HYBRID DEEP LEARNING AND REINFORCEMENT FRAMEWORK FOR PHYSICALLY CONSISTENT PRECIPITATION CONTROL ALONG THE MOROCCAN-IBERIAN COASTS

***Imad KATIBA*[1*], *Hanae BELMAJDOUB* [1,2] & *Khalid MINAOUI* [1]**

DOI: 10.21163/GT_2026.212.06

ABSTRACT

This study introduces a hybrid framework that integrates deep learning and reinforcement learning to optimize regional precipitation along the Moroccan and Iberian coasts of the North-Eastern Atlantic. A Bidirectional Long Short-Term Memory (BiLSTM) and Multi-Layer Perceptron (MLP) network was first trained to forecast monthly rainfall from multi-variable climate sequences combining key atmospheric and oceanic variables. The predictions were then coupled with a Twin Delayed Deep Deterministic Policy Gradient (TD3) controller capable of performing small, physically coherent adjustments to selected variables such as sea-level pressure, specific humidity, and near-surface winds. Through adaptive interactions with the predictive environment, the control agent learned strategies that enhance rainfall while maintaining atmospheric stability. The resulting framework demonstrates consistent improvements in simulated precipitation and reveals spatially coherent dominance patterns of sea-level pressure and humidity in rainfall modulation. The reinforcement-based control explicitly enforces stability and energetic constraints, ensuring physically admissible atmospheric responses. Overall, the proposed hybrid AI approach provides a physically interpretable foundation for regional water management and constitutes a decision-support framework for diagnosing precipitation sensitivity rather than an operational weather modification system. It can be extended to drought-prone or broader climatic domains to support future resilience and hydrological planning.

Keywords: *Deep learning; Reinforcement learning; BiLSTM; TD3; Climate control; Precipitation optimization; Moroccan–Iberian coasts; North-Eastern Atlantic; Atmospheric dynamics; Water resource management.*

1. INTRODUCTION

Coastal and oceanic regions off Morocco and Portugal are strongly influenced by large-scale ocean-atmosphere interactions that govern regional climate variability and its predictability. As highlighted by (Li et al., 2024), multiple coupled modes such as the El Niño-Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), and the Atlantic Multidecadal Oscillation (AMO) interact through complex teleconnections, ocean heat fluxes, and atmospheric bridges, generating anomalies that affect weather and climate patterns far beyond their regions of origin. Understanding these mechanisms remains a critical challenge for improving regional forecasts and assessing their socio-ecological impacts. Despite decades of advances in numerical weather prediction and global climate modeling, significant disparities persist in the capacity to deliver high-quality regional climate information, particularly in areas with limited computational or institutional resources. As noted by (Meque et al., 2021), the ability of National Meteorological and Hydrological Services to perform reliable modeling and deliver tailored climate services is often constrained by technical and financial limitations.

[1] *Laboratory of Research in Computer Science and Telecommunications, Faculty of Sciences, Mohammed V University, Rabat, Morocco.* [*]*Corresponding author: imad_katiba@um5.ac.ma (IK); hanae.belmajdoub@um5r.ac.ma (HB); khalid.minaoui@fsr.um5.ac.ma (KM),*

[2] *Faculty of Legal, Economic and Social Sciences – Souissi Campus, Mohammed V University, Rabat, Morocco.*

This emphasizes the need for innovative approaches that can complement traditional modeling frameworks, especially in regions where conventional models struggle to represent local dynamics accurately. In this context, surrogate modeling techniques based on machine learning have emerged as promising tools to reduce systematic biases in climate model outputs. For example, (Steininger et al., 2022) demonstrate how convolutional neural networks (CNNs) can be integrated into Model Output Statistics (MOS) schemes to refine climate projections and improve accuracy for key variables such as precipitation. Similarly, the availability of standardized benchmark datasets such as WeatherBench(Rasp et al., 2020) is facilitating transparent, reproducible comparisons among data-driven forecasting methods, accelerating progress in surrogate model design.

Several recent studies have shown that hybrid deep learning architectures, large-scale neural operators, and neuro-symbolic networks can extract fine-scale structures and identify early warning signals for climate tipping points (Pathak et al., 2022; Kurth et al., 2018; Sleeman et al., 2023). By combining physical modeling with advanced AI techniques, these approaches reveal emergent behaviors in coupled systems such as the atmosphere and ocean, enhancing both predictive skill and computational efficiency. Furthermore, as reviewed by (Ladi et al., 2022), machine learning methods have become integral to climate change mitigation and adaptation, particularly in coastal and urban contexts. Neural network-based approaches are now applied across topics ranging from geoengineering to land surface temperature management, demonstrating their versatility for addressing complex environmental challenges. Extending these advances to marine and atmospheric systems, (Ojo et al., 2022; Luo et al., 2022) show that Bayesian deep learning architectures can outperform traditional models in predicting sea surface temperature variability while providing robust uncertainty estimates. Such probabilistic formulations mitigate initialization bias and represent a promising pathway for near-term regional prediction, particularly across the North Atlantic corridor near Morocco and Portugal.

Parallel to these advances in predictive modeling, reinforcement learning (RL) and predictive control frameworks have emerged as powerful tools for optimizing complex environmental processes. However, value-based RL methods often suffer from function approximation errors that lead to overestimated value functions and suboptimal policies. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm (Fujimoto et al., 2018) addresses this issue by combining double Q-learning and delayed policy updates to improve stability and reduce overestimation bias. In related applications, neural network-driven predictive control systems have been successfully deployed in real-world energy and resource management contexts. For instance, (Routray et al., 2023; Al Sayed et al., 2024) demonstrate the effectiveness of RL in optimizing heating, ventilation, and air conditioning (HVAC) systems, while (Sierla et al., 2022) highlights the decisive role of reward design and action formulation in achieving stable, multi-objective control. More recently, (Durante et al., 2024) discussed the potential of multimodal AI agents equipped with adaptive perception to enhance the performance and resilience of RL systems operating in dynamic, uncertain environments.

In parallel, recent climate assessments, including the 2025 Global Climate Report by NOAA (NOAA National Centers for Environmental Information, 2025), continue to emphasize the growing frequency of precipitation anomalies and extreme hydro-climatic events. These findings highlight the urgent need for integrated frameworks capable not only of forecasting precipitation, but also of diagnosing the sensitivity of rainfall to its governing atmospheric drivers. Building upon these challenges, this study proposes a hybrid framework that couples deep learning-based precipitation prediction with reinforcement learning-driven control to investigate regional precipitation sensitivity and modulation. Unlike most existing reinforcement learning approaches applied to climate-related problems, which typically rely on abstract optimization settings or synthetic environments, the proposed framework operates directly on realistic hydro-atmospheric variables and explicitly enforces physical stability and energetic consistency during the control process. By jointly integrating prediction and physically constrained control within a unified structure, this approach provides new insights into how atmospheric pressure, humidity, and near-surface wind fields interact to regulate precipitation patterns along the Moroccan-Iberian Atlantic corridor.

The remainder of this paper is organized as follows: Section 2 introduces the study area and the datasets used. Section 3 describes the methodology and the proposed hybrid framework combining predictive modeling and reinforcement-based control. Section 4 presents and discusses the main results, including the regional patterns of precipitation modulation and variable influence. Finally, Section 5 concludes the study and outlines future research perspectives for regional hydroclimatic regulation.

2. STUDY AREA AND DATA PROCESSING

This study focuses on the North Atlantic sector, where atmospheric and oceanic variability jointly shape regional precipitation behavior. A coherent multi-source dataset was compiled to represent the physical state of this region over more than a decade of winter seasons. The analysis focuses on the winter months (December, January, February, and March) covering the period from December 2011 to March 2023, which corresponds to the most hydrologically active phase of the year. The study area extends from 20.6° N to 42.0° N in latitude and from 23.1° W to 4.8° W in longitude, encompassing the North-Eastern Atlantic domain and the adjacent coastal sectors of Morocco and the Iberian Peninsula, as illustrated in (**Fig. 1**), which delineates the spatial limits adopted for the study.

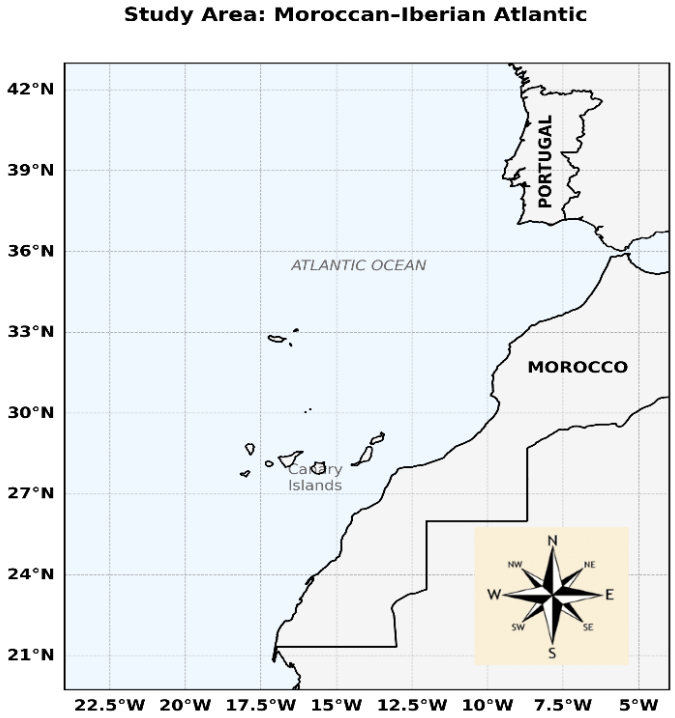

Fig. 1. Geographic extent of the Moroccan-Iberian Atlantic study area.

To ensure spatial consistency among all sources, each variable was interpolated onto a regular 0.25° grid. Precipitation data were obtained from the Integrated Multi-satellite Retrievals for GPM (IMERG) Version 07B Final Run (Huffman et al., 2023). Atmospheric variables, including sea-level pressure (SLP), 10 m wind components (U10M and V10M), specific humidity at 2 m (QV2M), and sea surface temperature (SST), were extracted from NASA's MERRA-2 reanalysis (Gelaro et al., 2017).Wind speed was calculated as the magnitude of the horizontal wind vector. Specifically, it was derived from the MERRA-2, 10 m wind components according to $WS = \sqrt{U10M^2 + V10M^2}$, ensuring consistency with the reanalysis convention.

In parallel, sea surface salinity (SSS) was obtained from the ECCO V4r4 dataset (Fukumori et al., 2021). Because these products have different native spatial resolutions, each field was resampled using bilinear interpolation applied through the SciPy griddata routine (Virtanen et al., 2020). This method estimates the value at each target grid cell as a weighted average of the four nearest source points, with weights proportional to their inverse distances along latitude and longitude. Compared to nearest-neighbor techniques, this approach produces smoother and physically continuous fields, preserving the relative spatial gradients of pressure, temperature, humidity, and salinity while aligning all variables in space and time. The operation therefore ensures both numerical stability and physical consistency, resulting in a homogeneous dataset suitable for subsequent analysis. All data were obtained from the NASA Earthdata portal (https://www.earthdata.nasa.gov).

The entire preprocessing workflow was designed to ensure internal coherence and numerical stability across the dataset. Missing values were filtered, temporal indices were standardized, and sliding sequences of four consecutive months were constructed for each grid point to provide dynamic inputs to the deep learning models. Precipitation data for March 2023 were isolated as independent targets for model validation. All variables were normalized using min-max scaling to guarantee comparability of physical magnitudes and stable convergence during subsequent computations. The resulting time series were organized into spatio-temporal matrices for each geolocation, ensuring full compatibility among atmospheric, marine, and hydrological parameters across the study domain.

An initial visualization of the spatial distribution of these variables for the first available month is presented (**Fig. 2**), which illustrates the regional gradients and large-scale contrasts that characterize the study area. This figure provides a comprehensive view of the atmospheric and oceanic structures used as the physical foundation for the predictive modeling and reinforcement-learning experiments described in the following sections.

3. METHODS

The methodology developed in this study integrates predictive modeling and reinforcement-based control to build a physically consistent framework for regional precipitation optimization. The approach relies on a hybrid architecture in which a deep learning predictor reproduces the temporal evolution of key atmospheric variables, while a reinforcement learning agent explores how small, physically admissible modifications to these variables can influence precipitation. The workflow operates in two sequential phases. First, a predictive model is trained to estimate future rainfall from multi-variable climate sequences derived from long-term reanalysis data. Then, the predicted values are used to guide a reinforcement learning agent that interacts with a virtual climate environment designed to preserve the internal dynamics of the atmosphere.

The agent's objective is to identify localized adjustment strategies that enhance precipitation while minimizing energetic cost and instability. This combined framework bridges statistical forecasting and adaptive control, allowing regional climate interactions to be represented not only as correlations but also as actionable mechanisms for hydrological modulation. This methodological design ensures that predictive learning and control adaptation remain coupled within a physically interpretable framework, where data-driven forecasts continuously inform reinforcement-based decisions. The following paragraphs describe the model architecture and the experimental configuration used for training and evaluation.

3.1. Model Architecture

The hybrid system consists of two main components: a predictive model based on a Bidirectional Long Short-Term Memory (BiLSTM) network (Graves & Schmidhuber, 2005), and a control module implemented through the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm. The *BiLSTM-MLP* predictor captures the temporal dependencies among the atmospheric and oceanic parameters that regulate precipitation dynamics.

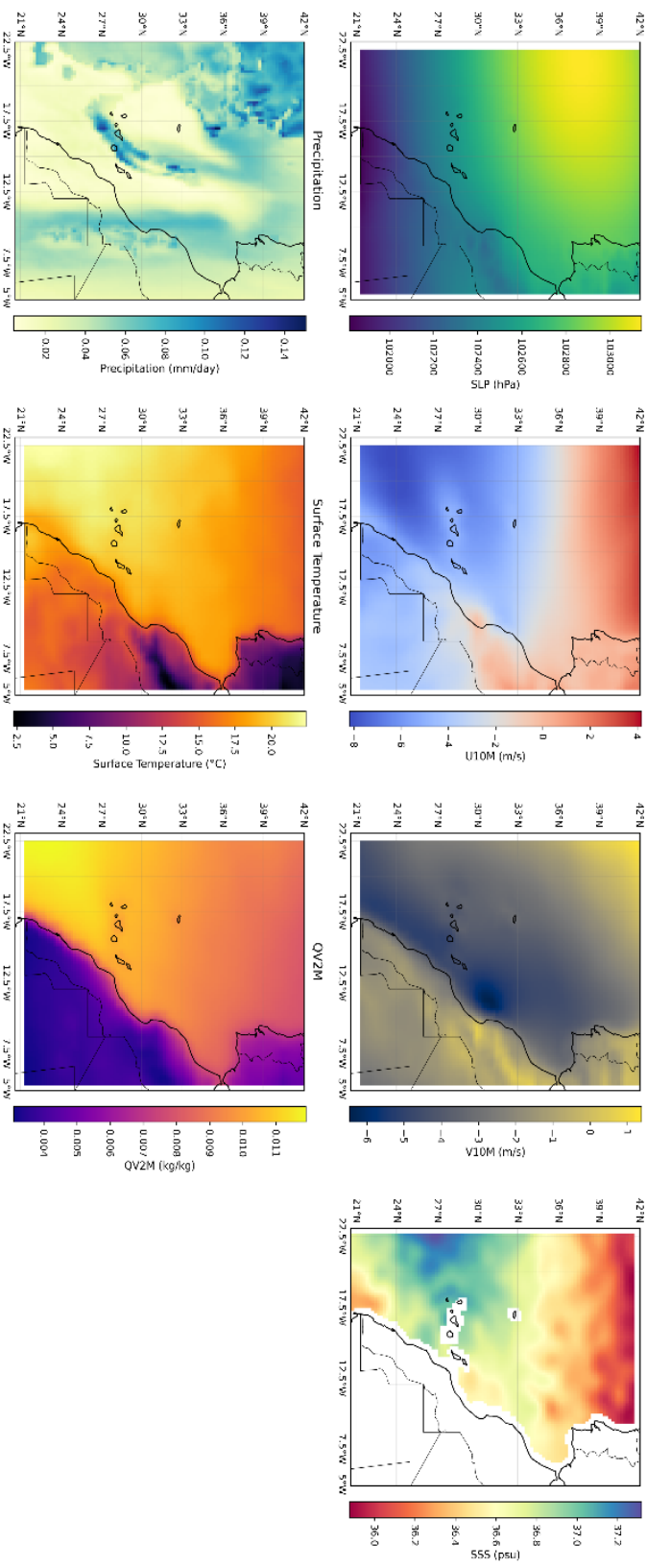

Fig. 2. Initial spatial distribution of the climate variables for the first month in the dataset.

For each grid point, the model processes four consecutive months of input data, including sea level pressure (SLP), specific humidity at two meters (QV2M), zonal and meridional wind components (U10M and V10M), wind speed, sea surface temperature (SST), and sea surface salinity (SSS). These variables describe the hydro-atmospheric state of the study region. The *BiLSTM* layer extracts both forward and backward temporal relationships, while a Multi-Layer Perceptron (MLP) (Rumelhart et al., 1986) refines the latent representations to produce the predicted precipitation for the following month. The overall structure of this predictive architecture is shown in (**Fig. 3**), which defines the baseline rainfall distribution used as reference for the control phase.

To ensure robust and stable predictions, the main hyperparameters of the BiLSTM-MLP network, such as the hidden-layer size, learning rate, and sequence length, were adjusted through a systematic grid-search procedure combined with early stopping on the validation root-mean-square error. This tuning process maintained an optimal balance between model complexity and generalization, preventing both underfitting and overfitting. Preliminary sensitivity tests indicated that excessively large hidden dimensions or high learning rates caused unstable convergence and degraded predictive performance, confirming the importance of careful hyperparameter selection to ensure numerical stability and physically consistent precipitation estimates.

After the predictive stage, the *BiLSTM* model is embedded within a reinforcement learning environment (Sutton & Barto, 2018) where the *TD3* agent acts as an adaptive controller. The agent performs small, continuous perturbations of limited amplitude (approximately 8%) on the variables SLP, QV2M, U10M, and V10M, with the aim of maximizing the predicted precipitation while maintaining physical coherence.

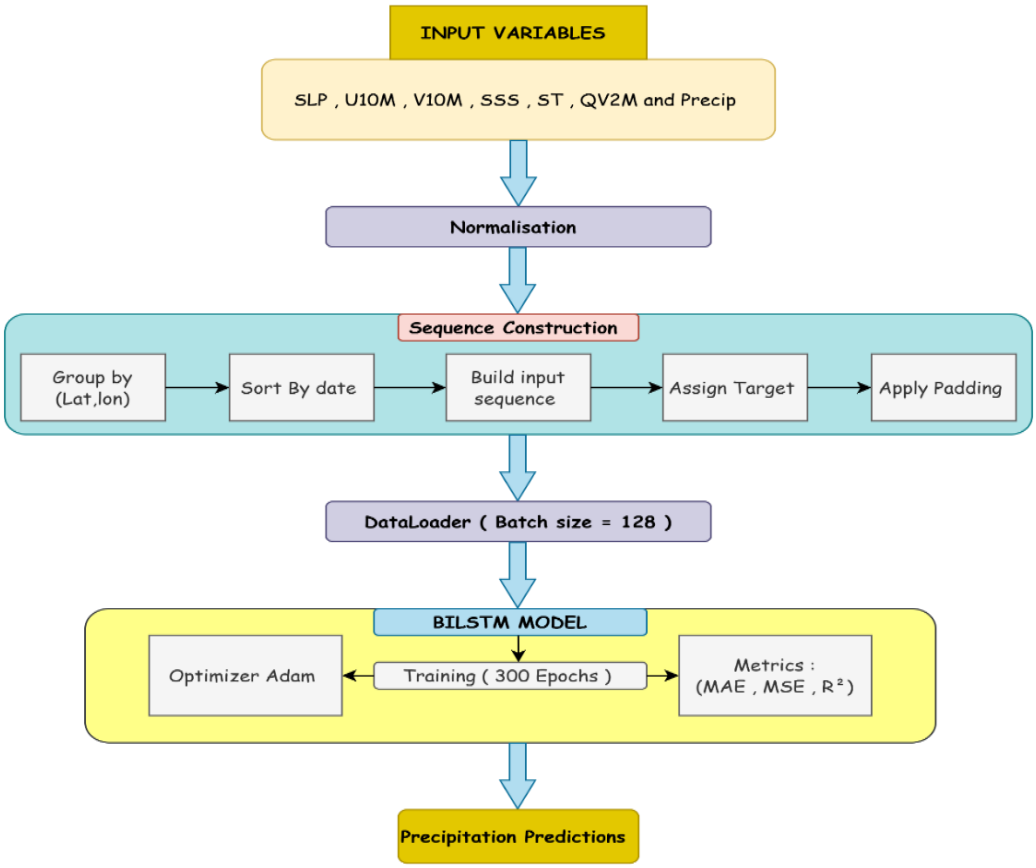

Fig. 3. Architecture of the *BiLSTM-MLP* predictor used for precipitation forecasting.

Each modification triggers a new forecast and a corresponding reward that balances three components: the precipitation gain, the energetic cost associated with the action, and a penalty term that limits abrupt or dynamically inconsistent variations. This formulation allows the agent to identify physically meaningful control strategies that enhance rainfall without violating atmospheric equilibrium. The integration of the predictive and optimization components is illustrated in (**Fig. 4**), which summarizes the TD3-based control framework coupled with *BiLSTM-MLP* forecasting.

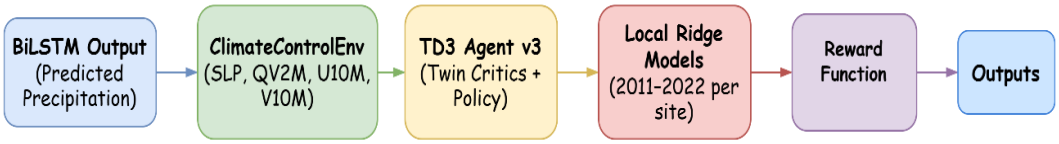

Fig. 4. Architecture of the TD3-based control framework coupled with BiLSTM-MLP forecasting.

3.2. Experimental Setup

All experiments were conducted on a GPU-enabled workstation to ensure efficient model training and evaluation. The entire workflow, including data harmonization, model training, and control optimization, was implemented in PyTorch and Stable-Baselines3 (Raffin et al., 2021). The BiLSTM predictor was trained for 300 epochs with a batch size of 128 using the Adam optimizer with a learning rate of 0.001 and an early-stopping criterion based on the validation root-mean-square error. The TD3 controller was subsequently trained for 75 000 time steps using Gaussian exploration noise and an experience-replay buffer to ensure stable policy learning. Each control episode consisted of eight consecutive steps, representing incremental adjustments of the same atmospheric configuration.

Evaluation was carried out across 3 659 spatial grid points covering the North-Eastern Atlantic domain under identical boundary conditions. The outputs included three complementary metrics that characterize both the physical and energetic aspects of the control process.

The precipitation gain (ΔP**)**, defined as:

$$\Delta P = P_{\mathrm{TD3}} - P_{\mathrm{ref}} \tag{1}$$

where P_{TD3} is the precipitation predicted under control and P_{ref} is the baseline (uncontrolled) precipitation, quantifies the rainfall enhancement (mm) induced by the TD3 agent.

The Reward Efficiency Index (REI), expressed as:

$$\mathrm{REI} = \frac{\Delta P}{C+\varepsilon} \ , where \ C = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|\Delta X_i|}{X_i}\right) \tag{2}$$

where C denotes the energetic cost of atmospheric perturbations applied to the variables $X_i \in \{\mathrm{SLP}, \mathrm{QV2M}, \mathrm{U10M}, \mathrm{V10M}\}$, and ε is a small regularization constant.

This ratio evaluates how effectively precipitation is increased relative to the energy required for control.

The Stability Index (SIS), computed as:

$$\mathrm{SIS} = 1 - \frac{\mathrm{Var}(\Delta X)}{\mathrm{Var}(X_{\mathrm{ref}})+\delta} \tag{3}$$

where $\mathrm{Var}(\Delta X)$ is the variance of the controlled perturbations applied by the TD3 controller, $\mathrm{Var}(X_{\mathrm{ref}})$ represents the variance of the unperturbed (reference) atmospheric fields, and δ is a small stabilizing term.

This index measures the smoothness and physical coherence of atmospheric adjustments. Values approaching 1 indicate stable and dynamically consistent control actions.

Together, these indicators allow both the magnitude and the physical plausibility of the reinforcement-based precipitation control to be quantitatively assessed. Spatial maps of dominant and blocking variables were also generated to highlight which atmospheric components exert the greatest influence on the optimized precipitation fields, providing a physically interpretable assessment of the regional control process.

4. RESULTS AND DISCUSSIONS

4.1. Predictive Model Evaluation

To evaluate the predictive capabilities of the proposed architectures, a set of experiments was performed using three deep learning models: *BiLSTM*, *MLP*, and *Transformer*. Each model was trained to forecast precipitation levels for March 2023 across all spatial grid points, using historical sequences of climate variables as input. The *BiLSTM* architecture employed a bidirectional recurrent structure followed by a dense output layer, while the *MLP* model relied on temporally averaged predictors. The *Transformer* model used self-attention mechanisms to capture long-range dependencies within the time series. All models were trained under identical configurations to ensure a fair and consistent comparison. The corresponding performance metrics, including the Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2 are summarized in **(Table 1)**.

Table 1.

Evaluation metrics of precipitation prediction models for March 2023.

Model	**MSE**	**MAE**	**RMSE**	**R^2**
BiLSTM	0.000007	0.001881	0.002645	0.9927
MLP	0.000076	0.005817	0.008717	0.9173
Transformer	0.000053	0.005912	0.007280	0.9427

As shown in **(Table 1)**, the BiLSTM model consistently outperformed both the MLP and Transformer architectures across all evaluation criteria. It achieved the lowest MSE and MAE values, together with the highest coefficient of determination (R^2 = 0.9927), indicating an excellent fit between predicted and observed precipitation. The small RMSE confirms its robustness and generalization capability. In contrast, the MLP model exhibited the weakest performance, revealing a limited ability to capture temporal dependencies. The Transformer model performed better than the MLP but remained less accurate than the BiLSTM, highlighting the advantage of recurrent structures for sequential climate prediction. This superior performance arises from the bidirectional recurrent memory of the BiLSTM, which allows it to capture delayed and nonlinear dependencies among key atmospheric drivers such as SLP, humidity, and wind components. Unlike the MLP, which treats each month as an independent input, the BiLSTM integrates information across consecutive time steps, reproducing the temporal inertia and feedback loops that control rainfall formation. Compared with the Transformer, whose attention layers require large training samples and may overfit small regional datasets, the BiLSTM maintains temporal continuity and stable convergence with limited data. The relationship between observed and predicted precipitation values for each model is depicted in **Fig. 5**.

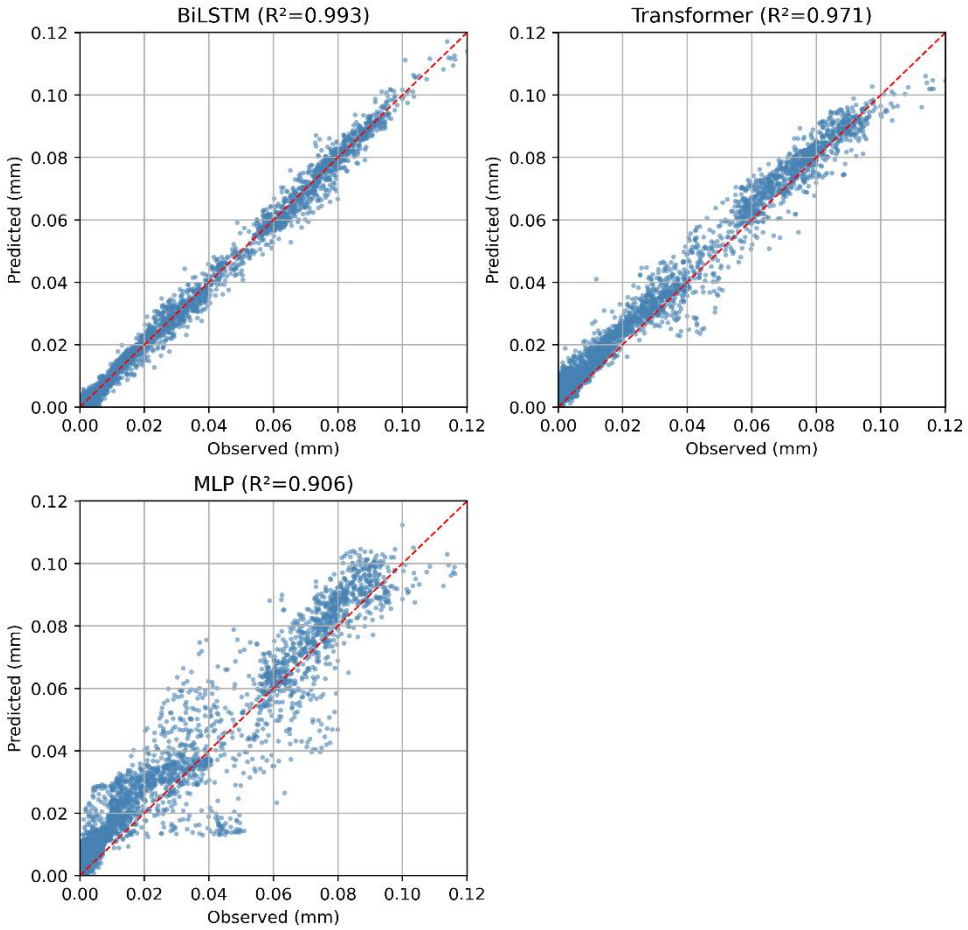

Fig. 5. Scatter plots comparing predicted and observed precipitation values for March 2023 using *MLP*, *BiLSTM*, and *Transformer* models.

The *BiLSTM* results show a nearly perfect alignment along the one-to-one diagonal, demonstrating both high predictive skill and low dispersion, whereas the *MLP* and *Transformer* display broader scatter, reflecting their reduced temporal sensitivity.

4.2. Reinforcement-Based Precipitation Control

After validating the predictive stage, the BiLSTM model was integrated within the TD3-based control framework to investigate adaptive precipitation modulation over 3,659 grid points across the North-Eastern Atlantic domain. The TD3 agent interacted with the predictive environment through multi-step episodes, applying small, continuous perturbations to the variables SLP, QV2M, U10M, and V10M while maximizing a reward function that combines precipitation gain, energetic cost, and physical stability. The optimization produced three key indicators: the precipitation variation (ΔP, in mm), the Reward Efficiency Index (REI), and the Stability Index (SIS), (**Table 2**). On average, the regional precipitation gain reached ΔP = 0.006 mm per grid point, with 63 % of locations showing positive increments. The mean REI was 0.0025 and the average SIS 0.0004, confirming that most optimized states remained dynamically stable. These results demonstrate that the controller successfully enhances rainfall potential within physically admissible limits, ensuring no artificial amplification or instability.

Table 2.

Summary of TD3 control performance indicators.

Indicator	Definition / Meaning	Average Value	Interpretation
ΔP (mm)	Precipitation gain per grid cell	0.0061	Positive rainfall enhancement
REI	Reward Efficiency Index	0.0025	Efficient gain-cost trade-off
SIS	Stability Index	0.0004	Physically stable configuration
% ΔP > 0	Percentage of points with positive rainfall change	63.4 %	Most sites improved after optimization

Spatial patterns derived from these indicators revealed coherent regional structures. The highest ΔP and REI values occurred between 20° N and 35° N, particularly off the Moroccan coast, where humidity flux and pressure gradients are most sensitive to small perturbations. In contrast, northern latitudes exhibited weaker responses, consistent with their reduced coupling between low-level winds and moisture convergence. The SIS distribution confirmed that areas of strong ΔP were not associated with instability peaks, reinforcing the physical realism of the optimization.

Beyond these aggregate indicators, the contribution analysis identified the dominant and blocking variables governing local rainfall responses. As illustrated in (**Fig. 6**), each color represents the variable exerting the strongest control on precipitation within a given grid cell. Sea-level pressure (SLP) emerged as the dominant influence across most of the domain, particularly along the subtropical corridor, while specific humidity (QV2M) dominated near the coastal regions. Wind components (U10M and V10M) showed localized dominance over transition zones where synoptic advection drives rainfall variability. These spatial patterns correspond closely to known atmospheric mechanisms operating in the Atlantic sector, confirming that the learned control strategies remain physically interpretable.

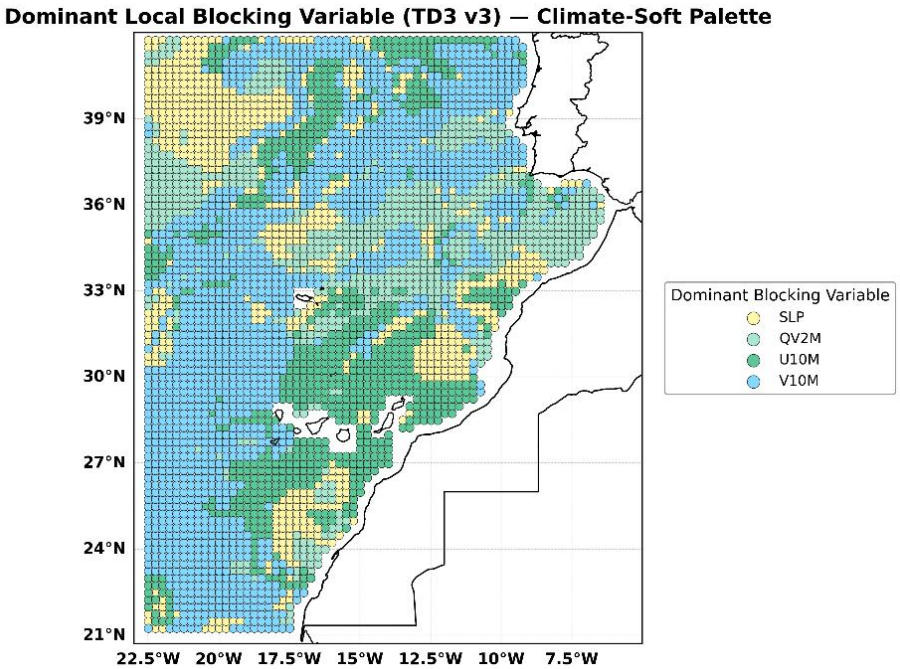

Fig. 6. Spatial distribution of dominant climatic variables influencing optimized precipitation across the study region.

The reinforcement-based results confirm that the hybrid *BiLSTM-TD3* framework achieves measurable precipitation enhancement while preserving atmospheric balance. The spatial coherence of ΔP, REI, and SIS, together with the dominance of physically meaningful variables, validates both the scientific realism and the operational potential of the proposed framework as a regional tool for hydro-climatic regulation.

4.3. Spatial manifestation of TD3-controlled precipitation strategies

Building on the reinforcement-based optimization results, the learned control strategies can be examined from a spatial and physical perspective to assess how the precipitation gains materialize across different climatic sub-regions of the North-Eastern Atlantic. Rather than focusing on aggregate indicators alone, this analysis investigates how the optimized atmospheric adjustments translate into geographically structured precipitation responses under real winter conditions.

A first-order spatial inspection reveals that precipitation enhancements are not uniformly distributed over the domain. Instead, precipitation variations exhibit coherent regional structures that closely follow the background hydro-climatic organization of the North-Eastern Atlantic. This spatial heterogeneity indicates that the reinforcement-based control does not impose homogeneous forcing, but rather interacts with the intrinsic sensitivity of the atmospheric system.

Along the Moroccan Atlantic coast, particularly during December 2018 and January 2020, precipitation gains appear more concentrated and spatially organized. These coastal areas are characterized by enhanced sensitivity to low-level moisture convergence and pressure gradients, allowing small atmospheric adjustments to produce measurable precipitation responses. Importantly, the enhanced precipitation patterns largely preserve the original spatial morphology, indicating that the control acts by reinforcing existing dynamical structures rather than generating artificial rainfall anomalies.

Moving offshore toward the central Atlantic, the precipitation response becomes weaker and more fragmented. In these regions, precipitation changes alternate between marginally positive and near-neutral values, reflecting a reduced coupling between surface-level atmospheric perturbations and precipitation processes. This behavior suggests that the reinforcement agent adapts its actions according to the local responsiveness of the atmospheric state, applying limited control where precipitation sensitivity is intrinsically low.

In the vicinity of the Azores region, precipitation changes remain consistently modest across all selected months. This result is physically consistent with the stabilizing influence of the Azores High, which constrains vertical motion and limits moisture uplift during winter conditions. The absence of strong precipitation enhancement in this area indicates that the control framework respects large-scale atmospheric constraints and does not attempt to override dominant synoptic regimes.

Further north, closer to the Iberian sector, precipitation responses show intermediate behavior, with localized gains appearing intermittently depending on the background circulation of each month. These spatial variations highlight the dependence of the control efficiency on transient synoptic configurations, reinforcing the idea that the learned strategies are state-dependent rather than spatially fixed.

This spatial manifestation of the learned control strategies is illustrated through a composite visualization of selected winter months (**Fig. 7**), comparing the observed precipitation fields before control, the TD3-enhanced precipitation after control, and the resulting absolute change (ΔP). While the Reward Efficiency Index (REI) and the Stability Index (SIS) provide a global assessment of control efficiency and physical coherence, the spatial analysis is primarily conducted using ΔP, as it directly reflects the geographical response of the atmospheric system to the applied control. The displayed patterns correspond to precipitation fields generated directly by the hybrid BiLSTM-TD3 framework, where the reinforcement agent operates through small, continuous perturbations of the atmospheric state while respecting physical stability constraints. The maps therefore represent the spatial realization of the learned control strategies rather than a post-processed or externally imposed adjustment.

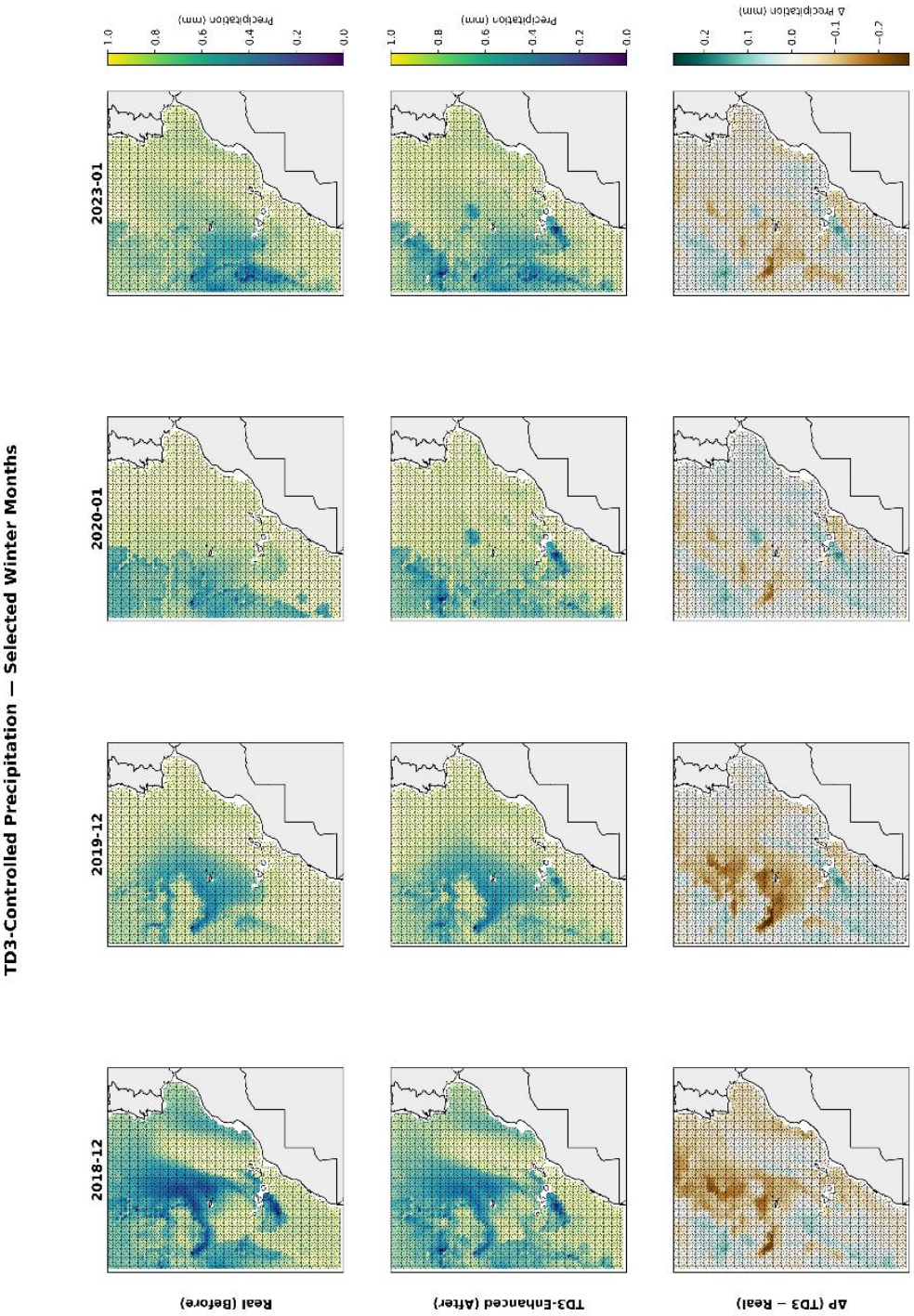

Fig. 7. Observed and TD3-controlled precipitation fields and associated changes (ΔP) for selected winter months from the dataset.

Overall, the spatial analysis confirms that the TD3-controlled precipitation strategies operate in a selective and physically consistent manner. Precipitation enhancement emerges primarily in regions where the atmospheric system exhibits sufficient dynamical sensitivity, while remaining limited in areas governed by strong stabilizing mechanisms. These results consolidate the physical interpretability of the proposed framework and confirm that the learned control strategies reinforce existing atmospheric dynamics rather than imposing artificial precipitation responses. From a broader perspective, this behavior highlights both the potential and the scope of applicability of the proposed framework. While the methodological structure of the BiLSTM-TD3 approach is generic and relies on physically meaningful variables common to many ocean-atmosphere systems, the spatially heterogeneous responses observed here indicate that effective application to other regions would require region-specific retraining and calibration to account for local climatic sensitivities and coupling mechanisms.

At the same time, it is important to emphasize that the present framework operates within a virtual predictive environment driven by reanalysis data and is not intended as a direct atmospheric actuation system. Rather, it should be viewed as a physically interpretable decision-support and scenario-exploration tool, designed to investigate how small, physically admissible perturbations could influence precipitation patterns under controlled conditions. These considerations define the current limitations of the approach while also outlining a clear pathway for future extensions toward broader hydro-climatic assessment and resilience-oriented applications.

5. CONCLUSION

The hybrid integration of deep learning and reinforcement learning provides a robust framework for physically consistent precipitation optimization along the Moroccan and Iberian coasts within the North-Eastern Atlantic. The BiLSTM predictor achieved highly accurate rainfall estimations, while the TD3 controller enhanced precipitation at more than 60% of grid points without compromising atmospheric stability. Spatial patterns revealed that sea-level pressure, humidity, and wind components exert the strongest control on rainfall modulation, confirming the physical realism of the learned strategies.

Beyond predictive performance, the proposed framework demonstrates that reinforcement learning can serve as a physically interpretable diagnostic tool to explore precipitation sensitivity under constrained atmospheric perturbations, rather than as an operational weather modification system. The spatial coherence of the optimized responses and stability indicators confirms that the control acts by modulating existing atmospheric structures while respecting large-scale dynamical constraints.

Although this study focuses on winter conditions over the North-Eastern Atlantic, the methodological structure is generic and can be extended to other coastal or drought-prone regions through region-specific retraining and calibration. The framework operates within a virtual environment driven by reanalysis data and should therefore be regarded as a decision-support and scenario-exploration tool. Future work will address multi-seasonal validation, higher-resolution datasets, and alternative reinforcement learning algorithms to further improve robustness and applicability.

ACKNOWLEDGEMENTS

The authors acknowledge the use of NASA Earthdata, ECCO, and MERRA-2 datasets that supported this research.

REFERENCES

Al Sayed K., Boodi A., Sadeghian Broujney R., et al (2024). Reinforcement learning for HVAC control in intelligent buildings: A technical and conceptual review. *Renewable and Sustainable Energy Reviews*, 187, 114268. https://doi.org/10.1016/j.rser.2024.114268

Durante Z., Huang Q., Wake N., et al (2024). Agent AI: Surveying the horizons of multimodal interaction. *arXiv preprint* arXiv:2401.03568. https://doi.org/10.48550/arXiv.2401.03568

Fukumori I., Wang O., Fenty I., et al (2021). ECCO Version 4 Release 4 (V4r4): A global ocean state estimate with improved representation of the Arctic and Southern Oceans. *Journal of Geophysical Research: Oceans*, 126(8), e2021JC017448. https://doi.org/10.1029/2021JC017448

Fujimoto S., van Hoof H., & Meger D. (2018). Addressing function approximation error in actor-critic methods. *arXiv preprint* arXiv:1802.09477. https://doi.org/10.48550/arXiv.1802.09477

Gelaro R., McCarty W., Suárez M. J., et al (2017). The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Graves A. & Schmidhuber J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, 2047–2052. https://doi.org/10.1109/IJCNN.2005.1556215

Huffman G. J., Bolvin D. T., Joyce R., et al (2023). Integrated multi-satellite retrievals for GPM (IMERG), version 07B, final run. *NASA Goddard Earth Sciences Data and Information Services Center (GES DISC).* https://doi.org/10.5067/GPM/IMERG/3B-MONTH/07

Kurth T., Treichler S., Romero J., et al (2018). Exascale deep learning for climate analytics. In *SC18 International Conference for High Performance Computing, Networking, Storage and Analysis.* IEEE, 649–660. https://doi.org/10.1109/SC.2018.00054

Ladi T., Jabalameli S., & Sharifi A. (2022). Applications of machine learning and deep learning methods for climate change mitigation and adaptation. *Environment and Planning B*, 49(4), 1285–1300. https://doi.org/10.1177/23998083221085281

Li G., Yu Z., Li Y., et al (2024). Interaction mechanism of global multiple ocean–atmosphere coupled modes and their impacts on South and East Asian monsoon: A review. *Global and Planetary Change*, 233, 104438. https://doi.org/10.1016/j.gloplacha.2024.104438

Luo X., Nadiga B. T., Park J. H., et al (2022). A Bayesian deep learning approach to near-term climate prediction. *Journal of Advances in Modeling Earth Systems*, 14(10), e2022MS003058. https://doi.org/10.1029/2022MS003058

Meque A., Gamede S., Motlhobogi T., et al (2021). Numerical weather prediction and climate modelling: Challenges and opportunities for improving climate services delivery in southern Africa. *Climate Services*, 23, 100243. https://doi.org/10.1016/j.cliser.2021.100243

NOAA National Centers for Environmental Information (2025). *Global Climate Report – March 2025.* https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202503#precip

Ojo O., et al (2022). Climate–environmental analysis using deep Bayesian networks for SST prediction. *(Preprint / Journal of Advances in Modeling Earth Systems).* https://doi.org/10.1029/2022MS003058

Pathak J., et al (2022). FourCastNet: Global high-resolution weather forecasting using adaptive Fourier neural operators. *arXiv preprint* arXiv:2202.11214.

Raffin A., Hill A., Gleave A., Kanervisto A., & Dormann N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *arXiv preprint* arXiv:2103.09575. https://github.com/DLR-RM/stable-baselines3

Rasp S., Pritchard M., & Gentine P. (2020). WeatherBench: A benchmark dataset for data-driven weather forecasting. *Geoscientific Model Development*, 13(2), 1073–1090. https://doi.org/10.5194/gmd-13-1073-2020

Routray A., et al (2023). Predictive control for smart HVAC and energy systems: A review. *Renewable and Sustainable Energy Reviews*, 187, 114268. https://doi.org/10.1016/j.rser.2024.114268

Rumelhart D. E., Hinton G. E., & Williams R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, 318–362. MIT Press.

Sierla S., Ihasalo H., & Vyatkin V. (2022). A review of reinforcement learning applications to control of heating, ventilation and air conditioning systems. *Energies*, 15(10), 3526. https://doi.org/10.3390/en15103526

Sleeman J., et al (2023). ACTD: Neurosymbolic architectures for early detection of tipping dynamics. *arXiv preprint* arXiv:2302.06852 [cs.AI]. https://arxiv.org/abs/2302.06852

Steininger M., Abel D., Ziegler K., & Krause A. (2022). ConvMOS: Climate Model Output Statistics with Deep Learning. *Data Mining and Knowledge Discovery*, 37(4), 1–31. https://doi.org/10.1007/s10618-022-00877-6

Sutton R. S. & Barto A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press. http://incompleteideas.net/book/the-book.html

Virtanen P., Gommers R., Oliphant T. E., et al (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2