

Towards an efficient and risk aware strategy for guiding farmers in identifying best crop management

Romain Gautron^{*1,2,3}, Dorian Baudry⁴, Myriam Adam^{5,6,7}, Gatien N. Falconnier^{1,2,8}, and Marc Corbeels^{1,2,9}

¹AIDA, Univ Montpellier, France.

²CIRAD, Montpellier, France.

³CGIAR Platform for Big Data in Agriculture, Alliance of Bioversity International and CIAT, Km 17, Recta Cali Palmira 763537, Colombia.

⁴Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9198-CRISAL, F-59000 Lille, France.

⁵CIRAD, UMR AGAP Institut, Bobo-Dioulasso 01, Burkina Faso.

⁶UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France.

⁷Institut National de l'Environnement et de Recherches Agricoles (INERA), Burkina Faso.

⁸International Maize and Wheat Improvement Centre (CIMMYT)-Zimbabwe, 12.5 km Peg Mazowe Road, Harare, Zimbabwe.

⁹International Institute of Tropical Agriculture, PO Box 30772, Nairobi, 00100, Kenya.

October 7, 2022

Abstract

Identification of best performing fertilizer practices among a set of contrasting practices with field trials is challenging as crop losses are costly for farmers. To identify best management practices, an “intuitive strategy” would be to set multi-year field trials with equal proportion of each practice to test. Our objective was to provide an identification strategy using a bandit algorithm that was better at minimizing farmers’ losses occurring during the identification, compared with the “intuitive strategy”. We used a modification of the Decision Support Systems for Agro-Technological Transfer (DSSAT) crop model to mimic field trial responses, with a case-study in Southern Mali. We compared fertilizer practices using a risk-aware measure, the Conditional Value-at-Risk (CVaR), and a novel agronomic metric, the Yield Excess (YE). YE accounts for both grain yield and agronomic nitrogen use efficiency. The bandit-algorithm performed better than the intuitive strategy: it increased, in most cases, farmers’ protection against worst outcomes. This study is a methodological step which opens up new horizons for risk-aware ensemble identification of the performance of contrasting 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 impact of cropping activities (Tilman et al., 2002). 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, might not work in the next season, when rainfall is less (Affholder, 1995). In fact, the performance of crop management at a given site has an underlying “hidden” distribution due to inter-annual weather variability, thus creating great uncertainty (Fosu-Mensah et al., 2012). Because crop management decisions are recurrent, i.e. they are repeated for each new crop growing season, the identification of optimal crop management falls into the category of sequential decision making under uncertainty (Gautron et al., 2022). 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 weather variability (Hochman and Carberry, 2011).

*romain.gautron@cirad.fr

There exist numerous decision support tools of widely ranging complexity for crop management, introduced to farmers with varying degrees of success (Gautron et al., 2022).

Machine learning (ML) and more generally artificial intelligence (AI) can help address sequential decision making under uncertainty. In particular, the bandit algorithm paradigm (Lattimore and Szepesvári, 2020) considers a decision-maker, called agent, who repeatedly faces a choice between contending actions, and has to iteratively improve its decisions with trials. The canonical bandit problem originates from clinical trials with sequential drug allocation (Thompson, 1933). At each time step, the agent chooses one action (i.e., one drug for a patient) amongst a set of possible actions. Each action provides a reward (i.e.; tumor cell reduction after taking the drug), drawn from a corresponding unknown reward distribution (i.e., the distribution of tumor cell reduction for the drug). The optimal action has the reward distribution with the highest mean reward (i.e., the highest mean tumor cell reduction). The objective of the agent is to sequentially choose actions such that the expected sum of rewards is maximized. Maximizing the total expected rewards is equivalent to minimizing the regret, which is a measure of the total losses that occur with sub-optimal actions (Robbins, 1952).

Iteratively, the agent refines his next decision based on all previous results. To know how a given action performs, a sufficient number of (possibly poor) rewards is required: this is the exploration phase. To maximize the expected sum of rewards, the previous actions that provided good results so far must be selected more frequently; this is the exploitation phase. Bandit algorithms aim at finding the right balance between exploration and exploitation. This *exploration-exploitation dilemma* is a reality for farmers when implementing crop management. Farmers typically want to minimize overall crop yield losses and typically explore the performance of promising new crop management practices on small test plots (Cerf and Meynard, 2006; Evans et al., 2017). They avoid potentially large crop yield losses from new management by managing a gradual transition between the current management and the promising new one(s), based on the results they obtain on the small test plots.

The objective of this paper is to develop a novel strategy to identify best crop management. We set as baseline an “intuitive strategy” which consists in identifying the best crop management through multi-year field trials in which a set of crop management practices is tested in an equiproportional way. We compare this “intuitive strategy” to a novel crop management identification strategy, based on a bandit algorithm. This novel identification strategy aims to minimize farmers’ yield losses occurring during the identification process, compared to the intuitive strategy. Thus, we test the hypothesis that bandit algorithm can help farmers to better identify the best crop management for their context, while further minimizing crop yield losses related to sub-optimal choices in new crop management.

Our case study considers the rainfed maize production in southern Mali, and we compare the performance of both crop management identification strategies based on maize growth simulations using a calibrated crop model in order to mimic real-world performance of crop management. The novel identification strategy does, however, not depend on model simulations, and ultimately aims at being applied in real field conditions. As for crop management, we focus on nitrogen fertilization. Tailoring nitrogen fertilizer recommendations to farmers’ contexts is known to be challenging. Indigenous soil nitrogen supply, depending to a large extent on past-season events, is not accurately known to farmers, whilst in-season nitrogen mineralization depends largely on weather events (Morris et al., 2018), themselves uncertain. Crop nitrogen requirements, such as with maize, are related to specific crop growth stages (Hanway, 1963) and excessive mineral nitrogen supply can induce nitrate leaching, especially in wet conditions (Meisinger and Delgado, 2002). Therefore, there are *a priori* no upfront optimal nitrogen fertilizer practices.

2 Methods

2.1 Virtual crop management identification problem

In our virtual crop management identification problem, a population or ensemble of farmers joined a participatory experiment to identify the best nitrogen fertilizer practices for maize production in their region, Koutiala in southern Mali. A total population of 500 farmers was considered. The distribution of soil types of the fields associated with the group of farmers was representative of the region (Table 1). A total population of 500 farmers was considered. Each farmer belonged to a cohort that corresponded to an ensemble of farmers growing maize on the same soil type. For each cohort, we wanted to identify the best nitrogen fertilizer practice from a set of recommended practices (see Table 3 and Section 2.1.1 for the performance metrics we considered).

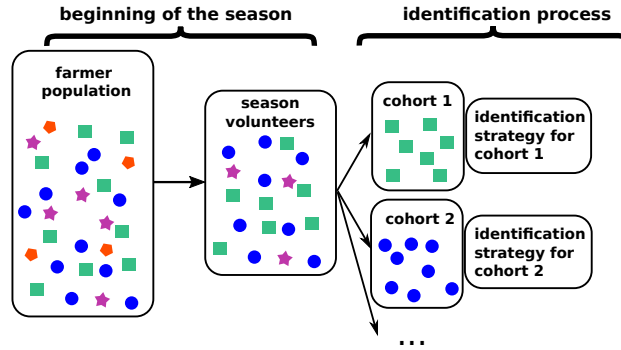


Figure 1: Yearly process to generate nitrogen fertilizer recommendations: at the beginning of the crop ping season. Individuals from the overall farmer population volunteered to test a fertilizer practice. Similar symbols represent a cohort, i.e., a group of farmers having fields with the same soil type. The group of volunteer farmers was broken down by cohort and researchers independently generated fertilizer recommendations for each cohort. Researchers did not control the number of volunteers from the respective cohorts In this example, only three of the four possible cohorts are found in the volunteer group.

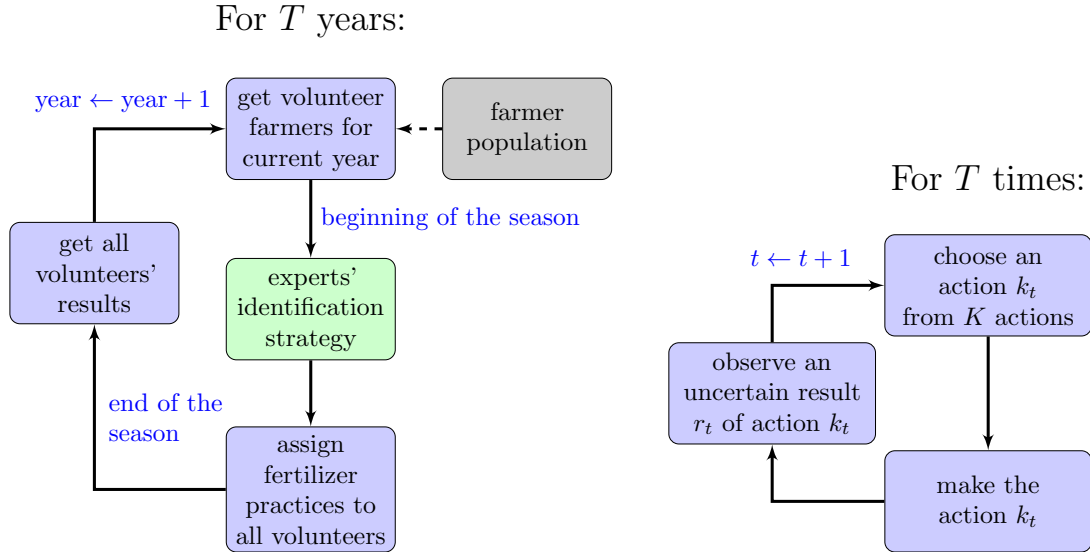
The research team set the additional objective to limit the crop yield losses of individual farmers that could arise from poor nitrogen fertilizer practice recommendations during the identification process.

At the beginning of each crop growing season, we assumed that a random number of farmers (uniformly obtained between 250 and 350) of the population of 500 farmers volunteered to apply the recommended fertilizer applications provided by the research team. Each year, the group of volunteers was variable in size and in the representation of cohorts, as could occur in reality (Figure 1). Thus, researchers did not control the composition of the group of volunteers. Each farmer indicated the fields and corresponding soils on which she/he planned to grow maize. Researchers then provided a fertilizer recommendation (Table 3) to each farmer for the ongoing season, depending on her/his soil i.e. cohort. At the end of the season, volunteer farmers shared their results in terms of maize grain yields with the research team, allowing to refine the recommendations for the next season. The whole process was repeated during 20 consecutive years following the same process (Figure 2a).

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

- Two split applications (practice 0): 15 days after planting (DAP) and 30 DAP.
- Three split applications (practice 4) :15 DAP, 30 DAP and 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 content (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 no stress, 1 maximal stress), hereby mimicking the use of a portable chlorophyll meter to monitor plant nitrogen content (e.g. Kalaji et al., 2017).

Practice 8 corresponded to the optimal fertilization for maize (70 kg N/ha) in the study area based on simulations (Huet et al., 2022) , i.e. the average of the N fertilizer rates that were observed to result in maximum positive return on fertilizer investment (Getnet et al., 2016). Finally, practice 9 (180 kg N/ha) corresponded to a nitrogen fertilizer practice that is likely excessive. For all these practices, the nitrogen fertilizer applied was assumed to be ammonium nitrate broadcasted on the soil surface.



(a) Diagram of the ensemble best fertilizer identification process. Each year, a group of volunteer farmers test fertilizer practices recommended by experts and contribute to identifying the best fertilizer practices for the region. At the end of each season, the farmers share their results with experts. The experts will use these results to improve their recommendations for the next growing season. The process repeats for a total number of T years.

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

Figure 2: Schematic representation of the ensemble best fertilization identification process and the canonical bandit problem.

Table 1: Main properties of the soil types of the fields of farmers growing maize in Koutiala, Mali (Adam et al., 2020). ‘SLOC.’ stands for soil organic matter (g C/ 100 g soil, mean value for the 0-30 cm topsoil); ‘SLDR’ stands for soil drainage rate (fraction/day); ‘SLDP’ stands for soil depth (cm); ‘Prop’ stands for the percentage of each soil type present in the study area.

Soil name	Texture	SLDR	SLOC	SLDP	Prop.
ITML840101	clay loam	0.60	0.20	110	7%
ITML840102	loam	0.60	0.45	100	9%
ITML840103	silty loam	0.60	0.27	160	21%
ITML840104	silty clay loam	0.25	0.70	105	4%
ITML840105	silty clay loam	0.40	0.35	120	24%
ITML840106	loam	0.60	0.30	110	27%
ITML840107	silty clay loam	0.25	0.60	105	8%

Table 2: Maize nitrogen fertilizer recommendations for maize in Koutiala, Southern Mali, that were considered in the virtual experiment. Whether or not rainfall and plant nitrogen stress were considered as factors for the fertilizer recommendation is indicated by Yes or No. ‘NSTRES’ stands for plant nitrogen stress and ‘DAP’ for days after planting.

index	max amount applied (kgN/ha)	max applications	rainfall threshold	NSTRES threshold	15 DAP N (kgN/ha)	30 DAP N (kgN/ha)	45 DAP N (kgN/ha)
0	135	2	No	No	15	120	0
1	135	2	No	Yes	15	120	0
2	135	2	Yes	No	15	120	0
3	135	2	Yes	Yes	15	120	0
4	135	3	No	No	15	60	60
5	135	3	No	Yes	15	60	60
6	135	3	Yes	No	15	60	60
7	135	3	Yes	Yes	15	60	60
8	70	2	No	No	23	0	47
9	180	3	No	No	60	60	60

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 Koutiala in southern Mali using *gym-DSSAT* (Gautron and Padrón González, 2022). *gym-DSSAT* is a modification of the DSSAT crop simulator (Hoogenboom et al., 2019) to allow a user to read DSSAT internal states and take daily fertilization decisions during the simulations (e.g. based on DSSAT internal states). For each soil type in Table 1 that was parametrized in DSSAT using the data from Adam et al. (2020), each simulated maize grain yield value is a sample of the 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 (Richardson and Wright, 1984; Soltani and Hoogenboom, 2003), which was calibrated using the 47-year-long weather records from N’tarla, about 30 km from Koutiala (Ripoche et al., 2015). The ‘sotubaka’ maize cultivar (from the DSSAT default cultivar list) was used for all model simulations as a representative of maize variety in southern Mali. Water and nitrogen stresses were simulated, but yield reduction through pests and diseases were not considered, neither was weed competition. In the model simulations, a different weather time series was generated for each growing season and for each recommendation using WGEN, inducing independent simulated maize yield responses to nitrogen fertilization. Section A of Supplementary Materials gives further details of the simulation settings.

We simulated 10^5 times the maize grain yield responses to a given fertilizer practice for the different soil types, which corresponds to 10^5 hypothetical growing seasons. These samples were used i) to ensure that simulated maize yield responses were in realistic expected ranges, ii) to qualitatively 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 application (i.e. no prior knowledge of the problem).

2.1.1 Performance indicators of fertilizer practices

A criterion to evaluate both the economic and environmental performance of a fertilizer practice π is Agronomic Nitrogen use Efficiency (ANE), as defined in Vanlauwe et al. (2011):

$$\text{ANE}^\pi := \frac{Y^\pi - Y^0}{N^\pi} \quad (1)$$

where Y^π is the crop yield obtained with the nitrogen fertilizer practice π which required a quantity N^π of nitrogen and Y^0 is the yield of the control obtained in the same conditions without nitrogen fertilization. Maximising ANE is a proxy of minimizing the quantity of nitrogen losses, e.g. through nitrate leaching.

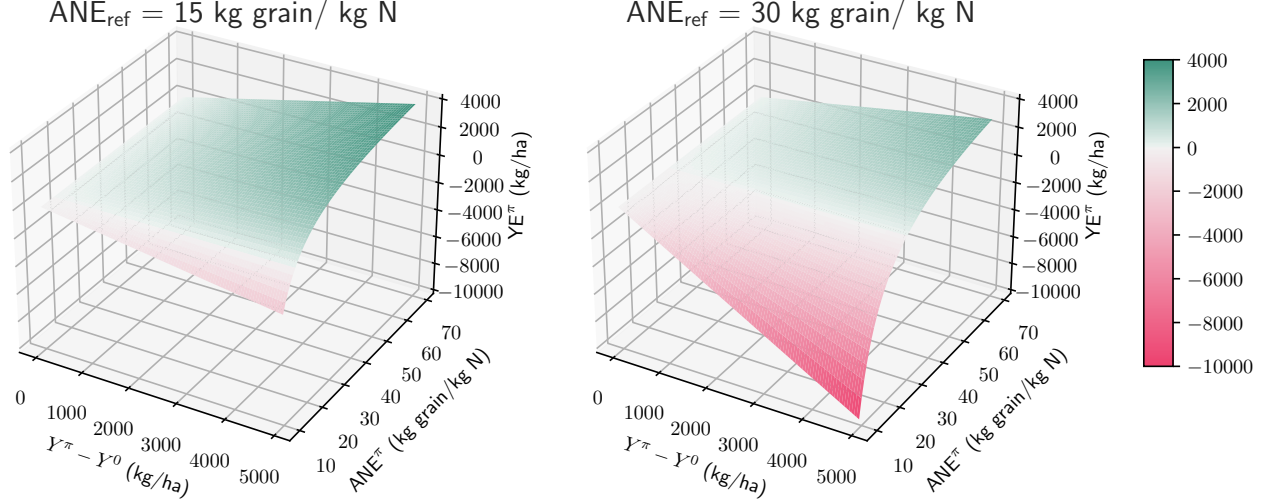


Figure 3: Yield Excess (YE^π , Equation 5) for $ANE_{\text{ref}} = 15$ kg grain /kg N and $ANE_{\text{ref}} = 30$ kg grain /kg N. Y^π is the maize grain yield obtained with nitrogen fertilizer practice π , Y^0 is the yield obtained with no nitrogen fertilization (control). ANE^π is the Agronomic Nitrogen use Efficiency of the nitrogen fertilizer practice π (Equation 1).

However, ANE has some limitations: for example, an ANE value of 25 kg grain/kg N can be achieved with a fertilizer input of 20 kg N/ha yielding a total yield gain of 500 kg/ha, or with an input of 60 kg N/ha yielding a total gain of 1500 kg/ha. For the same ANE, a farmer is likely to prefer the fertilizer practice that provides the greatest crop yield gain, i.e. with 60 kg N/ha. Similarly, choosing fertilizer practices only based on the associated crop yield gains is not satisfying. A similar yield gain can be achieved with different nitrogen fertilizer input rates which result in fairly different ANE: the practice with the highest efficiency must be preferred as it required less nitrogen fertilizer to achieve the same yield gain.

We built the Yield Excess (YE) indicator that favors the nitrogen fertilizer practice with the highest yield gain for those practices sharing the same ANE, and favors the practice with the highest efficiency for those practices sharing the same yield gain. YE of a nitrogen fertilizer practice π with respect to the reference practice π_{ref} of constant efficiency ANE_{ref} using the same quantity of nitrogen fertilizer as practice π , denoted N^π , is computed as follows:

$$YE^\pi := Y^\pi - Y^{\pi_{\text{ref}}} \quad (2)$$

$$= \underbrace{Y^\pi - Y^0}_{\text{yield gain of } \pi \text{ w.r.t. control}} - \underbrace{(Y^{\pi_{\text{ref}}} - Y^0)}_{\text{yield gain of } \pi_{\text{ref}} \text{ w.r.t. control}} \quad (3)$$

$$= Y^\pi - Y^0 - N^\pi \times ANE_{\text{ref}} \quad (4)$$

$$= (Y^\pi - Y^0) \times \underbrace{\left(1 - \frac{ANE_{\text{ref}}}{ANE^\pi}\right)}_{\text{penalization factor}} \quad (5)$$

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 constant ANE equal to ANE_{ref} and which uses the same quantity N^π of nitrogen fertilizer as π . YE^π increases with ANE^π (Figure 3). YE^π is negative and decreases with $Y^\pi - Y^0$ when $ANE^\pi < ANE_{\text{ref}}$ and is positive and increases with $Y^\pi - Y^0$ when $ANE^\pi \geq ANE_{\text{ref}}$. The YE of fertilizer practices with efficiency below ANE_{ref} are negatively affected by this metric. We chose $ANE_{\text{ref}} = 15$ kg grain/kg N for our model simulation experiments, the average ANE currently achieved by farmers across sub-Saharan Africa (Ten Berge et al., 2019; Vanlauwe et al., 2011).

Because farmers are usually risk averse (e.g. Cerf and Sebillotte, 1997; Menapace et al., 2013; Jourdain et al., 2020), they are likely to prefer, for example, a stable maize grain yield of 3000 kg/ha rather than a yield

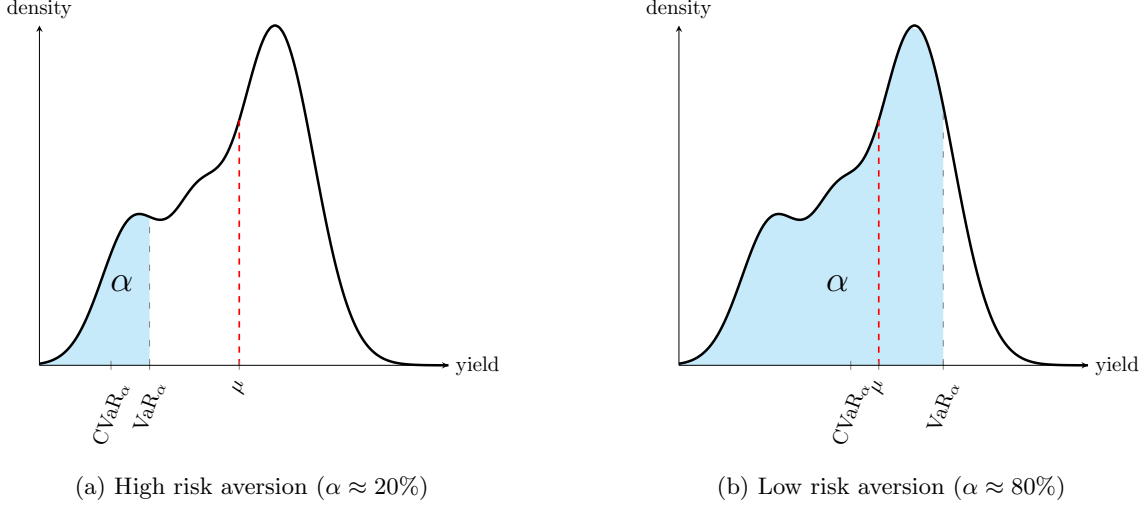


Figure 4: The Conditional Value-at-Risk (CVaR) of level α is the mean value of the blue area of the distribution of probability $0 < \alpha \leq 1$. VaR_α stands for Value-at-Risk of level α and is the quantile of probability α of the distribution. The more $\alpha \rightarrow 1$, the more risk neutral is the CVaR. μ represents the mean value of the distribution which equivalent to the CVaR of level $\alpha = 100\%$.

of 5000 kg/ha in half of the years, and of 1000 kg/ha in the other half of the years, while both distributions share the same expectation. To account for risk aversion, we computed the Conditional-Value-at-Risk (CVaR, Mandelbrot, 1997; Acerbi and Tasche, 2002), a risk-aware metric that originated from finance. The CVaR focuses on the lower tail of the distribution¹. For a (continuous) random variable X with cumulative distribution function F_X , we call Value-at-Risk (VaR) of level α the quantile of probability $\alpha \in (0, 1]$ of X , defined as:

$$\text{VaR}_\alpha(X) := \inf \{x \in \mathbb{R} : F_X(x) > \alpha\} \quad (6)$$

Then the CVaR of X of level $\alpha \in (0, 1]$ is the mean value of the left tail of X of probability α , defined as:

$$\text{CVaR}_\alpha(X) := \mathbb{E}[X|X \leq \text{VaR}_\alpha(X)] \quad (7)$$

A decision maker would choose the option with the highest CVaR for the considered level α . The more $\alpha \rightarrow 0^+$, the more the metric focuses on the worst observable yields. On the contrary, the more $\alpha \rightarrow 1$, the less risk averse is the measure. When $\alpha = 1$, the CVaR equals the usual expectation $\mathbb{E}[X]$, which is risk neutral (Figure 4). In our model simulation experiments, we chose $\alpha = 30\%$. The $\text{CVaR}_{30\%}$ represents the mean crop yield of the 30% worst observable years.

2.2 Identification of the best fertilizer practices

The canonical and batch bandit problems The ensemble identification of the best crop management practices with the constraint of minimizing farmers' crop yield losses occurring during the identification process (Section 2.1) can be modelled as a special type of bandit problems. The canonical bandit problem, which is the cumulated regret minimization (see Introduction), assumes that at each time step, a single trial is made and is followed by a single observation of a result, in a purely sequential mode. In contrast, the batch bandit setting (Perchet et al., 2015) assumes that at each time step an ensemble of trials are conducted in parallel, followed by the observation of an ensemble of results. Figure 2 illustrates on the one hand the ensemble identification process of best crop fertilizer practices (Figure 2a), modelled as a batch-bandit problem, and the on other hand the canonical bandit problem (Figure 2b).

In the canonical bandit problem, the agent goal is to maximize the expectation of the sum of rewards that were collected since the first decision. The agent objective can be formalized as maximizing $\mathbb{E} \left[\sum_{t=1}^T r_t \right]$ for any

¹Two definitions of the CVaR coexist in the literature, depending if an outcome is considered as a gain or a cost (Dowd, 2007). We adopted the gain point of view.

time horizon $T \geq 1$, with r_t the reward the agent has collected at time t . On the other hand, bandits that are *risk-aware* (Cassel et al., 2018), the agent maximizes a risk-aware measure of the collected rewards, such as the CVaR (Section 2.1.1), instead of the expectation of rewards. Our ensemble fertilizer decision problem can be described as a *risk-aware batch-bandit* decision problem.

The ensemble identification problem of best fertilizer practices In our virtual modeling experiment, for $t \in \{1, 2, \dots, T\}$, at each season t , researchers assigned each n_t volunteer farmers for season t with a nitrogen fertilizer practice $\pi \in \{1, 2, \dots, K\}$. Each farmer belonged to a cohort $c \in \{1, 2, \dots, C\}$. At the end of season t , researchers assemble rewards $Y_t = \{y_t^1, \dots, y_t^{n_t}\}$ as a result of the fertilizer practices of all farmers for season t . For each cohort $c \in \{1, \dots, C\}$, rewards are independently and identically distributed from unknown stationary distributions $\{\nu_1^c, \dots, \nu_K^c\}$. These reward distributions are the YE with $\text{ANE}_{\text{ref}} = 15$ kg grain/kg N associated to each of the ten recommended nitrogen fertilizer practices, for a given soil type. We denote $\mathcal{Y}_T = \bigcup_{t=1}^T Y_t$ the set of all rewards observed by all farmers between $t = 1$ and $t = T$. The objective of an identification strategy is to maximize, for a given CVaR level α and any time horizon $T \geq 1$:

$$\mathbb{E}[\text{CVaR}_\alpha(\mathcal{Y}_T)] \quad (8)$$

For each cohort $c \in \{1, \dots, C\}$, an optimal nitrogen fertilizer practice π_*^c is given by:

$$\pi_*^c = \underset{k}{\operatorname{argmax}} \text{CVaR}_\alpha(\nu_k^c) \quad (9)$$

Consequently, an optimal identification strategy always assigns nitrogen fertilizer practice π_*^c to all farmers belonging to cohort c .

2.2.1 Identification strategies

We expected fertilizer practices to perform differently within each cohort, i.e. each soil. For example, the optimal nitrogen fertilizer practices were expected to be different between a cohort growing maize on a shallow sandy soil and a cohort growing maize on a deep clayey soil. Consequently, the results of one cohort were not supposed to be directly relevant for another cohort. Each soil was considered as an independent identification problem, i.e. had its own independent identification strategy which did not share information with the identification strategies of other soils.

For a given soil, from one season to another, the identification strategy kept memory of all results observed during past seasons, for the same soil. In model simulation experiments, we considered two types of identification strategies: either the standard ETC (Explore-Then-Commit) strategy, previously referred as the ‘‘intuitive strategy’’, or BCB, the bandit-algorithm based identification strategy. For the seven soils in Table 1, the identification strategy types were either all ETC, or all BCB, but not a mix of both.

Intuitive identification strategy ETC 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 strategies. 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 Section B.2 of Supplementary Materials, we provide a simple adaptation of ETC to the batch setting (see Section 2.1) using the CVaR of rewards rather than the classical expectation. We considered ETC-3 and ETC-5, with respectively 3 and 5 years of exploration phases. During the exploration phase, fertilizer practices are randomly assigned in equal proportions to the farmers within the cohort.

Bandit based identification strategy BCB is a risk-aware bandit algorithm (Cassel et al., 2018) which uses the CVaR of rewards as decision criterion, in the batch bandit setting. BCB derives from the the work of Baudry et al. (2021a). We provide the pseudo-code of BCB and detail how it works in Supplementary Materials Section B.1. The general idea of the bandit algorithm is, for each season, to leverage the information acquired during all past seasons, such that the algorithm adapts to optimally manage the exploration-exploitation dilemma.

We provide a quick overview of the execution of BCB with algorithm 1. Considering the YE with $\text{ANE}_{\text{ref}} = 15$ kg grain/kg N as results, we set its maximum observable result to 4000 kg/ha for all fertilizer practices as

required for the execution of BCB (see first execution step of algorithm 1), based on Figure 3. As an additional feature, BCB provides a fair distribution of risky option trials amongst farmers at the cohort level. The bandit algorithm ranks each fertilizer practice according to its observed performance in the previous year. The algorithm then recommends first the practices that appear to yield best results to the farmers that have experienced worst results so far.

Algorithm 1 Simplified pseudo-code of BCB.

```

for fertilizer practice  $k \in \{1, \dots, K\}$  do
  Add maximum observable value to the results of fertilizer practice  $k$  // prior to any experiments
end
for season  $t \in \{1, \dots, T\}$  do
  for farmer  $f \in \{1, \dots, n\}$  do
    for fertilizer practice  $k \in \{1, \dots, K\}$  do
      Re-weight the rewards of the fertilizer practice  $k$  with random weights sampled from a
      Dirichlet distribution (Everitt and Skrondal, 2002)
      Score practice  $k$  with a noisy empirical measure of the CVaR at level  $\alpha$  of practice  $k$  from
      the re-weighted rewards
    end
    Recommend to the farmer  $f$  the fertilizer practice with the maximum score
  end
  Collect and store all results of the season for all fertilizer practices
end

```

2.2.2 Direct measure of performance of an identification strategy

We denote \widehat{C}_α 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 7 –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. $\mathcal{Y} = \{y_1, \dots, y_n\}$ such that $y_i \leq y_{i+1}$, and defining $q = y_{\lceil \alpha n \rceil}$ the empirical quantile of level α , we have:

$$\widehat{C}_\alpha(\mathcal{Y}) := q - \frac{1}{n\alpha} \sum_{i=1}^n \max(q - y_i, 0) \quad (10)$$

In a simulated problem, the quantity in Equation 8 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 $r \in \{1, \dots, R\}$ 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:

$$\mathbb{E}[\text{CVaR}_\alpha(\mathfrak{Y}_T)] \hat{=} \widehat{C}_\alpha(\mathfrak{Y}_T) \quad (11)$$

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 $\text{CVaR}_\alpha(\mathfrak{Y}_T)$ is computable.

2.2.3 Proxy measure of performance of identification strategy

While the quantity in Equation 8 can be estimated with Equation 11, it is intricate to analyze and derive statistical guarantees for this estimator. This is why, in the following, we introduce a proxy of this quantity called the cumulated CVaR regret, which is a central element behind the theoretical performance guarantees of bandit algorithms. The cumulated regret is also a convenient statistic to represent the performance of an algorithm, with little noise.

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 $\text{CVaR}_\alpha(\nu_*^c) = \max_{k \in \{1, \dots, K\}} \text{CVaR}_\alpha(\nu_k^c)$ 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 $t \in \{1, \dots, T\}$, we define the cumulated regret for the CVaR as in [Tamkin et al. \(2020\)](#):

$$\underbrace{R_\alpha^c(T)}_{\text{loss of the strategy}} := \underbrace{T \times \text{CVaR}_\alpha(\nu_*^c)}_{\text{score of the best possible strategy}} - \underbrace{\mathbb{E} \left[\sum_{t=1}^T \text{CVaR}_\alpha(\nu_{k(t)}^c) \right]}_{\text{score of the actual strategy}} \quad (12)$$

$$= \sum_{k=1}^K \underbrace{(\text{CVaR}_\alpha(\nu_*^c) - \text{CVaR}_\alpha(\nu_k^c))}_{\text{loss between the best option and the option } k \text{ for cohort } c} \times \underbrace{\mathbb{E}[N_k^c(T)]}_{\text{expected number of times option } k \text{ is chosen for cohort } c \text{ during the } T \text{ years}} \quad (13)$$

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 the best option is chosen within the T decisions, the smaller the cumulated regret is. 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:

$$R_\alpha(T) = \sum_{c=1}^C R_\alpha^c(T) \times \Pr(c), \text{ with } \sum_{c=1}^C \Pr(c) = 1 \quad (14)$$

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 13, and is thus supposed to be known. Note that, for a real-world problem, these quantities are unknown –else the decision problem would have been solved–. On the right hand side of Equation 13, the quantity $\mathbb{E}[N_k^c(T)]$ 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 14, the proportion of each soil, i.e. cohort, can be found in Table 1. Minimizing the cumulated regret maximizes the quantity in Equation 8, as shown by [Cassel et al. \(2018\)](#). For a given identification strategy, the smaller and less variable the mean cumulated regret of population (Equation 14), the more farmers are guaranteed to maximize their CVaR of YE.

Distribution of the cumulated regret of individual farmers The mean cumulated regret of the population given in Equation 14 does not indicate the distribution of individual farmer regrets. For each farmer f belonging to cohort c , the individual regret after T years for the CVaR of level $\alpha \in (0, 1]$ is computed as:

$$\tilde{R}_\alpha^{f,c}(T) := \sum_{k=1}^K \underbrace{(\text{CVaR}_\alpha(\nu_*^c) - \text{CVaR}_\alpha(\nu_{k(t)}^c))}_{\text{loss between the best option and the option } k} \times \underbrace{N_k^{f,c}(T)}_{\text{number of times option } k \text{ is chosen during } T \text{ years for farmer } f} \quad (15)$$

For each cohort c , the distribution of $\tilde{R}_\alpha^{f,c}(T)$ indicates how the potential losses due to bad recommendations are distributed amongst farmers.

3 Results

Table 3: Statistics of the optimal nitrogen fertilizer practices for each of the soil types presented in Table 1. For the corresponding optimal 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 2.1.1); \bar{X} : mean value of X ; Y^{π^*} : maize grain yield; ANE^{π^*} : Agronomic Nitrogen use Efficiency; YE^{π^*} : Yield Excess (Section 2.1.1); parentheses indicate standard deviations.

soil	π^*	\bar{N}^{π^*} (kg/ha)	$\text{CVaR}_{30\%}(Y^{\pi^*})$ (kg/ha)	\bar{Y}^{π^*} (kg/ha)	ANE^{π^*} (kg/kg)	$\text{CVaR}_{30\%}(\text{YE}^{\pi^*})$ (kg/ha)	$\bar{\text{YE}}^{\pi^*}$ (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)

3.1 Simulated responses to nitrogen fertilizer practices

Table 3 provides the statistics of the optimal nitrogen fertilizer practices for each soil type (Table 1), i.e. for each cohort, and Figure A.1 in Supplementary Materials shows the distribution of grain yield, ANE and YE responses. All responses showed values within the expected ranges for the considered growing conditions, 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, on average the YE (i.e. yield gain compared to the reference) for farmers ranged from 1200 kg/ha to 1800 kg/ha, and the $\text{CVaR}_{30\%}(\text{YE})$ (i.e. the mean crop YE of the 30% worst observable years) from 500 kg/ha to 1032 kg/ha.

There was no simple parametric assumption that could be made about YE, such as its probability distribution to be Gaussian (e.g. practice 5 in Figure A.1e). The thicker left tails for e.g. fertilizer practices 4 and 0 or the bi-modality of YE for practices 6 and 7 (Figure A.1e), further supported the use of the CVaR as a relevant risk measure. Indeed, the CVaR is most relevant for asymmetric and irregularly shaped distributions, such as thick-tailed or multi-modal distributions. For all soils, the optimal nitrogen fertilizer practices were either nitrogen fertilizer practice 0 or 8 i.e. nitrogen practices without threshold dependent top-dressing, and with a single nitrogen top-dressing application (Table 3).

The nitrogen fertilizer practices had different responses for the different soil types in terms of the grain yield and ANE (and consequently YE), and ranking of the practices was inconsistent across the soil types (Figure A.1). For instance, for the soil ITML840104 (silt clay loam of medium fertility), fertilizer practices 0 to 4 had similar YE (Figure A.1e). For the soil ITML840105 (silt clay loam of low fertility), practices 0, 1 and 4 were substantially better than practices 2 and 3 (Figure A.1f).

Threshold-based fertilizer practices behaved inconsistently across the soil types. As an example, for the bi-modal YE distribution of the fertilizer practice 1, most of the probability density was concentrated around 0 kg/ha for the soil ITML840104 (Figure A.1e) and around 1800 kg/ha for the soil ITML840105 (Figure A.1f). For the soil ITML840104 and practice 1, YE were mostly found around 0 kg/ha because most of the seasons, the nitrogen-stress threshold of 0.2 was not reached, and consequently no top-dressing occurred (Table 2). In such cases, only a basal-dressing of 15 kg N/ha was applied, instead of a total of 135 kg N/ha when the top-dressing was triggered. Consistently, for the same soil and fertilizer practice, the probability density of grain yield was concentrated around the low value of 1000 kg/ha (Figure A.1a). On the other hand, with the soil ITML840105, most of the seasons, the nitrogen-stress threshold of 0.2 was reached and practice 1 applied both basal and top-dressing. This corresponded to YE mostly found around 1800 kg/ha (Figure A.1f), and the corresponding grain yields were mostly found around 4000 kg/ha (Figure A.1b).

3.2 Identification of best fertilizer practices

In Section 3.2.2 and 3.2.3, we present respectively a direct measure of empirical performances of the nitrogen fertilizer practice identification strategies (see Section 2.2.2), and the regret as a proxy measure, both for the farmer population average and the individual farmer regret distribution (see Section 2.2.3). Section 3.2.1

provides a visual comparison of nitrogen fertilizer recommendations following respectively the BCB and ETC-5 identification strategies.

3.2.1 Sampling visualization

Figure 5 provides the average frequency with which the fertilizer practices were selected by the identification strategies, from the beginning of the experiment to time T , for soils ITML840105 and ITML840101. For the soil ITML840105, respectively for the BCB and ETC-5 strategies. After 20 years, BCB had selected the fertilizer practice 8, which was the optimal one (see Table 3), with an average proportion of 50%. The proportions of the optimal practice continuously increased from year 2 onwards (Figure 5a). During the first 5 years, ETC-5 uniformly sampled all fertilizer practices (Figure 5b), thus inducing potentially high losses for farmers. The proportion of the optimal practice started to increase from year 5 onwards. After year 20, ETC-5 sampled the optimal practice with an average proportion of 31%. For soil ITML840101, results are more contrasted. After year 20, both BCB has sampled the optimal strategy, which was fertilizer practice 0 (see Table 3) with an average proportion of 27% (Figure 5c) and ETC-5 (Figure 5d) with an average proportion of 26%. Note that in Figures 5c and 5d, the color differences are almost not perceptible for nitrogen fertilizer practices 0, 1 and 4, because all three practices showed similar performances. In Sections 3.2.2 and 3.2.3, we provide the results of statistics that account for all cohorts, i.e. soils.

3.2.2 Direct measure of performances of identification strategies

Figure 7 represents the evolution of the $\text{CVaR}_{30\%}(\text{YE})$ for all cohorts through the years (Equation 11). On average, farmers following the nitrogen fertilizer recommendations based on the BCB strategy had higher empirical CVaR at 30% of YE than farmers following those from ETC strategies, from the second year of the experiment onwards (Figure 6). The difference in performance between BCB and ETC is 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 respectively from the ETC-3 and the ETC-5 identification strategies. BCB allowed to identify faster the optimal 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, for the whole duration of the experiment.

3.2.3 Regret

Mean cumulated regret of the farmer population Figure 7 represents the evolution of the mean regret for all cohorts through the years (Equation 14). For $\alpha = 30\%$, BCB identification strategy outperformed ETC strategies, regardless of the number of years during which the strategy was applied. The difference in performance between BCB and ETC increases for the whole duration of the experiments. After 20 years, farmers following recommendations from BCB identification strategy experiences 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 all 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 BCB strategy was more robust (see quantile ranges in Figure 7) for this decision problem.

Individual cumulated regret distribution BCB prevented farmers from accumulating large individual cumulated regret during the participatory identification of the group (Figure 8): individual cumulated regrets for BCB were distributed towards lower values than for ETC strategies. With BCB, almost no individual cumulated regret was greater than 7.5 t/ha after 20 years, as opposed to ETC strategies. Consequently, BCB allowed a fairer sharing of identification mistakes in the population of farmers than ETC strategies.

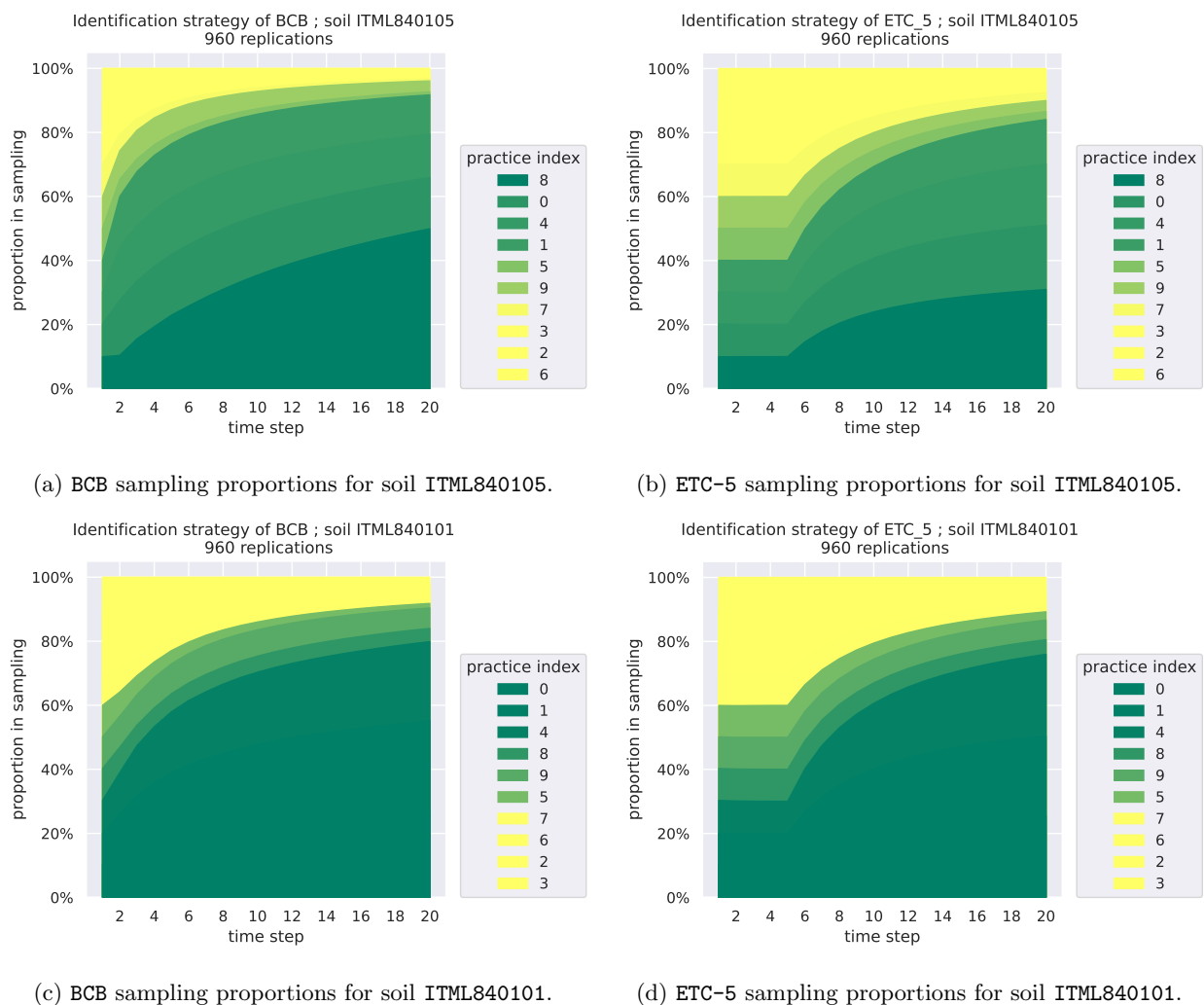


Figure 5: Averaged sampling proportions for soils ITML840105 and ITML840101, $T = 20$ years. 960 replications of the whole experiment were done. The fertilizer practices are ordered according to the true Conditional Value-at-Risk at level 30% (CVaR) of their Yield Excess (YE) with $ANE_{ref}=15$ kg grain/kg N ; the greener the color, the better a fertilizer practice is.

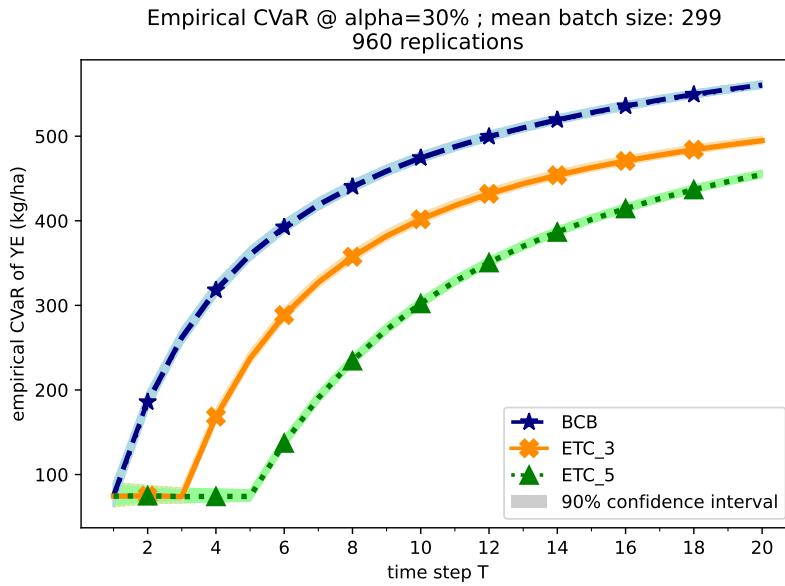


Figure 6: Empirical conditional Value-at-Risk (CVaR) at level 30% (CVaR) of maize yield excesses (YE) between $T = 0$ and the considered T ; $ANE_{ref} = 15$ kg grain/kg N. 960 replications of the whole experiment were done. One time step T is one year; ‘mean batch size’ is the number of farmers who have volunteered to participate at the trials, averaged over all years and all replications. Confidence intervals were computed following [Thomas and Learned-Miller \(2019\)](#).

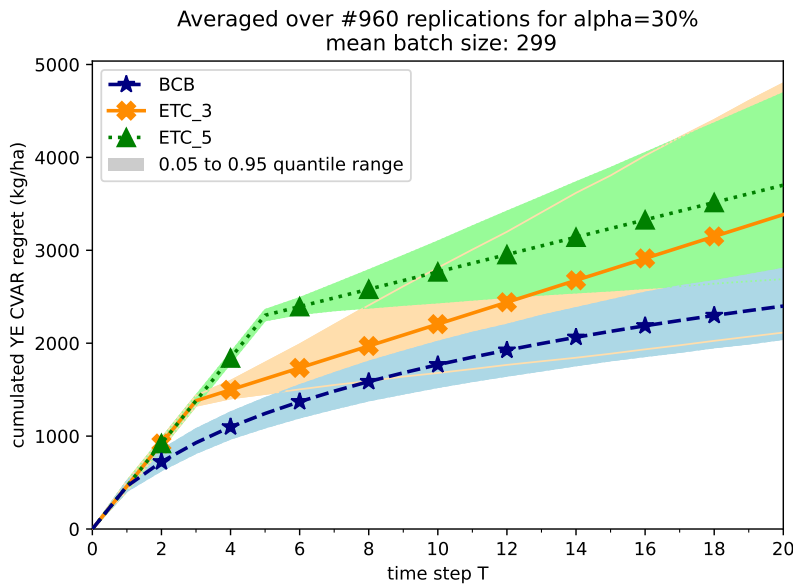


Figure 7: Mean cumulated regret of population, for the Conditional Value-at-Risk (CVaR) at level 30% of Yield Excess (YE); $ANE_{ref} = 15$ kg grain/kg N. The cumulated regret is averaged over the farmers’ population, between $T = 0$ and the considered T . 960 replications of the whole experiment were done. One time step T is one year, ‘mean batch size’ is the number of farmers who have volunteered to participate in the trials, averaged over all years and all replicates.

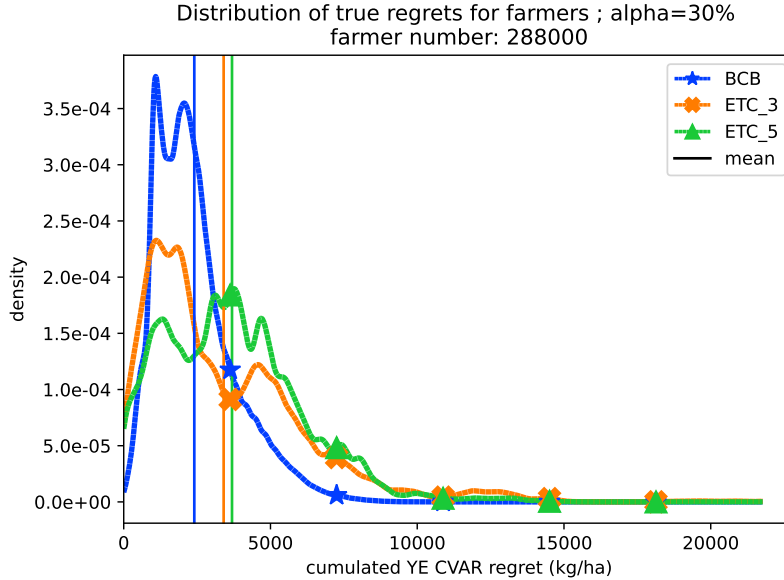


Figure 8: Distribution of individual cumulated regret after $T = 20$ years for Conditional Value-at-Risk at level 30% (CVaR) of the yield excess (YE) ; $ANE_{ref} = 15$ kg grain/kg N. The total number of farmers corresponds to a group of 300 farmers, with 960 replications of the whole experiment.

3.2.4 Sensitivity analysis

In Section C of Supplementary Materials, we present the same results than Sections 3.2.2 and 3.2.3 for higher CVaR levels of $\alpha = 50\%$ and $\alpha = 100\%$. The CVaR with the latter level recovers the usual expectation. For $\alpha = 50\%$, BCB showed similar performance than for $\alpha = 30\%$. For $\alpha = 100\%$, ETC-3 was the best performer, BCB and ETC-5 performed similarly. Nonetheless, BCB showed a smaller variance than both ETC-3 and ETC-5. The theoretical performance guarantee is presented in Section D of Supplementary Materials.

4 Discussion

4.1 Benefits from an adaptive identification strategy.

Practical perspective In multi-year multi-location on-farm trials, participating farmers simultaneously conduct field experiments with crops over multiple seasons to compare e.g. crop management practices (e.g. Naudin et al., 2010; Baudron et al., 2012; Falconnier et al., 2016). After a given number of years, results (often in terms of crop yields) are typically analyzed using mixed linear models (Laird and Ware, 1982), to take into account the design of an experiment with repeated measures, such as random effects associated with fields and farms. Best crop management practices are then identified based on this statistical analysis. In our simulated nitrogen fertilizer practice decision problem, we approximated multi-year on-farm trials with the ETC intuitive identification strategy. Both replicated on-farm trials and ETC consist of an exploration phase of a fixed duration (data collection), followed by an exploitation phase (application of the best identified practice after analysis of collected data). Consequently, both replicated on-farm trials and ETC can be considered as non-adaptive identification strategies: before the end of the exploration phase, the intermediary results are not exploited to gradually refine the experimental setup. In contrast, BCB refines its the recommendations every year, based on the results observed so far. The better a crop management option, the more its proportion in recommendations should increase with time. From a farmer’s perspective, this mean that the probability of highly sub-optimal recommendation decreases with time, as opposed to non-adaptive identification strategies during the exploration phase, which equi-proportionally recommend all crop management practices. Because with the bandit-algorithm-based identification strategy yield losses are reduced in most cases, compared to the non-adaptive identification strategies, the cost of the identification of best management practices is likely

to be reduced for the farmers. Another common method to generate crop management recommendation consists in the use of calibrated crop models and scenario analyses (e.g. Huet et al., 2022). This method can be complementary to the bandit-based approach. Candidate crop management practices can first be determined based on crop modeling results for the considered conditions, and then best crop management can be identified with the bandit algorithm..

Theoretical perspective ETC is theoretically proven to be a sub-optimal identification strategy without a calibration of the duration of the exploration phase that requires unavailable strong prior knowledge on the complexity of the decision problem (Lattimore and Szepesvári, 2020, Chapter 6). In numerical experiments, for $\alpha = 100\%$, ETC-3 best performed, probably because with these particular YE distributions and size of farmer group, 3 years of exploration was an optimal number. A slight change in the decision problem may induce that 3 or 5 years of exploration phase are no longer optimal (e.g. changing α to 30% or 50%). More generally, prior to an experiment, there is no guarantee than an arbitrate number of years of exploration of ETC will be optimal, and consequently there are no guarantees about the performance ETC, as opposed to BCB (see theoretical results in Section D). The main benefit of BCB over ETC is that it does not require a choice on parameters that require prior knowledge that is *a priori* not available. BCB neither requires strong assumptions about probability laws of reward distributions, as opposed to other common bandit algorithms. The only requisite for BCB is the knowledge of the maximum observable reward. In agronomy, such knowledge is easily available with expert knowledge: for instance, considering yield as reward, an expert can easily estimate a yield potential in the given crop growing conditions, for instance through modeling (Affholder et al., 2013).

4.2 Performances of fertilizer practices

For all soils, no optimal nitrogen fertilizer practice was threshold-based, nor shown split top-dressing. This does not discard a potential benefit from the threshold-based fertilizer practices, or split top dressing. Model simulations revealed that the effect of the nitrogen-stress or rainfall threshold depended on each soil, and the effect of the thresholds was not easy to anticipate. Consequently, the definition of the set of candidate fertilizer practices to explore must be carefully selected within the vast possible combination of practice parameters, e.g. application timing, rates, thresholds or number of split. In the experiments, the optimal values of practice parameters were not adjusted, because our focus was on designing a better generic identification method, rather than on designing refined parametrized fertilizer practices. For an application in real field conditions, we recommend these parameters to be adjusted based on simulations (see Section 4.1) and/or on prior small test plots. More generally, the design of fertiliser practices must include experts, local extensionists and farmers themselves (Cerf and Meynard, 2006; Hochman and Carberry, 2011). For instance, the maximum quantity of nitrogen fertilizer a farmer can apply may depend on the availability of fertilizer in the local market, and on the economic situation of each farmer.

4.3 Definition of farmers' objective

We defined the farmers' objective as maximizing the CVaR at level $\alpha = 30$ of the YE with $ANE_{\text{ref}} = 15$ kg grain/kg N. This quantity is interpretable as it represents a yield gain compared to a reference fertilizer practice, and it is easily calculable. The value of α allows to adjust the risk aversion level for a cohort of farmers. The value of ANE_{ref} defines an invariant economical and environmental trade-off which penalizes more or less the use of nitrogen fertilizer. Losses were defined as the expected performance difference between the best available nitrogen fertilizer practice, and the sub-optimal nitrogen fertilizer practices, in the face of the seasonal uncertainty.

However, we did not directly evaluate fertilizer practices by their economic return. Despite market risks being a reality, the economic return of maize nitrogen fertilization depends on many parameters changing through time, such as fertilizer subsidies, fertilizer market price, application costs, or harvest selling price. Because each year, the optimal nitrogen fertilizer practice is likely to change, such setting dramatically increases the complexity of the identification problem, and so does the required amount of data to identify best practices (we provide more details in Supplementary Materials, Section E). In any case, modelers should bear in mind the inherent limitations of the modeling of a farmer's objective, which always remains a proxy (McCown, 2002).

4.4 Limits and possible improvements

In our simulated crop management decision problem, we largely simplified the experimental structure of multi-year replicated field trials. First, for all simulations, weather time series were independent and identically distributed. Such assumption is unlikely to be true in the real world. During the same year, weather spatial correlations can be high, for instance in case of extreme weather events (Tack and Holt, 2016). Second, within the same cohort, all farmers were supposed to have exactly the same soil and cultivar, and to implement closely the fertilizer practice they were assigned. For real application, a farmer’s soil, site, year and other potential random effects should be properly considered. The bandit identification strategy we introduced should be extended to account for experimental structure and multiple factors at stake. For instance, contextual bandits (Lattimore and Szepesvári, 2020), which would allow to share information between decision contexts (here, the cohorts), might offer solutions.

As another limit, in simulations, we considered climate to be the same during the 20 years of the experiment. Such hypothesis is improbable in real conditions (e.g. Traore et al., 2017). Nevertheless, as Adam et al. (2020) has shown based on simulations, in Mali, improving current crop management, in particular nitrogen fertilization, may compensate the long-term effects of climate change, while addressing the urgent necessity of closing yield gaps. For a decision problem perspective, if climate changes through time, then optimal practices are likely to change with time. Such problem can be formalized as a non-stationary bandit problem (Lattimore and Szepesvári, 2020). To handle non-stationary, BCB can be equipped with a sliding window (Garivier and Moulines, 2011; Baudry et al., 2021b). This mechanism forces the bandit algorithm to overlook observations older than a given number of years, which consequently must regularly re-evaluate all fertilizer practices. Such approach reiterates the recommendations formulated by Adam et al. (2020): the bandit algorithm would handle climate change by regularly trying to improve current fertilizer practices.

5 Conclusion

We addressed the problem of the identification of best maize fertilizer practices, supported by virtual trials of a community of smallholder farmers. Our goal was to provide an identification strategy that minimized farmers’ yield losses occurring during field trials, compared to the ‘intuitive strategy’ which consisted of multi-year field trials with an equal proportion of each fertilizer practice tested each growing season. In simulated experiments mimicking the conditions of southern Mali as a case study, the bandit-based identification strategy we introduced showed better resulted in reducing farmers’ yield losses. We used model simulations to compare the identification strategies, but the bandit-based identification strategy does not depend on the simulations. This novel approach opens up new perspectives as an alternative to the usual multi-year on-farm trials. It can also complement the use of calibrated crop simulation models for the formulation of fertilizer recommendations. For instance, if parameterized fertilizer practices are explored, the choice of the parameters could be made based on prior simulations and/or small-scale field trials.

However, before confirming that the bandit-based identification strategy can be employed to identify best management practices in real conditions, several constraints must be addressed: (i) the structure of multi-year on-farm field trials with repeated measures and thus correlations, such that the effect, for the same year in a given region, of spatial correlation of weather series; (ii) the effect of climate change; (iii) the effect of farmer compliance, who may not strictly follow the fertilizer practice recommendations, which may induce extra noise. We briefly indicated how our approach can be extended to address the aforementioned constraints.

Software availability

All the numerical experiments in this paper are meant to be as reproducible as possible, and the code is open source. The Python code with the necessary packages, instructions and experimental data are provided in the following public GitLab repository: <https://gitlab.inria.fr/rgautron/batch-cvts/-/tree/master>. The simulations are performed with `gym-dssat` (https://gitlab.inria.fr/rgautron/gym_dssat_pdi), a modified version of the Decision Support System for Agrotechnology Transfer (DSSAT) software (<https://dssat.net/>).

References

- Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26:1487–1503.
- Adam, M., MacCarthy, D. S., Traoré, P. C. S., Nenkam, A., Freduah, B. S., Ly, M., and Adiku, S. G. (2020). 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.
- Affholder, F. (1995). Effect of organic matter input on the water balance and yield of millet under tropical dryland condition. *Field Crops Research*, 41(2):109–121.
- Affholder, F., Poeydebat, C., Corbeels, M., Scopel, E., and Tittonell, P. (2013). 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.
- Agrawal, S., Koolen, W. M., and Juneja, S. (2021). 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*.
- Baudron, F., Tittonell, P., Corbeels, M., Letourmy, P., and Giller, K. E. (2012). Comparative performance of conservation agriculture and current smallholder farming practices in semi-arid zimbabwe. *Field crops research*, 132:117–128.
- Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. (2021a). Optimal thompson sampling strategies for support-aware cvar bandits. In *International Conference on Machine Learning*, pages 716–726. PMLR.
- Baudry, D., Russac, Y., and Cappé, O. (2021b). On Limited-Memory Subsampling Strategies for Bandits. In *ICML 2021- International Conference on Machine Learning*, Vienna / Virtual, Austria.
- Cassel, A., Mannor, S., and Zeevi, A. (2018). A general approach to multi-armed bandits under risk criteria. In *Conference On Learning Theory*, pages 1295–1306. PMLR.
- Cerf, M. and Meynard, J.-M. (2006). Les outils de pilotage des cultures: diversité de leurs usages et enseignements pour leur conception. *Natures Sciences Sociétés*, 14(1):19–29.
- Cerf, M. and Sebillotte, M. (1997). Approche cognitive des décisions de production dans l’exploitation agricole [confrontation aux théories de la décision]. *Economie rurale*, 239(1):11–18.
- Dowd, K. (2007). *Measuring market risk*. John Wiley & Sons.
- Evans, K. J., Terhorst, A., and Kang, B. H. (2017). From data to decisions: helping crop producers build their actionable knowledge. *Critical reviews in plant sciences*, 36(2):71–88.
- Everitt, B. and Skrondal, A. (2002). *The Cambridge dictionary of statistics*, volume 106. Cambridge University Press Cambridge.
- Falconnier, G. N., Descheemaeker, K., Van Mourik, T. A., and Giller, K. E. (2016). Unravelling the causes of variability in crop yields and treatment responses for better tailoring of options for sustainable intensification in southern mali. *Field Crops Research*, 187:113–126.
- Fosu-Mensah, B., MacCarthy, D., Vlek, P., and Safo, E. (2012). 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.
- Garivier, A. and Moulines, E. (2011). On upper-confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory*, pages 174–188. Springer.
- Gautron, R., Maillard, O.-A., Preux, P., Corbeels, M., and Sabbadin, R. (2022). Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture*, 200:107182.

- Gautron, R. and Padrón González, E. J. (2022). gym-DSSAT - A crop model turned into a Reinforcement Learning environment.
- Getnet, M., Van Ittersum, M., Hengsdijk, H., and Descheemaeker, K. (2016). Yield gaps and resource use across farming zones in the central rift valley of ethiopia. *Experimental Agriculture*, 52(4):493–517.
- Hanway, J. (1963). Growth stages of corn (zea mays, l.) 1. *Agronomy Journal*, 55(5):487–492.
- Hochman, Z. and Carberry, P. (2011). Emerging consensus on desirable characteristics of tools to support farmers’ management of climate risk in australia. *Agricultural Systems*, 104(6):441–450.
- Hoogenboom, G., Porter, C., Boote, K., Shelia, V., Wilkens, P., Singh, U., White, J., Asseng, S., Lizaso, J., Moreno, L., et al. (2019). The dssat crop modeling ecosystem. *Advances in crop modelling for a sustainable agriculture*, pages 173–216.
- Huet, E., Adam, M., Traore, B., Giller, K., and Descheemaeker, K. (2022). 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.
- Jourdain, D., Lairez, J., Striffler, B., and Affholder, F. (2020). 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.
- Kalaji, H. M., Dabrowski, P., Cetner, M. D., Samborska, I. A., Lukasik, I., Brestic, M., Zivcak, M., Tomasz, H., Mojski, J., Kociel, H., et al. (2017). A comparison between different chlorophyll content meters under nutrient deficiency conditions. *Journal of Plant Nutrition*, 40(7):1024–1034.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.
- Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.
- Mandelbrot, B. B. (1997). The variation of certain speculative prices. In *Fractals and scaling in finance*, pages 371–418. Springer.
- Massart, P. (1990). The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Annals of Probability*, 18.
- McCown, R. L. (2002). Changing systems for supporting farmers’ decisions: problems, paradigms, and prospects. *Agricultural systems*, 74(1):179–220.
- Meisinger, J. J. and Delgado, J. A. (2002). Principles for managing nitrogen leaching. *Journal of soil and water conservation*, 57(6):485–498.
- Menapace, L., Colson, G., and Raffaelli, R. (2013). Risk aversion, subjective beliefs, and farmer risk management strategies. *American Journal of Agricultural Economics*, 95(2):384–389.
- Morris, T. F., Murrell, T. S., Beegle, D. B., Camberato, J. J., Ferguson, R. B., Grove, J., Ketterings, Q., Kyveryga, P. M., Laboski, C. A., McGrath, J. M., et al. (2018). Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. *Agronomy Journal*, 110(1):1.
- Naudin, K., Gozé, E., Balarabe, O., Giller, K. E., and Scopel, E. (2010). Impact of no tillage and mulching practices on cotton production in north cameroon: a multi-locational on-farm assessment. *Soil and Tillage Research*, 108(1-2):68–76.
- Perchet, V., Rigollet, P., Chassang, S., and Snowberg, E. (2015). Batched bandit problems. In Grünwald, P., Hazan, E., and Kale, S., 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.
- Richardson, C. W. and Wright, D. A. (1984). Wgen: A model for generating daily weather variables. *ARS (USA)*.

- Ripoche, A., Crétenet, M., Corbeels, M., Affholder, F., Naudin, K., Sissoko, F., Douzet, J.-M., and Tittonell, P. (2015). 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.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535.
- Soltani, A. and Hoogenboom, G. (2003). A statistical comparison of the stochastic weather generators wgen and simmeteo. *Climate Research*, 24(3):215–230.
- Tack, J. B. and Holt, M. T. (2016). The influence of weather extremes on the spatial correlation of corn yields. *Climatic Change*, 134(1-2):299–309.
- Tamkin, A., Keramati, R., Dann, C., and Brunskill, E. (2020). Distributionally-aware exploration for cvar bandits. In *NeurIPS 2019 Workshop on Safety and Robustness in Decision Making; RLDM 2019*.
- Ten Berge, H. F., Hijbeek, R., Van Loon, M., Rurinda, J., Tesfaye, K., Zingore, S., Craufurd, P., van Heerwaarden, J., Brentrup, F., Schröder, J. J., et al. (2019). Maize crop nutrient input requirements for food security in sub-saharan africa. *Global Food Security*, 23:9–21.
- Thomas, P. and Learned-Miller, E. (2019). Concentration inequalities for conditional value at risk. In *International Conference on Machine Learning*, pages 6225–6233. PMLR.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.
- Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., and Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, 418(6898):671–677.
- Traore, B., Descheemaeker, K., Van Wijk, M. T., Corbeels, M., Supit, I., and Giller, K. E. (2017). Modelling cereal crops to assess future climate risk for family food self-sufficiency in southern mali. *Field Crops Research*, 201:133–145.
- Vanlauwe, B., Kihara, J., Chivenge, P., Pypers, P., Coe, R., and Six, J. (2011). 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.

Supplementary Materials

A Maize simulations

The cultivation scenarios were based on the the conditions found in Southern Mali. The soils came from Adam et al. (2020) who compiled and supplemented with survey data the soils found in the literature for the location of Koutiala, Mali. The data of Adam et al. (2020) 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 Adam et al. (2020). During the simulations, the weather times series were generated using the WGEN weather model (see Richardson and Wright, 1984; Soltani and Hoogenboom, 2003). WGEN had been calibrated on 40 year long historical daily weather records from a weather station located in N’Tarla found in Ripoche et al. (2015), which was located about 20 km from Koutiala ; these historical weather records were the best available. The cultivars used in the simulation and its parametrization in DSSAT are presented in Table A.1 ; 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 A.2. For each soil, the initial soil nitrogen content was set according to the values found in Adam et al. (2020). 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 A.1 shows the simulated yield distributions for ITML840104 and ITML840105 soils.

Table A.1: 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

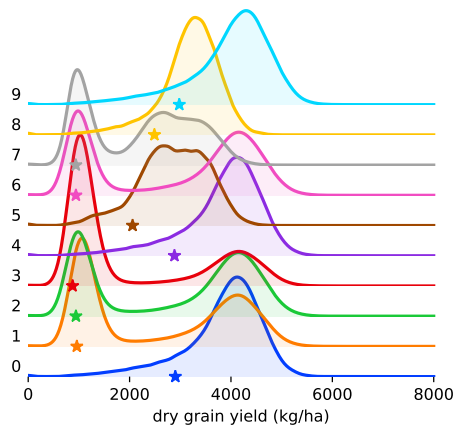
Table A.2: Automatic planting parametrization in DSSAT. PFRST: Starting date of the planting window; PLAST: End date of the planting window; PH2OL: Lower limit on soil moisture for automatic planting; PH2OU: Upper limit on soil moisture for automatic planting; PH2OD: Depth to which average soil moisture 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

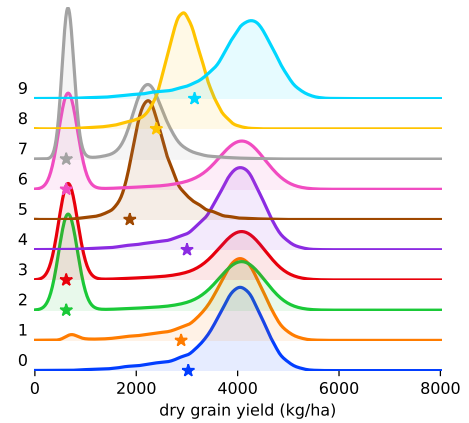
B Algorithms

B.1 Details about BCB

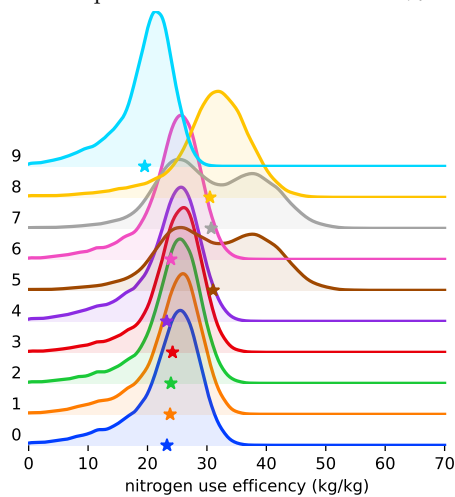
In algorithm B.1, we provide the detailed pseudo-code of BCB (BCB). As shown by Figure B.1, the higher the number of collected rewards, the less the weights sampled from Dirichlet distributions exhibit variance.



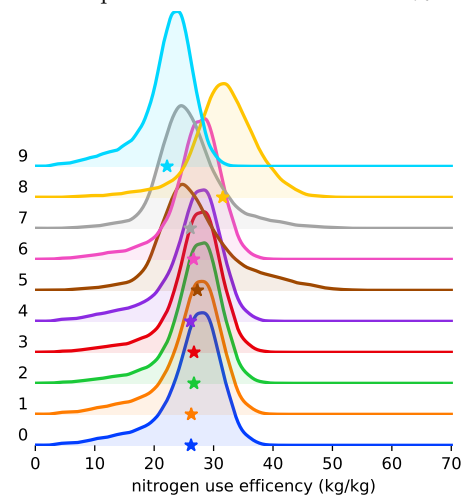
(a) Yield distributions for soil ITML840104. Stars represent the CVaR at level 30%.



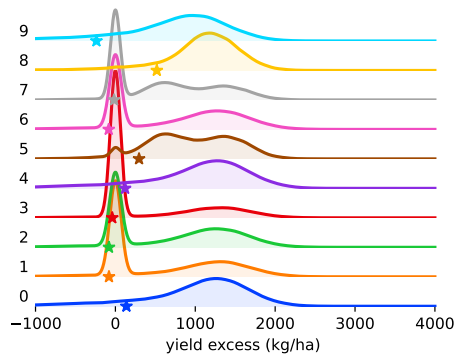
(b) Yield distributions for soil ITML840105. Stars represent the CVaR at level 30%.



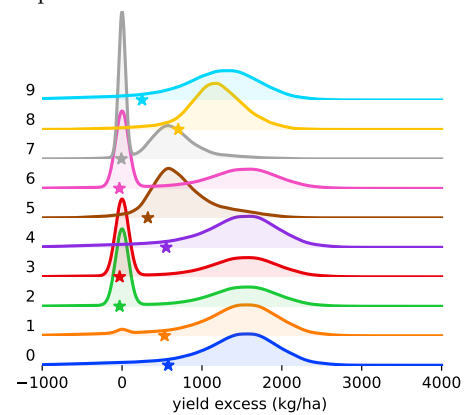
(c) Agronomic Nitrogen Efficiency (ANE) distributions for soil ITML840104. Stars represent the mean value.



(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 A.1: 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.

This variance directly relates to the noise introduced in the computation of the score of the different available actions.

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

// 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 $\mathcal{F}_t^c = \{f_1, \dots, f_{n_t}\}$; // 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 : $\hat{c}_{k,t-1} = \hat{C}_\alpha(\mathcal{X}_k)$

 end

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

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

 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 $\omega_k = \{w_1, \dots, w_{N_k}\} \sim \mathcal{D}_{N_k}$; // Dirichlet of concentration parameter $\underbrace{(1, \dots, 1)}_{N_k \text{ times}}$

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

 Sort \mathcal{X}_k in increasing order

 Compute $\tilde{c}_k = x_j - \frac{1}{\alpha} \sum_{i=1}^{N_k} w_i \max(x_j - x_i, 0)$; // assign a score to action k

 end

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

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

 end

 Sort the set of farmers \mathcal{F}_t^c according their increasing empirical regrets $l_{f,t-1}$

 Sort the set of actions \mathcal{A}_t according their increasing empirical CVaR $\hat{c}_{k,t-1}$

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

 Assign action a to farmer f ; // fair exploration

 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

Remark B.1 (First season). *Algorithm B.1 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 B.2) without changing the theoretical properties of our algorithm. Furthermore, the decision maker could also include any additional results collected before the*

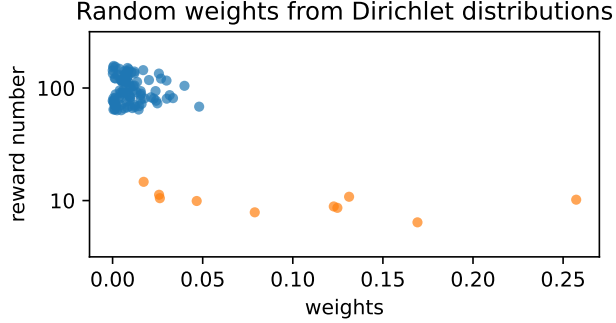


Figure B.1: Examples of weights sampled from Dirichlet distributions during BCB execution, respectively for 10 and 100 rewards. The greater the number of rewards, the less variance the weights show. The variance of weights is related to the noise level in the computation of the empirical CVaR of BCB.

experiment (if the practices has already been tested for some time) in the initialization of the algorithm.

B.2 Explore-Then-Commit (ETC)

We provide the pseudo-code of the Explore-Then-Commit (ETC) strategy with algorithm B.2. 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 B.1). 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 the scores, encouraging exploration even for sub-optimal practices, as it raises up their initial values when few rewards have been observed.

Algorithm B.2 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.: $\forall k \in \{1, \dots, K\} : N_k = 0$

// Do trials during t_{trials} years

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

// Beginning of the season t

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

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

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

end

$a_{\text{max}} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \hat{c}_k$;

// get the action that best performed during trials

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

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

// Beginning of the season t

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

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 : $r_{f,a_{\text{max}}}$

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

end

end

C Experiment complements

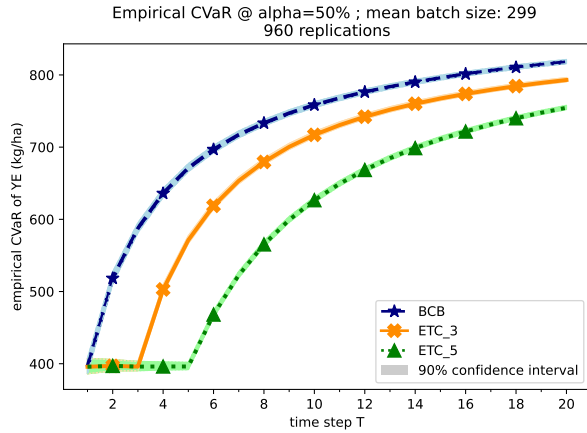
Following methods of Section 2 of the main text, we provide identification performances of identification strategies for CVaR levels $\alpha = 50\%$ and $\alpha = 100\%$ with Figures C.1, C.2 and C.3. For both CVaR levels, the YE is defined with $\text{ANE}_{\text{ref}} = 15$ kg N/kg grain.

D Theoretical Analysis

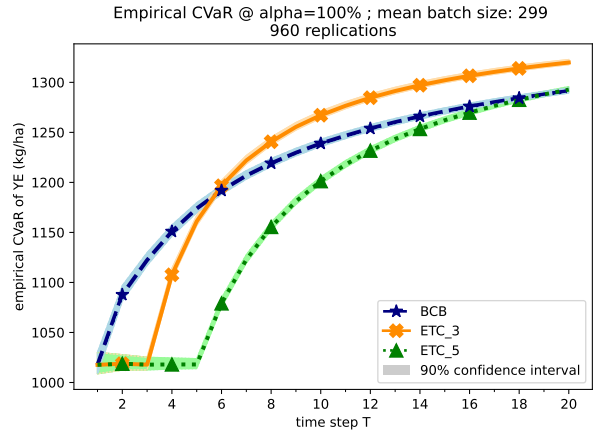
This section is devoted to the theoretical analysis of the BCB algorithm. We will mostly adapt the analysis of Baudry et al. (2021a), and show that the problem of learning with batched data of finite upper bounded size is no harder than the pure online learning problem considered in the original paper.

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 $c_1 = \operatorname{argmax}_{k=1, \dots, K} c_k$. Assume that BCB runs for T seasons, and that at each season the size of the batch is $n_T \leq F \in \mathbb{N}$. Then, for any $\varepsilon > 0$ small enough there exists some $\varepsilon_1 > 0, \varepsilon_2 > 0$ such that the regret of BCB satisfies*

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

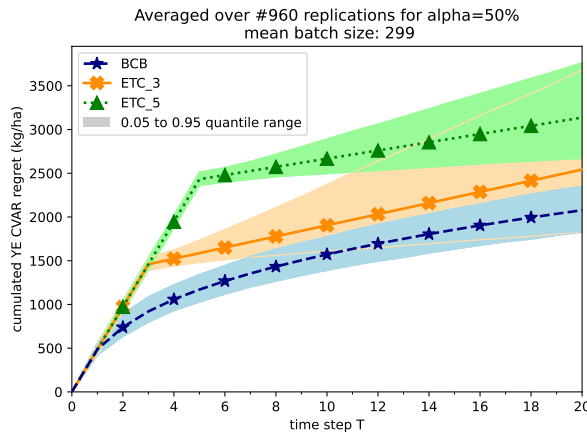


(a) $\alpha = 50\%$

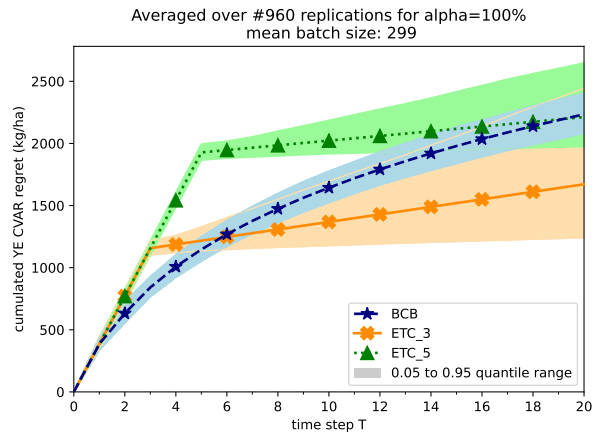


(b) $\alpha = 100\%$

Figure C.1: Farmers' empirical CVaR at level of all YE received between $T = 0$ and the considered T .

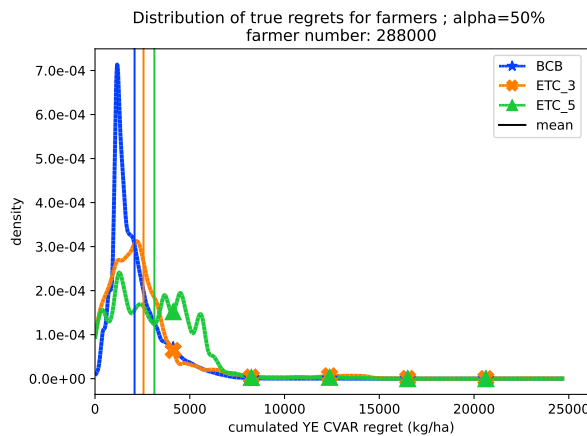


(a) $\alpha = 50\%$

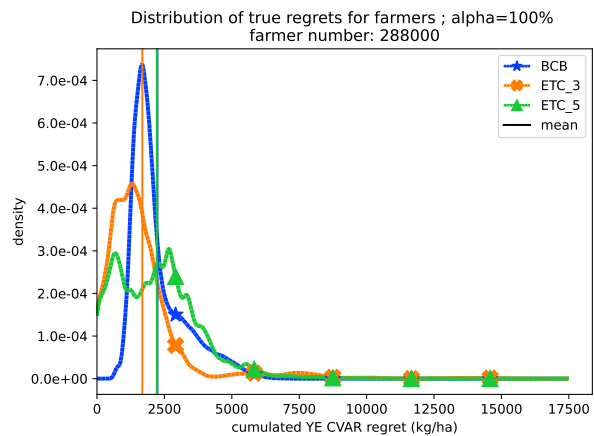


(b) $\alpha = 100\%$

Figure C.2: Cumulated regret averaged over the population for the CVaR at level of YE.



(a) $\alpha = 50\%$



(b) $\alpha = 100\%$

Figure C.3: Distribution of individual cumulated regret after $T = 20$.

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

It is interesting to compare this regret upper bound to the one obtained in the purely sequential setting, that we recall in Theorem D.2.

Theorem D.2 (α -CVaR Regret of B-CVTS with time horizon S_T (adapted from Theorem 3 in Baudry et al. (2021a))). *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 $c_1 = \operatorname{argmax}_K c_k$. Consider a number of data collected S_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*

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

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

First, we see that if F is indeed a constant (i.e do not depend on the time) then when T is large enough then F has not impact on the scaling of the regret. In our proof the main impact of the batch setting is an additive term F for each arm, hence the regret becomes close to the one of the sequential setting once $m_T^k \gg F$. Finally, if the number of farmers in each batch is exactly F at each step then $S_T = FT$ and, $m_T^k = n_{S_T}^k$, hence the asymptotically dominant (logarithmic) term is the same in the two settings.

These theoretical results show that learning with batch feedback does not introduce theoretical limitations in our setting, and so the BCB algorithm is theoretically grounded.

Proof of Theorem D.1. As in the proof of Baudry et al. (2021a) 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

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

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

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

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

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

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 ([Massart, 1990](#)) we obtain

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

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

Upper bounding A_2 The term A_2 is then handled with similar tricks, and the arguments used in [Baudry et al. \(2021a\)](#).

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

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 continuity of $\mathcal{K}_{\text{inf}}^{\alpha, \mathcal{D}}$ in its two arguments as proved in [Agrawal et al. \(2021\)](#), we obtain that for any $\varepsilon > 0$ small enough there exist some $\varepsilon_1, \varepsilon_2$ such that

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

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

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

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 [Baudry et al. \(2021a\)](#). 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:

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

where in the second term we count the number of recommendations provided by the algorithm, assigning those in the same batch an arbitrary order, $\tilde{c}_1(r)$ 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 [Baudry et al. \(2021a\)](#), 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. \square

E 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.

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

$$\text{VCR}^\pi = \frac{p_{\text{maize}}}{p_{\text{N}}} \times \frac{Y^\pi - Y^0}{N^\pi} \tag{E.1}$$

$$= \frac{p_{\text{maize}}}{p_{\text{N}}} \times \text{ANE}^\pi \tag{E.2}$$

where p_N is fertilizer unitary cost and p_{maize} unitary maize grain selling price. Remarking that each given year the ratio $\frac{p_{\text{maize}}}{p_N}$ is shared by all fertilizer practices. 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. 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.