Risk-Aware Bandits for Best Crop Management

Romain Gautron[1,2,3] **Dorian Baudry**[4] **Myriam Adam**[5,6,7] **Gatien N. Falconnier**[1,2,8]
Gerrit Hoogenboom[9] **Brian King**[3] **Marc Corbeels**[10]

R.Gautron@cgiar.org dorian.baudry@inria.fr

[1] AIDA, Université de Montpellier, Montpellier, France, [2] CIRAD, Montpellier, France,
[3] CGIAR Platform for Big Data in Agriculture, Alliance of Bioversity International and CIAT, Km 17, Recta Cali-Palmira 763537, Colombia,
[4] Oxford University. Formerly ENSAE Paris, and Inria Lille, University of Lille,
[5] CIRAD, UMR AGAP Institut, Bobo-Dioulasso 01, Burkina Faso,
[6] UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France,
[7] Institut National de l'Environnement et de Recherches Agricoles (INERA), Burkina Faso,
[8] International Maize and Wheat Improvement Centre (CIMMYT)-Zimbabwe
[9] Agricultural and Biological Engineering, University of Florida, USA,
[10] International Institute of Tropical Agriculture, Nairobi, Kenya

Abstract

Improving fertilizer practices through on-farm trials is challenging, especially in rain-fed farming due to weather uncertainty. However, it is crucial to test various practices to determine their performance, even if some may yield inferior results during the experiment. Our case study focuses on maize production in southern Mali, and we use the Decision Support System for Agrotechnology Transfer (DSSAT) crop model to simulate maize responses to nitrogen fertilization. We compare fertilizer practices using the Conditional Value-at-Risk (CVaR) of the Yield Excess (YE), a novel agronomic metric that considers both grain yield and nitrogen use efficiency. An "intuitive strategy" for practitioners, called *Explore-Then-Commit* (ETC) in the bandit literature, involves multi-year, multi-location field trials, where each practice is tested equally over several years. Inspired by a recent contribution, we propose the *Bounded–CVaR TS–Batch* (BCB) bandit algorithm, improving over ETC both theoretically and in crop model simulations. This study opens new horizons for risk-aware identification of best crop management practices' in real conditions.

1 Introduction

Identifying site-specific best-performing crop management is crucial for farmers to increase their income from crop production, but also for minimizing the negative environmental impacts of cropping activities [57]. However, due to weather variability, the identification of these practices can be challenging, in particular with rainfed farming: what worked best in a wet year or a year with sufficient rainfall, might not work in the next year, when rainfall is lower [3]. The performance of crop management at a given site has an underlying unknown distribution due to inter-annual weather variability, thus creating great uncertainty [20]. Because crop management decisions are repeated for each new crop growing season, the identification of best available crop management falls into the category of sequential decision making under uncertainty [25]. Computer-based decision support tools can allow farmers to make more informed (less uncertain) decisions about their cropping practices from one year to the next, and can facilitate farmers' risk management in the face of seasonal

38th Workshop on Aligning Reinforcement Learning Experimentalists and Theorists (ARLET 2024).

weather variability [28]. There exist numerous decision support tools of widely ranging complexity for crop management, that have been introduced to farmers with varying degrees of success [25].

In this study, we focus on nitrogen fertilization for rainfed maize production in southern Mali, using the Decision Support System for Agrotechnology Transfer (DSSAT) [29] crop model to simulate real-world performance. The objective is not to optimize nitrogen management itself but to refine the sequential selection strategy among pre-selected practices. While our example is nitrogen fertilization, our broader goal is to provide a method for identifying the best crop management practices from any set of predefined options, such as varietal choice or irrigation. This method is adaptable to real field conditions and do not rely on model simulations alone.

More specifically, we frame this problem as a Multi-Arm Bandit (MAB, see [39] for a survey). Indeed, a decision-maker (group of farmers) repeatedly faces a choice between contending actions (pre-selected practices), collect a reward (grain yield), and aims at iteratively improving their decision-making with trials in order to implement as often as possible the *best practice* among all candidates. In MAB, the typical objective is to sequentially choose actions such that the expected sum of rewards is maximized. This is equivalent to minimizing the regret, which measures the total losses that occur by testing *sub-optimal* practices [50]. This can be done by carefully balancing *exploration* (testing all practices to learn their reward) and *exploitation* (choosing in-trial best performing practices). The *exploration-exploitation dilemma* is a reality for farmers when implementing crop management. Farmers typically want to minimize overall crop yield losses and therefore may explore the performance of promising new crop management practices on small test field plots [14, 17]. Thus, they avoid potentially large crop yield losses from new practices by managing a gradual transition between the current practices and the promising new one(s), based on the results they obtain on the small test plots. Because trials are costly, bandit algorithms that minimize regret [8, 36, 12] are better-suited than pure exploration algorithms [22, 34] for the problem of improving crop management practices. In this article, we adapt a *risk-aware* bandit algorithm proposed in [9], reflecting the preferences of the farmer for multi-year experiments, that successfully tackles the realistic crop-management problem that we introduce.

2 Methodology

2.1 The virtual crop management problem

In our virtual crop management problem, a population of 500 virtual farmers from southern Mali joins a participatory experiment to improve their nitrogen fertilizer practices for maize production in their fields. The distribution of soil types of the fields of the group of virtual farmers is representative of the region (see Table 1 in appendix), and we design cohorts as group of farmers growing maize on the same soil type. For each cohort, we want to recommend as often as possible the best nitrogen fertilizer practice from a set of candidates (see Section 2.2 for the performance measures considered). The research team sets the additional objective to limit the maize yield losses of individual farmers that could arise from poor nitrogen fertilizer practice recommendations during the experiment.

At the beginning of each crop growing season, a random number of virtual farmers (uniformly obtained between 250 and 350) of the total population of 500 farmers volunteers to apply the recommended fertilizer policies provided by the research team. Each year, the group of volunteers is variable in size and in the representation of cohorts, as could occur in reality (Figure 1). Thus, researchers do not control the composition of the group of volunteers. Each virtual farmer indicates the fields and corresponding soil types on which she/he plans to grow maize. Following the recommendation strategies, researchers then provide a fertilizer practice among ten candidates to each virtual farmer for the ongoing growing season, depending on her/his soil type. At the end of the season, the farmers share their results in terms of maize grain yields with the research team, allowing to refine the recommendations for the next season. The whole experiment is repeated during 20 consecutive years following the same steps. Figure 6a in appendix illustrates this process, and corresponds to the steps described in Figure 1. We detail the experimental design in Appendix A, including the choice of the candidate practices, justified from an agronomic perspective.

Maize growth simulations. In order to get a proxy for real-world performances of the maize nitrogen fertilizer practices, we simulated maize growth responses to fertilization under the growing conditions of the Cercle of Koutiala in southern Mali using gym-DSSAT v0.0.7 [26] developed from

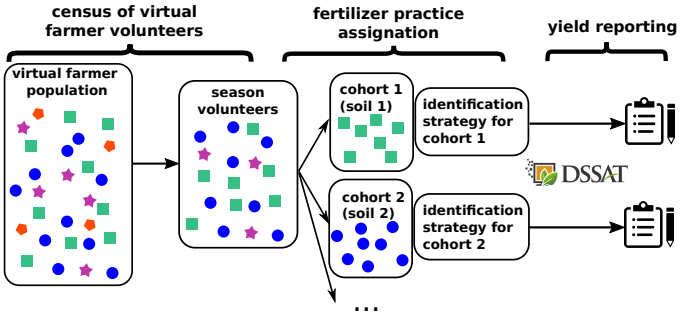


Figure 1: Set-up of the numerical experiment with ($n = 500$) virtual farmers grouped by cohorts ($c = 7$, Table 1, identified by symbols), sharing the same soil type.

the Decision Support System for Agrotechnology Transfer (DSSAT) `v4.7` crop model [29]. In the model simulations, a different weather time series is generated for each growing season but also for each farmer using the WGEN weather simulator [47], inducing sets of independent simulated maize yield responses to nitrogen fertilization. The modeling approach we embraced can be seen as distant farms encountering distinct weather patterns within the same year. The variability introduced by weather randomness is the source of uncertainty in the simulator. In Appendix B we further detail the DSSAT simulation settings. All the numerical experiments in this paper are meant to be fully reproducible, and the code is open source. The repository containing the Python code with the necessary packages, instructions and experimental data will be provided in the camera ready version.

2.2 Performance indicators of fertilizer practices

Yield Excess A popular indicator to evaluate both the economic and environmental performance of a nitrogen fertilizer practice π is Agronomic Nitrogen use Efficiency (ANE), as defined by [60]:

$$\mathrm{ANE}^{\pi} := \frac{\mathrm{Y}^{\pi} - \mathrm{Y}^{0}}{\mathrm{N}^{\pi}}, \tag{1}$$

where Y^{π} is the crop yield obtained with the fertilizer practice π with a quantity N^{π} of nitrogen, and Y^{0} is the yield of the control obtained in the same conditions without nitrogen fertilization. Maximizing ANE is a proxy of minimizing the quantity of nitrogen losses, e.g. through nitrate leaching. However, there are certain limitations associated with using ANE as an indicator for optimizing fertilizer rates. For example, an ANE value of 25 kg grain/kg N can be achieved with a fertilizer input of 20 kg N/ha resulting in a total yield gain of 500 kg/ha, or with an input of 60 kg N/ha resulting in a total gain of 1500 kg/ha. For the same ANE, a farmer prefers the fertilizer practice that provides the greatest crop yield gain, i.e. with 60 kg N/ha. Similarly, for a fixed crop yield the most efficient fertilizer practice should be preferred. Hence, we introduce the Yield Excess (YE) indicator to implement these preferences. The YE of a fertilizer practice π is defined with respect to a reference practice π_{ref} of fixed efficiency $\mathrm{ANE}_{\mathrm{ref}}$, using the same quantity of nitrogen fertilizer as practice π denoted by N^{π}, and is computed as follows,

$$\mathrm{YE}^{\pi} := \mathrm{Y}^{\pi} - \mathrm{Y}^{\pi_{\mathrm{ref}}} = \underbrace{\mathrm{Y}^{\pi} - \mathrm{Y}^{0}}_{\substack{\text{yield gain of } \pi \\ \text{w.r.t. control}}} - \underbrace{\left(\mathrm{Y}^{\pi_{\mathrm{ref}}} - \mathrm{Y}^{0}\right)}_{\substack{\text{yield gain of } \pi_{\mathrm{ref}} \\ \text{w.r.t. control}}} = \left(\mathrm{Y}^{\pi} - \mathrm{Y}^{0}\right) \times \underbrace{\left(1 - \frac{\mathrm{ANE}_{\mathrm{ref}}}{\mathrm{ANE}^{\pi}}\right)}_{\text{penalization factor}} \tag{2}$$

The YE of practice π with respect to the reference practice π_{ref} corresponds to the yield difference between the practice π and a reference practice that has a fixed ANE equal to $\mathrm{ANE}_{\mathrm{ref}}$ and which uses the same quantity N^{π} of nitrogen fertilizer as π. YE^{π} increases with ANE^{π} (Figure 2). YE^{π} is negative and decreases with $\mathrm{Y}^{\pi} - \mathrm{Y}^{0}$ when $\mathrm{ANE}^{\pi} < \mathrm{ANE}_{\mathrm{ref}}$ and is positive and increases with $\mathrm{Y}^{\pi} - \mathrm{Y}^{0}$ when $\mathrm{ANE}^{\pi} \geq \mathrm{ANE}_{\mathrm{ref}}$. The YE of fertilizer practices with efficiency below $\mathrm{ANE}_{\mathrm{ref}}$ are negatively affected by this metric. We chose $\mathrm{ANE}_{\mathrm{ref}} = 15$ kg grain/kg N for our study, i.e. the average ANE currently achieved by farmers across sub-Saharan Africa [54, 60].

Risk-awareness Because farmers are usually risk averse [e.g. 15, 43, 33], they are likely to prefer a stable maize grain yield excess of, for example, 3000 kg/ha rather than a yield of 5000 kg/ha in half

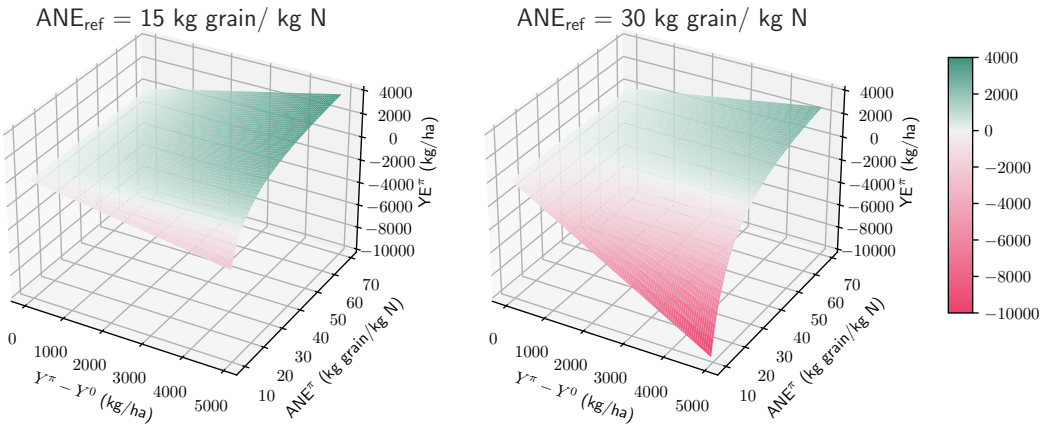
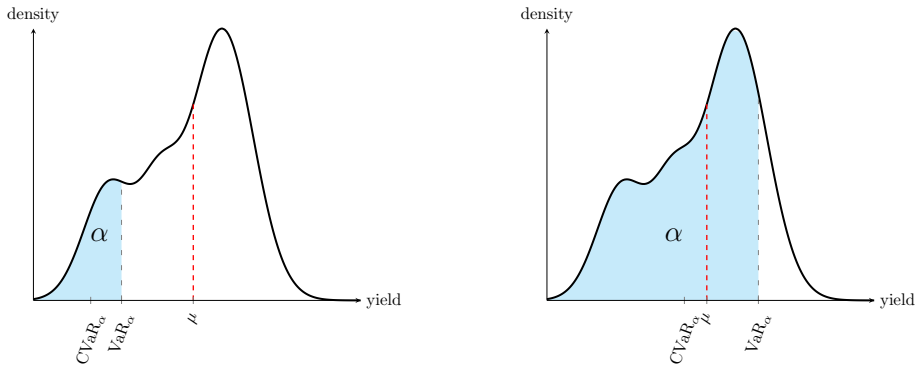


Figure 2: Yield Excess (YE^{π}, Equation 2) for $\text{ANE}_{\text{ref}} = 15$ kg grain /kg N (left) and $\text{ANE}_{\text{ref}} = 30$ kg grain /kg N (right) as a function of ANE_{π} and $\text{Y}^{\pi} - \text{Y}^{0}$.



(a) High risk aversion ($\alpha \approx 20\%$)

(b) Low risk aversion ($\alpha \approx 80\%$)

Figure 3: Conditional Value-at-Risk (CVaR) of level α in the case of high (a) and low(b) risk aversion. CVaR is the value of the blue area, the expectation of the example distribution is denoted by μ.

of the years, and of 1000 kg/ha in the other half of the years, while both distributions have the same expectation. To account for risk aversion, we consider the Conditional-Value-at-Risk (CVaR, [40, 1]), a popular risk-metric originated from the finance sector. Two definitions of the CVaR coexist in the literature, depending if an outcome is considered as a gain or a cost [16]. We adopt the gain point of view, in which the CVaR puts emphasis on the lower tail of a distribution. For a (continuous) random variable X with cumulative distribution function F_X, the CVaR is defined as follows

$$\text{CVaR}_{\alpha}(X) := \mathbb{E}[X | X \leq \text{VaR}_{\alpha}(X)] \,, \quad \text{with } \text{VaR}_{\alpha}(X) := \inf \{x \in \mathbb{R} : F_X(x) > \alpha\} \,. \tag{3}$$

A farmer is likely to prefer the practice with the highest CVaR for the considered level α. The more $\alpha \to 0^{+}$, the more the measure puts emphasis on the worst observable yields. On the contrary, the more $\alpha \to 1$, the less risk averse is the CVaR. When $\alpha = 1$, the CVaR equals the usual expectation $\mathbb{E}[X]$, which is risk neutral (see Figure 3 for an illustration). In our study, we choose $\alpha = 30\%$, representing the expected crop yield of the 30% lowest observable yields.

2.3 Formalization as a Multi-Armed Bandit

The maximization of farmers' YE when experimenting crop management practices can be modeled as a bandit problem [39, 8, 5] with specific features, that we detail in the following. First, contrarily to the canonical bandit problem where an observation directly follows each trial, observations are made at the end of each season (see Figure 6b), which is known as batched bandits [45, 21]. However, since the size of the batches is not controlled by the researchers this feature does not pose challenges, and simply motivates the use of a *randomized* bandit algorithm, to encourage the diversity of practices

tested at early stages of the experiment. Second, the objective of maximizing the CVaR of the Yield Excess indicator situates our problem in the literature on *risk-aware* bandits, and more precisely *CVaR bandits* [13, 9, 53]. These characteristics make `CVTS` [9] a natural candidate for our problem.

Formally, in our virtual experiment, for $t \in \{1, 2, \cdots, T\}$, in each season t, researchers assign a number n_t of volunteer farmers for season t with a nitrogen fertilizer practice $\pi \in \{1, 2, \cdots, K\}$. Each farmer belonged to a cohort $c \in \{1, 2, \cdots, C\}$. At the end of season t, researchers assemble rewards (YE of each trial) $Y_t = \{y_t^1, \ldots, y_t^{n_t}\}$ as a result of the fertilizer practices of all farmers for season t. For each cohort $c \in \{1, \cdots, C\}$, rewards are independently and identically distributed from unknown stationary distributions $\{\nu_1^c, \cdots, \nu_K^c\}$. These reward distributions are the YE with $\text{ANE}_{\text{ref}} = 15$ kg grain/kg N associated to each of the 10 recommended nitrogen fertilizer practices, for a given soil type. Following [9], for a given parameter α, a bandit policy selecting action $k_{t,n}$ for the n-th farmer in season t incurs the following α-CVaR regret

$$\mathcal{R}_T^\alpha = \sum_{t=1}^{T} \mathbb{E}\left[\sum_{n=1}^{n_t}\sum_{c=1}^{C} \mathbb{1}(c_n = c)\left(\text{CVaR}_\alpha(\nu_\star^c) - \text{CVaR}_\alpha(\nu_{k_{t,n}}^c)\right)\right], \tag{4}$$

where $\nu_\star^c = \underset{k}{\text{argmax}}\ \text{CVaR}_\alpha(\nu_k^c)$ is the distribution of the optimal practice for cohort c.

2.4 Algorithms for α-CVaR-regret minimization

We expect fertilizer practices to perform differently within each cohort, and assume that no model is available to share knowledge between cohorts. Hence, we treat the cohorts independently, presenting to each of them a replication of the same algorithm. In this study, we consider two algorithms: the standard `ETC` (Explore-Then-Commit) strategy, previously referred as the "intuitive strategy" for agronomists, and the more elaborated `BCB` (Bounded-CVaR-Thompson-Sampling Batch) strategy.

Intuitive strategy (`ETC`) Explore-Then-Commit (`ETC`) [23] provides a simple and intuitive solution to the exploration-exploitation dilemma. During an initial exploration phase of an arbitrary number of years, `ETC` equiproportionally test all nitrogen fertilizer practices. Thereafter, the exploitation phase starts and `ETC` chooses for the remaining time the fertilizer strategy that has shown best performance during the exploration phase. In Appendix C.2, we provide a natural adaptation of `ETC` to the batch setting (see Section 2.1) using the CVaR of rewards rather than the expectation. We consider `ETC-3` and `ETC-5`, with respectively three and five years for the exploration phase. During the exploration phase, fertilizer practices are randomly assigned in equal proportions to the farmers within the cohort.

Optimal Bandit strategy (`BCB`) `BCB`, adapts the *CVaR Thompson Sampling* algorithm (`CVTS`, [9]) to the batch setting of this paper. `CVTS` is itself an adaptation of the celebrated *Thompson Sampling* (`TS`) [56, 6, 37, 48] that achieves optimal guarantees in minimizing the α-CVaR-regret for bounded distributions with a known support. This is our main motivation for adapting this algorithm, instead of other candidates based on the *optimism principle* [13, 53]. In our setting, the YE is naturally bounded due to physical constraints.

An overview of the execution of `BCB` is shown in Algorithm 1, and a more detailed implementation can be found in Appendix C.1. For the execution of `BCB` (see first step of Algorithm 1), we set the maximum obtainable maize YE at 4000 kg/ha (Figure 2) for $\text{ANE}_{\text{ref}} = 15$ kg grain/kg N for all fertilizer practices.

3 Results: theory and experiments

3.1 Theoretical guarantees

It is proved that the `CVTS` algorithm is asymptotically optimal for CVaR-bandits [10]. The main difference with `BCB` is that it incorporates batched feedback and parallel learning across all cohorts. Our theoretical result proves that the batches do not alter the theoretical performance of the algorithm.

Algorithm 1 Simplified pseudo-code of BCB (Bounded-CVaR-Thompson-Sampling Batch), single cohort.

for each practice $k \in \{1, \dots, K\}$, add the maximum obtainable YE to the observations ;
// prior to any experiments, treat maximum YE as one observation for each practice
for season $t \in \{1, \cdots, T\}$ **do**
  for farmer $f \in \{1, \cdots, n_t\}$ **do**
    for fertilizer practice $k \in \{1, \cdots, K\}$ **do**
      Re-weight the YE collected for k with random weights drawn uniformly at random
      // Dirichlet distribution with a vector of ones as parameter [18]
      Evaluate the CVaR at level α of the resulting (noisy) empirical distributions
    end
    Recommend to the farmer f the fertilizer practice with the maximum noisy CVaR
    // Thompson Sampling principle: greedy w.r.t. the sampled model
  end
  Collect and store all results of the season for all fertilizer practices
end

Theorem 3.1 (α-CVaR regret of BCB)**.** *Let F denote the maximum number of farmers participating the experiment. Assume that there is only one cohort ($C = 1$). Then, BCB satisfies*

$$\mathcal{R}_T^{\alpha} \leq \mathcal{R}_T^{\alpha,\texttt{CVTS}} + \mathcal{O}\left(F \sum_{k:\Delta_k^{\alpha}>0} \Delta_k^{\alpha}\right), \quad \text{with } \Delta_k^{\alpha} = \mathit{CVaR}_{\alpha}(\nu_{\star}^{c}) - \mathit{CVaR}_{\alpha}(\nu_k^{c})$$

and $\mathcal{R}_T^{\alpha,\texttt{CVTS}}$ is the regret upper bound of the CVTS algorithm, provided in Theorem 3 of [9]. In particular, the asymptotic optimality of CVTS is preserved with the batched feedback.

Furthermore, the generalization to $C > 1$ cohorts is trivial due to their independent treatment. We prove Theorem 3.1 in Appendix D, by comparing a regret upper bound obtained for BCB (Theorem D.1) and the upper bound obtained in [9] for CVTS (Theorem D.2).

3.2 Simulated maize yield responses to nitrogen fertilizer practices

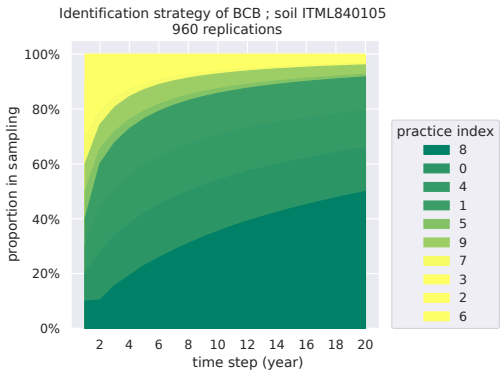
All simulated maize yield responses to nitrogen fertilization showed values within the expected ranges for the growing conditions in Koutiala, with an average grain yield varying from 3125 kg/ha for a sandy soil with low fertility (ITML84105) up to 3945 kg/ha for a loamy soil (ITML84106). When applying the most promising fertilization strategies, YE (i.e. yield gain compared to the reference) ranged from 1200 kg/ha to 1800 kg/ha, and $\text{CVaR}_{30\%}$(YE) (i.e. the mean crop YE of the 30% lowest yields) from 500 kg/ha to 1032 kg/ha. Table 5 provides the statistics of the best available nitrogen fertilizer practices for each soil type (Table 1), and Figure 7 in appendix shows the distributions of grain maize yield, ANE and YE responses.

For all soil types, the best available nitrogen fertilizer practices were either Practice 0 or 8 i.e. practices with a single nitrogen top-dressing application that is not threshold dependent (Table 5). Yet, the fertilizer practices had different responses for the different soil types in terms of grain yield and ANE (and consequently YE) and ranking of the practices were inconsistent across the soil types (Figure 7). For instance, for the soil ITML840104 (silt clay loam of medium fertility), Practices 0 to 4 all had similar YE values (Figure 7e), whilst, for the soil ITML840105 (silt clay loam of low fertility), Practices 0, 1 and 4 had substantially higher YE values compared to Practices 2 and 3 (Figure 7f).

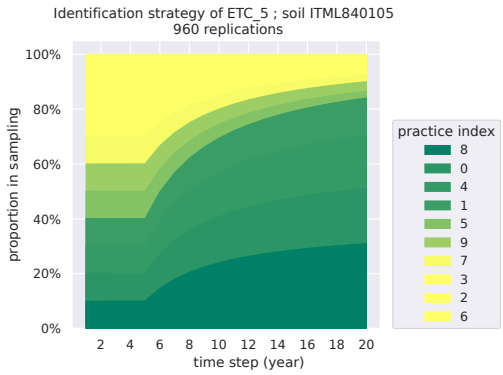
3.3 Empirical results with DSSAT

We now describe the empirical results obtained by repeating multiple times the experiment describes in Section 2.1, in order to compare the performance of ETC-3, ETC-5, and BCB in this context.

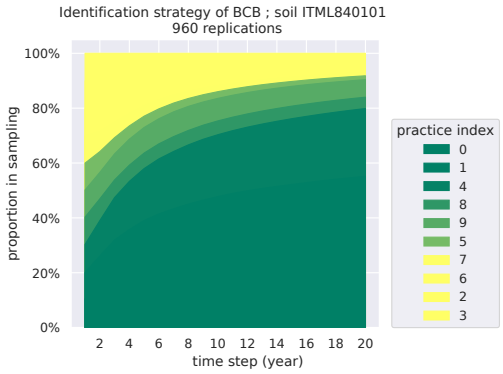
Proportion of best choices Figure 4 provides the average proportions at which the fertilizer practices were selected by the identification strategies, from the beginning of the experiment to



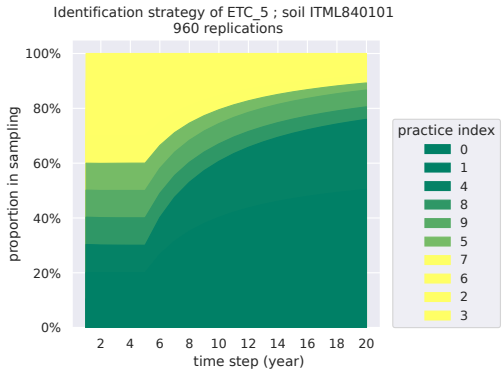
(a) BCB sampling proportions for soil ITML840105.



(b) ETC-5 sampling proportions for soil ITML840105.



(c) BCB sampling proportions for soil ITML840101.



(d) ETC-5 sampling proportions for soil ITML840101.

Figure 4: Averaged sampling proportions for soils ITML840105 and ITML840101, $T = 20$ years. Practices are ordered according to their true Conditional Value-at-Risk at level 30% of Yield Excess (YE); the greener the color, the better a fertilizer practice. Close colors indicate similarperformances.

time T, exemplified for soil types ITML840105 and ITML840101. For the soil ITML840105 (silt clay loam of low fertility), after a span of 20 years of experimentation, BCB selected the Practice 8, which was the best available one for this soil type (see Table 5), with an average proportion of 50%. ETC-5 also decided on the same practice, with an average proportion of 31%. For the soil ITML840101, BCB and ETC-5 similarly performed after 20 years of experimentation. For this soil type, BCB sampled the best available Practice 0 (Table 5) with an average proportion of 27%, ETC-5 selected the same practice with an average proportion of 26%. In the case of ETC-5, the constant and equal proportions of each management practice during the five first years seen in Figures 4b and 4d illustrate the equiproportional initial exploration phase used by the strategy.

Empirical CVaR of YE On average, farmers following the nitrogen fertilizer recommendations based on the BCB identification strategy had a higher empirical CVaR at 30% of YE than farmers following the recommendations from the ETC strategies, from the second year of the experiment onwards. Figure 5 (Left) shows the evolution of the CVaR at 30% of the YE for all cohorts (soil types) throughout the years (Equation 6). The difference in performance between BCB and ETC is relatively high during the initial years. For instance, at year 4, farmers following recommendations from the BCB identification strategy had a CVaR at 30% of YE of 318 kg/ha, compared to 168 kg/ha (47% less than BCB) and 74 kg/ha (77% less than BCB) for farmers following the recommendations from the ETC-3 and the ETC-5 strategies, respectively. Thus, BCB allowed to identify sooner the best available fertilizer practices and consequently further avoided low crop yield outcomes compared to ETC strategies. ETC strategies were adversely affected by their exploration phases during which all fertilizer practices were equiproportionally tested. In contrast, BCB had a continuously increasing empirical CVaR, during the whole duration of the experiment.

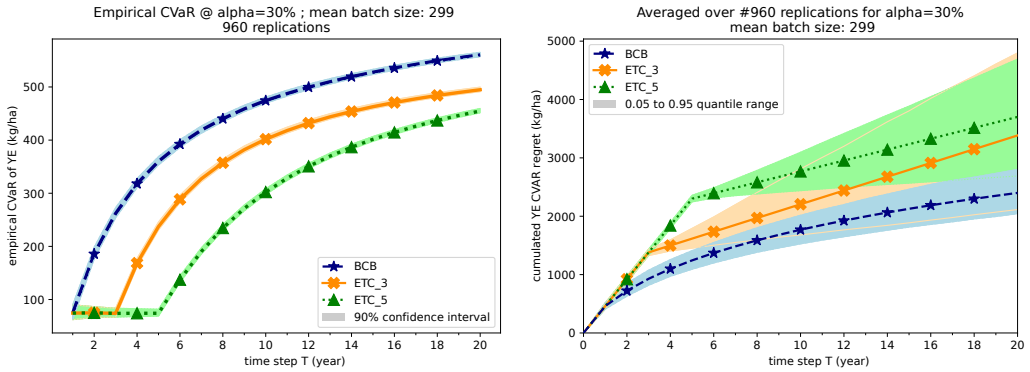


Figure 5: Empirical conditional Value-at-Risk (CVaR) at level 30% of maize Yield Excesses (YE) between $T = 0$ and $T = 20$ years (Left); and corresponding average cumulated regret over the virtual farmers' population (Right). Confidence intervals for the empirical CVaR were computed following [55].

CVaR regret For $\alpha = 30\%$, the BCB strategy outperformed ETC strategies, regardless of the number of years during which the strategy was applied. Figure 5 (Right) shows the evolution of the average cumulated regret for all cohorts throughout the years of the simulated experiment (Equation 9). The difference in performance between BCB and ETC increased for the whole duration of the experiment. After 20 years, farmers following recommendations from the BCB identification strategy experienced a mean cumulated regret of 2400 kg/ha, compared to 3385 kg/ha (41% more than BCB) and 3701 kg/ha (54% more than BCB) for farmers following the recommendations respectively from the ETC-3 and ETC-5 strategies. Consequently, farmers following BCB recommendations accumulated less regret compared to farmers following ETC recommendations. Furthermore, the variance of the cumulated regret (due to the different weather series in the experiments, for each season and each field trial, and the variability in cohorts each year) was smaller for BCB than for ETC, confirming that the BCB strategy was more robust (see quantile ranges in Figure 5) for this decision problem.

Sensitivity to the CVaR level α In Appendix F we present additional experiments for $\alpha = 50\%$ and for $\alpha = 100\%$ (expectation). For $\alpha = 50\%$ we obtain similar result to what we obtained with $\alpha = 30\%$, while for $\alpha = 100\%$ all algorithms seem to perform similarly for the time horizon considered. Nonetheless , BCB shows a smaller variance than both ETC-3 and ETC-5.

3.4 Discussion

In this section we discuss the results presented in previous sections and the perspectives they offer for the community of researchers in agronomy.

Benefits from an adaptive strategy The results presented in previous sections showcase the benefits of using tailored algorithms from the bandit literature to tackle multi-year multi-location on-farm trials similar to the one presented in this article. Indeed, we demonstrated both in theory and experiments that using our BCB algorithm, the better a crop management practice is, the more its representation among the tested practices grows over time. From a farmer's perspective, this means that the probability of testing sub-optimal recommendations decreases over time. This is in contrast with non-adaptive identification strategies, such as ETC that equi-proportionally recommend all crop management practices during the exploration phase. While the length of the exploration phase could be well-calibrated by chance, in general the ETC strategy is sub-optimal if the decision-maker does not have access to strong knowledge on the problem [39, Chapter 6]. On the contrary, BCB only requires to know the maximum observable reward. In agronomy, such knowledge is usually available through expert knowledge, either obtained through crop growth modeling or from field experiments conducted under optimal growing conditions [4]. Hence, the cost of the experiment to improve crop-management practices is likely to be reduced for the farmers when using bandit-based approaches with stronger guarantees. Another common method to generate crop management recommendation consists in the use of calibrated crop simulation models and scenario analyses [e.g. 30]. Although this

method has its limitations due to model uncertainty [61], it can be complementary to the bandit-based approach. For example, a set of candidate crop management practices can first be determined based on outcomes from crop modeling, and out of those, the true best option can then be identified in practice from field trials with the bandit algorithms for the experimental setup.

Definition of fertilization practices In this study, the values for the fertilizer practice attributes are likely not optimal, because our objective was on establishing an improved generic method for crop management experiments using bandit algorithms, rather than designing refined fertilizer recommendations. For an application in real field conditions, we recommend these attributes to be first estimated using existing expert knowledge and/or crop growth model simulations. The set of candidate practices can also comprise practices that are based on advanced methods, such as refined balance-based methods or machine learning-based methods for nitrogen fertilization [e.g., 44, 58]. More generally, the design of fertilizer recommendations must include experts, local agricultural extension officers and farmers themselves [14, 28]. Finally, it is also important to take into account that the quantity of mineral fertilizer a farmer can apply often depends on access to financial resources and markets [31].

Objective to maximize In Section 2.2, we advocated for the use of the CVaR of Yield Excess as a relevant performance metric for the problem considered in this study. In particular, the value of α allows to adjust the risk aversion level for a cohort of farmers, and the value of ANE_{ref} defines an invariant economic and environmental trade-off setting the boundaries of the performance of nitrogen fertilizer use. However, we did not evaluate fertilizer practices by their economic return that depends on many factors, such as fertilizer subsidies, fertilizer market price, application costs, and grain selling prices. Including those factors dramatically increases the complexity of the learning problem, and so does the required amount of data to identify the best practices (we provide more details in Appendix G). In this context, we must keep in mind the inherent constraints of modeling farmer's objectives and decisions, that always remains a proxy for real life situations and choices [42]. It is evident that farmers should play an active role in the formulation and validation of the objective to maximize, ensuring that mathematical terms are meaningfully translated into practical cases of crop management decision problems. Nonetheless, we emphasize that the `BCB` approach proposed in this paper is very flexible, and can be easily adapted to more sophisticated practices (e.g. including economic factors) and other performance metrics.

Limits and possible improvements The simulated crop management decision problem presented in this paper largely simplifies the experimental structure of multi-location, multi-year replicated field trials. First, weather time series are unlikely to be independent and identically distributed in the real world, because weather spatial correlations can be high, for instance in case of extreme weather events [52]. Second, within the same cohort, we assumed that farmers had identical soil type and maize cultivars, and were closely adhering to the assigned fertilizer practices. For the application of our methodology in real field conditions, variations in site conditions and other potential random effects should be properly considered, requiring some adaptation of `BCB`. Furthermore, information might be shared between cohorts by adopting a contextual/structured bandit approach [39], which would require to find a proper model of the rewards according to the characteristics of each soil type and the similarities between fertilizer practices. In the agronomy literature, mixed linear models [38] are typically used to account for random effects associated with the underlying structure of an experiment. While this additional complexity might improve the asymptotic performance of `BCB`, it is not clear that the gain could be observed within the limited time scale inherent to field trials. We leave the exploration of these directions for future work.

Finally, in the model simulations, we assumed average weather (rainfall, temperature) in southern Mali to remain the same throughout the 20 years of the experiment. Such hypothesis is unlikely in real conditions given climate change [e.g., 59]. Best available management practices are likely to change over time under climate change, as a response to the increasing occurrences of heat and water stress [2]. Such problem can be formalized as a non-stationary bandit problem [39]. To handle this, the `BCB` strategy can be equipped with a sliding window approach where the algorithms' decisions are based only on the most recent rewards, discarding older ones over time [24, 10].

4 Conclusion

Bandit algorithms aim at optimally balancing between exploration (gathering information) and exploitation (using the information to make good decisions) in uncertain decision problems with repeated choice between contending actions. In a simulated problem of testing fertilizer practices with virtual farmers, we compared the `BCB` bandit algorithm to the "intuitive strategy" of Explore-Then-Commit (`ETC`) in which the set of pre-defined practices are tested in an equiproportional way during a fixed number of years. During simulated field trials in southern Mali, `BCB` successfully minimized maize yield losses occured from testing worse performing fertilizer practices compared to the true best available practice, by up to 35% after 20 years. This novel approach opens up new perspectives as an alternative to the usual multi-year, multi-location on-farm trials. The bandit-based crop-management strategy shows promises in identifying best management practices in real field conditions, if variability in site conditions, possible correlations between site conditions, and other potential random effects are further considered in future works.

References

[1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26:1487–1503, 07 2002.

[2] M. Adam, D. S. MacCarthy, P. C. S. Traoré, A. Nenkam, B. S. Freduah, M. Ly, and S. G. Adiku. Which is more important to sorghum production systems in the sudano-sahelian zone of west africa: Climate change or improved management practices? *Agricultural Systems*, 185:102920, 2020.

[3] F. Affholder. Effect of organic matter input on the water balance and yield of millet under tropical dryland condition. *Field Crops Research*, 41(2):109–121, 1995.

[4] F. Affholder, C. Poeydebat, M. Corbeels, E. Scopel, and P. Tittonell. The yield gap of major food crops in family agriculture in the tropics: Assessment and analysis through field surveys and modelling. *Field Crops Research*, 143:106–118, 2013.

[5] S. Agrawal and N. Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1. JMLR Workshop and Conference Proceedings, 2012.

[6] S. Agrawal and N. Goyal. Further Optimal Regret Bounds for Thompson Sampling. In *Proceedings of the 16th Conference on Artificial Intelligence and Statistics*, 2013.

[7] S. Agrawal, W. M. Koolen, and S. Juneja. Optimal best-arm identification methods for tail-risk measures. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, 2021.

[8] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47, 2002.

[9] D. Baudry, R. Gautron, E. Kaufmann, and O. Maillard. Optimal thompson sampling strategies for support-aware cvar bandits. In *International Conference on Machine Learning*, pages 716–726. PMLR, 2021.

[10] D. Baudry, Y. Russac, and O. Cappé. On Limited-Memory Subsampling Strategies for Bandits. In *ICML 2021- International Conference on Machine Learning*, Vienna / Virtual, Austria, July 2021.

[11] A. Burnetas and M. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17(2), 1996.

[12] O. Cappé, A. Garivier, O.-A. Maillard, R. Munos, G. Stoltz, et al. Kullback–leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3):1516–1541, 2013.

[13] A. Cassel, S. Mannor, and A. Zeevi. A general approach to multi-armed bandits under risk criteria. In *Conference On Learning Theory*, pages 1295–1306. PMLR, 2018.

[14] M. Cerf and J.-M. Meynard. Les outils de pilotage des cultures: diversité de leurs usages et enseignements pour leur conception. *Natures Sciences Sociétés*, 14(1):19–29, 2006.

[15] M. Cerf and M. Sebillotte. Approche cognitive des décisions de production dans l'exploitation agricole [confrontation aux théories de la décision]. *Economie rurale*, 239(1):11–18, 1997.

[16] K. Dowd. *Measuring market risk*. John Wiley & Sons, 2007.

[17] K. J. Evans, A. Terhorst, and B. H. Kang. From data to decisions: helping crop producers build their actionable knowledge. *Critical reviews in plant sciences*, 36(2):71–88, 2017.

[18] B. Everitt and A. Skrondal. *The Cambridge dictionary of statistics*, volume 106. Cambridge University Press Cambridge, 2002.

[19] G. N. Falconnier, M. Corbeels, K. J. Boote, F. Affholder, M. Adam, D. S. MacCarthy, A. C. Ruane, C. Nendel, A. M. Whitbread, É. Justes, et al. Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-saharan africa. *Global change biology*, 26(10):5942–5964, 2020.

[20] B. Fosu-Mensah, D. MacCarthy, P. Vlek, and E. Safo. Simulating impact of seasonal climatic variation on the response of maize (zea mays l.) to inorganic fertilizer in sub-humid ghana. *Nutrient cycling in agroecosystems*, 94(2):255–271, 2012.

[21] Z. Gao, Y. Han, Z. Ren, and Z. Zhou. Batched multi-armed bandits problem. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[22] A. Garivier and E. Kaufmann. Optimal best arm identification with fixed confidence. In V. Feldman, A. Rakhlin, and O. Shamir, editors, *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, volume 49 of *JMLR Workshop and Conference Proceedings*, pages 998–1027. JMLR.org, 2016.

[23] A. Garivier, T. Lattimore, and E. Kaufmann. On explore-then-commit strategies. In D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 784–792, 2016.

[24] A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory*, pages 174–188. Springer, 2011.

[25] R. Gautron, O.-A. Maillard, P. Preux, M. Corbeels, and R. Sabbadin. Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture*, 200:107182, 2022.

[26] R. Gautron and E. J. Padrón González. gym-DSSAT - A crop model turned into a Reinforcement Learning environment, 3 2022.

[27] M. Getnet, M. Van Ittersum, H. Hengsdijk, and K. Descheemaeker. Yield gaps and resource use across farming zones in the central rift valley of ethiopia. *Experimental Agriculture*, 52(4):493–517, 2016.

[28] Z. Hochman and P. Carberry. Emerging consensus on desirable characteristics of tools to support farmers' management of climate risk in australia. *Agricultural Systems*, 104(6):441–450, 2011.

[29] G. Hoogenboom, C. Porter, K. Boote, V. Shelia, P. Wilkens, U. Singh, J. White, S. Asseng, J. Lizaso, L. Moreno, et al. The dssat crop modeling ecosystem. *Advances in crop modelling for a sustainable agriculture*, pages 173–216, 2019.

[30] E. Huet, M. Adam, B. Traore, K. Giller, and K. Descheemaeker. Coping with cereal production risks due to the vagaries of weather, labour shortages and input markets through management in southern mali. *European Journal of Agronomy*, 140:126587, 2022.

[31] T. S. Jayne, J. Govereh, M. Wanzala, and M. Demeke. Fertilizer market development: a comparative analysis of ethiopia, kenya, and zambia. *Food policy*, 28(4):293–316, 2003.

[32] J. Jones, G. Tsuji, G. Hoogenboom, L. Hunt, P. K. Thornton, P. Wilkens, D. Imamura, W. Bowen, and U. Singh. Decision support system for agrotechnology transfer: Dssat v3. *Understanding options for agricultural production*, pages 157–177, 1998.

[33] D. Jourdain, J. Lairez, B. Striffler, and F. Affholder. Farmers' preference for cropping systems and the development of sustainable intensification: a choice experiment approach. *Review of Agricultural, Food and Environmental Studies*, 101(4):417–437, 2020.

[34] M. Jourdan, R. Degenne, D. Baudry, R. de Heide, and E. Kaufmann. Top two algorithms revisited. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

[35] H. M. Kalaji, P. Dabrowski, M. D. Cetner, I. A. Samborska, I. Lukasik, M. Brestic, M. Zivcak, H. Tomasz, J. Mojski, H. Kociel, et al. A comparison between different chlorophyll content meters under nutrient deficiency conditions. *Journal of Plant Nutrition*, 40(7):1024–1034, 2017.

[36] N. Korda, E. Kaufmann, and R. Munos. Thompson sampling for 1-dimensional exponential family bandits. In *Advances in neural information processing systems*, pages 1448–1456, 2013.

[37] N. Korda, E. Kaufmann, and R. Munos. Thompson Sampling for 1-dimensional Exponential family bandits. In *Advances in Neural Information Processing Systems*, 2013.

[38] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[39] T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[40] B. B. Mandelbrot. The variation of certain speculative prices. In *Fractals and scaling in finance*, pages 371–418. Springer, 1997.

[41] P. Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Annals of Probability*, 18, 1990.

[42] R. L. McCown. Changing systems for supporting farmers' decisions: problems, paradigms, and prospects. *Agricultural systems*, 74(1):179–220, 2002.

[43] L. Menapace, G. Colson, and R. Raffaelli. Risk aversion, subjective beliefs, and farmer risk management strategies. *American Journal of Agricultural Economics*, 95(2):384–389, 2013.

[44] T. F. Morris, T. S. Murrell, D. B. Beegle, J. J. Camberato, R. B. Ferguson, J. Grove, Q. Ketterings, P. M. Kyveryga, C. A. Laboski, J. M. McGrath, et al. Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. *Agronomy Journal*, 110(1):1, 2018.

[45] V. Perchet, P. Rigollet, S. Chassang, and E. Snowberg. Batched bandit problems. In P. Grünwald, E. Hazan, and S. Kale, editors, *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, volume 40 of *JMLR Workshop and Conference Proceedings*, page 1456. JMLR.org, 2015.

[46] M. Piha. Optimizing fertilizer use and practical rainfall capture in a semi-arid environment with variable rainfall. *Experimental Agriculture*, 29(4):405–415, 1993.

[47] C. W. Richardson and D. A. Wright. WGEN: A model for generating daily weather variables. *ARS (USA)*, 1984.

[48] C. Riou and J. Honda. Bandit algorithms based on thompson sampling for bounded reward distributions. In *Algorithmic Learning Theory - 31st International Conference (ALT) 2012*, 2020.

[49] A. Ripoche, M. Crétenet, M. Corbeels, F. Affholder, K. Naudin, F. Sissoko, J.-M. Douzet, and P. Tittonell. Cotton as an entry point for soil fertility maintenance and food crop productivity in savannah agroecosystems–evidence from a long-term experiment in southern Mali. *Field crops research*, 177:37–48, 2015.

[50] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[51] A. Soltani and G. Hoogenboom. A statistical comparison of the stochastic weather generators WGEN and simmeteo. *Climate Research*, 24(3):215–230, 2003.

[52] J. B. Tack and M. T. Holt. The influence of weather extremes on the spatial correlation of corn yields. *Climatic Change*, 134(1-2):299–309, 2016.

[53] A. Tamkin, R. Keramati, C. Dann, and E. Brunskill. Distributionally-aware exploration for cvar bandits. In *NeurIPS 2019 Workshop on Safety and Robustness in Decision Making; RLDM 2019*, 2020.

[54] H. F. Ten Berge, R. Hijbeek, M. Van Loon, J. Rurinda, K. Tesfaye, S. Zingore, P. Craufurd, J. van Heerwaarden, F. Brentrup, J. J. Schröder, et al. Maize crop nutrient input requirements for food security in sub-saharan africa. *Global Food Security*, 23:9–21, 2019.

[55] P. Thomas and E. Learned-Miller. Concentration inequalities for conditional value at risk. In *International Conference on Machine Learning*, pages 6225–6233. PMLR, 2019.

[56] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[57] D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor, and S. Polasky. Agricultural sustainability and intensive production practices. *Nature*, 418(6898):671–677, 2002.

[58] J. Timsina, S. Dutta, K. P. Devkota, S. Chakraborty, R. K. Neupane, S. Bishta, L. P. Amgain, V. K. Singh, S. Islam, and K. Majumdar. Improved nutrient management in cereals using nutrient expert and machine learning tools: Productivity, profitability and nutrient use efficiency. *Agricultural Systems*, 192:103181, 2021.

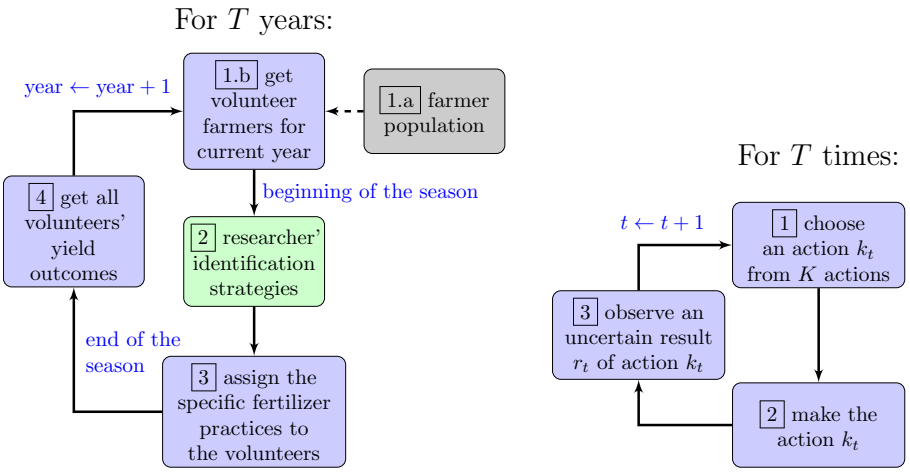
[59] B. Traore, K. Descheemaeker, M. T. Van Wijk, M. Corbeels, I. Supit, and K. E. Giller. Modelling cereal crops to assess future climate risk for family food self-sufficiency in southern mali. *Field Crops Research*, 201:133–145, 2017.

[60] B. Vanlauwe, J. Kihara, P. Chivenge, P. Pypers, R. Coe, and J. Six. Agronomic use efficiency of n fertilizer in maize-based systems in sub-saharan africa within the context of integrated soil fertility management. *Plant and soil*, 339(1):35–50, 2011.

[61] X. Yin, K. C. Kersebaum, C. Kollas, S. Baby, N. Beaudoin, K. Manevski, T. Palosuo, C. Nendel, L. Wu, M. Hoffmann, H. Hoffmann, B. Sharif, C. M. Armas-Herrera, M. Bindi, M. Charfeddine, T. Conradt, J. Constantin, F. Ewert, R. Ferrise, T. Gaiser, I. G. de Cortazar-Atauri, L. Giglio, P. Hlavinka, M. Lana, M. Launay, G. Louarn, R. Manderscheid, B. Mary, W. Mirschel, M. Moriondo, I. Öztürk, A. Pacholski, D. Ripoche-Wachter, R. P. Rötter, F. Ruget, M. Trnka, D. Ventrella, H.-J. Weigel, and J. E. Olesen. Multi-model uncertainty analysis in predicting grain n for crop rotations in europe. *European Journal of Agronomy*, 84:152–165, 2017.

A Experimental design

In this section we provide more detail on the design of the virtual crop-farming experiment presented in the paper.



(a) Best fertilizer practice identification process. At the start of the season, a number of farmers ($n = 250$ to 350) volunteer 1.b to test fertilizer practices recommended by the researcher following an identification strategy 2, 3. At the end of each season, the farmers share their yield outcomes with the experts 4. The experts will use these results to improve their fertilizer recommendations for the next growing season. The process is repeated for a total number of $T = 20$ years.

(b) Canonical bandit problem. For T times, an agent sequentially makes decisions on an action k_t from the set $\{1, \cdots, K\}$ of possible actions 1. After making the action k_t 2, the agent observes an uncertain result r_t 3. This result is sampled from a fixed distribution, unknown to the agent, which corresponds to the effect of action k_t.

Figure 6: Schematic representation of the ensemble best fertilization identification process (a) and the canonical bandit problem (b).

Nitrogen fertilizer practices. Ten nitrogen fertilizer practices were considered as recommendations in the virtual experiment (see Table 2). Practices 0 to 7 represent the following set of split fertilizer practice for a total amount of 135 kg N/ha applied:

- Two split applications (Practice 0): 15 kg N/ha at 15 days after planting (DAP), and 120 kg N/ha at 30 DAP.
- Three split applications (Practice 4): 15 kg N/ha at 15 DAP, 60 kg N/ha at 30 DAP and 60 kg N/ha at 45 DAP.
- Split applications according to the rainfall amount (Practices 2, 3 and 6, 7): 2nd and 3rd top-dressing applications only if the cumulated rainfall amount from the start of the season to 30 DAP exceeds the 30th percentile of historical rainfall i.e. 200 mm.
- Split applications according to plant nitrogen status (Practices 1, 3 and 5, 7): 2nd and 3rd top-dressing applications only if the simulated nitrogen stress factor (`NSTRES` in DSSAT, see below) exceeds 0.2 (0 standing for no stress, 1 for maximal stress) at 30 DAP, hereby mimicking the use of a portable chlorophyll meter to monitor plant nitrogen status [e.g. 35].

Split fertilizer applications were considered in order to adjust the amount of nitrogen applied to the likely crop demand as the season develops. This adjustment can rely on factors such as weather conditions and the crop performance [46].

Practice 8 corresponds to the recommended fertilizer application for maize (70 kg N/ha) in the study region, which was determined based on model simulations [30], i.e. the average of the nitrogen fertilizer rates that were expected to result in maximum positive return on fertilizer investment [27].

Practice 9 (180 kg N/ha) corresponds to a nitrogen fertilizer application that is likely excessive. In our model simulations (see below), the type of nitrogen fertilizer applied for all practices was set as ammonium nitrate broadcasted on the soil surface.

Table 1: Main properties of the soil types of the fields of farmers growing maize in Koutiala, Mali [2].

Soil name	**Texture**	SLDR	SLOC	SLDP	AWCH	pH	**Prop.**
ITML840101	clay loam	0.60	0.20	110	115	5.7	7
ITML840102	loam	0.60	0.45	100	124	5.5	9
ITML840103	silty loam	0.60	0.27	160	98	6.5	21
ITML840104	silty clay loam	0.25	0.70	105	101	5.5	4
ITML840105	silty clay loam	0.40	0.38	120	108	5.8	24
ITML840106	loam	0.60	0.30	110	115	5.7	27
ITML840107	silty clay loam	0.25	0.60	105	101	5.5	8

'SLDR': soil drainage rate (fraction/day); 'SLOC': soil organic matter (g C/ 100 g soil) in the 0-30 cm topsoil; 'SLDP': soil depth (cm); 'AWCH': soil available water-holding capacity (mm); 'pH' is the pH in water; 'Prop' stands for the percentage of each soil type present in the study area.

B Maize simulations

Simulator gym-DSSAT is a modification of the DSSAT crop simulator [29] to allow a user to read daily internal DSSAT states and, accordingly, to be able to take fertilization decisions on a daily basis. Evidence of the reliability of DSSAT in simulating maize responses to different nitrogen fertilization practices under the conditions of southern Mali is provided by [19, 30]. The soils (and associated model parameters) we used for simulations are the same as the ones used by [2] who calibrated DSSAT for sorghum under different plant densities and nitrogen fertilizer practices in southern Mali. For each soil type (Table 1) that was parameterized in DSSAT (soil parameter files *.SOL), each simulated maize grain yield value is a sample of the yield response distribution for the considered fertilizer practice. This response distribution is the result of weather variability, generated in our study by the stochastic weather generator WGEN [47, 51], which was parameterized using the 47-year weather records from the N'Tarla agricultural research station of the Institute of Rural Economics (12°35' N, 5°42' W, 302 m.a.s.l.), about 30 km from the city of Koutiala [49]. The 'Sotubaka' maize cultivar (original name 'Suwan 1 SR'), from the DSSAT default cultivar list) was used for all model simulations as a representative of the maize varieties grown in southern Mali. This cultivar was parameterized by the DSSAT team for the conditions of southern Mali [32]. Planting date was

Table 2: Maize nitrogen fertilizer practices for maize considered during the virtual experiment in Koutiala, Southern Mali. The inclusion of rainfall and plant nitrogen stress as threshold factors in the fertilizer practice is denoted by "Yes" or "No".

Index	Max. # of fertilizer applications	Rainfall threshold	NSTRES[1] threshold	Application at 15 DAP (kgN/ha)	Application at 30 DAP (kgN/ha)	Application at 45 DAP (kgN/ha)	Max. total amount applied (kgN/ha)
0	2	No	No	15	120	0	135
1	2	No	Yes	15	120	0	135
2	2	Yes	No	15	120	0	135
3	2	Yes	Yes	15	120	0	135
4	3	No	No	15	60	60	135
5	3	No	Yes	15	60	60	135
6	3	Yes	No	15	60	60	135
7	3	Yes	Yes	15	60	60	135
8	2	No	No	23	0	47	70
9	3	No	No	60	60	60	180

defined by an automatic rule (see Table 4) depending on soil water conditions. At the start of the simulations, the initial soil mineral nitrogen content was set to a fixed, depending on the soil type as in [2]. Still, the variability of the weather from the beginning of the simulation to the occurrence of the automatic planting (itself dynamic) induced a variable initial soil mineral nitrogen content at the planting date for each simulation. Water and nitrogen stresses were simulated but yield reduction through pests and diseases were not considered, neither was weed competition.

Model parameters The cultivation scenarios were based on the the conditions found in Southern Mali. The soils came from [2] who compiled and supplemented with survey data the soils found in the literature for the location of Koutiala, Mali. The data of [2] included soils' depth, texture, water capacity, bulk density, organic matter content, pH and initial mineral nitrogen content. Soil characteristics and proportions in the population were summarized in Table 1, based on [2]. During the simulations, the weather times series were generated using the WGEN weather model [see 47, 51]. WGEN had been parameterized on 47-year-long historical daily weather records from a weather station located in N'Tarla found in [49], which was located about 20 km from Koutiala ; these historical weather records were the best available. The cultivars used in the simulation and its parameterization in DSSAT are presented in Table 3 ; this cultivars comes with DSSAT default data and was representative of the cultivars used in Mali. The cultivars were already calibrated based on experiments carried out in Mali. The simulations were initiated on Day Of Year (DOY) 140 and the planting is automatically performed in a window ranging from DOY 155 to 185 ; we specified the parameters of the automatic planting with Table 4. For each soil, the initial soil nitrogen content was set according to the values found in [2]. The soil water content was set to crop lower limit, as a result of the end of the dry season at the usual planting dates. Because the simulations were initiated prior to planting date and because the weather was stochastically generated, the soil nitrogen mineral and water contents were uncertain at planting time. Each simulation was performed independently from the previous ones. At the beginning of the experiment, all the soils described in Table 1 were randomly distributed amongst the initial group of farmers following the proportions provided in Table 1. Figure 7 shows the simulated yield distributions for `ITML840104` and `ITML840105` soils.

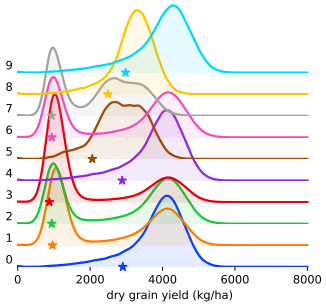
Table 3: Maize cultivar parametrization in DSSAT

name	**ecotype**	**P1**	**P2**	**P5**	**G2**	**G3**	**PHINT**
Sotubaka	`IB0001`	300.0	0.520	930.0	500.0	6.00	38.90

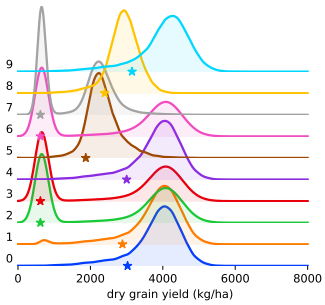
Table 4: Automatic planting parametrization in DSSAT. PFRST: Starting date of the planting window; PLAST: End date of the planting window; PH2OL: Lower limit on soil moisture for automatic planting; PH2OU: Upper limit on soil moisture for automatic planting; PH2OD: Depth to which average soil moisture is determined for automatic planting; PSTMX: Maximum temperature of planting; PSTMN: Minimum temperature of planting.

PFRST (DOY)	155
PLAST (DOY)	185
PH2OL (%)	40
PH2OU (%)	100
PH2OD (cm)	30
PSTMX (°C)	40
PSTMN (°C)	10

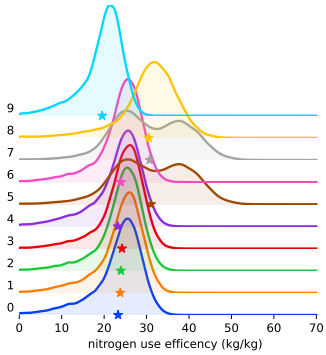
In the following Table 5 we present the numerical value associated with the optimal fertilizer practice for each soil type considered in this study. For the corresponding best available nitrogen fertilizer practice π^*, we define N^{π^*}: quantity of nitrogen fertilizer applied; $\text{CVaR}_{30\%}(X)$: conditional Value-at-Risk of X of level 30% (Section 2.2); $\bar{X}$: mean value of X; Y^{π^*}: maize grain yield; ANE^{π^*}: Agronomic Nitrogen use Efficiency; YE^{π^*}: Yield Excess (Section 2.2); values in parentheses indicate standard deviations.



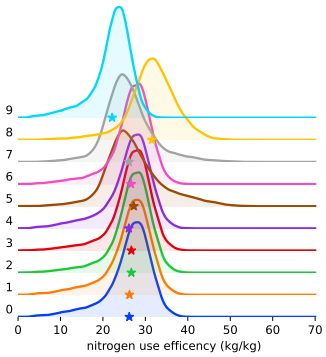
(a) Yield distributions for soil ITML840104. Stars represent the CVaR at level 30%.



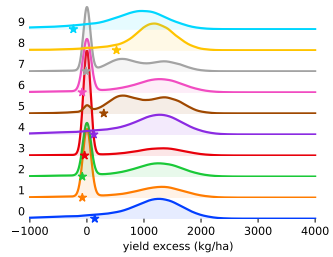
(b) Yield distributions for soil ITML840105. Stars represent the CVaR at level 30%.



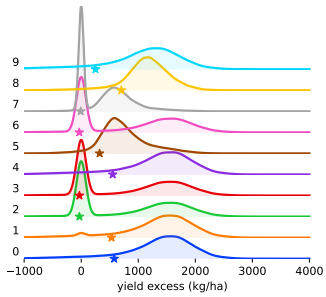
(c) Agronomic Nitrogen Efficiency (ANE) distributions for soil ITML840104. Stars represent the mean value.



(d) Agronomic Nitrogen Efficiency (ANE) distributions for soil ITML840105. Stars represent the mean value.



(e) Yield Excess (YE) distributions for soil ITML840104 with ANE_{ref}=15 kg grain/kg N. Stars represent the CVaR at level 30%.



(f) Yield Excess (YE) distributions for soil ITML840105 with ANE_{ref}=15 kg grain/kg N. Stars represent the CVaR at level 30%.

Figure 7: Simulated impact of maize fertilizer practices on grain yield, Agronomic Nitrogen use Efficiency (ANE), Yield Excess (YE) for 10^5 hypothetical years using a weather generator. Maize cultivar was the same for all simulations. Practices indexes are indicated on the left-hand side of each sub-figure.

Table 5: Statistics of the best available nitrogen fertilizer practices for each of the soil types presented in Table 1.

soil	π^*	$\bar{N}^{\pi^*}$ (kg/ha)	$CVaR_{30\%}$ (kg/ha)	$\bar{Y}^{\pi^*}$ (kg/ha)	$\bar{ANE}^{\pi^*}$ (kg/kg)	$CVaR_{30\%}$ (kg/ha)	$\bar{YE}^{\pi^*}$ (kg/ha)
ITML840101	**0**	120.0 (1.0)	3091	3874 (666)	30.0 (5.4)	**1032**	1795 (651)
ITML840102	**8**	69.8 (4.0)	2391	3150 (653)	33.2 (7.5)	**652**	1270 (529)
ITML840103	**8**	70.0 (0.4)	2539	3152 (526)	34.4 (6.8)	**808**	1356 (475)
ITML840104	**8**	69.9 (2.7)	2533	3339 (682)	31.7 (8.1)	**500**	1169 (565)
ITML840105	**8**	70.0 (1.2)	2467	3127 (570)	34.2 (7.3)	**757**	1346 (508)
ITML840106	**0**	120.0 (1.2)	3132	3945 (695)	28.9 (5.5)	**900**	1667 (660)
ITML840107	**8**	69.9 (2.7)	2472	3247 (659)	32.5 (8.0)	**565**	1226 (559)

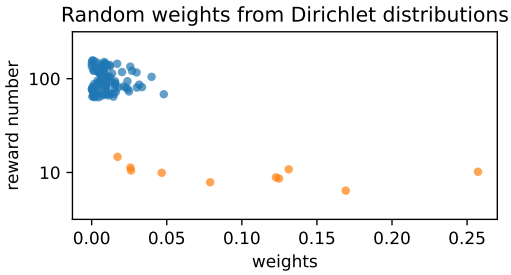


Figure 8: Examples of weights sampled from Dirichlet distributions during BCB execution, respectively for 10 and 100 rewards. The greater the number of rewards, the less variance the weights show. The variance of weights is related to the noise level in the computation of the empirical CVaR of BCB.

Practical implementation Since each call to DSSAT is costly, we simulated 10^5 times the maize grain yield responses to a given fertilizer practice for the different soil types prior to the experiment, which corresponds to 10^5 hypothetical growing seasons for each setting. We assume that these samples represent their respective distribution well enough so that sampling from them uniformly at random is close to sampling from the true underlying distribution. Since we report results from repeated simulations in the experiment section 3, this considerably speeds-up the cost to reproduce the experiments.

Furthermore, these samples were used i) to ensure that simulated maize yield responses were in realistic ranges, ii) to evaluate the complexity of the decision problem, and iii) to determine best nitrogen fertilizer practices whilst analyzing the performance of the crop management identification strategies. The samples were *not* provided to the algorithms prior to their learning.

C Algorithms

C.1 Details about BCB

In Algorithm 2, we provide the detailed pseudo-code of BCB (BCB). As shown by Figure 8, the higher the number of collected rewards, the less the weights sampled from Dirichlet distributions exhibit variance. This variance directly relates to the noise introduced in the computation of the score of the different available actions.

Remark C.1 (First season)**.** *Algorithm 2 is well defined for the first season as without data all CVaRs will be equal to the maximum observable result, making the algorithm choose each option arbitrarily at random. On average, each option will be equally explored. Note that we could replace this step by an equi-proportional exploration step (similar to Explore-Then-Commit, see C.2) without changing the theoretical properties of our algorithm. Furthermore, the decision maker could also include any additional results collected before the experiment (if the practices has already been tested for some time) in the initialization of the algorithm.*

C.2 Explore-Then-Commit (ETC)

We provide the pseudo-code of the Explore-Then-Commit (ETC) strategy with algorithm 3. The noise introduced by random weights and the presence of the maximum observable results in the histories manage the exploration/exploitation dilemma. BCB will favor fertilizer practices with higher CVaR compared to the others. But, the algorithm will still prevent the under-exploration of fertilizer practices by choosing them with a proper probability, even if e.g. poor YE have been observed due to rare unfavorable weather events. Indeed, with the extra randomness introduced by the random weighting of rewards, poor rewards may be re-weighted by smaller weights compared to higher rewards, yielding a good score. The amount of noise introduced by the random weights sampled from the Dirichlet distribution is related to variance of these random weights. The greater the number of rewards, the lesser the variance and consequently the lesser the noise (Figure 8). Thereby, the more a fertilizer practice was tried by the algorithm, the closer its score gets to the true CVaR of rewards. The presence of the maximum observable YE acts as an “optimistic bonus" in the computation of

Algorithm 2 BCB: identification strategy at cohort level (detailed)

Input: Level α, horizon T, K options, upper bounds $B_1, \dots, B_K$, $\mathcal{F}^c$ the set of all farmers in the cohort

Init.: $\forall k \in \{1, ..., K\}$: $\mathcal{X}_k = \{B_k\}$, $N_k = 0$; $\mathcal{F}_1^c = \{f_1, \cdots, f_{n_1}\}$; $t = 1$; $\mathcal{A}_1 = \{\emptyset\}$

// Beginning of first season

for $f \in \mathcal{F}_1^c$ **do**

  Randomly assign a crop management option $a \in \{1, \dots, K\}$ to the farmer f

  $\mathcal{A}_1 = \mathcal{A}_1 \cup \{a\}$

end

// End of first season

for $(a, f) \in (\mathcal{A}_1, \mathcal{F}_1^c)$ **do**

  Receive the result of the option a from farmer f: $r_{f,a}$

  Update $\mathcal{X}_a = \mathcal{X}_a \cup \{r_{f,a}\}$, $N_a = N_a + 1$

end

for $t \in \{2, \dots, T\}$ **do**

  // Beginning of season t

  Get $\mathcal{F}_t^c = \{f_1, \cdots, f_{n_t}\}$; // the set of farmers of the same cohort to provide recommendations

  for $k \in \{1, \dots, K\}$ **do**

    Update the empirical CVaR of action k: $\hat{c}_{k,t-1} = \widehat{C}_\alpha(\mathcal{X}_k)$

  end

  for $f \in \mathcal{F}_t^c$ **do**

    Update the empirical regret of farmer f: $l_{f,t-1} = \widehat{R}_f^\alpha(t-1)$

  end

  $\mathcal{A}_t = \{\emptyset\}$; // the set of recommendations to provide to the farmers

  for $f \in \mathcal{F}_t^c$ **do**

    for $k \in \{1, \dots, K\}$ **do**

      Draw $\omega_k = \{w_1, \cdots, w_{N_k}\} \sim \mathcal{D}_{N_k}$; // random weights drawn from a Dirichlet distribution of concentration parameter $\underbrace{(1, \cdots, 1)}_{N_k \text{ times}}$

      Search j the maximum index such that $\sum_{i=1}^{j} w_i \leq \alpha$

      Compute $\tilde{c}_k = x_j - \frac{1}{\alpha} \sum_{i=1}^{N_k} w_i \max(x_j - x_i, 0)$; // Compute the "noisy" empirical CVaR for action k

    end

    $a = \text{argmax}_{k \in \{1,\dots,K\}} \tilde{c}_k$

    $\mathcal{A}_t = \mathcal{A}_t \cup \{a\}$

  end

  for $(a, f) \in (\mathcal{A}_t, \mathcal{F}_t^c)$ **do**

    Assign action a to farmer f

  end

  // End of season t

  for $(a, f) \in (\mathcal{A}_t, \mathcal{F}_t^c)$ **do**

    Receive result of action a from farmer f: $r_{f,a}$

    Update $\mathcal{X}_a = \mathcal{X}_a \cup \{r_{f,a}\}$, $N_a = N_a + 1$

  end

end

the scores, encouraging exploration even for sub-optimal practices, as it raises up their initial values when few rewards have been observed.

Algorithm 3 ETC: identification strategy at cohort level

Input: Level α, horizon T, K options, $\mathcal{F}^c$ the set of all farmers in the cohort, t_{trials} the number of years of trials

Init.: $\forall k \in \{1, \cdots, K\} : N_k = 0$

// Do trials during t_{trials} years

for $t \in \{1, \cdots, t_{trials}\}$ **do**

  // Beginning of the season t

  Get $\mathcal{F}_t^c = \{f_1, \cdots, f_{n_t}\}$; // get the farmers willing to participate

  $\mathcal{A}_t = \{\emptyset\}$

    Fill $\mathcal{A}_t$ by uniformly distributing the K options to the farmers in $\mathcal{F}_t^c$

    // End of the season t

  for $(a, f) \in (\mathcal{A}_t, \mathcal{F}_t^c)$ **do**

    Receive the result of the option a from farmer f: $r_{f,a}$

    Update $\mathcal{X}_a = \mathcal{X}_a \cup \{r_{f,a}\}, N_a = N_a + 1$

  end

end

for $k \in \{1, \ldots, K\}$ **do**

  Compute the empirical CVaR of action k: $\hat{c}_{k,t-1} = C_\alpha(\mathcal{X}_k)$

end

$a_{\max} = \text{argmax}_{k \in \{1,\ldots,K\}} \hat{c}_k$; // get the action that best performed during trials

// After trial phase, always recommend the action that best performed during trials

for $t \in \{t_{trials} + 1, \cdots, T\}$ **do**

  // Beginning of the season t

  Get $\mathcal{F}_t^c = \{f_1, \cdots, f_{n_t}\}$

  for $f \in \mathcal{F}_t^c$ **do**

    Assign option $a_{\max}$ to the farmer f

  end

  // End of the season t

  for $f \in \mathcal{F}_t^c$ **do**

    Receive the result of the option $a_{\max}$ from farmer f: $r_{f,a_{\max}}$

    Update $\mathcal{X}_{a_{\max}} = \mathcal{X}_{a_{\max}} \cup \{r_{f,a_{\max}}\}, N_{a_{\max}} = N_{a_{\max}} + 1$

  end

end

D Theoretical Analysis

This section is devoted to the theoretical analysis of the BCB algorithm, in particular to the proof of Theorem 3.1. It is mostly adapted from the analysis of CVTS in [9], and shows that the problem of learning with batched data of finite upper bounded size is no harder than the pure online learning problem considered in the original paper. The proof of the result comes from deriving a regret upper bound for BCB, and comparing it to the regret upper bound of CVTS. The result presented in Theorem 3.1 is simplified, in order to convey the main idea that BCB preserves the theoretical guarantees of CVTS and that the batches only incur a second-order term for problem-dependent guarantees. We prove it using the results presented in the following Theorem D.1 and Theorem D.2. Before that, we recall from [9] the definition of the following quantity, defining the asymptotic optimality of CVaR-bandit algorithms.

Definition. *For any distribution F with bounded support $[O, B]$, for $B > 0$, and any $c \in [0, B]$, we define*

$$\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}(F, c) = \inf_{G:\, supp(G) \subset [0,B],\, CVaR_\alpha(G) > c} \text{KL}(F, G) \ .$$

Theorem 1 from [10] states how this quantity determines asymptotic optimality in CVaR bandits, in an analogous way to the Burnetas & Katehakis lower bound in standard bandits [11].

Theorem D.1 (α-CVaR Regret of BCB)**.** *Consider a bandit problem $(F_1, \dots, F_K) \in \mathcal{F}^K$, with respective CVaR$_\alpha$ denoted by $(c_1, \dots, c_K)$ with $c_1 = argmax_{k=1,\dots,K} c_k$. Assume that BCB runs for T seasons, and that at each season the size of the batch is $n_T \in [1, F]$, and $F \in \mathbb{N}$. Then, for any $\varepsilon > 0$ small enough there exists some $\varepsilon_1 > 0, \varepsilon_2 > 0$ such that the regret of BCB satisfies :*

$$\mathcal{R}_T^\alpha \leq \sum_{k=2}^{K} \Delta_k^\alpha \left(m_T^k + F + 2F \frac{e^{-2m_T^k \varepsilon_1^2}}{1 - e^{-2\varepsilon_1^1}} + C_{1,\varepsilon_2}^\alpha \right) ,$$

where $m_T^k = \frac{\log(FT)}{\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}(F_k, c_1) - \varepsilon}$ and C_{1,ε_2} is a constant depending only on the distribution F_1, the family $\mathcal{F}$ and ε_2.

Then, theorem D.2 is adapted from Theorem 3 in [9], presented in a form that simplifies comparison with the previous theorem.

Theorem D.2 (α-CVaR Regret of B-CVTS with time horizon S_T (adapted from Theorem 3 in [9]))**.** *Consider a bandit problem $(F_1, \dots, F_K) \in \mathcal{F}^K$, with respective CVaR$_\alpha$ denoted by $(c_1, \dots, c_K)$ with $c_1 = argmax_K c_k$. Consider a number of data collected $S_T = \sum_{t=1}^{T} n_t$. Then, for any $\varepsilon > 0$ small enough there exists some $\varepsilon_1 > 0, \varepsilon_2 > 0$ such that the CVaR-regret of B-CVTS satisfies*

$$\mathcal{R}_T^\alpha \leq \sum_{k=2}^{K} \Delta_k^\alpha \left(m_T^k + 2 \frac{e^{-2m_T^k \varepsilon_1^2}}{1 - e^{-2\varepsilon_1^1}} + C_{1,\varepsilon_2}^\alpha \right) ,$$

where $m_T^k = \frac{\log(FT)}{\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}(F_k, c_1) - \varepsilon}$ and C_{1,ε_2} is a constant depending only on the distribution F_1, the family $\mathcal{F}$ and ε_2.

We now prove Theorem 3.1 by comparing these two results.

Proof of Theorem 3.1. We remark that the two results admit the same first order term (logarithmic in T), which is already sufficient to show that BCB is asymptotically optimal because this result holds for CVTS. Similarly, the constant $C_{1,\varepsilon_2}^\alpha$ appears in the two bounds. Then, we observe that the second term in the bound of Theorem D.1 is the one that we introduced in Theorem 3.1. We finally observe that the remaining term in both upper bounds decreases to 0, since it contains a $\log(T)$ in the exponent in the numerator. This is still true when multiplying this term by F in Theorem D.1. □

In the proof, we used that $S_T \leq FT$ and that the upper bound F on the number of farmers is constant (i.e do not depend on the time). We now detail the proof of Theorem D.1, which mirrors the proof of Theorem D.2 in [9] with some additional ingredients needed to adapt it to the batch setting.

Proof of Theorem D.1. As in the proof of [9] we will decompose the expected number of pulls of each sub-optimal arm inside the cohort according to several possible events, corresponding to "good" scenarios (the empirical distributions accurately reflect the true distributions) and "bad" ones (the empirical distributions give a wrong idea of the true performance of some arms) for the trajectory of the bandit algorithms. We denote by T the number of seasons in the experiments and n_t the number of farmers at each season t for this cohort, and by F the total number of farmers available for the experiment. Then, the expected number of pulls of arm k during the total duration of the experiment inside the cohort is

$$\mathbb{E}[N_k(T)] = \mathbb{E}\left[\sum_{t=1}^{T} \sum_{f=1}^{n_t} \mathbb{1}(A_{t,f} = k) \right] ,$$

where $A_{t,f}$ denotes the recommendation to farmer f at season t.

The first step of the proof consists in considering the number of pulls of k when its sample size is larger (resp. smaller) than some fixed threshold m_T, that we will specify later.

$$\begin{aligned}\mathbb{E}[N_k(T)] =& \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k)\right]\\ \leq& \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\leq m_T)\right]\\ &+\mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_T}\mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T)\right]\end{aligned}$$

We now consider the first term and introduce the random variable $\tau = \{\sup_{t\leq T} : N_k(t-1)\leq m_T\}$. By construction, τ is the last season for which the total number of observations for arm k is smaller than m_T. Using the basic properties of τ we obtain that:

$$\begin{aligned}\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\leq m_T) \leq& \sum_{t=1}^{\tau}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\leq m_T)\\ &+\sum_{t=\tau+1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\leq m_T)\\ \leq& N_k(\tau)+\sum_{f=1}^{n_{\tau+1}}\mathbb{1}(A_{\tau,f}=k)\\ \leq& m_T+F\end{aligned}$$

As this result does not depend on the value of τ, we can then obtain:

$$\mathbb{E}[N_k(T)]\leq m_T+F+\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T)\right]}_{A}.$$

At this step, the only difference with the purely sequential bandit problem is the additional F. We now consider the term A, that we further analyze according to three events: (1) the empirical distribution of arm k is not close to its true distribution, (2) the empirical distribution of arm k is close to its true distribution but the "noisy" CVaR computed for arm k over-estimates its true CVaR, and (3) the "noisy" CVaR computed for the optimal arm 1 under-estimates its true CVaR. Classically in bandit analysis, we decompose the number of pulls of arm k according to these three events, as at least one of them must be true when $A_{t,f}=k$ holds, that is

$$\{A_t=k\}\subset\{F_{k,t-1}\notin\mathcal{B}_{\varepsilon_1}(F_k)\}\cup\{F_{k,t-1}\in\mathcal{B}_{\varepsilon_1}(F_k), \widetilde{c}_{k,t,f}\geq c_1-\varepsilon_2\}\cup\{\widetilde{c}_{1,t,f}\leq c_1-\varepsilon_2\},$$

where $\mathcal{B}_{\varepsilon_1}(F_k)$ is an ε_1-Levy ball around F_k, and $\varepsilon_1, \varepsilon_2$ are two small positive constants. This leads to

$$
\begin{aligned}
A \leq & \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T, F_{k,t-1}\notin\mathcal{B}_{\varepsilon_1}(F_k))\right]}_{A_1} \\
& + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T, F_{k,t-1}\in\mathcal{B}_{\varepsilon_1}(F_k), \widetilde{c}_{k,t,f}\geq c_1-\varepsilon_2)\right]}_{A_2} \\
& + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T, \widetilde{c}_{1,t,f}\leq c_1-\varepsilon_2)\right]}_{A_3} .
\end{aligned}
$$

Upper bounding A_2 Denoting by $\widehat{F}_{k,n}$ the empirical distribution of arm k after a total number of *pulls* n (instead of after season t), we obtain

$$
\begin{aligned}
A_1 &\coloneqq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T, F_{k,t-1}\notin\mathcal{B}_{\varepsilon_1}(F_k))\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(N_k(t-1)\geq m_T, F_{k,t-1}\notin\mathcal{B}_{\varepsilon_1}(F_k))\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k)\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{n=m_T}^{T}\mathbb{1}(N_k(t-1)=n, F_{k,t-1}\notin\mathcal{B}_{\varepsilon_1}(F_k))\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k)\right] ,
\end{aligned}
$$

with a union bound on the number of pulls. Under $N_k(t-1)=n$ it holds that $F_{k,t-1}=\widehat{F}_{k,n}$, and so we can further write that

$$
\begin{aligned}
A_1 &\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{n=m_T}^{T}\mathbb{1}(N_k(t-1)=n, \widehat{F}_{k,n}\notin\mathcal{B}_{\varepsilon_1}(F_k))\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k)\right] \\
&\leq \mathbb{E}\left[\sum_{n=m_T}^{T}\mathbb{1}(\widehat{F}_{k,n}\notin\mathcal{B}_{\varepsilon_1}(F_k))\sum_{t=1}^{T}\sum_{f=1}^{n_t}\mathbb{1}(A_{t,f}=k, N_k(t-1)=n)\right] \\
&\leq F\mathbb{E}\left[\sum_{n=m_T}^{T}\mathbb{1}(\widehat{F}_{k,n}\notin\mathcal{B}_{\varepsilon_1}(F_k))\right] \\
&= F\sum_{n=m_T}^{+\infty}\mathbb{P}(F_{k,n}\notin\mathcal{B}_{\varepsilon_1}(F_k))
\end{aligned}
$$

Finally, using the Dvoretzky–Kiefer–Wolfowitz inequality [41] we obtain:

$$\leq F \sum_{n=m_T}^{+\infty} 2e^{-2n\,\varepsilon_1^2}$$
$$\leq \frac{2Fe^{-2m_T\,\varepsilon_1^2}}{1-e^{-2\,\varepsilon_1^2}} \,.$$

This upper bound holds for any choice of m_T, ε_1, and we remark that if $m_T \to +\infty$ then $A_1 \to 0$.

Upper bounding A_2 The term A_2 is then handled with similar tricks, and the arguments used in [9].

$$A_2 \coloneqq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t} \mathbb{1}(A_{t,f}=k, N_k(t-1)\geq m_T, F_{k,t-1}\in\mathcal{B}_{\varepsilon_1}(F_k), \widetilde{c}_{k,t,f}\geq c_1-\varepsilon_2)\right]$$
$$\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{F} \mathbb{1}(N_k(t-1)\geq m_T, F_{k,t-1}\in\mathcal{B}_{\varepsilon_1}(F_k)) \times \mathbb{P}\left(\widetilde{c}_{k,t,f}\geq c_1-\varepsilon_2\,|\mathcal{F}_t\right)\right] \,,$$

where $\mathcal{F}_t$ is the canonical filtration, so the probability is obtained conditioning on the data observed before the beginning of the round. Using the the continuity of $\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}$ in its two arguments as proved in [7], we obtain that for any $\varepsilon > 0$ small enough there exist some $\varepsilon_1, \varepsilon_2$ such that:

$$A_2 \leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{F} \mathbb{1}(A_{t,f}=k, N_k(t-1)=n, F_{k,t-1}\in\mathcal{B}_{\varepsilon_1}(F_k)) e^{-m_T\left(\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}(F_k,c_1)-\varepsilon\right)}\right]$$
$$\leq F\times T\times e^{-m_T\left(\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}(F_k,c_1)-\varepsilon\right)} \,.$$

As we did not specify the choice of $\varepsilon_1, \varepsilon_2$ already we simply require them to be small enough to satisfy this condition. Then, we can calibrate m_T as

$$m_T = \frac{\log(T)+\log(F)}{\mathcal{K}_{\inf}^{\alpha,\mathcal{D}}(F_k,c_1)-\varepsilon} \,.$$

Furthermore, with this choice m_T will become the main term in the regret upper bound when T becomes large enough.

Upper bounding A_3 The final term is the one that leading to the most complicated part of the analysis in [9]. Fortunately, the batch setting will have no impact on this part, so we can directly reuse the results provided in this paper.

Indeed, we can re-write A_3 to make it equivalent to the corresponding term in the purely sequential problem:

$$A_3 = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{f=1}^{n_t} \mathbb{1}(\widetilde{c}_{1,t,f}\leq c_1-\varepsilon_2)\right] = \mathbb{E}\left[\sum_{r=1}^{S_T} \mathbb{1}(\widetilde{c}_1(r)\leq c_1-\varepsilon_2)\right] \,,$$

where in the second term we count the number of recommendations provided by the algorithm, assigning those in the same batch an arbitrary order, $\widetilde{c}_1(r)$ is then the noisy CVaR computed for arm 1 for this specific round. Furthermore, we write $S_T = \sum_{t=1}^{T} n_t \leq FT$. In [9], the authors obtain a constant upper bound for this term, depending only on ε_2 (and the upper bound of the support), and in particular not depending on the exact number of plays. We conclude that there exists some constant C_{1,ε_2} satisfying

$$A_3 \leq C_{1,\varepsilon_2} \,.$$

This result concludes our proof, and we refer the interested reader to the original paper for a complete proof and a detailed expression for C_{1,ε_2}. We further remark that contrarily to the previous terms, the upper bound of A_3 does not depend on F at all. □

E Performance measure

In this Section, we detail the computation of the performance measures we used to evaluate the identification strategies.

E.1 Direct measure of performance of an identification strategy

We denote $\widehat{C}_\alpha$ the expression of the empirical CVaR of level $\alpha \in (0, 1]$. The empirical CVaR is an estimate of the true CVaR as defined in Equation 3 –just as an average value is an estimate of the true mean of a distribution–. Assuming a sample $\mathcal{Y}$ of rewards sorted in an increasing order i.e. $\mathcal{Y} = \{y_1, \cdots, y_n\}$ such that $y_i \leq y_{i+1}$, and defining $q = y_{\lceil \alpha n \rceil}$ the empirical quantile of level α, we have:

$$\widehat{C}_\alpha(\mathcal{Y}) := q - \frac{1}{n\alpha} \sum_{i=1}^{n} \max(q - y_i, 0) \tag{5}$$

In a simulated problem, the CVaR can be estimated by repeatedly applying R times an identification strategy during T years, and then concatenating all results of all farmers from time $t = 1$ to time $t = T$ for all replications, and finally computing the empirical CVaR of the resulting set. In order to approximate all expectations, for all experiments, in practice we consider $R = 960$ (12 executions in parallel on an 80 core machine; for each one of the 960 experiments, the weather generator had a different random state). We denote $r \in \{1, \cdots, R\}$ the repetition index. We define $\mathfrak{Y}_T = \bigcup_{r=1}^{R} \mathcal{Y}_T^r$ i.e. the results of all farmers until year T for all replications. Then:

$$\mathbb{E}\left[\text{CVaR}_\alpha(\mathcal{Y}_T)\right] \mathrel{\widehat{=}} \widehat{C}_\alpha(\mathfrak{Y}_T) \tag{6}$$

The resulting quantity is an average measure of the results of the group. The more an identification strategy maximizes this quantity, the better it is. In a real-world problem, only one realization of $\text{CVaR}_\alpha(\mathcal{Y}_T)$ is computable.

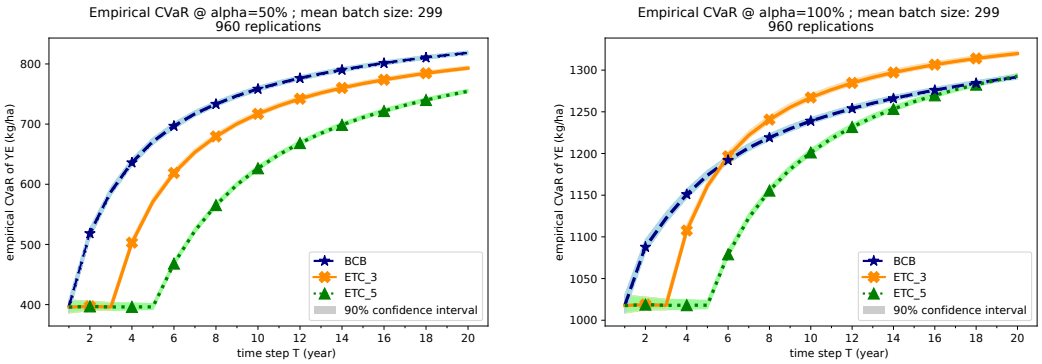
E.2 Proxy measure of performance of an identification strategy

While the CVaR can be estimated with Equation (6), it is complex to analyze and derive statistical guarantees for this estimator. This is why, we introduced a proxy of this quantity called the cumulated (CVaR) regret, which is a central element behind the theoretical performance guarantees of bandit algorithms.

Mean cumulated regret of the farmer population Considering a single cohort c, we suppose that we sequentially repeat T times the choice of one option k from an ensemble of K possible options. Here k is the index of the fertilizer practice. We denote $\text{CVaR}_\alpha(\nu_k^c)$ the CVaR of level α associated with the option k and cohort c, and $\text{CVaR}_\alpha(\nu_*^c) = \max_{k \in \{1,\cdots,K\}} \text{CVaR}_\alpha(\nu_k^c)$ the highest CVaR at level α of all options for cohort c i.e. the CVaR of the best option for cohort c. In expectation, for a farmer belonging to cohort c and following T years the recommendations of a given identification strategy selecting a fertilizer practice $k(t)$ each year $t \in \{1, \cdots, T\}$, we define the cumulated regret for the CVaR as in [53]:

$$\underbrace{R_\alpha^c(T)}_{\substack{\text{loss of the} \\ \text{strategy}}} := \underbrace{T \times \text{CVaR}_\alpha(\nu_*^c)}_{\substack{\text{score of the best} \\ \text{possible strategy}}} - \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \text{CVaR}_\alpha(\nu_{k(t)}^c)\right]}_{\substack{\text{score of the actual} \\ \text{strategy}}} \tag{7}$$

$$= \sum_{k=1}^{K} \underbrace{\left(\text{CVaR}_\alpha(\nu_*^c) - \text{CVaR}_\alpha(\nu_k^c)\right)}_{\substack{\text{loss between the best option} \\ \text{and the option } k \text{ for cohort } c}} \times \underbrace{\mathbb{E}\left[N_k^c(T)\right]}_{\substack{\text{expected number of times} \\ \text{option } k \text{ is chosen for cohort } c \\ \text{during the } T \text{ years}}} \tag{8}$$



(a) $\alpha = 50\%$ (b) $\alpha = 100\%$

Figure 9: Farmers' empirical CVaR at level of all YE received between $T = 0$ and the considered T.

For cohort c, the cumulated regret $R_\alpha^c(T)$ can be seen as a loss occurred with the considered strategy with respect to the best possible strategy –the one that always chooses the fertilizer practice with the best CVaR–. Equivalently, it can be interpreted as a measure of the expected total error due to sub-optimal actions made during a series of T decisions: the more often the best option is chosen within the T decisions, the smaller the cumulated regret. The mean cumulated regret of the total farmer population is given by the cumulated regret of each cohort, weighted by the probability of an individual to belong to this cohort:

$$R_\alpha(T) = \sum_{c=1}^{C} R_\alpha^c(T) \times \Pr(c), \text{ with } \sum_{c=1}^{C} \Pr(c) = 1 \tag{9}$$

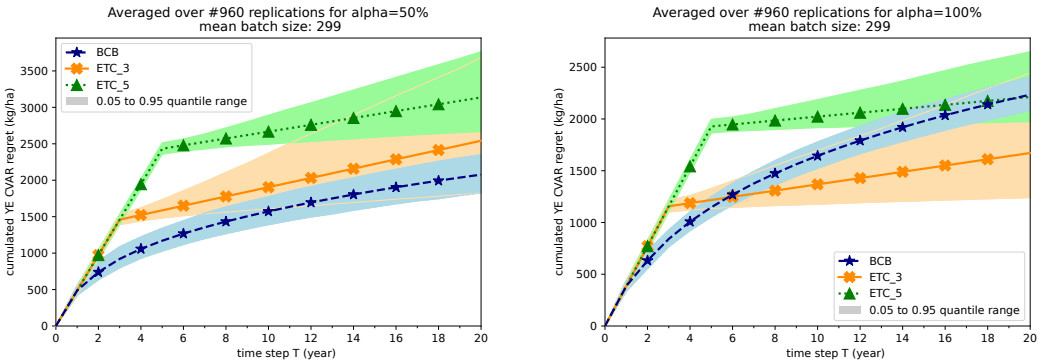
When extensively testing an identification strategy on a simulated problem, the CVaR of the different options can be approximated with a large enough number of samples or analytically computed, irrespective of the identification strategy. For each cohort, this corresponds to the left-hand side of Equation 8, and is thus supposed to be known. For a real-world problem, these quantities are unknown –else the decision problem would have been solved–. On the right hand side of Equation 8, the quantity $\mathbb{E}\left[N_k^c(T)\right]$ can be empirically approximated by repeatedly performing experiments with the identification strategy, and averaging the number of times each fertilizer practice has been chosen since time step T for each cohort. Finally, in Equation 9, the proportion of each soil, i.e. cohort, can be found in Table 1. Minimizing the cumulated regret maximizes the quantity in Equation **??**, as shown by [13]. For a given identification strategy, the smaller and less variable the mean cumulated regret of population (Equation 9), the more farmers are guaranteed to maximize their CVaR of YE.

F Experiment complements

Following methods of Section 2 of the main text, we provide identification performances of identification strategies for CVaR levels $\alpha = 50\%$ and $\alpha = 100\%$ with Figures 9, 10. For both CVaR levels, the YE is defined with $\text{ANE}_{\text{ref}} = 15$ kg N/kg grain.

G Alternative performance measure of fertilizer practices

We briefly discuss economical criteria we considered as performance indicators of fertilizer practices. A first indicator we considered was the gross margin. The cost of production of nitrogen fertilizer being indexed on the price of natural gas, it is subject to high volatility. As a consequence, an optimal practice is likely to be different each year and thus the decision problem would turn to be highly non-stationary. Such setting dramatically increases the complexity of the decision problem, and the chance of observing good identification performances are lowered.



(a) $\alpha = 50\%$ (b) $\alpha = 100\%$

Figure 10: Cumulated regret averaged over the population for the CVaR at level of YE.

Another economic measure could be the value:cost ratio (VCR), which is given for a fertilizer practice π as:

$$\mathrm{VCR}^{\pi} = \frac{p_{\mathrm{maize}}}{p_{\mathrm{N}}} \times \frac{Y^{\pi} - Y^{0}}{\mathrm{N}^{\pi}} \tag{10}$$

$$= \frac{p_{\mathrm{maize}}}{p_{\mathrm{N}}} \times \mathrm{ANE}^{\pi} \tag{11}$$

where p_{N} is fertilizer unitary cost and p_{maize} unitary maize grain selling price. We neglect a possible quality consideration that could motivate a different maize selling price between the fertilizer practices, for instance a difference of protein content in maize grains. Remarking that each given year the ratio $\frac{p_{\mathrm{maize}}}{p_{\mathrm{N}}}$ is shared by all fertilizer practices, then the decision problem is perfectly equivalent to choosing the fertilizer practice which maximizes the ANE. Thereby, the use of the cost:value ratio suffers from the same drawbacks as the ANE.