Journal of Data Analysis and Information Processing, 2026, 14(4), 392-427
https://www.scirp.org/journal/jdaip
ISSN Online: 2327-7203
ISSN Print: 2327-7211

Optimizing Maize Fertilization on Savanna Ferralsols through Deep Reinforcement Learning: A pH Zone Specialized, Calibrated, and Interpretable Policy

Mohamadou Yaya[1*], Blaise Omer Yenke[1], Moskolai Justin Ngossaha[2], Laurent Bitjoka[1], Samuel Kotva Goudoungou[1]

[1]University of Ngaoundere, Ngaoundere, Cameroon
[2]University of Douala, Douala, Cameroon
Email: *y.mohamadou@isdm-univ-dla.com, boyenke@univ-ndere.cm, jmoskolai@fs-univ-douala.cm, lbitjoka@univ-ndere.cm, Samuel.kotva-goudoungou@inria.fr

How to cite this paper: Yaya, M., Yenke, B.O., Ngossaha, M.J., Bitjoka, L. and Kotva Goudoungou, S. (2026) Optimizing Maize Fertilization on Savanna Ferralsols through Deep Reinforcement Learning: A pH Zone Specialized, Calibrated, and Interpretable Policy. *Journal of Data Analysis and Information Processing*, **14**, 392-427.
https://doi.org/10.4236/jdaip.2026.144020

Received: September 9, 2026
Accepted: November 21, 2026
Published: November 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).
http://creativecommons.org/licenses/by-nc/4.0/

Open Access

Abstract

Food security in sub-Saharan Africa depends heavily on staple crops grown by smallholder farmers. Yet maize yields remain structurally constrained by the low fertility of heavily weathered tropical soils, known as acidic Ferralsols, where nutrient availability is closely tied to pH and organic matter. In Cameroon's Adamaoua Region, fertilization still relies on generic recommendations that ignore field-level pedological heterogeneity. Existing decision-support tools rarely combine sequential decision logic with the well-established pH-dependent agronomic rule, motivating the use of reinforcement learning (RL), an approach still poorly validated on real field data in African agricultural contexts. This article proposes a maize fertilization recommendation system built around PPO (Proximal Policy Optimization), trained on a calibrated agronomic simulator based on the Mitscherlich-Baule-Liebig production model. The zone boundary is fixed a priori at pH 5.5. Below this threshold, the agent learns only the compost dose. Above it, it learns only the NPK and urea doses. This design focuses learning on fine dosing, not on rediscovering an already known threshold. The system takes soil parameters as input, generates a recommendation via the trained agent, then applies a leaching-prevention filter using weather data to set the application schedule, offering a choice between a more economical chemical formula and a costlier but more profitable restoration path. The learned PPO policy statistically outperforms four reference policies. A SHAP-based interpretability analysis, repeated across five seeds, shows agronomically coherent associations between inputs and outputs, although

cross-seed stability varies by output. These results position RL as a viable approach for site-specific fertilization recommendation in tropical agriculture.

Keywords

Reinforcement Learning, PPO, Precision Agriculture, Maize Fertilization, Ferralsol, pH-Dependent Decision Rule, Interpretability

1. Introduction

Maize (*Zea mays* L.) is the most widely cultivated cereal in the world. It is a major staple crop in Cameroon, where its production nonetheless remains limited by biological and environmental factors, particularly inadequate mineral fertilization [1]. In the Adamaoua region, the dominant soils are heavily weathered Ferralsols, characterized by pronounced acidity, a strong phosphorus fixation capacity by iron and aluminum oxides [2]-[4], and organic matter content that varies widely from one plot to another. These three factors interact in a non-trivial way: the same mineral fertilizer input can be highly profitable on a corrected soil rich in organic matter, and almost without effect, or even counterproductive once costs are deducted, on an unamended acidic soil. The recommendations currently issued by the Ministry of Agriculture and Rural Development (MINADER) prescribe fixed doses of NPK and urea, independent of the plot's pH and organic matter content. This "one-size-fits-all" approach ignores the pedological heterogeneity documented in the regional agronomic literature [5] [6] and can lead to input investments that are only marginally profitable or not profitable at all for the farmer, particularly on the most acidic soils, where mineral fertilizer applied without prior pH correction loses much of its effectiveness.

Maize is a staple crop in Cameroon. In Adamaoua, we work on Ferralsols. These soils are acidic and fix phosphorus. As a result, yields stay low even when farmers apply NPK. Given this pedological heterogeneity and the need to adapt the decision over the course of the crop cycle, this work turns toward reinforcement learning (RL). This paradigm, in which an agent learns a sequential decision policy through interaction with a simulated environment, is particularly well suited to site-specific fertilization, where each decision (compost, NPK, or urea dose) influences soil state, yield, and subsequent decisions. The state of the art (Section 2) situates this choice more precisely relative to existing work, both in precision agriculture and in the African context.

To address these limitations, this work proposes five contributions. First, a calibrated Mitscherlich-Baule-Liebig-type agronomic simulator that explicitly integrates the interaction between soil parameters (pH, N, P, K, OM, EC) and inputs (compost, NPK, urea). This simulator accounts for the rapid saturation of phosphorus, a phenomenon independently documented in the tropical pedological literature, as well as the effect of organic matter on nitrogen uptake. Second, a hybrid

rule/learning architecture in which a known agronomic rule, the pH 5.5 threshold, restricts the agent's action space so as to concentrate learning on fine dosing. Third, a quantification of the impact of two often-undocumented implementation choices, action scaling and observation normalization, on policy quality. This methodological result remains transferable to other applications of continuous RL in agriculture. Fourth, an interpretability validation using SHAP and LIME, repeated across five independent seeds to test decision stability. Finally, an operational deployment pipeline integrating weather forecasts and a dual economic/restoration recommendation, tested on a field workbook demonstrating the viability of moving from a research prototype to a tool usable by agricultural extension officers.

The remainder of this article is organized as follows: Section 2 presents the state of the art on reinforcement learning applied to agriculture and related work in the African context; Section 3 describes the methodological approach (agronomic simulator, formalization of the decision problem, reward function, learning algorithm and critical implementation corrections, reference policies, evaluation protocol, interpretability) as well as the proposed deployment architecture; Section 4 presents the study area, the case studies used for calibration, the results and their discussion; Section 5 concludes and opens perspectives.

2. Related Work

Reinforcement learning (RL) is a machine learning paradigm in which an agent learns a decision policy $\pi(a|s)$, a function that maps each observed states to an action a, through successive trials in a given environment. This agent seeks to maximize a cumulative reward while avoiding imitation of labeled examples [7]. Formally, the problem is framed as a Markov Decision Process (MDP) defined by a state space S, an action space A, a transition function (here, the agronomic simulator), and a reward function $r(s, a)$. Unlike supervised learning, RL does not require prior knowledge of the "correct" action for each state: the agent discovers it through exploration, guided solely by the reward signal, which here is the net economic margin obtained. This property is precisely what makes RL relevant to fertilization. Indeed, there is no database of labeled optimal doses for every combination of pH, nutrient content, and organic matter encountered in the field. Hence the need for a causal model (the agronomic simulator) linking a decision to an outcome, which the agent can explore.

This paradigm is particularly well suited to the problem of site-specific fertilization, where the "correct" dose depends on a non-linear interaction between several state variables (pH, nitrogen, phosphorus, potassium, organic matter) that no static reference table can cover exhaustively. The algorithm used in this work, Proximal Policy Optimization [8], is an on-policy RL algorithm with a continuous stochastic policy, widely adopted for its training stability on continuous action spaces. This is exactly the case for fertilizer doses to be applied, which are real valued quantities (kg/ha, t/ha) rather than discrete choices as illustrated in **Figure**

1. PPO limits the magnitude of policy updates at each iteration (via a clipped surrogate objective). This makes it less sensitive to hyperparameter choice than earlier policy gradient methods, a useful property in a context such as this one, where the computational budget available for hyperparameter search is limited.

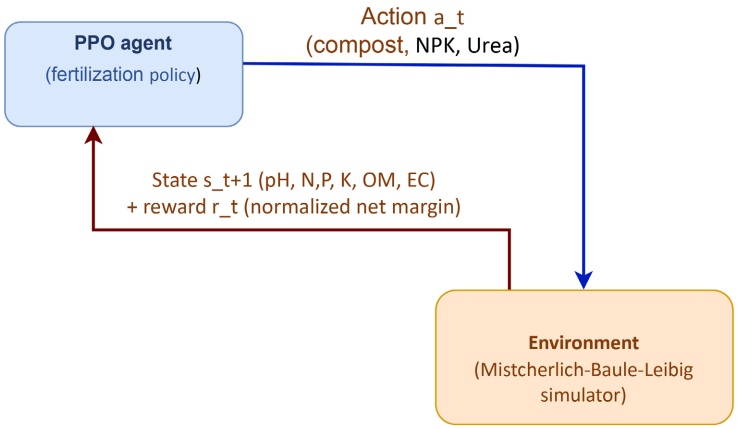


Figure 1. Reinforcement learning loop applied to the fertilization problem. The agent (PPO policy network) observes the soil state, chooses an action (input dose), and receives back a new state and a reward computed by the agronomic simulator.

The study closest to ours, both geographically and thematically, is that of Keegoui *et al.* [1], who evaluated AI generated maize fertilization recommendations on a ferralitic soil in Central Cameroon, comparing three treatments (control, IRAD recommendation, AI recommendation) across three improved varieties. However, that study neither develops nor documents its own model: the fertilization recommendation comes from a third party commercial mobile application (Croptune). This application is described as combining image analysis, big data, and cloud computing to detect nutritional deficiencies in real time from visual symptoms on the plant. No details of architecture, method, or decision logic were reported. The dose actually applied is moreover not the tool's raw output: the authors select, among the compound fertilizer formulations commercially available, the one that comes closest. This approach performs a one-shot prediction from a fixed set of input variables. It therefore contrasts with the approach of the present work, where the simulator, the zone rule, the learned policy, and every coefficient of the reward function are made explicit and reproducible. It does not rely on a policy that adjusts dynamically to soil state using a sequential decision agent capable of adjusting its recommendation according to soil characteristics and their evolution over the season. Nor does it explicitly specialize the recommendation by pH zone.

More broadly, continuous RL has been successfully applied to several agricultural decisions involving action spaces of a similar nature, with real valued quantities to be dosed rather than discrete choices. In irrigation, Saikai, Peake, and Chenu (2023) [9] trained a deep RL agent on high dimensional sensor data to schedule irrigation. This illustrates RL's ability to exploit a rich state (soil mois-

ture, weather) for a continuous decision. In nitrogen management, Wu, Tao, Zhao *et al.* (2022) [10] used a crop simulator to train an agent to dose nitrogen over the course of the crop cycle. This sequential decision structure is close to the one adopted here (dose adjusted to an evolving soil state). Tao, Zhao, Wu *et al.* (2022) extend this logic to integrated crop management (combined irrigation and fertilization). Yassine *et al.* (2022) [11] as well as Binas, Luginbuehl, and Bengio (2019) [12] apply RL to smart farm operation in a broader sense (multi-decision control of a simulated farm). This body of work confirms the maturity of continuous RL for sequential agricultural decisions. However, none focuses specifically on fertilization in the context of acidic tropical soils, where the pH by nutrient interaction imposes a zone based decision structure distinct from those studied, while also accounting for the soil restoration dimension.

Other studies combine machine learning and input recommendation specifically in Africa without resorting to RL: Musanase *et al.* (2023) [13], Lontsi Saadio Cedric *et al.* (2022) [14], and Bouni *et al.* (2022) [15] propose recommendation systems based on supervised learning (classification or regression) from soil analyses, following a one-shot diagnostic logic rather than a sequential decision policy. These approaches share with Keegoui *et al.* (2025) the limitation of not explicitly modeling how an early (planting stage) fertilization decision alters soil state and, consequently, the optimal decision at a later stage of the cycle, a dimension that only a sequential, MDP type formulation can natively capture.

At the pedological level, the regional and tropical literature has long documented the mechanisms that motivate the pH zone specialization adopted in this work. Heavily weathered Ferralsols, typical of Adamaoua, exhibit a strong phosphorus fixation capacity by iron and aluminum oxides, reducing its availability to the plant regardless of the dose applied (Hinsinger *et al.*, 2017; Owino *et al.*, 2015, on Ferralsols in Siaya, Kenya; Truong, 1996). Tabi *et al.* (2013) and Carsky *et al.* (1998) document, at the West and Central African regional scale, the field level pedological heterogeneity (pH, organic matter) that generic MINADER type recommendations ignore, as well as the structuring role of organic matter in nitrogen availability in tropical soil. These works converge on a common finding: the agronomic effectiveness of a mineral input strongly depends on the prior state of the soil, particularly its pH, which a priori justifies a zone differentiated decision policy rather than a single dose, and motivates the choice, in this work, to impose this differentiation as a structural rule rather than leaving it to the agent's discovery alone (Section 3.2).

The choice of a multiplicative production model to simulate yield also draws on an established line of agronomic models. The Mitscherlich-Baule-Liebig principle, on which the simulator used here is based (Section 3.1), combines several limiting factors through a product rather than a sum, reflecting the law of the minimum, according to which a severe deficit in a single factor cannot be compensated by an excess of another. QUEFTS (Quantitative Evaluation of the Fertility of Tropical Soils; Janssen *et al.*, 1990) [16], used in this work as an independent

cross-validation simulator (Section 4.6), applies a related logic, a law of the minimum on nitrogen, phosphorus, and potassium, specifically calibrated for tropical soils. The convergence of these two independently developed model families on the same principle of non-additive limiting factors reinforces the validity of this modeling choice for the Ferralsols studied here.

Several recent works (2022-2024) broaden the landscape of RL applied to crop management, independently of the specific context addressed here. Kallenberg *et al.* (2023) [17] present CropGym, a Gymnasium environment in which a PPO agent learns a weekly nitrogen fertilization policy for winter wheat, seeking to optimize yield against environmental impact. Turchetta *et al.* (2022) [18] propose CyclesGym, an RL framework for long-term crop management strategies, built on the first-principles simulator Cycles. Kelly, Foster and Schultz (2024), published in this same journal, evaluate the added value of deep RL for irrigation scheduling, with a baseline-comparison methodology close to the one adopted here.

These three works share a common limitation, one that is decisive for the context targeted by this article. None of them addresses acidic tropical soils, nor the phosphorus-fixation constraint that structures agronomic decision-making on Ferralsol. Kallenberg *et al.* (2023) [17] formulate the decision discretely (one dose from a finite weekly set) under a temperate climate, with no mechanism for pedological-zone specialization; their reward function trades off yield against environmental impact, not the net economic margin relevant to a smallholder farmer. Turchetta *et al.* (2022), broader in scope, report no interpretability-based validation, nor any cross-seed stability test. Kelly *et al.* (2024), while methodologically close (statistical comparison against baselines), address irrigation, a continuous but single-dimensional decision, without the multi-input interaction (compost, NPK, urea) or the a priori zone rule that structure the problem addressed here. None of the three reports a deployment pipeline tested in an African agricultural context, nor a dual economic/restorative recommendation. This convergent absence motivates, by contrast, the methodological contribution of this article: a hybrid rule/learning architecture specifically designed for the pH × input interaction on acidic tropical soil, validated through interpretability repeated across five seeds and deployed down to the level of a field workbook. This comparison should not, however, obscure the limitations specific to this work: the absence of real field-level agronomic validation, an evaluation that remains single-season, and a calibration resting on a limited number of field observations remain constraints that neither Kallenberg *et al.* (2023), nor Turchetta *et al.* (2022), nor Kelly *et al.* (2024) [19] have to confront in their respective contexts, given their access to more broadly established reference data or simulators. These limitations, together with those uncovered during the four-simulator cross-validation (Sections 4.7 - 4.12), are detailed in Section 5.1.

To our knowledge, no prior work combines continuous RL, explicit specialization through a priori agronomic rule (pH zone), validation through post-hoc interpretability repeated across multiple training seeds, and verification of policy

robustness against several independent mathematical formulations of yield (Section 4.8). The combination of these elements constitutes the main contribution of this article.

3. Methodological Approach and Deployment Architecture

Statement of the problem: maize fertilization on acidic Ferralsols must contend with a dual heterogeneity: pH, which governs the very availability of applied nutrients, and organic matter, which modulates their effectiveness. It turns out that the generic recommendations currently in circulation ignore these parameters. The task is therefore to produce, from a simple soil diagnostic, a sequential fertilization decision (at planting, then at the V6 stage) adapted to each plot, without assuming any pre-labeled database of optimal doses.

The design of this optimized fertilization model followed a structured and methodical approach. We began by determining the key factors influencing fertilization, namely information on soil condition, applied and extension-disseminated agricultural practices, and weather data. Next, we defined the inputs used for fertilization along with their associated technical pathways. Third, we clearly identified the model's objectives with respect to yield and net margin. The fourth step concerns the acquisition of data for calibration, followed by the design and implementation of the reinforcement learning model, with precise definition of states, actions, and reward functions. Finally, the validation of this model in order to ensure its robustness and effectiveness under various agronomic conditions.

In sum, this calibrated agronomic simulator (Section 3.1) serves as a training environment for a reinforcement learning agent (PPO), whose decision space is structured by a pH-zone rule set a priori (Section 3.2) rather than left to learning. The learning signal is thus concentrated on the fine-grained dosing of the inputs available within each regime. The policy thus trained is statistically validated against four reference policies (Sections 3.5 - 3.6), analyzed for interpretability (Section 3.7), then integrated into an operational deployment pipeline delivering a dual economic and restorative recommendation (Section 3.8).

3.1. Agronomic Simulator

Yield is modeled by a Mitscherlich-Baule-Liebig type multiplicative production function, which combines several limiting factors through a product rather than a sum. This reflects the agronomic principle whereby a severe deficit in a single factor (for example phosphorus) cannot be compensated by an excess of another factor (for example nitrogen):

$$Y = Y_{\max} \cdot \left(f_{\mathrm{N}} \cdot f_{\mathrm{P}} \cdot f_{\mathrm{K}} \cdot f_{\mathrm{MO}} \right)^{\beta} \cdot f_{\mathrm{pH}} \cdot f_{\mathrm{EC}}$$

where each factor $\in [0, 1]$ follows a Mitscherlich exponential saturation function, with a saturation coefficient cx specific to each nutrient. The calibration of these coefficients combines two complementary sources: the regional and tropical agronomic literature, and a direct numerical calibration on thirteen real field obser-

vations (four plots monitored by IRAD-Wakwa, including one over three consecutive seasons, and nine plots documented by the agricultural extension officer of Dang; Table 2), through minimization of the weighted squared error between simulated and observed yield (scipy.optimize.differential_evolution, population of 80 individuals, maximum 300 generations, final local refinement), reaching a root mean square error of 0.77 t/ha across the thirteen targets. This precision is deemed satisfactory given the natural variability of yield under smallholder conditions and the broadening of the calibration set relative to earlier versions of the model. $f_X f_X 1-\mathrm{e}^{-cx\cdot dispox}$

$$f(c,d)=1-\mathrm{e}^{-c\cdot d}$$

$$f_{\mathrm{pH}}(\mathrm{pH})=\begin{cases} F_{\mathrm{pH,min}} & \text{si } \mathrm{pH}\le \mathrm{pH}_{\mathrm{toxique}} \\ F_{\mathrm{pH,min}}+\left(f_{4.5}-F_{\mathrm{pH,min}}\right)\cdot\dfrac{\mathrm{pH}-4.0}{0.5} & \text{si } \mathrm{pH}_{\mathrm{toxique}}<\mathrm{pH}<4.5 \\ f_{4.5}+\left(1-f_{4.5}\right)\cdot\dfrac{\mathrm{pH}-4.5}{1.0} & \text{si } 4.5\le \mathrm{pH}<\mathrm{pH}_{\mathrm{optimal}} \\ 1.0 & \text{si } \mathrm{pH}\ge \mathrm{pH}_{\mathrm{optimal}} \end{cases}$$

With the adopted calibration:

- $\mathrm{pH}_{\mathrm{toxique}}=4.0$
- $F_{\mathrm{pH,min}}=0.2$
- $f_{4,5}=0.4816$
- $\mathrm{pH}_{\mathrm{optimal}}=5.5$

$$f_{\mathrm{EC}}(\mathrm{EC})=0.30+0.70\cdot\frac{\mathrm{EC}}{\mathrm{EC}_{\mathrm{seuil}}}\ \text{si } \mathrm{EC}<\mathrm{EC}_{\mathrm{seuil}},\ \text{sinon } 1.0$$

A subsequent recalibration, made necessary by the evolution of the yield formula during development (addition of the electrical conductivity factor EC, isolation of urea's own effect), first raised c_{N} to 0.10 through pure numerical adjustment (differential_evolution minimizing RMSE alone), bringing the root mean square error on the thirteen observations from 1.76 t/ha (original coefficients applied to the updated formula) down to 0.57 t/ha. This value, however, proved to saturate nitrogen at even a modest mineral fertilization level, structurally canceling urea's marginal effect regardless of the soil tested. This is a flaw invisible to RMSE alone, detected only by checking the behavior of the resulting policy (Section 3.4, flaw 11). The coefficients ultimately retained are $c_{\mathrm{N}} = 0.008$, $c_{\mathrm{P}} = 0.20$, and $c_{\mathrm{K}} = 0.073$. The value of c_{N} is consistent with Kihara *et al.* (2020), who report, across 368 trials in sub-Saharan Africa, a marginal response of 2.07 t/ha for nitrogen versus only 0.20 t/ha for potassium. This set of coefficients slightly worsens the RMSE to 0.77 t/ha but preserves a progressive marginal effect of urea over the whole range of observed nitrogen deficit. Unlike c_{N}, which reflects very slow saturation, the higher values of c_{P} and c_{K} reflect the rapid saturation of phosphorus and, to a lesser extent, potassium, linked to their fixation by the iron and aluminum oxides of heavily weathered ferralsols (Hinsinger *et al.*, 2017; Owino *et al.*, 2015; Truong, 1996) [3] [4].

Organic matter enters the model at two distinct levels: 1) as its own factor f_{MO} in the Baule product, and 2) as a multiplier amplifying the apparent effectiveness of applied nitrogen, phosphorus, and potassium (mo_efficiency_factor, increasing linearly with OM, capped at 1.8). This dual role reproduces the bell shaped dynamic observed empirically on the theoretical economic optimum (see Section 4.4): at moderate OM, amplification dominates and encourages greater investment in mineral inputs; beyond a certain threshold, the rapid saturation of f_P and f_K (higher coefficients) reduces the marginal gain, and nitrogen, which saturates more slowly, becomes the residual limiting factor, before it too saturates at very high OM. A complementary inter-seasonal dynamic models the progressive acidification of the soil by uncompensated mineral fertilizers, and the durable corrective effect of cumulative compost on pH.

$$\mu_{MO} = \text{clip}\left(1 + \alpha \cdot \max\left(MO - 1.5, 0\right), \mu_{min}, \mu_{max}\right)$$

$$pH_{eff} = \text{clip}\left(pH + \delta_{compost} \cdot C_{tha} + \delta_{MO} \cdot \max\left(MO - MO_{ref}, 0\right), 4.0, haut\right)$$

Symbol	Constant in code	Value
μ_{max}	MO_MULTIPLICATEUR_PLAFOND	1.8
μ_{min}	MO_MULTIPLICATEUR_PLANCHER	0.5
"high" (upper bound of pH_{eff})	MITS_Y_MAX	9.0

3.2. MDP Formalization and Zone Specialization Rule

The state observed by the agent is a 9-dimensional vector: s = [pH, N, P, K, OM, EC, predicted yield, cumulative compost, phase], where EC (electrical conductivity, mS/cm) was added to capture overall ionic soil availability, complementing individual N/P/K levels. The action is a 3-dimensional vector a = [compost dose, NPK dose, urea dose], with compost capped at 2 t/ha per IRAD's recommendation for the zone. A zone rule fixed a priori (not learned) filters the action actually applied according to the pH of the current state: below the threshold pH_threshold = 5.5, only the compost dimension is retained (NPK and urea are forced to zero); above it, only the NPK and urea dimensions are retained. This architectural choice stems from an empirical observation: left free on all three dimensions simultaneously, the agent learned the pH = 5.5 boundary itself very slowly, even though this is information already available in the agronomic literature, which diluted the learning signal available for fine dosing within each regime. **Figure 2** illustrates this hybrid rule-learning architecture.

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$$

$$s = \left[pH, N, P, K, MO, EC, \hat{Y}, C_{cum}, \phi\right] \in \mathbb{R}^9$$

$$a = \left[a_{compost}, a_{NPK}, a_{uree}\right] \in \left[-1, 1\right]^3$$

$$a_i^{reel} = \frac{a_i + 1}{2} \times a_i^{max}, a^{max} = \left[2.0\ t/ha, 200\ kg/ha, 150\ kg/ha\right]$$

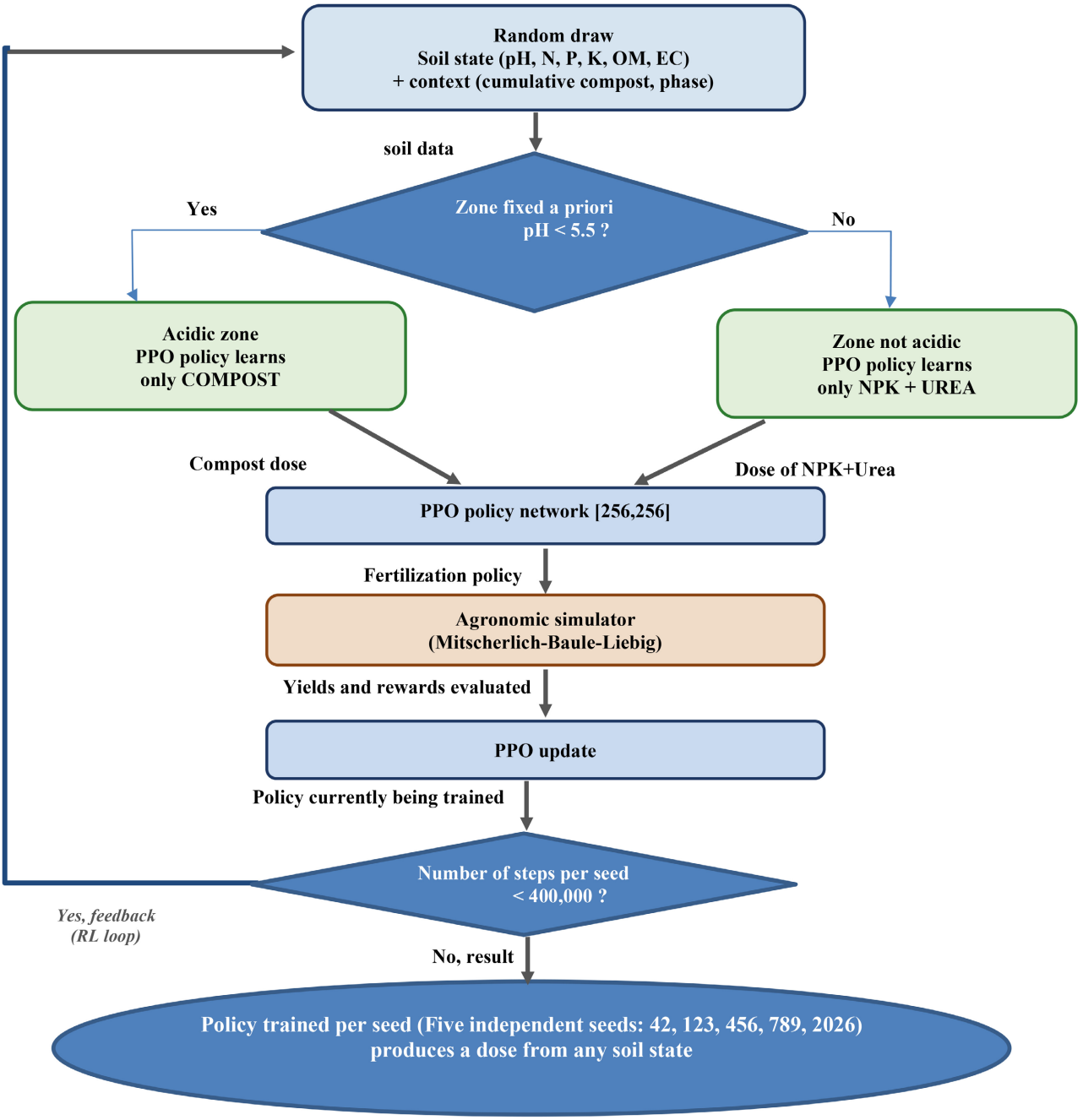


Figure 2. Training architecture of the hybrid rule-based and learning-based decision system.

The system state is observed before the action of each phase is chosen. In phase 1, at sowing, the agent observes the initial soil state, then chooses a compost or NPK dose. This action only updates the cumulative compost in the state. No yield is computed at this stage, and no reward is assigned yet. In phase 2, at growth stage six, the agent observes the updated state, including the cumulative compost from the previous phase. It then chooses a urea or additional compost dose. This second action triggers the final yield computation for the season, through the agronomic simulator. The reward for the whole episode is assigned at this precise

moment, at the end of phase 2.

Figure 3 presents the complete architecture of the agent's recommendation system during the exploitation phase. This architecture also incorporates the weather dimension before providing a final recommendation.

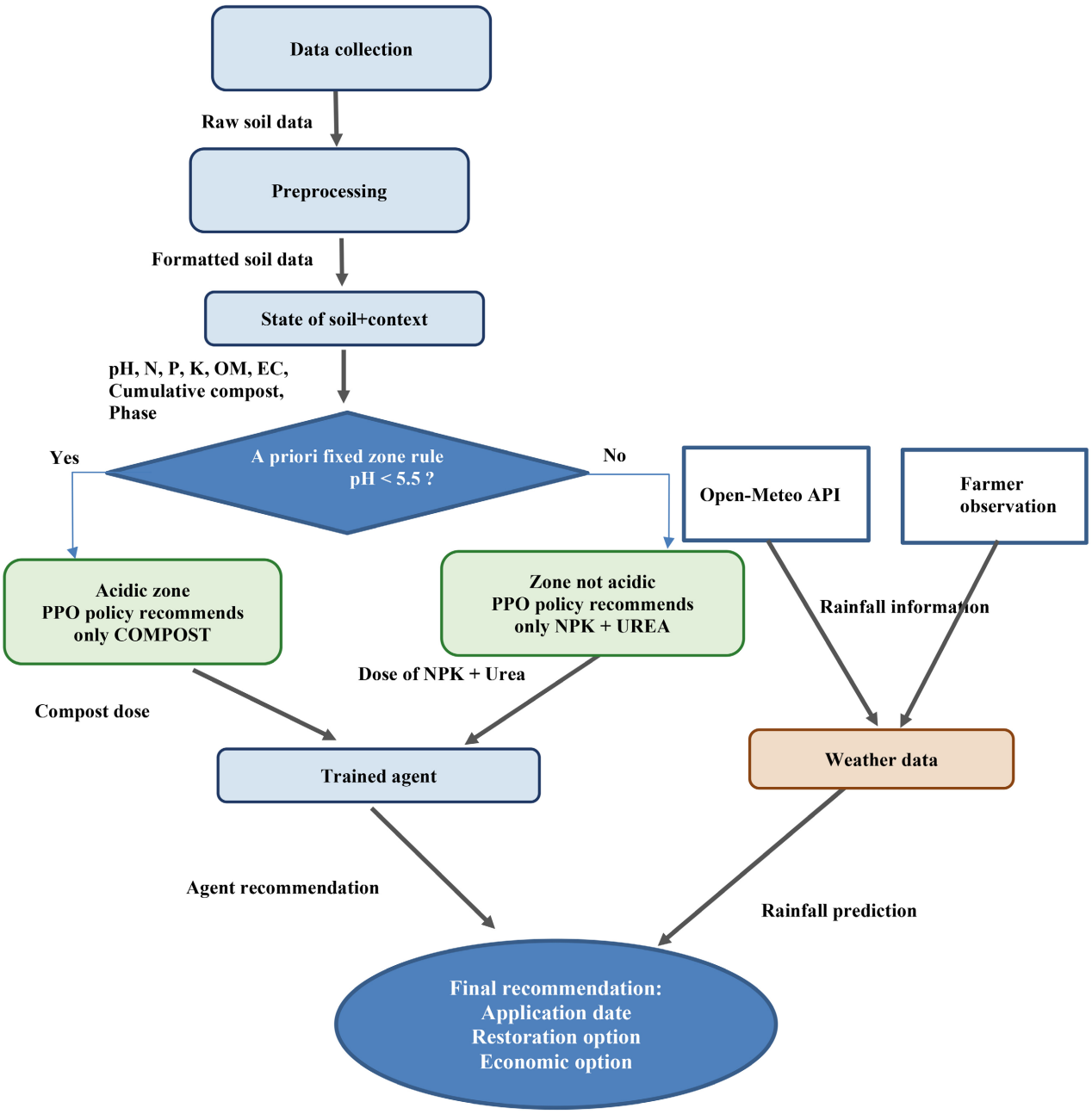


Figure 3. Complete architecture of the agent's recommendation system in deployment phase.

3.3. Reward Function

The reward combines the marginal revenue linked to the yield gain with the gross margin (revenue minus input costs minus penalties plus a soil restoration bonus), calculated from real prices observed on Cameroonian markets (Table 1).

Table 1. Reference unit prices used in the net margin calculation (FCFA).

Input / Product	Unit price	Source
Maize grain (sale)	120,000 FCFA/t	Ngaoundere market, avg. 2024-2025
NPK 20-10-10	600 FCFA/kg	Ngaoundere market
Urea 46% N	600 FCFA/kg	Ngaoundere market
Compost (composted manure)	30,000 FCFA/t	Real cost of smallholder production

$$C(a) = a_{\text{compost}}^{\text{réel}} \cdot p_{\text{compost}} + a_{\text{NPK}}^{\text{réel}} \cdot p_{\text{NPK}} + a_{\text{urée}}^{\text{réel}} \cdot p_{\text{urée}}$$

Each term uses the actual dose (in t/ha or kg/ha), not the network's raw setting (−1 to +1). This is necessary for the product with a price in FCFA/tonne or FCFA/kg to be dimensionally meaningful. $a^{\text{réel}}$

$$R_{\text{brut}} = Y \times 1000 \times p_{\text{maïs}}$$

$$M = R_{\text{brut}} - C(a)$$

The compost price deserves a clarification: it reflects the actual cost of smallholder production (labor, a maturation time of 3 to 6 months, transport logistics), rather than a subsidized price sometimes found in cooperatives (15,000 to 25,000 FCFA/t), which would underestimate the real opportunity cost for an independent producer. These prices, like any market price, are subject to seasonal and interannual variation; a price sensitivity analysis is a natural perspective for this work.

In sum, the reward is normalized by the local economic potential (a function of pH and OM) rather than by a global constant. This prevents the learning gradient from being dominated by states with high absolute economic stakes (rich soils), at the expense of states where the relative gap between a good and a bad decision is equally decisive for policy quality. This local potential is precomputed, prior to training, on a reference grid of 25 × 20 points (25 pH values between 4.0 and 7.0, step ≈ 0.125; 20 OM values between 0.5% and 5.0%, step ≈ 0.237), *i.e.* 500 pH × OM combinations. For each grid point, the potential is obtained through exact optimization (differential_evolution) of the input dose available in the corresponding pH zone (compost in the acidic zone; NPK and urea in the corrected zone), reusing directly the same cost, bonus, and penalty functions applied during training, rather than a duplicated and potentially desynchronized reward formula. The potential thus computed varies by a factor of 5 to 6 depending on pedological context, meaning that the best achievable outcome on the poorest soil is about 5 to 6 times smaller than the best achievable outcome on the richest soil, empirical, quantified proof that these two soils do not at all share the same "realistic ceiling." This confirms the need for adaptive rather than constant normalization.

The value between grid points is obtained through bilinear interpolation. For a soil whose pH and organic matter fall between four points of the grid, the potential is estimated as a weighted average of these four neighboring values. This method avoids a costly recalculation of the potential at every simulation step, while re-

maining continuous across the whole pedological domain covered.

$$B = C_{\text{cum}} \cdot \beta_{\text{compost}} \cdot \delta_{\text{MO}} + \Delta MO \cdot \beta_{\text{MO}},\ \ \delta_{\text{MO}} = \max\left(\frac{\text{MO}_{\text{cible}} - \text{MO}}{\text{MO}_{\text{cible}}}, 0\right)$$

β_{compost} = 20,000 FCFA/tonne, β_{MO} = 30,000 FCFA/% OM, $\text{OM}_{\text{target}}$ = 4.0%.

$$r = \text{clip}\left(\frac{M + B - P_{\text{penalite}}}{\Pi_{\text{local}}(\text{pH}, \text{MO})}, -5.5\right)$$

3.4. Learning Algorithm and Critical Implementation Corrections

The architecture adopted for the final training of the five seeds is a policy/value network with two hidden layers of 256 neurons each. The agent is thus trained using PPO (Stable-Baselines3 library), with an entropy coefficient of 0.02 and 400,000 training steps repeated across 5 independent random seeds (42, 123, 456, 789, 2026). A complementary controlled ablation compared this size to a lighter variant (128, 128), on 3 seeds each, independently of the entropy coefficient. Both network sizes yield statistically indistinguishable performance, ruling out network size as the cause of the residual cross-seed variance observed in the policy's fine dosing. This robustness to network size choice is a reassuring result: the residual variance stems from the flatness of the economic signal near the theoretical optimum, not from an undersized architecture.

Also, the entropy coefficient ultimately retained for training these five seeds is 0.02, not 0.005 as in the earlier version of the system. This increase became necessary after correcting an earlier reward bug (a duplicated bonus of 50.000 FCFA systematically saturating the clip): once this bug was fixed, the immediate cost signal dominated learning too early, and the agent froze on a zero-compost dose without exploring enough to discover the true optimum, located instead between 1.0 and 1.5 t/ha, a degenerate behavior confirmed on two independent seeds despite an otherwise sound reward function. The controlled ablation comparing 0.01 to 0.005 corresponds to a test round predating this correction, distinct from the value of 0.02 ultimately retained for the main results.

$$\pi_\theta(a \mid s) = \mathcal{N}\left(\mu_\theta(s), \sigma_\theta^2\right)$$

$$\rho_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)}$$

$$L^{\text{CLIP}}(\theta) = \hat{E}_t\left[\min\left(\rho_t(\theta)\hat{A}_t, \text{clip}\left(\rho_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right]$$

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1},\ \delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$$

(γ = 0.99; λ = 0.95; ε = 0.2), standard values taken from the literature, from the library used (Stable-Baselines3), and from the original PPO article (Schulman *et al.*, 2017).

During the development of this model, we encountered several difficulties. Six implementation flaws deemed major were identified and corrected during development and had a decisive impact on the quality of the learned policy, and con-

stitute a methodological result in their own right for this work:

1) Action scaling. We use Gymnasium's RescaleAction.

Stable-Baselines3 starts with exploration noise of 1.0. For NPK, the range is 0 - 200 kg/ha. So 1.0 is only 0.5% of the range. The agent almost does not explore. For compost, the range is 0 - 2 t/ha. There, 1.0 is 50% of the range. It explores well. Result: compost learned, NPK and urea stayed at zero. We fixed it by rescaling all actions to [−1, 1]. Now the network sees the same scale for each input.

2) Observation scaling. Same problem for observations.

Soil values have very different scales. pH is 4 - 7. Phosphorus is 2 - 14. Organic matter is 0.5 - 5. Without scaling, the network cannot learn a bell-shaped relation. For example, optimal urea first rises with organic matter, then falls. The network learned a flat line instead. We applied VecNormalize. We saved the mean and std from training and reused them at test time.

3) Urea efficiency bug.

We boost nitrogen from urea in phase 2 by 1.5× (Uhart & Andrade, 1995). In the first code, we boosted all nitrogen: soil + compost + NPK + urea. That hid the benefit of urea. NPK already brings N plus P and K for the same price. So the agent learned urea = 0 for all soils. We fixed it: the 1.5× now applies only to N from urea. Urea became useful again on low-N soils.

4) Compost bonus counted twice.

We give a bonus for compost on acidic soil. It was added twice: once in gross margin, once in final reward. Also it grew forever with cumulative compost, even on restored soil. The agent kept adding compost even at 4% OM. We removed the double count. Now the bonus fades out when organic matter approaches 4%.

5) Wrong normalization at inference.

SHAP and recommendation scripts used raw soil values. Training used normalized values. So SHAP underestimated the true effect by 18× for urea vs soil N. We fixed it by applying the saved VecNormalize stats at every inference call.

6) RMSE is not enough.

We calibrated c_N, c_P, c_K by minimizing RMSE. Two sets had almost same RMSE: 0.57 and 0.77 t/ha. First set killed urea effect completely. Second set kept it. RMSE alone did not show the bug. We now check two things after each calibration: RMSE and a simple behavior test (does urea dose fall when soil N rises?). Both must pass.

3.5. Reference Policies (Baselines)

Four fixed policies serve as comparison:

- B0: no input (pure subsistence).
- B1: full MINADER recommendation, applied blindly regardless of pH (NPK in phase 1, urea in phase 2).
- B2: compost alone (2 t/ha), an improved smallholder practice.
- B3: compost combined with the MINADER recommendation.

The zone filter is never applied to these reference policies. They represent real

smallholder practices, which may apply NPK/urea even on acidic soil. Not applying this filter is a necessary condition for the comparison with PPO to reveal the true cost of mineral fertilizer misuse on uncorrected soil.

3.6. Statistical Evaluation Protocol

Each policy is evaluated on 1000 simulation episodes. For PPO, this number is accumulated across the five independent training seeds (*i.e.* 5 × 1000 = 5000 episodes in total), while each of the four reference policies (B0 to B3), having no random component to vary, is evaluated on 1000 episodes only. The set of soils tested comes from a single, fixed evaluation seed, deliberately distinct from the seeds used for training. This same list of 1000 soils (same pH, N, P, K, OM, EC values) serves as an identical test bench for all policies compared, with none advantaged or disadvantaged by an easier or harder soil draw than the others. Thus, only the quality of the decision made on the same soil distinguishes the policies from one another.

The statistical significance of each performance gap is assessed using Welch's t-test, a variant of the classical t-test adapted to cases where the two groups compared do not share the same variance, a necessary precaution since PPO (5000 episodes) and each baseline (1000 episodes) have neither the same sample size nor, a priori, the same dispersion of results. Since four independent comparisons are conducted (PPO versus B0, B1, B2, and B3), a Bonferroni correction is applied to contain the overall risk of false detection: the usual significance threshold (α = 0.05) is divided by the number of comparisons (n = 4), giving an adjusted threshold α' = 0.0125 that each individual p-value must cross to be considered significant. Finally, beyond mere statistical significance, the magnitude of each gap is quantified using Cohen's d, which relates the difference in means to the natural dispersion of results within each group. This makes it possible to distinguish a real but negligible gap from a real and substantial one.

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}},\ s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$\alpha' = \frac{\alpha}{n},\ n = 4 \text{ comparaisons} \Rightarrow \alpha' = 0.0125$$

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$r_{rb} = 1 - \frac{2U}{n_1 n_2}$$

3.7. Interpretability

The final policy is analyzed using SHAP (Kernel Explainer) and LIME (Lime Tabular Explainer), each applied separately to the two pH zones to prevent pH from artificially dominating the attribution (since it determines which dose is non-zero). State variables that are structurally constant at the moment of the initial

decision (predicted yield, cumulative compost, phase) are excluded from the analysis: SHAP correctly assigns them zero importance, but LIME, in the absence of variance in the background data, assigned them artificially high weights. This numerical artifact was identified and corrected prior to the analysis. The entire procedure is repeated independently across the 5 training seeds, to assess the stability of the attribution beyond a single realization.

3.8. Deployment and Operating Architecture

Beyond experimental validation, this work proposes an operating pipeline designed to turn the trained model into a tool usable by an agricultural extension officer or an IRAD agent in the field, with no prior knowledge of RL. This pipeline comprises five successive steps, from soil measurement to the delivered recommendation, detailed below.

The pipeline is built around a structured Excel workbook (Figure 4), designed to remain accessible to field agents without computer training.

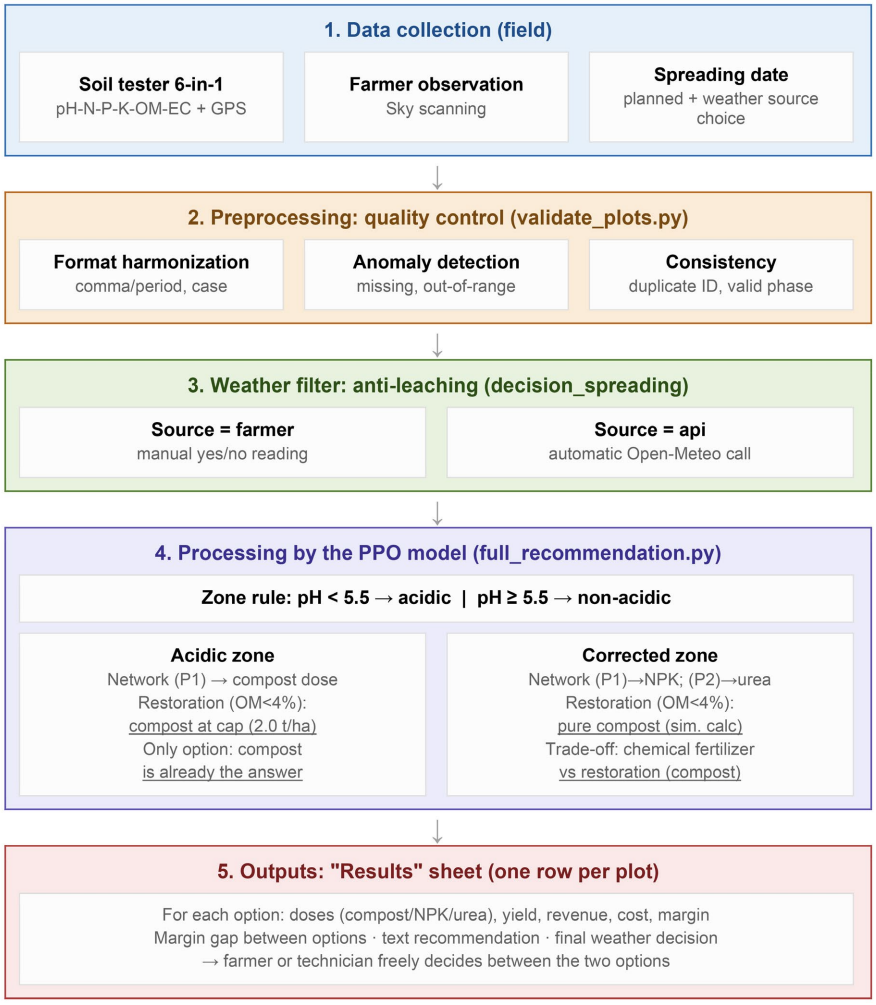


Figure 4. Complete processing pipeline, from field data collection to recommendation: data collection, preprocessing (quality control), weather filter, PPO decision, and output.

1) Collection: plot data (pH, N, P, K, OM, electrical conductivity EC, GPS coordinates, up to 50 plots per workbook) are collected using a handheld 6-in-1 soil tester and a GPS module. These data are then entered into the "Plots" sheet, and supplemented on the one hand with weather data from either the Open-Meteo API or the farmer's own observation (rain risk), and on the other hand with the planned application date.

2) Preprocessing: a quality control layer (valider_parcelles.py) validates each entry before any processing, enabling format harmonization and detection of missing or aberrant values. This separation between data transformation and cleaning was inspired by the architecture proposed by Tanzouak *et al.* (2019) [20] for integrating heterogeneous data in forecasting systems.

3) Weather filter: depending on the source chosen, the application decision results either from the farmer's stated observation or from an automatic call to the Open-Meteo API (free, no key required), targeting the 12-hour window following the planned application date rather than the script's execution time, with a fallback to manual entry in case of network failure.

4) Decision: the trained PPO model (recommandation_complete.py) produces an optimal option by pH zone (compost alone below 5.5; NPK then urea above); as long as organic matter has not reached a restoration target (4.0%), a complementary restoration option is computed. In the acidic zone, compost is already the only possible response. In the corrected zone, however, a pure compost alternative is computed directly via the simulator, since this scenario was never encountered by the policy during training (compost being always masked beyond the zone threshold there).

5) Output: the two options (doses, yield, revenue, cost, margin), the gap between them, and the final weather decision are written back into the "Results" sheet of the original workbook, leaving the farmer the final choice between immediate profitability and investment in the plot's long term fertility. In the acidic zone, NPK and urea are zero.

4. Results and Discussion

4.1. Study Area and Pedological Context

The study is conducted in the Adamaoua region (Cameroon), in partnership with the IRAD-Wakwa station, on soils typical of Ferralsols. These soils are heavily desaturated, acidic, with pH generally between 4.0 and 6.5 on smallholder plots, and organic matter content that varies widely (0.5% to 5%) depending on the plot's management history (fallow, prior organic inputs, erosion, etc.). The reference maize variety used to calibrate the ceiling yield is Shaba, whose yield potential (8 to 9 t/ha) serves as the simulator's upper bound.

4.2. Case Studies

This study was conducted precisely in Dang, site as the agricultural post used for our cases 5 to 13. A single point taken directly from Google Earth is used. This is

7.453508 N and 13.579961 E. For IRAD-Wakwa, the midpoint of the site is 7.268214 N and 13.535701 E.

IRAD-Wakwa observations correspond to a single occurrence per case. Observations from the agricultural post of Dang correspond to the average of two occurrences per case. Table 2 clearly highlights these case studies.

Table 2. Real field observations (IRAD-Wakwa and the Dang agricultural extension officer) used as targets for the simulator's numerical calibration. Cases involving compost (Cases 3 and 4) receive a higher weight, reflecting their importance in validating the pH correction mechanism.

Case	Source	Replication	pH (M)	N %	P mg/kg	K mg/kg	OM %	Yield
1	IRAD-Wakwa: pH 4.5, no input	$n = 1$	4.5	0.12	4.0	0.14	2.1	1.5 t/ha
2	IRAD-Wakwa: pH 4.5, NPK 200 kg/ha + urea 100 kg/ha	$n = 1$	4.5	0.12	4.0	0.14	2.1	2.0 t/ha
3	IRAD-Wakwa: pH 4.5, compost 2 t/ha	$n = 1$	4.5	0.12	4.0	0.14	2.1	5.0 t/ha
4	IRAD-Wakwa: pH 4.5 to 6.4, compost 2 t/ha/year over 3 years (cumulative)	$n = 1$	4.5 to 6.4	0.18	8.0	0.22	4.2	7.3 t/ha
5	Agricultural post, Dang: pH 5.0, no input	$n = 2$ (avg.)	5.0	0.14	5.0	0.16	2.4	1.7 t/ha
6	Agricultural post, Dang: pH 5.0, compost 2 t/ha	$n = 2$ (avg.)	5.0	0.14	5.0	0.16	2.4	5.2 t/ha
7	Agricultural post, Dang: pH 5.0, NPK 200 + urea 150 kg/ha	$n = 2$ (avg.)	5.0	0.14	5.0	0.16	2.4	2.3 t/ha
8	Agricultural post, Dang: pH 5.5, no input	$n = 2$ (avg.)	5.5	0.15	6.0	0.18	2.8	2.5 t/ha
9	Agricultural post, Dang: pH 5.5, compost 2 t/ha	$n = 2$ (avg.)	5.5	0.15	6.0	0.18	2.8	5.7 t/ha
10	Agricultural post, Dang: pH 5.5, NPK 200 + urea 150 kg/ha	$n = 2$ (avg.)	5.5	0.15	6.0	0.18	2.8	5.0 t/ha
11	Agricultural post, Dang: pH 6.2, no input	$n = 2$ (avg.)	6.2	0.17	7.5	0.21	3.8	4.2 t/ha
12	Agricultural post, Dang: pH 6.2, compost 2 t/ha	$n = 2$ (avg.)	6.2	0.17	7.5	0.21	3.8	6.5 t/ha
13	Agricultural post, Dang: pH 6.2, NPK 200 + urea 150 kg/ha	$n = 2$ (avg.)	6.2	0.17	7.5	0.21	3.8	6.0 t/ha

These thirteen observations directly confirm, on real data, the hierarchy already identified by the simulated baselines (Section 4.3): applying mineral fertilizer on uncorrected acidic soil (case 1) yields only a marginal gain (1.5 to 2.0 t/ha) at the cost of substantial input expense. Compost (case 2, then case 3 with a cumulative effect), by contrast, produces a markedly higher yield gain, together with a lasting rise in pH. The following coefficients (c_N, c_P, c_K, and the parameters of the pH/OM dynamics) are those that minimize the gap to these thirteen observations, under plausibility constraints drawn from the literature cited below. One observation spanning three consecutive seasons (IRAD-Wakwa plot, 2 t/ha of compost per year) documents in particular a pH restoration trajectory from 4.5 to 6.4 over three years, together with a doubling of organic matter (2.1% to 4.2%) and a final yield of 7.3 t/ha. This data represents the most direct empirical anchor available for the model's multi-year dynamics, although it is not exploited in the current single-season training campaign (see Section 5.1).

4.3. Comparative Performance

The analysis has been redone with a paired t-test. For each of the 1000 evaluation soils, the mean of PPO across the five seeds is compared to each baseline value on that same soil.

PPO statistically outperforms all four baselines (Table 3), with effect sizes classified as "large" for B0, B1, and B3 (d from 1.011 to 1.501), and "medium" for B2 ($d = 0.513$). The largest gap is now observed against B1 (blind MINADER alone, $d = 1.501$), confirming that undiscriminating application of mineral fertilizer on acidic soil, without a pH diagnosis, is frankly counterproductive. The narrowest gap, with B2 (compost alone, $d = 0.513$), is naturally explained: PPO itself now applies substantial compost doses (1.0 to 2.0 t/ha depending on organic matter, Section 4.5), bringing its strategy closer to that of B2. The four baselines show distributions with varying skewness (Figure 5): B0 (no input) is markedly positively skewed (skewness = +0.81, mean 0.192 above the median 0.165); B1 (blind MINADER) remains nearly symmetric (skewness = −0.01); B2 (compost alone) shows a slight negative skew (skewness = −0.35, mean 0.247 close to the median 0.252); B3 (compost + MINADER) shows the most pronounced skew (skewness = −1.09). A non-parametric Mann-Whitney test, less sensitive to these skews, confirms the significance and direction of the gap for all four comparisons ($p < 10^{-44}$ in every case), ruling out the hypothesis that PPO's superiority is driven only by a few extreme episodes.

The reward shown in Figure 6 differs from that in Table 3 in its normalization. During training, VecNormalize applies its own scale, based on a running variance, distinct from the normalization by local economic potential used at final evaluation. This difference explains the observed scale gap between the two, without calling into question the validity of either measure.

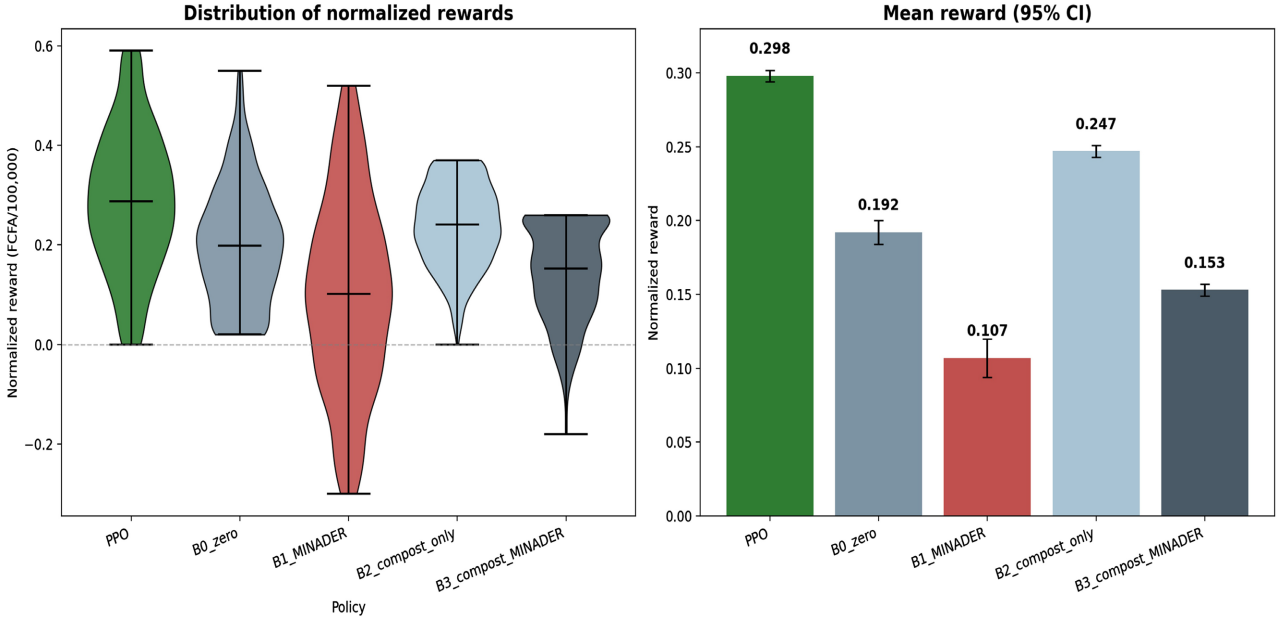


Figure 5. Distribution of normalized rewards, PPO versus the four reference policies (n = 5000/1000 episodes).

Table 3. Corrected statistical comparison (Paired t-test, n = 1000 shared soils, PPO averaged across five seeds, Bonferroni correction).

Policy	Mean	Standard deviation	Cohen's d vs PPO	p (paired)
PPO (method)	0.2956	0.0459	n.a.	n.a.
B0 (no input)	0.1958	0.1172	1.122	1.53×10^{-110}
B1 (blind MINADER)	0.1116	0.2046	1.241	2.95×10^{-127}
B2 (compost alone)	0.2493	0.0709	0.775	1.15×10^{-60}
B3 (compost and MINADER)	0.1561	0.0752	2.238	5.79×10^{-278}
B4 (zone rule only)	0.2257	0.0877	0.999	1.66×10^{-91}

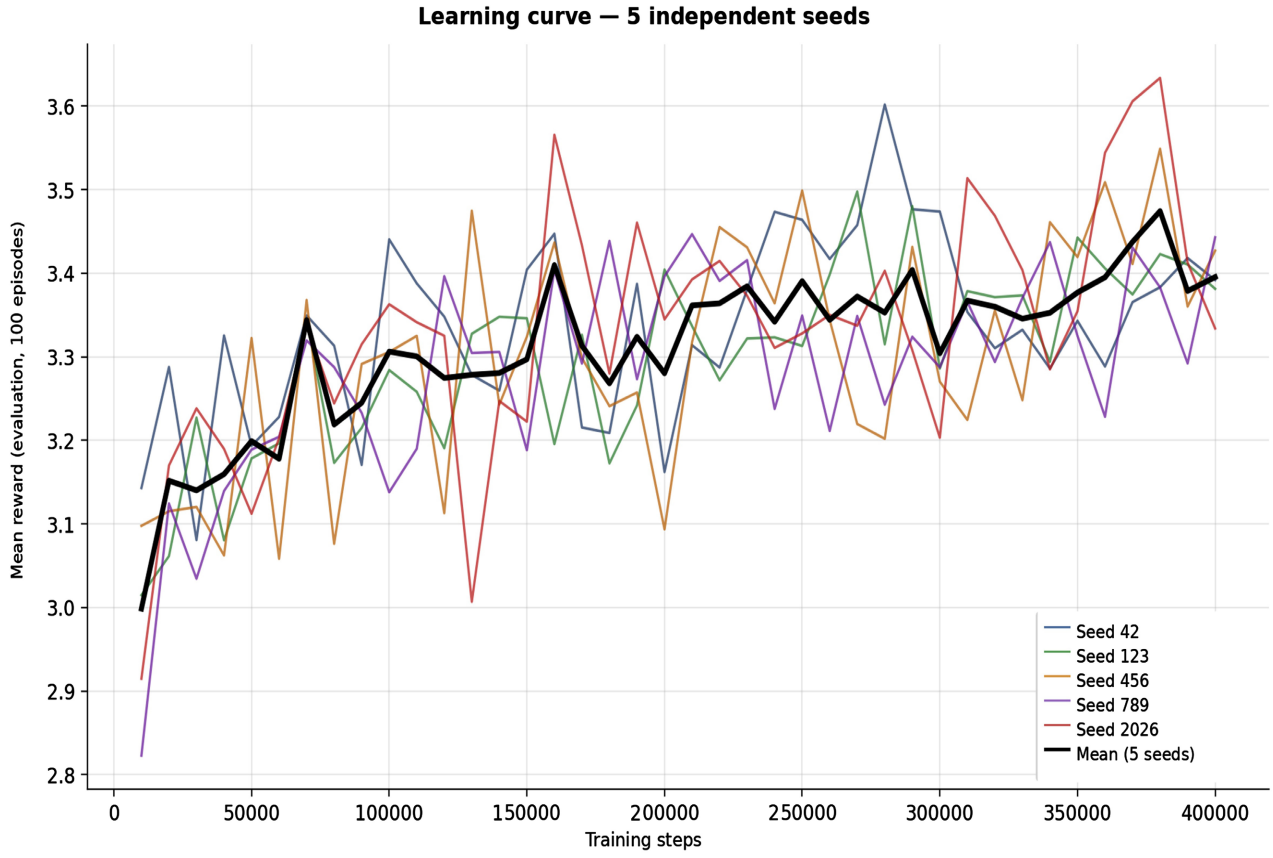


Figure 6. PPO learning curve (ensemble mean and per seed), 5 independent training seeds, 400,000 steps each. The reward shown here is normalized by VecNormalize during training, an internal scale used by the algorithm. It does not directly correspond to the economic reward normalized by local potential, reported in Table 3.

4.4. Learned Policy versus Theoretical Optimum

Figure 7 presents the ensemble mean policy (mean of the actual doses across the 5 seeds) as a function of pH and organic matter, and Figure 8(a) the corresponding theoretical economic optimum, computed independently of RL through exact optimization (differential_evolution) on the simulator alone.

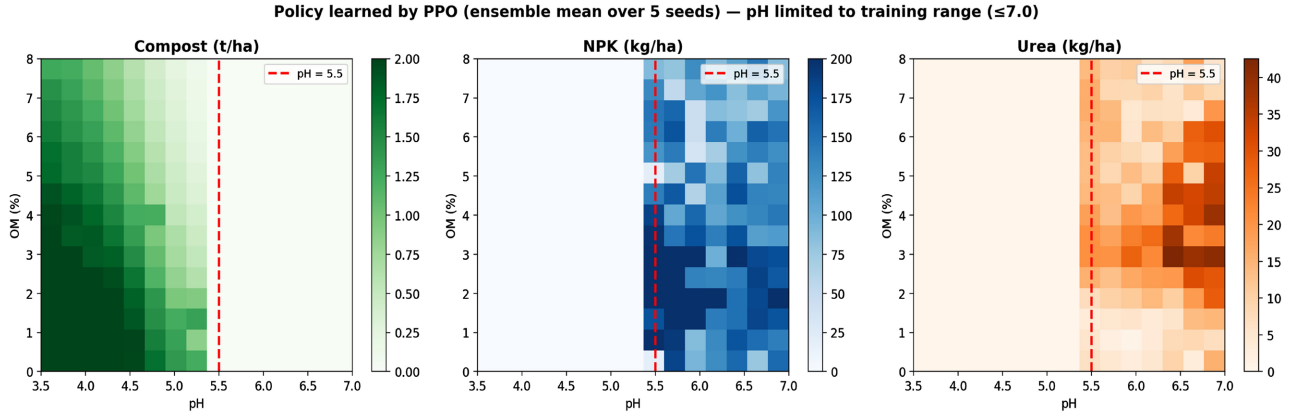
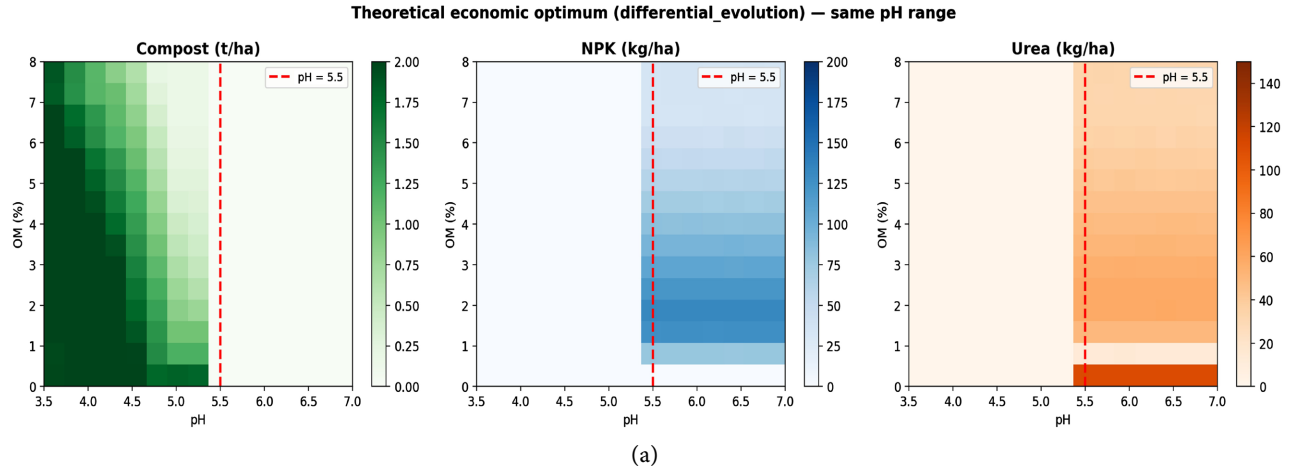
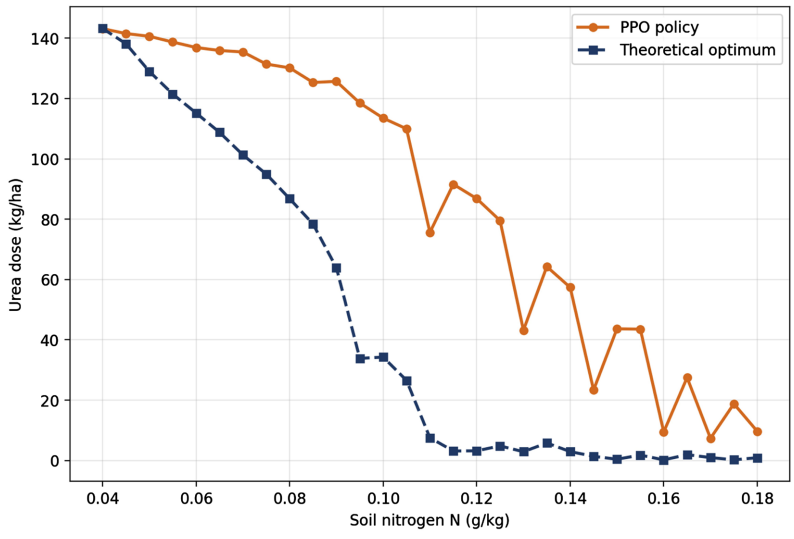


Figure 7. Policy learned by PPO (ensemble mean across 5 seeds): compost, NPK, and urea doses as a function of pH and organic matter.



(a)



(b)

Figure 8. (a) Theoretical economic optimum (independent computation via differential_evolution, without PPO), for direct comparison with **Figure 7**. (b) Direct comparison, at fixed pH (corrected zone), between the PPO policy and the theoretical optimum for urea dose as a function of soil nitrogen. Strong qualitative agreement (correlation r = 0.851); mean gap of 41.2 kg/ha in the intermediate deficit zone (maximum gap of 88.4 kg/ha), full convergence at both extremes.

On the pH × OM axes of Figure 7, the urea panel appears nearly empty across the entire grid. This result could wrongly suggest that the policy never uses urea. It is in fact an artifact of the choice of axes: this grid fixes soil nitrogen (N) at its median value, whereas the interpretability analysis (Section 4.5) establishes that N is precisely the factor that drives the urea decision, not pH or OM. At median N, the exact optimizer itself in fact yields an almost zero urea dose across the entire pH × OM grid. The PPO policy is therefore aligned with the theoretical optimum on this slice, which validates rather than contradicts the learned behavior. A complementary figure (urea as a function of N, at fixed pH and OM in the corrected zone) reveals the relationship actually driving the agent: a phase saturated at the ceiling (150 kg/ha) for N ≤ 0.08 g/kg, followed by a smooth, monotonic decrease down to a zero dose from N ≈ 0.17 g/kg onward, the expected nitrogen deficit behavior (Uhart & Andrade, 1995). Comparison with the corresponding theoretical optimum (Figure 8(b)) shows strong qualitative agreement (correlation r = 0.880 on the tested grid) but imperfect numerical convergence: the policy remains systematically more generous with urea than the exact optimum, with a mean absolute gap of 42.7 kg/ha and a maximum gap of 80.0 kg/ha, reached around N ≈ 0.10 g/kg. Both curves converge to zero at high N values, although the theoretical optimum reaches this ceiling earlier (N ≈ 0.14 g/kg) than the PPO policy (N ≈ 0.17 g/kg). This convergence gap, rather than a failure, illustrates the expected difference between an agent trained on a finite budget of episodes and an exact optimizer directly querying the simulator: the policy has learned the correct decision structure, progressively decreasing urea as soil nitrogen content rises, down to zero, without necessarily achieving exact fine-tuning across the entire range.

4.5. Interpretability and Cross-Seed Stability

To verify that the learned policy rests on coherent agronomic logic rather than a spurious correlation, the network's decision is broken down, variable by variable, using two independent interpretability methods, SHAP and LIME, whose agreement serves as a guarantee of the robustness of the results obtained.

The SHAP analysis (Table 4), repeated across the 5 seeds, now covers the three decisions actually made by the network: compost at the planting phase (P1, acidic zone), NPK (P1, corrected zone), and urea (P2, corrected zone). Compost at the grain filling phase (P2, acidic zone) remains a decision made by the network, but is fixed at 0 by construction only in the corrected zone. In the acidic zone, the learned policy converges empirically toward near zero values, without this being architecturally imposed. This result remains valid both economically and agronomically (Sections 3.4 and 5.1). Each output reveals a dominant variable that is consistent across seeds, with a degree of stability that varies by output. We find pH for compost (5/5 seeds, CV = 8.5%), consistent with its role as the trigger factor for the acidic zone; soil nitrogen for urea (5/5 seeds, CV = 7.5%), consistent with the correction of the simulator's nitrogen coefficient (Section 3.1), which restores a real and progressive marginal effect for this input, the most stable and

most pronounced result of the three decisions, with the gap to the second variable (organic matter) reaching a factor of 3 to 5 depending on the seed. For NPK, however, potassium and phosphorus compete closely and unstably for dominance: potassium dominates in 4 of the 5 seeds tested (K: 37.8, CV = 12.5%; P: 33.6, CV = 4.3%), but a cross-check using LIME on the same seeds confirms this dominance in only 3 of the 5 seeds, with the verdict reversing on at least one seed between the two methods, signaling that this result is sensitive not only to the sample of soils tested but also to the interpretability method itself. This instability nonetheless remains compatible with the converging content of the regional pedological literature (Section 2) on the strong joint fixation of phosphorus and, to a lesser extent, potassium in heavily weathered ferralitic soil. Both variables remain markedly more influential than pH, nitrogen, or organic matter for this particular decision.

Table 4. Stability of the SHAP attribution of the dominant variable for each output, across the 5 training seeds (CV = coefficient of variation; mean importance expressed in the unit specific to each input, t/ha for compost, kg/ha for NPK and urea).

Output	Variable	Dominance	Mean importance	Standard deviation	CV (%)
Compost (P1, planting)	pH	5/5 seeds (SHAP and LIME)	0.428	0.037	8.5%
Urea (P2)	N	5/5 seeds (SHAP and LIME)	54.65	4.114	7.5%
NPK (P1)	**K**	4/5 seeds (SHAP), 3/5 seeds (LIME)	37.84	4.741	12.5%
NPK (P1)	**P**	1/5 seeds (SHAP), 2/5 seeds (LIME)	33.56	1.452	4.3%

To check the robustness of these results independently of the interpretability method used, the same analysis was repeated using LIME (LimeTabularExplainer), on the same five seeds and the same three decisions, a method resting on a mathematical principle different from SHAP (local linear approximation rather than additive Shapley decomposition). Agreement between the two methods is total for compost and urea: pH dominates 5 seeds out of 5 for compost, nitrogen dominates 5 seeds out of 5 for urea, fully confirming the results obtained with SHAP for these two decisions. For NPK, however, LIME reveals the same instability already suspected with SHAP, but in a different light: potassium dominates in only 3 of 5 seeds there (versus 4 of 5 with SHAP), and the verdict diverges between the two methods on at least one seed (2026: K according to SHAP, P according to LIME). This double check confirms that potassium's dominance on NPK, unlike that of pH and nitrogen on the other two decisions, does not constitute a result robust to a change of method.

A closer look at dispersion itself sheds light on this instability: potassium, although dominant more often, varies noticeably more from one seed to another

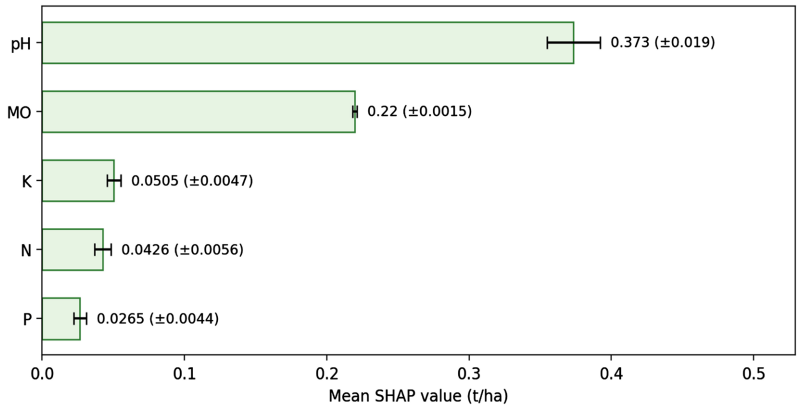
(CV = 12.5%) than phosphorus (CV = 4.3%). In other words, when phosphorus wins out, it does so with relatively constant magnitude; potassium, by contrast, wins out more frequently but with a logical contribution whose magnitude fluctuates more depending on the seed tested. This dual reading, frequency of dominance on one hand, regularity of magnitude on the other, suggests that potassium captures an effect more dependent on the precise pedological context of each sample, whereas phosphorus reflects a more systematic constraint, consistent with its particularly strong and well documented fixation in ferralitic soil (Section 2), regardless of fine variations in the composition of the soil tested.

We can clearly see that this stability varies by output: high and perfectly consistent for compost (pH dominant on 5/5 seeds, CV = 8.5%) and for urea (nitrogen dominant on 5/5 seeds, CV = 7.5%); more nuanced for NPK, where potassium dominates on 4 of the 5 seeds with SHAP but only 3 of 5 with LIME, the two methods diverging on at least one seed (2026: K according to SHAP, P according to LIME). This instability, confirmed by two independent methods, signals a rank that varies with initialization and with the interpretability method rather than an absence of signal, with the magnitude of potassium's attribution nonetheless remaining more variable (CV = 12.5%) than that of phosphorus (CV = 4.3%). This coexistence remains consistent with the regional pedological literature, which documents strong joint fixation of phosphorus and, to a lesser extent, potassium on ferralitic soils, with neither universally dominating the other.

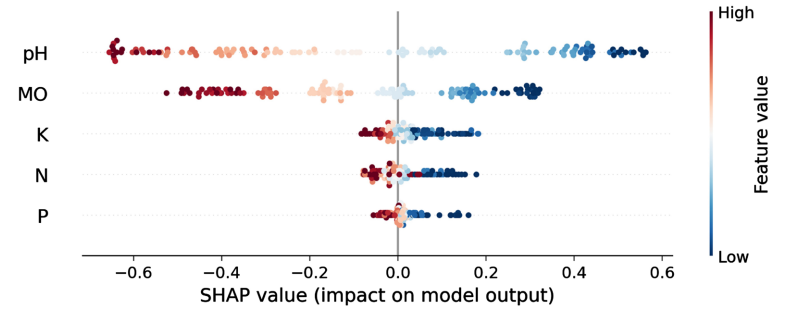
SHAP/LIME importance (mean ± SD across 5 seeds) for recommended dose. **Figure 9(b)** shows each variable's effect on the recommended compost dose in phase P1.

Each point represents a tested sample; its horizontal position indicates the variable's effect on that specific prediction, and its color the value of the variable for that sample (blue = low, red = high). **Figure 9(c)** presents mean variable importance (LIME) for the compost dose, for direct comparison with **Figure 9(a)** (SHAP).

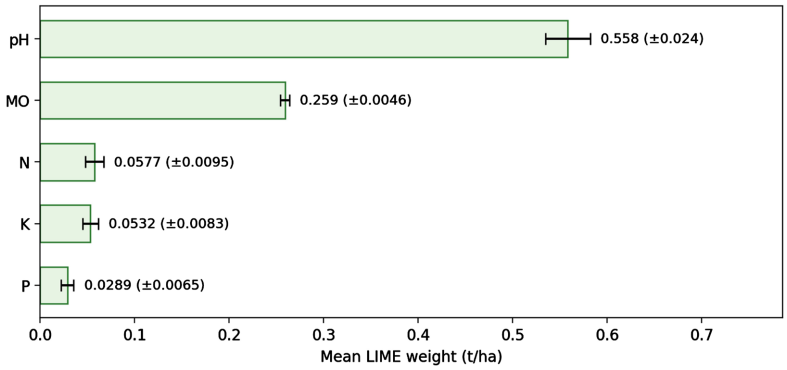
Figures 9-11 present the interpretability of the learned policy, one figure per input and per method. Panels a and b come from SHAP. Panel c comes from LIME.



(a)



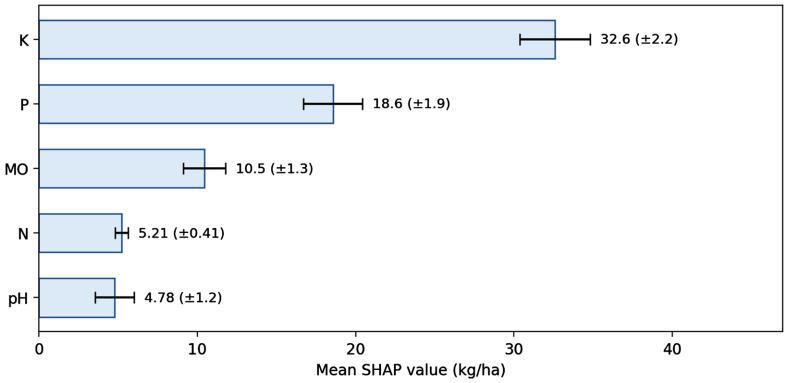
(b)



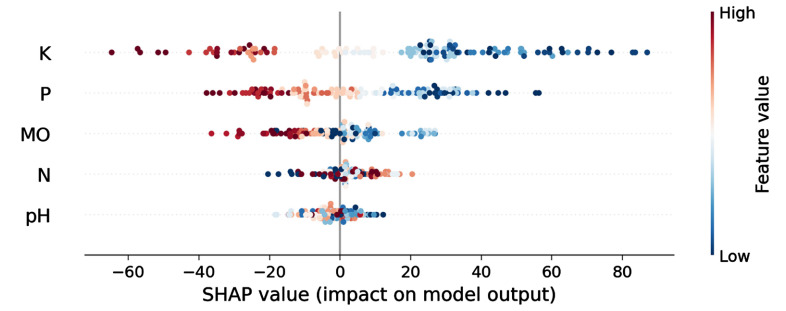
(c)

Figure 9. (a) Mean variable importance for the compost dose (pH zone < 5.5), (b) Effect of each variable on the compost dose recommended at the planting phase (P1). pH dominates by far, consistent with its role as the trigger for the initial decision, (c) Mean variable importance for the compost dose (LIME), for direct comparison with (a) (SHAP). SHAP/LIME importance (mean ± SD across 5 seeds) for recommended dose.

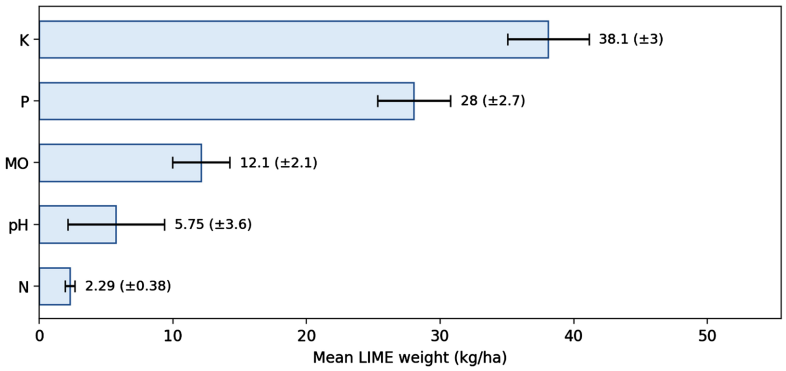
Figure 9(a) shows the mean importance of the variables (pH, OM, N, K, P) according to SHAP for the compost dose at sowing (P1), in the acidic zone (pH < 5.5). It is a bar chart, in t/ha. Figure 9(b) shows the effect of each variable on the same compost dose, according to SHAP. It is a scatter plot in which each point is a soil case, colored by the value of the variable, from blue (low) to red (high). Figure 9(c) shows the mean importance of the variables according to LIME for the same compost dose. It is a bar chart, for direct comparison with Figure 9(a).



(a)



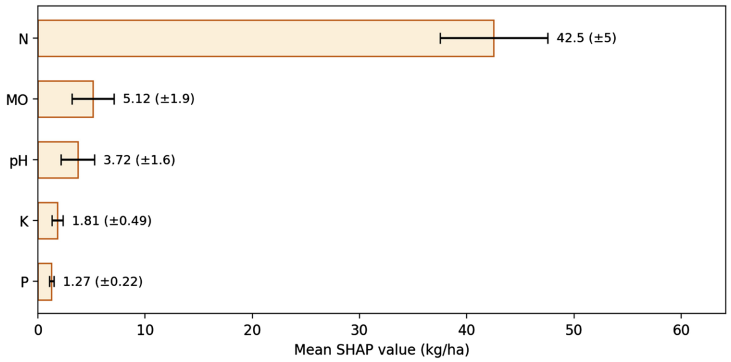
(b)



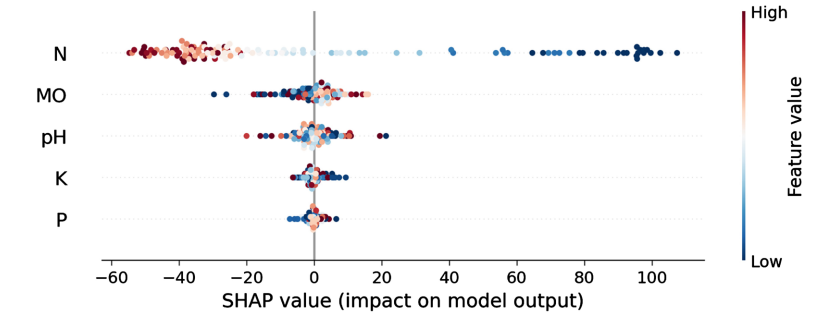
(c)

Figure 10. (a) Mean variable importance for the NPK dose (pH zone ≥ 5.5). (b) Effect of each variable on the recommended NPK dose. Potassium and phosphorus dominate, precisely the two nutrients that NPK supplies. (c) Mean variable importance for the NPK dose (LIME), for direct comparison with (a) (SHAP). Each point represents a tested sample; its horizontal position indicates the variable's effect on that specific prediction, and its color the value of the variable for that sample (blue = low, red = high). SHAP/LIME importance (mean ± SD across 5 seeds) for recommended dose.

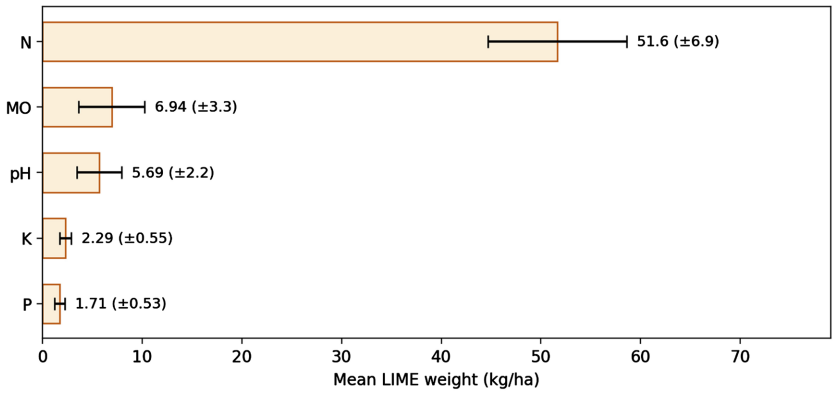
Figure 10(a) shows the mean importance of the variables according to SHAP for the NPK dose at sowing (P1), in the corrected zone (pH ≥ 5.5). It is a bar chart, in kg/ha. **Figure 10(b)** shows, as a scatter plot, the effect of each variable on this NPK dose according to SHAP. **Figure 10(c)** shows the mean importance of the variables according to LIME for the same NPK dose. It is a bar chart, for direct comparison with **Figure 10(a)**.



(a)



(b)



(c)

Figure 11. (a) Mean variable importance for the urea dose (pH zone ≥ 5.5). (b) Effect of each variable on the recommended urea dose (pH zone ≥ 5.5). Soil nitrogen dominates by a very wide margin. (c) Mean variable importance for the urea dose (LIME), for direct comparison with (a) (SHAP). SHAP/LIME importance (mean ± SD across 5 seeds) for recommended dose. Each point represents a tested sample; its horizontal position indicates the variable's effect on that specific prediction, and its color the value of the variable for that sample (blue = low, red = high).

Figure 11(a) shows the mean importance of the variables according to SHAP for the urea dose at growth stage V6 (P2), in the corrected zone (pH ≥ 5.5). It is a bar chart, in kg/ha. **Figure 11(b)** shows, as a scatter plot, the effect of each variable on this urea dose according to SHAP. **Figure 11(c)** shows the mean importance of the variables according to LIME for the same urea dose. It is a bar chart, for direct comparison with **Figure 11(a)**.

The three outputs thus display a coherent and distinct specialization: compost responds mainly to the soil's own state (pH, OM), since it does not directly supply N, P, or K in dominant quantity in the model. NPK responds to potassium and phosphorus deficits, the two elements it supplies. Urea responds almost exclusively to nitrogen deficit. This input to nutrient correspondence is imposed by no explicit rule of the system. It emerges entirely from learning, which reinforces the agronomic credibility of the policy.

4.6. External Validity of the Policy

The PPO policy was trained exclusively on the Mitscherlich-Baule-Liebig simula-

tor (Section 3.1). To verify that its conclusions are not an artifact of this training simulator, it was subjected to cross-validation on three independent simulators, not used for training: QUEFTS (Janssen *et al.*, 1990), recalibrated on the thirteen field observations, APSIM Next Generation, a crop simulator based on first principles, and WOFOST (van Diepen *et al.*, 1989, via the WOFOSTGym framework).

On the first two simulators, the result in the acidic zone remains consistent. On QUEFTS, broken down by pH zone with high statistical power (n = 10,000/2000 episodes per policy), PPO significantly outperforms the absence of input in the acidic zone (Cohen's d = 2.72, $p < 10^{-300}$). This result confirms the robustness of the learned compost strategy, independently of the simulator used to judge it.

On APSIM, broken down by pH zone, PPO significantly outperforms the three reference policies that actually fertilize (B1, B2, B3) in both zones (Cohen's d from 1.21 to 8.35 in the acidic zone and 1.34 to 8.03 in the corrected zone, all significant at the Bonferroni threshold). The relative order of the baselines remains agronomically consistent: compost outperforms NPK/urea in the acidic zone, the reverse in the corrected zone.

The third simulator, WOFOST, completes this picture. Calibrated with only the thirteen field observations, this simulator shows, on twelve independent soils (distributed as six per pH zone by random draw), a net margin ranking in favor of our policy. PPO significantly outperforms the three fertilizing baselines (B1, B2, B3) in both zones (Cohen's d from 1.49 to 15.58).

Scope and limitation of this validation. This four-independent-simulator triangulation (Mitscherlich, QUEFTS, APSIM, WOFOST) reduces the risk that PPO's superiority merely reflects an artifact of the training simulator, particularly against any reference policy that actually fertilizes, a result that is reproduced on all four simulators tested. However, it does not constitute a real agronomic validation in the strict sense: the four simulators remain models calibrated on a small number of field observations (thirteen, Section 4.2), not independent trials, and two of them (APSIM, WOFOST) show calibration limitations that remain unresolved, as documented above. The distinction between these three levels, simulator validation, policy validation within the simulator, and real agronomic validation, is developed in Section 4.7.

4.7. Three Levels of Validation

The validation of a system like this one plays out at three distinct levels, a distinction we consider useful to state explicitly rather than leaving it implicit.

Level 1, simulator validation: does the simulator faithfully reproduce actual yield, for a given set of doses? This is what the calibration RMSE measures (Section 3.1, 0.77 t/ha over thirteen observations). A good RMSE at this level guarantees nothing about the following two levels, a lesson already drawn from defect 6 documented in Section 3.4, where two calibrations with nearly identical RMSE produced qualitatively opposite policies.

Level 2, policy validation within the simulator: does the learned policy outper-

form reference policies, within the same simulator (or an independent simulator)? This is what **Table 3** (comparison to the four baselines on Mitscherlich) and Section 4.7 (cross-validation on QUEFTS, APSIM, and WOFOST) measure. This level validates the relative quality of the decision, not the absolute accuracy of the predicted yield.

Level 3, real agronomic validation: does the policy produce, on an actual plot, a gain actually measured in the field? None of the results presented in this article reach this level. The thirteen field observations (Section 4.2) are used to calibrate the simulators, not to validate the policy itself on decisions it would have recommended prospectively. This third level remains, at this stage, a future direction (Section 5.2), not an achieved result.

These three levels are independent: a well-calibrated simulator (level 1) can carry a policy that generalizes poorly (level 2, cf. the corrected zone on QUEFTS before correction, Section 4.7); a policy that is statistically superior across several simulators (level 2) remains, as long as it has not been tested in the field, a strong hypothesis rather than an agronomic conclusion (level 3).

4.8. Uncertainty and Sensitivity of the Simulator Calibration

Three coefficients (c_N, c_P, c_K) were adjusted so that the simulator fits the thirteen field observations as closely as possible. Two checks help determine whether these values can really be trusted, beyond the simple fact that they fit these thirteen specific cases well.

First test: what happens if the computer is left free to choose all three coefficients together?

The method is as follows. One observation is hidden, the computer is left to find the best values of c_N, c_P, and c_K on the remaining twelve, and we then check whether it correctly guesses the hidden observation. This is repeated twelve times, hiding a different observation each time.

We find that, every time, the computer converges back toward $c_N \approx 0.05$ never toward the 0.008 actually used in the final model. Yet we already know (Section 3.4) that $c_N = 0.05$ makes urea completely useless, regardless of the soil tested an absurd behavior. This test confirms that this is not an isolated accident: left to itself, the computer systematically falls back into this trap.

In short, we do not use $c_N = 0.008$ because the data demand it. We use it because we verified by hand that this choice produces agronomically sensible behavior, deliberately setting aside what pure optimization would have proposed.

Second test: once c_N is fixed at its true value, can c_P and c_K be trusted?

Using the same principle (hide one observation, fit on the other eleven, check against the hidden one, repeat twelve times), but this time with c_N locked at 0.008 and only c_P and c_K left free, we find that the resulting average error (0.77 t/ha) is almost identical to the one reported for the full model. This is a good sign, confirming that this figure is not a lucky coincidence.

But the two coefficients do not behave the same way: c_P stays stable from one

trial to the next (variation of about 10%). c_K, on the other hand, varies much more (variation of 45%) and this instability comes almost entirely from a single observation (a soil at pH = 6.2, with NPK and urea). Removing it sends c_K soaring to its maximum allowed value. This single observation therefore carries, on its own, almost all the available information about potassium.

In conclusion, c_P can be trusted. For c_K, however, caution is warranted not because it would be wrong, but because thirteen observations simply are not enough to pin it down with greater precision. That is enough to guide a dose (more or less fertilizer), but not enough to claim to know its true physiological value.

4.9. Comparison against a Direct Numerical Optimization (Non-RL Method)

Beyond the comparison against the four reference policies (Table 3) and the qualitative comparison against the theoretical optimum by bin (Section 4.4), a direct quantitative comparison situates the PPO policy against a non-RL optimization method: the exact, soil-by-soil search for the sequence of actions maximizing reward, via direct numerical optimization (differential_evolution) on the reward function itself, with no generalizable policy, no neural network, and no training cost.

Across 150 randomly drawn soils, independent of training, the PPO policy (ensemble mean across the 5 seeds) achieves a mean reward of 0.927 (standard deviation 0.138), against 1.074 (standard deviation 0.136) for the exact optimizer querying the simulator directly on each soil. The gap is clear and statistically significant (paired t-test, $t = 22.7$, $p = 2.7 \times 10^{-50}$; Cohen's $d = 1.07$, large effect), with a mean relative efficiency of 86.3% (median 87.0%): on average, PPO captures about 86% of the theoretically attainable reward on each soil tested.

Interpretation. This gap is not a sign of failed learning, but illustrates an expected structural difference between the two approaches. The oracle has an advantage that PPO never has: direct, full access to both the reward function and the simulator for each individually tested soil, with a computational budget dedicated to that single decision (a complete numerical search at every evaluation). PPO, by contrast, must produce an instantaneous decision, with no access to the simulator at decision time, drawn from a generalized policy learned once over a finite training budget (400,000 steps, Section 3.4) and valid across the entire pedological domain, rather than recomputed for each case. The cost of this generalization on about 14 points of relative efficiency on average here, is the expected counterpart of PPO's ability to produce an immediate recommendation, with no recomputation, even on a soil it never encountered during training, a property the oracle does not possess by construction: it must be run in full for every new soil.

Scope of this comparison. The oracle serves here as a theoretical upper bound, not a deployable competing method: its computation time per decision (a full numerical search, on the order of one to a few seconds per soil in this study) is incompatible with real-time operational use by an agricultural advisor, unlike the near-instantaneous inference of the already-trained PPO policy. This comparison

therefore quantifies an attainable performance ceiling, useful for situating the theoretical room for improvement of the learned policy, rather than a practical alternative to it.

4.10. Sensitivity Analysis of the pH Threshold and the Simulator's Structural Assumptions

We consider it useful to analyze the sensitivity of two factors: the pH threshold and the efficiency factor for nitrogen supplied by urea (EFFICIENCE_N_P2).

The pH = 5.5 threshold structures the entire architecture (Section 3.2) as an unlearned rule, imposed a priori following IRAD's recommendation. A sensitivity analysis checks whether this choice is arbitrary or reflects a genuine economic tipping point within the simulator itself.

A direct search for the crossover point, without any zone constraint, was carried out. Querying the full reward function (exact optimization, with no zone filter imposed) over a pH grid ranging from 4.3 to 6.7, with organic matter and nutrient contents fixed at representative median values, no crossover appears. Compost dominates the NPK/urea strategy across the entire range tested, including at pH = 6.7. An inspection of the reward mechanism reveals the cause. The restoration bonus (Section 3.3) depends solely on the organic-matter deficit relative to a restoration target (4.0%), never on pH. As long as organic matter remains below this target, regardless of pH, compost remains favored by this bonus, independently of its relative agronomic effectiveness at that particular point on the pH scale.

Temporarily removing this bonus to isolate the effect of simulated yield and input cost alone, a clear crossover appears, situated very precisely at the retained threshold. Compost remains the best strategy up to pH ≈ 5.0, after which NPK and urea become strictly superior from pH = 5.5 onward. At pH = 4.5, the reward reaches 0.926 for compost versus 0.681 for NPK + urea; at pH = 5.0, 1.060 versus 0.911; at pH = 5.5, the ranking reverses, with 1.066 for compost versus 1.078 for NPK + urea, a gap that holds at pH = 6.0 (1.066 versus 1.078) and pH = 6.5 (1.066 versus 1.078). This result confirms that the choice of the 5.5 threshold rests on a genuine tipping point in the simulated agronomic response itself, not on an arbitrary convention borrowed as-is from the literature without verification.

The direct methodological implication is that this result justifies, after the fact and more rigorously than a mere reference to the pedological literature, imposing the pH = 5.5 threshold as a structural rule rather than as a parameter left to learning. This confirms that the present sensitivity analysis recovers, through an independent route, a phenomenon already identified empirically during development.

The efficiency factor for nitrogen supplied by urea, fixed at 1.5, states that nitrogen from urea, applied at the V6 stage, is absorbed 1.5 times more effectively than nitrogen applied earlier. This value comes from a recognized scientific reference (Uhart & Andrade, 1995), not from an arbitrary choice.

Like any value drawn from the literature, it remains uncertain. We therefore tested what would happen if the true value were actually lower (1.2) or higher

(1.8), a margin of ±20% around 1.5.

For each of these three values, the optimal urea dose was recalculated on twelve different soils, varying only in their nitrogen content.

The result reveals a slight variation in dose. On average, it shifts by 11.5 kg/ha when using 1.2 instead of 1.5, and by only 5.8 kg/ha when using 1.8. In every case tested, the same logic holds: the more nitrogen available in the soil, the less urea is recommended.

4.11. Distinguishing the Simulator's Predictive Fit from the Behavioral Validity of the Policy

A good numerical fit (a good RMSE, for instance) never guarantees that the resulting policy behaves sensibly. Behavior must always be checked separately, and the numerical figure alone should never be relied upon. Four examples, encountered over the course of this work, illustrate this. First, two Mitscherlich calibrations had an almost identical score, yet one rendered urea completely useless while the other did not. Next, a QUEFTS calibration displayed an excellent score while canceling out urea's effect on almost every soil tested. Then, a free recalibration of the c_N value systematically fell back into the same trap as the first example, this time repeatedly rather than as an isolated occurrence. Finally, the restoration bonus, without any zone constraint, produced a mathematically correct but agronomically absurd optimum (compost recommended everywhere, even where it makes no sense).

In all four cases, only a check of the (agronomic, in our case) behavior, independent of the numerical score, made it possible to detect the problem. This work therefore recommends that such a check become a systematic step whenever a simulator is used to train a decision-making policy.

4.12. Contributions

The results of this study bring six main contributions to the literature on RL applied to precision agriculture in the African context.

This study contributes several elements to the literature on RL applied to fertilization in Africa. The hybrid rule-learning architecture shows that embedding a known agronomic threshold directly into the decision structure stabilizes learning on fine dosing, rather than leaving the agent to rediscover an already documented boundary. This principle is transferable to other thresholds used in agriculture, such as texture or phenological stage.

Documenting the six implementation flaws encountered constitutes a second contribution. Reference works in agricultural RL, such as Saikai *et al.* (2023) or Wu *et al.* (2022), do not describe this sensitivity to state and action scale. Yet on soils where kg/ha, pH units, and OM percentages coexist, this heterogeneity completely blocks exploration unless corrected by RescaleAction and VecNormalize.

Third, the stability of SHAP and LIME attribution across five seeds makes it possible to distinguish numerical noise from agronomic substance. The domi-

nance of pH for compost and of nitrogen for urea is confirmed on 5 out of 5 seeds with both methods. For NPK, the potassium-phosphorus competition remains sensitive to seed and method, which reflects well the joint fixation of these two elements on ferralsol. This practice, still uncommon in the interpretability literature applied to agriculture, makes it possible to distinguish superficial numerical noise from genuine instability in learning, reinforcing the scientific credibility of the resulting policy despite the residual variance observed in its fine tuning.

Fourth, the deployment architecture (Section 3.8) is itself a practical contribution rarely taken this far in the RL-agriculture literature. It comprises a structured Excel workbook, a data entry quality control layer, a dual source weather filter (API or stated observation, with automatic fallback in case of internet network failure), and a dual economic/restoration recommendation rather than a single imposed decision. This last choice deserves emphasis because it explicitly respects the farmer's decision-making autonomy. This dashboard makes visible the figures needed to arbitrate between immediate profitability and investment in long term soil fertility, rather than hiding the outcome behind a policy that would optimize a single implicit criterion.

Fifth, this work illustrates a methodological lesson transferable beyond this case study. Several policy outputs initially suspected of being faulty, such as a zero NPK dose on soil already rich in phosphorus, or the absence of a compost supplement at the grain filling phase, turned out, after systematic verification against an independent theoretical optimum and agronomic reasoning (compost mineralization time, diminishing returns), to be economically or agronomically correct corner solutions rather than learning flaws. This distinction, costly to establish but necessary, argues for systematic behavioral verification of any policy output deemed surprising, rather than a reflexive correction, before concluding that the model is malfunctioning.

Sixth, cross-validation across three simulators independent of the training process (QUEFTS, APSIM Next Generation, WOFOST) constitutes, to our knowledge, an approach still uncommon in the RL-for-agriculture literature, where the robustness of a policy is typically assessed on the single simulator used for training, at most supplemented by a statistical cross-validation on that same dataset. These three simulators rest on distinct mathematical formalisms. We note a law-of-the-minimum for QUEFTS, a first-principles day-by-day dynamic for APSIM and WOFOST, against the multiplicative production function of the training simulator. This makes it possible to test whether the superiority of the learned policy stems from a particular reward structure or holds up against genuinely independent modeling assumptions. The result obtained (PPO significantly outperforms the baselines that actually apply fertilizer in nearly all configurations tested, across all three simulators) thus strengthens the policy's credibility in a way that conventional cross-validation, internal to a single simulator, cannot achieve.

4.13. Discussion

These results fit within a still-limited body of literature on RL applied to fertilization in tropical contexts, and deserve to be confronted with its limitations as much as with its contributions. The most methodologically closely related works, namely Saikai *et al.* (2023), Wu *et al.* (2022), as well as Kallenberg *et al.* (2023) and Turchetta *et al.* (2022) discussed in the related-work section, share a common characteristic. None addresses acidic tropical soils, nor the phosphorus-fixation constraint that structures agronomic decision-making on Ferralsol. This absence is not merely a gap in geographic scope; it means that none of these works had to confront the problem of scale heterogeneity among state variables (pH, organic matter, kg/ha) documented in Section 4.6, nor the need for a hybrid rule/learning architecture to stabilize exploration. Our result on this point therefore does not contradict these works, but rather sheds light on a class of problems they simply did not have to solve.

On interpretability, however, our approach departs more clearly from common practice. The interpretability literature applied to agriculture most often relies on a single method (SHAP or LIME), on a single training seed, which does not allow a computational artifact to be distinguished from genuine agronomic grounding. The cross-method stability observed here (pH dominance for compost, nitrogen dominance for urea, across five seeds and two methods) strengthens the policy's credibility. The residual sensitivity observed on the potassium-phosphorus tradeoff for NPK, however, is a reminder that this stability is not guaranteed everywhere, and must be verified output by output, not assumed globally.

This work nonetheless carries limitations that should be explicitly acknowledged rather than left implicit. The calibration of the training simulator rests on only thirteen field observations. This restricted sample constrains the precision of several coefficients (Section 4.2) and, by construction, did not allow for genuine field level agronomic validation. The four simulators mobilized in this work, including those used for cross-validation, remain parametric models fitted on this same dataset, not independent trials. The evaluation also remains single season. The model does not capture the soil restoration value accrued over several years. This constitutes a limitation that the economic bonus mechanism (Section 3.3) attempts to compensate for without fully succeeding, as shown by its own side effect documented in Section 4.4.

These limitations do not invalidate the scope of the work, but they do specify its conditions of validity. The results shown here establish that an adaptive fertilization policy can be learned, interpreted, and rigorously verified in simulation. This is, moreover, a necessary but not sufficient precondition for real agronomic deployment. Moving from this multi-simulator validation to a field validation, across multiple seasons and multiple sites, constitutes the next logical and most demanding step of this research program.

5. Conclusion

Finally, we tested the model in a simple Excel tool. Extension workers enter soil

data from a portable tester. The tool checks the data, looks at the weather forecast, and gives two options: the cheapest dose for this season, or a higher compost dose that rebuilds soil fertility. We tried it with 12 plots around Wakwa. It works without internet if needed.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Keegoui, G., Messomo, X.Z., Egoume Guissana, A. and Fogang, J.A. (2025) Fertilisation du maïs (*Zea mays* L.) à base des recommandations de l'Intelligence Artificielle sur un sol férrallitique du Centre Cameroun. *International Journal of Biological and Chemical Sciences*, **19**, 277-288. https://doi.org/10.4314/ijbcs.v19i1.21

[2] Hinsinger, P., Ndour, Y., Becquer, T., Chapuis-Lardy, L. and Masse, D. (2017) Chapitre 26. Les enjeux liés au phosphore dans les sols tropicaux. In: Roose, E., Ed., *Restauration de la productivité des sols tropicaux et méditerranéens*, IRD Éditions, 329-341. https://doi.org/10.4000/books.irdeditions.24333

[3] Owino, C.O., Owuor, P.O. and Sigunga, D.O. (2015) Elucidating the Causes of Low Phosphorus Levels in Ferralsols of Siaya County, Western Kenya. *Journal of Soil Science and Environmental Management*, **6**, 260-267.

[4] Truong, B. (1996) Phosphorus Dynamics in Tropical Soils of Humid Zones, Consequences for Plant Nutrition. Fertilite du milieu et strategies paysannes sous les tropiques humides, CIRAD-SAR/CIRAD-FLHOR-DIR.

[5] Tabi, F.O., Bitondo, D., Yinda, G.S., Kengmegne, S.S.A. and Ngoucheme, M. (2013) Effect of Long-Term Integrated Soil Fertility Management by Local Farmers on Nutrient Status of a Typic Dystrandept under Potato-Based Cropping System. *International Research Journal of Agricultural Science and Soil Science*, **3**, 134-140.

[6] Carsky, R., Jagtap, S.S., Tian, G., Sanginga, N. and Vanlauwe, B. (1998) Maintenance of Soil Organic Matter and N Supply in the Moist Savanna Zone of West Africa. In: Lal, R., Ed., *Soil Quality and Agricultural Sustainability*, Ann Arbor Press, 223-236.

[7] Sutton, R.S. and Barto, A.G. (2018) Reinforcement Learning: An Introduction. 2nd Edition, MIT Press.

[8] Schulman, J., Wolski, F., Dhariwal, P., Radford, A. and Klimov, O. (2017) Proximal Policy Optimization Algorithms. arXiv: 1707.06347.

[9] Saikai, Y., Peake, A. and Chenu, K. (2023) Deep Reinforcement Learning for Irrigation Scheduling Using High-Dimensional Sensor Feedback. *PLOS Water*, **2**, e0000169. https://doi.org/10.1371/journal.pwat.0000169

[10] Wu, J., Tao, R., Zhao, P., Martin, N.F. and Hovakimyan, N. (2022) Optimizing Nitrogen Management with Deep Reinforcement Learning and Crop Simulations. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 19-20 June 2022. https://doi.org/10.1109/cvprw56347.2022.00178

[11] Yassine, H.M., Roufaida, K., Shkodyrev, V.P., Abdelhak, M., Zarour, L. and Khaled, R. (2022). Intelligent Farm Based on Deep Reinforcement Learning for Optimal Control. *International Symposium on iNnovative Informatics of Biskra* (*ISNIB*), Biskra, 7-8 December 2022. https://doi.org/10.1109/isnib57382.2022.10076088

[12] Binas, J., Luginbuehl, L.H. and Bengio, Y. (2019) Reinforcement Learning for Sustainable Agriculture. *Proceedings of the ICML* 2019 *Workshop on Climate Change*: *How Can AI Help*? (*CCAI*), 36*th International Conference on Machine Learning*, Long Beach. https://www.climatechange.ai/papers/icml2019/24

[13] Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A. and Kabandana, I. (2023) Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices. *Agriculture*, **13**, Article 2141. https://doi.org/10.3390/agriculture13112141

[14] Lontsi Saadio, C., Adoni, W.Y.H., Aworka, R., Zoueu, J.T., Kalala Mutombo, F. and Kimpolo, C.L.M. (2022) Crops Yield Prediction Based on Machine Learning Models: Case of West African Countries. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4003105

[15] Bouni, M., Hssina, B., Douzi, K. and Douzi, S. (2022) Towards an Efficient Recommender Systems in Smart Agriculture: A Deep Reinforcement Learning Approach. *Procedia Computer Science*, **203**, 825-830. https://doi.org/10.1016/j.procs.2022.07.124

[16] Janssen, B.H., Guiking, F.C.T., van der Eijk, D., Smaling, E.M.A., Wolf, J. and van Reuler, H. (1990) A System for Quantitative Evaluation of the Fertility of Tropical Soils (QUEFTS). *Geoderma*, **46**, 299-318. https://doi.org/10.1016/0016-7061(90)90021-z

[17] Kallenberg, M.G.J., Overweg, H., van Bree, R. and Athanasiadis, I.N. (2023) Nitrogen Management with Reinforcement Learning and Crop Growth Models. *Environmental Data Science*, **2**, e34. https://doi.org/10.1017/eds.2023.28

[18] Turchetta, M., Corinzia, L., Sussex, S., Burton, A., Herrera, J., Athanasiadis, I., *et al.* (2022) Learning Long-Term Crop Management Strategies with CyclesGym. *Advances in Neural Information Processing Systems* 35, New Orleans, 28 November-9 December 2022, 1396-11409. https://doi.org/10.52202/068431-0828

[19] Kelly, T.D., Foster, T. and Schultz, D.M. (2024) Assessing the Value of Deep Reinforcement Learning for Irrigation Scheduling. *Smart Agricultural Technology*, **7**, Article ID: 100403. https://doi.org/10.1016/j.atech.2024.100403

[20] Tanzouak, J., Yenke, B.O., Bame, N. and Ndoundam, R. (2019) An Efficient Data Analysis Based Flood Forecasting System (EDAFFS). *International Journal of Information Technology and Computer Science*, **11**, 38-47. https://doi.org/10.5815/ijitcs.2019.02.05