Evaluating the Predictive Added Value of Dynamics-Aware Spatial Partitions Under Tactical Distribution Shift

André N. Costa [1 2] **Vitor C. F. Gomes** [1] **Petter Ögren** [2]

Abstract

Classical Voronoi Diagrams (VD) provide interpretable spatial partitions for multi-agent systems, but Euclidean proximity is often an incomplete proxy for influence in aerospace engagements, where velocity, mass, and control effort affect spatial dominance. We evaluate whether Linear Quadratic Regulator (LQR)-derived Cost-Induced Voronoi (CIV) summaries provide target-relevant information beyond classical VD geometry and tactical scenario descriptors in high-fidelity air combat simulations. Formulating this as a fold-safe conditional added-value test, CIV features are residualized against the VD and scenario baseline within each training fold, and their explanatory contribution is evaluated using paired out-of-fold probabilistic losses over execution-level pre-hit summaries. The resulting added-value signal persists under scenario-cell clustered inference, negative controls, early flight-state controls, and leave-one-factor-out validation. Within the evaluated scenario family, CIV provides consistent pre-hit explanatory added value for binary target-damage occurrence, supporting its utility as a dynamics-aware representation of the active engagement phase.

1. Introduction

Spatial partitioning is a central representation for multi-agent coordination, coverage, and tactical analysis. Classical Voronoi Diagrams (VD) provide an interpretable way to assign regions of influence based on Euclidean proximity, and have been widely used in multi-agent coverage and sensing problems (Cortes et al., 2004; Aurenhammer et al., 2013; Okabe & Sugihara, 2012). However, Euclidean proximity is often an incomplete proxy for influence in aerospace engagements. In Beyond-Visual-Range (BVR) scenarios, an aircraft's ability to affect a region depends not only on position, but also on velocity, maneuverability, mass state, and control effort.

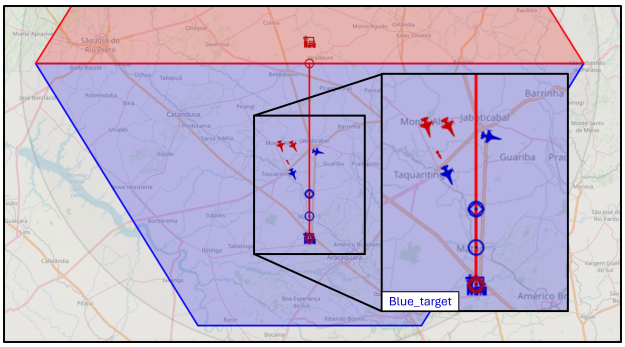

Figure 1. ASA BVR target-defense scenario used in the simulation study. Red aircraft (attackers) attempt to damage the defended Blue target, while Blue aircraft (defenders) perform a combat air patrol to prevent or delay target damage. The inset highlights the local engagement geometry around the defended target where spatial-control summaries are computed. For the experiments, we systematically vary the attack axis, red and blue shot philosophies, blue missile counts, and red aircraft counts across 64 tactical scenario configurations.

Cost-Induced Voronoi (CIV) regions address this limitation by replacing Euclidean distance with transition costs derived from a control model. In particular, Linear Quadratic Regulator (LQR)-based CIV metrics provide a tractable dynamics-aware partition in which second-order motion, velocity, and control effort affect the resulting regions of influence (Costa & Ögren, 2025a; Costa et al., 2026). These metrics are attractive because they preserve the interpretability of spatial partitions while incorporating agent dynamics. Yet their value in high-fidelity simulation is not guaranteed by the mathematical formulation alone: a simplified control-theoretic cost may or may not contain information about tactical outcomes beyond what is already captured by Euclidean VD geometry and scenario descriptors.

We formulate this evaluation as a conditional added-value problem for a physics-informed representation. Let Y denote the target-damage outcome, X a CIV-derived spatial-

[1]Institute for Advanced Studies (IEAv), São José dos Campos, São Paulo, Brazil. [2]Division of Robotics, Perception, and Learning (RPL), Royal Institute of Technology (KTH), Stockholm, Sweden. Correspondence to: André N. Costa <andrenc@kth.se>.

ICML 2026 Workshop on Hypothesis Testing, Seoul, South Korea, 2026. Copyright 2026 by the author(s).

control summary, and Z the baseline information set consisting of tactical scenario variables and Euclidean VD geometry. Because exact nonparametric conditional-independence testing is statistically limited when conditioning on high-dimensional continuous variables (Z) without strong functional assumptions (Shah & Peters, 2020), we instead evaluate a targeted added-value null. We test whether the component of X (the dynamics-aware CIV summary) not explained by Z improves held-out probabilistic scoring for the binary outcome Y. This approach transforms a theoretically "hard" independence test into a practical, fold-safe added-value problem.

This framing clarifies our hypothesis: we do not claim that the LQR surrogate is a calibrated controller or a complete tactical predictor, but rather that dynamics-aware partitions summarize target-relevant information not captured by geometric baselines. Rigorously testing this requires overcoming significant confounders, as apparent gains in simulation data can easily arise from temporal dependence or shared tactical structures rather than the representation itself. We evaluate this framework on data-farmed BVR simulations generated in the Aerospace Simulation Environment (ASA) (Dantas et al., 2024). The dataset contains 1,920 executions across 64 tactical scenario cells, varying attack axis, red and blue shot philosophy, blue missile count, and red aircraft count. Additionally, we varied the scenario random seed, which influences some decisions within aircraft behavior, such as missile launch instant and defensive maneuver turn direction, as well as other stochastic elements within the scenario like the missile kill probability. Because spatial dominance is most critical during the active engagement phase (before a target is damaged and dynamics fundamentally shift) VD and CIV utilities are aggregated into execution-level pre-hit summaries. To account for repeated executions within tactical regimes, inference is conducted at the scenario-cell level. Robustness is assessed using leave-one-factor-out validation, negative controls, and early flight-state covariates.

The contributions of this paper are threefold. First, we formulate the validation of dynamics-aware spatial partitions as a practical fold-safe added-value test for physics-informed representations when exact conditional-independence testing is infeasible. Second, we design an execution-level aggregation and scenario-cell inference protocol that avoids treating temporally dependent frames or repeated random seeds as independent evidence. Third, we show, in a high-fidelity ASA BVR scenario family, that LQR-CIV provides consistent pre-hit explanatory added value beyond classical VD and tactical scenario descriptors, with negative controls and tactical-factor holdout validation supporting the specificity and robustness of the signal.

2. Background

To contextualize the proposed conditional added-value test, this section reviews two relevant areas of prior work. First, we discuss the evolution of spatial partitioning from static geometric diagrams to dynamics-aware representations. Second, we review the application of these partitions to adversarial aerospace simulation, highlighting the critical gap between a theoretical mathematical partition and an empirically validated tactical metric.

2.1. Dynamics-aware spatial partitioning.

VDs are a standard tool for assigning regions of influence in multi-agent systems, with applications in coverage control, robotic coordination, and task allocation (Cortes et al., 2004; Okabe & Sugihara, 2012; Aurenhammer et al., 2013). Classical VDs, however, define proximity using Euclidean distance and therefore ignore the dynamic state and control limitations of physical agents. This restricts their use in aerospace settings, where velocity, maneuverability, mass state, and control effort dictate which agent can most effectively reach or influence a region. While generalized and weighted VDs partially address agent heterogeneity, they often remain tied to static geometric distance functions. CIV regions instead define influence through transition costs induced by a control model. In particular, LQR-based CIV formulations provide a tractable, dynamics-aware approximation in which second-order motion, velocity, and control effort are naturally embedded into the spatial partition (Costa & Ögren, 2025a).

2.2. Adversarial spatial dominance in tactical simulation.

Many aerospace problems are not purely cooperative coverage tasks, but adversarial interactions between opposing teams. In BVR engagements, the relevant quantity is not only which agent owns a geometric region, but whether a team can protect, deny, or threaten mission-relevant areas. Adversarial formulations extend individual partitions into team-level regions of influence, evaluating them by integrating a tactical-value density over the corresponding regions (Costa & Ögren, 2025b; Costa et al., 2026). ASA-based BVR studies have previously used simulation data to support engagement analysis and situational-awareness modeling (Dantas et al., 2022a,b), making spatial partitioning a natural representation for tactical dominance. However, a mathematical partition is not automatically a validated mission metric; a dynamics-aware partition may be elegant under its assumed surrogate model while still failing to add information in a high-fidelity, nonlinear simulation. The central empirical question is therefore whether CIV summaries explain tactical outcomes beyond what is already captured by classical Voronoi geometry and scenario descriptors.

3. Methodology

We evaluate whether CIV summaries contain target-relevant spatial-control information beyond classical VD geometry and tactical scenario descriptors. Let D_i^{target} denote the final continuous damage sustained by the defended blue target in execution i. For the primary analysis, this continuous damage measure was converted into a binary damage-occurrence label,

$$Y_i = \mathbf{1}\{D_i^{\text{target}} > 0\},$$

so that Y_i indicates whether the defended target sustained any nonzero damage during the simulation. This binarization yields a focused occurrence task and avoids introducing additional distributional assumptions for the continuous damage magnitude, which includes a point mass at zero and a continuous positive range.

Operationally, the null hypothesis is that adding the fold-residualized CIV summary does not improve out-of-fold probabilistic scoring relative to the VD scenario baseline. Positive evidence is interpreted as conditional added value of the dynamics-aware spatial representation, not as a claim of standalone early-warning capability.

3.1. Simulation Data and Spatial Representations

The data were generated in ASA using a BVR target-defense vignette (Dantas et al., 2024). Red aircraft attempted to damage a defended blue target, while blue aircraft attempted to prevent or delay target damage. The final dataset contained 1,920 simulation executions distributed across 64 scenario cells. Each scenario cell was defined by the combination of attack axis, blue shot philosophy, red shot philosophy, blue missile count, and red aircraft count; random seeds provided repeated executions within each cell.

Figure 1 provides a visual overview of the simulated target-defense vignette. The scenario consists of a defended blue target, blue aircraft assigned to protect or delay attacks against that target performing a combat air patrol, and red aircraft approaching from prescribed tactical axes carrying BVR missiles and general purpose unguided bombs. The inset highlights the local engagement region around the defended target, where the spatial-control summaries used in this paper are computed. The quantitative analysis is based on execution-level VD and CIV summaries aggregated from the simulation outputs.

For each execution-frame pair, we computed team-level spatial-control utilities using both classical VD and LQR-derived CIV representations (Okabe & Sugihara, 2012; Costa & Ögren, 2025a; Costa et al., 2026). The VD baseline assigned points in the analysis domain according to Euclidean proximity. CIV instead assigned points according to simplified control-theoretic transition costs, producing a dynamics-aware notion of spatial influence. We evaluated three CIV variants: shared-linear, isotropic, and anisotropic CIV. These variants are treated as sensitivity checks around the same central question: whether dynamics-aware spatial-control summaries add information beyond VD. Classical Euclidean VD was selected as the strict geometric baseline because it requires zero parametric tuning, completely eliminating the risk of data leakage or artificial baseline degradation that can occur when calibrating weights for Generalized or Power Voronoi Diagrams.

To focus the analysis on target-defense relevance, all utilities were weighted by a Gaussian density centered on the static position of the defended blue ground target. The primary frame-level feature was the red-minus-blue target-weighted advantage,

$$D_i^{(m)}(t) = U_{i,\text{red}}^{(m)}(t) - U_{i,\text{blue}}^{(m)}(t), \tag{1}$$

where m denotes the representation: VD, shared-linear CIV, isotropic CIV, or anisotropic CIV. Positive values indicate greater red spatial-control advantage around the defended target. Figure 2 illustrates how the CIV variants deform the induced partitions relative to classical VD for the same aircraft configuration in a particular simulation frame.

3.2. LQR-CIV Feature Construction

For each simulation frame, active aircraft were represented by planar position and velocity, $s_i = [p_{x,i}, p_{y,i}, v_{x,i}, v_{y,i}]^\top$. For a candidate spatial point q, the LQR-CIV cost of assigning that point to aircraft i was modeled as a quadratic transfer cost, $C_i(q) = e_i(q)^\top P_i e_i(q)$, where $e_i(q) = [p_i - q, v_i]^\top$. The matrix P_i is the infinite-horizon LQR cost-to-go matrix associated with a planar second-order surrogate model. In continuous-time LQR, this matrix is obtained as the stabilizing solution of the standard algebraic Riccati equation ($A^\top P_i + P_i A - P_i B R_i^{-1} B^\top P_i + Q_i = 0$), so that the resulting quadratic value function provides a control-aware proximity measure rather than a Euclidean distance (Lewis et al., 2012).

To ensure the representation did not encode an a priori side advantage, the same nominal LQR parameters were assigned to both teams: maneuver scale $a = 3.0$ (corresponding to approximately 3 km in the local horizontal plane) and a neutral control-effort multiplier $r = 1.0$. Frame-level mass effects were incorporated through an effective control penalty using current fuel and remaining payload information, whereas the shared-linear CIV variant disabled this scaling to isolate the effect of common linearized LQR geometry. All representations shared the same damage labels and preprocessing pipeline, and were computed on a 200x200 grid over the blue-zone domain using a 1.0 s velocity look-ahead and a 6.0 km target Gaussian σ centered at the ground target.

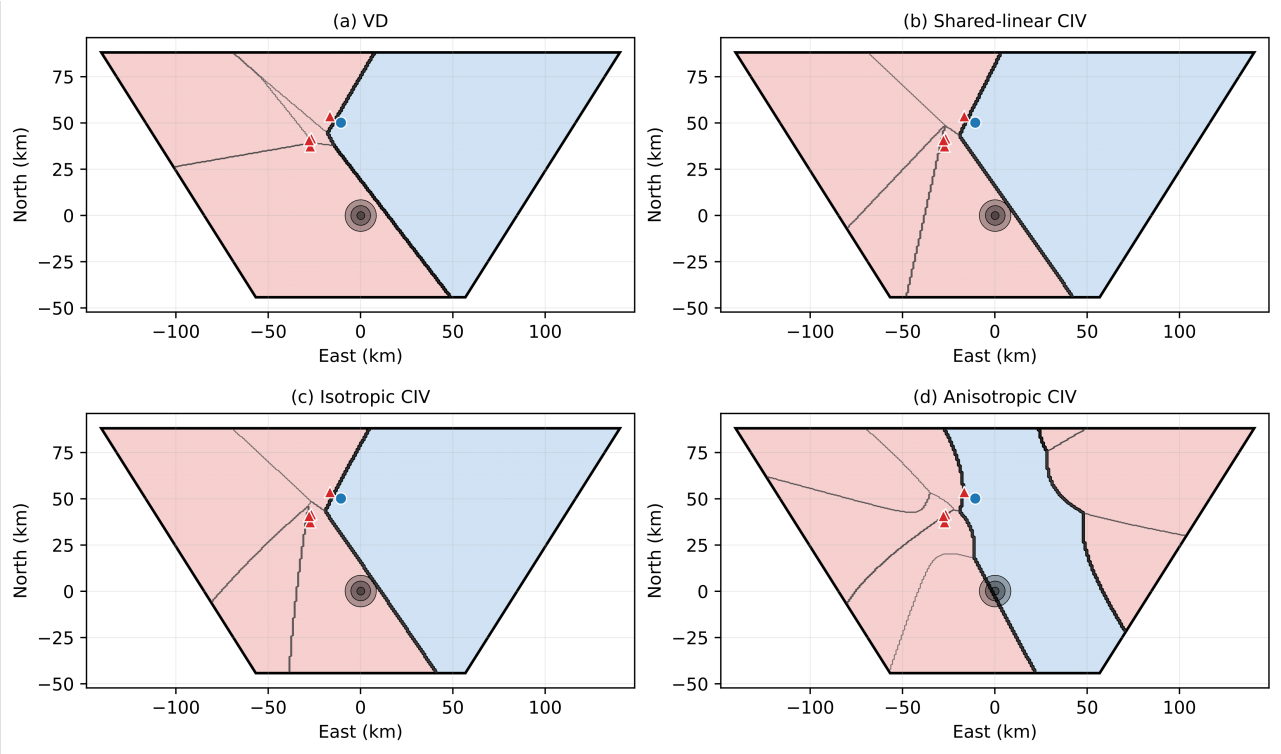

Figure 2. Representative engagement frame from the ASA BVR target-defense scenario. In this setup, Red aircraft (red triangles) attempt to damage a defended Blue target (black circle), while Blue aircraft (blue circle) maneuver to intercept or delay the attack. The partition boundaries represent team-level spatial dominance, weighted by a Gaussian density centered on the defended target to prioritize target-relevant influence. Subplot **(a)** displays the classical Euclidean VD, while subplots **(b)-(d)** illustrate how CIV variants deform these boundaries to account for agent velocity and maneuverability. To contrast the different partitioning methods, this frame highlights a specific tactical configuration: one Red aircraft in close proximity to a Blue fighter, alongside a cluster of three other Red aircraft. By utilizing LQR-derived transition costs, these partitions reflect the kinematic effect that a high-velocity agent can influence a different spatial region than one described only by instantaneous Euclidean position, which is the information used by classical VD.

Importantly, the LQR parameters were fixed before the added-value tests and were not tuned against the target-damage labels. Thus, the LQR surrogate is not presented as a calibrated aircraft controller. It is used as a physics-inspired spatial representation whose added value is evaluated statistically downstream.

3.3. Execution-Level Aggregation

Simulation frames within the same execution are strongly temporally dependent and are therefore not treated as independent samples. Instead, frame-level spatial utilities are aggregated into one sample per execution. The primary analysis uses the pre-hit window, defined as all frames strictly before the first target hit. If no hit occurs, the available mission trajectory is used.

For each representation m, the primary execution-level summary is the cumulative red-minus-blue spatial-control advantage,

$$A_i^{(m)} = \int_{\mathcal{T}_i} D_i^{(m)}(t)\, dt, \tag{2}$$

where $\mathcal{T}_i$ is the aggregation window for execution i. A duration-normalized mean was also computed as a sensitivity feature, but the main analysis uses cumulative area under the curve because it captures both the magnitude and persistence of spatial-control advantage.

While integrating over the entire pre-hit window might normally risk washing out the tactical signal with early, irrelevant maneuvering, the Gaussian target-weighting applied to $D_i^{(m)}(t)$ naturally mitigates this. Spatial-control advantages are heavily discounted when agents are far from the defended asset, meaning the cumulative integral is implicitly dominated by the active, target-relevant engagement phase.

This aggregation should be interpreted as a pre-damage explanatory summary for damage occurrence, not as a fixed-time operational forecast or a model of continuous damage severity. The analysis asks whether the spatial-control conditions observed before target damage contain information beyond VD and scenario descriptors. Consequently, the pre-hit analysis should not be interpreted as a deployable real-time warning system; it is a retrospective test of whether

the spatial-control trajectory before damage contains information associated with the eventual damage outcome.

3.4. Fold-Safe Explanatory Added-Value Test

The primary comparison is between a reduced model using tactical scenario variables and VD, and a full model that additionally includes the CIV component not explained by the reduced information set. The reduced model estimates the probability of any target damage, $P(Y_i = 1 \mid Z_i)$, where Z_i contains scenario variables and the VD summary. For each CIV variant, the CIV summary is first residualized against Z_i within each training fold, $X_i = h(Z_i) + \epsilon_i$, where h is a linear regression model operating on the continuous VD summary and one-hot encoded categorical scenario variables. The estimated residual is $\widehat{\epsilon}_i = X_i - \widehat{h}(Z_i)$. The residualization map $\widehat{h}$ is fit only on the training fold and then applied to the held-out fold, preventing leakage from the test data. This linear formulation prioritizes transparency, interpretability, and strict fold-safety, minimizing the risk of overfitting the residualization step on the available training folds. While this restriction implies that highly non-linear relationships between the baseline variables and the CIV features might not be perfectly partialled out, it provides a stable estimate of the residual information that is properly aligned with the linear capacity of the downstream logistic regression models. Extending this framework to non-linear residualization models (e.g., gradient boosting) to perform a stricter conditional independence test is a direction for future work.

The full model estimates

$$P(Y_i = 1 \mid Z_i, \widehat{\epsilon}_i). \tag{3}$$

Because the primary outcome is binary damage occurrence, both reduced and full models are implemented as logistic regressions using the same cross-validation splits. Performance is evaluated out of fold using log loss as the primary metric and Brier score as a secondary metric. For each execution, the paired loss improvement is

$$\Delta_i = \ell(Y_i, \widehat{p}_i^{\text{red}}) - \ell(Y_i, \widehat{p}_i^{\text{full}}), \tag{4}$$

where $\widehat{p}_i^{\text{red}}$ and $\widehat{p}_i^{\text{full}}$ are the predicted probabilities from the reduced and full models, respectively, and positive values indicate that the CIV-augmented model achieves better held-out probabilistic scoring.

Uncertainty is assessed using bootstrap confidence intervals (2,000 resamples) and paired sign-flip tests over the loss improvements (5,000 Monte Carlo resamples) (Tibshirani & Efron, 1993; Good, 2005). Because of this resampling scheme, the smallest reportable one-sided p-value is 2.0×10^{-4}. These p-values are interpreted as paired added-value tests under the chosen resampling unit, not as distribution-free tests of exact conditional independence. They are used to assess whether improvements are consistently positive, not to rank the relative strength of evidence among CIV variants; effect sizes and confidence intervals are emphasized for comparing variants.

3.5. Dependence-Aware Inference and Robustness Checks

Repeated executions from the same scenario cell share the same tactical configuration. Therefore, the primary inference is conducted at the scenario-cell level. We use grouped cross-validation so that all executions from a scenario cell remain in the same fold. Grouped cross-validation prevents training-test leakage across executions sharing the same tactical cell, whereas scenario-cell inference prevents repeated random seeds within a cell from being counted as independent evidence. We then average paired loss improvements within each scenario cell,

$$\Delta_c = \frac{1}{n_c} \sum_{i \in c} \Delta_i, \tag{5}$$

and perform inference over the scenario-cell improvements $\{\Delta_c\}$. This makes the scenario cell, rather than the individual frame or repeated seed, the primary inferential unit and avoids treating repeated stochastic executions within the same tactical regime as fully independent evidence.

By treating the scenario cell as the inferential unit, we explicitly assume that the paired predictive improvements across distinct combinations of tactical factors are exchangeable under the null hypothesis. This mitigates the severe overconfidence that would result from treating temporally dependent frames or random seeds sharing the same tactical regime as independent samples.

We include three robustness checks. First, negative controls repeat the added-value pipeline using shuffled target labels and placebo CIV features; these tests check whether the pipeline mechanically rewards arbitrary residualized covariates (Lipsitch et al., 2010). Second, early flight-state controls augment the reduced model with initial or fixed-window summaries of fuel, gross weight, altitude, thrust, g-load, and target-distance variables. This tests whether CIV added value persists beyond baseline physical-state differences. Third, leave-one-factor-out validation holds out entire values of tactical factors, such as attack axis, missile count, aircraft count, or shot philosophy, to assess robustness under structured tactical distribution shift.

The resulting claim is intentionally scoped. We test whether CIV provides conditional added value for pre-hit execution-level analysis of target damage within the ASA scenario family. We do not claim that CIV alone provides early mission forecasting or operational decision support. Prospective horizon design and real-time use are left for future work.

Table 1. Primary pre-hit added-value test. Positive Δ values indicate improvement of the full model over the reduced scenario+VD model. Confidence intervals and p-values are computed over scenario cells. All rows had $p_{\text{one}} = 2.0\times10^{-4}$.

CIV variant	ΔLL [95% CI]	ΔBrier [95% CI]
Isotropic	0.043 [0.027, 0.060]	0.013 [0.007, 0.019]
Shared-linear	0.043 [0.028, 0.058]	0.013 [0.009, 0.018]
Anisotropic	0.027 [0.017, 0.037]	0.009 [0.006, 0.013]

Table 2. Negative controls. Shuffled outcomes break the relationship between features and target damage, while placebo CIV replaces the CIV residual with an arbitrary shuffled covariate. The near-zero or slightly negative improvements, confidence intervals crossing zero, and large one-sided p-values indicate no systematic positive added value under the control conditions.

Control	Metric	Δ	95% CI	p_{one}
Shuffled outcome	LL	-0.0003	[-0.0010, 0.0004]	0.781
Shuffled outcome	Brier	-0.0001	[-0.0005, 0.0002]	0.762
Placebo CIV	LL	0.0000	[-0.0009, 0.0009]	0.502
Placebo CIV	Brier	-0.0001	[-0.0005, 0.0004]	0.583

4. Results

The final simulation results dataset contained 3,036,887 frame-level records, corresponding to 1,920 executions across 64 tactical scenario cells. The continuous target-damage variable was binarized into a damage-occurrence label for the primary analysis. The blue target sustained nonzero damage in 1,113 executions, yielding a positive rate of 57.97%. All primary analyses were performed after aggregating frame-level VD and CIV utilities into one pre-hit execution-level summary per simulation run.

4.1. Pre-Hit Added Value

Table 1 reports the primary pre-hit added-value test. The reduced model used tactical scenario variables and the cumulative VD red-minus-blue target-weighted advantage. The full models additionally included the fold-residualized CIV advantage. Positive values indicate that the CIV-augmented model achieved lower out-of-fold loss than the VD baseline.

All three CIV variants improved both log loss and Brier score relative to the reduced model. The largest improvements were obtained by the isotropic and shared-linear CIV variants, while anisotropic CIV produced a smaller but still positive improvement. This indicates that the CIV summaries contained target-relevant information not already explained by tactical scenario descriptors and VD geometry.

The effect was therefore not limited to a single scoring rule. Because log loss is sensitive to probabilistic confidence and Brier score measures squared probability error, the agreement between the two metrics suggests that the CIV residual improved probabilistic target-damage modeling rather than merely altering a decision threshold.

Interestingly, the simpler isotropic and shared-linear variants yielded larger improvements than the more complex anisotropic variant. Tactically, this is likely because BVR engagements involve frequent heading changes for missile evasion (e.g., cranking or notching). Because anisotropic CIV is highly sensitive to instantaneous heading, its execution-level integral may accumulate noise from transient maneuvers. In contrast, isotropic CIV captures the general energy and reachability state of the aircraft more stably over the course of the engagement.

While the sign-flip tests indicate consistently positive improvements, the effect size is also practically meaningful for post-simulation target-damage analysis. Because the reduced model already includes strong scenario descriptors such as missile and aircraft counts, an additional 0.043 log-loss (LL) improvement and an approximately 1.3 percentage-point reduction in Brier score from a compact spatial-control representation indicate that CIV captures systematic engagement information not summarized by VD geometry alone.

4.2. Robustness and Negative Controls

We next assessed whether the added-value signal was explained by baseline physical-state differences or by artifacts of the residualization pipeline. As a robustness check, we augmented the reduced model with initial and fixed-window summaries of fuel, mass, altitude, speed, thrust, g-load, and distance-to-target variables. The CIV residuals remained beneficial after these controls, with log-loss improvements of 0.046, 0.045, and 0.027 for isotropic, shared-linear, and anisotropic CIV, respectively, closely matching the primary results in Table 1. This suggests that the observed added value was not explained solely by baseline physical-state imbalances.

Table 2 reports the negative-control experiments. The shuffled-outcome control broke the relationship between features and target damage, while the placebo-feature control replaced the CIV residual with an arbitrary shuffled covariate. Both controls yielded near-zero or slightly negative improvements, confidence intervals crossing zero, and large one-sided p-values. This indicates that the added-value pipeline does not mechanically reward arbitrary residualized features or produce spurious gains when the relationship between features and target damage is destroyed.

Figure 3 summarizes the primary log-loss pattern: CIV variants produced stable positive scenario-cell improvements, whereas the negative controls remained centered near zero.

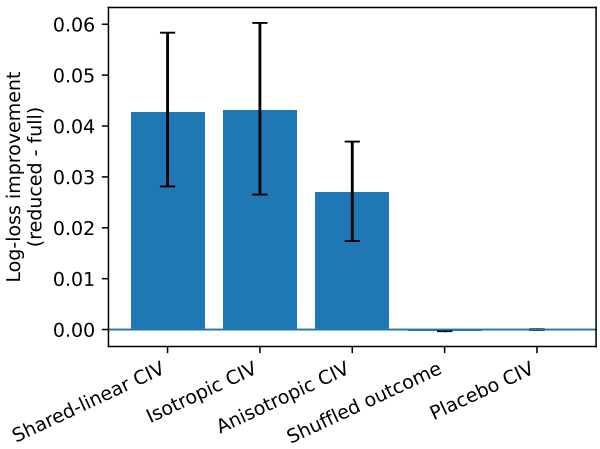

Figure 3. Scenario-cell clustered CIV added value and negative controls. Positive values indicate lower out-of-fold log loss for the CIV-augmented model. The shuffled-outcome and placebo-feature controls remain near zero, indicating that the pipeline does not mechanically reward arbitrary residualized features.

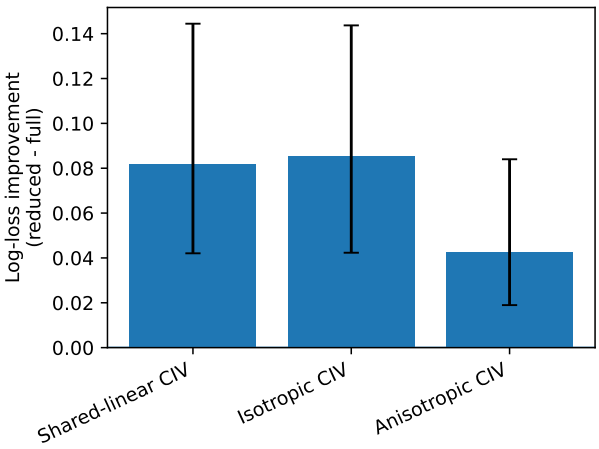

Figure 4. Leave-one-factor-out validation under structured tactical distribution shift. Positive values indicate lower out-of-fold log loss for the CIV-augmented model relative to the reduced scenario+VD model when values of a tactical factor are held out from training. The figure shows that CIV improvements remain positive across structured changes in tactical conditions, supporting the interpretation that the added-value signal is not driven only by repeated stochastic executions within fixed scenario cells.

4.3. Robustness Under Tactical-Factor Holdout

Leave-one-factor-out validation was used to test whether CIV added value persisted under structured tactical distribution shift. For each tactical factor, validation was performed by holding out one value of that factor at a time, training on the remaining executions, and evaluating on executions with the held-out value. The resulting improvements were then summarized at the factor level. This tested whether the CIV residual remained useful when generalizing across unseen values of attack axis, blue shot philosophy, red shot philosophy, blue missile count, and red aircraft count.

Figure 4 summarizes the resulting log-loss improvements across CIV variants, while Table 3 reports the best-performing CIV variant for each held-out factor. Across all held-out tactical factors, CIV produced positive out-of-fold improvements over the reduced VD model. The smallest gains occurred when holding out blue missile count and attack axis, while the largest gains occurred when holding out red aircraft count and red shot philosophy, with best observed improvements ranging from approximately 0.042 to 0.146 log-loss units.

These factor-holdout results should nevertheless be interpreted as robustness evidence within the evaluated ASA scenario family, rather than as a guarantee of universal out-of-distribution generalization.

Table 3. Best leave-one-factor-out result for each held-out tactical factor. For each factor, the table reports the CIV variant with the largest out-of-fold log-loss improvement among the evaluated CIV variants. Positive values indicate lower log loss for the CIV-augmented model relative to the reduced scenario+VD model. All reported rows had $p_{\text{one}} = 2.0\times10^{-4}$.

Held-out factor	Best CIV	ΔLL	95% CI
Attack axis	Shared-linear	0.046	[0.032, 0.059]
Blue missiles	Shared-linear	0.042	[0.028, 0.055]
Blue shot phil.	Shared-linear	0.052	[0.038, 0.065]
Red shot phil.	Isotropic	0.132	[0.110, 0.155]
Red aircraft	Shared-linear	0.146	[0.126, 0.167]

5. Conclusion

This paper evaluated whether LQR-derived CIV summaries provide target-relevant information beyond classical Voronoi geometry in high-fidelity BVR simulations. Using a fold-safe conditional added-value test, we residualized CIV summaries against scenario and Euclidean baselines to evaluate their explanatory contribution under held-out probabilistic scoring. The results support a focused conclusion: within the evaluated ASA BVR ground attack scenario family, CIV provides consistent pre-hit explanatory added value for binary target-damage occurrence. This effect persisted under scenario-cell clustered inference, early flight-state controls, negative controls, and tactical-factor holdout validation, supporting the interpretation that the dynamics-aware partition captures engagement-relevant kinematic information not properly summarized by static geometry alone.

Beyond the specific aerospace application, the study illustrates a practical validation problem relevant to physics-informed representations: exact conditional-independence testing is often difficult in simulation settings with continu-

ous variables, temporal dependence, and structured scenario variation. Rather than claiming a universal independence result, the proposed pipeline evaluates a targeted added-value null using fold-safe residualization, paired out-of-fold losses, and dependence-aware inference. This provides a practical way to test whether a candidate representation contributes information beyond a specified baseline while limiting leakage and overconfidence.

Because the primary analysis binarizes continuous target damage, the results establish added value for damage occurrence rather than for damage severity. Extending the same fold-safe added-value framework to continuous damage regression is left for future work. CIV should therefore be understood in this setting as an explanatory spatial representation of the active, target-relevant engagement phase. The results do not establish CIV as a standalone early-warning predictor, an operational decision-support policy, or a universally superior representation across aerospace domains.

Several directions follow naturally from this validation study. Methodologically, future work should evaluate non-linear residualization models as stricter sensitivity checks and extend the framework to continuous damage, fixed-horizon forecasting, and prospective validation. Representationally, richer control-cost approximations, including nonlinear or reachability-based formulations, may capture tactical structure beyond the LQR surrogate considered here. Finally, the observed added-value signal motivates future work on evaluating whether CIV utilities can serve as feedback signals for agent behavior guidance under prospective closed-loop experiments. Overall, this work connects control-theoretic spatial representations with statistical testing and provides a practical case study for evaluating physics-informed metrics in complex multi-agent simulations.

References

Aurenhammer, F., Klein, R., and Lee, D.-T. *Voronoi diagrams and Delaunay triangulations*. World Scientific Publishing Company, 2013.

Cortes, J., Martinez, S., Karatas, T., and Bullo, F. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, April 2004.

Costa, A. N. and Ögren, P. A control-theoretic framework for voronoi-like space partitioning in multi-agent drone systems with second-order costs. In *2025 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 1049–1056. IEEE, 2025a.

Costa, A. N. and Ögren, P. Optimizing the locations of opposing teams using adversarial voronoi regions. In *2025 European Control Conference (ECC)*, pp. 1047–1054. IEEE, 2025b.

Costa, A. N., Ögren, P., and Ribeiro, C. H. C. Agent utilities over generalized voronoi regions and their gradients, 2026. URL https://arxiv.org/abs/2606.17388.

Dantas, J. P., Costa, A. N., Gomes, V. C., Kuroswiski, A. R., Medeiros, F. L., Geraldo, D., Samersla, A. R., Silva, S. R., Caetano, A. F., Ferreira, Y. D., et al. Asa: A simulation environment for evaluating military operational scenarios. In *International Conference on Computational Science and Computational Intelligence*, pp. 203–215. Springer, 2024.

Dantas, J. P. A., Costa, A. N., Medeiros, F. L. L., Geraldo, D., Maximo, M. R. O. A., and Yoneyama, T. Supervised Machine Learning for Effective Missile Launch Based on Beyond Visual Range Air Combat Simulations. In *Proceedings of the Winter Simulation Conference*, WSC '22, Singapore, December 2022a.

Dantas, J. P. A., Maximo, M. R. O. A., Costa, A. N., Geraldo, D., and Yoneyama, T. Machine Learning to Improve Situational Awareness in Beyond Visual Range Air Combat. *IEEE Latin America Transactions*, 20(8), 2022b.

Good, P. *Permutation, parametric and bootstrap tests of hypotheses*. Springer, 2005.

Lewis, F. L., Vrabie, D., and Syrmos, V. L. *Optimal control*. John Wiley & Sons, 2012.

Lipsitch, M., Tchetgen, E. T., and Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388, 2010.

Okabe, A. and Sugihara, K. *Spatial analysis along networks: statistical and computational methods*. John Wiley & Sons, 2012.

Shah, R. D. and Peters, J. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020.

Tibshirani, R. J. and Efron, B. An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57(1):1–436, 1993.