

Drift-Aware Preference-Based Bayesian Optimization with Dynamic Multinomial-Logit Surrogates

A Confirmatory Semi-Synthetic Validation on NASA N-CMAPSS Maintenance Decisions

Muhammad Amir Saeed^{*1}, Sadia Saba^{†1}, and Antonio Candelieri^{‡1}

¹Department of Economics, Quantitative Methods and Business Strategies, University of Milano-Bicocca, Milan, Italy

Abstract

Preference-based Bayesian optimization (PBO) enables sequential optimization when candidate solutions can be compared but cannot be assigned reliable scalar values. Most PBO methods, however, assume stable decision-maker preferences. We introduce DA-MNL-PBO, a drift-aware framework in which interpretable multinomial-logit/Bradley-Terry preference coefficients evolve through a two-regime state-space model. A sequential Laplace approximation within an interacting multiple-model filter maintains stable and adaptive preference states, while a regime-aware acquisition rule combines optimization value with information about the preference coefficients and latent adaptation regime.

We evaluate DA-MNL-PBO in a semi-synthetic predictive-maintenance study using NASA N-CMAPSS run-to-failure trajectories. Real engine trajectories and remaining-useful-life information define maintenance-policy outcomes, while controlled time-varying stakeholder utilities generate pairwise preferences. Hyperparameters were frozen before confirmatory evaluation. Across two confirmatory subsets, four temporal conditions, three response-consistency levels, 200 paired replicates, and four methods, the study comprised 19,200 method-level runs. Under preference drift, DA-MNL-PBO achieved lower mean cumulative dynamic regret than static and sliding-window MNL-PBO in all 12 confirmatory drift conditions, reducing regret relative to static MNL-PBO by 6.9–71.1% (42.1% on average). Fixed exponential discounting performed better under all abrupt shifts, whereas DA-MNL-PBO performed better in five of six gradual-drift conditions. The method also produced meaningful shifts in maintenance recommendations after preference changes. These results support probabilistic drift-aware preference modeling particularly under gradual change, while demonstrating that fixed forgetting remains competitive under abrupt shifts. Because preferences are simulated, the results establish algorithmic adaptation rather than human-preference validity.

Keywords: preference-based Bayesian optimization; multinomial logit; Bradley–Terry model; non-stationarity; preference drift; interacting multiple model; dynamic regret; predictive maintenance; N-CMAPSS; human-in-the-loop optimization.

1 Introduction

Bayesian optimization (BO) is a sequential framework for optimizing expensive, noisy, or otherwise difficult-to-evaluate systems using a probabilistic surrogate and an acquisition rule that decides where the next evaluation should be spent (Shahriari et al., 2016). Classical BO

^{*}Corresponding author. Email: m.saeed@campus.unimib.it; ORCID: 0000-0003-1650-8194

[†]Email: s.saba5@campus.unimib.it; ORCID: 0009-0006-3166-8953

[‡]Email: antonio.candelieri@unimib.it; ORCID: 0000-0003-1431-576X

assumes that an objective value can be observed. In many engineering and human-in-the-loop settings, however, the most informative feedback is comparative: an operator can judge that one design, controller response, maintenance policy, or prompt configuration is preferable to another without being able to assign either option a stable numerical score. Preferential Bayesian optimization (PBO) formalizes this setting by learning an unobserved utility from pairwise or set-wise preference feedback and using that model to guide sequential search (González et al., 2017; Mikkola et al., 2020; Astudillo et al., 2023; Takeno et al., 2023; Xu et al., 2024; Kayal et al., 2025).

A central assumption in many PBO formulations is temporal stability: comparisons observed early and late in an optimization campaign are treated as evidence about the same latent utility. This assumption becomes fragile in prolonged decision processes. A maintenance planner may initially favor utilization and preservation of useful life, then become more risk averse when mission requirements change. A user tuning a robotic or control system may revise the balance between performance, comfort, and safety. In interactive AI, preferences can evolve with experience, context, or exposure to candidate outputs.

Recent work has increasingly recognized non-stationarity in preference-based and human-in-the-loop optimization. Dynamic Bradley–Terry models allow latent strengths or preference parameters to evolve over time (Bong et al., 2020); non-stationary dueling-bandit methods target dynamic regret under changing preferred alternative (Kleine Buening and Saha, 2023); and temporally weighted preference objectives have been proposed for preference drift in alignment settings (Son et al., 2025). Dynamic BO has likewise been studied in time-varying environments and human-in-the-loop control (Bogunovic et al., 2016; Kim and Sergi, 2026). Recent preference-aware Bayesian optimization has also addressed changing decision-maker preferences through online pairwise preference inference and sliding-window memory mechanisms (Ma et al., 2026).

The contribution pursued here is therefore narrower than introducing non-stationary preference optimization itself. We study a structured PBO setting in which interpretable MNL/Bradley–Terry preference coefficients are represented explicitly as dynamic latent states, uncertainty over stable versus adaptive evolution is maintained probabilistically, and information about that latent regime enters active comparison selection. This combination provides an alternative to committing in advance to a single fixed memory horizon or forgetting rate.

This paper develops DA-MNL-PBO—Drift-Aware Preference-Based Bayesian Optimization with Dynamic Multinomial-Logit Surrogates. The method extends an interpretable discrete-choice view of BO. Earlier work introduced multinomial-logit surrogates for categorical and mixed-variable optimization (Saeed and Candelieri, 2026), and subsequent work used MNL-based preference structures in multi-objective CO₂ storage design and cost-aware structured prompt optimization (Saeed, Eidsvik, et al., 2026; Saeed, Saba, et al., 2026). The present work changes the temporal assumption: the preference coefficients are states, not constants. A two-regime process distinguishes ordinary local evolution from a high-variance adaptive mode; an interacting multiple-model (IMM) filter (Blom and Bar-Shalom, 1988) and sequential Laplace updates maintain approximate Bayesian uncertainty; and the acquisition rule combines optimization value with information about both coefficients and regime state.

The empirical study is designed to separate methodological evidence from application realism. NASA N-CMAPSS (Arias Chao et al., 2021) provides high-frequency aircraft-engine run-to-failure trajectories under realistic flight conditions, but it is not a human-preference dataset. We therefore use it as a physical decision environment: a fixed remaining-useful-life (RUL) predictor converts engine histories into maintenance-policy outcomes, while known time-varying stakeholder utilities generate controlled pairwise choices. This semi-synthetic construction makes dynamic regret, adaptation delay, drift-signal specificity, and operational policy shifts observable by design; coefficient-recovery diagnostics are additionally available under the learner’s chosen observation-scale convention.

A first NASA development run exposed two experiment-design weaknesses: many nominal

maintenance policies were outcome-equivalent, and the initial economic and safety utility profiles shared the same optimum. Rather than tune the proposed algorithm on those outcomes, we corrected the decision experiment, retained the diagnostic record, and froze the final stakeholder scenario and a two-way cross-phase contrast criterion before confirmatory PBO evaluation on additional N-CMAPSS subsets. This design produces a useful distinction between *development evidence* (DS02-006) and *confirmatory evidence* (eligible additional subsets).

The contributions of this paper are fivefold:

1. We formulate PBO with a time-indexed, interpretable linear MNL/Bradley–Terry utility and a two-regime Markov state process, enabling direct interpretation of changing feature weights.
2. We derive a computationally lightweight sequential approximation that combines IMM mixing, logistic Laplace updates, and a continuous posterior adaptive-mode probability.
3. We introduce a composite acquisition rule that combines expected best-option utility with information about the current preference vector and the latent evolution regime, and we separately evaluate the incremental contribution of these information terms through post-confirmatory ablation.
4. We develop a pre-specified confirmatory semi-synthetic validation protocol on N-CMAPSS that freezes algorithmic parameters before confirmatory evaluation, removes outcome-equivalent policies, includes stationary and noise-only controls, and applies a pre-PBO eligibility criterion to candidate subsets.
5. We characterize the regimes in which additional dynamic structure is empirically useful: DA-MNL-PBO reduces mean dynamic regret relative to static and sliding-window MNL-PBO across the confirmatory drift conditions and shows its clearest comparative advantage under gradual preference evolution, whereas fixed exponential discounting remains a strong and often superior competitor after abrupt changes.

The final point is important. The purpose of the study is not to tune the proposed method until it wins every scenario. The confirmatory design is intended to reveal where adaptive regime modeling adds value, where a simpler forgetting mechanism suffices, and which inferential quantities are reliable enough to support substantive claims.

1.1 Research Questions and Confirmatory Hypotheses

The empirical study is organized around four pre-specified questions concerning the value and limitations of drift-aware preference modeling. Because the N-CMAPSS study is semi-synthetic, these hypotheses concern algorithmic behavior under controlled preference evolution rather than claims about observed human behavior.

H1 (stationary robustness). When stakeholder preferences remain stationary, maintaining a dynamic preference model should impose only a limited efficiency cost relative to a static full-history MNL-PBO model.

H2 (dynamic decision quality). When the latent preference vector changes over time, DA-MNL-PBO should reduce cumulative dynamic regret relative to a static full-history MNL-PBO model. Comparisons with sliding-window and exponentially discounted MNL-PBO are used to determine whether probabilistic regime adaptation provides additional benefit relative to simpler forgetting mechanisms and under which temporal regimes such benefit appears.

H3 (adaptive-mode responsiveness). The posterior probability assigned to the high-variance adaptive regime should increase more strongly following genuine preference drift than under stationary preferences. Because changes in response noise may also affect the posterior,

separation from a noise-only perturbation is treated as a specificity test rather than assumed a priori.

H4 (operational adaptation). Following a shift from the economic/operations preference profile toward the safety-oriented profile, the recommendations produced by DA-MNL-PBO should move in a safety-consistent operational direction, expressed through an earlier effective maintenance trigger, more frequent inspection, or both, depending on the policy geometry of the N-CMAPSS subset.

These hypotheses are evaluated using the primary and secondary endpoints defined in Section 5.5. The primary confirmatory endpoint is cumulative dynamic regret; the remaining quantities are used to characterize robustness, interpretability, and adaptation behavior.

2 Related Work and Research Gap

2.1 Preferential Bayesian optimization

Early PBO work formulated latent-function optimization through pairwise “duels” using Gaussian-process preference models (González et al., 2017). Projective preferential queries extended the interaction model to higher-dimensional settings (Mikkola et al., 2020). More recent methods have focused on acquisition design, inference quality, regret, and scalability. qEUBO gives expected utility of the best option a decision-theoretic interpretation and strong one-step properties (Astudillo et al., 2023); skew-GP work studies the accuracy and computation of preferential posterior inference (Takeno et al., 2023); principled PBO provides cumulative-regret guarantees using confidence sets (Xu et al., 2024); and recent BO-from-human-feedback analysis gives near-optimal regret bounds under Bradley–Terry–Luce feedback (Kayal et al., 2025). Local PBO methods extend trust-region and derivative-informed local search to preferential feedback in higher-dimensional problems (Menn et al., 2026).

Applications are also moving closer to real human-in-the-loop systems. Recent multi-task PBO work transfers preference information across controller-tuning tasks and users (Coutinho et al., 2026), while Bayesian preference inference has been combined with reinforcement learning from human feedback to reduce costly comparisons (Cercola et al., 2026). These developments strengthen the practical motivation for PBO, but the latent preference model is usually treated as stable during a single optimization campaign.

2.2 Time-varying optimization and dynamic preferences

Non-stationarity is more mature in standard bandit and BO settings. Time-varying GP bandit optimization explicitly discounts stale observations and analyzes regret under evolving reward functions (Bogunovic et al., 2016). In human-in-the-loop control, dynamic BO has recently been experimentally validated against standard BO when the input–output relation changes over time (Kim and Sergi, 2026). These studies establish a general principle: historical data are valuable only to the extent that the present objective is still compatible with the past.

Pairwise comparison models have an independent dynamic literature. Dynamic Bradley–Terry estimation allows latent strengths or rankings to change over time (Bong et al., 2020). Non-stationary dueling bandits target dynamic regret and adapt to changes in the preferred arm without assuming that the number of switches is known (Kleine Buening and Saha, 2023). In preference optimization for large language models, NS-DPO uses a dynamic Bradley–Terry view and temporal discounting to address preference drift (Son et al., 2025). These methods are highly relevant, but they differ from the structured PBO setting considered here: the present optimizer actively chooses informative comparisons from alternatives described by interpretable engineering features and maintains uncertainty for sequential Bayesian decision making.

A particularly relevant distinction is between *forgetting-based* and *state-based* adaptation. Sliding windows and exponential discounting modify the influence assigned to historical observa-

Table 1: Conceptual positioning of DA-MNL-PBO relative to neighboring approaches. The table distinguishes methodological capabilities rather than claiming overall superiority.

Approach	Time-varying preferences	Interpretable coefficients	Probabilistic regime	Active pair selection	Regime-aware acquisition
Static PBO	No	Model dep.	No	Yes	No
Sliding-window PBO	Implicit	Model dep.	No	Yes	No
Discounted PBO	Implicit	Model dep.	No	Yes	No
Dynamic Bradley–Terry	Yes	Yes	Model dep.	No/varies	No
Non-stationary dueling bandits	Yes	No/varies	No/varies	Yes	No
PABO (Ma et al., 2026)	Yes	Model dep.	No/varies	Yes	No/varies
DA-MNL-PBO	Yes	Yes	Yes	Yes	Yes

tions according to a fixed memory rule. Dynamic latent-state models instead place temporal evolution directly in the preference parameters and infer the current state from sequential evidence. The present work follows the latter strategy while retaining sliding-window and fixed-discount methods as explicit empirical baselines. Importantly, we do not assume that state-based adaptation must dominate simpler forgetting: one purpose of the evaluation is to identify the temporal regimes in which its additional structure is empirically justified.

2.3 Interpretable MNL surrogates and structured decision spaces

The MNL/Bradley–Terry representation is attractive when alternatives have meaningful attributes. In prior MNL-BO work by the authors, a multinomial-logit surrogate was used to represent categorical and mixed-variable designs while preserving directly interpretable utility coefficients (Saeed and Candelieri, 2026). Multi-objective CO₂ injection work further emphasized mixed engineering decisions and trade-off analysis (Saeed, Eidsvik, et al., 2026), while MNL-PCO extended the preference-based formulation to structured LLM prompt components under explicit evaluation cost (Saeed, Saba, et al., 2026). Those studies motivate a common question: if interpretable coefficients summarize what the optimizer currently values, how should those coefficients be updated when the decision criterion itself evolves?

2.4 Research gap

Existing approaches address important parts of non-stationary preference optimization, including temporal weighting or forgetting, dynamic pairwise-choice models, non-stationary dueling bandits, and time-varying Bayesian optimization. Our focus is a more specific methodological intersection. We seek a structured PBO framework that simultaneously provides: (i) interpretable time-varying MNL/Bradley–Terry coefficients over alternative attributes; (ii) active selection of pairwise comparisons; (iii) probabilistic uncertainty over stable versus accelerated coefficient evolution; and (iv) an acquisition mechanism that can value information about this latent evolution regime in addition to immediate optimization value.

DA-MNL-PBO is designed for this intersection. Its novelty is therefore not the general observation that preferences may change over time, nor the general use of forgetting to handle non-stationarity. Rather, it lies in integrating dynamic interpretable preference coefficients, regime-switching Bayesian state inference, and regime-aware active preference acquisition within a single PBO loop. The empirical study additionally evaluates when this additional structure is useful relative to simpler static, sliding-window, and exponentially discounted alternatives.

3 Problem Formulation

Let $\mathcal{X} = \{x_1, \dots, x_M\}$ denote a finite or discretized candidate set and let each candidate have an interpretable feature vector $\phi(x) \in \mathbb{R}^p$. At preference-query iteration t , the latent utility is

$$u_t(x) = \beta_t^\top \phi(x), \quad (1)$$

where the preference coefficient $\beta_t \in \mathbb{R}^p$ can change over time. For a pair x_i, x_j , define the difference vector

$$d_{ij} = \phi(x_i) - \phi(x_j). \quad (2)$$

The observed pairwise response $Y_t \in \{0, 1\}$ follows the Bradley–Terry/MNL probability

$$\mathbb{P}(Y_t = 1 \mid \beta_t, x_i, x_j) = \text{logit}^{-1}(\kappa d_{ij}^\top \beta_t) = \frac{\exp\{\kappa u_t(x_i)\}}{\exp\{\kappa u_t(x_i)\} + \exp\{\kappa u_t(x_j)\}}, \quad (3)$$

where $\kappa > 0$ controls response consistency. Larger κ yields more deterministic choices. This observation model is the pairwise special case of the MNL model (Bradley and Terry, 1952; Luce, 1959).

Only the product of the response-consistency scale and coefficient magnitude is identified from binary comparisons without an additional scale convention. In the frozen confirmatory implementation, the preference learner fixes its observation-model scale to one and estimates coefficients under that convention. The simulation parameter κ is used in the response-generating mechanism to control choice consistency, but it is not supplied to the learner as a known multiplier. Consequently, absolute coefficient magnitude is scale-dependent; the induced ordering of alternatives and the relative direction of the coefficient vector are the more directly interpretable quantities.

At each iteration the optimizer must (i) maintain a posterior approximation for the current β_t , (ii) choose a comparison that balances immediate decision value and information, and (iii) recommend

$$\hat{x}_t = \arg \max_{x \in \mathcal{X}} \mathbb{E}[u_t(x) \mid \mathcal{D}_{t-1}], \quad (4)$$

where $\mathcal{D}_{t-1}$ is the preference history.

Because the optimum can move, the primary performance criterion is dynamic regret. Define

$$x_t^* = \arg \max_{x \in \mathcal{X}} u_t(x), \quad r_t = u_t(x_t^*) - u_t(\hat{x}_t), \quad (5)$$

and cumulative dynamic regret

$$R_T = \sum_{t=1}^T r_t. \quad (6)$$

A method can therefore be penalized even if it learned an earlier optimum accurately but failed to adapt after the utility changed.

4 DA-MNL-PBO Methodology

4.1 Two-regime dynamic preference model

A single random walk uses one fixed adaptation rate. Instead, we introduce a latent regime $Z_t \in \{0, 1\}$: mode 0 represents ordinary local evolution, and mode 1 represents a higher-variance adaptive state. Conditional on the mode,

$$\beta_t \mid \beta_{t-1}, Z_t = k \sim \mathcal{N}(\beta_{t-1}, Q_k), \quad Q_1 \succ Q_0. \quad (7)$$

The regime follows a first-order Markov chain

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}. \quad (8)$$

The posterior adaptive-mode probability

$$\delta_t = \mathbb{P}(Z_t = 1 \mid \mathcal{D}_t) \quad (9)$$

is used as a continuous indicator of how much probability mass the model assigns to the high-variance adaptation regime. It is not assumed to be a calibrated changepoint probability.

4.2 IMM prediction and mixing

At $t - 1$, each mode has Gaussian approximation

$$\beta_{t-1} \mid Z_{t-1} = j, \mathcal{D}_{t-1} \approx \mathcal{N}(m_{t-1}^{(j)}, P_{t-1}^{(j)}), \quad (10)$$

with mode probability $c_{t-1}^{(j)}$. The predicted mode probability is

$$\bar{c}_t^{(k)} = \sum_j \pi_{jk} c_{t-1}^{(j)}. \quad (11)$$

The IMM mixing weights are

$$\mu_{j|k} = \frac{\pi_{jk} c_{t-1}^{(j)}}{\bar{c}_t^{(k)}}. \quad (12)$$

The mixed state and covariance entering mode k are

$$\bar{m}_{t-1}^{(k)} = \sum_j \mu_{j|k} m_{t-1}^{(j)}, \quad (13)$$

$$\bar{P}_{t-1}^{(k)} = \sum_j \mu_{j|k} \left[P_{t-1}^{(j)} + (m_{t-1}^{(j)} - \bar{m}_{t-1}^{(k)})(m_{t-1}^{(j)} - \bar{m}_{t-1}^{(k)})^\top \right], \quad (14)$$

followed by the state prediction

$$m_{t|t-1}^{(k)} = \bar{m}_{t-1}^{(k)}, \quad P_{t|t-1}^{(k)} = \bar{P}_{t-1}^{(k)} + Q_k. \quad (15)$$

This is the mechanism through which the high-variance mode can rapidly reduce the effective influence of stale comparisons while the stable mode preserves information when evidence for change is weak.

4.3 Sequential Laplace update

The Bernoulli-logistic observation is non-Gaussian. For mode k , let (m_k, P_k) denote its predicted Gaussian state and let $d_t = \phi(x_i) - \phi(x_j)$ denote the queried raw feature-difference vector. The frozen confirmatory implementation uses the unit observation-scale convention described above. The log posterior is

$$\ell_k(\beta) = -\frac{1}{2}(\beta - m_k)^\top P_k^{-1}(\beta - m_k) + y_t d_t^\top \beta - \log\{1 + \exp(d_t^\top \beta)\} + C. \quad (16)$$

With $p(\beta) = \text{logit}^{-1}(d_t^\top \beta)$,

$$\nabla \ell_k(\beta) = -P_k^{-1}(\beta - m_k) + d_t \{y_t - p(\beta)\}, \quad (17)$$

$$-\nabla^2 \ell_k(\beta) = P_k^{-1} + p(\beta)\{1 - p(\beta)\}d_t d_t^\top. \quad (18)$$

Newton iterations locate the mode $m_t^{(k)}$, and the inverse negative Hessian gives $P_t^{(k)}$. For mode-likelihood calculations, the logistic-normal integral is approximated by

$$\mathbb{E}[\text{logit}^{-1}(X)] \approx \text{logit}^{-1}\left(\frac{\mu}{\sqrt{1 + \pi\sigma^2/8}}\right), \quad X \sim \mathcal{N}(\mu, \sigma^2). \quad (19)$$

The predictive likelihoods update the mode probabilities by Bayes' rule. Posterior mean and covariance used for recommendation are obtained by moment matching across modes.

4.4 Drift-aware acquisition

The acquisition evaluates challenger comparisons against the current posterior-best option. For a candidate pair (x_i, x_j) , let

$$U_i = \beta^\top \phi(x_i), \quad U_j = \beta^\top \phi(x_j), \quad (20)$$

with posterior means μ_i, μ_j and difference variance $s_{ij}^2 = \text{Var}(U_i - U_j)$. The optimization-value component is the Gaussian expected best-option utility

$$A_t^{\text{opt}}(i, j) = \mu_i \Phi(z) + \mu_j \Phi(-z) + s_{ij} \varphi(z), \quad z = \frac{\mu_i - \mu_j}{s_{ij}}, \quad (21)$$

which is the linear-Gaussian analogue of qEUBO (Astudillo et al., 2023).

Let p_0, p_1 be predictive probabilities that x_i wins under modes 0 and 1, and let c_0, c_1 be the corresponding predicted mode probabilities. With $\bar{p} = c_0 p_0 + c_1 p_1$, the mutual information between the next binary response and regime is

$$A_t^{\text{drift}}(i, j) = H_B(\bar{p}) - c_0 H_B(p_0) - c_1 H_B(p_1), \quad (22)$$

where $H_B(p) = -p \log p - (1 - p) \log(1 - p)$. A local preference-information proxy is

$$A_t^\beta(i, j) = \frac{1}{2} \log \{1 + s_{ij}^2 \bar{p}(1 - \bar{p})\}. \quad (23)$$

The implemented acquisition is

$$A_t(i, j) = \tilde{A}_t^{\text{opt}}(i, j) + \lambda_d \tilde{A}_t^{\text{drift}}(i, j) + \lambda_\beta c_1 \tilde{A}_t^\beta(i, j), \quad (24)$$

where tildes denote within-candidate standardization. The factor c_1 increases parameter-relearning value when the filter already assigns material probability to the adaptive mode.

The implementation uses current-best-versus-challenger pairs. This reduces comparison construction from $O(M^2)$ to $O(M)$ per iteration while keeping the query centered on alternatives that can challenge the current recommendation.

4.5 Method architecture

Figure 1 summarizes the complete pipeline. The engineering layer is fixed before PBO: N-CMAPSS trajectories are aggregated, an RUL model is fitted on development trajectories, and a maintenance-policy outcome table is generated. The preference layer then operates only on the resulting policy features. This separation prevents preference-optimization outcomes from feeding back into the prognostic model.

Algorithm 1 DA-MNL-PBO: drift-aware preference-based Bayesian optimization

Require: candidate set $\mathcal{X}$, features $\phi(x)$, mode states $(m_0^{(k)}, P_0^{(k)})$, Π , Q_0, Q_1 , λ_d, λ_β , budget T

- 1: **for** $t = 1, \dots, T$ **do**
- 2: Perform IMM mixing and prediction for modes $k \in \{0, 1\}$
- 3: Moment-match the predicted mixture and obtain current posterior-best $\hat{x}_t$
- 4: **for all** $x \in \mathcal{X} \setminus \{\hat{x}_t\}$ **do**
- 5: Compute $A_t^{\text{opt}}(x, \hat{x}_t)$
- 6: Compute mode-specific predictive probabilities p_0, p_1
- 7: Compute $A_t^{\text{drift}}(x, \hat{x}_t)$ and $A_t^\beta(x, \hat{x}_t)$
- 8: Combine standardized terms using Eq. (24)
- 9: **end for**
- 10: Query the maximizing pair and observe Y_t
- 11: Apply a Laplace logistic update separately in both modes
- 12: Update mode probabilities and set $\delta_t = \mathbb{P}(Z_t = 1 \mid \mathcal{D}_t)$
- 13: Moment-match the posterior mixture and report the recommendation
- 14: **end for**
- 15: **return** query history, recommendations, coefficient states, uncertainty, and adaptive-mode probabilities

5 Experimental Design

5.1 Pre-NASA freeze and research-integrity protocol

The DA-MNL-PBO filter and acquisition configuration was selected before observing the final NASA outcomes. The frozen values are $Q_0 = 0.0005I$, $Q_1 = 0.15I$, transition matrix $\Pi = \begin{bmatrix} 0.995 & 0.005 \\ 0.020 & 0.980 \end{bmatrix}$, $\lambda_d = 0.20$, and $\lambda_\beta = 0.30$. The static, 25-comparison sliding-window, and fixed-discount ($\gamma = 0.94$) baselines were also frozen. A controlled pre-NASA stage used six synthetic temporal scenarios and 200 paired replicates per scenario to compare acquisition variants under a common filter and common random streams; its role was model selection, not confirmatory NASA evidence. An earlier 100-replicate abrupt-drift implementation check had reduced mean regret from 161.46 for static MNL-PBO to 85.20 for the dynamic prototype, but we do not use that preliminary number as the principal evidence in this paper.

Throughout this paper, *confirmatory* refers to computational evaluation conducted under a frozen, pre-specified analysis protocol on held-out decision environments after completion of the development stage. The term does not imply a preregistered clinical or human-subject confirmatory trial; rather, it distinguishes the frozen evaluation from development, diagnostic, and post-confirmatory exploratory analyses.

The first completed NASA run was intentionally retained as a diagnostic record. It showed that 300 nominal parameterizations can map to fewer distinct outcome vectors and that the original stakeholder profiles had insufficient operational contrast. We therefore (i) collapsed outcome-equivalent alternatives, (ii) revised and documented the stakeholder scenario, and (iii) created a pre-specified eligibility criterion for new subsets. Crucially, the DA-MNL-PBO hyperparameters were not changed. The final stakeholder scenario, candidate subsets, response-scale grid, query budget, metrics, and eligibility rule were written to a freeze manifest before confirmatory PBO runs.

5.2 N-CMAPSS engineering data

N-CMAPSS is a C-MAPSS-generated aircraft-engine run-to-failure benchmark designed to reflect realistic flight conditions (Arias Chao et al., 2021). Its HDF5 files contain operating-condition descriptors, measured and virtual sensor variables, health parameters, auxiliary engine/cycle

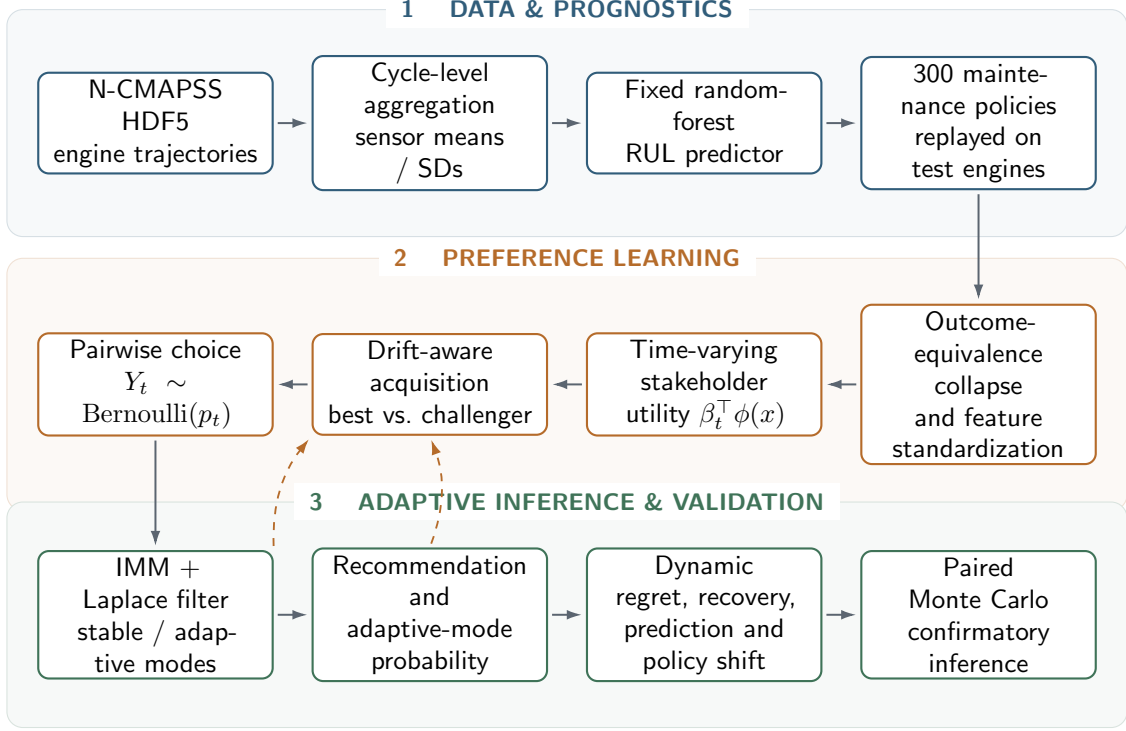


Figure 1: Architecture of the final DA-MNL-PBO validation pipeline. The prognostic/engineering layer is fitted before preference optimization. Pairwise preferences are semi-synthetic and generated from known time-varying stakeholder utility over real N-CMAPSS-derived policy outcomes.

indices, and RUL labels. We aggregate high-frequency development and test observations to engine-cycle level using means and standard deviations of measured sensors and operating conditions. The supplied development/test split is preserved.

A random-forest RUL regressor (Breiman, 2001) is fitted on development cycles and then frozen. It is infrastructure, not the proposed contribution. The test-set predictions are used only to evaluate maintenance policies. Across the four processed subsets, test RMSE ranged from 13.96 to 25.29 cycles and MAE from 10.86 to 21.45 cycles (Table 3 and figure 3). The confirmatory subsets contain 438 test cycles across six test engines (DS03-012) and 344 cycles across four engines (DS04).

5.3 Maintenance policy space and outcomes

A policy is

$$x = (I, h, s, w), \quad (25)$$

where I is inspection interval, h is an RUL threshold, s is an additive safety margin, and w is a smoothing window. The implemented trigger is

$$\widetilde{\text{RUL}}_t \leq h + s. \quad (26)$$

Because the trigger depends on $h + s$, we report the identifiable *effective trigger RUL* rather than interpreting h and s as independent operational effects. The grid is

$$I \in \{1, 2, 3, 5, 8\}, \quad h \in \{10, 20, 30, 40, 50\}, \quad s \in \{0, 5, 10, 15\}, \quad w \in \{1, 3, 5\}, \quad (27)$$

for 300 nominal policies.

Each policy is replayed over the test trajectories and summarized by five outcomes: late-maintenance risk $g_1(x)$, premature-maintenance loss $g_2(x)$, utilization $g_3(x)$, monitoring burden

Table 2: Frozen DA-MNL-PBO and confirmatory experimental settings. All algorithmic hyperparameters were selected before the N-CMAPSS confirmatory stage.

Setting	Frozen value
Stable process covariance	$Q_0 = 0.0005I$
Adaptive process covariance	$Q_1 = 0.15I$
Mode transition matrix	$\Pi = \begin{bmatrix} 0.995 & 0.005 \\ 0.020 & 0.980 \end{bmatrix}$
Initial mode probabilities	$(0.99, 0.01)$
Drift-information weight	$\lambda_d = 0.20$
Preference-information weight	$\lambda_\beta = 0.30$
Sliding-window baseline	25 comparisons
Fixed-discount baseline	$\gamma = 0.94$
Preference-query budget	$T = 160$
Abrupt event time	$\tau = 81$
Response scales	$\kappa \in \{0.75, 1.5, 3.0\}$
Monte Carlo replicates	200 per subset/condition/response scale
Conditions	stationary, noise-only, abrupt, gradual
Economic stakeholder weights	$(0.50, 1.00, 1.75, 0.50, 0.20)$
Safety stakeholder weights	$(2.00, 0.25, 0.10, 0.10, 1.75)$

$g_4(x)$, and uncertainty exposure at intervention $g_5(x)$. The preference feature vector is

$$\phi(x) = \text{standardize} [-g_1(x), -g_2(x), g_3(x), -g_4(x), -g_5(x)]^\top, \quad (28)$$

so larger values are preferable in every dimension.

Nominal policies with the same outcome vector are collapsed before preference optimization. The number of unique alternatives was 126 for DS02, 166 for DS03, 97 for DS04, and 175 for DS05 (Table 3 and figure 2).

5.4 Stakeholder preference scenarios and subset eligibility

The final economic/operations profile places relatively greater weight on utilization and useful-life preservation,

$$\beta^E \propto (0.50, 1.00, 1.75, 0.50, 0.20), \quad (29)$$

while the safety profile emphasizes late-risk avoidance and uncertainty exposure,

$$\beta^S \propto (2.00, 0.25, 0.10, 0.10, 1.75). \quad (30)$$

The vectors are normalized to unit Euclidean length before simulation.

The confirmatory eligibility rule was evaluated before PBO. Let x_E^* and x_S^* denote the unique optima under the two profiles. Define

$$C_E = \frac{U_E(x_E^*) - U_E(x_S^*)}{\max_x U_E(x) - \min_x U_E(x)}, \quad (31)$$

$$C_S = \frac{U_S(x_S^*) - U_S(x_E^*)}{\max_x U_S(x) - \min_x U_S(x)}. \quad (32)$$

A subset was confirmatory-eligible only if the optima were distinct and unique and $C_E > 0.10$ and $C_S > 0.10$. DS03 ($C_E = 0.126$, $C_S = 0.278$) and DS04 (0.462, 0.723) passed. DS05 failed because $C_E = 0.075$, and was excluded before PBO. DS02 is retained as development evidence because the stakeholder contrast was revised after its diagnostic run (Figure 2).

Table 3: N-CMAPSS engineering and confirmatory-design audit. The two cross-loss columns report the relative utility loss incurred by choosing the opposite phase optimum. Confirmatory inclusion required both losses to exceed the pre-specified 10% threshold, with distinct unique phase optima, before any confirmatory PBO run.

Subset	Role	Test units	Test cycles	RUL RMSE	RUL MAE	Unique policies	Econ. cross-loss	Safety cross-loss
DS02-006	Development	3	202	17.29	14.16	126	2.4%	34.4%
DS03-012	Confirmatory-included	6	438	13.96	10.86	166	12.6%	27.8%
DS04	Confirmatory-included	4	344	25.29	21.45	97	46.2%	72.3%
DS05	Confirmatory-excluded	4	327	16.03	12.63	175	7.5%	13.0%

5.5 Temporal conditions, baselines, and statistics

Each preference sequence contains $T = 160$ queries. For abrupt drift, $\beta_t = \beta^E$ for $t < 81$ and $\beta_t = \beta^S$ thereafter. Gradual drift interpolates between the two profiles over the horizon. The stationary control holds $\beta_t = \beta^E$. The noise-only control also holds $\beta_t = \beta^E$, but changes response consistency at the event time by multiplying the post-event κ by 0.45. The base response scales are $\kappa \in \{0.75, 1.5, 3.0\}$.

We compare four methods: static full-history MNL-PBO, DA-MNL-PBO, sliding-window MNL-PBO with window 25, and exponentially discounted MNL-PBO with discount 0.94. Within each subset/condition/ κ cell, 200 replicates use common random numbers across methods. The primary endpoint is cumulative dynamic regret. Secondary endpoints include post-event regret for abrupt drift, predictive log loss, Brier score, exact-best recovery, near-optimal recovery delay, adaptive-mode diagnostics, operational policy shifts, tail risk, and elapsed time. Coefficient RMSE and nominal coefficient coverage are retained only as scale-dependent descriptive diagnostics because absolute coefficient magnitude depends on the observation-scale convention.

Primary uncertainty summaries use Monte Carlo 95% intervals. Paired baseline-minus-DA differences are summarized using 5,000-resample bootstrap intervals for means and medians. Wilcoxon signed-rank tests are reported as sensitivity analyses with Benjamini–Hochberg multiplicity adjustment. Tail risk is described with the 95th percentile and CVaR95. The 0.50 adaptive-mode threshold is exploratory and is not treated as a calibrated alarm.

Response-consistency scale. The parameter κ controls stochasticity in the simulated pairwise responses through $\Pr(Y_t = 1) = \text{logit}^{-1}\{\kappa d_t^\top \beta_t\}$. In the frozen confirmatory implementation, κ is used by the data-generating response model but is not supplied to the preference learner as a known observation-scale multiplier; the sequential inference uses the unscaled feature-difference vector. Thus the experiments at $\kappa \in \{0.75, 1.5, 3.0\}$ evaluate the optimizer under progressively different levels of response consistency without assuming that the true consistency scale is known. A post-confirmatory sensitivity analysis additionally varies an assumed inference-scale multiplier to assess how the results depend on this modeling choice.

5.6 Post-confirmatory robustness analyses

After completion of the frozen confirmatory study, we conducted three additional analyses to probe mechanism and sensitivity. These analyses were not part of the pre-specified confirmatory test and are therefore reported as post-confirmatory exploratory evidence. First, an acquisition ablation retained the identical two-regime IMM/Laplace filter and removed, in turn, the regime-information term, the coefficient-information term, or both information terms. Second, a response-scale sensitivity analysis varied an assumed observation-model multiplier in $\{0.75, 1.0, 1.5, 3.0\}$ while retaining the original unit-scale implementation as the reference. Third, the fixed-discount baseline was evaluated over $\gamma \in \{0.85, 0.90, 0.94, 0.97, 0.99\}$; $\gamma = 0.94$ remained the frozen confirmatory baseline, while the best value selected separately within each scenario was treated only as a post-hoc oracle stress test.

N-CMAPSS validation design and pre-PBO confirmatory eligibility

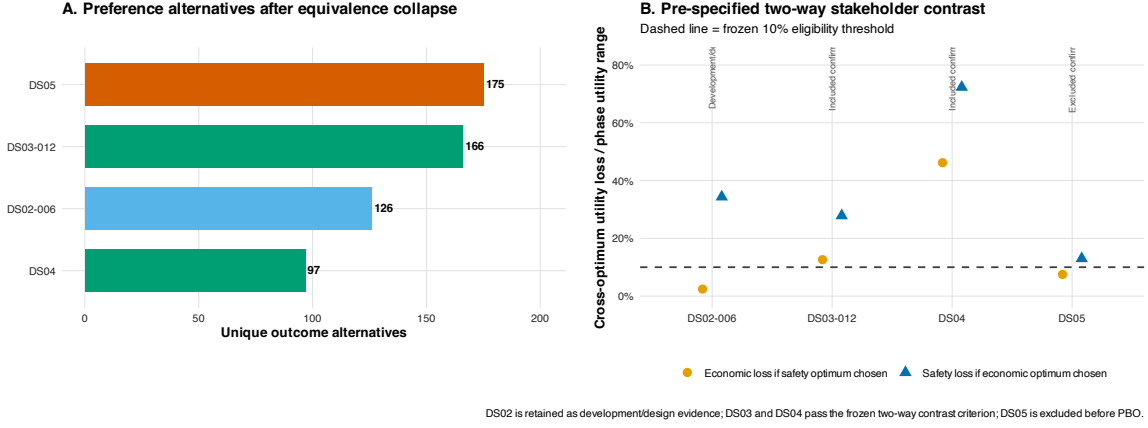


Figure 2: Validation design and pre-PBO eligibility. Outcome-equivalent policies are collapsed before preference optimization. The dashed 10% line is the frozen two-way cross-phase utility-loss threshold. DS03 and DS04 satisfy the rule; DS05 is excluded before PBO; DS02 remains development/design evidence.

Engineering layer: fixed RUL model before preference optimization

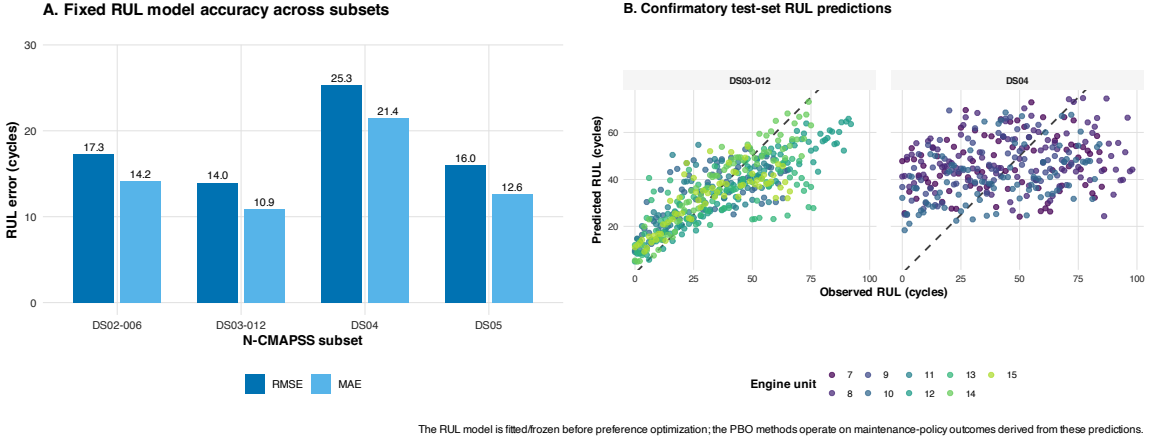


Figure 3: RUL engineering validation. Panel A reports fixed test RMSE and MAE across processed subsets; panel B compares observed and predicted RUL for the eligible confirmatory subsets. The RUL layer is fitted before preference optimization and is not updated by PBO.

All post-confirmatory analyses used the same two eligible confirmatory subsets, abrupt and gradual drift conditions, three data-generating response-consistency levels, 200 paired replicates per cell, and the original common-random-number seed construction. No DA-MNL-PBO hyperparameter was retuned. The three analyses added 28,800 method-level runs. A reproduction audit reran 60 frozen DA-MNL-PBO cases and reproduced cumulative dynamic regret to a maximum absolute difference of 3.84×10^{-13} .

6 Results

6.1 Engineering layer and confirmatory eligibility

Figure 2 shows that the final confirmatory design does not select datasets according to optimizer performance. Eligibility is determined entirely from policy-outcome geometry and the frozen stakeholder profiles before preference algorithms are run. This matters because DS05 is not “failed performance” evidence; it is a subset on which the pre-specified stakeholder contrast

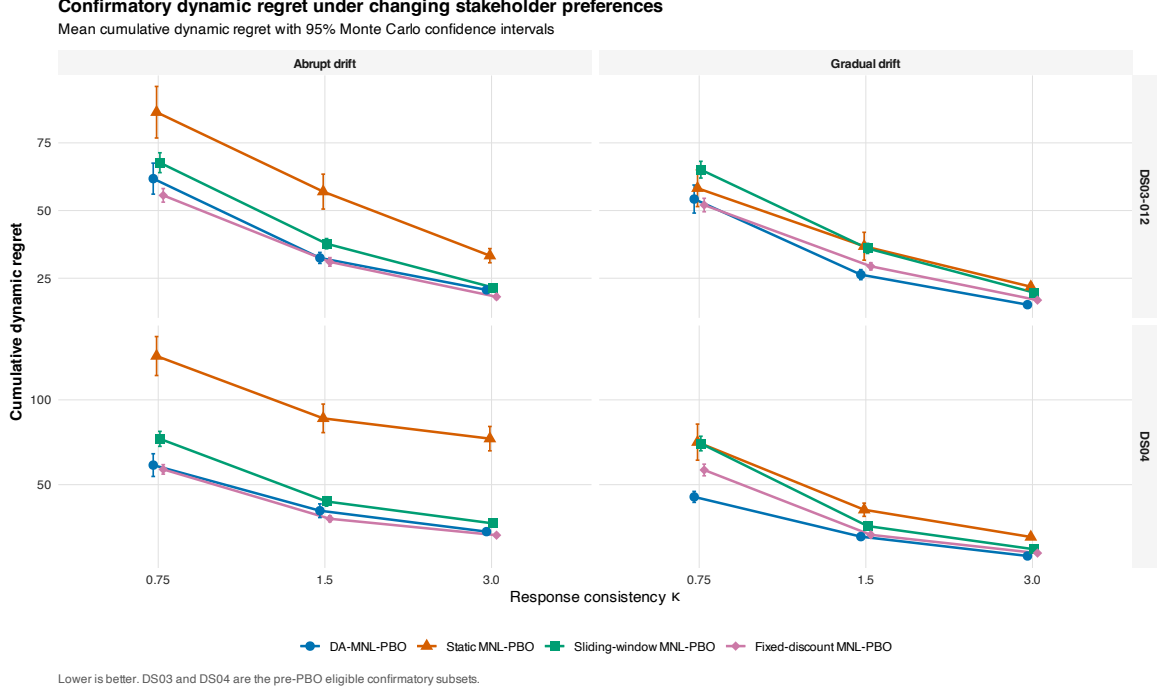


Figure 4: Mean cumulative dynamic regret with 95% Monte Carlo intervals for abrupt and gradual preference drift in the two confirmatory subsets. Lower is better.

is too weak in one direction to provide the intended confirmatory test. Figure 3 shows that the engineering layer is nontrivial but usable across subsets. DS04 has the largest RUL error, which is useful for stress-testing the maintenance-decision layer under a less favorable prognostic environment.

6.2 Primary endpoint: cumulative dynamic regret

DA-MNL-PBO has lower mean cumulative dynamic regret than static MNL-PBO in every one of the 12 confirmatory drift cells (Figure 4 and table 4). The relative reduction ranges from 6.9% (DS03 gradual, $\kappa = 0.75$) to 71.1% (DS04 abrupt, $\kappa = 3$), averaging 42.1%. The advantage over sliding-window MNL-PBO is also directionally consistent in all 12 cells, with an average mean-regret reduction of 20.4%.

The fixed-discount baseline is substantially harder to dominate. Under abrupt change it has lower mean regret than DA-MNL-PBO in all six confirmatory cells. For example, on DS04 at $\kappa = 1.5$, mean regret is 29.89 for fixed discount and 34.64 for DA-MNL-PBO. Under gradual change the pattern reverses in most settings: DA-MNL-PBO has lower mean regret in five of six cells, including 42.80 versus 58.66 on DS04 at $\kappa = 0.75$ and 7.96 versus 9.65 at $\kappa = 3$. The only gradual cell with a lower fixed-discount mean is DS03 at $\kappa = 0.75$ (52.04 versus 54.25).

This result identifies the empirical niche of the proposed approach more precisely than a blanket superiority claim. In these experiments, fixed exponential forgetting is highly effective after a sharp structural change when its discount is well matched to the environment. DA-MNL-PBO is comparatively strongest under gradual evolution, where evidence for changing utility accumulates over time. This pattern is consistent with the intended role of probabilistic allocation between stable and adaptive evolution, but the empirical comparison alone does not establish that mechanism as the unique cause of the observed advantage.

Table 4: Mean cumulative dynamic regret over 200 paired Monte Carlo replicates for the two confirmatory N-CMAPSS subsets. Lower values are better; the best mean within each scenario is bold.

Subset	Drift	κ	DA-MNL-PBO	Static	Sliding window	Fixed discount
DS03-012	Abrupt	0.75	61.79	86.32	67.67	55.61
DS03-012	Abrupt	1.5	32.50	56.99	37.76	31.04
DS03-012	Abrupt	3	20.70	33.33	21.57	18.15
DS03-012	Gradual	0.75	54.25	58.25	65.09	52.04
DS03-012	Gradual	1.5	26.33	36.82	36.01	29.40
DS03-012	Gradual	3	15.23	21.87	19.68	16.91
DS04	Abrupt	0.75	61.53	125.84	77.01	59.01
DS04	Abrupt	1.5	34.64	89.10	40.09	29.89
DS04	Abrupt	3	22.31	77.13	27.15	20.21
DS04	Gradual	0.75	42.80	75.07	74.25	58.66
DS04	Gradual	1.5	19.35	35.21	25.60	20.42
DS04	Gradual	3	7.96	19.15	11.91	9.65

Table 5: Cross-scenario paired-regret evidence for DA-MNL-PBO. Differences are defined as baseline regret minus DA-MNL-PBO regret, so positive values favor DA. “Positive CI” counts scenarios whose paired-bootstrap 95% CI for the mean lies fully above zero; BH significance uses the global Benjamini–Hochberg adjusted Wilcoxon sensitivity test.

Baseline	Positive mean	Positive CI	BH $p < 0.05$	Avg. paired diff.
Static MNL-PBO	12/12	11/12	11/12	26.31
Sliding-window MNL-PBO	12/12	10/12	11/12	8.70
Fixed-discount MNL-PBO	5/12	4/12	6/12	0.13

6.3 Paired effect sizes and robustness of the primary conclusion

The paired analysis strengthens the comparison with static and sliding-window baselines (Figure 5 and table 5). Against static MNL-PBO, the paired mean is positive in all 12 dynamic scenarios, the bootstrap mean interval lies entirely above zero in 11, and the globally BH-adjusted Wilcoxon test is below 0.05 in 11. The exception is DS03 gradual $\kappa = 0.75$, where the paired mean is +4.00 but the bootstrap interval spans $[-3.96, 12.14]$.

Against sliding-window PBO, all 12 paired means favor DA. Ten paired-bootstrap mean intervals are fully positive, while the BH-adjusted Wilcoxon sensitivity test is significant in 11. The two weakest mean-effect cases are DS03 abrupt $\kappa = 0.75$ (+5.88; bootstrap interval slightly crosses zero) and DS03 abrupt $\kappa = 3$ (+0.88; nonsignificant).

The fixed-discount comparison is intentionally mixed. Five of 12 paired means favor DA, concentrated in gradual conditions. Four gradual comparisons have bootstrap intervals fully above zero; several abrupt comparisons favor discount. This pattern is consistent with the mean-regret results and argues against interpreting the proposed adaptive-mode mechanism as a universal replacement for simple temporal discounting.

6.4 Stationary cost and drift-signal specificity

The stationary control answers whether a dynamic model pays an efficiency price when the decision maker does not drift. The penalty is modest rather than zero. On DS03 the mean DA-minus-static difference is +5.44 at $\kappa = 0.75$, -0.47 at 1.5, and +0.42 at 3. On DS04 it is +2.82, +2.05, and +1.67. Thus the evidence is consistent with H1 in the qualified sense intended by the design: adaptivity is not free, but the cost is small relative to the regret reductions observed when the target truly changes.

The noise-only control is more diagnostic for the latent regime. On both subsets, abrupt

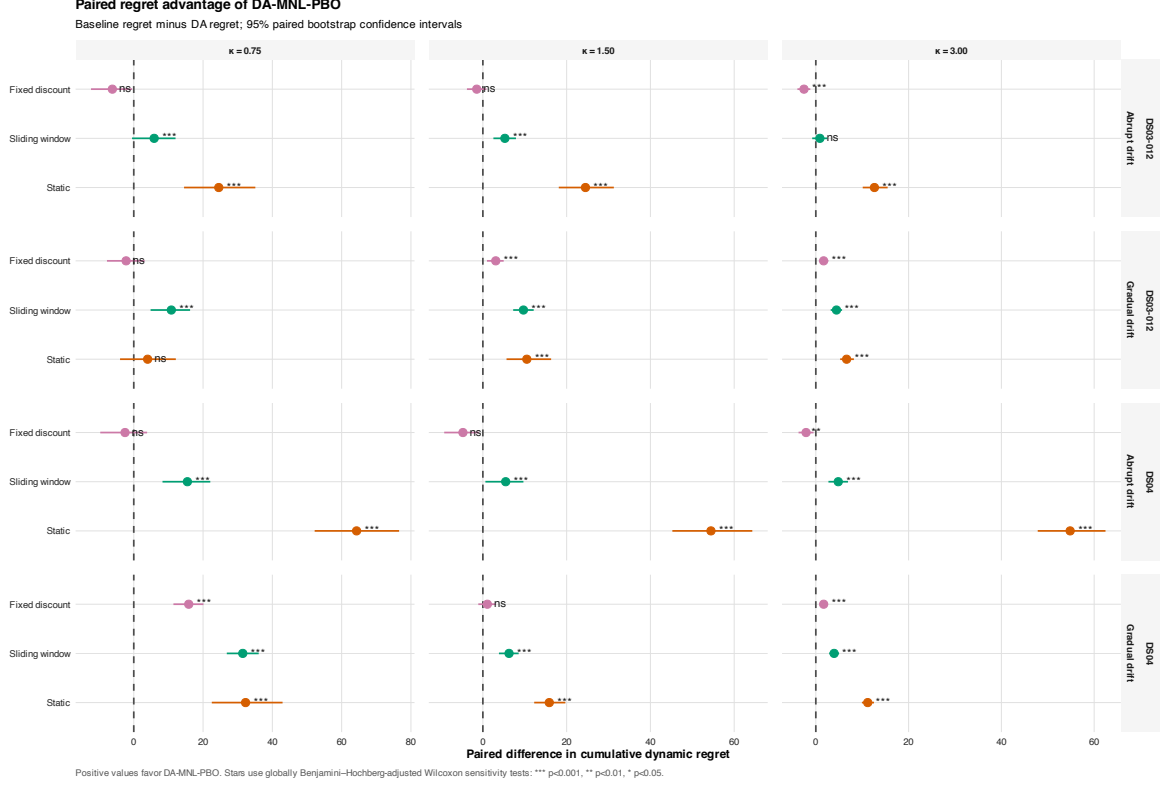


Figure 5: Paired baseline-minus-DA cumulative-regret differences with 95% bootstrap confidence intervals. Positive values favor DA-MNL-PBO. Stars denote globally Benjamini-Hochberg-adjusted Wilcoxon sensitivity tests.

drift produces a larger adaptive-mode response than stationarity. Separation from noise-only perturbation is stronger on DS04 than DS03. For DS04 at $\kappa = 3$, the mean post-minus-pre adaptive-mode probability is approximately 0.117 under abrupt change versus 0.018 under noise-only change and 0.004 under stationarity. In contrast, DS03 at $\kappa = 0.75$ produces similar abrupt and noise-only signals (approximately 0.124 and 0.118). We therefore interpret δ_t as a continuous evidence-weighting/adaptation signal, not a validated standalone drift detector.

6.5 Operational adaptation

The policy-level analysis translates latent preference adaptation into engineering actions (Figure 7 and table 6). On DS03, DA-MNL-PBO increases the effective trigger by 4.32, 2.37, and 0.65 RUL cycles as κ increases, while reducing the inspection interval by 2.49, 3.51, and 4.09 cycles. The phase-optimal trigger itself is 25 in both DS03 phases, so the strongest operational signature is more frequent inspection rather than a changed trigger. On DS04 the safety optimum is structurally more distinct: the economic and safety effective triggers are 35 and 40. DA-MNL-PBO shifts its recommended trigger by 13.76 to 22.04 cycles across the three response scales and simultaneously reduces the inspection interval by 3.27 to 5.86 cycles. The shift direction is consistently safety oriented.

This result is consistent with the operational version of H4, but also illustrates why the identifiable effective trigger $h + s$ is preferable to separate claims about the threshold and safety-margin parameters. Different nominal parameter pairs can generate the same maintenance trigger and therefore the same engineering outcomes.

Stationary robustness and drift-signal specificity

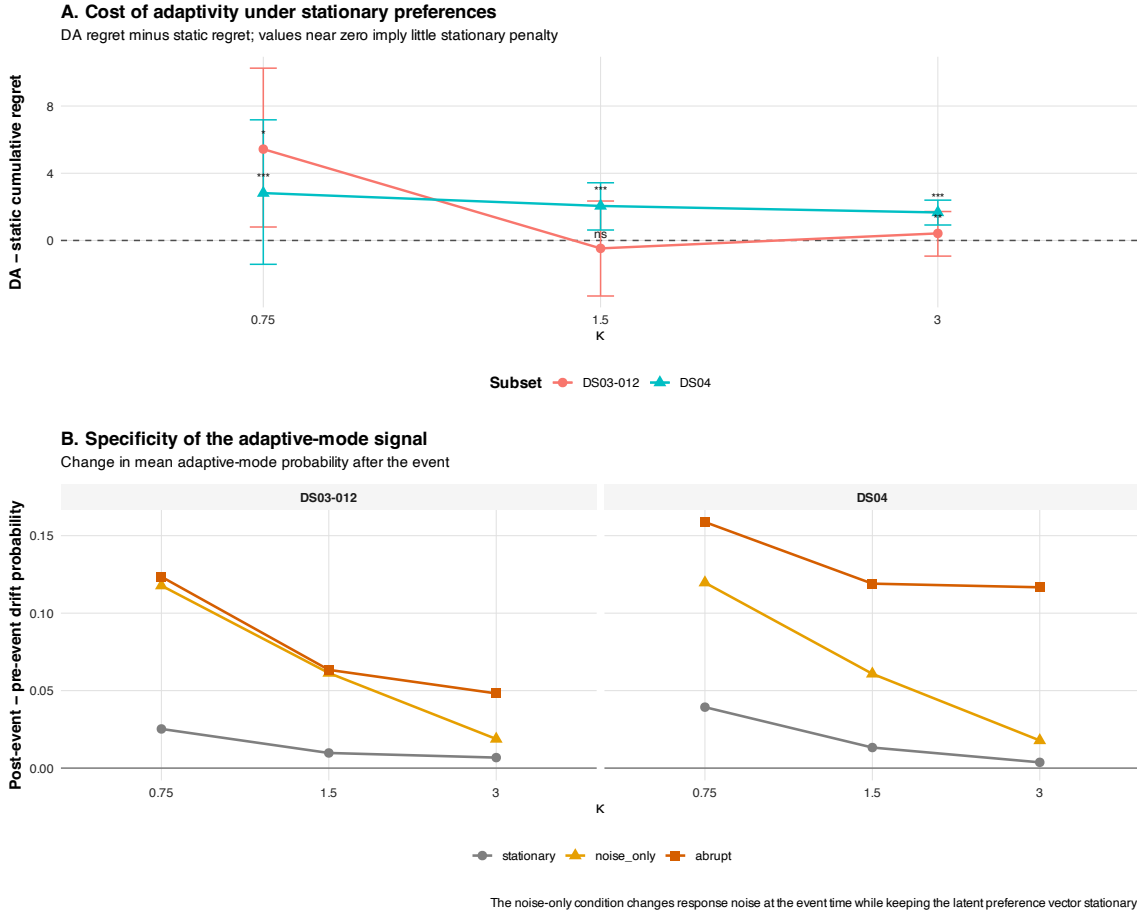


Figure 6: Stationary robustness and specificity of the adaptive-mode signal. Panel A reports DA-minus-static regret under stationary preferences; values near zero indicate little cost of maintaining adaptivity. Panel B reports the change in mean adaptive-mode probability after the event for stationary, noise-only, and abrupt conditions.

6.6 Tail risk and non-average behavior

The tail-risk results provide an important qualification to the mean-regret story. DA-MNL-PBO markedly reduces extreme losses relative to static MNL-PBO in many cells, but fixed discount often has lower CVaR95 in abrupt scenarios. For example, DS04 abrupt $\kappa = 0.75$ has CVaR95 values of 326.09 (static), 232.14 (DA), 156.54 (sliding), and 105.34 (discount). Under gradual drift, the ordering depends on subset and noise level. We therefore report tail metrics as complementary robustness diagnostics rather than evidence that the proposed method uniquely minimizes catastrophic risk.

Figure 9 condenses the primary result. Relative to static and sliding-window PBO, the heatmap is uniformly favorable across the confirmatory drift scenarios. Relative to fixed discount, the abrupt column is unfavorable and the gradual column is mostly favorable. This visual separation is central to the interpretation of the method.

6.7 Secondary predictive and inferential diagnostics

Supplementary Figs. S7–S20 report predictive log loss, Brier score, exact-best recovery, post-event regret, recovery delay, drift-threshold crossing, computational cost, regret–runtime trade-offs, and scale-dependent coefficient diagnostics. These metrics do not show uniform DA-MNL-PBO dominance. In particular, fixed-discount or static models can obtain better one-step predictive

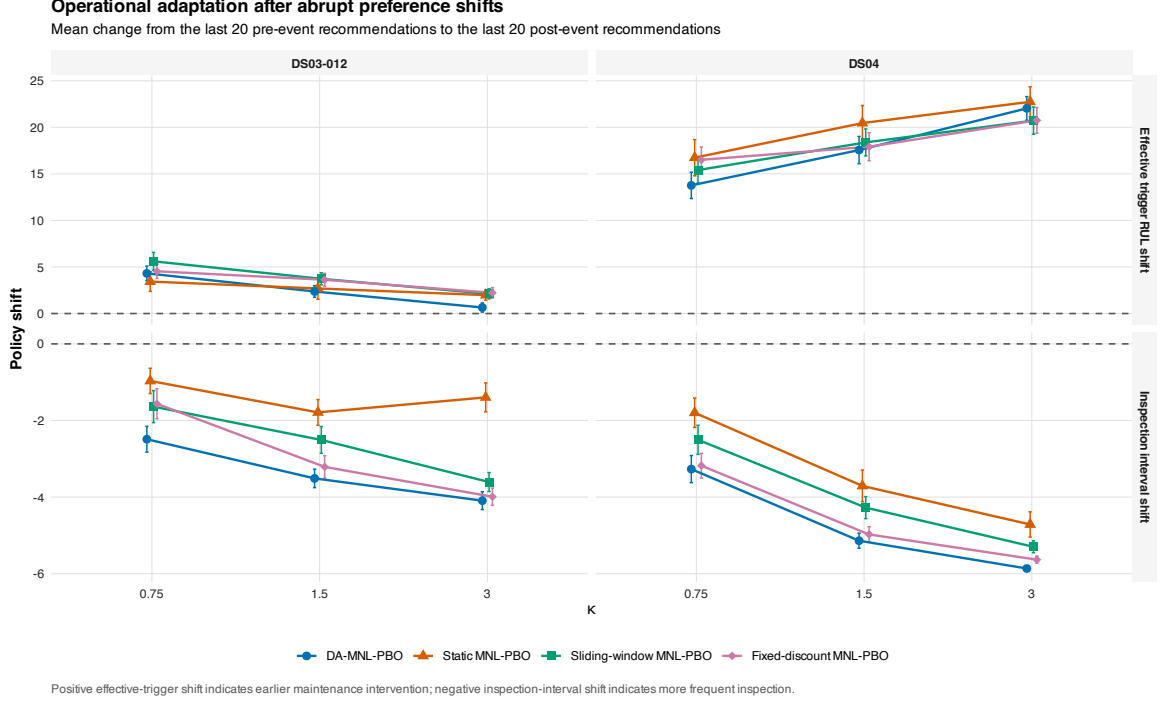


Figure 7: Operational shifts after abrupt stakeholder change. Positive effective-trigger shifts correspond to earlier maintenance intervention; negative inspection-interval shifts correspond to more frequent inspection.

scores in some cells even when DA-MNL-PBO has lower dynamic regret. This is not contradictory: the optimization endpoint rewards good sequential recommendations under a moving objective, whereas log loss rewards calibrated pairwise probability prediction.

Because the frozen learner fixes the observation scale to one while the data-generating response-consistency parameter varies, absolute coefficient magnitude is scale-dependent. We therefore do not interpret coefficient RMSE or nominal coefficient coverage across κ as calibrated recovery of a uniquely identified coefficient magnitude. The primary inferential claims are instead based on dynamic regret, predictive performance, recommendation recovery, operational adaptation, and the behavior of the adaptive-mode probability. Computation remains modest in this finite-candidate study: DA-MNL-PBO averages roughly 0.10–0.11 seconds per method run, compared with about 0.05–0.06 seconds for static MNL-PBO and 0.07–0.09 seconds for window/discount baselines on the analysis machine. The dynamic filter roughly doubles static per-run computation but remains small relative to the engineering-data preprocessing stage.

6.8 Post-confirmatory ablation and sensitivity analyses

The acquisition ablation showed that the incremental effect of the two information terms on the primary regret endpoint is context dependent rather than uniformly positive (Supplementary Fig. S35). Relative to removal of the regime-information term, the complete acquisition had lower mean cumulative regret in 7 of 12 dynamic cells; relative to removal of the coefficient-information term it had lower mean regret in 6 of 12 cells; and relative to the optimization-only acquisition it had lower mean regret in 8 of 12 cells. Paired bootstrap mean intervals were fully positive in 2, 0, and 1 cells, respectively, and no ablation comparison remained significant under the global Benjamini–Hochberg-adjusted Wilcoxon sensitivity analysis. The information terms nevertheless affected posterior predictive quality: the complete acquisition had lower mean predictive log loss and Brier score than the optimization-only acquisition in all 12 cells. These results do not support a claim that either information term independently guarantees lower dynamic regret;

Table 6: Operational adaptation of DA-MNL-PBO under abrupt preference change. Shifts compare the last 20 pre-event recommendations with the last 20 post-event recommendations. A positive effective-trigger shift means earlier intervention; a negative inspection-interval shift means more frequent inspection. Brackets are 95% Monte Carlo confidence intervals.

Subset	κ	Effective-trigger shift	Inspection-interval shift
DS03-012	0.75	4.32 [3.53,5.10]	-2.49 [-2.83,-2.15]
DS03-012	1.5	2.37 [1.74,2.99]	-3.51 [-3.75,-3.27]
DS03-012	3	0.65 [0.15,1.14]	-4.09 [-4.32,-3.86]
DS04	0.75	13.76 [12.35,15.17]	-3.27 [-3.62,-2.92]
DS04	1.5	17.56 [16.10,19.02]	-5.14 [-5.33,-4.94]
DS04	3	22.04 [20.77,23.30]	-5.86 [-5.92,-5.81]

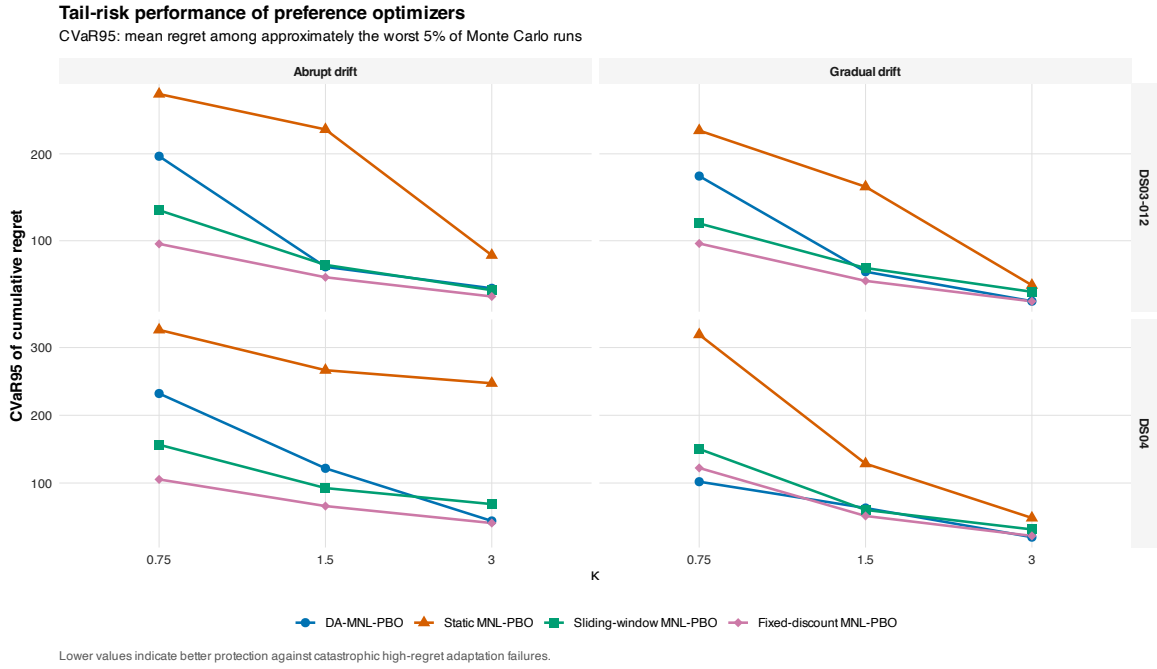
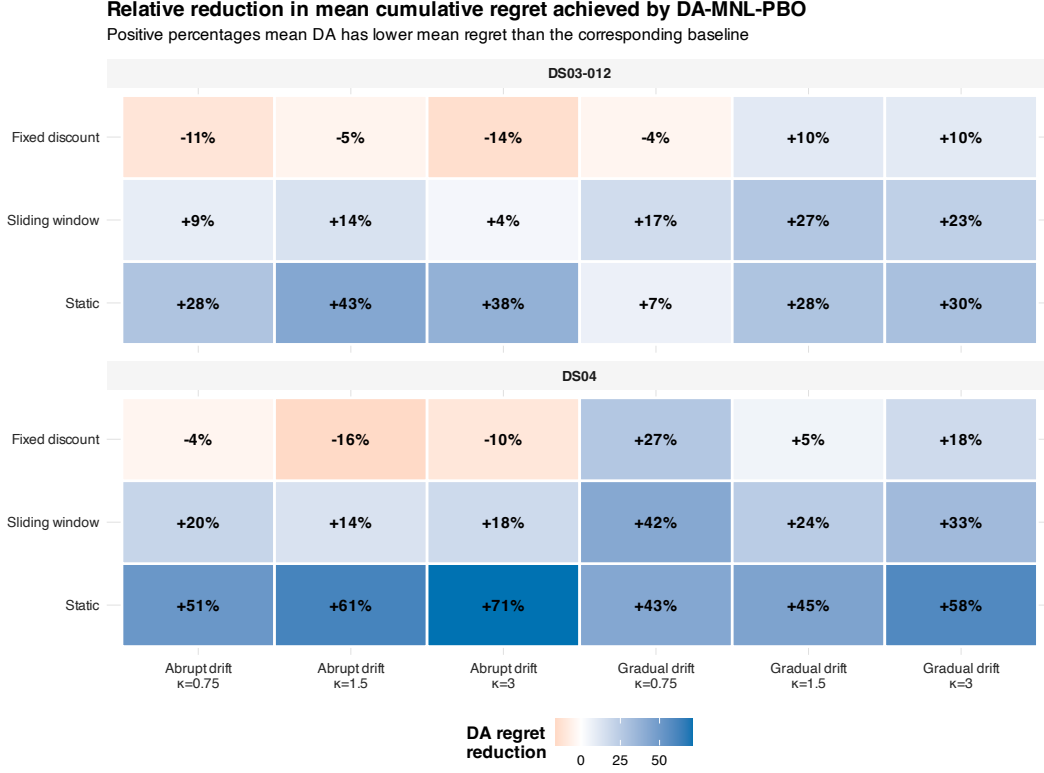


Figure 8: CVaR95 of cumulative dynamic regret, defined as mean regret among approximately the worst 5% of Monte Carlo runs. Lower values are better.

rather, their incremental value depends on the decision environment. Accordingly, the empirical evidence supports the complete drift-aware decision architecture more strongly than it supports a claim that either information component is individually necessary for regret reduction.

The response-scale analysis identified the learner’s observation-scale convention as a meaningful sensitivity dimension (Supplementary Fig. S36). Relative to the frozen unit-scale implementation, an assumed scale of 0.75 produced lower mean regret in 10 of 12 cells, whereas a scale of 1.5 produced higher mean regret in 10 of 12 cells and a scale of 3.0 produced higher mean regret in all 12 cells. Accordingly, the response-consistency parameter should not be interpreted as a known quantity supplied to the optimizer, and large observation-scale multipliers can materially degrade sequential decision quality. This analysis is exploratory and does not replace the unit-scale frozen confirmatory results.

The discount-factor sweep reinforced the regime-dependent comparison with exponential forgetting (Supplementary Fig. S37). Among the five tested values, the post-hoc cell-specific best discount was $\gamma = 0.94$ in 6 of 12 cells, $\gamma = 0.97$ in 5 cells, and $\gamma = 0.85$ in 1 cell, so no single tested discount was uniformly optimal. Even when fixed discounting was given the advantage of selecting its best tested value separately in each scenario, DA-MNL-PBO retained



This heatmap emphasizes effect magnitude; paired bootstrap inference is shown separately in Fig. 4.

Figure 9: Percentage reduction in mean cumulative regret of DA-MNL-PBO relative to each baseline. Positive values favor DA; negative values indicate a lower baseline mean.

lower mean regret in 5 of 6 gradual-drift cells, whereas the post-hoc oracle discount had lower mean regret in all 6 abrupt-shift cells. Thus the original conclusion is robust to reasonable discount-factor variation: the proposed method’s clearest empirical niche is gradual rather than abrupt preference change.

7 Discussion

7.1 What the results support

The strongest conclusion concerns dynamic decision quality under controlled preference evolution, rather than coefficient identification, human behavioral dynamics, or changepoint detection. Across two independently eligible confirmatory N-CMAPSS subsets, DA-MNL-PBO consistently reduces mean dynamic regret relative to a static full-history MNL surrogate. This confirms the core motivation: when the latent preference function changes, treating all historical comparisons as equally representative can create substantial stale-data bias. The same result is generally observed against a hard sliding window, suggesting that mechanically discarding all information beyond a fixed age is often less efficient than maintaining a probabilistic state model.

The fixed-discount baseline is more revealing. Its strong abrupt-shift performance shows that the advantage of a richer regime model is not simply “more forgetting.” A fixed discount can be highly effective when preferences move sharply and sustained forgetting is appropriate immediately. DA-MNL-PBO becomes relatively stronger under gradual change, where evidence for changing utility accumulates over time and the IMM can distribute posterior probability between stable and adaptive evolution rather than imposing one constant decay rate. The observed pattern is consistent with this modeling motivation, although the experiments do not

isolate regime inference as the sole causal explanation for the performance difference. More importantly, the comparison identifies when the additional modeling structure appears empirically worthwhile rather than asserting universal baseline dominance.

The post-confirmatory discount sweep sharpens this interpretation. The frozen $\gamma = 0.94$ was not uniquely privileged by the tested grid: a larger $\gamma = 0.97$ minimized mean regret in five cells and $\gamma = 0.85$ in one. Nevertheless, even a deliberately advantaged post-hoc oracle that selected the best tested discount separately within each scenario remained worse than DA-MNL-PBO in five of six gradual-drift cells, while outperforming it in all six abrupt-drift cells. The gradual-versus-abrupt distinction therefore does not arise solely from the particular frozen value $\gamma = 0.94$.

The acquisition ablation adds an important qualification. The information terms improve belief quality but do not produce a uniform primary-regret gain when removed one at a time. Their value is therefore best interpreted as context dependent within the complete dynamic decision rule, rather than as evidence that each term is individually necessary in every environment. This also motivates future work on adaptive or state-dependent acquisition weighting rather than fixed global values of λ_d and λ_β .

7.2 Interpretation of the adaptive-mode probability

The adaptive-mode posterior is useful but should be interpreted carefully. The stationary condition keeps it comparatively quiet, while abrupt changes typically increase it. The noise-only condition reveals that response inconsistency can also move the signal, especially when the underlying decision geometry makes preference and noise changes hard to distinguish. Therefore, δ_t is best understood as a posterior signal that current comparisons are more compatible with a high-variance evolution regime. It is not, in the present implementation, a calibrated probability that a discrete changepoint has occurred.

This distinction also explains why the algorithm can improve optimization even when the drift probability is not a perfect detector. The high-variance mode affects state uncertainty and the acquisition value continuously. Adaptation does not require crossing a hard threshold. A future version could add an explicit changepoint model or calibrate a decision threshold if operational alarm detection were itself the goal.

7.3 Engineering meaning of adaptation

The maintenance case study demonstrates why interpretable alternative features are useful. A change in stakeholder utility can be translated into changes in effective maintenance trigger, inspection frequency, utilization, and risk exposure. The DS03 and DS04 examples show that the same abstract “safety shift” need not map to the same operational action: when the two phase optima differ mainly in inspection interval, more frequent monitoring is the relevant response; when their effective triggers also differ, earlier intervention becomes visible. This is preferable to interpreting a generic latent GP movement without an attribute-level decision explanation.

The study also makes clear what N-CMAPSS contributes and what it does not. N-CMAPSS provides a realistic physical substrate for policy outcomes, including sensor trajectories and prognostic error. It does not provide observed human preference labels. The pairwise choice layer remains controlled and semi-synthetic, allowing exact dynamic-regret calculation and causal isolation of preference drift. Accordingly, the empirical claim concerns algorithmic adaptation in a realistic maintenance-decision environment, not verified human behavior in aviation maintenance.

7.4 Assessment of the evaluation hypotheses

H1 – stationary robustness. The evidence is consistent with H1. Maintaining adaptivity does not produce a large systematic deterioration under stationary preferences, although the

cost is not zero. The stationary penalty is therefore better described as modest than negligible.

H2 – dynamic decision quality. H2 is strongly supported relative to static full-history MNL-PBO and generally supported relative to the sliding-window baseline. DA-MNL-PBO has lower mean cumulative dynamic regret than both methods across all 12 confirmatory drift cells. Fixed exponential discounting remains a stronger, regime-dependent competitor: it performs better after abrupt changes, whereas DA-MNL-PBO is comparatively stronger under gradual drift. The post-confirmatory discount sweep shows that this qualitative distinction persists even when the discount factor is varied over a reasonable grid.

H3 – adaptive-mode responsiveness. H3 receives partial support. Abrupt preference drift produces a stronger adaptive-mode response than stationary preferences across the confirmatory subsets, but discrimination from a noise-only perturbation is subset and response-scale dependent. The adaptive-mode posterior should therefore be interpreted as a continuous adaptation signal rather than a calibrated changepoint probability.

H4 – operational adaptation. H4 is supported in the confirmatory maintenance environments. Following the safety-oriented preference shift, DA-MNL-PBO recommendations move toward earlier effective intervention, more frequent inspection, or both. The specific operational response depends on the geometry of the phase-optimal policies within each subset.

7.5 Relation to recent literature

The results complement several emerging directions in preference optimization and non-stationary sequential decision making. Recent theory establishes strong regret properties for PBO under stationary feedback (Kayal et al., 2025); local PBO improves the scalability of preferential search (Menn et al., 2026); and multi-task PBO transfers preference information across tasks and users (Coutinho et al., 2026). In parallel, dynamic pairwise-choice models and temporally weighted preference objectives address evolving preferences in ranking, bandit, and alignment settings (Bong et al., 2020; Kleine Buening and Saha, 2023; Son et al., 2025), while recent preference-aware BO considers adaptation to changing decision-maker preferences using explicit memory mechanisms (Ma et al., 2026).

The present contribution is complementary rather than a claim to originate non-stationary preference optimization. Its distinguishing feature is the use of interpretable preference coefficients as dynamic latent states within a PBO loop, together with probabilistic stable/adaptive regime inference and an acquisition component that can value information about that regime. Empirically, the results also clarify the limits of this additional structure: its clearest advantage appears under gradual preference evolution, whereas a well-matched fixed-discount rule remains difficult to improve upon after abrupt changes.

8 Limitations and Future Work

The study has several limitations that define the scope of inference.

1. **Semi-synthetic preferences.** N-CMAPSS provides realistic engine trajectories, not human preference observations. The preference paths are controlled, so conclusions about human behavioral dynamics require prospective user studies.
2. **Two confirmatory subsets.** Three additional subsets were pre-specified, but only DS03 and DS04 passed the frozen contrast criterion. This is methodologically appropriate, yet the confirmatory evidence still covers only two N-CMAPSS decision environments.
3. **Strong fixed-discount competitor.** Fixed exponential discounting outperforms DA-MNL-PBO in several abrupt-drift scenarios and often in abrupt tail risk. The proposed method should not be described as uniformly superior.

4. **Linear additive utility.** The MNL feature representation assumes an additive linear utility in standardized policy outcomes. Interactions, non-monotone preferences, and context-dependent attributes are not modeled.
5. **Approximate posterior inference.** Laplace/IMM inference is computationally convenient but does not represent arbitrary multimodality, and its uncertainty depends on the chosen observation-scale convention. We therefore do not claim calibrated coefficient credible intervals.
6. **Observation-scale identification and sensitivity.** Binary preference observations identify the product of response scale and coefficient magnitude unless an additional convention is imposed. The frozen learner fixes the observation scale to one, whereas the simulated response-consistency parameter varies. Post-confirmatory sensitivity analysis shows that the assumed learner scale can materially affect regret, particularly for large multipliers. Absolute coefficient magnitude, coefficient RMSE, and nominal coverage should therefore not be interpreted as scale-free inferential quantities.
7. **Context-dependent acquisition-component effects.** Post-confirmatory ablation does not show a uniform dynamic-regret advantage for each information term considered in isolation. The complete acquisition improves predictive scores relative to optimization-only querying, but its regret benefit is scenario dependent. Adaptive acquisition weights or principled weight learning are natural extensions.
8. **Drift signal is not a calibrated alarm.** The adaptive-mode posterior can react to pure noise changes. The 0.50 threshold is descriptive only.
9. **Finite candidate pool.** The acquisition is evaluated over a finite set of outcome-distinct policies and uses best-versus-challenger comparisons. Continuous or very large spaces may require local search, embeddings, or trust-region machinery such as recent local PBO approaches (Menn et al., 2026).
10. **RUL-model dependence.** The random forest is frozen and not the scientific contribution, but policy outcomes necessarily depend on prognostic quality. Alternative RUL models and controlled-error sensitivity analyses would strengthen engineering generality.
11. **Coverage of recent dynamic-preference optimizers.** The confirmatory baselines isolate three important memory assumptions—full-history stationarity, a hard sliding window, and exponential discounting—but do not exhaust the rapidly developing literature on preference-aware optimization under changing decision-maker preferences. Direct comparison with recent dynamic-preference BO/PBO frameworks would further clarify which gains arise specifically from probabilistic regime inference rather than from adaptation to non-stationarity more generally.
12. **No regret guarantee for the complete implemented algorithm.** Dynamic-regret theory exists for related non-stationary bandit and optimization settings, but the specific combination of IMM state evolution, Laplace preference inference, best-versus-challenger querying, and the composite acquisition used here is not accompanied by a formal regret bound. The reported dynamic-regret results are therefore empirical properties of the evaluated environments rather than consequences of a general theoretical guarantee. Establishing conditions under which regime-aware preference adaptation improves dynamic regret is an important theoretical direction.

Future work should therefore proceed along five directions. First, collect longitudinal human preference data in a controlled engineering or human-in-the-loop tuning experiment. Second, study identifiable or jointly learned response-scale formulations and adaptive acquisition weights

rather than fixed global information weights. Third, replace the two-regime random walk with richer state dynamics or explicit changepoint models and calibrate regime probabilities. Fourth, extend the utility beyond additive linear features, for example with hierarchical interactions or local nonlinear residual models while preserving interpretability. Fifth, evaluate the frozen method on larger N-CMAPSS subsets and other maintenance or control benchmarks, and compare against non-stationary dueling-bandit and GP-based dynamic PBO baselines where computationally feasible.

9 Conclusion

This paper introduced DA-MNL-PBO, a drift-aware preferential Bayesian optimization framework that treats interpretable MNL/Bradley–Terry coefficients as dynamic states. A two-regime IMM/Laplace filter provides both parameter uncertainty and a continuous adaptive-mode probability, and a composite acquisition rule uses optimization value, coefficient information, and regime information to choose comparisons.

The final N-CMAPSS study was deliberately structured as a confirmatory semi-synthetic experiment rather than a demonstration tuned to one dataset. Algorithmic hyperparameters were frozen before the NASA stage, outcome-equivalent policies were removed, stakeholder contrasts and subset eligibility were specified before confirmatory PBO, and stationary/noise-only controls were included. Across DS03 and DS04, DA-MNL-PBO consistently reduced mean dynamic regret relative to static and sliding-window MNL-PBO under preference drift. Its comparative advantage over fixed discount was strongest under gradual change, while fixed discount remained a strong or superior option under abrupt shifts. Post-confirmatory discount sensitivity showed that this regime-dependent pattern persists when the discount factor is varied: even a post-hoc cell-specific oracle discount remained worse than DA-MNL-PBO in five of six gradual cells but was better in all six abrupt cells. Acquisition ablation showed that the information terms have context-dependent rather than uniformly positive effects on dynamic regret, while improving predictive scores relative to optimization-only querying. The adaptive-mode signal was informative but not perfectly drift-specific, and response-scale sensitivity identified the observation-scale convention as an additional modeling limitation.

Taken together, the results support a deliberately qualified conclusion. Probabilistic dynamic preference modeling can improve sequential decision quality when stakeholder priorities evolve, but the value of additional temporal structure depends on how those priorities change. In the evaluated environments, DA-MNL-PBO is most compelling under gradual preference evolution, where it substantially reduces stale-data effects relative to static and fixed-window learning and remains competitive with, or improves upon, tuned exponential forgetting. After abrupt shifts, however, simple fixed discounting can adapt more effectively and should remain a serious baseline rather than being viewed as a method that regime modeling necessarily supersedes.

The contribution is therefore not a claim of universal dominance or of calibrated changepoint detection. It is a framework for representing interpretable preference coefficients as uncertain dynamic states and using that temporal uncertainty within sequential preference optimization. The semi-synthetic N-CMAPSS results establish the algorithmic plausibility of this approach under controlled preference evolution; validation with longitudinal human preferences, broader decision environments, contemporary dynamic-preference baselines, and theoretical analysis remains necessary to establish its generality.

Data, Code, and Reproducibility

The N-CMAPSS dataset is publicly available from NASA and is not redistributed with the manuscript (Arias Chao et al., 2021). The accompanying reproducibility package separates the frozen confirmatory outputs from all post-confirmatory analyses. It contains the final

analysis configuration, pre-analysis freeze manifest, subset-eligibility calculations, final validation note, plotting scripts, figure source data, machine-readable tables, and the post-confirmatory ablation and sensitivity outputs. The final R scripts record the seeds and frozen parameters. A reproduction audit of 60 frozen DA-MNL-PBO runs reproduced cumulative dynamic regret to a maximum absolute difference of 3.84×10^{-13} . The post-confirmatory analysis records the software environment (R 4.4.1 and package versions). Pairwise stakeholder preferences are simulated and should not be interpreted as observed human choices.

Ethics Statement

The study uses a public simulation-based engineering dataset and controlled synthetic preference responses. No human participants, personal data, or clinical records are used in the reported experiments.

CRedit authorship contribution statement

Muhammad Amir Saeed: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing—original draft, Visualization.

Sadia Saba: Methodology, Software, Validation, Visualization, Writing—review & editing.

Antonio Candelieri: Supervision, Methodology, Validation, Visualization, Writing—review & editing.

Acknowledgments

The authors thank the University of Milano-Bicocca research environment and the maintainers of the NASA N-CMAPSS benchmark.

References

- Arias Chao, Manuel, Chetan S. Kulkarni, Kai Goebel, and Olga Fink (2021). “Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics”. In: *Data* 6.1, p. 5. DOI: 10.3390/data6010005.
- Astudillo, Raul, Zhiyuan Jerry Lin, Eytan Bakshy, and Peter Frazier (2023). “qEUBO: A Decision-Theoretic Acquisition Function for Preferential Bayesian Optimization”. In: *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*. Vol. 206. Proceedings of Machine Learning Research. PMLR, pp. 1093–1114. URL: <https://proceedings.mlr.press/v206/astudillo23a.html>.
- Blom, Henk A. P. and Yaakov Bar-Shalom (1988). “The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients”. In: *IEEE Transactions on Automatic Control* 33.8, pp. 780–783. DOI: 10.1109/9.1299.
- Bogunovic, Ilija, Jonathan Scarlett, and Volkan Cevher (2016). “Time-Varying Gaussian Process Bandit Optimization”. In: *Proceedings of Machine Learning Research* 51. AISTATS 2016, pp. 314–323. URL: <https://proceedings.mlr.press/v51/bogunovic16.html>.
- Bong, Heejong, Wanshan Li, Shamindra Shrotriya, and Alessandro Rinaldo (2020). “Non-parametric Estimation in the Dynamic Bradley–Terry Model”. In: *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*. Vol. 108. Proceedings of Machine Learning Research. PMLR, pp. 3317–3326. URL: <https://proceedings.mlr.press/v108/bong20a.html>.

- Bradley, Ralph Allan and Milton E. Terry (1952). “Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons”. In: *Biometrika* 39.3/4, pp. 324–345. DOI: 10.2307/2334029.
- Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45, pp. 5–32. DOI: 10.1023/A:1010933404324.
- Cercola, Matteo, Valeria Capretti, and Simone Formentin (2026). “Efficient Reinforcement Learning from Human Feedback via Bayesian Preference Inference”. In: *IFAC Journal of Systems and Control* 35, p. 100398. DOI: 10.1016/j.ifacsc.2026.100398.
- Coutinho, João P. L., You Peng, Ricardo Rendall, Kaiwen Ma, Swee-Teng Chin, Ivan Castillo, and Marco S. Reis (2026). “Efficient Human-in-the-Loop MPC Tuning with Multi-Task Preferential Bayesian Optimization”. In: *Control Engineering Practice* 173, p. 106972. DOI: 10.1016/j.conengprac.2026.106972.
- González, Javier, Zhenwen Dai, Andreas Damianou, and Neil D. Lawrence (2017). “Preferential Bayesian Optimization”. In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 1282–1291. URL: <https://proceedings.mlr.press/v70/gonzalez17a.html>.
- Kayal, Aya, Sattar Vakili, Laura Toni, Da-Shan Shiu, and Alberto Bernacchia (2025). “Bayesian Optimization from Human Feedback: Near-Optimal Regret Bounds”. In: *Proceedings of the 42nd International Conference on Machine Learning*. Vol. 267. Proceedings of Machine Learning Research. PMLR, pp. 29449–29468. URL: <https://proceedings.mlr.press/v267/kayal25a.html>.
- Kim, GilHwan and Fabrizio Sergi (2026). “Validation of Dynamic Bayesian Optimization for a Non-Stationary Human-in-the-Loop Optimization Problem”. In: *IEEE Robotics and Automation Letters* 11.5, pp. 5733–5740. DOI: 10.1109/LRA.2026.3665072.
- Kleine Buening, Thomas and Aadirupa Saha (2023). “ANACONDA: An Improved Dynamic Regret Algorithm for Adaptive Non-Stationary Dueling Bandits”. In: *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*. Vol. 206. Proceedings of Machine Learning Research. PMLR, pp. 3854–3878. URL: <https://proceedings.mlr.press/v206/kleine-buening23a.html>.
- Luce, R. Duncan (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
- Ma, Yujie, Ludi Wang, Wenjuan Cui, Yuanchun Zhou, and Yi Du (2026). “Preference-Aware Bayesian Optimization for Interactive Decision Making”. In: *IEEE Transactions on Cybernetics* 56.4. Published online 25 November 2025, pp. 1970–1981. DOI: 10.1109/TCYB.2025.3633283.
- Menn, Johanna, Miriam Kober, Paul Brunzema, David Stenger, and Sebastian Trimpe (2026). *Local Preferential Bayesian Optimization*. DOI: 10.48550/arXiv.2606.02351. arXiv: 2606.02351 [cs.LG].
- Mikkola, Petrus, Milica Todorović, Jari Järvi, Patrick Rinke, and Samuel Kaski (2020). “Projective Preferential Bayesian Optimization”. In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. Proceedings of Machine Learning Research. PMLR, pp. 6884–6892. URL: <https://proceedings.mlr.press/v119/mikkola20a.html>.
- Saeed, Muhammad Amir and Antonio Candelieri (2026). “Bayesian Optimization for Categorical and Mixed Variables Using a Multinomial Logit Surrogate”. In: *Algorithms* 19.5, p. 361. DOI: 10.3390/a19050361.
- Saeed, Muhammad Amir, Jo Eidsvik, and Antonio Candelieri (2026). “Multi-Objective Bayesian Optimization Framework for CO₂ Injection Strategy Design: A Sleipner-Inspired Study”. In: *Geoenergy Science and Engineering*. DOI: 10.1016/j.geoen.2026.214834.
- Saeed, Muhammad Amir, Sadia Saba, and Antonio Candelieri (2026). *Cost-Aware Preference-Based Bayesian Optimization for Structured Prompt Configuration of Large Language Models*. SSRN preprint. DOI: 10.2139/ssrn.7308958.

- Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas (2016). “Taking the Human Out of the Loop: A Review of Bayesian Optimization”. In: *Proceedings of the IEEE* 104.1, pp. 148–175. DOI: 10.1109/JPROC.2015.2494218.
- Son, Seongho, William Bankes, Sayak Ray Chowdhury, Brooks Paige, and Ilija Bogunovic (2025). “Right Now, Wrong Then: Non-Stationary Direct Preference Optimization under Preference Drift”. In: *Proceedings of the 42nd International Conference on Machine Learning*. Vol. 267. Proceedings of Machine Learning Research. PMLR, pp. 56173–56203. URL: <https://proceedings.mlr.press/v267/son25a.html>.
- Takeno, Shion, Masahiro Nomura, and Masayuki Karasuyama (2023). “Towards Practical Preferential Bayesian Optimization with Skew Gaussian Processes”. In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. PMLR, pp. 33516–33533. URL: <https://proceedings.mlr.press/v202/takeno23b.html>.
- Xu, Wenjie, Wenbin Wang, Yuning Jiang, Bratislav Svetozarevic, and Colin N. Jones (2024). “Principled Preferential Bayesian Optimization”. In: *Proceedings of the 41st International Conference on Machine Learning*. Vol. 235. Proceedings of Machine Learning Research. PMLR, pp. 55305–55336. URL: <https://proceedings.mlr.press/v235/xu24y.html>.