The New Agronomists: Language Models are Experts in Crop Management

Jing Wu [*] [†] Zhixin Lai [*] [‡] Suiyao Chen [*] [§] Ran Tao [†] Pan Zhao [¶]

Naira Hovakimyan [†]

Abstract

Crop management plays a crucial role in determining crop yield, economic profitability, and environmental sustainability. Despite the availability of management guidelines, optimizing these practices remains a complex and multifaceted challenge. In response, previous studies have explored using reinforcement learning with crop simulators, typically employing simple neural-network-based reinforcement learning (RL) agents. Building on this foundation, this paper introduces a more advanced intelligent crop management system. This system uniquely combines RL, a language model (LM), and crop simulations facilitated by the Decision Support System for Agrotechnology Transfer (DSSAT). We utilize deep RL, specifically a deep Q-network, to train management policies that process numerous state variables from the simulator as observations. A novel aspect of our approach is the conversion of these state variables into more informative language, facilitating the language model's capacity to understand states and explore optimal management practices. The empirical results reveal that the LM exhibits superior learning capabilities. Through simulation experiments with maize crops in Florida (US) and Zaragoza (Spain), the LM not only achieves state-of-the-art performance under various evaluation metrics but also demonstrates a remarkable improvement of over 49% in economic profit, coupled with reduced environmental impact when compared to baseline methods. Our code is available at `https://github.com/jingwu6/LM_AG`.

1. Introduction

In today's agricultural landscape, addressing food security and sustainable farming practices is crucial, aligning with the United Nations' goal of Zero Hunger. The challenge of boosting food production for a global population expected to reach 9.6 billion by 2050, while minimizing negative environmental impacts like ecosystem degradation and greenhouse gas emissions, is paramount [43]. Key factors in crop management, particularly fertilization with nitrogen (N) and irrigation with water (W), significantly affect crop yields and environmental health [40, 49]. However, the previous best practices for these management aspects, derived from empirical experience and academic research [45, 51], face uncertainty in their effectiveness against changing climate and market conditions. Therefore, the adequacy of current strategies is questionable, highlighting a need for innovative, efficient, and adaptable management systems. These systems should be capable of devising optimal strategies suitable for varying conditions and objectives, such as maximizing economic profit or service utilization [2, 9, 10]. This research is anchored in this context, leveraging advanced AI methods to improve agricultural practices and tackle these critical challenges.

Reinforcement learning (RL) has shown exceptional capabilities in tasks that involve sequential decision-making (SDM), such as in robotics and gaming [11, 12, 24, 34]. This success suggests a significant potential for RL in optimizing crop management, which at its core is an SDM problem. Given the need for numerous interactions between the RL agent and the environment during policy training, field trial-based methods are impractical. Consequently, the use of crop models to simulate both the crop and its environment, providing a platform for interaction with the RL agent, appears to be the most feasible approach [5, 36].

Recently, the authors of [47, 52] proposed to train management policies for crop management using deep RL with DSSAT [21] and Gym-DSSAT [41], one of the most widely used crop models in the world. Their trained policies, both under full and partial observations, outperformed baseline policies by achieving higher yields or similar yields with reduced nitrogen (N) fertilizer input. However, there are limitations to these approaches. Firstly, the models primarily employed Multilayer Perceptrons (MLPs), which, while effective, have limited fitting power compared to more complex architectures. This limitation could potentially con-

[*]These authors contributed equally to this work.
[†]University of Illinois at Urbana-Champaign
[‡]Cornell University
[§]University of South Florida
[¶]University of Alabama

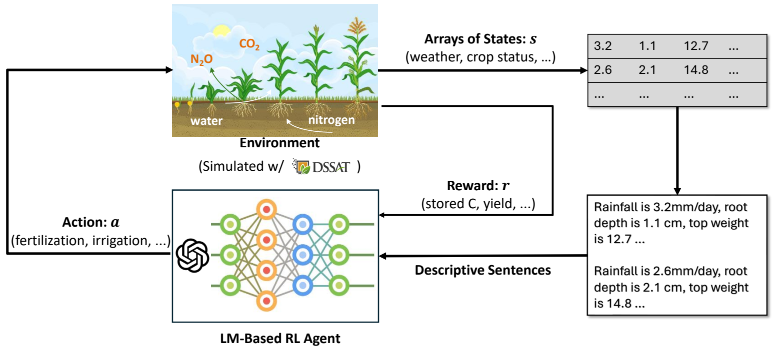

Figure 1. Framework and pipeline of the intelligent crop management system using LM-based RL

strain the models' ability to capture the intricate dynamics of crop growth and management fully. Secondly, the reliance on MLPs limits the incorporation of additional descriptive features for state representations in the model. These features could include various environmental, soil, and crop growth parameters that are crucial for precise agricultural decision-making. This gap in the model's design raises a critical question: Can language models (LMs) serve as viable alternatives for RL agents in these crop management tasks? The limitations of the existing models and this pivotal question motivate the present paper.

In this paper, we present an intelligent crop management framework, depicted in Figure 1, that incorporates a powerful LM, and crop simulations via DSSAT and Gym-DSSAT. Concretely, we transform the states from simulation tools, typically arrays of numbers, into more descriptive sentences. This conversion enables a significant shift in our approach: we replace the traditional MLP-based RL agent with an LM-based RL agent. This new agent leverages LMs to encode these descriptive state sentences into embeddings, thereby capturing a more informative and nuanced understanding of the states. Meanwhile, we notice that LMs have shown distinctive cognitive capabilities, which include advanced thinking [50], robust memory functions [37], reflective skills [27, 44] and reasoning [53]. As a result, the RL agent should be equipped with the ability to comprehend complex aspects of crop growth and simulation environments. We, therefore, anticipate that the incorporation of LMs will markedly improve the performance of the RL agent in crop management tasks.

To demonstrate the effectiveness of the proposed method, we conducted case studies simulating maize crops management in Florida, USA, and Zaragoza, Spain. This choice of locations aligns with the settings used in previous studies [47]. In both scenarios, the policies trained by our framework exhibited superior performance compared to the previous state-of-the-art approaches; the baseline was derived from either maize production guidelines recommended by agricultural experts as well as survey results on actual management practices of maize farmers. Additionally, continuing in the vein of established research, our investigation also includes the training of RL-based policies with well-recognized reward functions [47]. These functions are designed to represent different balances among key factors: crop yield, resource utilization, and environmental impact, particularly focusing on nitrate leaching during the crop growth cycle. In summary, the primary contributions of our work can be delineated as follows:

- We investigate a critical yet under-explored question: Can LMs serve as better alternatives for RL agents in crop management tasks to offer more nuanced and effective solutions and advance the state of intelligent crop management systems?
- To the best of our knowledge, this work marks the first attempt to integrate descriptive language to represent agricultural states and to employ LMs in the pursuit of optimal crop management policies.
- We empirically demonstrate that our proposed framework exceeds the performance of existing state-of-the-art approaches in various key aspects, including crop yield, resource utilization, and environmental impact.

2. Related Work

2.1. Crop Management with RL

Considering crop management as a Markov decision process (MDP), initial efforts in applying reinforcement learning (RL) to derive optimal management strategies from simulators have emerged, but this field is still developing. An early attempt to use a basic RL approach for wheat management in France was documented by [16]. Another study by [46] focused on optimizing irrigation for maize in Texas, US. However, these studies had limitations, such as narrow state and action spaces. For example, [46] included only a single state variable for RL training. In [3], the researchers applied the proximal policy optimization (PPO) algorithm for fertilizer and irrigation policy optimization, but their results did not significantly surpass existing baseline methods in simulations. More comprehensive research in this area is represented by [52], which examined N fertilization for maize in Florida and Iowa. Subsequent studies have expanded on this approach using different crop models [23, 30]. Addressing the challenges of partially observed crop management, authors of [47] explored the use of imitation learning, training an RL agent with a broad set of state features and then applying it to a subset of these features.

2.2. Crop Models for RL

The necessity for crop models arises from the practical challenges of conducting real-world farming experiments, which are often laborious, time-consuming, and expensive. These models are crucial for assessing the impact of climate change and various management practices on crop production [56]. Among the numerous crop simulation models developed, APSIM and DSSAT are particularly notable for their widespread use and continuous updates, providing accurate estimations of crop production in relation to multiple factors [4, 6, 22, 47, 52]. Traditional crop models, however, typically require the pre-definition of management practices before simulations, a limitation when compared to the dynamic decision-making capability of reinforcement learning (RL). To bridge this gap, efforts have been made to integrate RL with crop models, enabling real-time decision-making during simulations. Innovations like the CropGym environment and interfaces based on the SIMPLE crop model for russet potatoes, both utilizing the Open AI Gym framework [7, 35, 56], demonstrate the feasibility of RL in crop management. Despite this progress, some of these models oversimplify essential crop and environmental details. In contrast, Gym-DSSAT, built on the robust DSSAT model, allows for detailed, daily interactions between the RL agent and the simulated environment, a significant advancement in optimizing nitrogen and irrigation management [41, 47, 52].

2.3. LM for Decision-making

In recent years, there has been a surge in studies utilizing pre-trained LMs as decision-making agents. These models' remarkable capabilities have been harnessed across various domains, generating improved control plans for diverse robots and agents [1, 8, 19, 20, 28, 29, 33, 39]. Notably, researchers of [25] developed LM-based agents for user interface (UI) interactions, while ReAct [54] integrated action decisions with natural language reasoning, demonstrating promising results.

To the best of our knowledge, our work represents the first endeavor to leverage the LMs in formulating optimal management strategies for crop models in agriculture.

3. Methods

3.1. Problem Formulation

In this study, we approach nitrogen fertilization and irrigation management as a finite Markov Decision Process (MDP), following the paradigm of previous work [47, 52]. Each day, denoted as day t, involves the agent receiving the environmental state, s_t, and subsequently selecting an action a_t from the action space $\mathcal{A}$. This selection is guided by a policy $\pi(s_t, \theta_t)$, where θ_t symbolizes the policy parameters on that particular day, and notably, the policy in this context is a pretrained language model. The state s_t encompasses vital data pertaining to weather, plant growth, and soil conditions, as simulated for that day. The action a_t is composed of two key decisions: the quantity of nitrogen fertilizer, denoted as N_t, and the amount of irrigation water, W_t, to be applied. The effectiveness of these decisions is quantified by the reward $r_t(s_t, a_t)$, which is calculated based on the outcomes of s_t and a_t, defined as:

$$r_t(s_t, a_t) = \begin{cases} w_1 Y - w_2 N_t - w_3 W_t - w_4 N_{l,t} & \text{if harvest at } t, \\ -w_2 N_t - w_3 W_t - w_4 N_{l,t} & \text{otherwise,} \end{cases} \tag{1}$$

where $w_1, w_2, w_3, w_4, Y, N_{l,t}$ denote four custom weight factors, yield at harvest and the amount of nitrate leaching on a given day, respectively. Both Y and $N_{l,t}$ are derived from the state variable s_t. The design of the reward function, characterized by the weights w_1, w_2, w_3, w_4, is pivotal in steering the agent's strategy. The agent's objective is to identify the optimal policy $\pi(s_t, \theta_t)$ that selects action a_t to maximize the total future discounted return. This return, defined as $R_t = \sum_{\tau=t}^{T} \gamma^{\tau-t} r_\tau$, captures the accumulated reward from the current action a_t to the future rewards, discounted by factor γ.

3.2. LM-based RL Agent

To harness the full potential of language models (LM) in comprehending crop models and identifying optimal management strategies, we made adaptations to the state variables from the simulation tool, specifically Gym-DSSAT. Traditionally, the state in such simulations is represented by an array of variables reflecting various crop and environmental conditions, like rainfall and root depth. However, this format does not provide a direct correlation between the variables and their descriptive meanings, posing a challenge for RL agents to interpret each variable independently. To overcome this, we transformed the raw data into a more language-friendly format. Each variable name and its corresponding value were combined into coherent sentences. This approach essentially transforms the state data into a format that is more accessible and interpretable by LMs, allowing for a more intuitive and efficient exploration of management practices.

In our approach, we have innovated by substituting the traditional MLPs with a distilled and pre-trained BERT model from [42] serving as the RL agent. This advanced model is utilized to encode the concatenated sentences, which represent the state variables, into feature embeddings. Following this encoding process, we introduce a few fully connected layers connected to the distilled BERT encoder. These layers are responsible for transforming the generated feature embeddings into a format that aligns with the action space of the RL agent. This novel architecture not only leverages the linguistic understanding of BERT but also ensures that the complex relationships within the crop management data are effectively captured and translated into actionable insights.

3.3. Policy Training with LM

In this study, we use the Deep Q-Network (DQN) from [34] to train our agent. The goal is to learn an optimal policy that maximizes the future discounted return, denoted as R_t. A novel aspect of our approach is the integration of the distilled BERT model to represent the action-value function, also known as the Q function, within the DQN framework. This Q function, formally defined as $Q^{\pi}(s,a) = \mathbb{E}[R_t|s_t = s, a_t = a, \pi]$, is essential for calculating the expected future discounted return from state s when action a is taken, following policy π.

The objective is to refine the parameters of the Q-network to pinpoint the optimal Q function, $Q^{\star}(s,a)$, which indicates the highest return possible from state s by taking action a and adhering to the optimal policy. For selecting the optimal action in state s_t, we employ a greedy policy defined as $a_t^{\star} = \max_{a \in \mathcal{A}} Q^{\star}(s_t, a)$. Training the Q-network, which effectively means training the policy, involves minimizing the following loss function:

$$L_i(\theta_i) \triangleq \underset{(s,a,r,s')}{\mathbb{E}} \left[r + \gamma \max_{a' \in \mathcal{A}} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i) \right]. \tag{2}$$

Here, s, a, r, s' denote the state, action, reward, and next state, respectively, with γ representing the discount factor, and θ_i^- representing the parameters of a target network defined earlier. The tuples (s, a, r, s') for the loss function are randomly sampled from the replay buffer, a collection of prior state-action-reward-next state tuples accumulated during training.

3.4. Crop Simulations with Gym-DSSAT

Similar to [47, 52], we leverage Gym-DSSAT [41], a Gym interface for DSSAT that enables the agent to interact with the simulated environment (i.e., reading the weather, soil, and crop information and applying management practices) on a daily basis. For more details about DSSAT and Gym-DSSAT, readers can refer to Section 2.2.

4. Experiments and Results

In this section, various experiments are conducted on real-world datasets to demonstrate the effectiveness and superiority of the proposed framework. The experiment settings are introduced in Section 4.1, where the setup and techniques used in the experiments are detailed. Following this, the training and evaluation details are illustrated in Section 4.2, providing the necessary details to reproduce the work of the paper. Then, we present the evaluation results, where the performance of our proposed method is compared against existing baselines and SoTA approaches in Section 4.3. Lastly, ablation studies are conducted for policy training in Section 4.4.

4.1. Experimental Setup

The experiments focusing on training policies for nitrogen and irrigation management in maize crops were conducted through two distinct case studies, both utilizing real-world data. The first of these case studies was set in a simulated environment replicating Florida, USA, in 1982, while the second case study was based on the simulated conditions of Zaragoza, Spain, in 1995. The primary objective of these case studies was to test and demonstrate the viability and advantages of the proposed framework, rather than preparing it for immediate real-world application. For those interested in the specifics of deploying this framework in practical settings, further details are provided in Section 5.

For each case study, DQN was used to train the LM-based RL agent under full observation. The performance of all trained policies was evaluated in simulation, and compared with baseline policies and previous state-of-the-art methods as mentioned in [47]. The baseline for the Florida study was based on a maize production guide for Florida farmers [51], and for the Zaragoza study it was derived from survey data on maize farming practices in Zaragoza [31, 45].

The framework was implemented to train the RL agent under full observation. This approach involved testing with five different reward functions, each designed to demonstrate the adaptability of the framework to various trade-offs. These trade-offs include balancing crop yield, N fertilizer use, irrigation water use, and environmental impact. This variety in reward functions allows the framework to be evaluated across a range of scenarios and objectives, showcasing its flexibility in addressing different agricultural management priorities.

4.2. Implementation Details and Evaluation Metrics

Implementation Details. The RL agent in our study employs a combination of DistilBERT and a three-layer fully connected neural network for feature adaptation. The process begins with DistilBERT encoding the state inputs into

	w_1 (Y)	w_2 (N_t)	w_3 (W_t)	w_4 $(N_{l,t})$	Note
RF 1	0.158	0.79	1.1	0	Economic profit
RF 2	0.158	0.79	0	0	Free water
RF 3	0.158	0	1.1	0	Free N fertilizer
RF 4	0.158	1.58	1.1	0	Doubled N price
RF 5	0.2	1	1	5	With N Leaching

Table 1. Weights used in each reward function (RF) defined by equation (1)

768-dimensional embeddings. Notably, the parameters of DistilBERT are trained end-to-end in this model. After this initial encoding, the embeddings are passed through fully connected layers, one with 512 units and the other with 256 units. The final layer in this sequence is responsible for mapping these processed embeddings to the action space, completing the flow from the input state to the actionable output in the RL framework. The discrete action space is defined as follows:

$$\mathcal{A} = \{40k\frac{\text{kg}}{\text{ha}} \text{ N fertilizer } \& \ 6k\frac{\text{L}}{m^2} \text{ Irrigation water}\}, \tag{3}$$

where $k = 0, 1, 2, 3, 4$, resulting in a total of 25 possible actions. This action space design incorporates standard quantities of N fertilizer and irrigation water that are typically applied by farmers in a single day. It also allows for a wide range of options, aiding the discovery of effective policies. The discount factor is meticulously set at 0.99. To facilitate the neural network's updates, Pytorch is employed alongside the Adam optimizer [26], characterized by an initial learning rate of 1e-5 and a batch size of 512. This setup is strategically chosen to optimize the learning process while ensuring efficient computation.

The direct application of DistilBERT's tokenizer to numerical values introduces significant training instability. Concretely, numerical values are often segmented into multiple tokens, resulting in considerable variance for small numerical differences. For instance, the number 360 tokenizes into [9475], while 361 splits into [4029, 2487], indicating a disproportionate representation of adjacent numbers. This inconsistency can amplify instability during training. Additionally, the tokenization of decimal numbers compounds this issue. For example, 0.1 translates into [1014, 1012, 1015], where '0' and the decimal point are tokenized separately, leading to unnecessary token proliferation and computational inefficiency.

To address the tokenization challenges with numerical values in our model, we have developed a straightforward yet effective data preprocessing technique. This method involves normalizing numerical values to fit within the range of [0, 300] and subsequently utilizing only the integer portion for tokenization. The decision to cap the range ensures that each normalized number corresponds to a single token, thereby simplifying and stabilizing the tokenization process. Additionally, focusing solely on the integer part helps to minimize the number of tokens used. We achieve a succinct representation comprising 27 distinct tokens, which includes 25 feature-specific tokens plus two special tokens ([CLS] and [SEP]). This streamlined token set not only improves the stability of the training process but also enhances its computational efficiency, which is crucial for the complex task of crop management optimization using RL and language models.

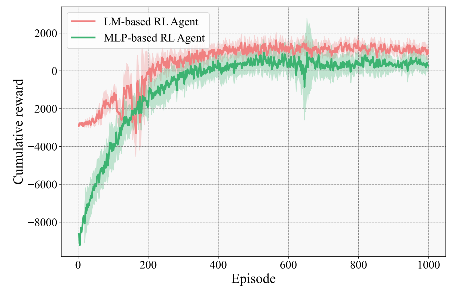

Figure 2. Cumulative reward versus episodes for policy training under RF1

Evaluation Metrics. In each case study, we employed reward functions in line with the approach described in previous research [47]. Specifically, five distinct reward functions for r_t derived from Equation (1) were utilized to train the RL agent. For each reward function, a single trained policy was selected for evaluation. The parameters for each reward function (RF) are detailed in Table 1.

RF1 quantifies the economic profit ($/ha) that farmers accrue, calculated based on the estimated prices of maize and the costs of N fertilizer and irrigation water, as referenced from [32] and [51]. RF2-RF4 represent variations of economic profit under different scenarios. Specifically, RF2 addresses the hypothetical situation where irrigation water is free of cost; RF3 considers the case where N fertilizer is free; and RF4 models a scenario in which the price of N fertilizer is doubled.

In contrast to RF1-RF4, which focus solely on economic profit, RF5 incorporates an additional environmental aspect, specifically nitrate leaching. Nitrate leaching is a significant environmental concern as it contributes to problems like eutrophication of water bodies and soil degradation [14]. RF5 is structured to balance yield, N fertilizer, and irrigation use while assigning a substantially higher weight to nitrate leaching. This approach aims to minimize nitrate leaching while still achieving favorable economic outcomes.

4.3. Results of Experiments

The evaluation outcomes for the trained policies in both the Florida and Zaragoza case studies are detailed in Table 2,

Florida Case	N Input (kg/ha) ↓	Irrigation (L/m^2) ↓	Yield (kg/ha) ↑	RF1 ↑	RF2 ↑	RF3 ↑	RF4 ↑	RF5 ↑
Empirical Baseline	360	394	10772	984	1417	1269	700	338
Policy1: Traditional Agent	200	**120**	10852	1425	1557	1538	1267	1673
Policy1: LM-based Agent (Ours)	**122**	192	**11402**	**1464**	**1675**	**1590**	**1337**	**1748**
Policy2: Traditional Agent	200	732	11244	813	1619	971	655	1020
Policy2: LM-based Agent (Ours)	**160**	**510**	**11474**	**1126**	**1687**	**1252**	**999**	**1330**
Policy3: Traditional Agent	19920	**108**	10865	-1.4e4	-1.4e4	1598	-3.0e4	-4.9e4
Policy3: LM-based Agent (Ours)	**10000**	264	**13152**	**-6.1e3**	**-5.8e3**	**1788**	**-1.4e4**	**-3.8e4**
Policy4: Traditional Agent	160	102	**10358**	1398	**1510**	1524	1272	1635
Policy4: LM-based Agent (Ours)	**160**	**36**	10192	**1428**	1468	**1555**	**1302**	**1647**
Policy5: Traditional Agent	200	138	10926	1417	1568	1575	1259	1651
Policy5: LM-based Agent (Ours)	**160**	**60**	**11280**	**1590**	**1656**	**1716**	**1463**	**1841**

Table 2. The evaluation results of our trained policies, comparing them with previous SoTA methods and baseline policies. 'Policy x' refers to the policy optimized using the reward function 'RF x'. The 'RF x' column details the cumulative rewards for each policy, calculated in accordance with 'RF x'. Details of each reward function can be found in Table 1. The best value is highlighted in **bold**.

Table 3, and Figure 2. It's important to note that these results may not entirely reflect the optimal potential of the policies due to the random initialization of the Q-network and its episodic updates. Additionally, further refinement through hyperparameter tuning might yield more competitive outcomes. However, such tuning was intentionally avoided in this study to maintain a focus on generalizability and fair evaluation. Despite these deliverable-introduced constraints, the chosen policies still illustrate the effectiveness of the LM-based RL agent in enhancing crop management strategies. These policies also effectively demonstrate how different RFs can influence training outcomes.

The evaluation results, as detailed in Table 2 and Table 3, indicate that the proposed LM-based RL agent outperforms previous SoTA and empirical baselines in most metrics and scenarios. Notably, the LM-based RL agent consistently utilizes lower amounts of nitrogen and generally requires less irrigation, yet it manages to secure higher yields. These improvements are consistent across various reward functions that prioritize different optimization objectives, underscoring the agent's adaptability and robustness in optimizing for diverse agricultural goals. The findings validate the previous hypothesis that language models have a heightened capacity to decipher complex crop management scenarios and simulate environments, ultimately leading to the discovery of more optimal management practices. Compared with the baseline policies, the RL-trained policies achieve a 49% and a 67% increase in terms of profit, i.e., RF1, and almost a 445% and a 37% increase in terms of RF5 for the Florida case and Zaragoza case, respectively. Notably, the enormous negative values of the cumulative rewards of Trained Policy 3 from both case studies are the results of the large amounts of N input, which are not punished during training with RF 3.

Consistent with prior studies [47], the choice of reward function significantly influences the strategy of policies trained with LM-based RL agents. For instance, when trained with RF2, which posits irrigation water as a free resource, Trained Policy 2 tends to maximize irrigation while minimizing nitrogen input. This approach leads to the highest yield and cumulative reward as per the criteria of RF2. In contrast, RF3 assumes zero cost for nitrogen fertilizer, prompting Trained Policy 3 to favor high nitrogen use and minimal irrigation in both case studies. Under RF4, which reflects a doubled cost of nitrogen fertilizer in comparison to RF1, Trained Policy 4 leads to a reduction in nitrogen use. Despite the reduced nitrogen input, this policy still achieves a substantial yield and notably saves over 64% of water resources, indicating the agent's capability to find a balance between cost efficiency and agricultural output.

In general, the results presented showcase the state-of-the-art capabilities of the LM-RL framework in optimizing crop management. This optimization is proven to be effective under various criteria, across different geographic locations, and within diverse environmental conditions. The framework's adaptability is highlighted by its ability to consistently apply LM-RL training to discover optimal management policies that align with specific targets, as dictated by the design of the chosen reward function. This flexibility and effectiveness affirm the potential of LM-RL as a powerful tool for agricultural management and decision-making.

4.4. Ablation Studies

4.4.1 Training Separately on Fertilization and Irrigation

In our previous research endeavors, we concurrently optimized N fertilization and irrigation practices, subsequently comparing these results against both established baseline practices and previous SoTAs. To further elucidate the efficacy of this joint optimization approach, this section in-

Zaragoza Case	N Input (kg/ha) ↓	Irrigation (L/m^2) ↓	Yield (kg/ha) ↑	RF1 ↑	RF2 ↑	RF3 ↑	RF4 ↑	RF5 ↑
Empirical Baseline	250	752	10990	712	1539	909	514	1176
Policy1: Traditional Agent	240	**330**	10477	1103	1466	1292	913	1525
Policy1: LM-based Agent (Ours)	**160**	354	**10806**	**1192**	**1581**	**1318**	**1065**	**1617**
Policy2: Traditional Agent	200	1068	10923	393	1568	551	235	888
Policy2: LM-based Agent (Ours)	**160**	**1032**	**10856**	**453**	**1588**	**580**	**327**	**964**
Policy3: Traditional Agent	10640	**324**	10626	-7083	-6727	1323	-1.5e4	-8839
Policy3: LM-based Agent (Ours)	**10000**	342	**10903**	**-6553**	**-6177**	**1347**	**-1.4e4**	**-8161**
Policy4: Traditional Agent	120	**336**	9601	1053	1422	1147	958	1464
Policy4: LM-based Agent (Ours)	**160**	348	**10250**	**1110**	**1493**	**1268**	**984**	**1542**
Policy5: Traditional Agent	200	390	10589	1086	1515	1244	928	1528
Policy5: LM-based Agent (Ours)	**160**	**362**	**10660**	**1160**	**1558**	**1286**	**1033**	**1610**

Table 3. The evaluation results of our trained policies, comparing them with previous SoTA methods and baseline policies. 'Policy x' refers to the policy optimized using the reward function 'RF x'. The 'RF x' column details the cumulative rewards for each policy, calculated in accordance with 'RF x'. Details of each reward function can be found in Table 1. The best value is highlighted in **bold**.

Fertilization	Irrigation	RF1↑
Baseline N Fertilization	Baseline Irrigation	984
Baseline N Fertilization	Training Irrigation	1376
Training N Fertilization	Baseline Irrigation	1157
Training N Fertilization	Training Irrigation	**1464**

Table 4. Performance comparison of the trained policies on both N fertilization and irrigation with the trained policies on either N fertilization or irrigation. The best values are shown in **bold**.

troduces an ablation study wherein the management policies for N fertilization and irrigation were trained independently. Specifically, while one practice was subject to optimization, the other adhered to established baseline methods. For instance, when optimizing an N management policy, the irrigation management followed the predefined baseline protocol, and vice versa. To be specific, experiments were conducted within the framework of the Florida case study, utilizing RF1 to guide the optimization process. The results, delineated in Table 4, provide a clear indication of the advantages inherent in the simultaneous optimization of N fertilization and irrigation management, as opposed to the independent optimization of each practice. This finding reveals that synergistically managing nitrogen fertilization and irrigation together yields superior agricultural outcomes compared to optimizing each practice in isolation.

4.4.2 Exploration of Framework

In order to investigate the most effective framework of RL agents, an ablation study was conducted. This study aimed to ascertain the impact of the framework's structure on management practices. Aligning with the setup of our previous experiments, we present the results for the Florida case using the reward function RF1. The outcomes, as depicted in Table 5, indicate that employing a three-layer MLP yields the best results with a traditional RL agent. However, a notable decline in performance is observed when scaling the agent size from an MLP to a ResNet152 [17]. This performance drop suggests the occurrence of overfitting within the RL framework, implying that simply increasing the size of the neural network does not necessarily enhance the exploration of optimal management practices.

Contrastingly, the use of LMs, such as Distilled Bert, demonstrated a different trend. Not only did the LM exhibit improved performance, but it also provided valuable insights. The results suggest that LMs possess a unique ability to comprehend the underlying patterns and logic of crop and environmental models. This capability enables them to pinpoint more optimal solutions while successfully circumventing the issue of overfitting, which was observed with larger neural network models.

5. Path to Deployment

The effectiveness of management policies trained within the DSSAT-simulated environment may not directly translate to real-world scenarios. This potential discrepancy arises from uncertainties in weather conditions and differences between the crop models used for training and actual agricultural systems. This phenomenon, known as the *sim-to-real gap* [57], highlights a common challenge in applying RL policies, developed and refined in simulated settings, to practical, real-world environments.

5.1. Closing the Sim-To-Real Gap

To enhance the robustness of our trained management policies against the challenges posed by the *sim-to-real gap*,

Model Architecture	# of Parameters	RF1↑
Three-layer MLP	0.2M	1425
Five-layer MLP	0.5M	1312
ResNet18	11.0M	510
ResNet50	25.6M	230
ResNet101	44.7M	107
ResNet152	60.4M	110
Distilled Bert	60.3M	**1464**

Table 5. Performance comparison of different frameworks as RL agents. The best values are shown in **bold**.

we plan to incorporate *domain and dynamics randomization* techniques, as suggested in previous studies [38, 48]. This approach involves introducing variations in critical parameters of the model and randomizing weather conditions during policy training. Such perturbations are intended to "force" the policies to become resilient to uncertainties in both the model and weather conditions.

While the primary focus of our current work is to establish the LM-based RL framework for crop management and to assess its effectiveness, we acknowledge the importance of addressing the robustness of these policies in real-world scenarios. Therefore, we aim to delve into this aspect in a forthcoming study, which will specifically target and evaluate the robustness of our LM-based RL policies against real-world variabilities and uncertainties.

5.2. Policy Evaluation with Measurement Noises

In order to assess the robustness of our method against random measurement noises, we conducted experiments following previous work [47]. In practical scenarios, farmers rely on weather forecasts and soil moisture measurements to make informed decisions. However, these data sources often contain inaccuracies due to forecast errors and sensor limitations. To simulate this real-world scenario, we tested LM-based RL under policy 1 from the Florida case study by introducing random measurement noises to key observable state variables each day in the simulation. These noise values were determined based on the real-world accuracy data of weather forecasts and commonly used soil moisture meters [13, 15, 18, 55]. For each variable of added noise, the policy's performance was evaluated 400 times, with the average cumulative reward and standard deviation reported. The results, detailed in Table 6, indicate that temperature and rainfall data inaccuracies have the most significant impact on policy performance, while other variables have minimal effects. Such an observation is consistent with previous research [47]. Notably, even with accumulated noise with multiple variables, the trained policy managed to achieve an average cumulative reward of 1248.8. While 15.3% lower than the reward in a noise-free environment, it is still considerably higher than that of the baseline policy. These findings demonstrate that the policies trained using our method can yield relatively satisfactory and robust results compared to baseline approaches, even under real-world scenarios.

Variables	Noises	RF 1	STD	Decrease (%)
Empirical Baseline	N/A	984.4	N/A	N/A
No Noise	N/A	1463.9	N/A	N/A
Soil water content	-+0.02	1463.9	0.0	0.0
Soil water content	-+0.05	1462.2	1.9	0.1
Temperature	-+1	1443.7	89.4	1.3
Temperature	-+2	1289.0	361.0	11.9
Solar Radiation	-+2%	1468.5	0.7	0
Solar Radiation	-+10%	1468.8	7.6	0
Rain Fall	90 % Acc.	1416.5	220.7	3.2
Leaf Area Index	-+10%	1457.1	1.2	0.4
Leaf Area Index	-+20%	1451.8	5.8	0.8
Soil water content +Temperature +Solar Radiation + Rain Fall + Leaf Area Index	-+0.02 -+2 -+2% 90 % Acc. -+20%	1248.8	386.8	15.3

Table 6. Performance of the LM-based RL with Policy1 under measurement noises evaluated with RF1. The decrease (%) is calculated with respect to RF1, where no noise was applied.

6. Conclusion

In this paper, we address the crucial challenge of optimizing crop management to maximize yield while minimizing management costs and environmental impacts. We present an innovative framework that combines deep reinforcement learning, language models, and crop simulations using Gym-DSSAT. The experimental results clearly demonstrate that Language Model-based Reinforcement Learning agents surpass baseline models and significantly outperform existing SoTA methods. This enhanced performance stems from the LM-RL agents' capacity to dynamically adjust their strategies according to different reward function designs, coupled with their ability to think and infer like expert agronomists. This dual capability enables them to maximize rewards in a variety of scenarios. Crucially, the framework has proven effective even in the presence of measurement noise in observable state variables, which is particularly promising for real-world applications.

We aspire for our work to serve as a proof of concept for the potential of LMs as adept agronomists, sparking interest and motivating further exploration in this area. The ultimate goal is to encourage researchers and practitioners to investigate and implement more advanced language models in practical agricultural settings. We believe that such advancements could significantly contribute to the evolution of agricultural technology, leading to smarter, more efficient, and sustainable farming practices worldwide.

References

[1] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*, 2022. 3

[2] Iffat Ara, Lydia Turner, Matthew Tom Harrison, Marta Monjardino, Peter DeVoil, and Daniel Rodriguez. Application, adoption and opportunities for improving decision support systems in irrigated agriculture: A review. *Agricultural Water Management*, 257:107161, 2021. 1

[3] Chace Ashcraft and Kiran Karra. Machine learning aided crop yield optimization. *arXiv preprint arXiv:2111.00963*, 2021. 2

[4] Senthold Asseng, Frank Ewert, Cynthia Rosenzweig, James W Jones, Jerry L Hatfield, Alex C Ruane, Kenneth J Boote, Peter J Thorburn, Reimund P Rötter, Davide Cammarano, et al. Uncertainty in simulating wheat yields under climate change. *Nature climate change*, 3(9):827–832, 2013. 3

[5] J-M Attonaty, M-H Chatelin, F Garcia, and S Ndiaye. Using extended machine learning and simulation technics to design crop management strategies. In *EFITA First European Conference for Information Technology in Agriculture, Copenhagen (Denmark)*, 1997. 1

[6] Simona Bassu, Nadine Brisson, Jean-Louis Durand, Kenneth Boote, Jon Lizaso, James W Jones, Cynthia Rosenzweig, Alex C Ruane, Myriam Adam, Christian Baron, et al. How do various maize crop models vary in their responses to climate change factors? *Global change biology*, 20(7):2301–2320, 2014. 3

[7] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016. 3

[8] Boyuan Chen, Fei Xia, Brian Ichter, Kanishka Rao, Keerthana Gopalakrishnan, Michael S Ryoo, Austin Stone, and Daniel Kappler. Open-vocabulary queryable scene representations for real world planning. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11509–11522. IEEE, 2023. 3

[9] Suiyao Chen. Some recent advances in design of bayesian binomial reliability demonstration tests. *Doctoral dissertation, University of South Florida*, 2020. 1

[10] Suiyao Chen, Xinyi Liu, Yulei Li, Jing Wu, and Handong Yao. Deep representation learning for multi-functional degradation modeling of community-dwelling aging population. *arXiv preprint arXiv:2404.05613*, 2024. 1

[11] Yikun Cheng, Pan Zhao, Fanxin Wang, Daniel J Block, and Naira Hovakimyan. Improving the robustness of reinforcement learning policies with $\mathcal{L}_1$ adaptive control. *IEEE Robotics and Automation Letters*, 7(3):6574–6581, 2022. 1

[12] Yikun Cheng, Pan Zhao, and Naira Hovakimyan. Safe and efficient reinforcement learning using disturbance-observer-based control barrier functions. In *Learning for Dynamics and Control Conference*, pages 104–115. PMLR, 2023. 1

[13] Douglas R Cobos and Decagon Devices. Why does my soil moisture sensor read negative? *Online]. Available: http://manuals. decagon. com/RetiredandDiscontinued/Slicksandcontent/ Presentations/SoilMoisture301.pdf*, 2010. 8

[14] HJ Di and KC Cameron. Nitrate leaching in temperate agroecosystems: sources, factors and mitigating strategies. *Nutrient cycling in agroecosystems*, 64:237–256, 2002. 5

[15] Eric Floehr. Weather forecast accuracy analysis. In *Proc of the 9th Python in Science Conference, SciPy*, pages 36–39, 2010. 8

[16] Frédérick Garcia. Use of reinforcement learning and simulation to optimize wheat crop technical management. In *Proceedings of the International Congress on Modelling and Simulation*, pages 801–806, 1999. 2

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 7

[18] Detlev Heinemann, Elke Lorenz, and Marco Girodo. Forecasting of solar radiation. *Solar energy resource management for electricity generation from local level to global scale. Nova Science Publishers, New York*, pages 83–94, 2006. 8

[19] Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *International Conference on Machine Learning*, pages 9118–9147. PMLR, 2022. 3

[20] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*, 2022. 3

[21] James W Jones, Gerrit Hoogenboom, Cheryl H Porter, Ken J Boote, William D Batchelor, LA Hunt, Paul W Wilkens, Upendra Singh, Arjan J Gijsman, and Joe T Ritchie. The DSSAT cropping system model. *European Journal of Agronomy*, 18(3-4):235–265, 2003. 1

[22] James W Jones, John M Antle, Bruno Basso, Kenneth J Boote, Richard T Conant, Ian Foster, H Charles J Godfray, Mario Herrero, Richard E Howitt, Sander Janssen, et al. Brief history of agricultural systems modeling. *Agricultural systems*, 155:240–254, 2017. 3

[23] Michiel GJ Kallenberg, Hiske Overweg, Ron van Bree, and Ioannis N Athanasiadis. Nitrogen management with reinforcement learning and crop growth models. *Environmental Data Science*, 2:e34, 2023. 3

[24] Elia Kaufmann, Antonio Loquercio, Rene Ranftl, Alexey Dosovitskiy, Vladlen Koltun, and Davide Scaramuzza. Deep drone racing: Learning agile flight in dynamic environments. In *Conference on Robot Learning*, pages 133–145, 2018. 1

[25] Geunwoo Kim, Pierre Baldi, and Stephen McAleer. Language models can solve computer tasks. *Advances in Neural Information Processing Systems*, 36, 2024. 3

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[27] Zhixin Lai, Jing Wu, Suiyao Chen, Yucheng Zhou, Anna Hovakimyan, and Naira Hovakimyan. Language models are

free boosters for biomedical imaging tasks. *arXiv preprint arXiv:2403.17343*, 2024. 2

[28] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9493–9500. IEEE, 2023. 3

[29] Shengjie Liu, Jing Wu, Jingyuan Bao, Wenyi Wang, Naira Hovakimyan, and Christopher G Healey. Towards a robust retrieval-based summarization system. *arXiv preprint arXiv:2403.19889*, 2024. 3

[30] Malvern Madondo, Muneeza Azmat, Kelsey Dipietro, Raya Horesh, Michael Jacobs, Arun Bawa, Raghavan Srinivasan, and Fearghal O'Donncha. A swat-based reinforcement learning framework for crop management. *arXiv preprint arXiv:2302.04988*, 2023. 3

[31] Wafa Malik, Ramon Isla, and Farida Dechmi. Dssat-ceres-maize modelling to improve irrigation and nitrogen management practices under mediterranean conditions. *Agricultural Water Management*, 213:298–308, 2019. 4

[32] German Mandrini, Cameron Mark Pittelkow, Sotirios Archontoulis, David Kanter, and Nicolas F Martin. Exploring trade-offs between profit, yield, and the environmental footprint of potential nitrogen fertilizer regulations in the us midwest. *Frontiers in plant science*, 13, 2022. 5

[33] Oier Mees, Jessica Borja-Diaz, and Wolfram Burgard. Grounding language with visual affordances over unstructured data. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11576–11582. IEEE, 2023. 3

[34] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015. 1, 4

[35] Hiske Overweg, Herman NC Berghuijs, and Ioannis N Athanasiadis. Cropgym: a reinforcement learning environment for crop management. *arXiv preprint arXiv:2104.04326*, 2021. 3

[36] Mike W Palmer, Julia Cooper, Catherine Tétard-Jones, et al. The influence of organic and conventional fertilisation and crop protection practices, preceding crop, harvest year and weather conditions on yield and quality of potato (Solanum tuberosum) in a long-term management trial. *European Journal of Agronomy*, 49:83–92, 2013. 1

[37] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023. 2

[38] Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3803–3810, 2018. 8

[39] Shreyas Sundara Raman, Vanya Cohen, Eric Rosen, Ifrah Idrees, David Paulius, and Stefanie Tellex. Planning with large language models via corrective re-prompting. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*, 2022. 3

[40] BVS Reddy, P Sanjana Reddy, F Bidinger, and Michael Blümmel. Crop management factors influencing yield and quality of crop residues. *Field Crops Research*, 84(1-2):57–77, 2003. 1

[41] Gautron Romain, Preux Philippe, Bigot Julien, Maillard Odalric-Ambrym, Emukpere David, et al. gym-dssat: a crop model turned into a reinforcement learning environment. *arXiv preprint arXiv:2207.03270*, 2022. 1, 3, 4

[42] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019. 4

[43] Tim Searchinger, Richard Waite, Craig Hanson, Janet Ranganathan, Patrice Dumas, Emily Matthews, and Carni Klirs. Creating a sustainable food future: A menu of solutions to feed nearly 10 billion people by 2050. final report. *World Resources Institute*, 2019. 1

[44] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[45] Ahmed Skhiri and Farida Dechmi. Impact of sprinkler irrigation management on the del reguero river (spain). i: Water balance and irrigation performance. *Agricultural Water Management*, 103:120–129, 2012. 1, 4

[46] Lijia Sun, Yanxiang Yang, Jiang Hu, Dana Porter, Thomas Marek, and Charles Hillyer. Reinforcement learning control for water-efficient agricultural irrigation. In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, pages 1334–1341, 2017. 2

[47] Ran Tao, Pan Zhao, Jing Wu, Nicolas F Martin, Matthew T Harrison, Carla Ferreira, Zahra Kalantari, and Naira Hovakimyan. Optimizing crop management with reinforcement learning and imitation learning. *arXiv preprint arXiv:2209.09991*, 2022. 1, 2, 3, 4, 5, 6, 8

[48] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30, 2017. 8

[49] Qianning Wang, Chenglin Wang, Zhixin Lai, and Yucheng Zhou. Insectmamba: Insect pest classification with state space model. *arXiv preprint arXiv:2404.03611*, 2024. 1

[50] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022. 2

[51] David Wright, Ian Small, Cheryl Mackowiak, Zane Grabau, Pratap Devkota, and Silvana Paula-Moraes. Field corn production guide: Ss-agr-85/ag202, rev. 8/2022. *EDIS*, 2022(4), 2022. 1, 4, 5

[52] Jing Wu, Ran Tao, Pan Zhao, Nicolas F Martin, and Naira Hovakimyan. Optimizing nitrogen management with deep reinforcement learning and crop simulations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and*

Pattern Recognition Workshops, pages 1712–1720, 2022. 1, 2, 3, 4

[53] Siheng Xiong, Ali Payani, Ramana Kompella, and Faramarz Fekri. Large language models can learn temporal reasoning. *arXiv preprint arXiv:2401.06853*, 2024. 2

[54] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022. 3

[55] Pengcheng Zhang, Yangyang Jia, Jerry Gao, Wei Song, and Hareton Leung. Short-term rainfall forecasting using multi-layer perceptron. *IEEE Transactions on Big Data*, 6(1):93–106, 2018. 8

[56] Chuang Zhao, Bing Liu, Liujun Xiao, Gerrit Hoogenboom, Kenneth J Boote, Belay T Kassie, Willingthon Pavan, Vakhtang Shelia, Kwang Soo Kim, Ixchel M Hernandez-Ochoa, et al. A simple crop model. *European Journal of Agronomy*, 104:97–106, 2019. 3

[57] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE symposium series on computational intelligence (SSCI)*, pages 737–744. IEEE, 2020. 7