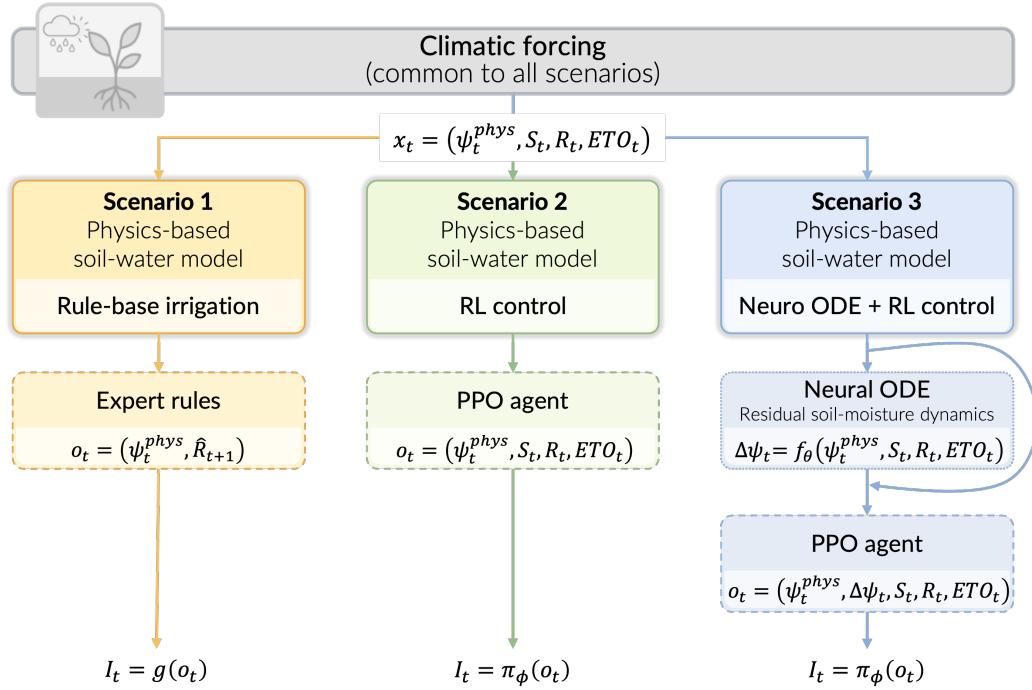


# Graphical Abstract

**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<sup>a</sup>, Philippe Berton<sup>b</sup>, Fabien Dauriac<sup>b</sup>

<sup>a</sup>*Université de Technologie de Tarbes Occitanie Pyrénées, Laboratoire Génie de Production, 47 Avenue d'Azereix, Tarbes, 65000, France*

<sup>b</sup>*Rives et Eaux du Sud-Ouest, BP 449 Chemin de Lalette, Tarbes, 65000, France*

---

## Abstract

Making decisions in physical systems often depends on process models. These models can be easy to understand, but may not match the real system exactly. At the same time, standard reinforcement learning (RL) methods can handle uncertainty but do not always behave reliably. This problem worsens when the system is only partially observed, responds slowly, or is affected by random external factors, as in this study. We suggest a hybrid RL approach that combines physics-based rules with learned corrections to control changing systems. Our method clearly separates hidden physical states from the measurements of sensors. This adds learned corrections to the environment model, so the controller can adjust when the model does not match reality. RL is used to make decisions, while the learned corrections fix errors in the system's behavior. This is done without losing physical accuracy or stability. We compare three decision-making approaches under the same random external factors: rule-based control, standard RL, and hybrid RL with learned corrections via Neural Ordinary Differential Equations (ODE). We measured stress, efficiency, and robustness. The results show that model-based RL is more efficient but can lead to extreme system states. The hybrid RL with corrections greatly reduces these extreme states while maintaining most of the efficiency, shifting the balance between robustness and efficiency. Although we tested this on an irrigation system, the method can also be used for predictive maintenance, building energy management, and other physical systems that are only partly observed and have delays or uncertainty.

### *Keywords:*

Intelligent irrigation, Environmental decision support systems,  
Control-aware modelling, Physics-informed reinforcement learning, Hybrid

## 1. Related Work

This section reviews previous research on irrigation modelling and control, emphasizing the progression from process- and rule-based methods to contemporary learning-based and hybrid physics-informed approaches. The present study integrates neural ordinary differential equations with adaptive reinforcement learning to enhance the robustness of irrigation control under changing climatic conditions, an area that has received limited attention in prior research. The objective is to situate this work within the broader context of environmental modelling and to identify the specific gaps addressed by the proposed method.

### 1.1. *Process-based modelling of irrigation systems*

Process-based models are fundamental tools in irrigation modelling, representing interactions among soil, water, and plants through basic physical principles and evapotranspiration formulas (Bo et al., 2025). Examples include simple bucket models and FAO-56 methods, as implemented in tools such as AquaCrop (Raes et al., 2009). These models are transparent, require minimal data, and are therefore valuable for scenario comparison and policy research (Amarasinghe et al., 2020).

However, process-based models rely on simplified assumptions regarding soil heterogeneity, root water uptake, and boundary conditions. Their predictive accuracy declines under highly variable or extreme weather conditions, as they are unable to capture all relevant complexities and uncertainties (Verbruggen et al., 2025). Recent studies indicate that, although these physical models are essential, they are insufficient on their own for effective irrigation management in a changing climate (Fatichi et al., 2016; Seneviratne et al., 2021).

### 1.2. *Rule-based and heuristic irrigation control*

Rule-based irrigation methods, frequently derived from expert recommendations or agricultural guidelines, remain widely utilized. These systems employ fixed thresholds for soil moisture, crop stress, or water loss. Their simplicity, transparency, and ease of implementation make them advantageous in data-scarce environments (Fereres and Soriano, 2007).

However, these rule-based systems lack adaptability to changing climatic conditions (Padilla-Nates et al., 2025). Research demonstrates that fixed rules are often ineffective under unpredictable rainfall or when soil response is delayed, motivating the exploration of adaptive control methods (Jones et al., 2022). Consequently, rule-based control is now primarily regarded as an initial approach rather than a sustainable solution for climate-resilient irrigation.

### *1.3. Reinforcement learning for irrigation and water management*

Reinforcement learning (RL) is a versatile framework for sequential decision-making in dynamic and uncertain environments (Sutton and Barto, 2018). In water management, RL has been applied to reservoir operation, canal control, and irrigation planning. Numerous studies indicate that RL can outperform fixed rule-based systems (Yang et al., 2021; Giuliani et al., 2021).

Despite these successes, data-driven RL methods encounter significant challenges in environmental applications. They typically require extensive training data, may violate physical constraints, and can yield solutions that are difficult to interpret or trust (Li et al., 2022). Recent reviews highlight that ensuring physical consistency and reliability remains a major obstacle for RL in environmental systems (Rolnick et al., 2022; Reichstein et al., 2019).

In environmental modelling, adaptive control and policy search methods have been investigated to support decision-making under uncertainty and multiple objectives. Policy search approaches offer systematic means to balance robustness, efficiency, and risk in water management (Giuliani et al., 2016). Specialized algorithms, such as the Borg framework, are capable of identifying diverse optimal control strategies for multi-objective environmental problems (Hadka and Reed, 2013).

Within this context, reinforcement learning is regarded as a valuable complement rather than a replacement for existing decision-making approaches in environmental management. RL shares conceptual similarities with policy search and multi-objective optimization but offers advantages in leveraging temporal information and managing delayed system responses (Padilla-Nates et al., 2025).

### *1.4. Physics-informed and hybrid learning approaches*

To address these challenges, physics-informed machine learning has gained prominence. This approach incorporates physical knowledge into data-driven

models. Research demonstrates that integrating physical constraints enhances model performance, stability, and interpretability, particularly in data-limited scenarios (Willard et al., 2022; Karniadakis et al., 2021).

Neural Ordinary Differential Equations (Neural ODEs) provide an effective means of integrating data-driven and physical models. They enable neural networks to approximate unknown or unmodeled components in continuous-time systems (Rackauckas et al., 2020). In environmental science, Neural ODEs and related techniques are increasingly employed to augment, rather than replace, physical models, thereby maintaining interpretability while improving accuracy (Rackauckas et al., 2021; Beucler et al., 2021).

### *1.5. Hybrid modelling for control under uncertainty*

Physics-informed learning has been extensively studied for system identification and prediction, but its integration with reinforcement learning for control in environmental modelling remains limited. Recent research on safe and model-based RL underscores the importance of incorporating physical constraints to enhance control reliability and prevent unsafe or unrealistic solutions (Berkenkamp et al., 2017; Perkins et al., 2023).

Few studies have systematically compared rule-based control, reinforcement learning, and hybrid neurophysical control within a unified modelling framework. The present work addresses this gap by evaluating these methods under identical physical and meteorological conditions, thereby ensuring fair and reproducible results. Each approach utilizes the same foundational data, such as evapotranspiration rates, soil type, and initial moisture, and is subjected to the same weather scenarios. The application of Neural ODEs to capture unmodeled dynamics and enhance irrigation control robustness under climate variability remains underexplored (Höge et al., 2022). Accordingly, this study investigates control strategies that increase model complexity within a physics-based irrigation context.

Policy search and multi-objective optimization methods have been applied to irrigation management (Giuliani et al., 2016, 2021); however, this study distinguishes itself through its control setup and model construction. While most policy search approaches emphasize fixed or seasonal strategies, the present work focuses on daily control with real-time feedback and delayed system responses. The proposed framework explicitly separates the effects of learning-based control and model correction by comparing rule-based, RL, and hybrid neuro-physical methods under consistent conditions. This design

facilitates a clear assessment of how learning and model adjustments influence system performance, stability, and interpretability—factors not directly addressed in previous studies. To evaluate these aspects, the study employs specific metrics: performance (irrigation efficiency and water savings), stability (temporal variability of actions and system response to new inputs), and interpretability (simplicity and clarity of control policies). These measures provide a comprehensive evaluation of the new approach’s ability to address the limitations of traditional methods (Huang et al., 2025).

## 2. Related Work

This section reviews the literature on irrigation modelling and control, emphasizing the shift from process- and rule-based approaches to learning-based and hybrid physics-informed methods. It introduces the present study’s approach, which integrates neural ordinary differential equations with adaptive reinforcement learning to enhance irrigation control robustness under climatic variability, a gap not extensively addressed in prior research. The objective is to situate this work within established environmental modelling paradigms and to 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 (Bo et al., 2025). 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 (Amarasinghe et al., 2020).

However, process-based models rely on simplified assumptions regarding soil heterogeneity, root water uptake, and boundary conditions. Their performance often deteriorates under highly variable or extreme climatic conditions, where unmodeled dynamics and parameter uncertainty become significant (Verbruggen et al., 2025). Recent reviews indicate that, although physically based models are essential, their use alone is often 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 (Fereres and Soriano, 2007).

Heuristic controllers are inherently static and frequently fail to adapt to nonstationary climatic conditions (Padilla-Nates et al., 2025). Comparative studies demonstrate that threshold-based rules often perform suboptimally when rainfall variability or delayed soil responses are pronounced, motivating the investigation of adaptive control strategies (Jones et al., 2022). Consequently, rule-based control is increasingly regarded as 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 face significant challenges in environmental applications. These methods typically require large amounts of training data, may violate physical constraints, and can generate policies that are difficult to interpret or trust (Li et al., 2022). 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 is best regarded as a complementary adaptive control paradigm rather than a replacement for existing environmental decision-support methodologies. RL shares conceptual foundations with policy search and many-objective optimization, while offering enhanced capacity to leverage temporal structure and delayed system responses (Padilla-Nates et al., 2025).

#### *2.4. Physics-informed and hybrid learning approaches*

To address 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, particularly 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 systematically compared rule-based control, reinforcement learning, and hybrid neurophysical control within a unified environmental modelling framework. This work addresses this gap by structuring the comparison across these methodologies under identical physical and climatic conditions to ensure rigor and reproducibility. Identical physical forcing uses the same baseline evapotranspiration rates, soil properties, and initial moisture levels, while climatic forcing applies the same sequence of weather events

to each control strategy. The potential of learning residual dynamics via Neural ODEs to enhance irrigation control robustness under climatic variability has received limited attention (Höge et al., 2022). This gap motivates the present investigation of control strategies that 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. To assess these aspects, this study employs defined metrics. Performance is evaluated using measures of irrigation efficiency and water savings. Stability is assessed by the variation in control actions over time and the system's responsiveness to changing environmental inputs. Interpretability is evaluated by examining the transparency and simplicity of the control policies derived through the framework. These metrics provide a comprehensive assessment of the effectiveness of the proposed approach in addressing the limitations of traditional methods (Huang et al., 2025).

### 3. Materials and Methods

This section describes the materials and methods employed in the study. First, the system dynamics, study scope, and data sources are clarified. The irrigation management problem is then formally defined. The subsequent sections present the physics-based soil–water environment used to simulate daily dynamics and introduce the three control scenarios under investigation: rule-based control, reinforcement learning, and hybrid neuro-physical reinforcement learning. The section concludes with a detailed account of the experimental design, training procedures, and evaluation protocols imple-

mented to ensure reproducibility and fair comparison across control strategies.

### 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.

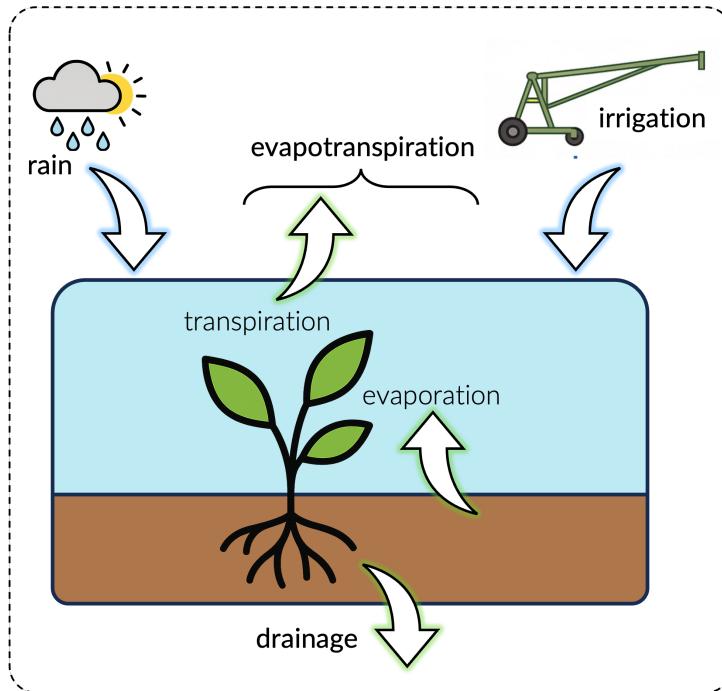


Figure 1: Schematic representation of the soil–water balance processes considered in the physical model including water inputs from rainfall and irrigation, losses due to evapotranspiration (evaporation and transpiration), and drainage from the root zone.

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 evapo-

transpiration 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 simulation 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 represented as a single, homogeneous root-zone compartment with an effective storage capacity and a monotonic retention relationship linking soil-water storage to matric potential. This bucket-type model omits vertical heterogeneity, preferential flow paths, layered hydraulic properties, and lateral redistribution. These simplifications are notable, as vertical heterogeneity and related processes significantly influence water retention and drainage in real soils. In agricultural practice, neglecting such complexity may result in suboptimal irrigation scheduling, since layered soils can cause uneven percolation and impact root water uptake. While incorporating these processes would increase model complexity due to higher dimensionality, nonlinearities, and delayed, history-dependent soil moisture responses, this study prioritizes tractability while acknowledging the associated trade-off in realism (Vereecken, 2007; Fatichi et al., 2016). Previous quantitative assessments indicate that such simplifications can lead to under- or overestimation of water availability by 10–20%

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*

Simulated data generated by a physics-based soil–water balance model driven by stochastic climatic forcing were used in this study. This simulation-based approach facilitates controlled experimentation, systematic comparison of control strategies, and reproducibility across diverse hydroclimatic conditions, while mitigating confounding factors such as data scarcity, sensor noise, and site-specific calibration requirements.

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.

While this study centers on simulated data, the proposed framework is designed for transferability to real-world irrigation systems. In operational contexts, soil–water storage is generally not directly observable; instead, soil–water status is monitored using tensiometers that measure soil matric potential  $\psi_t$ . For field deployment, simulated  $\psi_t$  can be replaced with tensiometer readings via a standardized protocol that integrates these measurements into the control framework. Rainfall inputs may be sourced from on-site rain gauges or open-access meteorological datasets, such as those provided by national weather services or global reanalysis products. The observation structure adopted, based on  $\psi_t$  and climatic variables, is intended to reflect these practical constraints.

Successful field deployment requires addressing several specific considerations. Tensiometers must be calibrated and maintained to ensure accurate readings. Real-time measurements should be integrated into a centralized data aggregation system accessible to the control framework. Challenges include maintaining sensor connectivity and data accuracy across large agricultural areas. Investment in robust communication networks or cloud-based platforms may be necessary to support reliable data flow.

Potential discrepancies between simulated and field conditions, including soil heterogeneity and microclimatic variations, necessitate location-specific calibration of control model parameters. Stakeholder collaboration is essential to address operational challenges such as equipment costs, training requirements, and user engagement. A pilot study is recommended to validate system performance in real-world settings prior to broader implementation.

### *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  $t \in \{0, \dots, T - 1\}$ . Irrigation decisions are made once per day, consistent with the temporal resolution of the meteorological forcing and the operational granularity targeted in this study.

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. (The term latent signifies that it is not directly measurable in typical operational settings.) Instead, soil-water status is monitored using tensiometers that provide soil matric potential (tension)  $\gamma C9_t$  (cbar), which serves as the primary observable variable.

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

$$\psi_t = f_{\text{ret}}(S_t), \quad S_t = f_{\text{ret}}^{-1}(\psi_t), \quad (1)$$

where  $f_{\text{ret}}(\cdot)$  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 sensor-accessible. Addressing partial observability involves exploring the use of filtering techniques, such as Kalman filtering, or implementing recurrent policies that leverage historical observation data. Preliminary results indicate that incorporating belief-state estimation can enhance control accuracy under scenarios with limited data availability. However, challenges remain in designing efficient filtering strategies that can operate with the constraints of real-time decision-making and the variability of field conditions.

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.

### 3.3.2. Climate drivers

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

$$d_t := (R_t, ET0_t, Kc_t), \quad (2)$$

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

$$d_t \sim \mathcal{P}_d. \quad (3)$$

### 3.3.3. Physics-based dynamics (mass balance)

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

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

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  $\text{clip}(\cdot)$  enforces physical bounds. Drainage is computed as

$$D_t := D(S_t), \quad (5)$$

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

$$ET_{c,t} = K c_t ET0_t f_{ET}(\psi_t), \quad (6)$$

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).

### 3.3.4. Sequential decision-making and objective

At each day  $t$ , the controller selects an irrigation action

$$I_t \in [0, I_{\max}], \quad (7)$$

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:

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

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

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

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  $\pi$  is evaluated by the expected discounted return:

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

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  $\psi_t$  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.

Scenario 1 couples the physics-based environment with a deterministic rule-based policy representative of common operational practices. Serving as the zero-learning control, it provides a baseline experiment against which more advanced learning-based methodologies can be contrasted. 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  $S_0 = S_{fc}$ , with  $\psi_0 = f_{ret}(S_0)$ .

This scenario serves as an interpretable, low-cost baseline that is robust by design but non-adaptive; its parameters remain fixed across seasons and do not optimize trade-offs under variable conditions.

### 3.5.1. 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).

### 3.5.2. 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:

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

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

The residual model then predicts an additive correction as follows:

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

and the hybrid prediction is

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

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

*Residual model architecture.* In our implementation,  $f_\theta$  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

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

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

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

In the current implementation,  $\psi_{t+1}^{\text{?ref}}$  is generated by injecting stochastic perturbations into the physical update (emulating unresolved processes and observation noises). The perturbation magnitudes and types are selected to mirror real-world uncertainties typically encountered in soil-water dynamics, such as variabilities in evapotranspiration response, drainage sensitivity, and irrigation efficiency. These choices are informed by empirical studies on soil heterogeneity and environmental variability. The parameters were optimized using Adam with a robust regression loss (Smooth L1). After pretraining,  $f_\theta$  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  $\eta_I$  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  $\psi_{t+1}^{\text{ref}}$  is then obtained by propagating the perturbed model forward in time, while the nominal physical prediction  $\psi_{t+1}^{\text{phys}}$  is computed using the unperturbed model. The residual target is defined as

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

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 describes the experimental design and evaluation protocol implemented to ensure fair, transparent, and reproducible comparisons

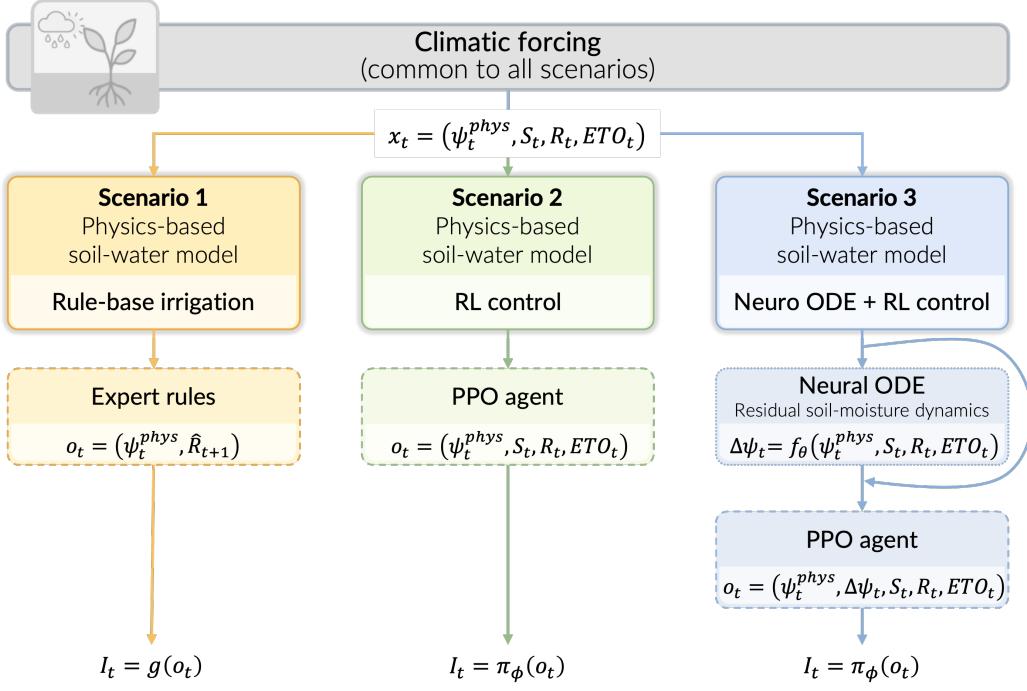


Figure 2: Overview of the three irrigation control scenarios considered in this study. Scenarios are ordered by increasing learning involvement and model–controller coupling. All scenarios share a common climatic forcing and a physics-based soil-water model, while differing in the irrigation decision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

among the three control scenarios. The design isolates the effects of the control strategy from those of the physical model and climatic forcing. All experiments utilized identical environment dynamics, weather generation procedures, and evaluation metrics, differing only in the control formulation. The protocol emphasizes configuration-driven reproducibility, controlled randomness, and consistent performance assessment across independent runs. Several hyperparameters, including learning rate, discount factor, and clipping range, were tuned based on stability and performance criteria using an iterative grid search. Stability was evaluated through convergence behavior and variance reduction, resulting in a robust configuration across scenarios. All reported hyperparameters correspond to a stable region identified during preliminary tuning and are not optimized for individual scenarios. 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  $\eta_I$ ), (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  $ET0_t$ , 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  $\eta_I$ . Unless specified otherwise, default soil parameters are used.

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

$$S_0 = S_{fc}, \quad (18)$$

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

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

where  $\psi_t$  denotes the current soil-water tension,  $I_{\max}$  is the maximum admissible daily irrigation depth, and  $\hat{R}_{t+1}$  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

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

where  $f_{\text{ET}}(\psi_t)$  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

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

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

This procedure is repeated sequentially for  $t = 0, \dots, T - 1$ , 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

$$\mathbf{o}_t = [\psi_t, S_t, R_t, \text{ET0}_t], \quad (22)$$

where  $\psi_t$  denotes soil-water tension,  $S_t$  the soil-water storage,  $R_t$  the rainfall input, and  $ET0_t$  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:

$$I_t \in [0, I_{\max}], \quad (23)$$

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 ( $\psi_t$ ) 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  $\tau_{\text{opt}}$ ; (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 ( $t = 0, \dots, T - 1$ )
$T$	day	Length of the growing season (time horizon)
<i>Soil-water state variables</i>		
$S_t$	mm	soil-water storage in the root zone (latent physical state)
$S_{\text{max}}$	mm	Maximum soil-water storage (soil capacity)
$S_{\text{fc}}$	mm	soil-water storage at field capacity
$\psi_t$	cbar	Soil matric potential (tension), observable via tensiometers
$f_{\text{ret}}$	—	soil-water retention function linking $S_t \leftrightarrow \psi_t$
<i>Hydrological fluxes</i>		

*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 $t$
$\hat{R}_{t+1}$	mm	One-day-ahead rainfall forecast used by rule-based control (Scenario 1)
$ET0_t$	$\text{mm day}^{-1}$	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
<i>Mass balance and dynamics</i>		
$\eta_I$	–	Irrigation efficiency coefficient
$f_{\text{phys}}$	–	Physics-based soil-water balance model
$f_{\text{res}}$	–	Learned residual dynamics (Neural ODE component)
<i>Decision-making and learning</i>		
$\mathbf{o}_t$	–	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 $I_t = g(\psi_t, I_{\max}, \hat{R}_{t+1}; \kappa)$
$\kappa$	–	Parameters of the irrigation rule (e.g., tension thresholds or comfort band limits)
$\pi(\cdot)$	–	Control policy (rule-based or learned)
$\pi_\theta$	–	Parametric RL policy with parameters $\theta$ (Scenarios 2–3)
<i>Reinforcement learning formulation</i>		
$r_t$	–	Immediate reward at day $t$
$\gamma$	–	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
<i>Neural ODE residual model (Scenario 3)</i>		
$f_\theta$	–	Neural network parameterizing residual correction
$\Delta\psi_t$	cbar	Residual correction to soil-water tension
$\psi_{t+1}^{\text{phys}}$	cbar	Physical model prediction of soil tension
$\psi_{t+1}$	cbar	Hybrid prediction: $\psi_{t+1}^{\text{phys}} + \Delta\psi_t$
<i>Performance indicators</i>		
$\bar{\psi}$	cbar	Mean soil matric potential over the season
$\tau_{\text{opt}}$	%	Fraction of days within optimal tension range
$I_{\text{tot}}$	mm	Total irrigation volume over the season
$D_{\text{tot}}$	mm	Total drainage loss over the season
Eff	–	Water-use efficiency metric ( $ETc/(I + R)$ )

#### 4. Results and discussions

In this study, we focus on examining various irrigation control scenarios in the context of semi-arid agricultural conditions. An important aspect of this context is climate variability and change, which can significantly impact the effectiveness and adaptability of control strategies. Non-stationarity in climatic conditions, such as regime shifts and long-term trends, presents a critical challenge to the robustness of irrigation policies. By addressing these challenges upfront, we aim to prepare the readers to better appreciate the subsequent discussion on the robustness of the learned irrigation control policies under varied and unforeseen climatic shifts.

The importance of effective irrigation control cannot be overstated, especially for farmers facing challenges related to yield losses and water-scarcity pressures. These stakes are particularly high in semi-arid regions where water management is crucial for the sustainability of agricultural practices. 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. To provide a clear overview of this process, a brief outline of the primary steps followed in the study is offered: (1) Defining the control scenarios and parameters, (2) conducting simulations using consistent configurations, and (3) analyzing model performance through specific indicators. 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 [Bennett2013PerformanceEMS]. 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.

The findings indicate that rule-based control maintains soil-water tension primarily within or near the agronomically optimal range for the crop. Specifically, this range is defined as soil-water tension between 20 and 60 cbar, a span characterized as the comfort zone for plant growth. 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, water losses occur primarily through two mechanisms: infiltration at the surface and free drainage. Infiltration involves water seeping into the soil, which can lead to losses beyond the root zone, especially when excessive irrigation coincides with rainfall events. Free drainage, on the other hand, refers to the gravitational flow of water beyond the plant root zone, which is most pronounced when soil-water storage is at or near capacity. 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 3: Seasonal dynamics of rule-based irrigation control (Scenario 1): soil-water tension, soil-water storage, irrigation and rainfall inputs, and hydrological fluxes.

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

#### 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 opti-

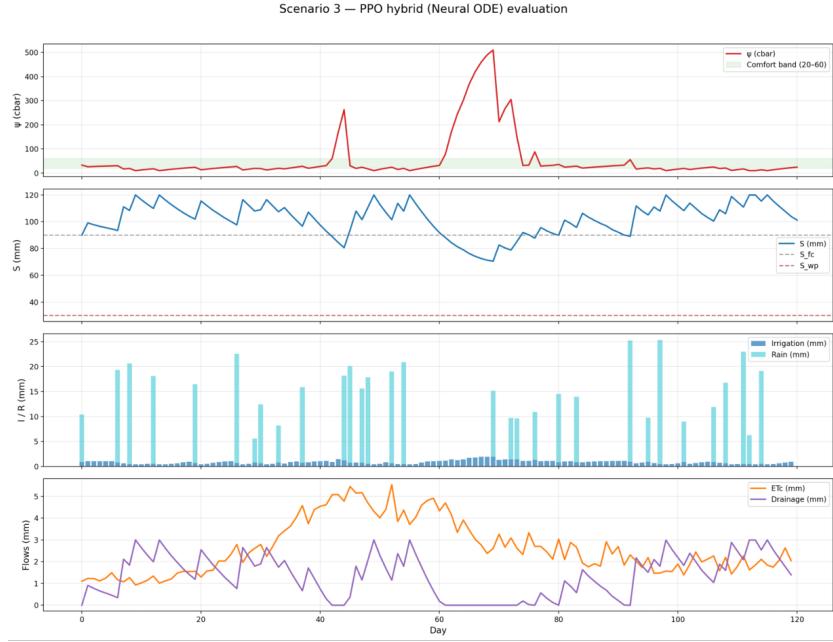


Figure 5: Seasonal dynamics of hybrid PPO control with Neural ODE-augmented dynamics (Scenario 3).

mal 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 ( $\theta_{ET}^{crit} = 80$  cbar) further shapes controller behavior by establishing a narrow boundary between non-stress and stress conditions. Within this range, learning-based controllers can im-

plement 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.

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

#### 4.5. Comparative analysis across scenarios

The following cross-scenario synthesis distills these patterns. 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 ( $\psi_{fc} \approx 33$  cbar,  $\psi_{ET}^{crit} = 80$  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.

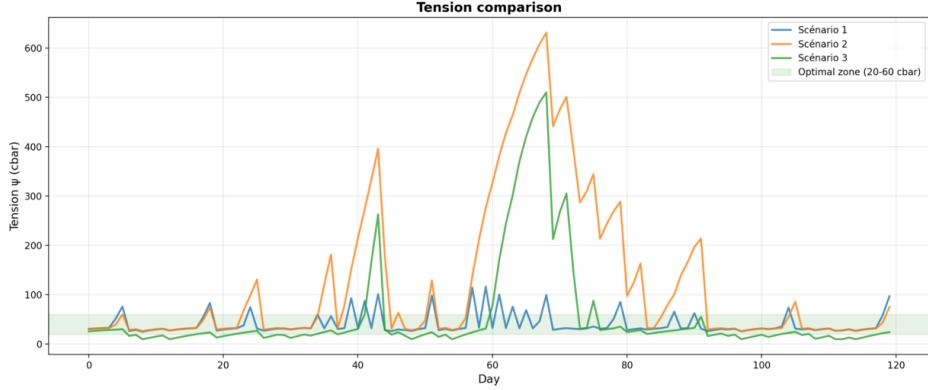


Figure 6: Comparison of soil-water tension dynamics across the three scenarios. The shaded area indicates the agronomically optimal tension range.

#### 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  $\psi_{ET}^{\text{crit}}$  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 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.

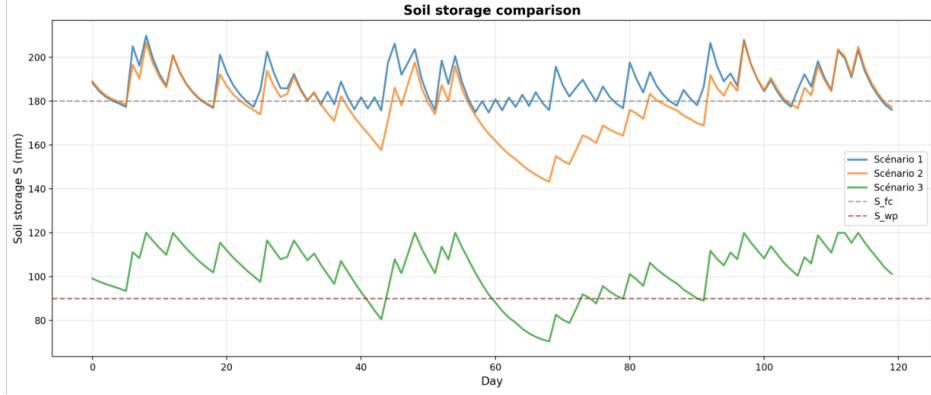


Figure 7: Comparison of soil-water storage trajectories across the three scenarios, relative to field capacity and wilting point.

#### 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 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

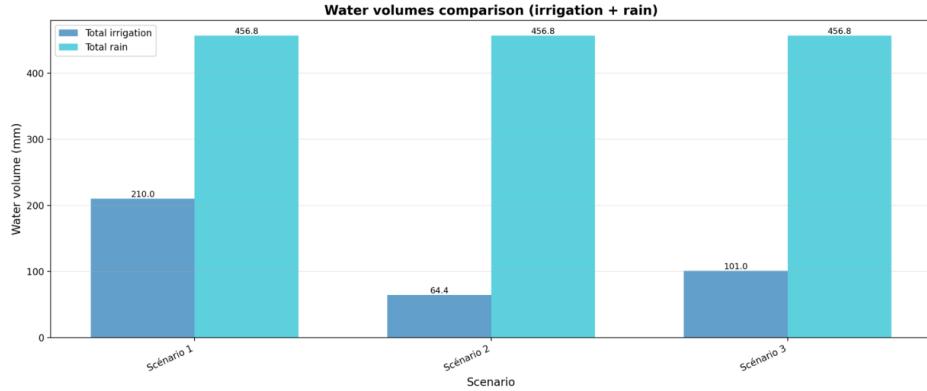


Figure 8: Comparison of cumulative irrigation and rainfall volumes across scenarios.

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.

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.

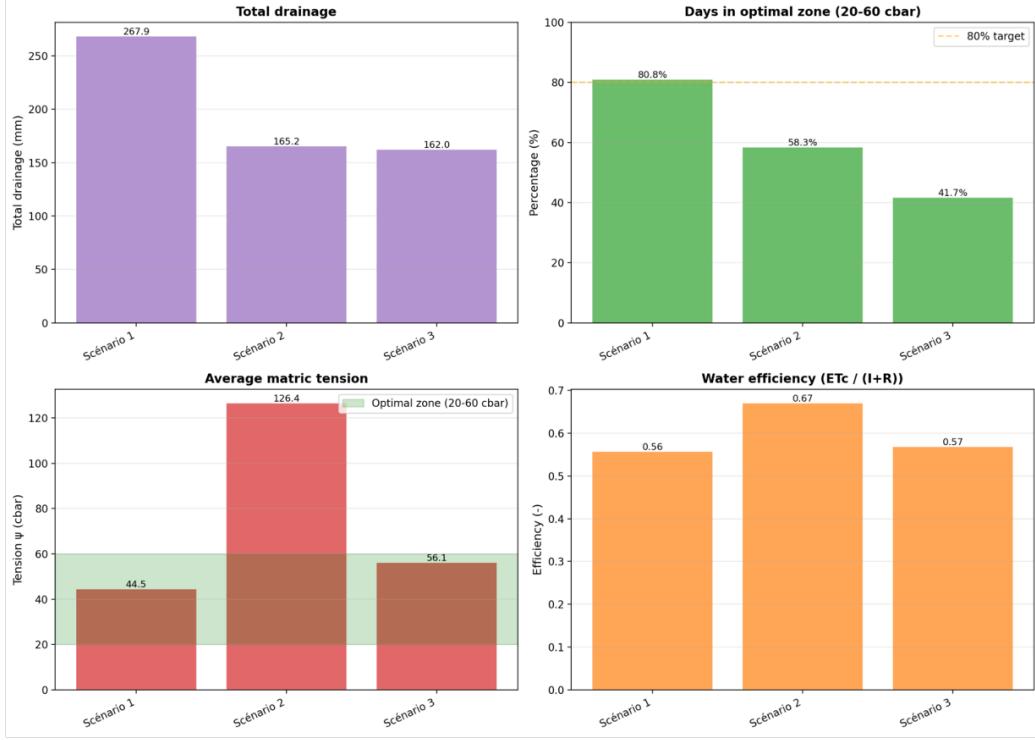


Figure 9: Comparison of aggregated performance indicators: total drainage, percentage of days in the optimal tension range, mean soil-water tension, and water-use efficiency.

#### 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 designed 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.

A specific policy recommendation derived from these findings is that hybrid controllers are particularly effective in regions where seasonal rainfall is 500 mm or less. In such conditions, the hybrid approach provides a balanced mechanism to capitalize on limited water resources while mitigating the risk of extreme plant stress. This translates to policy-ready guidance for regions facing stricter water constraints and greater climatic variability, extending

the impact of this research beyond academic exploration into real-world agricultural decision-making.

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 limitation could lead to missed midday stress events, for example, when high temperatures and solar radiation peaks cause sudden increases in evapotranspiration, stressing plants significantly by mid-afternoon. Such events may not be captured when sampling only occurs at a daily interval, potentially resulting in suboptimal irrigation decisions that may increase the risk of over-irrigation or plant stress. 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. To address these risks, 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 sys-

tematic 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 high-evaporative-demand or infrequent-rainfall scenarios. 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. However, uncertainty in the residual correction is not explicitly quantified or propagated in the current configuration. This lack of uncertainty modeling could potentially affect the robustness of the policy, as unaccounted-for uncertainties may introduce variability in decisions taken based on the residuals. Future research intends to explore strategies for incorporating uncertainty quantification within the residual learning framework to enhance policy reliability under variable 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. We welcome partnerships to validate these controllers in on-farm trials, inviting agronomists and practitioners to collaborate on field data collection and application. This collaboration aims to enhance real-world validation and uptake of our methodologies, bridging the gap between theoretical development and practical application.

## 5. Conclusion and perspectives

This section sums up the main findings and what they mean for smarter irrigation control. The results show that using physics-informed reinforcement learning can make irrigation more efficient and better able to handle changing weather, giving a strong way to manage water resources well.

This study looked at how physics-informed reinforcement learning could help make irrigation smarter when the weather is unpredictable. Picture a

vineyard in a place with sudden spring storms, where uneven rain can hurt crops or waste water. In this situation, three main control methods matter: a rule-based strategy, reinforcement learning that works with a simple physical model, and a hybrid approach where the physical model is improved by a learned correction. To compare these, we set up scenarios with the same soil settings for all methods and used weather data showing the area's changing rainfall. Each method started with the same conditions and rules to make sure the comparison was fair and could be repeated. This let us clearly see how more learning affects irrigation results. The results showed clear differences in how well the three methods managed soil water, especially in tough situations like our example vineyard.

The findings show that no single method was best in every way. Rule-based control helps avoid plant stress and is easy to understand in soils that can hold a moderate amount of water and have safe limits. But newer model-based control methods, like DDRMPC, have cut down on water use and improved control compared to traditional rule-based methods. Physics-based reinforcement learning saves more water by taking into account delayed water movement and build-up in the soil. However, with the chosen soil and weather settings, this extra efficiency also made the system more likely to face extreme stress, showing that the model can be sensitive to changes and may not always predict weather changes well.

The hybrid neuro-physical approach deals with this trade-off by lowering extreme stress and water loss while keeping most of the efficiency benefits of learning-based control. This improvement was steady across the different soil and weather conditions we tested, showing it is more reliable in these cases, though not always better in every situation. By fixing regular mistakes from simple physics, like late stress or missed water loss, the neural part improves the physical model. The hybrid approach could also work for many types of crops, like grains and fruits, especially in places with changing weather. Its ability to adjust to different soils and climates shows it could help improve water use in many farming situations.

Besides the numbers, this study explains when learning-based controllers do better than simple rules. Learning-based methods are especially useful when irrigation needs to balance several goals at once, like reducing plant stress, saving water, and cutting losses when the weather changes. In these cases, fixed limits made for certain soils or weather often do not work well. Learning-based methods are made to balance stress, water use, and losses by adjusting to changes in soil and weather. The Neural ODE makes the system

more reliable when the model is simplified by learning extra details on top of a clear physical model, so it stays consistent and easy to understand.

From a practical point of view, the best control method depends on the soil, weather changes, and how much data is available. Rule-based methods work well in simple, low-risk places with steady soils and few sensors, so they are good for small farmers who do not have advanced technology and use steady routines. On the other hand, learning-based and hybrid methods become more useful when water demand goes up, rain becomes less predictable, and soils react in more complex ways. These methods are best for orchard managers and tech-focused farming companies that can use advanced data tools and flexible strategies. In these cases, being able to adjust to unknowns and random changes is more important than the extra computer power and data needed, especially when step-by-step training or using knowledge from other systems is possible.

The limits found in this study point to several directions for future research. First, the current daily state setup does not fully capture long-term effects from deeper soils, layers, or slow water movement. Using neural-controlled differential equations is a solid way to model changes over time from sensor data that comes in at different times. Some research questions are: How can these equations be changed to better catch long-term effects in different soil layers? What happens to irrigation efficiency when deeper soil layers are added to these models? Second, relying on simple random weather generators limits how well the system can plan ahead. Adding weather forecasts or climate model data would help test how strong the system is when conditions change. Future work could ask: What water management improvements come from using better forecasting? How do these methods compare in accuracy and usefulness to simple random generators? Third, not having a clear world model makes it hard to plan and test what-if situations, showing the need to build models that learn how the system works, its uncertainty, and how to control it all together. Research could look at experiments that use these models to predict irrigation results or find out when what-if analysis can best help with irrigation decisions.

In short, this study shows that physics-informed learning gives a strong and flexible way to handle the trade-offs in smart irrigation when things are uncertain. Instead of replacing physical models or expert knowledge, hybrid neuro-physical methods offer an extra way to build adaptive, reliable, and easy-to-understand systems for water management that can handle changing climates, using clear soil and weather information.

## 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

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

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  $S_{fc}^{(\ell)}$ , and wilting point  $S_{wp}^{(\ell)}$ .

Daily forcing includes rainfall  $R_t$ , irrigation  $I_t$ , reference evapotranspiration  $ET0_t$ , and crop coefficient  $Kc_t$ . The irrigation efficiency is denoted by  $\eta_I$ .

### *Step 1: Infiltration into the upper layer*

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

$$U_t = \eta_I I_t + R_t. \quad (\text{A.1})$$

The intermediate storage after infiltration is

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

and surface runoff is defined as

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

*Step 2: Soil evaporation from the upper layer*

Potential crop evapotranspiration is computed as

$$\text{ETc}_t^{\text{pot}} = K c_t \text{ET0}_t. \quad (\text{A.4})$$

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

$$E_t^{\text{pot}} = \lambda_E \text{ETc}_t^{\text{pot}}. \quad (\text{A.5})$$

Evaporation is limited by soil moisture availability through a stress factor

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

leading to actual evaporation

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

The updated upper-layer storage is

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

*Step 3: Transpiration and root water uptake*

Potential transpiration is defined as

$$T_t^{\text{pot}} = (1 - \lambda_E) \text{ETc}_t^{\text{pot}}. \quad (\text{A.9})$$

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

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

The effective transpiration stress is

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

yielding actual transpiration

$$T_t = T_t^{\text{pot}} \bar{f}_T. \quad (\text{A.12})$$

Layer-wise uptake is allocated as

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

where  $\varepsilon$  is a small constant to avoid division by zero.

Updated storages after transpiration are

$$S_t^{(1,+++)} = \max(0, S_t^{(1,++)} - T_t^{(1)}), \quad (\text{A.14})$$

$$S_t^{(2,+)} = \max(0, S_t^{(2)} - T_t^{(2)}). \quad (\text{A.15})$$

#### *Step 4: Vertical redistribution between layers*

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

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

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

Storages are updated as

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

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

#### *Step 5: Deep drainage loss*

Drainage from the lower layer is computed as

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

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

#### *Step 6: State update*

The final state for the next day is given by

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

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

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),  $ET0_t$  ( $\text{mm day}^{-1}$ ), and  $Kc_t$ .

*Fluxes.* Crop evapotranspiration is computed as

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

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

*Update.*

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

### *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.*

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

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),  $\ell = 1, \dots, L$ . Layer 1 is near-surface; deeper layers represent the root zone and sub-root zone.

*State.*

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

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

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

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

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

where  $k_\ell$  controls delayed redistribution. Bottom drainage is

$$D_t = Q_t^{(L \rightarrow L+1)}. \quad (\text{B.7})$$

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

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

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

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

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

$$S_{t+1}^{(1)} = \text{clip} \left( S_t^{(1)} + I_{\text{in},t}^{(1)} - Q_t^{(1 \rightarrow 2)} - E_t - T_t^{(1)}, 0, S_{\text{max}}^{(1)} \right), \quad (\text{B.10})$$

$$S_{t+1}^{(\ell)} = \text{clip} \left( S_t^{(\ell)} + Q_t^{(\ell-1 \rightarrow \ell)} - Q_t^{(\ell \rightarrow \ell+1)} - T_t^{(\ell)}, 0, S_{\text{max}}^{(\ell)} \right), \quad \ell = 2, \dots, L. \quad (\text{B.11})$$

*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  $\rho^{(\ell)}(t)$ :

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

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 inter-layer fluxes using unsaturated conductivity  $K(\theta)$  and matric head gradients:

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

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:

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

*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:

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

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)):

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

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

$$ET_{c,t}(S_t) = Kc_t ET0_t f_{\text{ET}}(\psi_t), \quad \psi_t = f_{\text{ret}}(S_t). \quad (\text{D.2})$$

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

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

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

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

- 3. Drainage nonnegativity.** The drainage function satisfies  $D(S) \geq 0$  for all  $S \in [0, S_{\max}]$ .

- 4. Retention curve properties.** The mapping  $f_{\text{ret}} : [0, S_{\max}] \rightarrow [\psi_{\min}, \psi_{\max}]$  is continuous, strictly monotone, and invertible on its image. Moreover, both  $f_{\text{ret}}$  and  $f_{\text{ret}}^{-1}$  are Lipschitz on their respective domains with constants  $L_{\text{ret}}$  and  $L_{\text{ret}}^{-1}$ .

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*

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

**Proof 1.** *By definition,  $\text{clip}(x, 0, S_{\max}) \in [0, S_{\max}]$  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*

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

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

$$ET_{c,t}(S_t) = K c_t ET0_t f_{\text{ET}}(\psi_t) \leq K c_{\max} ET0_{\max}.$$

*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  $\psi_{t+1}$  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  $\psi_{t+1}$  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  $f_{\text{ret}}$  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  $f_{\text{ret}}^{-1}$ . 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  $L_{\text{ET}}$  (uniformly in  $t$  through bounded  $Kc_t \text{ET}0_t$ ). Then the pre-clipped map satisfies

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

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.

## 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.
- Amarasinghe, K., Rodolfa, K., Lamba, H., Ghani, R., 2020. Explainable machine learning for public policy: Use cases, gaps, and research directions. *Data & Policy* 2, e20. doi:10.1017/dap.2020.19.
- 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.
- Bo, Y., Yan, Y., Chen, J., Liang, Y., Zhang, Y., Liu, Y., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., 2025. Process-based modeling framework for sustainable irrigation management at the regional scale: integrating rice production, water use, and greenhouse gas emissions. *Geoscientific Model Development* 18, 3799–3818. doi:10.5194/gmd-18-3799-2025.
- Faticchi, 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.
- Fereres, E., Soriano, M.A., 2007. Deficit irrigation: A global perspective. *Agricultural Water Management* 80, 1–22. doi:10.1016/j.agwat.2005.07.014.
- 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.
- Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., Fenicia, F., 2022. Improving hydrologic models for predictions and process understanding using neural ODEs. *Hydrology and Earth System Sciences* 26, 5085–5106. doi:10.5194/hess-26-5085-2022.
- Huang, Z., Wang, Y., Hui, C., XiaoCheng, 2025. An intelligent water-saving irrigation system based on multi-sensor fusion and visual servoing control. arXiv preprint arXiv:2510.23003. doi:10.48550/arXiv.2510.23003.
- 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.
- Li, Z., Zhang, Y., Wang, J., Chen, X., Li, J., 2022. Deep learning for intelligent irrigation decision-making: A review. *Agricultural Water Management* 265, 107561. doi:10.1016/j.agwat.2022.107561.
- Monteith, J.L., 1965. Evaporation and environment. *Symposia of the Society for Experimental Biology* 19, 205–234.
- Padilla-Nates, J.P., Garcia, L.D., Lozoya, C., Orona, L., Cortes-Perez, A., 2025. Greenhouse irrigation control based on reinforcement learning. *Agronomy* 15, 2781. doi:10.3390/agronomy15122781.
- 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., Supkar, 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., Dornmann, 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>.
- 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.
- Verbruggen, W., Wårllind, D., Horion, S., Meunier, F., Verbeeck, H., Schurgers, G., 2025. Implementing a process-based representation of soil water movement in a second-generation dynamic vegetation model: application to dryland ecosystems (LPJ-GUESS-RE v1.0). *EGUphere* , 1–28doi:10.5194/egusphere-2025-1259.
- 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.