

To Measure or Not: A Cost-Sensitive, Selective Measuring Environment for Agricultural Management Decisions with Reinforcement Learning

Hilmy Baja, Michiel Kallenberg, Ioannis N. Athanasiadis

Artificial Intelligence Group, Wageningen University and Research
 hilmy.baja@wur.nl, michiel.kallenberg@wur.nl, ioannis.athanasiadis@wur.nl

Abstract

Farmers rely on in-field observations to make well-informed crop management decisions to maximize profit and minimize adverse environmental impact. However, obtaining real-world crop state measurements is labor-intensive, time-consuming and expensive. In most cases, it is not feasible to gather crop state measurements before every decision moment. Moreover, in previous research pertaining to farm management optimization, these observations are often assumed to be readily available without any cost, which is unrealistic. Hence, enabling optimization without the need to have *temporally complete* crop state observations is important. An approach to that problem is to include measuring as part of decision making. As a solution, we apply reinforcement learning (RL) to recommend opportune moments to simultaneously measure crop features and apply nitrogen fertilizer. With realistic considerations, we design an RL environment with explicit crop feature measuring costs. While balancing costs, we find that an RL agent, trained with recurrent PPO, discovers adaptive measuring policies that follow critical crop development stages, with results aligned by what domain experts would consider a sensible approach. Our results highlight the importance of measuring when crop feature measurements are not readily available.

Code — <https://github.com/WUR-AI/CropGym-ToMeasureOrNot>

Introduction

As the global population continues to grow and the impacts of climate change become more prominent, optimal farm management decisions play a crucial role in our future sustenance. Effective farm management practices are key to not only ensure food security but also mitigate the apparent environmental risks related to agriculture (Lipper et al. 2014). Environmental risks come from mismanagement of farming activities, such as over-applying fertilizer and pesticides (Martínez-Dalmau, Berbel, and Ordóñez-Fernández 2021). In practice, farmers rely on their experience and in-field observations to make better-informed management decisions in order to minimize costs and reduce environmental impacts. For instance, before fertilizing, a farmer might mea-

sure the soil (or leaf) nitrogen content and based on that decide when and how much fertilizer to apply (Berghuijs et al. 2024).

However, acquiring a large amount of in-field agricultural data is labor intensive, time-consuming and expensive (Wu et al. 2022a). In most cases, the cost and inconvenience of collecting and processing the data outweigh its informational usefulness for the farmer (Thompson et al. 2019). Furthermore, this issue is compounded in regions with low data availability and high data collection costs, as the infrastructure required for gathering such data is limited, presenting a large hurdle for applying data-driven optimization (Cravero et al. 2022). By integrating data-collection recommendations as part of decision-making, data collection can be optimized and executed only when it is the most beneficial for the farmer. Specifically, it is of importance to have a system that simultaneously recommends optimal moments in a growing season to obtain measurements of crop states for the objective of cost-effective and optimal management.

The unrealistic assumption that feature observations are readily available has prompted recent work in reinforcement learning (RL) pertaining to active feature measuring for balancing measurement costs (Bellinger et al. 2021a). Previous work has shown early success in this setting, however, it assumed uniformity in the feature costs (Yin et al. 2020). In the real-world, recognizing that certain features are more expensive to measure than others is essential, as it potentially influences the policies that an RL agent learns.

In this work, we experiment on the effect of cost in a *measure-and-control* paradigm and design an RL environment in which an agent interacts with a crop growth model (CGM) environment during a growing season. It is important to acknowledge that crop features have different measuring costs. The ability to optimize fertilization with fewer crop state measurements is important to lower the hurdle of applying data-driven methods for improving agricultural management activities. We summarize our contributions below:

- We propose an RL paradigm of costly measurements for crop management, incorporating crop state measurements as part of decision making;
- We design and provide code for an RL environment, coupled with the WOFOST (Berghuijs et al. 2024) crop growth model, that allows an agent to simultaneously learn a control policy (fertilization, irrigation, etc.) and

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

a measuring policy. The RL environment is highly configurable, enabling the user to define their own objectives and RL spaces;

- We defined a financial focused reward function that includes the definition of feature measurement costs;
- We test our approach *in silico* in a case study in the Netherlands, optimizing fertilization for winter wheat in a rain-fed environment. Experimental results show that costs indeed affect the RL agent’s optimization capabilities, where higher costs prove detrimental to the agent’s performance. Nevertheless, with realistic costs, the RL agent manages to achieve better performance compared to a baseline of standard practice.

Related Work

This work fits in two growing RL research areas: RL for agriculture and costly active-measuring and -feature acquisition in RL.

RL for crop management. Generally, early RL research in agriculture utilizes simulated environments (Gautron et al. 2022a; Goldenits et al. 2024). The bulk of prior research builds a Gym interface that wraps around a CGM and obtain management policies *in silico*. Notable examples include CropGym (Overweg, Berghuijs, and Athanasiadis 2021; Kallenberg et al. 2023), gym-DSSAT (Gautron et al. 2022b), SWATGym (Madondo et al. 2023), and CyclesGym (Turchetta et al. 2022). Related to partially observable crop management, Tao et al. (2022) employed imitation learning to let an RL agent trained with large number of state features to work with a subset of these state features. Each work demonstrated a good potential for RL in crop management.

However, these prior works assume *temporally complete feature observations*, which is almost always not the case due to the cost, time and effort of acquiring these observations. Turchetta et al. (2022) identified these costly observations as a core challenge for RL in agriculture.

Costly measuring in RL. Using deep RL, there have been several studies of costly measurements (Bellinger, Crowley, and Tamblyn 2023; Krale, Simão, and Jansen 2023). In the context of active measuring, Nam, Fleming, and Brunskill (2021) formalized an MDP framework called action-contingent noiselessly observable MDP (ACNO-MDP) that, similarly, defines an explicit cost to observe the complete observation space. These prior works assume *completely missing feature observations* in certain time steps, which does not perfectly fit the formalization of our problem, but nonetheless still describes the general joint measuring and control approach.

Costly active feature acquisition in RL. *Active feature acquisition* (AFA) describes a paradigm where an agent can select features to acquire to improve model accuracy. There has been a fair amount of prior work in the machine learning domain about AFA (Shim, Hwang, and Yang 2018; Kossen et al. 2023). In RL, the work of Yin et al. (2020) includes the formalization of a POMDP extension they call AFA-POMDPs. They employ a sequential variational auto encoder (seq-VAE), that is pre-trained with fully observable features offline, to learn a simultaneous sequential measur-

ing and control policy. However, they did not investigate the realistic cost of individual features. They evaluate the medical sepsis RL environment (from Oberst and Sontag 2019), where they assume the cost of acquiring different features to be the same, despite the monetary cost of measuring heart rate and glucose being vastly different. This is a critical aspect considering the premise of cost-sensitive feature acquisition.

To our knowledge, there has been no prior work for costly feature acquisition in RL for crop management.

Problem Setup

In this section, we formalize the setting of this work. Moreover, we elaborate in detail the challenge in crop management and the technical implementation of the RL environment.

MDP Setting

Crop management problems involve sequential decision-making, which can be formalized into a Markov Decision Process (MDP). A standard MDP can be described with the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$, where \mathcal{S} is the state space and \mathcal{A} is the action space. \mathcal{T} and \mathcal{R} are the environment’s transition function $\mathcal{T}(s_{t+1}|s_t, a_t)$ and reward function $\mathcal{R}(s_t, a_t, s_{t+1})$, respectively. In crop management problems, as with many real-world environments, the agent is not privy of the complete environment state. This is where most real-world MDPs fall into: partially observable MDPs (POMDPs, Cassandra 1998). A POMDP is formalized into the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, \gamma \rangle$. In addition to the standard MDP elements, \mathcal{O} is introduced as the space of possible observations $o \in \mathcal{O}$, which has an observation function of $\mathcal{O}(o_t|s_t, a_t)$.

Coined by Yin et al. (2020), *active feature acquisition POMDPs* (AFA-POMDPs) describe an MDP setting where an agent can selectively acquire features at different time steps. It is characterized with the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{R}, \mathcal{C}, \gamma \rangle$. AFA-POMDPs extend POMDPs by extending the action space $\mathcal{A} = \mathcal{A}_c \times \mathcal{A}_m$ and adding a cost function \mathcal{C} . The action space consists of the control actions \mathcal{A}_c concatenated with measuring actions \mathcal{A}_m , and its Cartesian product shows the possible action outputs. An action output is denoted as $\mathbf{a}_c \in \mathcal{A}_c$ and $\mathbf{a}_m \in \mathcal{A}_m$. \mathcal{C} refers to the cost of unmasking observations. In our setup, we assume that \mathbf{a}_m contains a fixed number of features of size N_m .

To allow for cost-sensitive, selective measurements, we adapt the above framework by defining:

$$\mathcal{C}(\mathbf{a}_m) = \sum_{i=0}^{N_m} c_i a_{m_i}$$

\mathbf{c} is a vector that lists explicit costs for individual features i , denoted by c_i . Despite the action space extension, the transition function is similar to the standard MDP with a slight change $\mathcal{T}(s_{t+1}|s, \mathbf{a}_c)$, because we assume measurement actions do not affect state transitions. The objective of this setup is as follows:

$$\max_{\pi_c, \pi_m} \mathbb{E} \left[\sum_{t=0}^T \gamma^t \left(\mathcal{R}(s_t, \mathbf{a}_{c,t}) - \sum_{i=0}^{N_m} c_i a_{m_i,t} \right) \right]$$

Here, T is a constant, as we defined a fixed sowing and harvest date. Hence, we set γ to 1 and the objective effectively becomes the maximum expected total accumulated reward in a trajectory. The explicit reward function \mathcal{R} , including the cost term, is defined in the section RL Environment.

In the context of crop management, the elements of the tuples could be mapped as follows: \mathcal{S} is the whole range of crop and environmental states, some which are hidden to the agent. Elements in \mathcal{O} are a subset of \mathcal{S} ; crop and environmental states that the agent can observe. \mathcal{A}_c represents the levels of fertilization and \mathcal{A}_m is the possible measurement actions that directly correspond to N_m . \mathcal{T} is a simulation step of a CGM, T is the simulation duration, \mathcal{C} is the cost of measuring crop features and \mathcal{R} is the amount of projected yield.

Crop Management

As with many biological problems, crop management is highly complex with many factors that influence the process, outcome and yield of the crop growth. Generally, in the agricultural domain, this problem is denoted as $G \times E \times M$, which means that crop processes are influenced by the Genotype, Environment and Management (Martín, Olesen, and Porter 2014). In this work, we focus on arable (*i.e.*, open-field) rain-fed winter wheat. Therefore, we do not have any control of the environmental conditions. We assume the soil type is uniform through the whole farm. Moreover, we assume the absence of yield-reducing factors such as pests and diseases. Next, we opt to focus on one variety of winter wheat typically grown in the Netherlands (and western Europe), so we keep the genotype factor fixed. In the rain-fed field, we do not apply irrigation. This leaves us only with control over nitrogen (N) fertilization for the management.

N is a yield-limiting factor for crops (Chukalla et al. 2020). Knowing the crop or soil state before applying N fertilizer could enable timely fertilization actions, avoiding waste of resources and detrimental environmental effect. Moreover, in dry years, especially for rain-fed environments, the crop experiences water stress and water takes over as the yield-limiting factor, regardless of the N fertilization (Eck 1988).

The RL environment we developed is an interface of the Python Crop Simulation Environment (PCSE, de Wit 2023), which houses several crop growth models. In this work, we utilize the World Food Studies (WOFOST, Van Diepen et al. 1989; De Wit et al. 2019) CGM. WOFOST is a process-based CGM that simulates crop growth. It is capable of simulating nitrogen- and water-limited yield production, including the processes that are calibrated to a certain crop variety and location (Berghuijs et al. 2024). WOFOST has been proven to be a robust and reliable CGM, validated through its integral role in the European MARS crop yield forecasting system (Van der Velde and Nisini 2019).

RL Environment

RL Spaces: The observation space dictates the features that the agent can observe and measure. Table 1 shows the features of the CGM with continuous values for the crop and weather features. Additionally, we add another feature with continuous values called 'Random' that acts as a distraction feature, which will be explained in the Experiments and Results section. Figure 1 shows a schematic diagram describing the observation and action space processing.

The action space is a vector of discrete and binary values. The discrete value corresponds to the agent's control action of applying nitrogen to the crop, which are 7 levels: $\{0, 10, 20, 30, 40, 50, 60\} kg/ha$. The binary values correspond to the agent's action of measuring. Table 1 shows six features that we define as 'measurable'. The environment has a weekly time step. We aggregated the time series data following Kallenberg et al. (2023): the sequence of weather with a length of 3×7 (*i.e.*, daily rain, temperature and solar irradiance) was processed into an average pooling layer, resulting in a vector size of 3×1 . The crop features that had a length of 6×7 were shrunk to 6×1 by taking its last entry. The *Random* feature is also appended to the end of the crop feature at this moment. The vectors are then concatenated and flattened, a vector of masks corresponding to the measuring action of the agent at that time step is subsequently appended to the vector, resulting in a vector with a length of 16 features.

Constraints: The nitrogen management regulations in the Netherlands limit the amount nitrogen fertilizer that the farmers are allowed to apply (Flach and Selten 2021) due to potential detrimental environmental impact. This limit also affects the potential production of the yield. Hence, to keep realistic considerations, we impose a constraint for the agent to apply a maximum total of $200 kgN/ha$.

Reward function: The agent's main objective is maximizing yield by balancing the information obtained from costly measurements. We designed a natural financial reward function that assumes one unit of reward to be equal to 1 *kg* of winter wheat yield. R follows:

$$R_t = (TWSO_t - TWSO_{t-1}) - \beta N_t - D - \sum_{i=0}^{N_m} c_i a_{m_i,t}$$

where t is a weekly time step, $TWSO$ (total weight storage organ) is the wheat yield in *kg/ha*, and N is the amount of Nitrogen in *kg/ha*. The parameter β is a ratio of the price of 1 *kg* of wheat yield and 1 *kg* of N fertilizer, so we set $\beta = 2$ to mimic the respective prices (Wageningen Economic Research 2023a,b). D is a deployment cost that penalizes the agent for going out the field to apply N , with a penalty of 10 if $N_t > 0$. c_i is the cost of measuring a specific feature (*i.e.* *LAI*, *SM*, *etc.*) and $a_{m_i,t}$ denotes a measurement action a_{m_i} at time step t . We describe the rationale for the costs in the next section.

RL agent. For this experiment, we employ a PPO agent with recurrent networks (LSTM-PPO), as implemented by the library Stable Baselines 3 (Raffin et al. 2021). We take

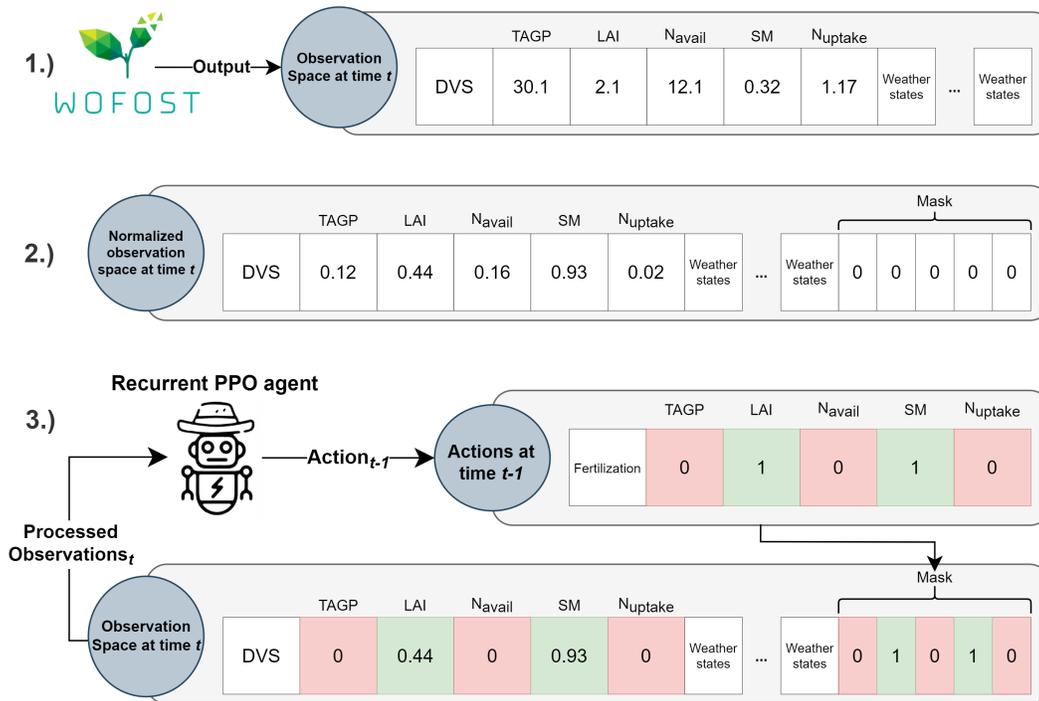


Figure 1: A schematic diagram describing the observation and action space processing, without the *Random* feature. These masks indicate whether the agent decided to measure a specific feature. 1.) Outputs of WOFOST are processed and flattened into the observation space vector. 2.) The observations are normalized with standardization and observation masks are initialized. 3.) The agent receives a vector comprising of the normalized features and masks. Normalization parameters were derived from multiple random runs. The agent pays a cost for measuring features denoted by the green shade (*LAI* and *SM*). If a feature is not measured, the agent does not pay its cost and the mask along with the normalized feature shows a value of zero.

advantage of the recurrent properties of the agent to discover temporal dependencies in the environment (Ni, Eysenbach, and Salakhutdinov 2021). Also, in their experiments, Bellinger et al. (2021b) proved that masks in the observation space work well for costly measuring with an actor-critic agent. The agent has two separate LSTM networks for actor and critic, each with 2 hidden layers with size of 256 and an activation function of \tanh . We set the agent’s learning rate to $1e-4$. The rest of the hyperparameters we kept same as the default.

Design Rationale and Assumptions

Feature selection and observation. WOFOST has a plethora of crop states/features that describe various crop and soil processes. The RL agent only observes a subset of these features, which makes the environment inherently partially observable. Though, a large portion of the features are highly correlated and redundant. Hence, we handpicked 6 crop features that are considered most important for the task of nitrogen fertilization, out of which, 5 are *measurable* (Table 1). *TAGP* describes the biomass of the crop, which is highly correlated to yield. *LAI* describes the area of leaves, thus explaining the photosynthesis capabilities of the crop. *NuptakeTotal* describes the total amount of N the crop took from the soil. *SM* describes the moisture around the crop’s root, related to potential water stress for the crop.

NAVAIL describes the amount of N in the soil. We give the agent full access to the remaining crop feature: development stage (*DVS*). This is akin to the farmer going out and visually checking a crop if it has emerged from the ground ($DVS \geq 0$), has reached the flowering stage ($DVS \geq 1$) or matured ($DVS = 2$). *DVS* grows monotonically from -0.1 to 2, hence it can be used to infer not only the stages of crop growth, but also the progression of time during the crop growth period. So, we also use *DVS* as a proxy for time and let the agent learn critical crop stages for adaptive measuring and fertilizing strategies. The agent also has full access to the weekly weather information, as this akin to a farmer checking the weather of last week from their nearest weather station.

Feature measurement costs. We defined explicit cost to obtain measurements for different features. We base the cost of an estimation on how expensive it is to obtain a measurement of the feature (shown under Table 1). *SM* and *LAI* are cheaper since it is possible to measure through sensors (Hummel, Sudduth, and Hollinger 2001), various remote sensing (Hasegawa et al. 2010; Schmutge 1983) or optical methods (Bréda 2003). *NuptakeTotal* can be measured with non-destructive sensors (Ulissi et al. 2011), which requires considerable labour to do for a whole field. *NAVAIL* can be measured with soil tests (Dahnke and Johnson 1990), requiring the soil to be sent to a lab, consequently incurring

Feature	Description [units]	Fully Observable	Realistic Cost
DVS	Development Stage [-]	✓	-
TAGP	Above Ground Biomass [kg/ha]	×	25
LAI	Leaf Area Index [-]	×	5
NAVAIL	Soil Nitrogen Content [kg/ha]	×	20
SM	Root Zone Soil Moisture [-]	×	5
NuptakeTotal	Total Nitrogen Uptake [kg/ha]	×	20
IRRAD	Solar Irradiance [$J/m^2/day$]	✓	-
TMIN	Minimum Temperature [$^{\circ}C/day$]	✓	-
RAIN	Daily Rainfall [cm/day]	✓	-
Random	Random Variable [-]	×	10

Table 1: Observable crop and weather features and the respective cost to observe. ‘Random’ is a distraction feature, which has no correlation to any of the other features. If a feature is fully observable, the agent will receive an observation of it in each time step. We also define the explicit cost for three different experimental setups. One unit of cost is roughly the price of 1 kg of wheat.

monetary cost and time. *TAGP* is an important feature for estimating yield and growth process of crops (Ma et al. 2022). The work of Kuyah and Rosenstock (2015) evaluated methods and the cost to measure above-ground biomass. Measuring *TAGP* requires considerable labor, and could be done through destructive or non-destructive methods.

Assumptions. We have made a few assumptions regarding the implementations and experimental design:

- *Measurements are noiseless.* Noise is almost always present due to either sensor, human or environmental factors. However, in this work we assume the measurements of the RL agent are noiseless.
- *Measurement actions do not alter the underlying model state.* All measurements are assumed to be done non-destructively.
- *Measurements are only repeatable costs.* Some measurements can be obtained by paying a *one-time cost*, such as buying a soil sensor for *SM*. However, we assume all measurements are *repeatable costs*, which is realistic due to the labor and time required to process the acquired measurements.
- *If measured, the feature values are immediately available.* Some measurements require processing or waiting time, especially for soil tests.

Experiments and Results

In this section we would like to answer the following questions:

1. How does different measurement costs affect a recurrent PPO agent’s ability to optimize yield?
2. How adaptive is the RL agent’s measuring policies between cold and hot years?
3. Does the agent learn to ignore features with no information?

Training conditions: We conduct experiments where the recurrent PPO agent has a joint task of measuring and

applying N fertilizer in the WOFOST CGM environment. Training was done in with semi-fine soil and climate conditions of the Netherlands with 3 coordinates: (52, 5.5), (51.5, 5), (52.5, 6) ($^{\circ}N, ^{\circ}E$) obtained from the NASA POWER weather dataset (Sparks 2018). The agent trained on the odd years from 1990 to 2022 (n=16), and we save the even years for evaluation. We set the simulation length to be fixed for 47 weeks, from sowing (October 1st) until harvesting (September 1st). We evaluate the agent with even years (n=16), and weather from the Netherlands with coordinates (52.5, 5.5) ($^{\circ}N, ^{\circ}E$). In total, one evaluation run contains 16 episodes.

The initial soil conditions (*i.e.*, how much moisture and N nutrients were in the soil when sowing) affects the agent’s learned policy. An agent will learn that the soil conditions always start with a similar amount, potentially remembering this and discouraging it from measuring. Hence, to add per episode variability to the RL environment, for soil moisture and nitrogen content, we set it with values from a randomized normal distribution generator, with a mean and standard deviation of 15 (in kg/ha for nitrogen content). The values are clipped to be bounded between 0 and 100. The generator is seeded, therefore repeatable across different seeds.

Training scenarios: To understand the effect of measuring cost in maximizing yield, we train agents with several cost scenarios. First, a *Realistic* cost scenario, which was explained in section Design Rationale and Assumptions and the explicit costs shown in Table 1. Second, a *Flat-cost* scenario, where each feature has a cost of 10 (kg of wheat) to measure, a bit lower than the average cost of *Realistic* features. Third, a *No-cost* scenario, where the agent is free to perform measure actions. And fourth, an *Exp-cost* scenario, a very expensive *flat* cost of 60 for measuring each feature, forcing the agent to converge to a policy of maximizing reward without measuring. To compare the performance of the RL agents to a simple fertilizing policy, we add a non-measuring fixed-fertilizing policy that fertilizes 3 times in fixed dates, each two months apart from January to May. Each fertilization has an amount of 66.67 kg/ha , with a cumulative total of 200 kg/ha adhering to the constraint

Scenario	LAI	SM	NuptakeTotal	TAGP	NAVAIL	Random
No-cost	23.9 (4.0)	25.2 (5.6)	19.9 (4.4)	24.2 (4.2)	26.9 (4.7)	23.5 (4.3)
Flat-cost	2.2 (1.7)	1.8 (1.6)	2.2 (1.3)	2.0 (1.5)	1.5 (1.2)	1.0 (0.9)
Realistic	5.5 (2.5)	4.8 (2.5)	0.6 (0.9)	0.2 (0.3)	0.3 (0.4)	0.2 (0.3)
Exp-cost	0.0 (0.0)	0.0 (0.0)	0.2 (0.3)	0.0 (0.0)	0.1 (0.1)	0.0 (0.0)

Table 2: Number of times the agent performed a measuring action for the corresponding feature in a one-year period. We report the average (and mean absolute deviation) across years and seeds. The maximum amount of measuring actions an agent can perform in one year is 47.

Scenario	Yield [tons/ha]	
	Median (IQR)	95% CI
No-cost	7.86 (0.97)	(6.92, 9.40)
Flat-cost	7.59 (1.17)	(6.88, 8.98)
Realistic	7.46 (1.09)	(6.45, 9.13)
Exp-cost	6.63 (4.35)	(0.93, 8.94)
All-observed	7.86 (0.96)	(6.99, 9.44)
None-observed	7.21 (3.82)	(0.93, 8.81)
Standard-practice	7.30 (1.09)	(6.55, 8.65)
Random-spread	0.65 (0.28)	(0.26, 1.06)

Table 3: Yield for different cost scenarios across seeds and evaluation years in the Netherlands.

applied in our experiments. Additionally, we add a *random-spread* baseline that randomly spreads 200 kg/ha Nitrogen fertilizer throughout the growing seasons. We treat *Exp-cost* as a baseline of what an agent can achieve without measuring in an AFA-POMDP setting. On the other hand, we treat *No-cost* as the upper bound for attainable yield. Further, we show two scenarios of RL agents trained outside the AFA-POMDP framework, with only a fertilizing action, without the task of measuring. Consequently, these scenarios have an easier learning process due to the smaller action and observation space. *All-observed* was trained with complete features. *None-observed* was trained with only *DVS* and the weather features, without any crop features. We train each scenario with 1.5M steps and repeat the experiments 10 times with consistent seeds.

Attained yield and measuring frequency of each scenario: In Table 3, we report the yield across different seeds and evaluation years. We estimate the 95% confidence interval around the sample median by bootstrapping. Additionally, we compute the inter-quartile range from the sample to show variance between years. In Table 2 we report the measuring frequency by showing the average number of measurement actions performed by the agent in a one-year period. Moreover, Figure 2 shows the temporal measuring policy of each cost scenario. We designed a reward function that maximizes yield through N fertilization, though we constrain the amount of N it can apply. At first glance to Table 3, we immediately notice that scenarios with lower cost achieve better yield. All scenarios learned to apply the maximum allowed amount of N fertilizer, *i.e.*, 200 kg/ha.

Notwithstanding, each scenario achieved different yields. We provide analyses below:

Between the four cost scenarios, *No-cost* achieves a higher median yield compared to the others and the lowest yield variance. The agent typically measures roughly half of the available steps in a year. Also, its performance is almost identical to *All-observed*, which hints to the potential redundancy for an agent in a crop management environment to have *temporally complete* observations.

There is a notable difference in achieved yield between *Flat-cost* and *Realistic*, with the former being ahead. Though, both are lower than *No-cost*, indicating that the presence of measuring cost prohibits the agent to obtain the required state information to perform optimal and timely fertilization. Moreover, the result of *Flat-cost* suggests that, with the unrealistic assumption that each features have the same cost to measure, the agent can utilize the more expensive features to better estimate the overall state. In *Realistic*, the difference in cost is mirrored in the measurement frequencies of certain features.

From an agronomic point of view, the measuring policy is understandable; the agent measures total N uptake (*NuptakeTotal*) twice as much as N soil content (*NAVAIL*), despite the similar cost, since it gives a better idea of how much N the crop took from the soil.

The expensive flat cost scenario (*Exp-cost*) performed the worse and has the largest variance between episodes, highlighting the importance of measuring. We note that the *Exp-cost* still converged to measuring some features in rare cases (*i.e.* *NuptakeTotal* and *NAVAIL*). The agent measured features that are related to N fertilization to potentially check whether more N is needed.

The agent trained in scenario *All-observed* achieved virtually identical results as *No-cost*. Scenario *None-observed* performed worse than *Realistic*, showing that the observations of these crop states are necessary for the agent to make informed decisions for timely fertilization actions.

In scenarios that achieve lower yields, a significant amount of fertilizer is wasted and not converted into yield. This excess N application will have a significant environmental impact. These results highlight the significance of measuring to achieve more effective and optimal fertilization policies.

Random feature measuring policy: We added *Random* as a distraction feature to investigate how an agent will deal with a feature that carries no information. We set the *Realistic* cost to 10, same as *Flat-cost*. A seeded normal distribu-

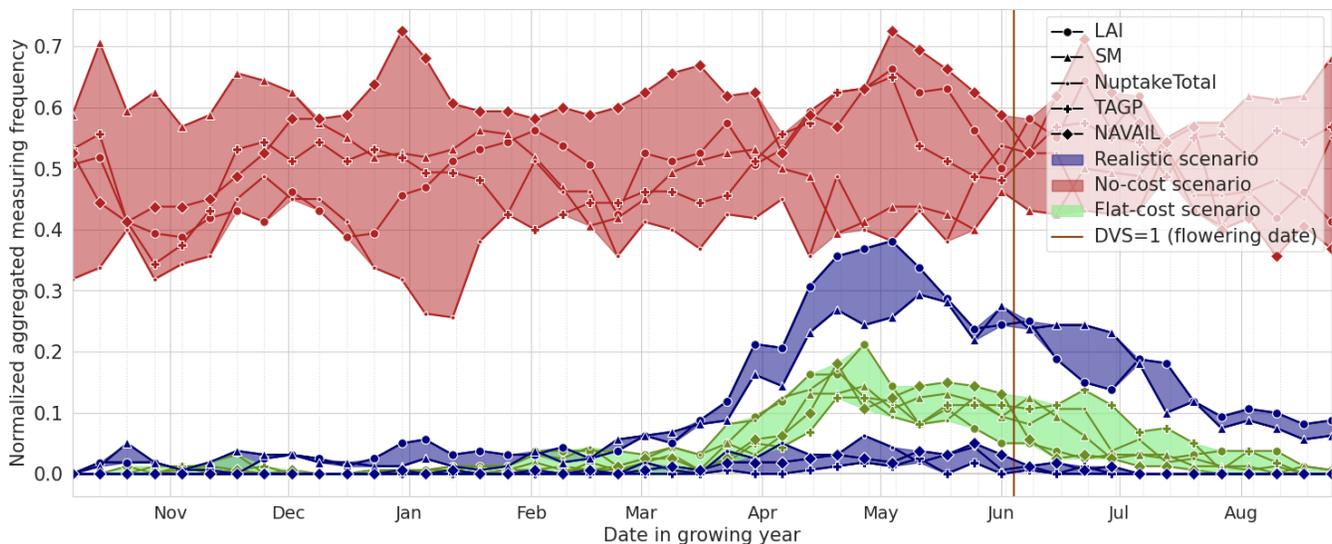


Figure 2: The normalized aggregated measuring frequency calculated from all seeds and years. We show shaded bands to emphasize the measuring action frequency throughout the growing year. The different colors represent the cost scenarios, and the different marker shapes represent measured features. The vertical line shows the average flowering day.

tion generator sets the values of *Random*, with a mean and standard deviation of 10. In the scenarios with costs (in Table 2), we notice that *Random* is the feature measured the least. However, in *No-cost*, it is measured about the same amount compared to other features. We notice in the *Realistic* scenario, the *Random* variable is measured as often as biomass (*TAGP*), despite the large difference in cost, suggesting that the agent is able to distinguish features that carry valuable information.

Temporal measuring policy: We present Figure 2 to show when the agent decides to measure in a growing year. The agents can fully observe the crop development stage (*DVS*), which we use as a proxy for time. We notice with *No-cost*, there is no clear pattern of measuring, which can be associated with the agent having no incentive to balance measuring costs with fertilization costs. In the scenarios with costs, the agent typically starts measuring close to the crop’s flowering stage, which is a critical stage for grain (yield) formation.

Adaptive policy: In this section we show how the agent adapts its fertilization based on its measurements and crop development. In colder years, the development of the plant is delayed, so there is a need to adjust management. We report Figure 3, showing how a cold and normal year affects the measuring and fertilization policy in two scenarios. Years 1990 and 2010 had yearly cumulative minimum temperatures of $1980.11^{\circ}C$ and $1488.36^{\circ}C$, respectively. Note that in 2010, *LAI* experienced delayed development due to a colder winter. Since the agent always observes *DVS* and the weather, it expects delayed crop development and adjust its measuring schedule for colder years. Accordingly, after obtaining the crop state information, the agent starts fertilizing later to adapt to the delayed development. Note that in 1990,

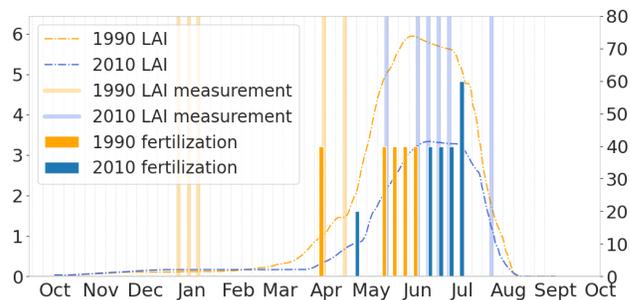


Figure 3: *LAI*, measurement and fertilization actions for 1990 and 2010 growing year period for the *Realistic* scenario. The y-axis (left) shows *LAI* development and secondary y-axis (right) shows bars depicting applied fertilizer actions in $[kg/ha]$. Measurement actions are shown with transparent vertical lines.

it chose to measure three times in January, which is generally the emergence period for winter wheat, signifying the start of stem and leaf (*LAI*) development. This demonstrates that the agent learned a policy that is able to anticipate growth throughout the season.

Discussion and Limitations

Realistic cost policy: The scenarios *No-cost* and *Flat-cost* achieved the best performance as they allow access to relatively complete crop state information (Wu et al. 2022b). However, these cost scenarios are unrealistic; crop state measurements are not always readily available, nor do they have uniform cost. The *Realistic* scenario agent, which had to pay a realistic cost for measuring features, still managed to perform well. The agent frequently measured *LAI*, which

is strongly related to nitrogen leaf content in different crop stages (Yin et al. 2003), and *SM* (soil moisture), which is a major yield-limiting factor (Day and Intalap 1970; Fahad et al. 2019). With these results, we demonstrate the importance of measuring when crop feature data is not readily available or difficult to obtain consistently.

RL Algorithm: We employed LSTM-PPO for our problem setup. Yin et al. (2020) utilize LSTM-A3C to train their RL policy, of which they employ a seq-VAE that is pre-trained with fully observable features that they feed to the LSTMs hidden states. In contrast, we do not employ pre-training in our approach and let the agent learn from a *tabula rasa*. A performance increase is probable if we pre-train our LSTM-PPO networks with fully observable features. Nevertheless, we show that our simple approach obtains measuring policies that enable good yield.

Assumptions: In section Design Rationale and Assumptions we explain the rationale of our design choices and defined some limiting assumptions of our proposed solution. Future work may build on our findings by relaxing these assumptions such as incorporating temporal delays or adding noise to measurements.

Scalability: A limitation of our approach is the need to possibly retrain the RL agent for different sites and setups. Scaling up is possible by including various sites in the agent’s training. Though, it is challenging from both domain perspectives: generalization remains a challenge in RL and ML (Cobbe et al. 2019; Li et al. 2022); and site-specific N management demands detailed site knowledge of in-field N soil variability for accurate recommendations (Schut and Giller 2020). In this work, we evaluate our approach in a representative case study with a well-calibrated CGM, thus we leave further exploration of this challenge for future work.

Sim2real gap: The experiments in this work are done *in silico*, hence there exists a gap between simulation and reality. Various things are not simulated in modern CGMs: yield-reducing factors such as pests and diseases (Donatelli et al. 2017), genetic variability (Hirel et al. 2007), among others, which introduce distribution shifts between simulation and reality. RL algorithms that are robust to distributional shifts (Turchetta et al. 2022) and randomization (Tobin et al. 2017) can help narrow this gap. Nevertheless, a CGM can achieve high accuracy if calibrated specifically to the year and conditions of the location it is simulating (Ahmed et al. 2020; He et al. 2017). For future work we intend to test and evaluate our developed system in field trials.

Conclusion

In this work, we propose an RL approach of integrating data collection to decision making by obtaining a measuring policy that balances measuring costs with fertilization. Inspired by the problem of difficult and costly data collection in agriculture, we design an RL environment with realistic considerations by adapting the framework of *AFA-POMDPs*. We evaluate our approach *in silico* in a case study in the Netherlands, with WOFOST, a thoroughly validated CGM.

Our test includes different cost scenarios. We find that the RL agent discovers adaptive measuring policies that coincide with critical crop development stages and learns that some features are more valuable than others. While cost does indeed affect the discovered measuring policy, with realistic costs the agent is still capable of achieving good yield by utilizing cheaper measurements.

Our work highlights the importance of measuring when crop feature measurements are not readily available. By integrating measurement recommendations in the decision making process, we can minimize unnecessary data collection. This approach is an important step to lower the hurdle of applying data-driven optimization for crop management. Ultimately, our work allows for the ability to optimize yield while reducing required data samples, further realizing better crop management policies for mitigating adverse environmental impact and sustaining global food demands.

Acknowledgments

The authors thank Ron van Bree for constructive discussions during the design stage of this work, and Aike Potze for the useful comments during the writing stage of the manuscript. This work was supported by the European Union Horizon Research and Innovation programme, Smart Droplets (Project Code: 101070496).

References

- Ahmed, M.; Ahmad, S.; Raza, M. A.; Kumar, U.; Ansar, M.; Shah, G. A.; Parsons, D.; Hoogenboom, G.; Palosuo, T.; and Seidel, S. 2020. Models calibration and evaluation. *Systems modeling*, 151–178.
- Bellinger, C.; Coles, R.; Crowley, M.; and Tamblyn, I. 2021a. Active Measure Reinforcement Learning for Observation Cost Minimization. In *Canadian Conference on AI*.
- Bellinger, C.; Crowley, M.; and Tamblyn, I. 2023. Dynamic observation policies in observation cost-sensitive reinforcement learning. *arXiv preprint arXiv:2307.02620*.
- Bellinger, C.; Drozdyuk, A.; Crowley, M.; and Tamblyn, I. 2021b. Scientific Discovery and the Cost of Measurement–Balancing Information and Cost in Reinforcement Learning. *arXiv preprint arXiv:2112.07535*.
- Berghuijs, H. N.; Silva, J. V.; Reidsma, P.; and de Wit, A. J. 2024. Expanding the WOFOST crop model to explore options for sustainable nitrogen management: A study for winter wheat in the Netherlands. *European Journal of Agronomy*, 154: 127099.
- Bréda, N. J. J. 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany*, 54(392): 2403–2417.
- Cassandra, A. R. 1998. A survey of POMDP applications. In *Working notes of AAAI 1998 fall symposium on planning with partially observable Markov decision processes*, volume 1724.
- Chukalla, A. D.; Reidsma, P.; van Vliet, M. T.; Silva, J. V.; van Ittersum, M. K.; Jomaa, S.; Rode, M.; Merbach, I.; and

- van Oel, P. R. 2020. Balancing indicators for sustainable intensification of crop production at field and river basin levels. *Science of the Total Environment*, 705.
- Cobbe, K.; Klimov, O.; Hesse, C.; Kim, T.; and Schulman, J. 2019. Quantifying Generalization in Reinforcement Learning. arXiv:1812.02341.
- Cravero, A.; Pardo, S.; Sepúlveda, S.; and Muñoz, L. 2022. Challenges to use machine learning in agricultural big data: a systematic literature review. *Agronomy*, 12(3): 748.
- Dahnke, W. C.; and Johnson, G. V. 1990. *Testing Soils for Available Nitrogen*, chapter 6, 127–139. John Wiley & Sons, Ltd. ISBN 9780891188629.
- Day, A.; and Intalap, S. 1970. Some effects of soil moisture stress on the growth of wheat (*Triticum aestivum* L. em Thell.) 1. *Agronomy Journal*, 62(1): 27–29.
- de Wit, A. 2023. The Python Crop Simulation Environment.
- De Wit, A.; Boogaard, H.; Fumagalli, D.; Janssen, S.; Knapen, R.; van Kraalingen, D.; Supit, I.; van der Wijngaart, R.; and van Diepen, K. 2019. 25 years of the WOFOST cropping systems model. *Agricultural systems*, 168: 154–167.
- Donatelli, M.; Magarey, R.; Bregaglio, S.; Willocquet, L.; Whish, J.; and Savary, S. 2017. Modelling the impacts of pests and diseases on agricultural systems. *Agricultural Systems*, 155: 213–224.
- Eck, H. V. 1988. Winter wheat response to nitrogen and irrigation. *Agronomy journal*, 80(6): 902–908.
- Fahad, M.; Ahmad, I.; Rehman, M.; Waqas, M. M.; and Gul, F. 2019. Regional wheat yield estimation by integration of remotely sensed soil moisture into a crop model. *Canadian Journal of Remote Sensing*, 45(6): 770–781.
- Flach, B.; and Selten, M. 2021. Netherlands: Dutch Parliament Approves Law to Reduce Nitrogen Emissions. Global Agricultural Information Network Report NL2020-0069, Foreign Agricultural Service, US Department of Agriculture.
- Gautron, R.; Maillard, O.-A.; Preux, P.; Corbeels, M.; and Sabbadin, R. 2022a. Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture*, 200: 107182.
- Gautron, R.; Padrón, E. J.; Preux, P.; Bigot, J.; Maillard, O.-A.; and Emukpere, D. 2022b. gym-DSSAT: a crop model turned into a Reinforcement Learning environment. arXiv:2207.03270.
- Goldenits, G.; Mallinger, K.; Raubitzek, S.; and Neubauer, T. 2024. Current applications and potential future directions of reinforcement learning-based Digital Twins in agriculture. *Smart Agricultural Technology*, 100512.
- Hasegawa, K.; Matsuyama, H.; Tsuzuki, H.; and Sweda, T. 2010. Improving the estimation of leaf area index by using remotely sensed NDVI with BRDF signatures. *Remote Sensing of Environment*, 114(3): 514–519.
- He, D.; Wang, E.; Wang, J.; and Robertson, M. J. 2017. Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234: 136–148.
- Hirel, B.; Le Gouis, J.; Ney, B.; and Gallais, A. 2007. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *Journal of Experimental Botany*, 58(9): 2369–2387.
- Hummel, J. W.; Sudduth, K. A.; and Hollinger, S. E. 2001. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Computers and electronics in agriculture*, 32(2): 149–165.
- Kallenberg, M. G.; Overweg, H.; van Bree, R.; and Athanasiadis, I. N. 2023. Nitrogen management with reinforcement learning and crop growth models. *Environmental Data Science*, 2: e34.
- Kossen, J.; Cangea, C.; Vértes, E.; Jaegle, A.; Patraucean, V.; Ktena, I.; Tomasev, N.; and Belgrave, D. 2023. Active Acquisition for Multimodal Temporal Data: A Challenging Decision-Making Task. arXiv:2211.05039.
- Krale, M.; Simão, T. D.; and Jansen, N. 2023. Act-Then-Measure: Reinforcement Learning for Partially Observable Environments with Active Measuring. *arXiv preprint arXiv:2303.08271*.
- Kuyah, S.; and Rosenstock, T. S. 2015. Optimal measurement strategies for aboveground tree biomass in agricultural landscapes. *Agroforestry systems*, 89: 125–133.
- Li, B.; Jin, J.; Zhong, H.; Hopcroft, J. E.; and Wang, L. 2022. Why Robust Generalization in Deep Learning is Difficult: Perspective of Expressive Power. *ArXiv*, abs/2205.13863.
- Lipper, L.; Thornton, P.; Campbell, B. M.; Baedeker, T.; Braimoh, A.; Bwalya, M.; Caron, P.; Cattaneo, A.; Garrity, D.; Henry, K.; et al. 2014. Climate-smart agriculture for food security. *Nature climate change*, 4(12): 1068–1072.
- Ma, C.; Liu, M.; Ding, F.; Li, C.; Cui, Y.; Chen, W.; and Wang, Y. 2022. Wheat growth monitoring and yield estimation based on remote sensing data assimilation into the SAFY crop growth model. *Scientific Reports*, 12(1): 5473.
- Madondo, M.; Azmat, M.; Dipietro, K.; Horesh, R.; Jacobs, M.; Bawa, A.; Srinivasan, R.; and O'Donncha, F. 2023. A SWAT-based Reinforcement Learning Framework for Crop Management. *arXiv preprint arXiv:2302.04988*.
- Martín, M. M.-S.; Olesen, J. E.; and Porter, J. R. 2014. A genotype, environment and management (GxExM) analysis of adaptation in winter wheat to climate change in Denmark. *Agricultural and Forest Meteorology*, 187: 1–13.
- Martínez-Dalmau, J.; Berbel, J.; and Ordóñez-Fernández, R. 2021. Nitrogen Fertilization. A Review of the Risks Associated with the Inefficiency of Its Use and Policy Responses. *Sustainability*, 13(10).
- Nam, H. A.; Fleming, S.; and Brunskill, E. 2021. Reinforcement learning with state observation costs in action-contingent noiselessly observable Markov decision processes. *Advances in Neural Information Processing Systems*, 34: 15650–15666.
- Ni, T.; Eysenbach, B.; and Salakhutdinov, R. 2021. Recurrent Model-Free RL is a Strong Baseline for Many POMDPs. *CoRR*, abs/2110.05038.

- Oberst, M.; and Sontag, D. A. 2019. Counterfactual Off-Policy Evaluation with Gumbel-Max Structural Causal Models. *CoRR*, abs/1905.05824.
- Overweg, H.; Berghuijs, H. N. C.; and Athanasiadis, I. N. 2021. CropGym: a Reinforcement Learning Environment for Crop Management. *CoRR*, abs/2104.04326.
- Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; and Dormann, N. 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 22(268): 1–8.
- Schmugge, T. J. 1983. Remote Sensing of Soil Moisture: Recent Advances. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21(3): 336–344.
- Schut, A. G.; and Giller, K. E. 2020. Soil-based, field-specific fertilizer recommendations are a pipe-dream. *Geoderma*, 380: 114680.
- Shim, H.; Hwang, S. J.; and Yang, E. 2018. Joint active feature acquisition and classification with variable-size set encoding. *Advances in neural information processing systems*, 31.
- Sparks, A. H. 2018. nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for R. *Journal of Open Source Software*, 3(30): 1035.
- Tao, R.; Zhao, P.; Wu, J.; Martin, N. F.; Harrison, M. T.; Ferreira, C.; Kalantari, Z.; and Hovakimyan, N. 2022. Optimizing crop management with reinforcement learning and imitation learning. *arXiv preprint arXiv:2209.09991*.
- Thompson, N. M.; Bir, C.; Widmar, D. A.; and Minert, J. R. 2019. Farmer perceptions of precision agriculture technology benefits. *Journal of Agricultural and Applied Economics*, 51(1): 142–163.
- Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; and Abbeel, P. 2017. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *CoRR*, abs/1703.06907.
- Turchetta, M.; Corinzia, L.; Sussex, S.; Burton, A.; Herrera, J.; Athanasiadis, I.; Buhmann, J. M.; and Krause, A. 2022. Learning long-term crop management strategies with CyclesGym. *Advances in Neural Information Processing Systems*, 35: 11396–11409.
- Ulissi, V.; Antonucci, F.; Benincasa, P.; Farneselli, M.; Tosti, G.; Guiducci, M.; Tei, F.; Costa, C.; Pallottino, F.; Pari, L.; et al. 2011. Nitrogen concentration estimation in tomato leaves by VIS-NIR non-destructive spectroscopy. *Sensors*, 11(6): 6411–6424.
- Van der Velde, M.; and Nisini, L. 2019. Performance of the MARS-crop yield forecasting system for the European Union: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015. *Agricultural Systems*, 168: 203–212.
- Van Diepen, C. v.; Wolf, J. v.; Van Keulen, H.; and Rappoldt, C. 1989. WOFOST: a simulation model of crop production. *Soil use and management*, 5(1): 16–24.
- Wageningen Economic Research. 2023a. agrimatie.nl. Price development of fertilizer. Accessed: 2023-09-16.
- Wageningen Economic Research. 2023b. agrimatie.nl. Price development of seeds and grains. Accessed: 2023-09-16.
- Wu, B.; Zhang, M.; Zeng, H.; Tian, F.; Potgieter, A. B.; Qin, X.; Yan, N.; Chang, S.; Zhao, Y.; Dong, Q.; Boken, V.; Plotnikov, D.; Guo, H.; Wu, F.; Zhao, H.; Deronde, B.; Tits, L.; and Loupian, E. 2022a. Challenges and opportunities in remote sensing-based crop monitoring: a review. *National Science Review*, 10(4): nwac290.
- Wu, J.; Tao, R.; Zhao, P.; Martin, N. F.; and Hovakimyan, N. 2022b. Optimizing nitrogen management with deep reinforcement learning and crop simulations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1712–1720.
- Yin, H.; Li, Y.; Pan, S. J.; Zhang, C.; and Tschitschek, S. 2020. Reinforcement learning with efficient active feature acquisition. *arXiv preprint arXiv:2011.00825*.
- Yin, X.; Lantinga, E. A.; Schapendonk, A. H.; and Zhong, X. 2003. Some quantitative relationships between leaf area index and canopy nitrogen content and distribution. *Annals of botany*, 91(7): 893–903.