SUPERVISING THE DEPLOYED PREDICTION IN FINITE CONTEXT ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Context-adapted dynamics models deploy predictions fitted to observed support; perturbation training can instead supervise the midpoint of predictions fitted to altered supports. We ask when this supervision choice changes learning beyond allowed hyperparameter retuning. At fixed disagreement regularization, finite-step adaptation lets midpoint supervision compensate or amplify learning bias. Under smoothness assumptions, we bound prediction changes outside allowed tuning directions; a two-mode example establishes exact separation from any shared scalar penalty. Across three settings from two dynamics families, central supervision reduces error relative to midpoint supervision by 3.61–4.12% in matched comparisons and 2.52–2.71% after independent retuning under the primary noisy-support protocol. Using prediction responses to select an objective reduces equal-budget error by 1.90% relative to fixed centered transfer-noise regularization and 1.51% relative to one-probe central consistency, charging online selection but excluding reusable development calibration. Supervision placement matters for learning and objective choice under computational budgets.

1 INTRODUCTION

Contextual dynamics models fit a low-dimensional context to a short, noisy support trajectory, then use a shared representation to predict from new initial states (Kirchmeyer et al., 2022; Blanke and Lelarge, 2024; Nzoyem et al., 2025; 2026). Perturbing the support changes the fitted context and hence the deployed prediction.

This creates a choice during training: supervise the prediction fitted to observed support, or predictions fitted to perturbed supports. These are different uses of augmentation (Botev et al., 2022; Yang et al., 2023). Under squared loss, averaging two perturbed losses supervises their prediction midpoint and penalizes disagreement. We compare this *midpoint supervision* with *central supervision* of the deployed prediction, keeping the disagreement penalty fixed. Here, central denotes observed, possibly noisy support; finite adaptation uses a fixed number of executed fitting steps.

Can changing supervision do more than adjusting existing hyperparameters? The allowed adjustments, such as changing a shared penalty, define the *tuning family*. In a scalar model, moving supervision may only change effective shrinkage. With several prediction components, it can instead move those components in proportions that the allowed adjustments cannot reproduce. This distinction tells us whether to change the objective or its tuning.

A *supervision response* is the change in the learned deployed prediction induced by changing supervision. We separate it into a component explained by allowed tuning changes and a residual. The residual measures what tuning cannot locally reproduce; its alignment with prediction error gives the first-order risk effect relative to the matched tuning response. These responses guide selection among centered transfer-noise regularization (TNR), midpoint supervision, and one-probe central consistency.

Contributions. We characterize when changing supervision moves learned deployed predictions beyond a specified tuning family, holding disagreement regularization fixed. Standard objective identities and sensitivity tools yield a prediction-space criterion, a finite-neighborhood bound, and an exact two-mode example. We then use short-training responses as calibrated features for budgeted objective selection; held-out evaluation tests their practical value.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

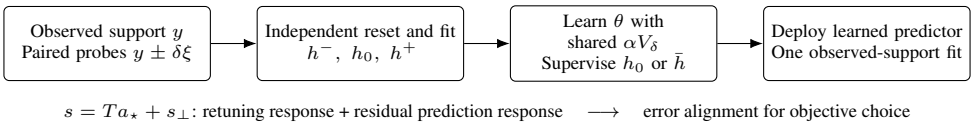

Figure 1: Supervision changes during learning; deployment uses one central fit. Projection separates the supervision response into a retuning component and a residual informing objective choice.

2 Related work

Context adaptation. Meta-learning adapts parameters or contexts (Finn et al., 2017; Zintgraf et al., 2019). Dynamics models share structure across systems (Yin et al., 2021; Li et al., 2023), with extensions for continual adaptation (Missel and Wang, 2025), physical structure (Jing et al., 2026), and latent dynamics (Vermani et al., 2025). Contextual self-modulation shares information across environment contexts (Nzoyem et al., 2026). We perturb support within one environment and compare supervision locations at a fixed adaptation interface.

Augmentation and consistency. Noise and derivative penalties encourage smoothness (Bishop, 1995; Simard et al., 1991; Rifai et al., 2011; Sajjadi et al., 2016; Miyato et al., 2019); kernel and group analyses explain averaging and invariance (Dao et al., 2019; Chen et al., 2020). Botev et al. (2022) compare loss and prediction averaging with explicit output consistency. Yang et al. (2023) compare consistency and augmented training, including tuning under misspecification; Lin et al. (2024) analyze augmentation's spectral regularization. We hold disagreement regularization fixed and ask which changes in learned deployed predictions allowed retuning can reproduce after finite adaptation. An objective identity or fixed-predictor expansion alone does not resolve this learning question. The tuning family is part of the claim: enlarging it can remove separation even when a fixed-coefficient risk gap remains.

Learning sensitivity. Optimization differentiation, influence functions, and differentiable solvers provide sensitivity tools (Domke, 2012; Franceschi et al., 2018; Koh and Liang, 2017; Pineda et al., 2022; Blondel et al., 2022). Related work studies bilevel bias (Vicol et al., 2022), unrolled gradient estimation (Vicol, 2023), and simulator unrolling (List et al., 2025). We project supervision responses against tuning responses in deployed prediction space, retain direct adapter effects, and calibrate residuals for selection.

3 Supervision as an objective intervention

3.1 The deployed predictor

An environment defines $\dot{x} = F_e(x, u)$ and support $S = \{(t_i, y_i)\}_{i=0}^{m-1}$ with all d state coordinates observed. Using a shared rollout R_θ, the adapter fits context $c \in \mathbb{R}^r$ and support origin z by reducing

$$L_S(\theta, c, z) = \frac{1}{md} \sum_i \|R_\theta(c, z; t_i, u) - y_i\|^2 + \frac{\lambda}{r} \|c\|^2, \qquad (c_K, z_K) = A_{\theta,K}(S). \tag{1}$$

Each support resets $(c_0, z_0) = (0, y_0)$. Query metadata $v = (x_0^Q, T_Q, u_Q)$ gives an independent initial state, times, and controls in the same environment. Only context transfers:

$$h_{\theta,K}(y; v) = \operatorname{vec} R_\theta(c_K, x_0^Q; T_Q, u_Q), \qquad y = \operatorname{vec}(Y_S). \tag{2}$$

Risk is $R_K(\theta) = \mathbb{E}\,\|h_{\theta,K}(y; v) - q\|^2 / n_Q$, with n_Q scored scalar outputs excluding the query origin. Labels q train θ but are unavailable to adaptation and deployment. The primary adapter uses $K = 3$ Levenberg–Marquardt (LM) attempts and $\lambda = .01$. Differentiation includes resets, solves, accepted updates, and rollout with locally fixed branches; rejected attempts consume budget. Appendix B.1 specifies LM and a branch-free gradient control.

3.2 SUPERVISION AND RESPONSE REGULARIZATION

For a symmetric probe $\xi = \Omega^{1/2}u$ with independent Rademacher coordinates in u, independently refit $y^{\pm} = y \pm \delta\xi$, $\delta > 0$. Suppressing θ, K, v, define

$$h_0 = h(y), \qquad \bar{h} = (h^+ + h^-)/2, \qquad b_\delta = \bar{h} - h_0,$$
$$V_\delta = \frac{\|h^+ - h^-\|^2}{4\delta^2 n_Q}, \qquad E_\delta = \frac{\|b_\delta\|^2}{\delta^2 n_Q}, \qquad \ell(a,q) = \frac{\|a-q\|^2}{n_Q}. \tag{3}$$

Here V_δ measures disagreement between the perturbed predictions, while b_δ measures how far their midpoint moves from the central prediction. TNR supervises the central prediction: $J_0 = \ell(h_0, q) + \alpha V_\delta$. To vary supervision and the midpoint penalty separately, define

$$J_{\eta,\kappa} = (1-\eta)\ell(h_0,q) + \eta\ell(\bar{h},q) + \alpha V_\delta + \kappa\alpha E_\delta = J_0 + \eta C_\delta + \kappa\alpha E_\delta, \tag{4}$$

where $C_\delta = \ell(\bar{h},q) - \ell(h_0,q)$, $0 \le \eta \le 1$, and $\alpha, \kappa \ge 0$. Increasing η moves supervision from central to midpoint; κ adds a midpoint-displacement penalty. Every prediction branch is differentiated.

Proposition 1 (Exact objective accounting)**.** *For arbitrary executed predictions under squared loss,*

$$\tfrac{1}{2}\{\ell(h^+,q) + \ell(h^-,q)\} = \ell(h_0,q) + \delta^2 V_\delta + C_\delta, \tag{5}$$
$$C_\delta = \{2\langle h_0 - q, b_\delta\rangle + \|b_\delta\|^2\}/n_Q, \tag{6}$$
$$\frac{\|h^+ - h_0\|^2 + \|h^- - h_0\|^2}{2\delta^2 n_Q} = V_\delta + E_\delta. \tag{7}$$

The supervised-loss mixture in Equation (4) *equals*

$$\ell(h_0 + \eta b_\delta, q) + \eta(1-\eta)\delta^2 E_\delta. \tag{8}$$

Expanding squared norms proves the identities. Paired augmentation has $\eta = 1, \kappa = 0, \alpha = \delta^2$; mixing it with central loss at weight ω gives $\eta = \omega, \alpha = \omega\delta^2$. Our intervention decouples supervision and disagreement; the tuned-augmentation baseline retains their natural coupling.

Interior mixtures add the even term in Equation (8); the direct-prediction control $\widetilde{J}_{\eta,\kappa} = \ell(h_0 + \eta b_\delta, q) + \alpha V_\delta + \kappa\alpha E_\delta$ shares their endpoints. Affine support-to-query maps have $b_\delta = 0$ and provide an exact null. For bounded symmetric probes and four bounded derivatives, $G = \partial_y h$, $H_j = \partial_y^2 h_j$ give

$$\mathbb{E}V_\delta = \operatorname{tr}(G\Omega G^\top)/n_Q + O(\delta^2), \qquad \mathbb{E}C_\delta = \delta^2 \sum_j (h_{0j} - q_j)\operatorname{tr}(H_j\Omega)/n_Q + O(\delta^4). \tag{9}$$

Sensitivity and residual–curvature alignment are distinct. Probes add within-environment support error at fixed q; $\delta^2\Omega$ is separate from central corruption. One-probe consistency (CC_1) uses

$$J_{1\mathrm{p}} = \ell(h_0,q) + \alpha\|h(y+\delta\xi) - h_0\|^2/(\delta^2 n_Q). \tag{10}$$

Objective key. TNR supervises h_0 with penalty αV_δ; **midpoint** supervises $\bar{h}$ with the same penalty; **antithetic central consistency ($\mathbf{CC}_\pm$)** supervises h_0 with $\alpha(V_\delta + E_\delta)$. Their (η, κ) values are $(0,0), (1,0), (0,1)$, respectively. $\mathbf{CC}_1$ uses one perturbed fit: symmetry gives $\mathbb{E}_\xi J_{1\mathrm{p}} = \mathbb{E}_\xi J_{0,1}$, not equality for each probe. CC_1 and optimized midpoint use two fits per training episode; TNR and $\mathrm{CC}_\pm$ use three. All deploy one central fit.

4 LEARNED RISK BEYOND SCALAR RETUNING

4.1 LEARNING BIAS AND ADAPTATION DEPTH

Let $F_{0,\delta}$ be the population central objective and $F_{1,\delta} = F_{0,\delta} + \mathbb{E}C_\delta$ its midpoint counterpart. The fixed-predictor expansion in Equation (9) becomes a learned-risk effect through the response of fitted parameters.

Proposition 2 (Learning response and radius order)**.** *Suppose $F_{0,\delta}$ has an isolated local minimizer $\theta_{0,\delta}$ with uniformly positive-definite Hessian $H_{0,\delta}$, a smooth nearby minimizer branch, and locally bounded derivatives. Assume uniformly $\mathbb{E}C_\delta = \delta^2 C_2 + O(\delta^4)$ in value and the required derivatives. Then*

$$\theta_{1,\delta} - \theta_{0,\delta} = -\delta^2 H_{0,\delta}^{-1} \nabla C_2(\theta_{0,\delta}) + O(\delta^4), \tag{11}$$

$$R(\theta_{1,\delta}) - R(\theta_{0,\delta}) = -\delta^2 \nabla R(\theta_{0,\delta})^\top H_{0,\delta}^{-1} \nabla C_2(\theta_{0,\delta}) + O(\delta^4). \tag{12}$$

If $F_{0,\delta} = R$ and θ_0 is a strict local population minimum with Hessian $H_R \succ 0$,

$$R(\theta_{1,\delta}) - R(\theta_0) = \frac{\delta^4}{2} \nabla C_2(\theta_0)^\top H_R^{-1} \nabla C_2(\theta_0) + O(\delta^6). \tag{13}$$

Regularization can leave a nonzero deployed-risk gradient, allowing midpoint supervision to compensate or amplify bias at second order in radius. This term vanishes at an unregularized population minimum. Appendix A.2 proves the result and treats finite outer optimization.

A controlled system isolates the role of adaptation depth. Let $C_\star \sim N(0,1)$, $Y = C_\star + \epsilon$, $\epsilon \sim N(0,\sigma^2)$ independently, and query target $q = C_\star + \gamma C_\star^2$. The model has vector field $(1-u)c + u\theta(c + \gamma c^2)$, with $u = 0$ on support and $u = 1$ on the unit query interval. Gradient adaptation on $(c-Y)^2/2$ from zero gives $c_K = s_K Y$, $s_K = 1 - (1-\tau)^K$, $0 < \tau < 1$. Deployment predicts $\theta f_K(Y)$, where $f_K = s_K Y + \gamma s_K^2 Y^2$; scalar probes shift its midpoint by θd_K, $d_K = \gamma s_K^2 \delta^2$.

Write $A = \mathbb{E}f_K^2$, $B = \mathbb{E}(f_K q)$, $\mu = \mathbb{E}f_K$, $m_q = \mathbb{E}q$, $D = \mathbb{E}(s_K + 2\gamma s_K^2 Y)^2$, and $r = \alpha D$.

Proposition 3 (Finite-adaptation population boundary)**.** *The central and midpoint population minimizers satisfy*

$$\theta_0 = \frac{B}{A+r}, \qquad \theta_1 = \frac{B + d m_q}{A + r + 2d\mu + d^2}, \qquad d = d_K. \tag{14}$$

For deployed risk $R(\theta) = \mathbb{E}(\theta f_K - q)^2$,

$$R(\theta_1) - R(\theta_0) = (\theta_1 - \theta_0)\{A(\theta_1 + \theta_0) - 2B\}, \tag{15}$$

$$\theta_1 - \theta_0 = \frac{d\{m_q(A+r) - B(2\mu + d)\}}{(A+r)(A + r + 2d\mu + d^2)}. \tag{16}$$

Both objectives are quadratic. For $B, r > 0$, midpoint supervision improves risk exactly when $0 < \theta_1 - \theta_0 < 2(B/A - \theta_0)$, compensating shrinkage without excessive overshoot. At $(\tau, \sigma, \gamma, \delta, \alpha) = (.3, .3, .6, .5, .1)$, changing $K = 1$ to $K = 3$ reverses the risk gap from $-.001190$ to $.005661$. Yet whenever $0 < \theta_1 \le B/A$, central training with $\alpha_{\text{eq}} = (B/\theta_1 - A)/D$ reproduces the midpoint solution exactly. A scalar fixed-penalty gap can therefore be entirely explained by retuning.

4.2 Prediction-space retuning separation

To compare supervision and tuning responses, let $F(\theta, \eta, \beta)$ be a smooth population objective with supervision parameter η and continuous tuning settings $\beta \in \mathbb{R}^p$ (nuisance parameters). At $(0, \beta_0)$ assume an isolated minimizer θ_0 with Hessian $H \succ 0$. For deployment $h(\theta, \beta; x)$, $x = (y, v)$, define $\langle a, b \rangle_P = \mathbb{E}\langle a, b \rangle / n_Q$. Parameter responses $u = -H^{-1}\partial_\eta \nabla_\theta F$ and $v_j = -H^{-1}\partial_{\beta_j} \nabla_\theta F$ induce

$$s(x) = D_\theta h(x) u, \qquad t_j(x) = D_\theta h(x) v_j + \partial_{\beta_j} h(x). \tag{17}$$

The direct derivative accounts for nuisance settings that change the deployed adapter. Put $G_{ij} = \langle t_i, t_j \rangle_P$, $z_i = \langle t_i, s \rangle_P$, $a_\star = G^\dagger z$, and $s_\perp = s - \sum_j (a_\star)_j t_j$.

Theorem 1 (Prediction-space retuning separation)**.** *Assume twice continuously differentiable solution and prediction maps with square-integrable, locally dominated derivatives. Set $h_t = h(\theta(t, \beta_0), \beta_0)$ and $g(\beta) = h(\theta(0, \beta), \beta)$. Then*

$$\inf_{a \in \mathbb{R}^p} \lim_{t \to 0} \frac{\|h_t - g(\beta_0 + ta)\|_P^2}{t^2} = \|s_\perp\|_P^2 = \|s\|_P^2 - z^\top G^\dagger z. \tag{18}$$

A positive value excludes first-order reproduction by any differentiable local retuning curve through β_0. For squared deployed risk,

$$R(h_t) - R(g(\beta_0 + ta_\star)) = 2t\langle h_0 - q, s_\perp \rangle_P + O(t^2). \tag{19}$$

Implicit differentiation and least-squares projection prove the result, including singular Gram matrices (Appendix A.3). The residual captures changes unavailable to local retuning; error alignment determines their risk effect. Retuning here matches predictions, whereas experiments select endpoints by validation risk.

The local response also gives a quantitative finite-neighborhood result. Write $Ta = \sum_j a_j t_j$, $d_{A_0} = \min_{\|a\| \le A_0} \|s - Ta\|_P$, and $M_{A_0} = (L_\eta + A_0^2 L_\beta)/2$.

Corollary 1 (Finite-neighborhood lower bound). *On the same smooth branch, suppose* $\left\|\partial_t^2 h_t\right\|_P \le L_\eta$ *for* $t \in [0, t_0]$ *and* $\left\|D_\beta^2 g\right\|_{\mathrm{op},P} \le L_\beta$ *throughout the admissible ball* $\|\beta - \beta_0\| \le A_0 t_0$. *Then, for* $0 < t \le t_0$,

$$\inf_{\|\beta - \beta_0\| \le A_0 t} \|h_t - g(\beta)\|_P \ge [t d_{A_0} - M_{A_0} t^2]_+, \qquad [x]_+ = \max\{x, 0\}. \tag{20}$$

The operator norm acts on unit Euclidean directions. Uniform Taylor remainders prove the bound (Appendix A.4). Since $d_{A_0} \ge \|s_\perp\|_P$, a positive residual yields a nonzero separation range even for tuning directions varying with t.

4.3 Two-mode incompatibility and adaptation pathways

Take two independent copies of the controlled system with shared α, mode-specific $(\tau_j, \sigma_j, \gamma_j)$, and $s_{K,j} = 1 - (1 - \tau_j)^K$.

Corollary 2 (Exact separation from a shared scalar penalty). *For* $B_j, D_j, \theta_{1,j} > 0$, *a nonnegative common penalty* α' *reproduces the midpoint deployed predictor if and only if*

$$\alpha_{\mathrm{eq},1} = \alpha_{\mathrm{eq},2} \ge 0, \qquad \alpha_{\mathrm{eq},j} = \frac{B_j/\theta_{1,j} - A_j}{D_j}. \tag{21}$$

If the coefficients differ, the infimum squared prediction discrepancy over all $\alpha' \ge 0$ *is strictly positive.*

At $K = 3, \alpha = .1, \delta = .5$ and mode triples $(.3, .3, .6)$ and $(.12, .5, .9)$, the required penalties are .118741 and .096015. Midpoint risk changes are .005661 and $-.000942$; the minimum mean prediction discrepancy is 1.4950911×10^{-4} near $\alpha' = .108108$ (Figure 2). Separate mode penalties reproduce these responses; a shared penalty cannot.

For independent query metadata, the support-to-query map $h = R \circ A$ has curvature

$$D_y^2 h[\xi, \xi] = D_c R\, D_y^2 A[\xi, \xi] + D_c^2 R[D_y A\, \xi, D_y A\, \xi]. \tag{22}$$

The first term is adapter curvature; the second is query curvature transported through fitted context. The solved model isolates the latter: even a linear adapter changes modes at different rates with depth. Neural controls retain each pathway separately.

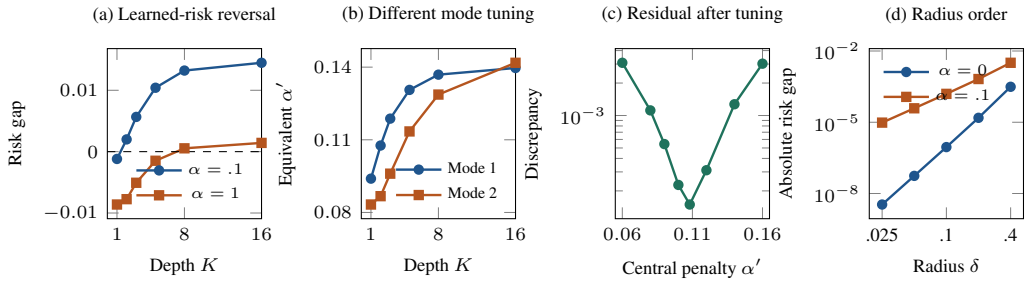

Figure 2: Exact population results (parameters: Section 4). (a) Midpoint-minus-central risk reverses with depth or regularization. (b) Modes require different midpoint-equivalent penalties. (c) At $K = 3$, common-penalty tuning leaves minimum mean prediction discrepancy 1.4951×10^{-4} near $\alpha' = .108108$. (d) Small-radius gaps approach order four without regularization, two with it. Panels (a,d) use mode 1.

5 RETUNING-AWARE OBJECTIVE SELECTION

Local projection asks which prediction changes allowed tuning can reproduce. Independent endpoint search over $(\alpha, \delta, \lambda, \text{outer rate})$ at fixed depth tests whether an error advantage survives retuning. *Online selection* chooses among development-fixed rules before deployment, without repeating that search. These comparisons answer different questions: local prediction separation does not establish the best achievable risk after a finite hyperparameter search.

Our selector, retuning-aware alignment (RAA), trains a prefix, measures responses on labeled validation episodes, predicts candidate risks, then restarts the chosen objective with the remaining budget. A learning *bank* contains training and validation environments with its own pretraining, selection, and optimizer seeds. Development banks calibrate the selector; held-out banks evaluate it. The online budget covers this selection and fitting. Deployment then adapts only to observed support, without query labels; it does not repeat labeled objective selection.

5.1 TRANSPORTED PREDICTION RESPONSES

We differentiate a short continuation of prefix training to obtain selection features. Theorem 1 motivates the projection; its smooth-population assumptions do not certify final objective rankings. For update $w_{t+1} = U_t(w_t, \eta, \beta)$, with optimizer state in w_t, sensitivity obeys

$$v_{t+1} = \partial_w U_t \, v_t + \partial_\eta U_t. \tag{23}$$

Nuisance responses follow analogous recursions (Domke, 2012; Franceschi et al., 2018); endpoint prediction derivatives include the direct adapter terms in Equation (17).

At central update 400, we take 16 unclipped outer-SGD steps with a fixed-step gradient adapter and zero initial sensitivities. On 64 labeled diagnostic episodes, we project the supervision response onto changes in $(\log \alpha, \log \lambda, \log \tau_{\text{inner}})$. A development-fixed relative Gram-eigenvalue cutoff 10^{-4} defines the retained span. With central continuation prediction $\widehat{h}$, risk $\widehat{R} = \left\|\widehat{h} - q\right\|^2_{\widehat{P}}$, and residual $\widehat{s}_\perp$, the features are

$$\mathcal{A} = \frac{2\langle \widehat{h} - q, \widehat{s} \rangle_{\widehat{P}}}{\widehat{R}}, \qquad \mathcal{A}_\perp = \frac{2\langle \widehat{h} - q, \widehat{s}_\perp \rangle_{\widehat{P}}}{\widehat{R}}, \qquad \mathcal{E}_\perp = \frac{\|\widehat{s}_\perp\|^2_{\widehat{P}}}{\|\widehat{s}\|^2_{\widehat{P}}}. \tag{24}$$

Responses point from central to midpoint supervision. Positive $\mathcal{A}$ predicts a local risk increase, and negative $\mathcal{A}$ a decrease; $\mathcal{A}_\perp$ uses the same sign convention relative to the matched retained tuning response. $\mathcal{E}_\perp$ is the response-energy fraction outside that span (zero for a zero response), not a preference for either objective. Prefix history is not differentiated. Development banks calibrate these SGD/GD features for the longer LM/Adam fit; their signs do not guarantee its endpoint ranking.

5.2 BUDGETED OBJECTIVE CHOICE

The candidates are TNR, midpoint, and CC_1, with hyperparameter rules fixed on development banks (provisional settings for author confirmation are listed in Appendix B.6). A transported κ response and a total-variance–cost feature characterize CC_1. Together with Equation (24), they augment metadata: depth, radius, noise scale, central validation error, inner residual, and gradient norm. Development-only ridge regressions predict candidate relative risk and clean-error degradation; standardization and ridge strength use leave-one-development-bank-out validation.

The selector chooses the lowest predicted-risk candidate meeting the predicted clean-error tolerance. It falls back to CC_1 if features are invalid or no candidate is predicted feasible. This preference does not certify test-time feasibility. All selectors share a twelve-minute post-prefix ceiling. After selection, fitting restarts from shared pretraining within the total 180-minute online budget. Prefix training, diagnosis, restart, and final fitting are charged; fixed objectives use the full budget for fitting. Development calibration is excluded as a reusable cost (Appendix B.6); the reported savings concern online use of an already calibrated selector.

Sampling efficiency also affects objective choice: between-episode variation can offset antithetic variance reduction. Appendix A.7 derives the total-variance–cost criterion; equal-time retraining tests its learning consequence.

6 EVALUATION

6.1 TASKS AND COMPARISONS

We use three settings: 40-state Lorenz–96 (L96) with scalar or five-factor hidden forcing, and periodic viscous Burgers with hidden transport, diffusion, and forcing. Models fit eight context coordinates to eight support observations with 5% independent Gaussian noise; query origins are independent and clean. Appendix B.2 specifies representations, horizons, and integration. Primary probes use $\Omega = .05^2 I$ in training-standardized coordinates, scaled by δ.

Each bank contains 64 training and 32 validation environments and uses two optimizer seeds. Eight primary banks share a 256-environment test cohort; four replication banks use a separate cohort. After 3,200 pretraining updates, training contexts are discarded and matched arms restart from the shared bank-specific representation for 2,400 updates.

Matched intervention holds the representation, depth, initialization, random streams, (α, δ), and outer rate fixed; midpoint retains an unused central fit to match work. **Independent retuning** allows each endpoint at most 64 configurations and 48 device-hours at $K = 3$, then fresh 2,400-update fitting. **Equal-time comparisons** allow either two hours of final fitting after independent search or 180 minutes including online objective selection. Shared pretraining is excluded equally.

Selection is calibrated on four scalar-L96 banks and evaluated on six new banks each from five-factor L96 and Burgers. Twelve conditions cross $K \in \{1, 3, 8\}$, $\delta \in \{.1, .5\}$, and central-noise scale $\{0, .05\}$: 144 conditions clustered within twelve banks. Selectors share diagnostic data, time allowances, and development-fixed candidates. Test labels and endpoint rankings remain unavailable until choices are locked.

Normalized MSE averages scored outputs in training-standardized coordinates; seeds are averaged within bank. Paired crossed bootstrap resamples banks and common test environments independently. The three primary contrasts use Bonferroni-adjusted simultaneous intervals; other intervals are pointwise 95%. Decision inference resamples whole banks within task. Failures and fallbacks remain in all-case summaries. Appendix B details data generation, adaptation, search, and inference.

6.2 SUPERVISION EFFECTS SURVIVE INDEPENDENT RETUNING

Table 1: Supervision effects and equal-time performance (normalized MSE; lower is better). A: error cells report reference / TNR; blue rows show relative gains. B: independently selected objectives. Bold marks lower paired/task means, not statistical significance.

A. Matched and independently retuned supervision

Reference / TNR	Scalar L96	Five-factor L96	Burgers
Matched midpoint	.0969 / **.0932**	.1440 / **.1388**	.0340 / **.0326**
Gain (%) [interval]	3.82 [.4,7.1]	3.61 [.2,7.0]	4.12 [.1,8.0]
Retuned midpoint	.0952 / **.0928**	.1412 / **.1376**	.0332 / **.0323**
Gain (%) [interval]	2.52 [.5,4.5]	2.55 [.3,4.8]	2.71 [.2,5.3]
Matched $CC_{\pm}$	.0935 / **.0932**	**.1384** / .1388	.0328 / **.0326**
Gain (%) [interval]	.32 [-1.0,1.6]	$-.29$ [-1.7,1.1]	.61 [-1.5,2.7]

B. Independent selection, then two-hour fitting

Objective	Scalar L96	Five-factor L96	Burgers	Updates
Tuned augmentation	.0990	.1530	.0352	2180
Optimized midpoint	.0980	.1468	.0349	2740
$CC_{\pm}$	.0959	.1444	.0337	1840
CC_1	**.0955**	**.1440**	.0339	2690
TNR	.0957	.1448	**.0335**	1890

A: gain $= 100(1 - R_{\mathrm{TNR}}/R_{\mathrm{ref}})$; matched midpoint intervals are simultaneous across tasks, others pointwise crossed 95%. Matched rows fix coefficients and work; retuned rows use independent search. B: search ceiling of 26 configurations/24 device-hours per objective; updates are for scalar L96. All deploy one central fit. TNR–CC_1 intervals span zero (Appendix Table 5).

Central supervision improves noisy error across all three settings: gains are 3.61–4.12% with matched settings and 2.52–2.71% after independent retuning (Table 1A). Most banks favor central supervision (Figure 3a). Replication retains positive mean gains, although the five-factor interval includes zero (Appendix Table 3).

TNR and $CC_{\pm}$ have small, task-dependent differences with intervals spanning zero (Table 1A). Thus the consistent distinction is supervision placement. Crossed supervision–κ controls and interior mixtures also separate this effect from the even-response penalty (Appendix Table 4).

6.3 Prediction geometry and adaptation depth

The diagnostic finds supervision changes outside its retained tuning span: .29, .34, and .26 of response energy remains in scalar L96, five-factor L96, and Burgers (Figure 3b). These fractions include direct adapter effects but concern only the short continuation, not all possible retuning.

At fixed $\alpha = .3$, $\delta = .5$, and 5% corruption, scalar-L96 smooth-adapter gains are -1.20%, 2.00%, and 3.10% at depths 1, 3, and 8 (panel c). The sign reverses without changing the data law or query horizon, as the population boundary anticipates. Per-depth paired coefficient selection instead yields positive, nonmonotone gains (Appendix Table 4B).

Episode-centered controls preserve the central prediction while linearizing either map. Adapter-only and query-only curvature retain .37 and .48 of full midpoint-displacement energy; linearizing both gives exactly zero (panel d). Cross terms and higher-order interactions prevent additive interpretation of these energies. Figure 2 separately establishes second- and fourth-order radius behavior in the exact model.

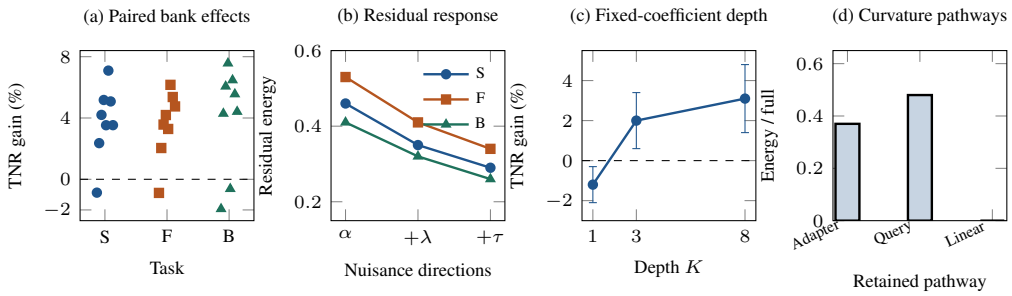

Figure 3: Mechanism controls. (a) Eight paired bank gains per task (Table 3). (b) Residual energy after successive nuisance directions. (c) Smooth-adapter depth effects at $\alpha = .3, \delta = .5$; bars: bank-based 95% intervals. (d) Isolated curvature pathways. S/F/B: scalar L96, five-factor L96, Burgers.

6.4 Objective choice under equal time

Cheaper updates matter under a time budget. With independent selection and equal final fitting time, CC_1 has the lowest mean on both L96 settings and TNR on Burgers; their paired intervals span zero (Table 1B). CC_1 completes more updates. Including between-episode variation also makes pairing's Burgers variance–cost advantage nearly neutral (Appendix Table 5).

RAA has the best sign-accuracy and regret point estimates among learned selectors (Table 2). Regret measures excess risk over the best of three completed candidates. That retrospective comparison is unconstrained, so it includes any cost of the clean-error preference.

The benefit survives charging online costs. In fresh 180-minute fits, RAA's risk ratio to always-TNR is .981 [.969,.994], and its paired gain over CC_1 is 1.51% [.1,2.9]. Each feature ablation increases mean regret, suggesting contributions from residual features, direct adapter derivatives, and variance–cost information (Figure 4a).

The pooled risk gain of 1.90% differs from the 2.05% average of task-relative gains because task reference risks are unequal (.1320 and .0330; panel d). Both summaries favor RAA.

Table 2: Objective selection on 144 conditions in twelve held-out banks. Regret uses fixed-work endpoints; risk uses fresh 180-minute fits including online overhead.

Policy	Sign BA ↑	Regret (%) ↓	Risk / TNR ↓	Risk 95% CI
Always-TNR	.50	1.92	1.000	Reference
Always-midpoint	.50	3.35	1.019	[1.008,1.031]
Always-CC_1	—	1.52	.996	[.988,1.004]
Metadata	.66	1.18	.994	[.984,1.004]
Short-run	.69	.91	.991	[.980,1.002]
Transported alignment	.75	.73	.989	[.978,1.000]
Retuning-aware (RAA)	**.79**	**.48**	**.981**	[.969,.994]

Blue marks RAA; bold marks best point estimates. BA: balanced midpoint–TNR sign accuracy. Regret uses the unconstrained retrospective oracle; risk CIs use paired whole-bank resampling. Selectors share development data, labels, and diagnostic ceilings.

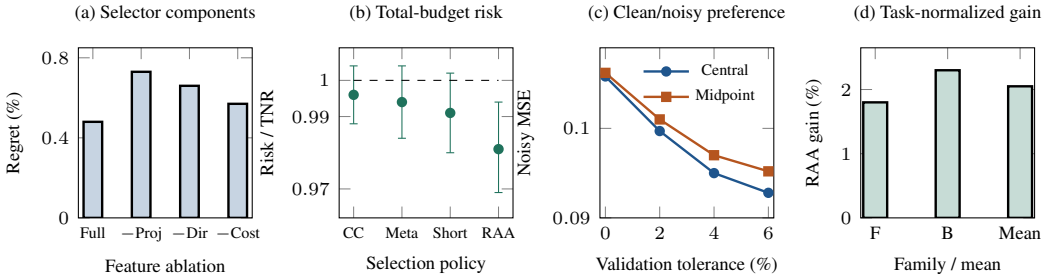

Figure 4: Decision evidence. (a) Removing projected features (Proj), direct derivatives (Dir), or variance–cost information (Cost). (b) Total-budget risk; bars: whole-bank 95% intervals. CC/Meta/Short: always-CC_1/metadata/short runs. (c) Scalar-L96 clean/noisy frontier. (d) Five-factor/Burgers relative gains and their unweighted mean; unlike (b), these are point estimates.

6.5 Observation regimes and operating trade-offs

On clean scalar L96, TNR raises error by 5.17% over unregularized training. Validation tolerances from 0% to 6% trace a clean/noisy trade-off without certifying test-time feasibility (Figure 4c). Adding query-input perturbations improves TNR over support-only perturbation for noisy origins and continuation from the observed terminal state (Appendix Table 7).

Midpoint also outperforms TNR under fourfold covariance-trace overestimation and frozen long-horizon deployment (Appendix Table 7). Central supervision's primary advantage therefore depends on the observation law and deployment horizon.

7 Discussion and Conclusion

Changing supervision can move learned predictions beyond effective shrinkage. Our local criterion, finite-neighborhood bound, and two-mode example specify when allowed tuning cannot reproduce that change (Yang et al., 2023; Lin et al., 2024).

Executed adaptation must be included when interpreting robustness (Vicol et al., 2022; List et al., 2025). Central supervision retains an advantage after independent search under the primary noisy-support protocol, while one-probe consistency benefits from cheaper updates. RAA uses response analysis to choose objectives after charging diagnostic costs (Cygert et al., 2026; Alfarra et al., 2024). Residual magnitude alone does not rank objectives: error alignment and learning cost determine its practical value.

The theory concerns smooth branches, and the diagnostic is calibrated on fully observed simulated dynamics. Partial observations, shared sensor bias, process noise, and other adapters require corresponding observation models and development coverage. Clean-input and covariance trade-offs make the deployment distribution part of objective choice.

Reproducibility statement

Appendix A gives proofs and analytical constructions. Appendix B specifies data generation, executed adaptation, learning banks, selection, budgets, and inference. Appendix C reports supporting comparisons. Code for anonymous review is available at `https://anonymous.4open.science/r/tnr-source-artifact-C290/`.

AI use statement

Generative AI assisted with manuscript proofreading, figure preparation, and LaTeX formatting. The authors are responsible for the accuracy of the scientific claims and the final research record.

References

Motasem Alfarra, Hani Itani, Alejandro Pardo, Shyma Yaser Alhuwaider, Merey Ramazanova, Juan Camilo Perez, Zhipeng Cai, Matthias Müller, and Bernard Ghanem. Evaluation of test-time adaptation under computational time constraints. In *International Conference on Machine Learning*, volume 235 of PMLR, pp. 976–991, 2024.

Christopher M. Bishop. Training with noise is equivalent to Tikhonov regularization. *Neural Computation*, 7(1):108–116, 1995.

Matthieu Blanke and Marc Lelarge. Interpretable meta-learning of physical systems. In *International Conference on Learning Representations*, 2024.

Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-Lopez, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and modular implicit differentiation. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Aleksander Botev, Matthias Bauer, and Soham De. Regularising for invariance to data augmentation improves supervised learning. *arXiv preprint arXiv:2203.03304*, 2022.

Shuxiao Chen, Edgar Dobriban, and Jane H. Lee. A group-theoretic framework for data augmentation. *Journal of Machine Learning Research*, 21(245):1–71, 2020.

Sebastian Cygert, Damian Sójka, Tomasz Trzciński, and Bartłomiej Twardowski. Realistic evaluation of test-time adaptation: Unsupervised model selection. In *International Conference on Computer Vision Theory and Applications*, volume 1, pp. 75–86, 2026.

Tri Dao, Albert Gu, Alexander Ratner, Virginia Smith, Chris De Sa, and Christopher Ré. A kernel theory of modern data augmentation. In *International Conference on Machine Learning*, volume 97 of PMLR, pp. 1528–1537, 2019.

Justin Domke. Generic methods for optimization-based modeling. In *Artificial Intelligence and Statistics*, volume 22 of PMLR, pp. 318–326, 2012.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, volume 70 of PMLR, pp. 1126–1135, 2017.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *International Conference on Machine Learning*, volume 80 of PMLR, pp. 1568–1577, 2018.

Cheng Jing, Uvini Balasuriya Mudiyanselage, Woojin Cho, Minju Jo, Anthony Gruber, and Kookjin Lee. Meta-learning structure-preserving dynamics. In *International Conference on Machine Learning*, 2026.

Matthieu Kirchmeyer, Yuan Yin, Jeremie Dona, Nicolas Baskiotis, Alain Rakotomamonjy, and Patrick Gallinari. Generalizing to new physical systems via context-informed dynamics model. In *International Conference on Machine Learning*, volume 162 of PMLR, pp. 11283–11301, 2022.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning*, volume 70 of PMLR, pp. 1885–1894, 2017.

Qiaofeng Li, Tianyi Wang, Vwani Roychowdhury, and M. Khalid Jawed. Metalearning generalizable dynamics from trajectories. *Physical Review Letters*, 131:067301, 2023.

Chi-Heng Lin, Chiraag Kaushik, Eva L. Dyer, and Vidya Muthukumar. The good, the bad and the ugly sides of data augmentation: An implicit spectral regularization perspective. *Journal of Machine Learning Research*, 25(91):1–85, 2024.

Bjoern List, Li-Wei Chen, Kartik Bali, and Nils Thuerey. Differentiability in unrolled training of neural physics simulators on transient dynamics. *Computer Methods in Applied Mechanics and Engineering*, 433:117441, 2025.

Ryan Missel and Linwei Wang. Continual slow-and-fast adaptation of latent neural dynamics (CoS-Fan): Meta-learning what-how & when to adapt. In *International Conference on Learning Representations*, 2025.

Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1979–1993, 2019.

Roussel Desmond Nzoyem, David A. W. Barton, and Tom Deakin. Neural context flows for meta-learning of dynamical systems. In *International Conference on Learning Representations*, 2025.

Roussel Desmond Nzoyem, David A. W. Barton, and Tom Deakin. Reevaluating meta-learning optimization algorithms through contextual self-modulation. In *Conference on Lifelong Learning Agents*, volume 330 of PMLR, pp. 500–524, 2026.

Luis Pineda, Taosha Fan, Maurizio Monge, Shobha Venkataraman, Paloma Sodhi, Ricky T. Q. Chen, Joseph Ortiz, Daniel DeTone, Austin Wang, Stuart Anderson, Jing Dong, Brandon Amos, and Mustafa Mukadam. Theseus: A library for differentiable nonlinear optimization. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Salah Rifai, Xavier Glorot, Yoshua Bengio, and Pascal Vincent. Adding noise to the input of a model trained with a regularized objective. *arXiv preprint arXiv:1104.3250*, 2011.

Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Patrice Simard, Bernard Victorri, Yann LeCun, and John Denker. Tangent Prop: A formalism for specifying selected invariances in an adaptive network. In *Advances in Neural Information Processing Systems*, volume 4, 1991.

Ayesha Vermani, Josue Nassar, Hyungju Jeon, Matthew Dowling, and Il Memming Park. Meta-dynamical state space models for integrative neural data analysis. In *International Conference on Learning Representations*, 2025.

Paul Vicol, Jonathan P. Lorraine, Fabian Pedregosa, David Duvenaud, and Roger B. Grosse. On implicit bias in overparameterized bilevel optimization. In *International Conference on Machine Learning*, volume 162 of PMLR, pp. 22234–22259, 2022.

Paul Vicol. Low-variance gradient estimation in unrolled computation graphs with ES-Single. In *International Conference on Machine Learning*, volume 202 of PMLR, pp. 35084–35119, 2023.

Shuo Yang, Yijun Dong, Rachel Ward, Inderjit S. Dhillon, Sujay Sanghavi, and Qi Lei. Sample efficiency of data augmentation consistency regularization. In *International Conference on Artificial Intelligence and Statistics*, volume 206 of PMLR, pp. 3825–3853, 2023.

Yuan Yin, Ibrahim Ayed, Emmanuel de Bezenac, Nicolas Baskiotis, and Patrick Gallinari. LEADS: Learning dynamical systems that generalize across environments. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Luisa Zintgraf, Kyriacos Shiarli, Vitaly Kurin, Katja Hofmann, and Shimon Whiteson. Fast context adaptation via meta-learning. In *International Conference on Machine Learning*, volume 97 of PMLR, pp. 7693–7702, 2019.

A Proofs and analytical constructions

A.1 Squared-loss identities and support expansions

Put $a = (h^+ - h^-)/2$ and $b = \bar{h} - h_0$. Averaging the squared norms of $h^\pm - q = \bar{h} - q \pm a$ cancels the cross terms, proving Equation (5) after division by n_Q. Expanding $\|h_0 + b - q\|^2 - \|h_0 - q\|^2$ proves Equation (6); applying the same cancellation to $h^\pm - h_0 = b \pm a$ proves Equation (7). Finally,

$$(1-\eta)\,\|h_0 - q\|^2 + \eta\,\|h_0 + b - q\|^2 = \|h_0 + \eta b - q\|^2 + \eta(1-\eta)\,\|b\|^2 .$$

All identities concern executed predictions and remain valid across LM branch changes.

For the smooth expansions, assume symmetric bounded ξ, covariance Ω, and four bounded derivatives of h on the probed neighborhood. Taylor expansion gives $h(y \pm \delta\xi) = h_0 \pm \delta G\xi + \delta^2 D^2 h[\xi,\xi]/2 \pm \delta^3 D^3 h[\xi,\xi,\xi]/6 + O(\delta^4)$. Thus the divided odd response is $G\xi + O(\delta^2)$ and $b = \delta^2 D^2 h[\xi,\xi]/2 + O(\delta^4)$. Taking expectations yields Equation (9). Unbounded probes require corresponding moment and domination assumptions. Conditioning on an acceptance branch does not itself preserve probe symmetry.

A.2 Population moments, radius order, and finite learning

Let $s = 1 - (1-\tau)^K$ and $v = 1 + \sigma^2$. The Gaussian moments are

$$\begin{aligned} A &= s^2 v + 3\gamma^2 s^4 v^2, & B &= s + \gamma^2 s^2 (3 + \sigma^2), \\ \mu &= \gamma s^2 v, & m_q &= \gamma, \\ D &= s^2 + 4\gamma^2 s^4 v, & Q &= \mathbb{E}q^2 = 1 + 3\gamma^2, \qquad d = \gamma s^2 \delta^2. \end{aligned} \tag{25}$$

In particular, $\mathbb{E}[Y^2 C_\star^2] = 3 + \sigma^2$, and odd joint moments vanish. The antithetic midpoint of $f_K(Y \pm \delta)$ is $f_K(Y) + d$, while its divided odd difference is $s + 2\gamma s^2 Y$. Therefore

$$F_0(\theta) = (A + r)\theta^2 - 2B\theta + Q, \qquad F_1(\theta) = (A + r + 2d\mu + d^2)\theta^2 - 2(B + d m_q)\theta + Q.$$

Their quadratic coefficients are positive for $K \geq 1$ and $0 < \tau < 1$. Differentiating proves Equation (14). Subtracting the two fractions proves Equation (16); subtracting $R(\theta) = A\theta^2 - 2B\theta + Q$ at the two endpoints gives Equation (15).

Set $M = A + r$ and $L = m_q M - 2\mu B$. Then

$$\Delta\theta = \frac{d(L - Bd)}{M(M + 2\mu d + d^2)} = \frac{L}{M^2} d + O(d^2), \qquad R'(\theta_0) = -2rB/M.$$

It follows that

$$R(\theta_1) - R(\theta_0) = -\frac{2rBL}{M^3} d + O(d^2).$$

With $r = 0$, the linear risk term vanishes and the gap is exactly $A(\Delta\theta)^2$, with expansion $(m_q A - 2\mu B)^2 d^2 / A^3 + O(d^3)$. Since $d = O(\delta^2)$, these are the stated second- and fourth-order regimes. Leading coefficients can cancel, so the unregularized order is fourth or smaller.

Proof of Proposition 2. Let $\Delta_\delta = \theta_{1,\delta} - \theta_{0,\delta}$. Smooth local dependence on the objective perturbation and the uniform inverse-Hessian bound first give $\Delta_\delta = O(\delta^2)$. Expanding perturbed stationarity at $\theta_{0,\delta}$ yields

$$0 = H_{0,\delta}\Delta_\delta + \delta^2 \nabla C_2(\theta_{0,\delta}) + O(\|\Delta_\delta\|^2 + \delta^2\,\|\Delta_\delta\| + \delta^4).$$

Solving gives Equation (11). Expanding R to second order gives Equation (12), since the displacement is $O(\delta^2)$. If $F_{0,\delta} = R$, the linear risk term is zero and the Hessian is H_R. Substituting the displacement into $\frac{1}{2}\Delta_\delta^\top H_R \Delta_\delta$ gives Equation (13); the displacement remainder contributes $O(\delta^6)$ and the bounded third-order risk remainder has the same order. The assumptions refer to smooth population objectives on the stated branch.

Finite outer optimization. For the loss-mixture path, write $A_\eta = A + r + \eta(2d\mu + d^2)$ and $B_\eta = B + \eta d m_q$. Starting from zero, T population-gradient steps of size a give

$$\theta_{\eta,T} = [1 - (1 - 2aA_\eta)^T]B_\eta/A_\eta.$$

The condition $0 < a < 1/\max_{\eta\in[0,1]} A_\eta$ ensures convergence but does not require monotone iterates. At fixed finite T, endpoint displacement is $O(d)$ and the central risk gradient need not vanish even at $r = 0$, allowing a second-order radius term. This is optimization bias, distinct from finite-sample generalization.

A.3 Prediction-space retuning separation

Let $\widetilde{h}_{t,a} = g(\beta_0 + ta)$ in Theorem 1, with expectations over the joint episode law and labels used only for risk. Since H is nonsingular, the implicit function theorem applied to $\nabla_\theta F = 0$ at $(\theta_0, 0, \beta_0)$ gives a twice differentiable local solution map when F has the corresponding third derivatives. Alternatively, the stated solution-map smoothness and second-order domination assumptions suffice.

Differentiating stationarity in η and β_j gives $Hu + \partial_\eta \nabla_\theta F = 0$ and $Hv_j + \partial_{\beta_j} \nabla_\theta F = 0$. Differentiating deployment then gives s and t_j in Equation (17). The direct term $\partial_{\beta_j} h$ cannot be omitted for nuisance parameters that act through support fitting or query rollout.

Let $T : \mathbb{R}^p \to L_2(P)$ send a to $\sum_j a_j t_j$. Then $G = T^*T$ and $z = T^*s$. If $a \in \ker G$, $\|Ta\|_P^2 = a^\top G a = 0$, so $a^\top z = \langle Ta, s\rangle_P = 0$. Thus $z \in \operatorname{range} G$, including singular cases. Completing the quadratic norm gives

$$\inf_a \|s - Ta\|_P^2 = \|s\|_P^2 - z^\top G^\dagger z, \qquad a_\star = G^\dagger z.$$

For each fixed a, dominated second derivatives give $h_t = h_0 + ts + O_{L_2}(t^2)$ and $\widetilde{h}_{t,a} = h_0 + tTa + O_{L_2}(t^2)$. The limit in Equation (18) is therefore $\|s - Ta\|_P^2$; taking its infimum proves the claim. Any differentiable nuisance curve through β_0 has a finite tangent a and the same first-order form, proving non-reproduction.

For square-integrable targets and the same domination,

$$R(h_0 + tq_1 + O_{L_2}(t^2)) = R(h_0) + 2t\langle h_0 - q, q_1\rangle_P + O(t^2).$$

Subtracting the expansions for $q_1 = s$ and $q_1 = Ta_\star$ proves Equation (19). Prediction-nearest retuning need not minimize risk, and a positive residual norm alone gives no risk sign.

Finite training counterpart. With fixed minibatches, initialization, and continuous optimizer state, differentiating $w_{t+1} = U_t(w_t, \eta, \beta)$ gives Equation (23) and the nuisance recursions. These finite-training derivatives replace stationary-solution sensitivities in the projection; Adam's moment buffers belong in the state. Appendix B.6 specifies the smooth diagnostic continuation.

A.4 Finite-neighborhood separation and constrained tuning

For Corollary 1, Taylor's integral remainder in $L_2(P)$ gives

$$h_t = h_0 + ts + r_\eta(t), \qquad g(\beta_0 + ta) = h_0 + tTa + r_\beta(t, a),$$

with $\|r_\eta(t)\|_P \le L_\eta t^2/2$ and $\|r_\beta(t,a)\|_P \le L_\beta t^2 \|a\|^2/2$. Every candidate in the radius-$A_0 t$ ball has such a representation with $\|a\| \le A_0$. The bound on the combined remainder is uniform over that compact set. Therefore

$$\|h_t - g(\beta_0 + ta)\|_P \ge t\|s - Ta\|_P - M_{A_0} t^2.$$

Nonnegativity and minimization prove Equation (20), including tuning choices whose direction depends on t. Continuity gives a minimum in the definition of d_{A_0}. Minimizing over a subset of all directions implies $d_{A_0} \ge \|s_\perp\|_P$. When $M_{A_0} = 0$, the linear lower bound holds throughout the assumed neighborhood. For the two modes in Section 4, set $\alpha_0 = .1$, $A_0 = .2$, and $t_0 = .2$, with $t = \eta$ and probe radius fixed at $\delta = .5$. The admissible interval stays within $\alpha' \ge .06$. Explicit derivatives of $\theta_j(\eta, \alpha) = (B_j + \eta d_j m_{q,j})/(A_j + \alpha D_j + \eta(2d_j\mu_j + d_j^2))$ give conservative bounds

$L_\eta \le .001632$ and $L_\beta \le 2.112$, with $d_{A_0} \simeq .0128154$. Equation (20) is nonzero throughout $0 < t \le .2$; at $t = .2$ its lower bound is approximately 8.41×10^{-4} in prediction norm.

At an active constraint such as $\alpha = 0$ with $\alpha \ge 0$, feasible first-order changes use an additive penalty coordinate and satisfy $a_\alpha \ge 0$. The prediction-nearest comparison then becomes constrained least squares, rather than the unrestricted pseudoinverse formula. A logarithmic penalty response is inactive at zero and does not represent these one-sided additive moves. The empirical diagnostic uses its specified logarithmic family and records inactivity, so its residual is relative to that retained family.

Exact projection onto a fixed nuisance span is invariant under nonsingular changes of its basis. Truncation by a relative Gram-eigenvalue threshold can depend on the basis scaling. The numerical projection is consequently defined together with its logarithmic coordinates and threshold; it approximates the stated span rather than certifying separation from all feasible hyperparameter changes.

A.5 Exact finite-radius separation in two modes

For independent C_j, ϵ_j, define a two-dimensional controlled vector field with support velocities c_j and query velocities $\theta_j(c_j + \gamma_j c_j^2)$. Diagonal gradient step sizes τ_j on the support objective give $c_{K,j} = s_{K,j} Y_j$. The outer objective averages the two scalar losses and uses a common α. Multiplying the whole objective by two leaves the minimizer unchanged, so each coordinate has Equation (14).

Because $A_j = \mathbb{E} f_j^2 > 0$, equality of deployed predictors in mean square requires equality of both coefficients. Solving $B_j/(A_j + \alpha' D_j) = \theta_{1,j}$ gives the necessary and sufficient equivalent coefficient in Equation (21). It follows that a common nonnegative coefficient exists exactly when both equivalent coefficients coincide and are nonnegative.

The prediction discrepancy from tuning a common central coefficient is

$$d_P^2(\alpha') = \frac{1}{2} \sum_{j=1}^{2} A_j \left\{ \frac{B_j}{A_j + \alpha' D_j} - \theta_{1,j} \right\}^2 .$$

It is continuous on $[0, \infty)$ and tends to $\frac{1}{2} \sum_j A_j \theta_{1,j}^2 > 0$ as $\alpha' \to \infty$. If the equivalent coefficients differ, it is positive at every finite point. A compactness argument then gives a strictly positive global infimum. For the parameters in the main text, bounded scalar minimization and direct evaluation give $\alpha' \simeq .108108$ and $d_P^2 \simeq 1.4950911 \times 10^{-4}$. The positivity is established analytically; the numerical minimizer only locates the minimum shown in Figure 2.

Separate mode penalties remove this incompatibility; more generally, a larger nuisance span can contain the supervision response in Theorem 1.

A.6 Curvature pathways and observation interfaces

For $h(y) = R(A(y))$, two applications of the chain rule give Equation (22). The reset belongs inside A; treating the initial observation as a constant drops a pathway. Episode-centered controls define

$$A_{y_0}^{\mathrm{lin}}(y') = A(y_0) + D_y A(y_0)(y' - y_0), \quad R_{c_0}^{\mathrm{lin}}(c') = R(c_0) + D_c R(c_0)(c' - c_0),$$

where y_0 denotes the current full central support vector and $c_0 = A(y_0)$ in this paragraph. The query-only control uses $R \circ A^{\mathrm{lin}}$, the adapter-only control uses $R^{\mathrm{lin}} \circ A$, and their composition is affine in y'. All coincide with the full central prediction at the anchor. The affine composition has zero midpoint displacement at every radius. At finite radius, the isolated pathway energies need not sum to the full energy because of cross terms and higher-order interactions.

For continuation, $h(y) = H(y, g(y))$ also has the query-origin pathway. Its first derivative is $G_{\mathrm{cont}} = \partial_y H + \partial_x H \, \partial_y g$. Full probes recompute the origin from the same perturbed terminal observation. For independently noisy query origins, leading variance includes $G \Sigma_y G^\top + J_x \Sigma_x J_x^\top$; correlated support/origin noise adds the two cross-covariance terms. Duplicating an independent perturbation at a shared terminal origin changes this observation law.

A.7 Antithetic gradients and total variance

At a fixed model state, write $g(e,\xi) = \nabla_\theta J_{1\mathrm{p}}$, $g_s = [g(e,\xi) + g(e,-\xi)]/2$, and $g_a = [g(e,\xi) - g(e,-\xi)]/2$. For each episode, $[J_{1\mathrm{p}}(\xi) + J_{1\mathrm{p}}(-\xi)]/2 = J_{0,1}(\xi)$. Under differentiation–expectation interchange and finite second moments, the same identity holds for gradients. Define

$$B_{\mathrm{ep}} = \operatorname{tr}\operatorname{Cov}_e\{\mathbb{E}_\xi[g \mid e]\}, \qquad W = \mathbb{E}_e \operatorname{tr}\operatorname{Cov}_\xi(g_s \mid e), \qquad Q_a = \mathbb{E}_{e,\xi}\left\|g_a\right\|^2.$$

The symmetric gradient is even in ξ and g_a is odd, so $\mathbb{E}_\xi g_a = 0$ and their cross covariance vanishes. Therefore

$$\operatorname{Cov}_\xi(g \mid e) = \operatorname{Cov}_\xi(g_s \mid e) + \mathbb{E}_\xi[g_a g_a^\top \mid e].$$

The law of total covariance gives paired and one-probe variance traces $B_{\mathrm{ep}} + W$ and $B_{\mathrm{ep}} + W + Q_a$. For independent fixed-state gradient estimates, variance under a time budget is proportional to cost times single-estimate variance. With paired and single-probe costs c_s, c_1, pairing therefore improves variance per budget exactly when

$$c_s(B_{\mathrm{ep}} + W) < c_1(B_{\mathrm{ep}} + W + Q_a). \tag{26}$$

The ordering concerns pre-clipping gradients; clipping and Adam need not preserve it for final updates.

B Systems and evaluation protocols

B.1 Executed adapters and differentiation

For $w = (c, z)$ the support residual is

$$r_S(w) = \begin{bmatrix} \operatorname{vec}(R_\theta(c, z; T_S, u) - Y_S)/\sqrt{md} \\ \sqrt{\lambda/r}\, c \end{bmatrix}.$$

Its squared norm is Equation (1), without a factor of one half. At every attempt, LM forms $J = \partial_w r_S$ and solves $(J^\top J + \mu I)\Delta = -J^\top r_S$. The predicted reduction is computed from the linearized residual $r_S + J\Delta$. A candidate is rejected if residuals or steps are nonfinite, if either reduction is nonpositive or numerically unresolved, or if the actual/predicted reduction ratio is below 10^{-4}. Acceptance multiplies damping by .5 and rejection by 4, with clamps $[10^{-8}, 10^8]$. Rejection retains (c, z) and consumes one attempt. Every central and perturbed branch resets $(c, z, \mu) = (0, y_0, .01)$ independently.

Differentiation includes residual and Jacobian derivatives in each executed solve: if $A\Delta = -g$, then $\dot\Delta = -A^{-1}(\dot g + \dot A\Delta)$. Damping multipliers and branch choices are locally fixed; Jacobians and resets remain differentiated rather than replaced by stationary-optimum sensitivities.

The smooth adapter is $w_{k+1} = w_k - \tau\nabla_w L_S$ with the same reset and fixed-step RK4. The selected-depth control searches $\tau \in \{.01, .03, .1\}$ and $K \in \{3, 5, 8\}$. The fixed-coefficient sign test fixes $\tau = .03, \lambda = .01, \alpha = .3, \delta = .5$ and varies $K \in \{1, 3, 8\}$ at a common data law and query horizon. Both objectives restart from common pretraining on eight independent scalar-L96 banks, with two seeds and a separate 256-environment cohort. Table 4B instead reports per-depth paired selection.

B.2 Data-generating systems and representations

The 40-state Lorenz–96 system is

$$\dot x_j = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F_j,$$

with cyclic indices. Scalar forcing is $F_j = F$, $F \sim \mathrm{Unif}[6, 10]$. Five-factor forcing is

$$F_j = F_0 + \sum_{k=1}^{2}\{a_k \sin(2\pi kj/40) + b_k \cos(2\pi kj/40)\},$$

with $F_0 \sim \mathrm{Unif}[7, 9]$ and other coefficients independently uniform on $[-.75, .75]$. Initial states are the forcing state plus independent Gaussian deviation .1, followed by 20 time units of burn-in. Independent trajectories provide support and query origins within each environment. Reference integration uses DOP853 at relative/absolute tolerances $10^{-9}/10^{-11}$. Support times are

$0, .05, .10, .15, .20, .25, .30, .40$; query times are every $.05$ through $.8$, excluding the supplied origin. Learned RK4 steps are at most $.005$; refinement uses $.0025$ with complete re-adaptation.

The field follows

$$\partial_t u + au\partial_x u = \nu\partial_{xx} u + b\sin(2\pi x), \qquad x \in [0, 1],$$

with periodic boundary conditions, $a \sim \mathrm{Unif}[.7, 1.3]$, $\log \nu \sim \mathrm{Unif}[\log .01, \log .05]$, and $b \sim \mathrm{Unif}[-.2, .2]$. Initial functions contain four Fourier modes with independent sine/cosine coefficients of standard deviation $.5/k^2$. Support and query functions are independent within environment. Reference integration uses 128 spatial points, Fourier derivatives with two-thirds dealiasing, and RK4 step at most 2×10^{-5}. Observations retain 64 points; eight support times span $[0, .2]$, and eight equally spaced scored query times extend through $.4$. Reference step halving and 256-point refinement are separate from learned RK4 refinement from .001 to .0005.

L96 uses the context-informed dynamics adaptation (CoDA) representation with eight context coordinates and two width-128 tanh layers (Kirchmeyer et al., 2022). The field uses $f_\theta(x, c) = f_{0,\theta}(x) + B_\theta(x)c$ with eight context coordinates and the same hidden widths. Architecture, trainable dimensions, integration steps, and normalization are fixed within every objective comparison. Training-only coordinate scales are frozen before validation or test. Damping and inner penalties are defined in these selected context coordinates.

B.3 Independent banks and primary selection

A primary bank regenerates 64 training and 32 validation environments, shared pretraining, selection, and two final optimizer seeds. Eight primary banks share a 256-environment test cohort; replication uses four new banks and another cohort. Training, validation, and test environments are disjoint. Support and query origins are independent within environment, with 32 corruption draws averaged per support/query panel. These nested observations are not independent training replicates.

Pretraining has 3,200 supervised-rollout updates using training-environment contexts; these contexts are discarded before adaptation-based fine-tuning. Outer Adam uses moments $(.9, .999)$, epsilon 10^{-8}, and norm-one clipping. Each update samples four environments, two supports per environment, and four queries per support. Initialization, minibatches, probes, and corruption streams are paired across compared objectives. The final matched arms restart from shared pretraining and stop after 2,400 updates.

The primary mechanistic grid is $\alpha \in \{0, .01, .03, .1, .3, 1\}$, $\delta \in \{.1, .25\}$, and outer rate $\{10^{-4}, 2 \times 10^{-4}\}$, with 800 updates per trial. Collapsing duplicate zero-penalty TNR radii gives 22 TNR and 24 midpoint trials. Paired selection minimizes mean endpoint validation MSE within a 6% clean-error tolerance relative to the central-only reference. If no regularized pair is feasible, both arms use the central-only fallback. A fallback remains in an all-case summary, but is not itself evidence of a nontrivial midpoint intervention. Table-specific inclusion and fallback counts require the author confirmation described in Appendix B.8. Matched midpoint retains an unused central fit to match work.

B.4 Independent endpoint retuning and competing objectives

Independent endpoint selection considers $\alpha \in \{0, .003, .01, .03, .1, .3, 1\}$, $\delta \in \{.1, .25, .5\}$, $\lambda \in \{.003, .01, .03\}$, and outer rate $\{10^{-4}, 2 \times 10^{-4}\}$ at $K = 3$. Each objective receives at most 64 distinct configurations and 48 device-hours per bank. A fixed seeded schedule covers central-only fallbacks and the primary grid before the remaining product grid, collapsing exact duplicates. Each trial has at most 800 updates and 40 minutes, including terminal validation. Independent validation selection uses the last valid terminal score and the same clean-error tolerance; nonfinite trials are ineligible and test scores are unused. Two fresh seeds then restart for 2,400 updates from shared pretraining. The clean/noisy frontier repeats selection at tolerances $0, 2, 4, 6$ percent.

The practical equal-time comparison gives each objective at most 26 configurations and 24 device-hours of search, followed by a fresh two-hour fit. It includes TNR, optimized midpoint, CC_1, $\mathrm{CC}_\pm$, tuned augmentation, full-pipeline JVP, VAT, and an output-even penalty. Compiling, pilot normalization, failures, validation, and I/O count toward search; shared pretraining is excluded equally. Timing uses one A100 80GB, float32, TF32 disabled, and eager execution.

Paired augmentation mixes central loss and average perturbed loss, coupling odd regularization and supervision shift. Its grid uses radii $\{.1, .25, .5, 1\}$, weights $\{.1, .5, 1\}$, and two outer rates, plus fallbacks. Full-pipeline JVP is $\|\partial_y h\,\xi\|^2 / n_Q$ through reset, fitting, and rollout. The output-even penalty is $\|h^+ + h^- - 2h_0\|^2 / (\delta^4 n_Q)$. Training-only pilot episodes normalize competing penalty magnitudes. VAT uses a whitened ball, one direction power step, and a stopped central reference; direction search and all backward passes count toward its budget. Every objective deploys one fit.

B.5 Projection and curvature controls

Prediction derivatives use the same 64 episodes at each retained checkpoint, with inner products averaged over all scored outputs and episodes. Gram eigenvectors below 10^{-4} times the largest eigenvalue are dropped using a development-fixed threshold. A zero response has residual-energy fraction zero. Appendix A.4 describes the distinction between exact and truncated spans.

The smooth continuation transports $\log \alpha$, $\log \lambda$, and $\log \tau_{\text{inner}}$ directions successively, including learned-parameter motion and direct adapter dependence. Residual fractions divide aggregate residual by full response energy. At zero outer penalty, the log-penalty direction is zero and a metadata indicator records inactivity. Invalid derivatives invoke the fallback in Algorithm 1.

Curvature controls use frozen primary checkpoints with complete re-adaptation. The episode-centered linearizations in Appendix A share the central prediction and probe covariance; displacement-energy ratios pool squared norms across episodes and banks. These diagnose the probe map; their energies need not add, as explained there.

B.6 Diagnostic calibration and budgeted decisions

RAA uses four scalar-L96 development banks, disjoint from primary and replication banks. Held-out five-factor L96 and Burgers each use six new banks and independent endpoint test cohorts. Depths $1, 3, 8$, probe radii $.1, .5$, and central corruption scales $0, .05$ give 12 conditions per bank. Candidate rules are fixed on development banks and selected online without further hyperparameter search, separately from the independently retuned endpoints in Table 1A.

The proposed rule uses the same $\alpha = .1$, $\lambda = .01$, and outer Adam rate 10^{-4} for TNR, midpoint, and CC_1; these values lie within the reported search grids. Depth K and radius δ follow the condition, and all candidates use the same executed LM adapter. TNR and midpoint differ only in supervision at a fixed αV_δ; CC_1 uses Equation (10). The proposed clean-error tolerance is 6% relative to an unregularized, central-supervision reference in the same bank and condition, with the same adaptation depth and fitting budget. These defaults specify a reproducible candidate family but do not establish which family produced Table 2.

All learned selectors receive a TNR prefix through update 400, 64 labeled validation episodes, and a 12-minute post-prefix ceiling. Metadata comprises K, δ, noise scale, central validation error, inner residual, and outer-gradient norm. The short-run comparator evaluates paired candidate forks at their last completed common update within the ceiling. The transported comparator adds normalized supervision and even-penalty alignments to metadata; RAA also uses residual alignment, residual-energy fraction, and total-variance–cost information.

The smooth continuation uses 16 unclipped outer-SGD steps of size 10^{-4}, fixed minibatches, the condition's depth, and inner step size .03. Sensitivities start at zero at update 400 and differentiate both continuation and deployed prediction. Alignment features follow Equation (24); the variance–cost feature uses Appendix B.7. Separate ridge regressions predict midpoint/TNR and CC_1/TNR endpoint risk differences and clean-error degradation. Standardization and ridge strength from $\{0, .01, .1, 1, 10\}$ use leave-one-development-bank-out validation. Exact-zero gaps enter regression but not binary sign scoring; other gaps are retained regardless of magnitude. Balanced sign accuracy averages the two midpoint–TNR sign-specific recalls.

For *fixed-work endpoint regret*, all three candidates have 2,400 updates and two seeds. Condition-i regret is $100\{R_i(a_i)/\min_a R_i(a) - 1\}$, averaged over conditions within bank and then banks. The retrospective oracle minimizes risk over all completed candidates without a clean-error constraint. Regret thus includes selection error and any cost of the selector's clean-error preference; neither selection nor fallback certifies test-time feasibility.

Input: shared representation checkpoint, support/query diagnostic panel, condition metadata, development-frozen candidate rules and calibration, total budget B.

1. Start the timer and train the central prefix through update 400.
2. Compute metadata and the chosen selector's features using only the diagnostic panel, under the post-prefix ceiling.
3. Predict candidate risk and clean-error degradation; choose the lowest-risk predicted-feasible candidate. Use CC_1 if features are invalid or none is feasible.
4. Lock the choice without accessing endpoint test labels. Restart the chosen final learner from shared pretraining and spend the actual remaining budget.
5. Score its central deployed prediction on the independent test cohort. Reveal archived candidate endpoints only for retrospective decision evaluation.

Algorithm 1: Budgeted objective selection with frozen calibration.

For *end-to-end risk*, a fresh 180-minute fit includes prefix training, actual diagnosis, feature construction, calibration evaluation, restart, and final fitting. Saved diagnostic time is available for final updates; fixed objectives use all 180 minutes for fitting. Shared pretraining is excluded equally, and reusable development calibration lies outside the online budget. Risk ratios pool normalized MSE across matched banks and conditions before division; CC_1 gains use the same resampled bank-condition pairs.

Separate transported-alignment study. Table 6A uses four held-out banks per family, 96 conditions, a 32-step continuation, and an eight-minute ceiling. It is not pooled with the 16-step RAA study.

B.7 Gradient variance, observations, and numerical checks

At fixed shared checkpoints, 64 episodes and 128 independent unoriented probe pairs per episode estimate full pre-clipping gradients. Conditional one-probe and paired covariance traces are computed within episode and pooled across banks. Between-episode covariance subtracts the finite-probe contribution from the covariance of episode-mean gradients; corrected estimates and any nonnegativity truncation are recorded. Total ratios combine between-episode and conditional traces. Timings match model state, batch size, and differentiation path.

Correlated support noise has covariance $\Omega_{(i,j),(i',j')} = .05^2(.8)^{|i-i'|}\mathbf{1}[j = j']$. Frozen stress keeps IID-trained parameters and probes. Calibrated models instead retrain on correlated observations. Paired sensor differences identify covariance only for conditionally independent centered errors with equal covariance; 256 training-only windows are split 192/64 to choose shrinkage, then refitted. Shared systematic sensor error is not identified by such differences. Trace/shape stress retains nominal coefficients while changing supplied probe covariance and retraining.

Noisy independent query origins have standardized scale .05 independent of support noise. Full probes perturb both pathways. Continuation uses the observed support terminal state as the query origin and recomputes it from the same perturbed support in each branch. Strongest-found adversarial evaluation uses ten projected-gradient steps, five restarts, transfer attacks, and the unchanged input, retaining the largest scored loss. These losses are empirical attacks, not certified maxima.

B.8 Uncertainty and failure accounting

Normalized MSE excludes the query origin and uses training-only coordinate scales. Corruptions and queries are averaged within support and environment, and seeds within bank. The crossed bootstrap independently resamples banks and common environments while preserving method pairing, using 20,000 resamples. Pointwise percentiles are .025, .975; simultaneous primary bounds use $.05/(2 \times 3)$ and $1 - .05/(2 \times 3)$. These approximate intervals require environment-level scores; aggregate bank means alone do not determine them.

Decision uncertainty resamples whole banks within family, retaining all conditions: the 144 conditions are clustered in twelve held-out banks. Primary means include leave-one-bank-out estimates. All-case evaluation retains failed fits, infeasible selections, and fallbacks. Nonfinite query losses receive 10^6;

finite losses are not winsorized. Missing features invoke the fixed fallback. Rejected adapter attempts, invalid diagnostic calls, nonfinite predictions, and failed fits retain their respective denominators.

C SUPPORTING COMPARISONS

Bold marks preferred point estimates; blue identifies selected summaries or methods, and amber identifies operating boundaries. Neither denotes statistical significance.

Halving the learned integration step with complete frozen-model re-adaptation changes scalar TNR MSE by .8%. Separate finer-step retraining changes the midpoint comparison gain from 3.82% to 3.49%.

Table 3: Replication and bank heterogeneity. Primary means average two seeds over common test environments; replication uses four new banks and a new cohort. TNR gains over midpoint have pointwise crossed 95% intervals. LOBO gives the pooled-mean gain range when omitting one primary bank.

	Scalar L96		Five-factor L96		Burgers	
Bank	**Midpoint**	**TNR**	**Midpoint**	**TNR**	**Midpoint**	**TNR**
1	.0918	.0926	.1350	.1362	.0310	.0316
2	.0975	.0952	.1470	.1440	.0350	.0335
3	.0952	.0912	.1400	.1350	.0330	.0310
4	.1024	.0971	.1550	.1485	.0370	.0342
5	.0907	.0875	.1370	.1325	.0320	.0322
6	.1001	.0930	.1460	.1370	.0340	.0318
7	.0984	.0934	.1490	.1410	.0360	.0340
8	.0991	.0956	.1430	.1362	.0340	.0325
Primary pooled	.0969	.0932	.1440	.1388	.0340	.0326
LOBO gain (%)	[3.33,4.45]		[3.24,4.21]		[3.57,4.90]	
Replication	.0977	.0943	.1451	.1402	.0343	.0330
Replication gain (%)	3.48 [.3,6.6]		3.38 [-.1,6.9]		3.79 [.1,7.5]	

At $\eta = .25, .5, .75$ and $\kappa = 0$, scalar loss-mixture MSEs are .0938, .0946, and .0957; direct-prediction MSEs are .0937, .0945, and .0956. Both paths start at .0932 and end at .0969. Their small interior differences agree with the even term in Equation (8), isolated by the crossed controls below.

Table 4: Response-penalty and adaptation controls. (A) Central/midpoint supervision crossed with even-response penalties at common selected coefficients. (B) Scalar retraining with per-depth paired selection; Figure 3c instead fixes coefficients across depths. Intervals are crossed 95%.

A. Even-response penalties

		κ	κ	κ
Task	**Sup.**	**0**	**.25**	**1**
Scalar L96	Central	**.0932**	**.0933**	**.0935**
	Midpoint	.0969	.0970	.0972
Five-factor L96	Central	**.1388**	**.1382**	**.1384**
	Midpoint	.1440	.1435	.1438
Burgers	Central	**.0326**	**.0325**	**.0328**
	Midpoint	.0340	.0339	.0342

B. Selected-depth controls

Adapter	K	**Mid.**	**TNR**	**Gain (%)**	**95% CI**
LM	1	.1091	.1079	1.10	[-1.5,3.7]
LM	3	.0969	.0932	3.82	[1.2,6.3]
LM	5	.0968	.0941	2.79	[.2,5.4]
LM	8	.1018	.1001	1.67	[-.8,4.2]
Smooth GD	1	.1204	.1195	.75	[-1.8,3.3]
	3	.1015	.0990	2.46	[.1,4.9]
	5	.0990	.0967	2.32	[-.1,4.8]
	8	.1025	.1010	1.46	[-1.2,4.1]
Amortized	None	.1035	.1023	1.16	[-.9,3.3]

The amortized encoder has a different parameter budget and tests nonlinearity without iterative fitting. Selected smooth-GD gains are 2.72% [.3,5.1] for scalar L96 and 2.67% [.1,5.2] for five-factor L96.

Table 5: Equal-time fitting and gradient variation. (A) Independent selection followed by two-hour fitting; all methods deploy one central fit. Updates and strongest-found attack losses are for scalar L96. (B) Matched-checkpoint variance and cost, including between-episode variation B_{ep}/W.

A. Equal-time competing objectives

Objective	Scalar	Five-factor	Burgers	Updates	Scalar attack
Tuned augmentation	.0990	.1530	.0352	2180	—
Optimized midpoint	.0980	.1468	.0349	2740	.2320
$CC_{\pm}$	.0959	.1444	.0337	1840	.2190
CC_1	**.0955**	**.1440**	.0339	2690	.2203
TNR	.0957	.1448	**.0335**	1890	.2210
Full-pipeline JVP	.0974	.1460	.0345	—	.2390
VAT	.0995	.1491	.0352	—	**.1870**
Output-even penalty	.0982	.1467	.0349	—	—

B. Variance–cost ratios

Ratio	Scalar	Five-factor	Burgers
$V_1/V_{\pm}$	1.31	1.24	1.63
$c_{\pm}/c_1$	1.42	1.45	1.37
Cond. VC	1.084	1.169	.840
B_{ep}/W	.70	.90	.60
Total VC	1.201	1.287	.983

TNR gains over CC_1 are $-.21\%$, $-.56\%$, and 1.18%, with intervals [-1.5,1.1], [-2.1,1.0], and [-1.0,3.3]. Variance–cost (VC) ratios below one favor pairing. Conditional variance-ratio intervals are [1.19,1.46], [1.12,1.40], and [1.43,1.88]. Total VC is $(c_{\pm}/c_1)(1 + B_{\mathrm{ep}}/W)/(V_1/V_{\pm} + B_{\mathrm{ep}}/W)$.

Table 6: Diagnostic cohorts and selector components. (A) Transported alignment: 96 conditions in eight banks, 32-step continuation, eight-minute ceiling. (B,C) RAA cohort from Table 2: residuals use all three retained nuisance directions; BA is sign accuracy, regret uses fixed-work endpoints, and risk uses total-budget fitting. Panel C gives pooled regret; the direct-adapter ablation omits direct adapter terms from nuisance responses.

A. Transported-alignment study

Diagnostic	Spearman ρ	95% interval	BA	95% interval
Fixed-map shift	.22	[.03,.39]	.56	[.46,.65]
Midpoint displacement	.15	[-.05,.33]	.54	[.45,.64]
One-step alignment	.39	[.20,.55]	.64	[.54,.73]
Short-run gap	.43	[.24,.60]	.65	[.55,.74]
Transported alignment	**.59**	[.40,.72]	**.76**	[.66,.84]

Midpoint wins in 28/96 conditions. The paired transported-minus-short-run BA difference is .11 with interval [.03,.19].

B. Retuning-aware selection by family

Family	Residual	BA	Regret (%)	Risk/TNR
Five-factor L96	.34	.78	.51	.982
Burgers	.26	.80	.45	.977

C. Selector ablations

Selector	Regret (%)
Full retuning-aware	**.48**
No projected residual features	.73
No direct adapter derivatives	.66
No total-variance–cost feature	.57

Higher correlation/BA and lower regret/risk are better; residual energy is descriptive. Family TNR reference MSEs are .1320 and .0330; pooling gives RAA/TNR ratio .981, distinct from the unweighted mean of family ratios despite equal bank counts.

Table 7: Observation and deployment controls. (A) Frozen and separately retrained scalar predictors; all four methods share the observation interface within each row. (B) All objectives except support-only TNR perturb the complete input and share the deployed interface.

A. Covariance, clean-input, and horizon boundaries

Condition	Augmentation	Midpoint	$CC_{\pm}$	TNR
Clean inputs	.0920	.0909	.0915	.0916
Frozen correlated observations	.1358	.1341	.1299	.1291
Retrained, oracle covariance	.1327	.1316	.1279	.1273
Retrained, paired calibration	.1341	.1330	.1300	.1295
Supplied covariance trace $\times 4$	.1410	.1414	.1423	.1420
Frozen $K = 8, H = 3.2$	1.91	1.90	1.93	1.95
Retrained $K = 3, H = 3.2$	1.69	1.68	1.68	1.69
Shuffled support	.1752	.1728	.1711	.1719

Unregularized clean MSE is .0871. Shuffled support tests retained conditioning, not physical-parameter identifiability.

B. Query-input pathways

Query interface	Aug.	Midpoint	$CC_{\pm}$	TNR support-only	TNR full
Clean independent	.0978	.0969	.0935	**.0932**	**.0932**
Noisy independent	.1260	.1243	.1210	.1241	**.1206**
Continuation	.0824	.0810	.0770	.0815	**.0767**