenishment of root-zone storage.

In contrast, Scenario 2 exhibited pronounced stress peaks, with soil-water tension occasionally exceeding the critical stress threshold. These extremes

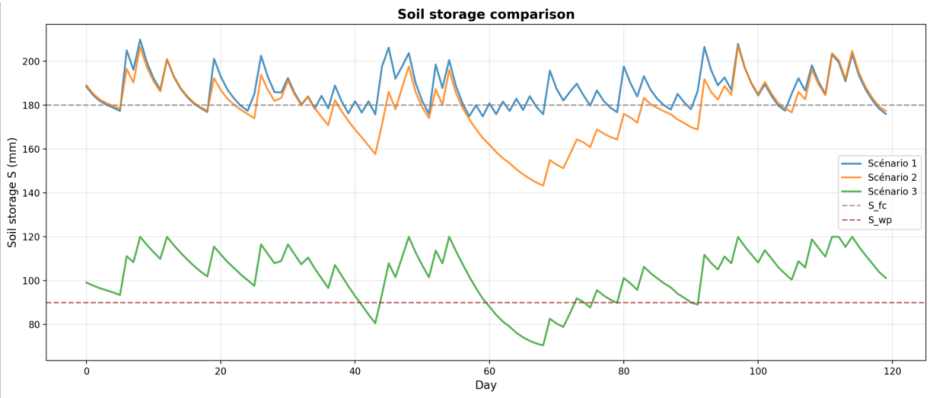


Figure 7: Comparison of soil-water storage trajectories across the three scenarios, relative to field capacity and wilting point.

result from the interaction among the finite root zone capacity, delayed hydrological responses embedded in the soil retention curve, and the reinforcement learning agent's incentive to minimize irrigation under stochastic rainfall. When rainfall realizations deviate from expected patterns, the simplified physical model underestimates the risk of cumulative depletion, which leads to delayed corrective actions.

Scenario 3 reduces the magnitude of extreme tension excursions compared to Scenario 2. Residual neural correction partially compensates for structural mismatches in evapotranspiration reduction and drainage response, thereby reducing abrupt transitions into high-stress regimes. However, the hybrid controller does not fully replicate the conservative behavior of Scenario 1, as its objective is to reshape rather than eliminate the efficiency–robustness trade-off.

4.5.2. Comparison of soil-water storage

Figure 7 shows the corresponding soil-water storage trajectories.

The rule-based controller maintains soil-water storage near field capacity (S_{fc}), reflecting its conservative irrigation logic and resulting in limited exposure to water stress. However, with a drainage coefficient of $k_d = 0.30$, this approach consistently leads to deep percolation losses when rainfall or irrigation surpasses short-term evapotranspiration demand.

Physics-based reinforcement learning (Scenario 2) permits greater soil-water depletion, especially during prolonged dry periods with elevated $ET0_t$. The reward structure, which penalizes irrigation volumes more than transient

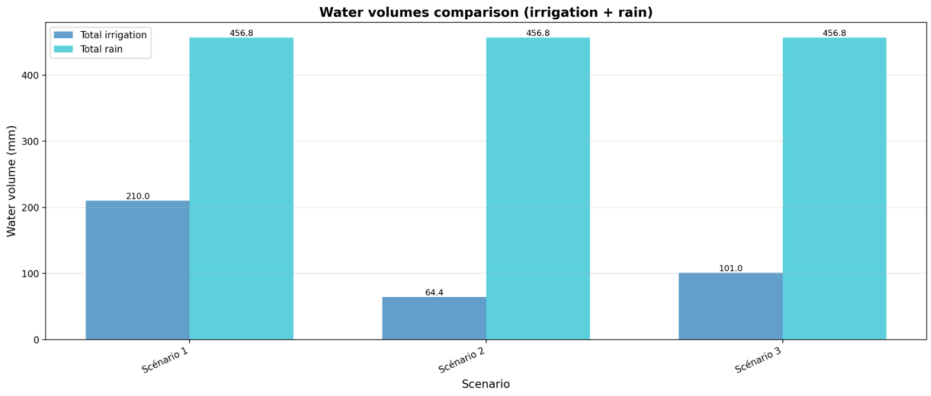


Figure 8: Comparison of cumulative irrigation and rainfall volumes across scenarios.

stress, encourages this behavior. As a result, the agent utilizes the entire dynamic range of the soil reservoir, occasionally approaching the wilting point.

The hybrid controller (Scenario 3) moderates these dynamics. Residual corrections modify the physical model's response to storage depletion and drainage, resulting in smoother soil-water trajectories that prevent both excessive depletion and systematic over-irrigation.

4.5.3. Comparison of cumulative water volumes

Figure 8 presents a comparison of cumulative irrigation and rainfall volumes across the scenarios.

Rainfall contributions remained identical across all scenarios. Thus, variations in cumulative irrigation volumes are attributable solely to differences in control strategies. Scenario 1 applied the greatest irrigation depth, consistent with its conservative threshold logic and prioritization of stress avoidance under uncertain rainfall conditions.

Scenario 2 achieved the lowest irrigation volume by leveraging rainfall variability and soil storage capacity to enhance water-use efficiency. However, under the chosen parameterization, this approach increases stress variability within the system. Scenario 3 represents an intermediate case, demonstrating that the hybrid correction adjusts irrigation demand without reverting to an excessively conservative strategy.

4.5.4. Comparison of aggregated performance indicators

Figure 9 summarizes the aggregated performance indicators.

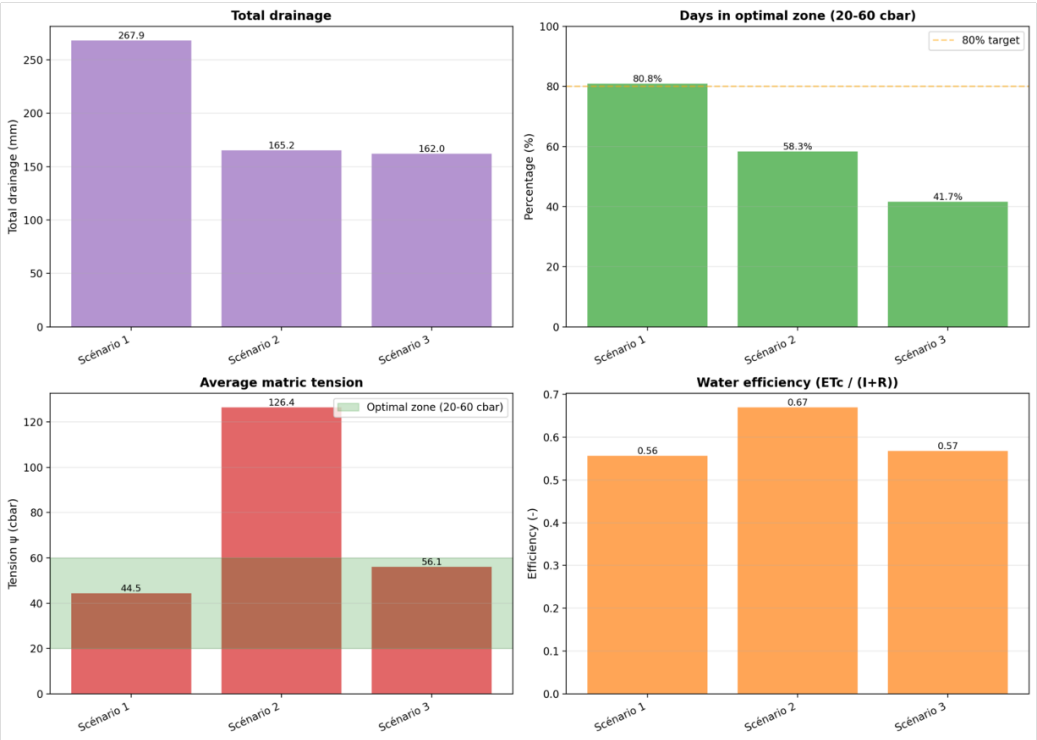


Figure 9: Comparison of aggregated performance indicators: total drainage, percentage of days in the optimal tension range, mean soil-water tension, and water-use efficiency.

Rule-based control (Scenario 1) maximizes the proportion of days within the optimal tension range; however, this approach incurs substantial drainage losses due to maintaining storage near S_{fc} in soils with notable percolation sensitivity. Physics-based reinforcement learning (Scenario 2) attains the highest water-use efficiency by permitting greater fluctuations in soil-water tension, but this results in the least favorable stress indicators.

The hybrid neuro-physical approach (Scenario 3) reduces both drainage losses and the occurrence of extreme stress events. By addressing systematic biases in the physical model response, especially under conditions of high evaporative demand or near-capacity storage, this method achieves a more balanced compromise between efficiency and robustness.

4.5.5. Comparative synthesis

A consistent pattern was observed across all indicators, closely associated with the chosen soil and climatic parameters. Rule-based control is de-

signed to prioritize stress avoidance, which benefits shallow or highly stress-sensitive soils, but leads to inefficient water use when drainage sensitivity is pronounced. Conversely, physics-based reinforcement learning emphasizes efficiency by leveraging soil storage and rainfall variability, though this increases risk under simplified dynamics and stochastic influences.

Hybrid neuro-physical reinforcement learning modifies this trade-off by correcting model discrepancies relevant to control, rather than increasing the physical model's resolution. Notably, the Neural ODE component does not seek to maximize the duration within agronomic comfort zones; instead, it mitigates delayed or exaggerated responses resulting from simplified representations of evapotranspiration and drainage.

Operationally, these results indicate that hybrid controllers are especially appropriate in scenarios where basic physical knowledge exists, but soil behavior and climatic forcing are uncertain. Rule-based strategies continue to be suitable for low-risk or highly regulated environments, while learning-based and hybrid methods are increasingly pertinent under stricter water constraints, greater climatic variability, and changing soil conditions. Practical limitations such as training costs and data availability highlight the need for future research on incremental learning, sensor-driven calibration, and cross-site transfer.

4.5.6. Limitations

Several limitations persist and delineate clear avenues for future research. These constraints are directly related to the soil and climatic parameterization employed in this study, which was deliberately selected to facilitate controlled and reproducible comparisons among control strategies.

- The controllers depend on Markovian state representations with a fixed daily resolution. While this approach aligns with the daily availability of irrigation decisions and meteorological data, it restricts the modeling of long-term dependencies, delayed hydrological responses, and irregular temporal sampling. This constraint is especially significant for soils with greater root-zone depths or slower drainage, where the effects of past deficits can persist over several days. Future research will investigate Neural Controlled Differential Equations to more accurately capture continuous-time dynamics from sparse or irregular sensor data.
- Climatic forcing is modeled using a simplified stochastic weather generator parameterized by seasonal reference evapotranspiration patterns

and rainfall occurrence probabilities. While this approach enables systematic stress testing under controlled variability, it fails to represent multiyear persistence, regime shifts, or compound extremes. As a result, anticipatory behavior is constrained, particularly in scenarios with high evaporative demand or infrequent rainfall. Incorporating time-series foundation models, such as PatchTST, for weather forecasting could facilitate more realistic climate-aware decision-making and support evaluations under non-stationary climatic conditions.

- The lack of an explicit world model limits long-term planning and counterfactual analysis. Although the current approach emphasizes reactive control under uncertainty, advanced planning capabilities may become essential under stricter water constraints, heightened drainage sensitivity, or deeper soil profiles. Future work envisions extending latent world-model architectures that jointly learn system dynamics, uncertainty, and control. These enhancements aim to improve the realism and robustness of stress testing, rather than to ensure consistent performance gains.

Therefore, the robustness examined in this study pertains specifically to stochastic variability and model mismatch under a fixed climatic parameterisation. It should not be construed as evidence of generalization across non-stationary climate regimes or long-term distribution shifts.

Effect of increased physical complexity. The soil-water balance model in this study was deliberately simplified to isolate the effects of control strategies under consistent soil and climatic conditions. The chosen parameterisation assumes a homogeneous root zone, a single storage variable, and aggregated representations of drainage and evapotranspiration stress. Introducing further physical complexity, such as soil heterogeneity, layered profiles, preferential flow, or dynamic crop growth influencing root water uptake and evapotranspiration, would increase state uncertainty, extend memory effects, and intensify nonlinear interactions.

Under these conditions, rule-based strategies (Scenario 1) would likely require frequent retuning of tension thresholds and irrigation doses across different soils and seasons. This sensitivity is especially evident in soils with higher drainage coefficients or narrower optimal tension ranges, where minor parameter mismatches can cause oscillatory irrigation behavior and increased losses. Physics-based reinforcement learning (Scenario 2) may also become

more susceptible to structural model mismatches when key nonlinearities are omitted, such as misrepresentation of evapotranspiration reduction or redistribution processes. In the absence of explicit mechanisms to address parameter uncertainty, such as domain randomization, this may result in increased plant stress.

The hybrid approach (Scenario 3) is anticipated to remain advantageous primarily by correcting systematic residual errors resulting from simplified physics, such as delayed stress onset or underestimated drainage. However, its benefits are expected to appear as enhanced robustness and safer behavior under model mismatches, rather than as universal superiority across all performance metrics. These considerations support the adoption of a progressive validation protocol in future work, where physical fidelity is incrementally increased and policies are evaluated for transfer performance and robustness to physically meaningful perturbations.

Climatic forcing and non-stationarity. The stochastic weather generator used in this study introduces variability through seasonal patterns and stochastic noise in rainfall occurrence and reference evapotranspiration. Although this configuration allows for controlled scenario comparisons, it does not explicitly capture regime shifts, interannual persistence, or long-term climatic trends. As a result, the robustness of the learned policies under non-stationary climatic conditions is only partially evaluated.

A more rigorous evaluation would test controllers across distinct climatic regimes, such as arid, temperate, and tropical environments, using historical weather records or synthetic trajectories generated from climate model outputs. In these scenarios, fixed rule-based strategies would likely require frequent recalibration, especially when rainfall frequency or evaporative demand diverges from the assumptions underlying the rule parameters. Learning-based controllers would encounter distribution shifts relative to their training data, which could degrade performance unless adaptation mechanisms are implemented.

The hybrid neuro-physical approach is anticipated to be especially effective in these contexts, as residual learning can address persistent biases resulting from unmodeled climate effects. However, this benefit depends on explicitly addressing nonstationarity during training or adaptation. Future research will therefore prioritize cross-climate evaluation and transfer, consistent with EMS best practices for robustness analysis under climatic uncertainty.

Residual neural correction. The residual neural component models discrepancies not captured by simplified soil-water balance equations. In the current implementation, this component is trained in a supervised manner before reinforcement learning, utilizing simulated trajectories with stochastic perturbations to the physical model to emulate unmodeled effects and observation noise. The residual model remains fixed during policy training and is applied in inference mode within the environment.

The residual function f_{res} is implemented as a lightweight multilayer perceptron operating in discrete time. Its inputs include current soil-water tension, control action, rainfall, and reference evapotranspiration, and its output provides an additive correction to the physical prediction. While this design maintains physical interpretability and computational efficiency, it does not explicitly model uncertainty in the correction or adapt online to changing conditions.

Future work will investigate residual learning strategies that incorporate uncertainty quantification, online adaptation, and continuous-time formulations, particularly as higher-frequency sensor data become available.

Transferability to other environmental systems. The proposed framework demonstrates transferability through its modelling and control structure, rather than by directly reusing a trained policy. Application to other environmental systems necessitates identification of four analogous components: (i) a process-based model f_{phys} that describes dominant system dynamics, such as mass balance equations for reservoir storage or atmospheric transport models for air quality; (ii) control actions that represent management decisions; (iii) stochastic drivers that capture exogenous forcing; and (iv) an objective function that encodes system-specific trade-offs.

Within this structure, learning-based or hybrid controllers can reshape trade-offs under uncertainty by complementing simplified physical models with data-driven adaptation. Thus, transferability resides in the general decision-support paradigm that integrates process knowledge, stochastic forcing, and adaptive control, rather than in the direct deployment of a specific irrigation policy or neural architecture across domains.

5. Conclusion and perspectives

This study examined the potential of physics-informed reinforcement learning for intelligent irrigation control under stochastic climatic forcing. Three

control paradigms were systematically compared within a common physics-based soil-water environment: a rule-based heuristic strategy, reinforcement learning interacting directly with a simplified physical model, and a hybrid neuro-physical approach in which the physical dynamics were augmented by a learned residual correction. The comparison was conducted under identical soil parameterisations and meteorological forcing, enabling a controlled assessment of how increasing levels of learning influence irrigation performance.

The results indicate that no single strategy outperformed all others across every performance criterion. Rule-based control offers strong guarantees for stress avoidance and interpretability, especially in soil configurations with moderate storage capacities and conservative tension thresholds. However, its reactive nature results in excessive irrigation and drainage losses, particularly when drainage sensitivity or rainfall variability increases. Physics-based reinforcement learning significantly enhances water-use efficiency by leveraging delayed hydrological responses and cumulative effects present in the soil-water balance. Nevertheless, under the selected soil and meteorological parameters, this efficiency gain is accompanied by increased vulnerability to extreme stress episodes, highlighting sensitivity to structural model mismatch and limited anticipation of climatic variability.

The hybrid neuro-physical approach modifies this trade-off by reducing extreme stress events and drainage losses while retaining most of the efficiency gains achieved by learning-based control. Notably, this improvement was observed across the examined ranges of soil storage, drainage behavior, and stochastic meteorological forcing, suggesting enhanced robustness within the considered parameterisation rather than universal superiority. By addressing systematic discrepancies introduced by simplified physics, such as delayed stress onset or underestimated losses, the residual neural component serves to complement the physical model. Reference Appendix C further clarifies how the proposed formulation differs in intent and application from classical Richards-equation-based and ecohydrological models, emphasizing its function as a control-aware abstraction rather than a purely descriptive simulator.

In addition to quantitative performance, this study clarifies when and why learning-based controllers outperform heuristic rules. Learning-based approaches are particularly advantageous when irrigation objectives require balancing multiple, often competing, goals such as stress mitigation, water conservation, and loss reduction under variable climatic conditions. In these scenarios, fixed thresholds tailored to specific soil or weather conditions of-

ten lack robustness, whereas learning-based policies can adapt to evolving system trajectories influenced by soil properties and meteorological factors. The Neural ODE serves a corrective function by learning residual dynamics atop an interpretable physical model, thereby enhancing robustness to model simplifications while preserving physical consistency and transparency.

Operationally, the findings indicate that the suitability of control paradigms depends on soil characteristics, climatic variability, and data availability. Rule-based strategies are effective in low-risk, low-complexity environments with stable soils and limited sensing infrastructure. In contrast, learning-based and hybrid approaches are increasingly relevant as evaporative demand rises, rainfall variability intensifies, and soils display pronounced nonlinear responses through drainage and stress mechanisms. In these contexts, adaptability to parameter uncertainty and stochastic influences becomes more valuable than the additional computational and data demands, especially when incremental training or transfer learning can be utilized.

The limitations identified in this study suggest several avenues for future research. First, the current daily Markovian state representation does not adequately capture long-term dependencies arising from deeper soils, layered profiles, or delayed redistribution processes. Integrating neural-controlled differential equations provides a principled approach for modeling continuous-time dynamics based on sparse and asynchronous sensor data. Second, dependence on simplified stochastic weather generators restricts anticipatory capabilities; integrating data-driven weather forecasting models or climate model outputs would facilitate more rigorous robustness assessments under non-stationary conditions. Third, the lack of an explicit world model constrains planning and counterfactual analysis, highlighting the need for future development of latent world-model architectures that jointly learn system dynamics, uncertainty, and control.

In summary, this study demonstrates that physics-informed learning constitutes a robust and flexible framework for addressing the trade-offs inherent in intelligent irrigation under uncertainty. Rather than supplanting physical modeling or agronomic expertise, hybrid neuro-physical approaches provide a complementary pathway toward adaptive, robust, and interpretable decision-support systems for climate-resilient water management, grounded in explicit soil and meteorological parameterization.

Appendix A. Daily integration scheme for the two-layer bucket model

This appendix details the daily numerical integration of the intermediate two-layer bucket soil–water model used as an extension of the single-reservoir formulation. The model is designed to capture vertical heterogeneity, delayed redistribution, and distinct evaporation and transpiration processes while remaining compatible with daily closed-loop control and reinforcement learning.

State variables and parameters

The soil profile is discretized into two layers. At day t, the state is given by

$$\mathbf{S}_t = \left(S_t^{(1)},\, S_t^{(2)}\right),$$

where $S_t^{(1)}$ and $S_t^{(2)}$ (mm) denote the soil-water storage in the upper and lower layers, respectively. Each layer $\ell \in \{1, 2\}$ is characterized by a maximum storage $S_{\max}^{(\ell)}$, field capacity $S_{\mathrm{fc}}^{(\ell)}$, and wilting point $S_{\mathrm{wp}}^{(\ell)}$.

Daily forcing includes rainfall R_t, irrigation I_t, reference evapotranspiration $\mathrm{ET0}_t$, and crop coefficient Kc_t. The irrigation efficiency is denoted by η_I.

Step 1: Infiltration into the upper layer

Daily water inputs are first applied to the upper soil layer:

$$U_t = \eta_I I_t + R_t. \tag{A.1}$$

The intermediate storage after infiltration is

$$S_t^{(1,+)} = \min\left(S_t^{(1)} + U_t,\; S_{\max}^{(1)}\right), \tag{A.2}$$

and surface runoff is defined as

$$\mathrm{Runoff}_t = \max\left(0,\; S_t^{(1)} + U_t - S_{\max}^{(1)}\right). \tag{A.3}$$

Step 2: Soil evaporation from the upper layer

Potential crop evapotranspiration is computed as

$$\mathrm{ETc}_t^{\mathrm{pot}} = Kc_t\,\mathrm{ET0}_t. \tag{A.4}$$

A fixed fraction $\lambda_E \in [0,1]$ is assigned to potential soil evaporation:

$$E_t^{\mathrm{pot}} = \lambda_E\,\mathrm{ETc}_t^{\mathrm{pot}}. \tag{A.5}$$

Evaporation is limited by soil moisture availability through a stress factor

$$f_E^{(1)}(S) = \mathrm{clip}\left(\frac{S - S_{\mathrm{wp}}^{(1)}}{S_{\mathrm{fc}}^{(1)} - S_{\mathrm{wp}}^{(1)}},\, 0,\, 1\right), \tag{A.6}$$

leading to actual evaporation

$$E_t = E_t^{\mathrm{pot}}\, f_E^{(1)}\left(S_t^{(1,+)}\right). \tag{A.7}$$

The updated upper-layer storage is

$$S_t^{(1,++)} = \max\left(0,\; S_t^{(1,+)} - E_t\right). \tag{A.8}$$

Step 3: Transpiration and root water uptake

Potential transpiration is defined as

$$T_t^{\mathrm{pot}} = (1 - \lambda_E)\,\mathrm{ETc}_t^{\mathrm{pot}}. \tag{A.9}$$

Each layer contributes to transpiration according to a root fraction $\rho^{(\ell)}$, with $\rho^{(1)} + \rho^{(2)} = 1$. A layer-specific stress factor is defined as

$$f_T^{(\ell)}(S) = \mathrm{clip}\left(\frac{S - S_{\mathrm{wp}}^{(\ell)}}{S_{\mathrm{fc}}^{(\ell)} - S_{\mathrm{wp}}^{(\ell)}},\, 0,\, 1\right). \tag{A.10}$$

The effective transpiration stress is

$$\bar{f}_T = \rho^{(1)} f_T^{(1)}\left(S_t^{(1,++)}\right) + \rho^{(2)} f_T^{(2)}\left(S_t^{(2)}\right), \tag{A.11}$$

yielding actual transpiration

$$T_t = T_t^{\mathrm{pot}}\,\bar{f}_T. \tag{A.12}$$

Layer-wise uptake is allocated as

$$T_t^{(\ell)} = T_t \, \frac{\rho^{(\ell)} f_T^{(\ell)}\left(S_t^{(\ell)}\right)}{\sum_{j=1}^{2} \rho^{(j)} f_T^{(j)}\left(S_t^{(j)}\right) + \varepsilon}, \tag{A.13}$$

where ε is a small constant to avoid division by zero.

Updated storages after transpiration are

$$S_t^{(1,+++)} = \max\left(0,\ S_t^{(1,++)} - T_t^{(1)}\right), \tag{A.14}$$

$$S_t^{(2,+)} = \max\left(0,\ S_t^{(2)} - T_t^{(2)}\right). \tag{A.15}$$

Step 4: Vertical redistribution between layers

Delayed percolation from the upper to the lower layer is modeled as

$$Q_{12,t} = k_{12} \, \max\left(0,\ S_t^{(1,+++)} - S_{\text{fc}}^{(1)}\right), \tag{A.16}$$

where $k_{12} \in [0, 1]$ is a redistribution coefficient.

Storages are updated as

$$S_t^{(1,\text{final})} = S_t^{(1,+++)} - Q_{12,t}, \tag{A.17}$$

$$S_t^{(2,++)} = \min\left(S_t^{(2,+)} + Q_{12,t},\ S_{\text{max}}^{(2)}\right). \tag{A.18}$$

Step 5: Deep drainage loss

Drainage from the lower layer is computed as

$$D_t = k_d \, \max\left(0,\ S_t^{(2,++)} - S_{\text{fc}}^{(2)}\right), \tag{A.19}$$

where $k_d \in [0, 1]$ is the drainage coefficient.

Step 6: State update

The final state for the next day is given by

$$S_{t+1}^{(1)} = S_t^{(1,\text{final})}, \tag{A.20}$$

$$S_{t+1}^{(2)} = S_t^{(2,++)} - D_t. \tag{A.21}$$

This six-step integration preserves mass balance, enforces physical bounds, and introduces delayed responses through vertical redistribution, while remaining computationally efficient and numerically stable for daily control applications.

Appendix B. Stepwise integration pathway toward ecohydrological formulations

This appendix outlines a stepwise pathway to increase ecohydrological realism in the irrigation environment while preserving a controlled increase in model complexity. The steps are ordered from the current lumped bucket representation to a multilayer Richards-equation-based formulation. Each step specifies the governing state variables, flux parameterisations, and a practical discrete-time integration procedure compatible with daily decision cycles.

Appendix B.1. Step 0: Single-layer root-zone bucket (baseline)

State and drivers. Let S_t (mm) denote root-zone storage. Daily climate drivers are R_t (mm), $ET0_t$ (mm day^{-1}), and Kc_t.

Fluxes. Crop evapotranspiration is computed as

$$ET_{c,t} = Kc_t \, ET0_t \, f_{\mathrm{ET}}(\psi_t), \tag{B.1}$$

with $\psi_t = f_{\mathrm{ret}}(S_t)$. Drainage is $D_t = D(S_t)$ (e.g., activated above field capacity).

Update.

$$S_{t+1} = \mathrm{clip}(S_t + \eta_I I_t + R_t - ET_{c,t} - D_t,\ 0,\ S_{\max}) \, . \tag{B.2}$$

Appendix B.2. Step 1: Explicit partition of evapotranspiration (E–T separation)

To align with ecohydrological conventions, evapotranspiration is decomposed into soil evaporation E_t and plant transpiration T_t.

Partition.

$$ET_{c,t} = E_t + T_t, \qquad E_t = (1 - f_c) \, K_{e,t} \, ET0_t, \qquad T_t = f_c \, K_{cb,t} \, ET0_t \, f_T(\psi_t), \tag{B.3}$$

where f_c is fractional cover, $K_{e,t}$ is an evaporation coefficient, and $K_{cb,t}$ is a basal crop coefficient. The stress response $f_T(\psi_t) \in [0, 1]$ acts primarily on transpiration.

Update. Replace $ET_{c,t}$ in the Step 0 balance by $E_t + T_t$.

Appendix B.3. Step 2: Multi-layer bucket (intermediate ecohydrological structure)

We discretize the soil profile into L layers with storages $S_t^{(\ell)}$ (mm), $\ell = 1, \ldots, L$. Layer 1 is near-surface; deeper layers represent the root zone and sub-root zone.

State.

$$\mathbf{S}_t = \left(S_t^{(1)}, \ldots, S_t^{(L)}\right), \qquad \psi_t^{(\ell)} = f_{\text{ret}}^{(\ell)}\left(S_t^{(\ell)}\right). \tag{B.4}$$

Infiltration and runoff. Let $P_t = R_t + \eta_I I_t$ be total input. Optionally include runoff Q_t:

$$I_{\text{in},t}^{(1)} = \max(P_t - Q_t,\, 0). \tag{B.5}$$

Vertical redistribution (parameterised percolation). Define inter-layer drainage fluxes $Q_t^{(\ell\to\ell+1)}$ such as

$$Q_t^{(\ell\to\ell+1)} = k_\ell \, \max\left(S_t^{(\ell)} - S_{\text{fc}}^{(\ell)},\, 0\right), \tag{B.6}$$

where k_ℓ controls delayed redistribution. Bottom drainage is

$$D_t = Q_t^{(L\to L+1)}. \tag{B.7}$$

Evaporation and transpiration allocation across layers. Evaporation extracts from the top layer only:

$$E_t \leftarrow \text{removes water from } S_t^{(1)}. \tag{B.8}$$

Transpiration is distributed by root fractions $\rho^{(\ell)}$ (with $\sum_\ell \rho^{(\ell)} = 1$) and layer stress:

$$T_t^{(\ell)} = T_t \, \frac{\rho^{(\ell)}\, f_T(\psi_t^{(\ell)})}{\sum_{j=1}^{L} \rho^{(j)}\, f_T(\psi_t^{(j)}) + \varepsilon}. \tag{B.9}$$

Discrete update (explicit daily). For $\ell = 1, \ldots, L$:

$$S_{t+1}^{(1)} = \text{clip}\Big(S_t^{(1)} + I_{\text{in},t}^{(1)} - Q_t^{(1\to2)} - E_t - T_t^{(1)},\; 0,\; S_{\max}^{(1)}\Big), \tag{B.10}$$

$$S_{t+1}^{(\ell)} = \text{clip}\Big(S_t^{(\ell)} + Q_t^{(\ell-1\to\ell)} - Q_t^{(\ell\to\ell+1)} - T_t^{(\ell)},\; 0,\; S_{\max}^{(\ell)}\Big), \quad \ell = 2, \ldots, L. \tag{B.11}$$

Notes. This step introduces ecohydrological complexity (vertical heterogeneity, delayed response, root uptake depth profiles) while preserving numerical stability and computational efficiency suitable for reinforcement learning.

Appendix B.4. Step 3: Root-zone growth and time-varying rooting depth

Let the effective rooting depth $Z_r(t)$ vary in time (e.g., logistic growth), inducing time-varying root fractions $\rho^{(\ell)}(t)$:

$$Z_r(t) = Z_{r,\min} + (Z_{r,\max} - Z_{r,\min})\,(1 - \exp(-\kappa t))\,, \tag{B.12}$$

and define $\rho^{(\ell)}(t)$ by allocating root density over layers whose depth intervals fall within $[0, Z_r(t)]$. This increases memory effects and shifts transpiration extraction deeper as the season progresses.

Appendix B.5. Step 4: Richards-equation-inspired fluxes (semi-physical closure)

To move closer to Richards dynamics without solving a PDE, define interlayer fluxes using unsaturated conductivity $K(\theta)$ and matric head gradients:

$$Q_t^{(\ell\to\ell+1)} \approx K(\theta^{(\ell)}) \left(\frac{h^{(\ell)} - h^{(\ell+1)}}{\Delta z} + 1 \right) \Delta t, \tag{B.13}$$

where $\theta^{(\ell)}$ is volumetric water content derived from storage, and $h^{(\ell)}$ is pressure head. This introduces physically interpretable gradients while retaining a layered ODE-like update.

Appendix B.6. Step 5: Full multilayer Richards equation (reference ecohydrological model)

A full ecohydrological reference can be defined by the 1D Richards equation with a root-uptake sink:

$$\frac{\partial \theta(z,t)}{\partial t} = \frac{\partial}{\partial z} \left[K(\theta) \left(\frac{\partial h}{\partial z} + 1 \right) \right] - S_{\text{root}}(h, z, t). \tag{B.14}$$

Discretisation (outline). Using L layers, define $\theta_t^{(\ell)}$ and $h_t^{(\ell)}$ per layer. Compute conductivities $K^{(\ell)}$ and interface fluxes $q^{(\ell+1/2)}$ using hydraulic functions (e.g., van Genuchten–Mualem). Update via an implicit scheme for stability:

$$\boldsymbol{\theta}_{t+1} \text{ solves } \boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t = \Delta t\, \mathbf{F}(\boldsymbol{\theta}_{t+1}) - \Delta t\, \mathbf{S}_{\text{root}}(\boldsymbol{\theta}_{t+1}), \tag{B.15}$$

typically requiring Newton iterations. Boundary conditions represent infiltration at the surface and free drainage or fixed head at the bottom.

Use in the present framework. This step is most suitable as a high-fidelity "reference simulator" for offline benchmarking, stress testing, and residual-learning targets, rather than for large-scale RL training loops.

Appendix B.7. Summary

The described pathway enables a systematic transition from a single-bucket representation (Step 0) to an ecohydrologically grounded, layered model that incorporates delayed responses and heterogeneity (Steps 2 to 4), culminating in a Richards-equation-based reference (Step 5). This approach facilitates transparent model development, controlled increases in complexity, and reproducible assessment of controller robustness as physical realism increases.

Appendix C. Purpose and control coupling of alternative soil–water modelling formulations

This appendix clarifies the conceptual purpose, physical realism, and degree of control coupling for the various soil–water modelling formulations examined in this study. The aim is to position the proposed model in relation to classical Richards-equation-based and ecohydrological formulations, and to highlight the respective advantages and limitations of each approach within closed-loop irrigation control frameworks.

Appendix C.1. Proposed physics-based and hybrid control-aware formulation

The soil–water model used in this study is a reduced, discrete-time representation of root-zone water balance, specifically designed for closed-loop irrigation control. It integrates mass conservation, physically interpretable fluxes such as evapotranspiration, drainage, and storage, and incorporates stress-modulated plant response within a daily decision-making framework consistent with operational irrigation practices.

A key characteristic of this formulation is its control awareness. Rather than providing a high-fidelity description of soil hydrodynamics at fine spatial or temporal scales, the model is designed to: (i) remain numerically stable under repeated policy exploration, (ii) support large numbers of simulation episodes required by reinforcement learning, (iii) expose meaningful and actionable state variables to the controller, and (iv) allow systematic isolation of control effects from physical modelling assumptions.

The hybrid extension augments this structure with a learned residual correction that compensates for systematic modelling errors while maintaining the underlying physical structure. This design allows learning-based controllers to explicitly address model discrepancies, thereby enhancing robustness under stochastic climatic conditions without compromising interpretability or computational efficiency.

Advantages.

The model aligns closely with daily irrigation decision cycles and the availability of sensor data. It demonstrates computational efficiency that is suitable for reinforcement learning applications and sensitivity analyses. There is an explicit separation among physical dynamics, control logic, and learned correction components. The approach maintains stable behavior under stochastic forcing and during exploratory control actions.

Limitations.

Vertical soil heterogeneity and fine-scale redistribution processes are represented through parameterization rather than explicit resolution. Memory effects extending beyond the daily time step are approximated using storage dynamics instead of continuous-time processes. Predictive fidelity is reduced under conditions dominated by strong vertical gradients or preferential flow.

Appendix C.2. Richards-equation-based soil–water models

Richards-equation-based models offer a physically rigorous framework for describing unsaturated flow in soils by resolving pressure head gradients, hydraulic conductivity, and moisture redistribution in continuous space and time. These models are extensively used in hydrology and soil physics to investigate infiltration, redistribution, and drainage under well-defined boundary and initial conditions.

However, Richards-based formulations are primarily descriptive rather than control-oriented. Their numerical complexity, stiffness, and sensitivity to parameter uncertainty present significant challenges for direct integration into closed-loop control or reinforcement learning frameworks.

Advantages.

These models exhibit high physical fidelity and provide explicit representation of vertical soil processes. They are capable of capturing delayed responses, sharp wetting fronts, and nonlinear hydraulic behavior. These models are suitable for use as reference models in benchmarking or offline validation studies.

Limitations for control.

They incur high computational costs and impose restrictive time-step requirements. Numerical instability may occur under exploratory or extreme control actions. There is limited compatibility with large-scale policy learning or online decision-making processes. It is challenging to provide interpretable, low-dimensional state representations for use by controllers.

Consequently, within the present framework, Richards-equation-based models are most appropriately regarded as high-fidelity reference simulators or as sources of training data for residual learning, rather than as operational control environments.

Appendix C.3. Classical Ecohydrological Models

Ecohydrological models serve as an intermediate approach between fully physical Richards-equation-based formulations and simplified bucket-type models. These models typically represent the soil–plant–atmosphere continuum using multiple soil layers, explicit root water uptake profiles, evapotranspiration partitioning, and simplified vertical redistribution processes. They are often formulated as systems of coupled ordinary differential equations.

These models are well suited for long-term ecosystem analysis, water balance studies, and investigations of climate–vegetation interactions, where the primary objective is to understand system behavior rather than to optimize control actions. Consequently, they have traditionally been developed as descriptive or predictive tools, not as components of closed-loop decision-making systems.

Advantages.

- Explicit and physically meaningful representation of soil–plant–atmosphere interactions.

- Improved handling of vertical heterogeneity, delayed redistribution, and depth-dependent root water uptake compared to single-layer bucket models.
- Enhanced interpretability through process-level decomposition, enabling diagnostic insight into the relative contributions of evaporation, transpiration, and layer-wise water uptake.
- Strong grounding in ecohydrological theory and empirical observations, facilitating scientific analysis and model validation.

Limitations for control integration.

- Increased state dimensionality raises the complexity of policy learning and state-space exploration for reinforcement learning, even when individual state variables remain physically interpretable.
- Longer memory effects and internal redistribution dynamics reduce controller responsiveness at daily decision scales and complicate credit assignment.
- Computational cost grows rapidly with the number of layers and nonlinear process parameterizations, limiting suitability for large-scale policy training and repeated stochastic evaluation.
- Limited native support for tight coupling with control objectives, uncertainty-aware exploration, and reproducible benchmarking under exploratory control actions.

Appendix C.4. Comparative perspective and rationale for model choice

The modelling approach adopted in this study represents a deliberate balance between physical realism and compatibility with control frameworks. Although Richards-equation-based and ecohydrological models provide enhanced descriptive capabilities, their direct integration with reinforcement learning frameworks is frequently impractical.

In contrast, the proposed formulation preserves key ecohydrological mechanisms while maintaining explicit compatibility with closed-loop control and learning. Embedding physical constraints within a control-oriented structure and incorporating learned residual dynamics enables systematic exploration

of irrigation strategies under uncertainty, while retaining numerical stability and interpretability.

Therefore, the proposed model serves as a complementary decision-support abstraction, specifically designed for adaptive irrigation control, benchmarking, and policy evaluation under stochastic climatic conditions, rather than as a substitute for detailed ecohydrological simulation.

Appendix D. Stability and well-posedness of the control-aware soil–water model

This appendix presents a mathematical analysis of the proposed daily soil–water dynamics, establishing the following properties:

(i) *Well-posedness (existence and uniqueness)*: for any admissible initial condition and any bounded climatic forcing and irrigation input, the state update equations admit a *unique* solution at each daily time step, yielding a well-defined state trajectory over the entire simulation horizon.

(ii) *Positive invariance of the physical domain*: the soil–water storage remains within a physically meaningful bounded set for all time, i.e., if the initial storage lies within the admissible domain $[0, S_{\max}]$, then all subsequent states generated by the update equations remain in this domain, ensuring mass conservation and physical consistency.

(iii) *Bounded-input bounded-state stability (BIBS)*: under bounded climatic forcing and bounded irrigation actions, the resulting soil-water state remains bounded for all time. Specifically, the system dynamics exhibit an input-to-state-like stability property, such that variations in external forcing result in proportionate and bounded variations in soil-water storage, thereby preventing unphysical divergence or numerical instability.

These properties justify the model's use as a numerically stable closed-loop environment for control and reinforcement learning.

Appendix D.1. Model definition

We consider the daily storage update (Eq. (4)):

$$S_{t+1} = \mathrm{clip}(S_t + U_t - L_t(S_t),\ 0,\ S_{\max})\,, \tag{D.1}$$

where

$$U_t := \eta_I I_t + R_t \geq 0$$

is the total daily input (irrigation + rainfall), and

$$L_t(S_t) := ET_{c,t}(S_t) + D(S_t) \geq 0$$

is the total loss, combining evapotranspiration and drainage. Evapotranspiration is defined as

$$ET_{c,t}(S_t) = Kc_t\, ET0_t\, f_{\mathrm{ET}}\big(\psi_t\big), \qquad \psi_t = f_{\mathrm{ret}}(S_t). \tag{D.2}$$

In the hybrid setting (Scenario 3), the simulator maintains S_t and $\psi_t = f_{\mathrm{ret}}(S_t)$, and applies a residual correction in tension space:

$$\psi_{t+1} = \psi_{t+1}^{\mathrm{phys}} + \Delta\psi_t, \qquad S_{t+1} = f_{\mathrm{ret}}^{-1}(\psi_{t+1}), \tag{D.3}$$

with $\Delta\psi_t = f_\theta(\psi_t, I_t, R_t, ET0_t)$.

Appendix D.2. Assumptions

We state standard boundedness and regularity assumptions consistent with operational agro-hydrological practice.

1. **Bounded inputs.** There exist finite constants $I_{\max}$, $R_{\max}$, $ET0_{\max}$ and $Kc_{\max}$ such that

   $$0 \leq I_t \leq I_{\max}, \quad 0 \leq R_t \leq R_{\max}, \quad 0 \leq ET0_t \leq ET0_{\max}, \quad 0 \leq Kc_t \leq Kc_{\max}.$$

2. **Stress factor boundedness.** The evapotranspiration stress reduction satisfies

   $$0 \leq f_{\mathrm{ET}}(\psi) \leq 1 \quad \text{for all admissible } \psi.$$

3. **Drainage nonnegativity.** The drainage function satisfies $D(S) \geq 0$ for all $S \in [0, S_{\max}]$.
4. **Retention curve properties.** The mapping $f_{\mathrm{ret}} : [0, S_{\max}] \to [\psi_{\min}, \psi_{\max}]$ is continuous, strictly monotone, and invertible on its image. Moreover, both f_{ret} and f_{ret}^{-1} are Lipschitz on their respective domains with constants L_{ret} and L_{ret}^{-1}.
5. **Residual boundedness (hybrid only).** The learned residual is bounded:

   $$|\Delta\psi_t| \leq \Delta\psi_{\max} \quad \text{for all admissible inputs.}$$

   In practice this is enforced by architecture choice (e.g., tanh output) and/or explicit clipping.

Existence and uniqueness follow directly from the explicit nature of the update equations. The state transition is defined as a deterministic, single-valued mapping $S_{t+1} = F(S_t, I_t, d_t)$ composed of continuous elementary functions (addition, multiplication, thresholding, and clipping). As no implicit equation or fixed-point problem is involved, a unique next-state value exists for every admissible input and initial condition.

Appendix D.3. Positive invariance and boundedness

Theorem 1 (Positive invariance of the storage domain). *For any initial condition $S_0 \in [0, S_{\max}]$ and any admissible input sequence, the trajectory generated by* (D.1) *satisfies*

$$S_t \in [0, S_{\max}] \quad \textit{for all } t \geq 0.$$

Proof 1. *By definition,* $\mathrm{clip}(x, 0, S_{\max}) \in [0, S_{\max}]$ *for any real* x*. Therefore, if* $S_t \in [0, S_{\max}]$*, then* S_{t+1} *defined by* (D.1) *belongs to* $[0, S_{\max}]$*. Since* $S_0 \in [0, S_{\max}]$*, the result follows by induction.*

Interpretation. Theorem 1 formalizes that the simulator is *physically admissible by construction*: storage cannot become negative and cannot exceed the prescribed capacity even under exploratory actions, which is essential for stable RL training.

Appendix D.4. Bounded-input bounded-state stability (BIBS)

Theorem 2 (BIBS stability of the physical update). *Under Assumptions 1–3, the storage trajectory of* (D.1) *is bounded for any bounded input sequence. In particular,*

$$0 \leq S_t \leq S_{\max} \quad \forall t,$$

and the losses satisfy the uniform bound

$$0 \leq ET_{c,t}(S_t) \leq Kc_{\max} ET0_{\max}, \qquad 0 \leq D(S_t) \leq \sup_{S \in [0, S_{\max}]} D(S).$$

Proof 2. *The state bound is Theorem 1. For evapotranspiration, by* (D.2) *and Assumption 2,*

$$ET_{c,t}(S_t) = Kc_t ET0_t f_{\mathrm{ET}}(\psi_t) \leq Kc_{\max} ET0_{\max}.$$

Drainage is nonnegative by Assumption 3, and bounded on the compact interval $[0, S_{\max}]$ *if* D *is continuous (standard for threshold/linear forms used here).*

Appendix D.5. Stability of the hybrid residual coupling

We now show that the hybrid correction in tension space preserves boundedness and well-posedness, provided the residual is bounded and the retention mapping is invertible on a bounded interval.

Theorem 3 (Well-posedness and boundedness of the hybrid update). *Assume Theorem 1 holds for the physical predictor and Assumptions 4–5 hold. If ψ_{t+1} in* (D.3) *is projected onto $[\psi_{\min}, \psi_{\max}]$ (equivalently, if $\Delta\psi_t$ is bounded tightly enough such that $\psi_{t+1} \in [\psi_{\min}, \psi_{\max}]$), then the hybrid update is well-defined and yields*

$$S_{t+1} \in [0, S_{\max}] \quad \forall t.$$

Proof 3. *By Theorem 1, the physical predictor yields $S_{t+1}^{\text{phys}} \in [0, S_{\max}]$, hence $\psi_{t+1}^{\text{phys}} = f_{\text{ret}}(S_{t+1}^{\text{phys}}) \in [\psi_{\min}, \psi_{\max}]$ by Assumption 4. If ψ_{t+1} is ensured to remain in $[\psi_{\min}, \psi_{\max}]$, then $S_{t+1} = f_{\text{ret}}^{-1}(\psi_{t+1})$ is well-defined and belongs to $[0, S_{\max}]$ by invertibility of f_{ret} on this interval.*

Practical remark (implementation detail).. In a control-aware simulator, it is appropriate to enforce the projection

$$\psi_{t+1} \leftarrow \text{clip}(\psi_{t+1}^{\text{phys}} + \Delta\psi_t,\ \psi_{\min}, \psi_{\max})$$

before applying f_{ret}^{-1}. This preserves physical admissibility without altering the conceptual role of the residual.

Appendix D.6. Incremental stability (optional Lipschitz bound)

To quantify sensitivity, we provide a simple incremental bound. Let S_t and $\tilde{S}_t$ be two trajectories driven by the same exogenous forcing but possibly different actions, and define $\delta_t := |S_t - \tilde{S}_t|$.

Proposition 1 (One-step incremental bound (physical update)). *Assume $D(\cdot)$ and $f_{\text{ET}}(f_{\text{ret}}(\cdot))$ are Lipschitz on $[0, S_{\max}]$ with constants L_D and L_{ET} (uniformly in t through bounded Kc_tET0_t). Then the pre-clipped map satisfies*

$$|S_{t+1}^{\text{raw}} - \tilde{S}_{t+1}^{\text{raw}}| \leq (1 + L_D + L_{ET})\,|S_t - \tilde{S}_t| + \eta_I |I_t - \tilde{I}_t|,$$

and after clipping,

$$|S_{t+1} - \tilde{S}_{t+1}| \leq |S_{t+1}^{\text{raw}} - \tilde{S}_{t+1}^{\text{raw}}|,$$

since $\text{clip}(\cdot, 0, S_{\max})$ *is non-expansive.*

Proof 4. *Triangle inequality gives the bound for the raw update. Non-expansiveness of clipping follows from the fact that projection onto a convex interval is 1-Lipschitz.*

Interpretation. Proposition 1 formalizes the numerical stability of the environment under exploratory control. The one-step map is Lipschitz, and the clipping operation is non-expansive, which together prevent uncontrolled amplification of small perturbations.

Appendix D.7. Summary

The proposed environment demonstrates the stability necessary for closed-loop benchmarking. The soil–water state remains physically bounded for all admissible actions and climatic forcing, ensuring positive invariance. Hydrological fluxes remain bounded under bounded drivers, reflecting bounded-input bounded-state (BIBS) stability. Additionally, the hybrid residual coupling preserves well-posedness under mild boundedness and invertibility assumptions.

Furthermore, the stability and well-posedness analysis establishes continuous dependence of the soil–water state on initial conditions, control inputs, and climatic forcing. This property is a necessary prerequisite for sensitivity analysis. Explicit analytical parametric sensitivity metrics are not derived in this appendix. Instead, sensitivity is investigated empirically through controlled scenario analyses, consistent with the control-oriented and benchmarking-focused scope of the model.

Collectively, these properties support the control-aware purpose of the formulation. The focus is on ensuring stable, interpretable, and reproducible closed-loop simulations under uncertainty, rather than replicating all fine-scale soil hydrodynamics.

Appendix E. Stability and well-posedness of the ecohydrological soil–water model

This appendix presents a mathematical analysis of a class of eco-hydrological soil–water models that extend the single-reservoir formulation by incorporating vertical layering, root water uptake distribution, and delayed redistribution processes. The purpose is to establish that, under standard regularity assumptions, such models remain suitable for closed-loop simulation and control-oriented experimentation.

Specifically, we establish the following properties:

(i) *Well-posedness (existence and uniqueness)*: for any admissible initial condition and bounded climatic forcing and irrigation inputs, the eco-

hydrological update equations admit a unique solution at each daily time step.

(ii) *Positive invariance of the physical domain*: soil–water storage in each layer remains within a physically meaningful bounded domain for all time.

(iii) *Bounded-input bounded-state (BIBS) stability*: under bounded forcing and bounded control actions, all state variables and hydrological fluxes remain bounded, preventing numerical blow-up or unphysical behavior.

These properties justify the use of ecohydrological formulations as numerically stable reference environments or intermediate models for control-oriented studies, while highlighting the increased complexity relative to the control-aware formulation adopted in the main text.

Appendix E.1. Model definition

We consider an L-layer ecohydrological soil profile with daily time discretization. Let

$$\mathbf{S}_t = \left(S_t^{(1)}, \ldots, S_t^{(L)}\right)$$

denote the vector of soil–water storages (mm) in each layer. The daily update is written component-wise as

$$S_{t+1}^{(\ell)} = \mathrm{clip}\Big(S_t^{(\ell)} + U_t^{(\ell)} - E_t^{(\ell)} - T_t^{(\ell)} + Q_t^{(\ell-1\to\ell)} - Q_t^{(\ell\to\ell+1)},\ 0,\ S_{\max}^{(\ell)}\Big), \tag{E.1}$$

for $\ell = 1, \ldots, L$, with the conventions $Q_t^{(0\to1)} = 0$ and $Q_t^{(L\to L+1)} = D_t$ (deep drainage).

Here:

- $U_t^{(1)} = \eta_I I_t + R_t$ and $U_t^{(\ell)} = 0$ for $\ell > 1$;
- $E_t^{(1)}$ denotes soil evaporation from the top layer;
- $T_t^{(\ell)}$ denotes root water uptake (transpiration) from layer ℓ;
- $Q_t^{(\ell\to\ell+1)}$ denotes delayed vertical redistribution.

Appendix E.2. Assumptions

We impose standard boundedness and regularity assumptions commonly adopted in ecohydrological modelling.

1. **Bounded inputs.** Irrigation, rainfall, and climatic drivers are uniformly bounded:
$$0 \le I_t \le I_{\max}, \quad 0 \le R_t \le R_{\max}, \quad 0 \le ET0_t \le ET0_{\max}.$$
2. **Storage bounds.** Each layer has a finite capacity $S_{\max}^{(\ell)}$.
3. **Stress functions.** Evaporation and transpiration stress factors $f_E^{(\ell)}(\cdot)$ and $f_T^{(\ell)}(\cdot)$ satisfy
$$0 \le f_E^{(\ell)}(\cdot), f_T^{(\ell)}(\cdot) \le 1,$$
and are continuous and Lipschitz on $[0, S_{\max}^{(\ell)}]$.
4. **Redistribution fluxes.** Vertical fluxes $Q_t^{(\ell \to \ell+1)}$ are continuous, nonnegative, and bounded on compact domains (e.g., linear or threshold-based percolation).
5. **Root distribution.** Root fractions $\rho^{(\ell)}$ satisfy $\rho^{(\ell)} \ge 0$ and $\sum_{\ell=1}^{L} \rho^{(\ell)} = 1$.

Appendix E.3. Well-posedness

Theorem 4 (Existence and uniqueness). *For any admissible initial state* $\mathbf{S}_0$ *and any bounded input sequence, the ecohydrological update* (E.1) *admits a unique solution at each time step.*

Proof 5. *Each update equation is explicit and single-valued, composed of continuous functions (addition, multiplication, stress functions, redistribution fluxes) followed by clipping onto a closed interval. No implicit equations or fixed-point problems arise. Therefore, a unique next-state value exists for each layer and each time step.*

Appendix E.4. Positive invariance

Theorem 5 (Positive invariance of layered storage). *If* $S_0^{(\ell)} \in [0, S_{\max}^{(\ell)}]$ *for all* ℓ*, then*
$$S_t^{(\ell)} \in [0, S_{\max}^{(\ell)}] \quad \forall t \ge 0, \ \forall \ell.$$

Proof 6. *The clipping operator in* (E.1) *projects each component onto its admissible interval. The result follows by induction on* t*.*

Appendix E.5. Bounded-input bounded-state stability

Theorem 6 (BIBS stability). *Under the stated assumptions, the ecohydrological model is bounded-input bounded-state stable: bounded climatic forcing and bounded irrigation actions imply bounded storages and bounded fluxes.*

Proof 7. *By positive invariance, all $S_t^{(\ell)}$ remain bounded. Stress factors are bounded by construction, implying bounded evaporation and transpiration. Redistribution and drainage fluxes are bounded on compact domains. Therefore, all state variables and fluxes remain bounded for all time.*

Appendix E.6. Incremental stability (optional)

Under additional Lipschitz assumptions on stress and redistribution functions, the ecohydrological update is locally Lipschitz in the state vector, yielding continuous dependence on initial conditions and inputs. This property ensures numerical robustness but does not imply global contraction due to nonlinear stress and redistribution mechanisms.

Appendix E.7. Summary

Ecohydrological soil–water models admit well-defined, physically admissible, and bounded solutions under standard assumptions. The state remains positively invariant, fluxes remain bounded under bounded forcing, and the dynamics depend continuously on initial conditions and inputs.

However, compared to the control-aware single-reservoir formulation, ecohydrological models introduce increased state dimensionality, stronger nonlinearities (via stress functions and redistribution), and longer memory effects. These features enhance physical interpretability but complicate policy learning and closed-loop control.

Accordingly, ecohydrological formulations are best suited as high-fidelity reference models or intermediate validation layers in progressive control frameworks, rather than as primary environments for large-scale reinforcement learning.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56, FAO, Rome.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling & Software 40, 1–20. doi:10.1016/j.envsoft.2012.09.011.

Berkenkamp, F., Schoellig, A.P., Krause, A., 2017. Safe model-based reinforcement learning with stability guarantees. Advances in Neural Information Processing Systems 30, 908–918.

Beucler, T., et al., 2021. Implicit learning of convective organization explains precipitation stochasticity. Nature 597, 672–677. doi:10.1038/s41586-021-03860-w.

Fatichi, S., et al., 2016. Ecosystem and land surface modelling in a changing climate. Hydrology and Earth System Sciences 20, 455–478. doi:10.5194/hess-20-455-2016.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., Reed, P.M., 2016. Coping with deep uncertainty in water management: Policy search under uncertainty. Environmental Modelling & Software 81, 60–74. doi:10.1016/j.envsoft.2016.02.006.

Giuliani, M., et al., 2021. Reinforcement learning and control of water systems: An overview. Environmental Modelling & Software 141, 105045. doi:10.1016/j.envsoft.2021.105045.

Hadka, D., Reed, P.M., 2013. Borg: An auto-adaptive many-objective evolutionary computing framework. Environmental Modelling & Software 37, 97–111. doi:10.1016/j.envsoft.2012.07.004.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21, 602–614. doi:10.1016/j.envsoft.2006.01.004.

Jones, H.G., et al., 2022. Smart irrigation systems: A review of control strategies and technologies. Agricultural Water Management 260, 107300. doi:10.1016/j.agwat.2021.107300.

Karniadakis, G.E., et al., 2021. Physics-informed machine learning. Nature Reviews Physics 3, 422–440. doi:10.1038/s42254-021-00314-5.

Monteith, J.L., 1965. Evaporation and environment. Symposia of the Society for Experimental Biology 19, 205–234.

Perkins, S., et al., 2023. Safe reinforcement learning for real-world control systems: A survey. IEEE Transactions on Artificial Intelligence 4, 1–18. doi:10.1109/TAI.2022.3220730.

Rackauckas, C., Ma, Y., Martensen, J., Warner, P., Zubov, K., Supekar, S., Skinner, D., Ramadhan, A., Edelman, A., Perdikaris, P., 2020. Universal differential equations for scientific machine learning. Proceedings of the National Academy of Sciences 117, 29041–29048. doi:10.1073/pnas.2001336117.

Rackauckas, C., et al., 2021. Scientific machine learning through physics-informed neural networks and universal differential equations. Computing in Science & Engineering 23, 18–31. doi:10.1109/MCSE.2020.3042241.

Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water. FAO, Rome.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N., 2021. Stable-baselines3: Reliable reinforcement learning implementations. Journal of Machine Learning Research 22, 1–8. URL: `https://www.jmlr.org/papers/v22/20-1364.html`.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling & Software 22, 1543–1556. doi:10.1016/j.envsoft.2007.02.004.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204. doi:10.1038/s41586-019-0912-1.

Rodríguez-Iturbe, I., Porporato, A., 2004. Ecohydrology of Water-Controlled Ecosystems. Cambridge University Press.

Rolnick, D., et al., 2022. Tackling climate change with machine learning. ACM Computing Surveys 55, 1–96. doi:10.1145/3485128.

Seneviratne, S.I., et al., 2010. Investigating soil moisture–climate interactions in a changing climate. Earth-Science Reviews 99, 125–161.

Seneviratne, S.I., et al., 2021. Weather and climate extreme events in a changing climate. Nature Climate Change 11, 964–974. doi:10.1038/s41558-021-01092-9.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. 2 ed., MIT Press, Cambridge, MA.

Vereecken, H., et al., 2007. Modeling soil processes: Review, key challenges, and new perspectives. Vadose Zone Journal 6, 749–762.

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2022. Integrating physics-based modeling with machine learning: A survey. Nature Reviews Physics 4, 366–382. doi:10.1038/s42254-022-00437-4.

Yang, T., et al., 2021. Reinforcement learning for water resources management: A review. Water Resources Research 57, e2020WR028838. doi:10.1029/2020WR028838.