Budgeted Online Active Learning with Expert Advice and Episodic Priors

Kristen Goebel[1], William Solow[1], Paola Pesantez-Cabrera[2], Markus Keller[3], Alan Fern[1]

[1]School of Electrical Engineering and Computer Science, Oregon State University
[2]School of Electrical Engineering and Computer Science, Washington State University
[3]Irrigated Agriculture Research and Extension Center, Washington State University
{goebelk, soloww, afern}@oregonstate.edu, {p.pescantezcabrera, mkeller}@wsu.edu

Abstract

This paper introduces a novel approach to budgeted online active learning from finite-horizon data streams with extremely limited labeling budgets. In agricultural applications, such streams might include daily weather data over a growing season, and labels require costly measurements of weather-dependent plant characteristics. Our method integrates two key sources of prior information: a collection of preexisting expert predictors and episodic behavioral knowledge of the experts based on unlabeled data streams. Unlike previous research on online active learning with experts, our work simultaneously considers query budgets, finite horizons, and episodic knowledge, enabling effective learning in applications with severely limited labeling capacity. We demonstrate the utility of our approach through experiments on various prediction problems derived from both a realistic agricultural crop simulator and real-world data from multiple grape cultivars. The results show that our method significantly outperforms baseline expert predictions, uniform query selection, and existing approaches that consider budgets and limited horizons but neglect episodic knowledge, even under highly constrained labeling budgets.

Code — https://tinyurl.com/BOAL-AgAID

1 Introduction

Consider a vineyard manager forecasting grapevine cold hardiness during dormancy—a property dictated by the weather history. Accurate forecasts guide frost mitigation to prevent crop loss, but ground-truth measurements are costly and slow, requiring field sampling and lab work. With only 2–5 samples per season possible, two questions arise: when should samples be taken, and how can we learn from so few?

This scenario illustrates the broad problem of *budgeted online active learning (BOAL)*, where a function is learned via strategic label queries under the following constraints:

1. **Online querying**: Query decisions must be made sequentially as data arrives
2. **Finite horizon**: A fixed time window for data collection
3. **Strict query budgets**: A fixed query budget with a focus on small-budget scenarios with 2–10 labels

These constraints arise across domains such as agricultural monitoring, environmental sensing, industrial control, and personalized medicine. As discussed in Section 2, existing active learning methods do not jointly address online querying, finite horizons, and fixed-budgets—particularly very small budgets. We bridge this gap with a framework that exploits two forms of prior knowledge: expert advice and historical episodic data.

First, we adapt the classic paradigm of online learning from expert advice, equipping the learner with a set of preexisting predictors likely to include a “good enough” option. This allows rapid learning by quickly identifying strong experts, even with few labels. While active learning with experts has been explored, prior work has not addressed strict budgets within finite horizons.

Second, we propose query selection methods that leverage historical episodes of the input stream (e.g., decades of weather). Building on recent work in prophet-inequality problems and introducing a simple adaptive strategy, we use prior data to guide the key choice: sample now or wait for a better time. To our knowledge, this is the first use of episodic data for query selection in BOAL.

We evaluate our approach on BOAL tasks derived from a multi-crop agricultural simulator and decades of real grape cold-hardiness data. Results show consistent gains over baseline predictors, uniform query selection, and methods that ignore historical data, even with very small sample sizes. In summary, our contributions are:

1. A general BOAL framework for extremely small budgets.
2. The first use of the learning-from-experts paradigm in BOAL.
3. The first use of prior unlabeled episodic data for online query selection.
4. Empirical validation on realistic simulated and real agricultural data.

2 Related Work

Most active learning research follows a pool-based paradigm, where unlabeled data is freely sampled for labeling. Early work explored online settings (Atlas, Cohn, and Ladner 1990), leading to algorithms such as query-by-committee (Seung, Opper, and Sompolinsky 1992; Freund et al. 1997). However, none address all aspects of our setting: online querying, finite horizons, and strict budgets. We summarize the closest categories below.

Online Active Model Selection. Related pool-based approaches (e.g., (Hara et al. 2024)) select from a fixed set of models but do not apply to online problems. Online variants exist (Karimi et al. 2021; Liu et al. 2022) but do not explicitly handle budget or horizon constraints and target classification, not real-valued labels essential to our applications.

Active Learning with Expert Advice. Online expert advice has been widely studied (Littlestone and Warmuth 1994; Freund and Schapire 1997; Cesa-Bianchi et al. 1997; Herbster and Warmuth 1998), with active learning variants aiming to limit labels (Cesa-Bianchi, Lugosi, and Stoltz 2005; Mitra and Gopalan 2020; Zhao, Hoi, and Zhuang 2013; Hao et al. 2018; Kumar, Abernethy, and Saligrama 2022; Castro, Hellström, and van Erven 2024; Truong, Etesami, and Kiyavash 2021). Most focus on classification, with few handling real-valued labels, and none enforce fixed budgets over finite horizons.

Budget-Aware Online Active Learning. Prior work combines online learning with budgets but ignores finite horizons (Zhang et al. 2018). More relevant are methods for sample selection along robot trajectories (Das et al. 2015; Luo, Nam, and Sycara 2017), which address both constraints via submodular secretary algorithms (Bateni, Hajiaghayi, and Zadimoghaddam 2013). These methods are non-adaptive, as observed labels do not inform future queries. An adaptive variant (Fujii and Kashima 2016) was later proposed, but it does not exploit prior episodic knowledge and is only tested on synthetic i.i.d. streams.

No prior work leverages unlabeled stream statistics to inform online query selection. In contrast, we leverage that historical data using prophet inequalities and a new empirical thresholding method, yielding gains under tight budgets and non-i.i.d. streams.

3 Budgeted Online Active Learning

We study *budgeted online active learning (BOAL)* over a data stream $x_{1:T} = (x_1, \ldots, x_T)$ with known horizon T, where each $x_t \in \mathbb{R}^d$ is a d-dimensional feature vector. The goal is to approximate a target function $f : x_{1:t} \to \mathbb{R}$ mapping the input history to a value. For instance, in agriculture, x_t may represent daily weather, T the season length, and f a weather-dependent plant property.

Our framework learns f via selective label queries under strict budget and time constraints. We focus on settings with a small query budget $B \ll T$ (e.g., 2–5 queries) and no ability to query past inputs. This is motivated by applications, such as, agriculture where each label requires costly field and lab work. The online active learning protocol proceeds as follows:

1. At each time $t \in \{1, \ldots, T\}$, the algorithm observes x_t.
2. It decides whether to query y_t, ensuring total queries do not exceed B.
3. If queried, $y_t = f(x_{1:t})$ is revealed.

This setting poses a challenge due to the interplay of three factors: online decision-making, a finite horizon, and a severely limited query budget. Unlike batch active learning, missed queries at time t cannot be recovered. This creates a core dilemma: *Given the current history, should the algorithm use a query now or wait for a potentially better opportunity before the horizon ends?* Our main contribution is a novel approach that leverages prior knowledge to address this tradeoff under extreme budget constraints.

4 Generic Approach

We now present the generic algorithm underlying our BOAL approach. Sections 5 and 6 describe our specific instantiation leveraging expert advice and episodic priors.

In pool-based active learning (Settles 2009), all unlabeled instances are available upfront, and selection is based on instance scores—often reflecting model uncertainty—to guide labeling. The model is updated immediately after each query. In contrast, the online setting prohibits maximizing over all scores: the algorithm must decide at each step whether to query the current instance, with no chance to revisit. We refer to this problem of selecting the highest-scoring instance from a stream as ONLINEMAX.

When all we know about a sequence of scores $(s_{t_0}, \ldots, s_{t_h})$ is that they arrive in random order with unknown values, the classic secretary algorithm (Dynkin 1963) optimally solves ONLINEMAX as follows:

1. Observe the first $t' = \lfloor (t_h - t_0 + 1)/e \rfloor$ scores.
2. $S_{\max} = \max\{s_1, \ldots, s_{t'}\}$
3. For $t \in \{t' + 1, \ldots, T\}$
   if ($s_t > S_{\max}$ or $t = t_h$) then select t and stop.

Despite its simplicity, this strategy achieves the optimal $1/e$ approximation ratio in expectation relative to the best offline value. Closer to online budgeted active learning, the *submodular secretary problem* (Bateni, Hajiaghayi, and Zadimoghaddam 2013) extends this setting by selecting B instances to maximize a submodular set function. The proposed algorithm applies the secretary strategy B times over evenly partitioned segments, achieving an expected approximation ratio of $(1 - 1/e)/7$ under the standard random-arrival model.

Our generic approach, like prior BOAL methods (Fujii and Kashima 2016; Luo, Nam, and Sycara 2017), follows the structure of the submodular secretary algorithm. As shown in Algorithm 1, we initialize a prediction model $\hat{f}$ and divide the stream into B segments of size T/B (assuming T is divisible by B). In each segment, an ONLINEMAX algorithm selects an instance based on a scoring function SCORE (Section 5), which may depend on the current model and prior stream data. The selected instance is used to update $\hat{f}$ before proceeding, allowing the algorithm to adapt over time.

Both prior methods are instances of this framework, with Luo et al. (2017) updating the model only after collecting all B labels. They train models from scratch using Gaussian Processes or linear SVMs with uncertainty-based SCORE functions—approaches ill-suited to extremely small budgets. In contrast, we address this limitation via expert advice (Section 5), enabling effective learning from few queries.

Algorithm 1: Generic Budgeted Online Active Learning

Input: Time horizon T, query budget B, data stream $x_{1:T}$
Output: Updated prediction model $\hat{f}$
Requires: ONLINEMAX (Sec. 6); UPDATE; SCORE (Sec. 5)

```
Initialize prediction model f̂
for i ← 1 to B do
    Initialize OnlineMax;
    for t ← ⌊T/B⌋(i − 1) + 1 to ⌊T/B⌋i do
        if OnlineMax(Score(x_t)) = select or t = ⌊T/B⌋i then
            y_t ← Query(x_t)
            Update(f̂, (x_t, y_t)) break;
return f̂
```

Prior evaluations are also limited: Luo et al. (2017) focus on sample selection along predefined robot trajectories, while Fujii and Kashima (2016) test on synthetic i.i.d. streams from classification benchmarks.

Another key distinction from prior work lies in the choice of ONLINEMAX. Previous methods rely on the classic secretary algorithm, which ignores prior knowledge of stream statistics and may result in poor query choices. For instance, if we can estimate the expected maximum score within a segment, the secretary algorithm may still select a much lower-scoring instance, wasting a query. Our approach, detailed in Section 6, addresses this by leveraging prior stream data to make more informed and efficient decisions.

5 Leveraging Expert Advice

To handle extremely limited budgets, we adopt the classic online learning framework of learning from expert advice (Cesa-Bianchi et al. 1997). This approach uses a set of expert predictors to enable effective learning from very few samples by rapidly identifying a good expert mixture. Below, we describe the framework and our specific instantiations of the UPDATE and SCORE subroutines used in Algorithm 1.

5.1 Classic Framework

The expert advice framework considers online prediction over a stream $x_{1:T} = (x_1, \ldots, x_T)$ with labels $y_t^* \in \mathcal{Y}$, without assumptions on stream or label characteristics. The learner is given N expert predictors $F = \{f_1, \ldots, f_N\}$, where each $f_i : x_{1:t} \to \mathcal{Y}$. While no assumptions are made about the experts, F is typically chosen to include at least one "good enough" predictor. At each time step t, the learner:

1. Observes x_t
2. Selects an expert $f_i \in F$ and predicts $y_t = f_i(x_{1:t})$
3. Observes the loss $l_{i,t} = l(f_i(x_{1:t}), y_t^*)$ for each expert $i \in \{1, \ldots, N\}$

The learner's cumulative loss is $L_T = \sum_{t=1}^{T} l(y_t, y_t^*)$, and each expert's is $L_{i,T} = \sum_{t=1}^{T} l_{i,t}$. The goal is to minimize expected regret, which is the difference between L_T and the best expert's loss, $\min_i L_{i,T}$.

A simple yet powerful algorithm for this framework is HEDGE (Freund and Schapire 1997), which maintains a weight $w_{j,t}$ for each expert j at time t, initialized to 1. At each step, the learner forms a distribution over experts:

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}},$$

and uses it to make a prediction. In our experiments with real-valued labels, we use the weighted average:

$$y_t = \sum_i p_{i,t} \cdot f_i(x_{1:t}). \tag{1}$$

After observing the losses, the weights are updated via:

$$w_{i,t+1} = w_{i,t} \cdot \exp(-\eta \cdot l_{i,t}), \tag{2}$$

where η is a learning rate. Poorly performing experts receive exponentially lower weight, and over time, the distribution favors better experts. With a suitable η, HEDGE guarantees expected cumulative regret of $O(\sqrt{T \log N})$, which approaches the performance of the best expert as T grows.

5.2 Active Framework

Here we instantiate the UPDATE and SCORE functions in Algorithm 1 for the experts framework.

UPDATE. Our online active learning setting differs from the classic expert framework in that labels are revealed only when a query is issued. We adapt the HEDGE algorithm for the UPDATE subroutine in Algorithm 1 as follows: when a query is made at time t, we observe expert losses and apply the weight update from Equation 2; otherwise, we set all expert losses to zero, leaving weights unchanged.

SCORE. Algorithm 1 also requires a SCORE subroutine suited for regression, as our applications involve real-valued labels. As noted in Section 2, prior work on active learning with expert advice for regression is limited. One such method (Truong, Etesami, and Kiyavash 2021) uses the maximum pairwise difference $\max_{i,j} |f_i(x_{1:t}) - f_j(x_{1:t})|$ as the score. While it offers theoretical guarantees, it can be uninformative in practice, since it ignores expert weights.

Instead, we draw inspiration from the query-by-committee framework (Seung, Opper, and Sompolinsky 1992), which uses ensemble disagreement as a score. For regression, this has been instantiated via prediction variance (Krogh and Vedelsby 1994). Since HEDGE maintains expert weights, we define score as the weighted variance:

$$\text{Score}(x_{1:t}) = \sum_i p_{i,t} \left(f_i(x_{1:t}) - y_t\right)^2, \tag{3}$$

where y_t is the weighted prediction (Equation 1). This emphasizes disagreement among well-performing experts while reducing the influence of low-weighted ones.

Budget	0	2				3				4				10			
Crop, Label	Base	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS
Maize, NAVAIL	3.87	2.79	2.51	2.59	2.39	2.42	2.43	2.29	2.43	2.58	2.53	2.27	2.28	2.28	2.33	2.23	2.23
Maize, GRLV	2.83	2.80	2.47	2.14	2.23	2.50	2.45	2.08	2.13	2.30	2.39	2.04	2.09	2.06	2.06	1.96	1.98
Millet, NAVAIL	5.25	2.89	2.79	2.87	2.68	2.74	2.72	2.67	2.75	2.80	2.75	2.69	2.67	2.71	2.72	2.68	2.68
Millet, GRLV	1.82	1.77	1.47	1.24	1.27	1.63	1.56	1.23	1.25	1.43	1.48	1.24	1.27	1.31	1.30	1.24	1.25
Sorghum, NAVAIL	2.95	2.37	2.21	1.45	1.53	1.75	2.13	1.48	1.46	1.73	1.87	1.34	1.46	1.33	1.50	1.25	1.21
Sorghum, GRLV	2.81	2.06	2.19	1.72	1.88	1.80	2.07	1.52	1.74	1.70	1.81	1.42	1.62	1.34	1.44	1.24	1.34
Wheat, NAVAIL	4.03	2.49	2.14	1.96	1.96	2.05	1.96	1.85	2.08	2.16	2.08	1.82	1.97	1.91	1.93	1.87	1.82
Wheat, GRLV	1.32	1.29	0.94	0.78	0.83	1.26	1.10	0.77	0.81	1.11	1.11	0.80	0.82	0.95	0.90	0.83	0.83

Table 1: Comparison of RMSE of label predictions for the different sampling methods. The most closely related prior work is SA (Fujii and Kashima 2016). Each method is evaluated on four crops and two labels. The column labels denote 1) the number of samples, 2) the sampling method.

6 Leveraging Episodic Priors

Here we describe our instantiations for ONLINEMAX in Algorithm 1.

In many BOAL applications, historical data provides sample episodes from the process generating the input stream $x_{1:T}$. For example, when streams represent weather data, we often have access to past weather episodes aligned to the same temporal structure. Such data enables computation of statistics about expert behavior (e.g., mean disagreement scores), which can guide query selection. However, as discussed in Section 4, prior BOAL work relies on the classic secretary algorithm for ONLINEMAX, which assumes no prior knowledge and cannot exploit such information.

To address this, below we propose two approaches that leverage historical sub-episodes aligned with the current ONLINEMAX call in Algorithm 1. Specifically, for a stream segment $(x_{t_0}, \ldots, x_{t_h})$, we sample K historical episodes in the same range:

$$E(t_0, t_h) = \left\{x^1_{t_0:t_h}, \ldots, x^K_{t_0:t_h}\right\}.$$

In the weather example, this corresponds to retrieving historical weather data for dates t_0 through t_h. Notably, our methods use only input features from this prior data and does not require labels.

6.1 Prophet-Secretary Algorithm (PSA)

The prophet inequality problem, like the secretary problem, addresses online selection but assumes greater knowledge of the input stream. The classic setting (Krengel and Sucheston 1977) assumes T known independent distributions, adversarially ordered, and admits a simple stopping rule achieving the optimal 0.5 approximation ratio relative to a prophet with full hindsight (Samuel-Cahn 1984). More recent extensions achieve stronger guarantees by leveraging additional stream knowledge.

We focus on the prophet-secretary variant (Esfandiari et al. 2017; Ehsani et al. 2018), where the T known distributions are observed in random order. Although it assumes independence and does not exploit correlation structure, it supports effective algorithms based on a single stream statistic. In contrast, approaches that require detailed knowledge (e.g., Markov models (Truong and Wang 2019)) are impractical in settings like weather prediction, where accurate long-range estimation is difficult.

The PSA algorithm (Ehsani et al. 2018) uses prior episodes $E(t_0, t_h)$ to estimate the expected maximum score:

$$\text{OPT} = \frac{1}{K}\sum_{k=1}^{K} \max_t\{\text{Score}(x^k_t)\}.$$

This estimate converges quickly via standard concentration bounds. Given OPT, PSA compares each score s_t in the stream $(s_{t_0}, \ldots, s_{t_h})$ to a decreasing threshold:

1. $\tau_t = \text{OPT} \cdot \left(1 - \exp\left(\frac{t - t_h}{t_h - t_0 + 1}\right)\right)$
2. For $t \in \{t_0, \ldots, t_h\}$: if $s_t > \tau_t$, select t and stop.

PSA achieves the best known $(1 - 1/e)$ approximation ratio for the prophet-secretary setting. Notably, each call to PSA produces a different threshold sequence, since OPT is computed from the samples $E(t_0, t_h)$ specific to that segment.

6.2 Empirical Threshold Selection (ETS)

We propose a simple alternative that uses historical data to select a single threshold τ as a stopping rule for ONLINEMAX. Given historical episodes $E(t_0, t_h)$ and a candidate threshold τ, we estimate its expected performance as:

1. Compute the expert-based score function along each historical episode in $E(t_0, t_h)$.
2. Let s^k_τ be the first score in $s^k_{t_0:t_h}$ exceeding τ.
3. Return $s^*_\tau = \frac{1}{K}\sum_k s^k_\tau$

ETS evaluates a discrete set of thresholds $\mathcal{T} = \{\tau_1, \ldots, \tau_M\}$ and selects the best-performing one: $\tau^* = \arg\max_{\tau \in \mathcal{T}} s^*_\tau$. It then applies this threshold to the current input stream, stopping at the first score that exceeds τ^*.

Like PSA, ETS can be applied dynamically at each ONLINEMAX call, adapting to changing expert weights and score distributions. Unlike PSA, it explicitly optimizes for empirical performance on historical data. This comes at the cost of higher computation, as multiple thresholds must be simulated over all prior episodes.

7 Experimental Results

We evaluate our method against baseline and existing approaches on several simulated agricultural tasks across different crops, as well as a real-world grape cold-hardiness prediction problem, described below. All source code, data, and scripts to replicate experiments are publicly available.

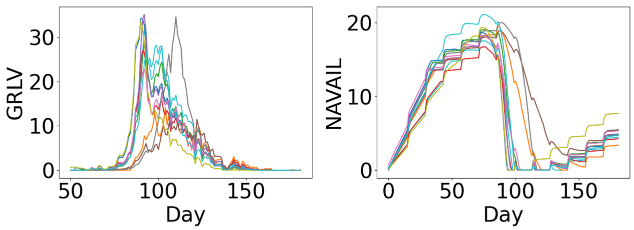

Figure 1: Leaf growth rate (GRLV) and available nitrogen (NAVAIL) from set of maize experts for a single season. Each line gives the predictions of a different expert, which are other maize crops with different simulator parameters.

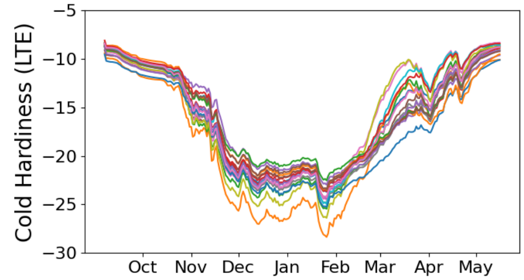

Figure 2: Cold-hardiness predictions from set of experts for a single season. Each line gives the predictions of a different expert, which are other grape cultivar models.

7.1 Agricultural Simulator Problems

Crop growth models (Boote, Jones, and Pickering 1996) are widely used to formalize and validate agronomic understanding of field processes. We use the WOFOSTGym simulator (Solow, Saisubramanian, and Fern 2025; van Diepen et al. 1989), a high-fidelity model for annual crops under nutrient and water fluxes. A well-calibrated WOFOSTGym model can serve as a digital twin, providing estimates of crop or soil variables that are costly or difficult to measure—making it a natural setting for BOAL-guided data collection to support site-specific agromanagement.

We use WOFOSTGym both to generate data for simulated BOAL tasks and to define a set of expert models. Specifically, we use four parameterizations (de Wit et al. 2019) representing wheat, maize, sorghum, and millet. Simulations are run using historical weather data (1984–2023) at 52°N, 5°W from the NASA POWER database, under an open-loop agromanagement policy alternating biweekly between irrigation and fertilization.

To introduce expert variability, we generate 15 WOFOSTGym models per crop by applying $\pm 10\%$ uniform noise to a subset of crop-specific parameters, reflecting realistic soil and crop differences (see Figure 1). For evaluation, we target two key internal state variables: (1) *nitrogen available in the subsoil (NAVAIL)* and (2) *growth rate of living leaves (GRLV)*. Both are informative for agromanagement yet costly to measure directly.

7.2 Cold-Hardiness Prediction

Cold temperatures can severely damage fruit crops, with risk depending on their cold hardiness—a property that varies throughout dormancy and across grape cultivars. Accurate estimates of cold hardiness are critical for managing costly frost-mitigation measures, but direct measurements are expensive, time-consuming, and sensitive to daily weather variation. While cultivar-specific prediction models exist (Ferguson et al. 2014; Saxena et al. 2023), they cover only a small subset of cultivars, leaving many without reliable estimates.

We use a dataset collected from 1988–2022 at the WSU Irrigated Agriculture Research and Extension Center (Prosser, WA), which includes cold-hardiness measurements for 18 grape cultivars. Bud samples were collected biweekly during dormancy to determine ground-truth cold hardiness, with each cultivar contributing between 4 and 30 seasons of data. Daily weather data for each season is obtained via the AgWeatherNet API (WSU 2023). The dataset includes dormancy-season records from September 7 to May 15, excluding seasons with over 10% missing weather data.

Since ground-truth labels are only available biweekly, we train a neural network to interpolate daily labels for each target cultivar, enabling BOAL queries on any day. These interpolated labels are used only for simulating the query process; evaluation is performed on real measurements. To ensure label quality, we restrict experiments to the 8 cultivars with the most accurate prediction models. For each BOAL task, we define a set of 17 experts based on a multi-task cold-hardiness model (Saxena et al. 2023) trained on all cultivars except the target. Implementation details are in Appendix A.

7.3 Experimental Protocol

We evaluate several BOAL methods on both simulated agricultural tasks and the real-world grape cold-hardiness problem. For each crop–target pair, we conduct multiple BOAL experiments, each corresponding to a single growing season (WOFOSTGym) or dormancy period (cold hardiness). In WOFOSTGym, one of 15 models is selected as the target and the remaining 14 serve as experts; results are averaged over all 15 targets. For grape cold hardiness, we evaluate on 8 cultivars, using models trained on the remaining 17 as experts. This setup reflects realistic scenarios where the target is not perfectly modeled by any expert, but at least one is expected to offer useful guidance.

We test query budgets of 2–5 (cold hardiness) and 2–10 (WOFOSTGym), reflecting realistic data collection constraints. For each target and budget, we conduct 37 BOAL runs using different historical years of weather data. Performance is measured using root mean square error (RMSE) with respect to the target function, averaged across the 37 runs. For methods using prior data, the remaining 36 seasons serve as the source of historical episodes, ensuring that test and training years do not overlap.

All experiments use the HEDGE algorithm with learning rate $\eta = 1$. We evaluate five methods: 1) **Base** – Uniformly weighted experts with no queries. 2) **UNI** – Queries issued uniformly in time. 3) **SA** – The Secretary Algorithm-

Budget	0	2				3				4				5			
Cultivar	Base	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS
Chardonnay	1.35	1.40	1.35	1.37	1.37	1.39	1.36	1.33	1.35	1.38	1.36	1.33	1.34	1.35	1.35	1.33	1.33
Grenache	1.57	1.48	1.48	1.43	1.45	1.47	1.46	1.43	1.43	1.45	1.48	1.42	1.42	1.43	1.48	1.41	1.44
Merlot	1.52	1.63	1.55	1.52	1.53	1.55	1.53	1.50	1.52	1.54	1.52	1.51	1.50	1.52	1.52	1.50	1.51
Mourvedre	2.00	2.13	1.77	1.79	1.76	1.79	1.76	1.72	1.73	1.78	1.74	1.73	1.74	1.76	1.74	1.74	1.75
Pinot Gris	1.50	1.67	1.67	1.64	1.67	1.71	1.69	1.65	1.68	1.69	1.70	1.64	1.63	1.64	1.69	1.62	1.65
Sangiovese	1.88	1.81	1.50	1.51	1.48	1.65	1.49	1.46	1.45	1.60	1.52	1.48	1.45	1.56	1.50	1.51	1.46
Syrah	1.31	1.47	1.38	1.37	1.37	1.38	1.37	1.32	1.36	1.37	1.35	1.33	1.34	1.35	1.34	1.32	1.35
Viognier	1.51	1.50	1.46	1.51	1.43	1.55	1.45	1.49	1.43	1.50	1.50	1.46	1.41	1.47	1.49	1.44	1.42
Mean	1.58	1.64	1.52	1.52	1.51	1.56	1.51	1.49	1.49	1.54	1.52	1.49	1.48	1.51	1.51	1.49	1.49

Table 2: Comparison of RMSE of cold-hardiness predictions for the different sampling methods. The column labels denote 1) the number of samples, 2) the sampling method.

Budget	0	2				3				4				5			
Cultivar	Base	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS	UNI	SA	PSA	ETS
Chardonnay	1.75	1.74	1.67	1.71	1.69	1.67	1.66	1.62	1.65	1.65	1.63	1.63	1.62	1.63	1.63	1.65	1.63
Grenache	1.77	1.68	1.74	1.77	1.69	1.71	1.70	1.69	1.64	1.65	1.72	1.62	1.63	1.65	1.69	1.67	1.61
Merlot	1.74	1.86	1.75	1.74	1.72	1.79	1.72	1.71	1.71	1.73	1.75	1.7	1.71	1.72	1.73	1.69	1.71
Mourvedre	1.87	2.01	1.89	1.78	1.77	1.82	1.77	1.79	1.75	1.85	1.86	1.74	1.75	1.81	1.81	1.76	1.74
Pinot Gris	1.51	1.61	1.52	1.55	1.53	1.66	1.58	1.54	1.52	1.60	1.57	1.50	1.49	1.52	1.61	1.52	1.50
Sangiovese	1.74	1.75	1.39	1.42	1.44	1.57	1.42	1.38	1.41	1.50	1.42	1.41	1.43	1.48	1.40	1.41	1.40
Syrah	1.29	1.47	1.46	1.45	1.44	1.43	1.45	1.42	1.42	1.43	1.44	1.41	1.42	1.42	1.43	1.42	1.42
Viognier	1.38	1.34	1.40	1.41	1.35	1.38	1.37	1.36	1.31	1.33	1.35	1.35	1.31	1.31	1.34	1.31	1.30
Mean	1.63	1.68	1.60	1.60	1.58	1.63	1.59	1.56	1.55	1.59	1.59	1.54	1.54	1.57	1.58	1.55	1.54

Table 3: Comparison of RMSE of cold-hardiness predictions for the different sampling methods with a worse set of experts. The column labels denote 1) the number of samples, 2) the sampling method.

based method from (Fujii and Kashima 2016), without prior episodic data. 4) **PSA** – Algorithm 1 instantiated with the prophet-secretary algorithm. 5) **ETS** – Algorithm 1 instantiated with our empirical threshold selection method.

7.4 Agricultural Simulator Results

Overall Benefit of BOAL. Table 1 reports results for each WOFOSTGym crop–target pair across methods and budgets. Across all budget levels, active learning methods consistently outperform UNI, even with very few queries—demonstrating the value of selective querying. Notably, our ETS method frequently achieves the lowest losses among methods. Below, we assess statistical significance using the *paired Wilcoxon Signed-Rank Test* (Wilcoxon 1945).

Impact of Queries. We first compare all query methods to Base, which issues no queries. At a 0.05 significance level, all query-based methods significantly outperform Base across all tasks and budgets. This highlights that the expert weighting mechanism can effectively leverage even a small number of labels. The improvement is especially notable given that the unqueried expert ensemble (Base) already performs well.

Impact of Selective Queries. We next compare informed query methods (SA, PSA, ETS) to uniform querying (UNI). At a 0.05 significance level, PSA and ETS significantly outperform UNI in all cases. SA performs better than UNI at 0.05 significance in over half of the settings, but is worse or indistinguishable in the rest. These results show that conditioning query decisions on observed history improves efficiency, though SA's limitations suggest that informed selection alone may be insufficient without prior knowledge.

Impact of Episodic Prior Knowledge. Finally, we compare PSA and ETS—our methods that incorporate prior episodic data—to SA, which does not. PSA and ETS outperforms SA in all but one and three cases respectively at 0.05 significance. These results demonstrate the added value of episodic prior knowledge in improving query efficiency and final predictive performance.

7.5 Real-World Cold-Hardiness Results

We repeat the experiments on real-world grape cold-hardiness. Table 2 reports results for each target cultivar. As with WOFOSTGym, we observe clear benefits from informed querying and leveraging episodic prior data. According to the Wilcoxon signed-rank test, all methods significantly outperform Base in all but one case (UNI with 2 queries). Similarly, all informed methods beat UNI in all but one case (SA with 5 queries), and all episodic prior methods beat SA in all but one case (PSA with 2 queries). However, these gains are smaller and diminish more rapidly with increasing budget compared to WOFOSTGym. This is likely due to noise in the real-world test labels, which limits achievable RMSE improvement.

Lower Quality Experts. We also evaluate performance using a weaker set of experts. As described in Appendix A, we use our multi-task model to sample 17 virtual cul-

Budget	2				3				4				10			
Crop, Label	SA	PSA	ETS	Max	SA	PSA	ETS	Max	SA	PSA	ETS	Max	SA	PSA	ETS	Max
Maize, NAVAIL	20.1	16.9	26.9	28.7	16.0	16.9	15.8	19.8	12.2	12.7	14.2	15.3	6.4	6.6	6.8	7.0
Maize, GRLV	9.3	28.0	34.4	47.0	5.9	17.0	26.0	33.1	7.4	13.6	17.6	22.0	6.2	6.9	7.5	9.6
Millet, NAVAIL	20.5	14.1	21.5	21.8	15.4	13.9	17.5	18.0	13.4	10.3	12.3	12.5	6.6	6.4	6.6	6.6
Millet, GRLV	4.3	8.4	11.8	14.4	1.7	5.9	7.6	9.9	2.4	4.4	5.2	7.0	2.3	2.5	2.5	3.0
Sorghum, NAVAIL	7.8	11.7	15.1	18.8	6.3	9.3	10.1	11.1	5.7	7.6	7.6	9.4	3.4	3.8	4.0	4.2
Sorghum, GRLV	12.8	27.9	30.5	30.5	7.0	18.4	21.5	20.5	8.1	14.1	15.5	17.9	5.2	5.8	7.0	8.8
Wheat, NAVAIL	19.9	13.3	26.0	27.0	15.9	14.0	15.1	17.1	13.1	10.1	14.4	15.0	6.7	6.3	6.9	7.0
Wheat, GRLV	2.9	6.9	9.2	13.2	0.6	4.6	6.7	8.6	0.6	3.4	4.6	6.0	1.7	1.9	2.1	2.5

Table 4: Comparison of average SCORE value for different sampling methods. Each method is evaluated on four crops and two labels. The column labels denote 1) the number of samples collected, 2) the sampling method used, either SA, PSA, ETS, or Max Oracle, where Max Oracle has the full season information and will always pick the highest SCORE value.

Budget	2				3				4				5			
Cultivar	SA	PSA	ETS	Max	SA	PSA	ETS	Max	SA	PSA	ETS	Max	SA	PSA	ETS	Max
Chardonnay	1.21	1.53	2.15	2.83	1.42	1.45	1.71	2.22	1.06	1.55	1.81	1.87	1.08	1.57	1.64	1.64
Grenache	1.03	1.61	2.43	3.25	1.39	1.46	1.81	2.24	0.78	1.45	1.77	1.77	0.74	1.38	1.67	1.66
Merlot	1.12	1.42	2.38	3.17	1.18	1.38	1.76	2.17	0.73	1.38	1.72	1.72	0.74	1.31	1.53	1.56
Mourvedre	1.06	1.51	2.36	3.21	1.18	1.34	1.75	2.22	0.61	1.31	1.72	1.77	0.63	1.29	1.60	1.59
Pinot Gris	1.39	2.08	2.41	3.00	1.79	1.81	1.95	2.57	1.32	1.93	2.19	2.28	1.31	1.94	2.03	2.09
Sangiovese	1.11	1.60	2.47	3.47	1.35	1.45	1.79	2.33	0.72	1.40	1.74	1.77	0.70	1.51	1.68	1.63
Syrah	1.06	1.50	2.41	3.22	1.24	1.36	1.82	2.26	0.62	1.38	1.74	1.79	0.61	1.35	1.65	1.62
Viognier	1.15	1.71	2.72	3.23	1.30	1.54	2.04	2.57	0.73	1.58	2.03	2.06	0.81	1.47	1.86	1.83
Mean	1.14	1.62	2.42	3.17	1.36	1.47	1.83	2.32	0.82	1.50	1.84	1.88	0.83	1.48	1.71	1.70

Table 5: Comparison of average SCORE value for different sampling methods for cold-hardiness prediction. The column labels denote 1) the number of samples collected, 2) the sampling method used, either SA (Fujii and Kashima 2016), PSA, ETS, or Max Oracle, where Max Oracle has the full season information and will always pick the sample with the highest SCORE value.

tivars from the learned embedding space. These virtual experts generally perform worse than the original set.

Table 3 shows results for this setting. **Base** performance confirms that the new experts are weaker, though still viable. While all methods perform worse overall, relative performance is consistent. In all but one case (SA vs. UNI with 4 queries), the same significant differences are observed, indicating robustness to changes in expert quality.

7.6 Investigating OnlineMax

The above results show that our methods leveraging prior episodic data outperform those that do not. Here, we investigate whether this improvement stems from better ONLINEMAX performance or other factors. Recall that the goal of ONLINEMAX is to select a query with a high active learning score (Equation 3) over a given interval. Table 4 (WOFOST-Gym) and Table 5 (cold hardiness) report the average selected score across all ONLINEMAX calls for each method: SA, PSA, ETS, and a Max Oracle that selects the highest score in hindsight as an upper bound.

We observe that PSA and ETS consistently achieve higher average scores than SA, with ETS often outperforming PSA—especially at smaller budgets. These differences align closely with the BOAL performance results, suggesting that the improved outcomes stem directly from more effective query selection. Notably, ETS often approaches the Max Oracle, indicating near-optimal performance in selecting high-score samples. We also find that absolute score values decrease with larger budgets, which reflects reduced expert disagreement as weights are updated over time.

To further understand SA's weaker performance, we examined specific failure cases and identified two recurring patterns. First, if the maximum score occurs in SA's observational window, it is never selected—since SA waits until the end of the interval. Second, if scores remain low during observation and rise quickly afterward, SA selects the first modestly higher score, which may still be far from optimal. In both cases, the lack of stream statistics prevents SA from adapting. In contrast, PSA and ETS avoid these pitfalls by incorporating prior episodic information.

8 Summary

We studied active learning under online, finite-horizon constraints with very small query budgets. Our novel approach integrates the learning-from-experts paradigm with prior unlabeled episodic data. Across both realistic simulations and real-world agricultural datasets, our method significantly outperformed baseline expert predictions, uniform querying, and a prior BOAL method that does not leverage episodic data. During 2024-2025 dormancy we used our BOAL approach to collect four samples from four vineyards with the goal of providing those vineyards with customized cold-hardiness models. Multiple years will be required to assess the potential benefits of this approach over the baseline.

Acknowledgements

This research was supported by USDA-NIFA award No. award No.2021-67021-35344 under the National AI Institute: Agricultural AI for Transforming Workforce and Decision Support (AgAID). The authors thank Lynn Mills and Alan Kawakami for the collection of cold-hardiness data.

References

Atlas, L.; Cohn, D.; and Ladner, R. 1990. Training connectionist networks with queries and selective sampling. *Advances in neural information processing systems*, 2.

Bateni, M.; Hajiaghayi, M.; and Zadimoghaddam, M. 2013. Submodular secretary problem and extensions. *ACM Transactions on Algorithms (TALG)*, 9(4): 1–23.

Boote, K. J.; Jones, J. W.; and Pickering, N. B. 1996. Potential Uses and Limitations of Crop Models | Agronomy Journal.

Castro, R.; Hellström, F.; and van Erven, T. 2024. Adaptive selective sampling for online prediction with experts. *Advances in Neural Information Processing Systems*, 36.

Cesa-Bianchi, N.; Freund, Y.; Haussler, D.; Helmbold, D. P.; Schapire, R. E.; and Warmuth, M. K. 1997. How to use expert advice. *Journal of the ACM (JACM)*, 44(3): 427–485.

Cesa-Bianchi, N.; Lugosi, G.; and Stoltz, G. 2005. Minimizing regret with label efficient prediction. *IEEE Transactions on Information Theory*, 51(6): 2152–2162.

Das, J.; Py, F.; Harvey, J. B.; Ryan, J. P.; Gellene, A.; Graham, R.; Caron, D. A.; Rajan, K.; and Sukhatme, G. S. 2015. Data-driven robotic sampling for marine ecosystem monitoring. *The International Journal of Robotics Research*, 34(12): 1435–1452.

de Wit, A.; Boogaard, H.; Fumagalli, D.; Janssen, S.; Knapen, R.; van Kraalingen, D.; Supit, I.; van der Wijngaart, R.; and van Diepen, K. 2019. 25 years of the WOFOST cropping systems model. *Agricultural Systems*, 168: 154–167.

Dynkin, E. B. 1963. The optimum choice of the instant for stopping a Markov process. *Soviet Mathematics*, 4: 627–629.

Ehsani, S.; Hajiaghayi, M.; Kesselheim, T.; and Singla, S. 2018. Prophet secretary for combinatorial auctions and matroids. In *Proceedings of the twenty-ninth annual acm-siam symposium on discrete algorithms*, 700–714. SIAM.

Esfandiari, H.; Hajiaghayi, M.; Liaghat, V.; and Monemizadeh, M. 2017. Prophet secretary. *SIAM Journal on Discrete Mathematics*, 31(3): 1685–1701.

Ferguson, J. C.; Moyer, M. M.; Mills, L. J.; Hoogenboom, G.; and Keller, M. 2014. Modeling Dormant Bud Cold Hardiness and Budbreak in Twenty-Three Vitis Genotypes Reveals Variation by Region of Origin. *American Journal of Enology and Viticulture*, 65(1): 59–71.

Freund, Y.; and Schapire, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1): 119–139.

Freund, Y.; Seung, H. S.; Shamir, E.; and Tishby, N. 1997. Selective sampling using the query by committee algorithm. *Machine learning*, 28: 133–168.

Fujii, K.; and Kashima, H. 2016. Budgeted stream-based active learning via adaptive submodular maximization. *Advances in Neural Information Processing Systems*, 29.

Hao, S.; Hu, P.; Zhao, P.; Hoi, S. C.; and Miao, C. 2018. Online active learning with expert advice. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(5): 1–22.

Hara, S.; Matsuura, M.; Honda, J.; and Ito, S. 2024. Active model selection: A variance minimization approach. *Machine Learning*, 1–19.

Herbster, M.; and Warmuth, M. K. 1998. Tracking the best expert. *Machine learning*, 32(2): 151–178.

Karimi, M. R.; Gürel, N. M.; Karlaš, B.; Rausch, J.; Zhang, C.; and Krause, A. 2021. Online active model selection for pre-trained classifiers. In *International Conference on Artificial Intelligence and Statistics*, 307–315. PMLR.

Krengel, U.; and Sucheston, L. 1977. Semiamarts and finite values. *Bull. Amer. Math. Soc.*, 83(4): 745–747.

Krogh, A.; and Vedelsby, J. 1994. Neural network ensembles, cross validation, and active learning. *Advances in neural information processing systems*, 7.

Kumar, B.; Abernethy, J. D.; and Saligrama, V. 2022. Activehedge: Hedge meets active learning. In *International Conference on Machine Learning*, 11694–11709. PMLR.

Littlestone, N.; and Warmuth, M. K. 1994. The weighted majority algorithm. *Information and computation*, 108(2): 212–261.

Liu, X.; Xia, F.; Stevens, R. L.; and Chen, Y. 2022. Contextual active online model selection with expert advice. In *Proceedings of the ICML Workshop on Adaptive Experimental Design and Active Learning in the Real World*.

Luo, W.; Nam, C.; and Sycara, K. 2017. Online decision making for stream-based robotic sampling via submodular optimization. In *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 118–123. IEEE.

Mitra, S.; and Gopalan, A. 2020. On adaptivity in information-constrained online learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 5199–5206.

Samuel-Cahn, E. 1984. Comparison of threshold stop rules and maximum for independent nonnegative random variables. *the Annals of Probability*, 1213–1216.

Saxena, A.; Pesantez-Cabrera, P.; Ballapragada, R.; Lam, K.-H.; Keller, M.; and Fern, A. 2023. Grape cold hardiness prediction via multi-task learning. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23, 15717–15723. Washington, DC, USA: AAAI Press. ISBN 978-1-57735-880-0.

Settles, B. 2009. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Seung, H. S.; Opper, M.; and Sompolinsky, H. 1992. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, 287–294.

Solow, W.; Saisubramanian, S.; and Fern, A. 2025. WOFOSTGym: A Crop Simulator for Learning Annual and Perennial Crop Management Strategies. In *Reinforcement Learning Conference (RLC) 2025*.

Truong, A.; Etesami, S. R.; and Kiyavash, N. 2021. Selective Labeling in Learning with Expert Advice. In *2021 American Control Conference (ACC)*, 2537–2542. IEEE.

Truong, V.-A.; and Wang, X. 2019. Prophet Inequality with Correlated Arrival Probabilities, with Application to Two Sided Matchings. *arXiv: Optimization and Control*.

van Diepen, C.; Wolf, J.; van Keulen, H.; and Rappoldt, C. 1989. WOFOST: a simulation model of crop production. *Soil Use and Management*, 5(1): 16–24.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. Biom. Bull., 1, 80–83.

WSU. 2023. AgWeatherNet. https://weather.wsu.edu.

Zhang, Y.; Zhao, P.; Cao, J.; Ma, W.; Huang, J.; Wu, Q.; and Tan, M. 2018. Online adaptive asymmetric active learning for budgeted imbalanced data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2768–2777.

Zhao, P.; Hoi, S. C. H.; and Zhuang, J. 2013. Active Learning with Expert Advice. In Nicholson, A. E.; and Smyth, P., eds., *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence, UAI 2013, Bellevue, WA, USA, August 11-15, 2013*. AUAI Press.