ERPA: EARLY PREDICTION OF ADSORBATE–METAL DISTANCE CHANGES ON OXIDE SURFACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Can the initial atomic motion, forces, and energy changes reveal how much the adsorbate–metal distance will change by the end of oxide-surface relaxation? We investigate how these early observations predict the absolute initial-to-final adsorbate–metal distance change during density functional theory (DFT) geometry optimization. Adsorbate identity, surface-metal element, material composition, surface orientation, and initial geometry also help predict this change, making the contribution of the early response the central challenge. We introduce Early Relaxation-Path Endpoint Response Analysis (ERPA) to predict this distance-change magnitude from early displacement, force, and energy descriptors, and assess their predictive value against a static-context reference and a geometry-scrambled control. Descriptors use nominal frames 5 and 10 of OC22 trajectories; shorter trajectories contribute their final stored frame. The scrambled control reassigns geometric increments within adsorbate–anchor groups in each split while retaining the recipient's initial, force, and energy descriptors. In processed held-out surface validation on the geometry-screened subset, the paired representation reduces mean absolute error (MAE) by 0.108173 relative to the static reference and by 0.115808 Å relative to geometry scrambling. Positive static-control gains extend across O, O_2, OH, HO_2, and H_2O. Force-ratio, energy-drop, and distance-change strata connect these gains to recognizable relaxation responses. A separately refitted strict pre-endpoint evaluation uses observations at indices 4 and 9 only for trajectories with at least 11 frames; on 1022 geometry-screened trajectories in processed held-out surface validation, Path reaches MAE 0.259915 Å, with gains of 0.094553 and 0.097738 Å against Static and Shuffle, respectively. Together, these comparisons show that a system's own early geometry, force, and energy response supports prediction of its final adsorbate–metal distance-change magnitude before the stored endpoint.

1 INTRODUCTION

Can the first steps of a density functional theory (DFT) geometry optimization reveal how much the adsorbate–metal distance on an oxide surface will change by the trajectory endpoint? A DFT geometry-optimization trajectory records the starting configuration and its subsequent response through atomic positions, forces, and energies (Hohenberg and Kohn, 1964; Kohn and Sham, 1965; Hjorth Larsen et al., 2017; Tran et al., 2023).

The central challenge is to distinguish what early relaxation reveals from what the starting configuration already predicts. Adsorbate identity, anchor element, composition, orientation, and initial geometry provide information about the starting configuration. Early motion provides a second source of information: how each starting configuration actually responds to optimization. We first compare response-based prediction with a static reference, then test whether geometric increments retain their value after exchange among systems with the same adsorbate and anchor categories. Keeping the recipient's initial, force, and energy descriptors in place makes the key question precise: does its own geometric response improve prediction?

AdsorbML ranks candidate adsorption configurations after learned-force relaxation, and AdsorbDiff learns adsorbate placement before learned-force relaxation and DFT evaluation (Lan et al., 2023; Kolluru and Kitchin, 2024). With ERPA, we predict the absolute initial-to-final adsorbate–metal distance

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.

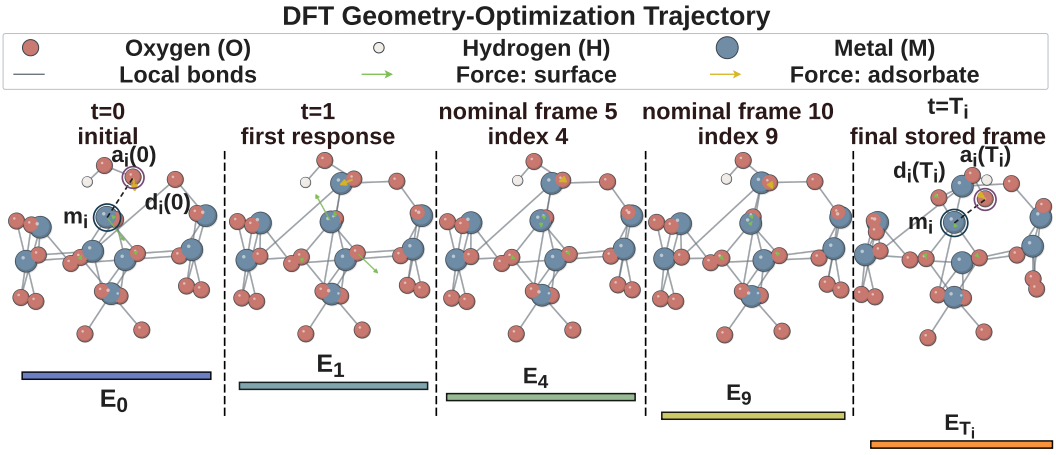

Figure 1: **Task formulation using an oxide-surface DFT relaxation trajectory.** An illustrative oxide adsorbate–surface system is shown from the initial structure ($t = 0$) through the early observed frames (nominal frames 5 and 10, at indices 4 and 9, with an illustrative intermediate frame) to the final stored state ($t = T_i$), together with atomic forces and total-energy progression. Red, ivory and blue spheres represent oxygen, hydrogen and metal atoms; gray lines show local bonds. Green and yellow arrows show projected surface-atom and adsorbate forces, respectively, using a common scale with clipped lengths. The horizontal bars schematically indicate total energy at the labeled observation indices. Dark and purple rings identify the anchor m_i and selected adsorbate atom $a_i(t)$; dashed lines join these atoms to mark $d_i(0)$ and $d_i(T_i)$. The anchor retains its initial atom index and moves with the configuration; the closest assigned adsorbate oxygen, or closest assigned adsorbate atom when oxygen is absent, is selected separately in each frame. Frames $0 : k$ supply the observed inputs, with the available-frame rule for short trajectories, while frame T_i defines the local endpoint $Y_i = |d_i(T_i) - d_i(0)|$. The task is to determine whether the early trajectory contains endpoint-relevant information beyond static chemical, surface, and initial-state information.

change from the initial state and observations at indices 4 and 9, reaching MAE 0.259915 Å on 1,022 geometry-screened held-out-surface trajectories in the strict pre-endpoint evaluation (Section 4.7). To explain this early prediction, ERPA combines controlled comparisons of relaxation-response representations with endpoint-centered interpretation of the resulting gains. The comparison measures the gain over static-context prediction and tests the value of correctly paired early geometry. We interpret gains for the same distance-change endpoint through adsorbate chemistry, physical response strata, absolute prediction errors, and trajectory-level consistency.

Figure 1 follows the local adsorbate–surface geometry from the initial structure through the observed frames to the endpoint.

2 Related Work

ERPA uses controlled representation comparisons to quantify what early relaxation contributes to final distance-change prediction. Duan et al. (2022) use evolving geometric and electronic-structure descriptors to classify whether molecular-catalyst geometry optimizations reach the intended geometry and electronic state. For oxide surfaces, we predict the absolute initial-to-final adsorbate–metal distance change as a continuous quantity, with MAE reductions of 0.103431–0.189925 Å relative to the static reference across O, O_2, OH, HO_2, and H_2O in the held-out adsorbate–anchor-resolved subset (Table 2). Matched static and geometry-scrambled controls test the value of the observed response; linear, mean, atom-set, local-graph, and physical-reference baselines measure prediction accuracy under their respective evaluation settings. Chemistry and response strata connect these comparisons to adsorption environments and physical variables. Extended modeling and statistical background appears in Appendix A.

3 METHODS

3.1 TASK AND ENDPOINT DEFINITION

We use OC22 oxide-catalyst relaxation trajectories to predict the absolute initial-to-final change in local adsorbate–metal distance (Tran et al., 2023). The supplied trajectory archive retains frames that fail the electronic-iteration or maximum-force filters used in the OC22 benchmark; our geometry filters operate on this untrimmed archive. We define the endpoint at the final stored frame of each retained trajectory.

For each trajectory i, the ERPA comparison can be written as a supervised comparison dataset with a local endpoint response:

$$\begin{aligned}\mathcal{D} &= \{(S_i, X_{i,0:k}, Y_i)\}_{i=1}^{N},\\ Y_i &= |\Delta d_{A-M,i}|\\ &= |d_{\mathrm{final}}(a_i(T_i), m_i) - d_{\mathrm{initial}}(a_i(0), m_i)|\,.\end{aligned}$$

Here S_i denotes static chemistry, surface context, and initial numerical descriptors, $X_{i,0:k}$ denotes early coordinate, force, and energy observations from the first relaxation frames, $a_i(t)$ denotes the adsorbate-side atom selected at frame t, and m_i denotes the surface metal anchor assigned from the initial structure. The primary endpoint is the absolute initial-to-final change in adsorbate–metal distance. For trajectory i, the reported descriptors use both $t_{5,i} = \min(4, T_i)$ and $t_{10,i} = \min(9, T_i)$, where $T_i = L_i - 1$, corresponding to the nominal frame-5 and frame-10 observations. For trajectories with $L_i \leq 10$, the frame-10 descriptors use the final stored frame; the frame-10 absolute distance change then equals Y_i because both use the same distance routine.

The anchor is selected from the initial structure as the nearest assigned surface metal atom associated with the adsorbate and retains its atom index across frames, while its coordinates are taken from each frame. The adsorbate-side atom is selected separately in each frame: the distance routine uses the closest assigned adsorbate oxygen when present, and otherwise the closest assigned adsorbate atom. Anchor displacement and surface-oxygen displacement provide contextual endpoints for neighboring motion; the primary adsorbate–metal distance-change endpoint directly describes local adsorption-site geometry. Coordinates, distances, displacements, MAE and ΔMAE are in Å; energies are in eV, forces in eV/Å, and force ratios are dimensionless. The extraction preserves the ASE trajectory units; training normalization applies to predictor inputs, while targets and reported errors retain their length scale.

The path context S_i^{path} contains adsorbate and anchor categories, atom count, log-transformed initial maximum force, and initial adsorbate–metal distance. Feature definitions, construction details, and subset counts are given in Appendix B.1.

Subset and observation protocols. The adsorbate-anchor resolved subset (Anch) retains successfully extracted adsorbate records with no extraction issues or assignment flags (atom-count mismatch, initial distance above 3.2 Å, or centroid below/high above the anchor). Geom additionally removes every flagged geometry-sanity record, including final adsorbate or surface-oxygen centroid displacement above 5 Å, and is nested within Anch. Tables 1–6 retain this original available-frame protocol and their original model-specific populations.

A separate strict pre-endpoint evaluation filters training, ID and Surf records to $L_i \geq 11$ and common complete numerical inputs before fitting any model or permuting geometry. Both observation indices, 4 and 9, then precede $T_i = L_i - 1$. Seven ridge representations are refitted within each of Geom and Anch, with prespecified $L_i \geq 21$ and $L_i \geq 51$ sensitivity cohorts. Appendix E.3 reports fitting details and the separate results.

3.2 ERPA

ERPA predicts the final distance-change magnitude from a system's early response and uses static and geometry-scrambled controls to assess the predictive value of that response. It compares a static predictor using S_i with a response predictor using S_i^{path} and the observed trajectory descriptors. The static reference also includes bulk composition and surface orientation. Their MAE difference measures the advantage of the response representation over this specified static reference; geometry

scrambling then tests the importance of its pairing with the recipient system:

$$\begin{aligned}\Delta_{\text{path}} &= \mathcal{R}^{\mathcal{E}}_{\text{MAE}}(f_{\theta_{\text{static}}}(S_i)) \\ &\quad - \mathcal{R}^{\mathcal{E}}_{\text{MAE}}\Big(g_{\theta_{\text{path}}}(S_i^{\text{path}}, X_{i,0:k})\Big) \\ &= \text{MAE(control)} - \text{MAE(model)}.\end{aligned}$$

The predictors are fitted on the corresponding training split and evaluated on validation set $\mathcal{E}$. A positive Δ_{path} records the response predictor's MAE advantage over the specified control. Lower MAE against both Static and Shuffle links the final distance-change magnitude to the system's early coordinate, force, and energy response through a predictive association under these controls.

We compare an early relaxation-response model, which uses early coordinate, force, and energy descriptors from the first relaxation frames, against two primary controls. The static chemistry-context control uses adsorbate symbols, surface-anchor symbols, bulk composition, surface orientation, atom count, initial maximum force, and initial adsorbate–metal distance. The within-context path-scrambling control uses the same context as the trajectory model but replaces only the paired early geometric-increment fields with a context-preserving scrambled block; the recipient's initial numerical, force, and energy descriptors remain in place.

The three predictors can be summarized as

$$\begin{aligned}\hat{Y}_i^{\text{static}} &= f_{\theta_{\text{static}}}(S_i), \\ \hat{Y}_i^{\text{path}} &= g_{\theta_{\text{path}}}(S_i^{\text{path}}, X_{i,0:k}), \\ \hat{Y}_i^{\text{shuffle}} &= g_{\theta_{\text{shuffle}}}(S_i^{\text{path}}, \tilde{X}_{i,0:k}).\end{aligned}$$

Here $\tilde{X}_{i,0:k}$ denotes the early-path representation with its geometric-increment descriptors reassigned and its force and energy descriptors retained, and θ_{shuffle} denotes the predictor fitted with the same split protocol after path scrambling. The central comparison requires $\hat{Y}^{\text{path}}$ to improve over both $\hat{Y}^{\text{static}}$ and $\hat{Y}^{\text{shuffle}}$.

The scrambled geometric-increment block is constructed by a within-group permutation,

$$\tilde{X}_i^{\text{inc}} = X_{\pi(i)}^{\text{inc}}, \qquad G_{\pi(i)} = G_i,$$

where G_i contains the split, adsorbate category, and surface-anchor element, and X^{inc} contains the early adsorbate, surface-oxygen, anchor, and adsorbate–metal geometric increment fields. Thus, the early geometry is reassigned within the matched context group while the recipient's initial, force, and energy fields remain unchanged. The permutation preserves the donor field values and may contain fixed points; singleton groups remain unchanged.

ID denotes same-distribution development validation and Surf the processed held-out surface validation; membership follows trajectory filenames in the source lists (Appendix B.1).

3.3 Row Qualification Criteria and Decision Rule

Appendix B.2 gives the qualification criteria, indicator form, and paired-error definitions.

3.4 Effective Response Coordinates

We interpret prediction gains through three early physical descriptors—force ratio, energy drop, and local distance change—termed effective response coordinates. NequIP uses equivariant representations and MACE uses higher-body-order messages to learn the energies and forces associated with atomic configurations (Batzner et al., 2022; Batatia et al., 2022). We combine DFT-observed geometric increments with energy changes and force ratios to predict the final adsorbate–metal distance-change magnitude, reducing MAE by 0.100336 Å relative to our early energy–force representation in the geometry-screened ($L_i \geq 11$) held-out-surface comparison (Table 9). The force ratio measures the remaining maximum force relative to its initial value. Energy change records the change in the scalar optimization objective. Distance change records motion in the local coordinate used to define the endpoint. Reading their strata together connects prediction gains to recognizable response states. Appendix B.3 gives the local relaxation formulation.

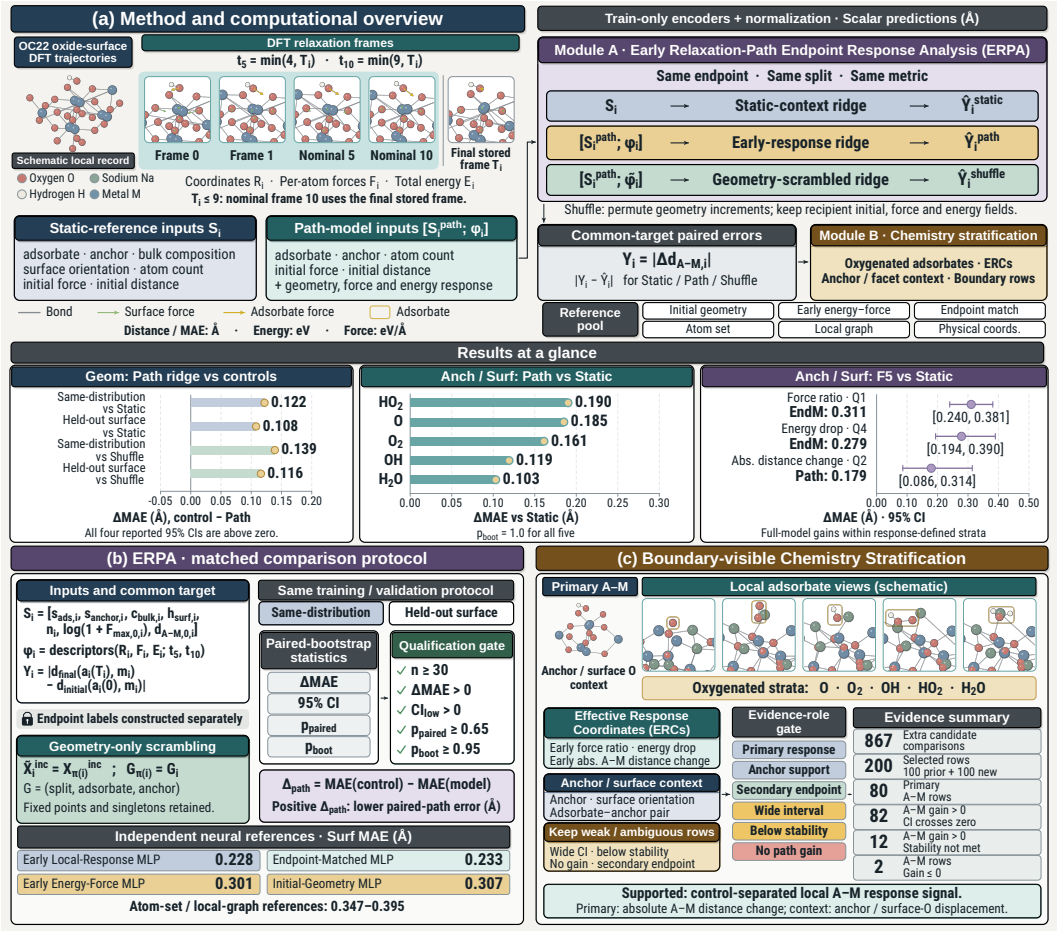

Figure 2: ERPA workflow. (a) Initial and observed coordinates, forces and energies define response descriptors at both nominal frames 5 and 10, with the available-frame fallback shown explicitly. (b) Static, Path and Shuffle ridge predictors share the distance-change endpoint and evaluation split. Shuffle reassigns only geometric increments within split, adsorbate and anchor groups, retaining recipient initial, force and energy fields. Endpoint labels are constructed from initial and final stored structures and enter training objectives and paired errors. (c) Chemistry and response strata organize the comparisons; endpoint identity and numerical qualification determine their roles. Numerical results are shown separately in Figures 3–4, Figure 5, and Tables 1–6.

3.5 Model Families, Baselines, and Implementation

The matched-control comparisons use tabular ridge regression, alongside separate PyTorch MLP and local-graph reference baselines in Table 6 (Paszke et al., 2019).

The endpoint-matched (EndM) ridge and MLP use the adsorbate and anchor categories, initial fields, both energy drops and force ratios, and only the two endpoint-specific geometric increments; Path additionally includes the other local geometric increments. For the primary endpoint, EndM uses the frame-5 and frame-10 absolute adsorbate–metal distance changes. E/F uses the initial fields and both energy drops and force ratios, with adsorbate and anchor categories. Full model and baseline specifications appear in Appendix B.4.

4 Results

Does the observed response improve prediction beyond the starting configuration, and does that improvement depend on keeping each system's geometry, forces, and energies together? We address

these questions first, then follow the gains across adsorbates and response regimes, quantify absolute endpoint error, and test prediction with observations strictly before the endpoint.

4.1 Static Controls

The observed relaxation response predicts the final distance change more accurately than the static chemistry and initial-state reference. Adsorbate identity, surface metal element, material composition, surface orientation, atom count, initial maximum force, and initial distance all carry endpoint information.

Table 1 in Appendix C reports the global control rows, and Figure 1 defines the task, inputs, endpoint, and comparison scope. Under the geometry-screened trajectory subset, the early relaxation-response model improves on the static reference by $\Delta\mathrm{MAE} = 0.122416$ for same-distribution validation and 0.108173 for Surf. The positive intervals show that these average gains persist across resampled paired errors. The next comparison tests whether the geometric increments must remain paired with their own system.

The same direction holds in the more inclusive adsorbate-anchor resolved subset, where the Surf gain against Static is 0.159056. The coordinate–force–energy response therefore improves distance-change prediction over a reference that already uses chemical, surface, and initial-state information. Relative to the corresponding Static MAE, the Surf Path and EndM reductions are approximately 29–34% across Geom and Anch.

4.2 Within-Context Path-Scrambling Controls

Does prediction benefit from this particular system's geometric response? We test this by reassigning geometric increments while retaining each recipient's initial, force, and energy descriptors.

The early relaxation-response model improves on the path-scrambling control by $\Delta\mathrm{MAE} = 0.138515$ for same-distribution validation and 0.115808 for Surf. The Surf gain remains positive in the more inclusive adsorbate-anchor resolved subset, at 0.149 (Table 1). Fisher et al. (2019) compare a predictor's losses with original and switched inputs to quantify its reliance on a group of variables. We obtain lower distance-change MAE with correctly paired responses than with a refitted control that exchanges only geometric increments within split–adsorbate–anchor groups while retaining each system's initial, force, and energy descriptors (Table 1).

Figure 3 brings together the global, adsorbate, and response-stratum comparisons.

4.3 Prediction Gains across Oxygenated Adsorbates

Gains over Static extend across O, O_2, OH, HO_2, and H_2O for the same local endpoint (Table 2); Figure 2 summarizes the ERPA comparison and its evidence roles. In the Surf adsorbate-anchor resolved subset, all five species meet the primary-evidence criteria, with $p_{\mathrm{boot}} = 1.0$ for each; their exact sample counts, gains, intervals, and p_{paired} values are retained in Table 2.

The gain over the static reference is positive for all five species, and every confidence interval in Table 2 remains above zero. The paired-response model also lowers absolute error for 66.3–69.8% of trajectories within each species. HO_2 and O have the largest absolute reductions, whereas H_2O combines a smaller gain with a comparatively narrow positive interval. The observed response thus improves distance-change prediction across all five species, with the gain's magnitude and precision varying by adsorbate.

4.4 Effective Response Coordinates

We examine prediction gains in response strata defined by quartiles of ranked descriptor values across the input cache; Table 3 evaluates the Surf trajectories belonging to those strata. For the endpoint-matched ridge model, the low force-ratio stratum has a gain of 0.310553 [0.240024, 0.381381], and the upper energy-drop quartile has a gain of 0.278650 [0.193617, 0.389668]. For the early-response ridge model, the second quartile of absolute distance change has a gain of 0.179490 [0.086074, 0.313928]. Within these response-defined strata, the full-model gains connect final distance-change

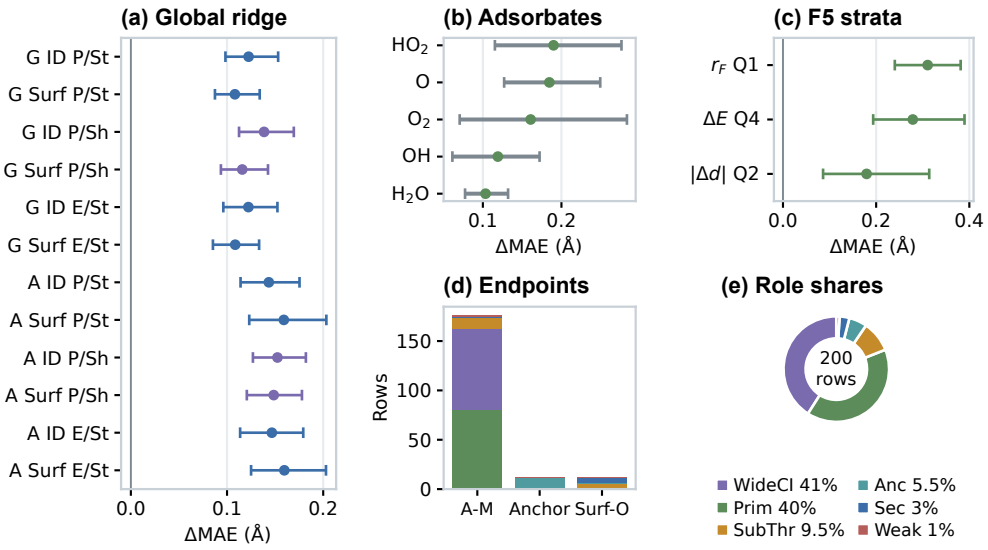

Figure 3: Core evidence under the original available-frame protocol. In (a–c), points and error bars show ΔMAE and 95% bootstrap intervals. (a) All 12 global ridge comparisons; blue denotes Static and purple Shuffle. Geom/Anch denote geometry-screened/adsorbate-anchor resolved subsets; ID/Surf denote same-distribution development/processed held-out surface validation. Path/EndM denote early-response/endpoint-matched ridge; labels give subset, split and model/control, using G/A for Geom/Anch, P/E for Path/EndM and St/Sh for Static/Shuffle. (b) Oxygenated adsorbates, Anch Surf Path/Static (Table 2). (c) Anch Surf strata at nominal frame 5 (F5): force ratio r_F Q1 and energy drop ΔE Q4 use EndM/Static; absolute distance change $|\Delta d|$ Q2 uses Path/Static; Q1–Q4 are ranked quartiles. (d–e) All 200 comparison rows by endpoint and role (Table 5). A-M, Anchor and Surf-O denote distance change, anchor displacement and surface-oxygen centroid displacement. Prim, Anc and Sec denote primary, anchor-support and secondary-endpoint roles; WideCI denotes positive gains with intervals crossing zero, SubThr other positives below stability requirements, and Weak nonpositive gains. Figures 4 and 5 share these role colors.

prediction to the remaining maximum force relative to its initial value, energy change, and motion along the target-related local coordinate.

4.5 Gain Consistency and Endpoint Roles

Prediction gains vary in mean size, interval precision, and trajectory-level consistency; we examine all three through the same paired-error analysis. Lower-confidence rows define the measured scope alongside the primary rows.

Tables 4 and 5 report the boundary rows and endpoint-level roles; Figures 4 and 5 provide the corresponding model, control, and chemistry views. The evidence tables contain 200 selected rows: 100 earlier rows and 100 expanded rows selected from an additional 867-candidate stratum-comparison pool. The remaining adsorbate–metal rows comprise 82 positive estimates with intervals crossing zero, 12 positive rows below the stability threshold, and 2 weak or negative rows.

Appendix C discusses these lower-confidence comparisons.

Figure 5 presents the chemistry comparisons and evidence roles.

4.6 Baseline Pool and Raw-Error Scale

The matched controls establish a prediction gain; the neural references quantify absolute distance-change error in processed held-out surface validation under the original available-frame evaluation. The early-response MLP reaches MAE 0.228, compared with 0.307 for initial geometry and 0.301

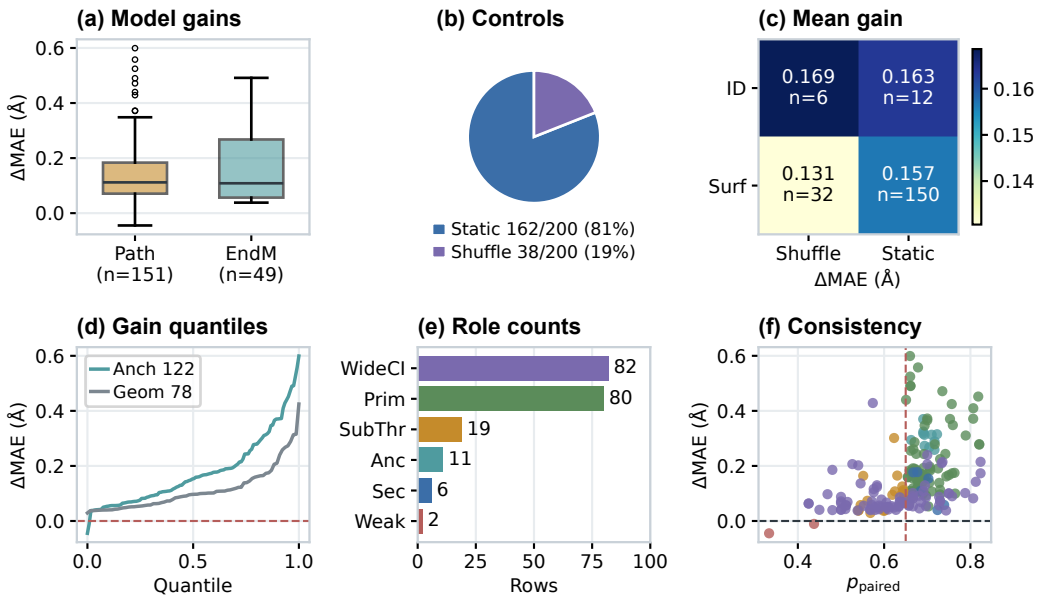

Figure 4: Model and control overview of the 200 selected comparison rows under the original available-frame protocol. Abbreviations and role colors follow Figure 3; n denotes comparison rows. (a) Unweighted gain distributions for Path (151 rows) and EndM (49): lines mark medians, boxes the 25th–75th percentiles, whiskers the most extreme observations within 1.5 interquartile ranges, and small points the remaining observations. (b) Static and Shuffle contribute 162 and 38 of the 200 rows. (c) Unweighted mean gains by split/control, with each cell's row count. (d) Sorted gains against evenly spaced quantiles for Anch (122 rows) and Geom (78). Panels (a), (c) and (d) include all endpoint families and the remaining comparison categories. (e) Complete role counts. (f) Gain versus p_{paired}, the fraction of trajectories with lower model absolute error. Dashed lines mark two components of the qualification rule; colors show the complete assigned role.

for energy–force descriptors (Table 6), reducing error by approximately 26% and 24%, respectively. The endpoint-matched MLP reaches 0.233. Atom-set and local-graph implementations in the same comparison have higher errors, placing the response representation on an absolute prediction scale.

4.7 Strict Pre-Endpoint Validation

Using observations strictly before the final stored frame, ERPA predicts the distance-change magnitude at that frame more accurately than both Static and Shuffle. In the Geom $L_i \geq 11$ cohort, train/ID/Surf contain 9758/979/1022 trajectories. Path ridge gives ID MAE 0.265847 against 0.375656 for Static and 0.395300 for Shuffle. On Surf, Path MAE is 0.259915 against 0.354468 and 0.357652, giving gains of 0.094553 [0.076246, 0.113587] and 0.097738 [0.080517, 0.115895] Å. On the same 1022 trajectories, Path also reduces MAE by 0.092802 [0.075472, 0.111932] Å relative to Static-AA, which uses the same initial fields and adsorbate/anchor categories, and by 0.100336 [0.082867, 0.117047] Å relative to E/F (Table 9). Across both subsets, both validation views and all three length cutoffs, all 24 Static/Shuffle gain intervals are positive. The paired-positive fraction measures how widely the improvement is shared across trajectories; two Static comparisons have $p_{\text{paired}} < 0.65$ (Appendix E.3). The endpoint-matched ridge reference focuses its geometric inputs on the two target-distance increments and gives lower Surf MAE than Path in all six cohorts. These refits support early prediction of the stored distance-change endpoint using observations strictly before the final stored frame. An additional input-force screen reduces the gain magnitudes while preserving positive intervals (Appendix E.3).

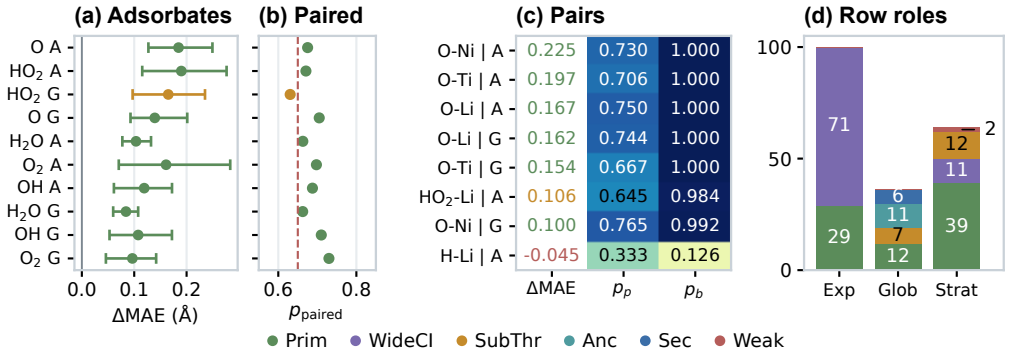

Figure 5: Chemistry-strata comparisons under the original available-frame protocol. (a–b) All ten oxygenated-adsorbate comparisons, with identical row order, show gains with 95% bootstrap intervals and p_{paired}. Each compares Path ridge with Static on Surf; labels specify the adsorbate and subset. The dashed line marks $p_{\text{paired}} = 0.65$. (c) All eight adsorbate–anchor comparisons use the same model, control and split. A and G denote Anch and Geom. Gains in Å occupy a separate column; shading encodes fractions on a common 0–1 scale. $p_p = p_{\text{paired}}$ and $p_b = p_{\text{boot}}$, the fraction of bootstrap mean gains above zero. The negative H–Li gain and below-threshold HO_2 comparisons remain visible. (d) All 200 rows by record group: Glob, Strat and Exp denote global-control, stratified-control and expanded-comparison rows. Role colors and abbreviations follow Figure 3. Figure A.1 gives the anchor and facet details.

5 Discussion

The central finding is that the final distance-change magnitude is predictable from how an adsorbed configuration begins to respond during relaxation. Initial descriptors characterize the starting configuration. Observed displacement records how that configuration moves, while forces and energies describe the accompanying response. The geometry-scrambling comparison ties this gain to the system-specific response: prediction improves when geometric increments remain paired with their own system's initial, force, and energy descriptors. ERPA uses this pairing to turn the initial relaxation response into a quantitative prediction of the final local distance-change magnitude. The strong signed-distance reference is especially informative because the endpoint itself measures local distance change. The results support a representation that combines direct geometric progress with the accompanying force and energy response, while the Surf and cross-validation references retain their respective evaluation roles.

Further comparison-record interpretation appears in Appendix D.

6 Conclusion

We introduced ERPA to predict the absolute initial-to-final adsorbate–metal distance change from early atomic motion, forces, and energy changes during oxide-surface DFT geometry optimization. The paired response lowers MAE relative to both the static reference and geometric reassignment in same-distribution development validation and held-out-surface validation. Its advantage over static descriptors extends across five oxygenated adsorbates, and response-defined strata relate the gains to recognizable changes in force, energy, and local geometry. Together, these findings identify the predictive value of keeping geometric response associated with the system from which it arises. For the geometry-screened cohort ($L_i \geq 11$), Path ridge achieves 0.259915,Å MAE on 1022 held-out-surface trajectories using indices 4 and 9, confirming that early DFT relaxation responses predict the final adsorbate–metal distance change.

AI use statement

In this work, no generative AI tools were used.

REFERENCES

A. P. Bartok, M. C. Payne, R. Kondor, and G. Csanyi, "Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons," Physical Review Letters 104, 136403 (2010). https://doi.org/10.1103/PhysRevLett.104.136403.

I. Batatia et al., "MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields," Advances in Neural Information Processing Systems 35, 11423–11436 (2022). https://doi.org/10.52202/068431-0830.

S. Batzner et al., "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," Nature Communications 13, 2453 (2022). https://doi.org/10.1038/s41467-022-29939-5.

J. Behler and M. Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," Physical Review Letters 98, 146401 (2007). https://doi.org/10.1103/PhysRevLett.98.146401.

L. Chanussot et al., "Open Catalyst 2020 (OC20) Dataset and Community Challenges," ACS Catalysis 11, 6059–6072 (2021). https://doi.org/10.1021/acscatal.0c04525.

C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, "Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals," Chemistry of Materials 31, 3564–3572 (2019). https://doi.org/10.1021/acs.chemmater.9b01294.

C. Duan, A. Nandy, H. Adamji, Y. Roman-Leshkov, and H. J. Kulik, "Machine Learning Models Predict Calculation Outcomes with the Transferability Necessary for Computational Catalysis," Journal of Chemical Theory and Computation 18(7), 4282–4292 (2022). https://doi.org/10.1021/acs.jctc.2c00331.

B. Efron, "Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics 7, 1–26 (1979). https://doi.org/10.1214/aos/1176344552.

B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC (1993). https://doi.org/10.1007/978-1-4899-4541-9.

A. Fisher, C. Rudin, and F. Dominici, "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously," Journal of Machine Learning Research 20(177), 1–81 (2019).

J. Gasteiger, J. Gross, and S. Guennemann, "Directional Message Passing for Molecular Graphs," International Conference on Learning Representations (2020). https://openreview.net/forum?id=B1eWbxStPH.

J. Gasteiger, F. Becker, and S. Guennemann, "GemNet: Universal Directional Graph Neural Networks for Molecules," Advances in Neural Information Processing Systems 34, 6790–6802 (2021). https://proceedings.neurips.cc/paper/2021/hash/35cf8659cfcb13224cbd47863a34fc58-Abstract.html.

C. R. Harris et al., "Array programming with NumPy," Nature 585, 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2.

T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer (2009). https://doi.org/10.1007/978-0-387-84858-7.

A. Hjorth Larsen et al., "The atomic simulation environment–a Python library for working with atoms," Journal of Physics: Condensed Matter 29, 273002 (2017). https://doi.org/10.1088/1361-648X/aa680e.

P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," Physical Review 136, B864–B871 (1964). https://doi.org/10.1103/PhysRev.136.B864.

J. D. Hunter, "Matplotlib: A 2D Graphics Environment," Computing in Science & Engineering 9, 90–95 (2007). https://doi.org/10.1109/MCSE.2007.55.

R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1137–1143 (1995).

W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," Physical Review 140, A1133–A1138 (1965). https://doi.org/10.1103/PhysRev.140.A1133.

A. Kolluru and J. R. Kitchin, "AdsorbDiff: Adsorbate Placement via Conditional Denoising Diffusion," Proceedings of the 41st International Conference on Machine Learning, PMLR 235, 25042–25057 (2024). https://proceedings.mlr.press/v235/kolluru24a.html.

J. Lan et al., "AdsorbML: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials," npj Computational Materials 9, 172 (2023). https://doi.org/10.1038/s41524-023-01121-5.

I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," International Conference on Learning Representations (2019). https://openreview.net/forum?id=Bkg6RiCqY7.

I. C. Man et al., "Universality in Oxygen Evolution Electrocatalysis on Oxide Surfaces," ChemCatChem 3, 1159–1165 (2011). https://doi.org/10.1002/cctc.201000397.

W. McKinney, "Data Structures for Statistical Computing in Python," Proceedings of the 9th Python in Science Conference (2010). https://doi.org/10.25080/Majora-92bf1922-00a.

A. Musaelian et al., "Learning local equivariant representations for large-scale atomistic dynamics," Nature Communications 14, 579 (2023). https://doi.org/10.1038/s41467-023-36329-y.

J. K. Norskov et al., "Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode," Journal of Physical Chemistry B 108, 17886–17892 (2004). https://doi.org/10.1021/jp047349j.

A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," Advances in Neural Information Processing Systems 32, 8024–8035 (2019). https://proceedings.neurips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

J. Rossmeisl et al., "Electrolysis of water on oxide surfaces," Journal of Electroanalytical Chemistry 607, 83–89 (2007). https://doi.org/10.1016/j.jelechem.2006.11.008.

K. T. Schuett et al., "SchNet: A deep learning architecture for molecules and materials," Journal of Chemical Physics 148, 241722 (2018). https://doi.org/10.1063/1.5019779.

K. T. Schuett, O. T. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 9377–9388 (2021). https://proceedings.mlr.press/v139/schutt21a.html.

A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," Multiscale Modeling & Simulation 14, 1153–1173 (2016). https://doi.org/10.1137/15M1054183.

M. Shuaibi et al., "Rotation Invariant Graph Neural Networks using Spin Convolutions," arXiv:2106.09575 (2021). https://doi.org/10.48550/arXiv.2106.09575.

J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost," Chemical Science 8, 3192–3203 (2017). https://doi.org/10.1039/C6SC05720A.

N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research 15, 1929–1958 (2014).

C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," BMC Bioinformatics 9, 307 (2008). https://doi.org/10.1186/1471-2105-9-307.

A. P. Thompson et al., "Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials," Journal of Computational Physics 285, 316–330 (2015). https://doi.org/10.1016/j.jcp.2014.12.018.

R. Tran et al., "The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts," ACS Catalysis 13, 3066–3084 (2023). https://doi.org/10.1021/acscatal.2c05426.

T. Xie and J. C. Grossman, "Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties," Physical Review Letters 120, 145301 (2018). https://doi.org/10.1103/PhysRevLett.120.145301.

A Extended Related Work

Large-scale atomistic datasets have made relaxed-structure and energy prediction central problems in scientific machine learning for catalysis. OC20 and OC22 define benchmark settings in which adsorbate-surface structures, relaxation trajectories, and endpoints can be studied at scale (Tran et al., 2023; Chanussot et al., 2021). Most benchmark uses of these resources emphasize final energies, relaxed structures, forces, or surrogate modeling of atomistic configurations. AdsorbML uses machine-learning potentials to identify low-energy adsorbate–surface configurations (Lan et al., 2023). AdsorbDiff predicts adsorbate positions and orientations through conditional denoising diffusion, followed by machine-learning force-field relaxation and DFT evaluation (Kolluru and Kitchin, 2024). These approaches motivate using the initial relaxation response to predict how much the local adsorbate–metal distance changes by the stored endpoint.

Geometry-aware neural networks provide the main modeling background for atomistic prediction. SchNet uses continuous-filter interactions (Schuett et al., 2018); DimeNet, GemNet, and SpinConv encode directional or angular geometry (Gasteiger et al., 2020; 2021; Shuaibi et al., 2021); NequIP, PaiNN, MACE, and Allegro use equivariant representations, with MACE adding higher-body-order messages (Batzner et al., 2022; Schuett et al., 2021; Batatia et al., 2022; Musaelian et al., 2023). Crystal and materials graph networks provide related representations for composition and structural environments (Xie and Grossman, 2018; Chen et al., 2019). These representations connect atomic positions to local interactions, motivating our examination of changes in geometry. Graph predictors provide separate reference baselines for the controlled tabular comparisons in this study.

The work is also related to classical and modern interatomic-potential modeling, including high-dimensional neural network potentials, Gaussian approximation potentials, moment tensor potentials, spectral neighbor descriptors, and ANI-style neural potentials (Behler and Parrinello, 2007; Bartok et al., 2010; Shapeev, 2016; Thompson et al., 2015; Smith et al., 2017). These methods represent local atomic environments to approximate energies, with coordinate derivatives supplying forces. Their geometry–energy–force relationship motivates examining displacement, energy change, and force response together for local endpoint prediction.

Finally, the paper intersects with statistical model comparison and confounding-control methodology. Bootstrap methods quantify variation in estimated error differences (Efron, 1979; Efron and Tibshirani, 1993), and validation protocols organize model selection and evaluation (Kohavi, 1995). For correlated predictors, the permutation scheme determines which associations are disrupted (Strobl et al., 2008). Paired errors connect the mean gain to trajectory-level consistency.

B Additional Method Details

B.1 Feature Definitions

The data object is a DFT geometry-optimization trajectory: a sequence of atomic structures produced while an adsorbed configuration relaxes toward a local minimum. In the processed OC22 trajectories used for this analysis, each usable trajectory exposes multi-frame atomic coordinates, per-frame energies, per-atom forces, adsorbate identity, surface and anchor information, and geometry at the final stored frame.

The static context and early path block are separated at the feature level:

$$
\begin{aligned}
S_i &= [s_{\mathrm{ads},i}, s_{\mathrm{anchor},i}, c_{\mathrm{bulk},i}, h_{\mathrm{surf},i}, n_i, \log(1+F_{\mathrm{max},0,i}), d_{A-M,0,i}],\\
S_i^{\mathrm{path}} &= [s_{\mathrm{ads},i}, s_{\mathrm{anchor},i}, n_i, \log(1+F_{\mathrm{max},0,i}), d_{A-M,0,i}],\\
X_{i,0:k} &= [\mathbf{R}_{i,0:k}, \mathbf{F}_{i,0:k}, E_{i,0:k}].
\end{aligned}
$$

Here $s_{\mathrm{ads},i}$ and $s_{\mathrm{anchor},i}$ denote adsorbate identity and surface-anchor element, $c_{\mathrm{bulk},i}$ and $h_{\mathrm{surf},i}$ denote material composition and surface orientation, n_i is the atom count, $F_{\mathrm{max},0,i}$ is the initial maximum per-atom force magnitude, and $d_{A-M,0,i}$ is the initial adsorbate–metal distance. The reported static chemistry-context comparator receives S_i, while the early-response predictor receives S_i^{path} and early descriptors; material-extended early-response (Path+material) and reduced-context static (Static-AA) variants are included in Tables 8 and 9. Static-AA retains the initial numerical fields and adsorbate/anchor categories.

Feature construction follows a fixed analysis protocol. Static categorical fields are encoded from training rows and applied unchanged to validation rows. Initial numerical fields include atom count, log-transformed initial maximum force, and initial adsorbate–metal distance. Early descriptors are computed from the cached frame-5 and frame-10 observations and normalized with training-split statistics. The descriptor set includes early energy change, maximum-force ratios, adsorbate-centroid displacement, anchor displacement, surface-oxygen displacement, and local adsorbate–metal distance change. Final-frame labels and row-role labels are excluded from the feature block.

The validation setting retains the processed cache assignments from the OC22 S2EF trajectory lists `train_s2ef_t`, `val_id_s2ef_t`, and `val_ood_s2ef_t` (Tran et al., 2023). In this cache, Surf bulk identifiers and composition element sets are disjoint from training, while all Surf Miller-index triples occur in training. The same-distribution split is used for model selection and its development-validation performance is reported alongside the Surf results.

Complete Path rows number 9907/999/1035 for Geom and 11382/1129/1170 for Anch in train/ID/Surf. Of these, 149/20/13 and 171/23/19, respectively, use the final stored frame for nominal frame 10. The separate five-fold trajectory-hash CV references use 489 Geom trajectories from the pilot sample and 11941 from the larger cache, pooling the source splits before assigning folds.

B.2 Row Qualification Criteria and Decision Rule

We assign each comparison a role using its endpoint, sample size, mean gain, confidence interval, trajectory-level consistency, and bootstrap stability. A fixed qualification rule distinguishes primary evidence from lower-confidence positive rows. The rule is applied uniformly across evaluated strata, linking each evidence role to the same thresholds in the tables.

A row qualifies as primary evidence for adsorbate–metal response when it has bootstrap diagnostics and satisfies sample size $n \geq 30$, positive $\Delta\mathrm{MAE}$, a positive lower bound of the 95% confidence interval, paired positive fraction at least 0.65, and bootstrap positive fraction at least 0.95.

The qualification rule referenced in Section 3.3 has the indicator form

$$\begin{aligned}\mathbb{I}_{\mathrm{primary}} =& \mathbb{I}[n \geq 30]\,\mathbb{I}[\Delta\mathrm{MAE} > 0]\,\mathbb{I}[\mathrm{CI}_{\mathrm{low}} > 0] \\ & \cdot \mathbb{I}[p_{\mathrm{paired}} \geq 0.65]\,\mathbb{I}[p_{\mathrm{boot}} \geq 0.95].\end{aligned}$$

Here $\mathbb{I}[\cdot]$ equals 1 when the condition holds and 0 otherwise. Rows with complete bootstrap diagnostics can enter the primary evidence role when their endpoint is adsorbate–metal distance change. Qualified anchor and surface-oxygen rows retain supporting and secondary-endpoint roles, respectively; positive rows with intervals crossing zero are classified before the remaining below-stability positives, and nonpositive gains receive the weak role. Let $\delta_i = e_i^{\mathrm{control}} - e_i^{\mathrm{path}}$, where $e_i^m = |Y_i - \hat{Y}_i^m|$. Its mean is $\Delta\mathrm{MAE}$, the fraction with $\delta_i > 0$ is p_{paired}, and the fraction of bootstrap means above zero is p_{boot}: these measure average improvement, trajectory-level consistency, and resampling stability, respectively.

B.3 Effective Response Coordinates

Here $\Delta E_{0\to k} = E_0 - E_k$ is the early energy drop, and $F_{\max,t} = \max_j \|\mathbf{F}_{j,t}\|_2$ is the maximum per-atom force magnitude at frame t. The ratio uses the observed-frame maximum divided by the initial maximum; values with missing forces or $F_{\max,0} < 10^{-12}$ are treated as missing.

The formulation combines measured early response coordinates with a local description of relaxation:

$$\begin{aligned}r_{\mathrm{early}} =& \big[\Delta E_{0\to k}, F_{\max,k}/F_{\max,0}, |\Delta d_{A-M}(k)|, \\ & \|\Delta\mathbf{r}_{\mathrm{ads}}(k)\|_2, \|\Delta\mathbf{r}_{\mathrm{anchor}}(k)\|_2\big], \\ \mathbf{R}_{t+1} =& \mathbf{R}_t + \eta_t \Phi_t(\mathbf{F}_{0:t}, \mathbf{R}_{0:t}, q_t), \\ \mathbf{F}_t =& -\nabla E(\mathbf{R}_t).\end{aligned}$$

The displacement vectors describe adsorbate-centroid and anchor motion, alongside surface-oxygen-centroid motion in the full descriptor set. The update expression is descriptive: real DFT geometry optimizers can depend on force history $\mathbf{F}_{0:t}$, coordinate history, line search, constraints, and optimizer state q_t. Near a local quadratic approximation in the movable-coordinate subspace,

$$\mathbf{F}_0 \approx -\mathbf{H}(\mathbf{R}_0 - \mathbf{R}^*),$$

where $\mathbf{R}^*$ is a nearby stationary configuration and $\mathbf{H}$ the energy Hessian in the movable-coordinate subspace. Here η_t is the step scale, Φ_t the optimizer-dependent update direction, and t indexes geometry-optimization steps. Within this local approximation, the initial force depends on both the displacement from the stationary configuration and the local energy curvature; observed displacement records the optimizer's actual motion, and energy change records the change in its scalar objective. This relation motivates combining observed motion with force and energy changes to predict the final distance-change magnitude. The coordinates give a physical description of early geometry, force, and energy changes in the processed relaxation trajectories.

B.4 Model Families, Baselines, and Implementation

Ridge regression uses an unpenalized bias term, training-split normalization, and same-distribution validation RMSE to select regularization from $\{0.01, 0.1, 1, 10, 100, 1000\}$ (Hastie et al., 2009; Kohavi, 1995).

The tabular MLP uses trajectory features, mean squared error, and AdamW optimization (Paszke et al., 2019; Loshchilov and Hutter, 2019), with checkpoint selection by same-distribution validation MAE; its two hidden layers have widths 128 and 64, ReLU activations, and dropout 0.1 (Srivastava et al., 2014). The PyTorch atom-set and local-graph references provide alternative representations of the same endpoint prediction task: the atom-set model mean-pools atom embeddings, while the original local-graph model uses 12-nearest-neighbor graphs, two message-passing layers of width 96, and mean pooling. Equivariant atomistic representations and interatomic-potential models provide the broader modeling background (Batzner et al., 2022; Schuett et al., 2021; Batatia et al., 2022; Musaelian et al., 2023; Xie and Grossman, 2018; Chen et al., 2019; Behler and Parrinello, 2007; Bartok et al., 2010; Shapeev, 2016; Thompson et al., 2015; Smith et al., 2017).

The neural comparison also includes initial-geometry MLP, early energy-force MLP, endpoint-matched MLP, atom-set networks at the initial frame and at frames 5 and 10, and local-graph networks at the initial frame and at frames 5 and 10. These predictors place the paired early response beside simpler feature sets and neural references.

The baseline pool contains mean controls, linear references, neural tabular references, atom-set references, local-graph references, and additional physical-reference controls. The physical-reference controls include surface anchor element, surface facet, initial coordination, oxygen-neighbor shell, adsorbate-context group means, initial force magnitude bins, initial adsorbate–metal distance bins, frame-5 and frame-10 energy-drop bins, force-ratio bins, adsorbate-motion bins, anchor-motion bins, surface-oxygen-motion bins, and interpretable ridge rows over early geometry, force, energy, and response-proxy features. The tabular ridge and MLP fits remove rows missing the target or a required numerical field. Categorical levels are learned from training counts; rare or missing categories are pooled when that level exists, and otherwise unseen validation categories receive the all-zero reference encoding. Retained levels require at least 20 training records for the matched-control ridge fits and 8 for the MLPs.

The physical group-mean references fit endpoint means on each training fold; unseen or undersized groups fall back through successively shorter prefixes of the grouping key and then to the training-fold global mean. Binned means estimate quantile edges and bin means from training-fold rows with both the coordinate and target available, falling back to their global mean for undersized bins; evaluation likewise requires both fields. Pilot group means use an adsorbate mean and then the global mean as fallbacks. Pilot ridge fits use $\alpha = 10$; retained categories and mean groups require at least four training records. The pilot and larger-sample CV scores pool held-out predictions across their five trajectory-hash folds.

C Extended Results

ERPA tests whether the early response lowers distance-change MAE against a static-context reference and a within-context geometry-scrambling control under predefined same-distribution development and processed held-out surface validation. The static reference quantifies endpoint information available from chemistry and the initial state. The scrambling control reassigns early geometric increments within adsorbate–anchor groups while retaining each recipient's initial, force, and energy descriptors. The geometry-screened comparison in processed held-out surface validation gives MAE

Table 1: Values are rounded to three decimal places. Global controls for the primary adsorbate–metal endpoint. Positive gains against static and shuffled-path controls quantify information added by correctly paired early trajectories. Path and EndM denote early local-response and endpoint-matched ridge models; Static includes initial numerical fields and adsorbate, anchor, composition, and facet categories, and Shuffle reassigns early geometric increments. Geom denotes the geometry-screened subset, Anch the adsorbate-anchor resolved subset, ID the same-distribution split, and Surf the held-out surface split.

Subset	Split	Model	Control	n	ΔMAE (Å)	95% CI
Geom	ID	Path	Static	999	0.122	[0.098,0.153]
Geom	Surf	Path	Static	1035	0.108	[0.087,0.134]
Geom	ID	Path	Shuffle	999	0.139	[0.112,0.169]
Geom	Surf	Path	Shuffle	1035	0.116	[0.094,0.143]
Geom	ID	EndM	Static	999	0.122	[0.096,0.152]
Geom	Surf	EndM	Static	1035	0.108	[0.085,0.133]
Anch	ID	Path	Static	1129	0.143	[0.114,0.175]
Anch	Surf	Path	Static	1170	0.159	[0.123,0.203]
Anch	ID	Path	Shuffle	1129	0.152	[0.127,0.182]
Anch	Surf	Path	Shuffle	1170	0.149	[0.121,0.178]
Anch	ID	EndM	Static	1129	0.147	[0.114,0.179]
Anch	Surf	EndM	Static	1170	0.160	[0.125,0.203]

Table 2: Oxygenated-adsorbate rows for the primary endpoint in the held-out adsorbate-anchor resolved subset, comparing early local-response ridge with the static chemistry-context control. p_{paired} is the fraction of trajectories with lower absolute error for Path. Positive intervals across all five species connect the early-path advantage to chemically resolved adsorption-site relaxation.

Oxygenated adsorbate	n	ΔMAE (Å)	95% CI	p_{paired}
HO_2	161	0.189925	[0.115303, 0.276389]	0.671
O	394	0.184616	[0.127079, 0.249360]	0.675
O_2	86	0.160829	[0.070499, 0.283338]	0.698
OH	160	0.119076	[0.061202, 0.172163]	0.688
H_2O	175	0.103431	[0.077333, 0.132082]	0.663

gains of 0.108173 over the static reference and 0.115808 Å over geometry scrambling. The advantage over static descriptors also appears in all five oxygenated-adsorbate categories. Response-defined strata connect the gains to force ratio, energy change, and motion along the local distance coordinate.

The global prediction gain leads to the next question: how does the response's predictive value vary across adsorbates and physically defined response regimes? Lower-confidence rows and secondary endpoints describe the precision, consistency, and local geometric scope of the evidence, while the primary endpoint remains the absolute adsorbate–metal distance change.

C.1 Prediction Gains across Oxygenated Adsorbates

The global gain raises a chemical question: does the response representation remain useful across different oxygenated adsorbates? Oxygen-containing surface species provide a familiar organizing context in oxide electrocatalysis, where their adsorption energetics are central to mechanistic analysis (Norskov et al., 2004; Man et al., 2011; Rossmeisl et al., 2007). We examine O, O_2, OH, HO_2, and H_2O as separate strata of the same local distance-change endpoint.

Table 3: Effective response-coordinate strata in the held-out adsorbate-anchor resolved subset. Force-ratio, energy-drop, and local distance motion describe gains of the full early local-response (Path) or endpoint-matched (EndM) ridge model over Static within response quartiles (Q1–Q4).

Coord.	Range	n	Model	Control	ΔMAE (Å)	95% CI
F5 $\lvert\Delta d_{A-M}\rvert$	Q2	306	Path	Static	0.179490	[0.086074, 0.313928]
F5 ΔE	Q4	368	EndM	Static	0.278650	[0.193617, 0.389668]
F5 force ratio	Q1	275	EndM	Static	0.310553	[0.240024, 0.381381]

C.2 Effective Response Coordinates

The response can also be read through early coordinate, force, and energy descriptors, which are standard objects in atomistic geometric learning (Schuett et al., 2018; Gasteiger et al., 2020; 2021; Shuaibi et al., 2021; Batzner et al., 2022). After the control and oxygenated-adsorbate comparisons establish where the signal is visible, effective response coordinates identify the response regimes in which it can be interpreted. Each reported gain remains a joint model comparison within its response stratum.

The direct distance-change reference then tests how accurately a simple local geometric summary predicts the final distance-change magnitude. The larger-sample physical-reference CV row reports MAE 0.273 and $R^2 = 0.639$ under five-fold trajectory-hash cross-validation; Surf neural references provide the separate cross-model comparison.

Table 4: Surface-oxygen displacement comparisons. All eight original rows are retained; gains and intervals are in Å. Sub-thr rows have $p_{\text{paired}} < 0.65$; qualified rows retain their secondary-endpoint role.

Subset	Split	Model	Control	n	ΔMAE	95% CI	p_{paired}	p_{boot}	Role
Geom	ID	Path	Static	999	0.037030	[0.022122, 0.051513]	0.599	1.000	Sub-thr
Geom	Surf	Path	Static	1035	0.037603	[0.021921, 0.053029]	0.540	1.000	Sub-thr
Geom	ID	Path	Shuffle	999	0.039811	[0.023420, 0.055220]	0.724	1.000	Secondary
Geom	Surf	Path	Shuffle	1035	0.058100	[0.045318, 0.071668]	0.739	1.000	Secondary
Geom	ID	EndM	Static	999	0.038210	[0.022157, 0.052748]	0.602	1.000	Sub-thr
Geom	Surf	EndM	Static	1035	0.038922	[0.024808, 0.054626]	0.541	1.000	Sub-thr
Anch	ID	Path	Static	1129	0.095038	[0.054776, 0.134855]	0.640	1.000	Sub-thr
Anch	Surf	Path	Static	1170	0.176622	[0.129212, 0.236522]	0.668	1.000	Secondary

Table 5: Evidence-role counts by endpoint family in the 200 selected rows. The hierarchy concentrates primary evidence in adsorbate–metal response and assigns anchor and surface-oxygen motion supporting or secondary roles. Prim denotes primary evidence, Wide-CI intervals crossing zero, and Sub-thr below-stability positive rows.

Endpoint	Role	Rows
A-M $\lvert\Delta d\rvert$	Prim	80
A-M $\lvert\Delta d\rvert$	Wide-CI	82
A-M $\lvert\Delta d\rvert$	Sub-thr	12
A-M $\lvert\Delta d\rvert$	Weak	2
Anc Δr	support	11
Anc Δr	Sub-thr	1
Surf-O Δr	secondary	6
Surf-O Δr	Sub-thr	6

C.3 Gain Consistency and Endpoint Roles

Anchor displacement and surface-oxygen displacement provide local context around the primary endpoint under their own supporting and secondary roles. Table 5 distinguishes these endpoint families, and Table 4 gives the surface-oxygen gains and qualification roles.

Wide-interval rows show positive estimates with limited precision; below-threshold rows show favorable directions with lower trajectory-level consistency; weak rows contain little or negative gain. A positive mean can therefore coexist with limited interval precision or lower trajectory-level consistency. For example, the supplementary comparison table (`all_200_evidence_rows.csv`) identifies negative Anch Surf Path/Static gains for the H–Li adsorbate–anchor stratum and the Li_4O_4 bulk-composition stratum.

The expansion in Figure 5 selects 29 qualifying primary-endpoint rows, cycling through grouping fields toward a cumulative target of 80, with gain, paired-positive fraction, and sample count ranked in descending order within each field, and 71 other rows ranked by qualification category, gain, and sample count.

Figure A.1 provides the anchor and facet views. Chemistry strata compare the same local endpoint across adsorbate identities, while response strata organize comparisons by observed force, energy, or geometric change. Each gain remains attached to its model, control, and evaluation protocol.

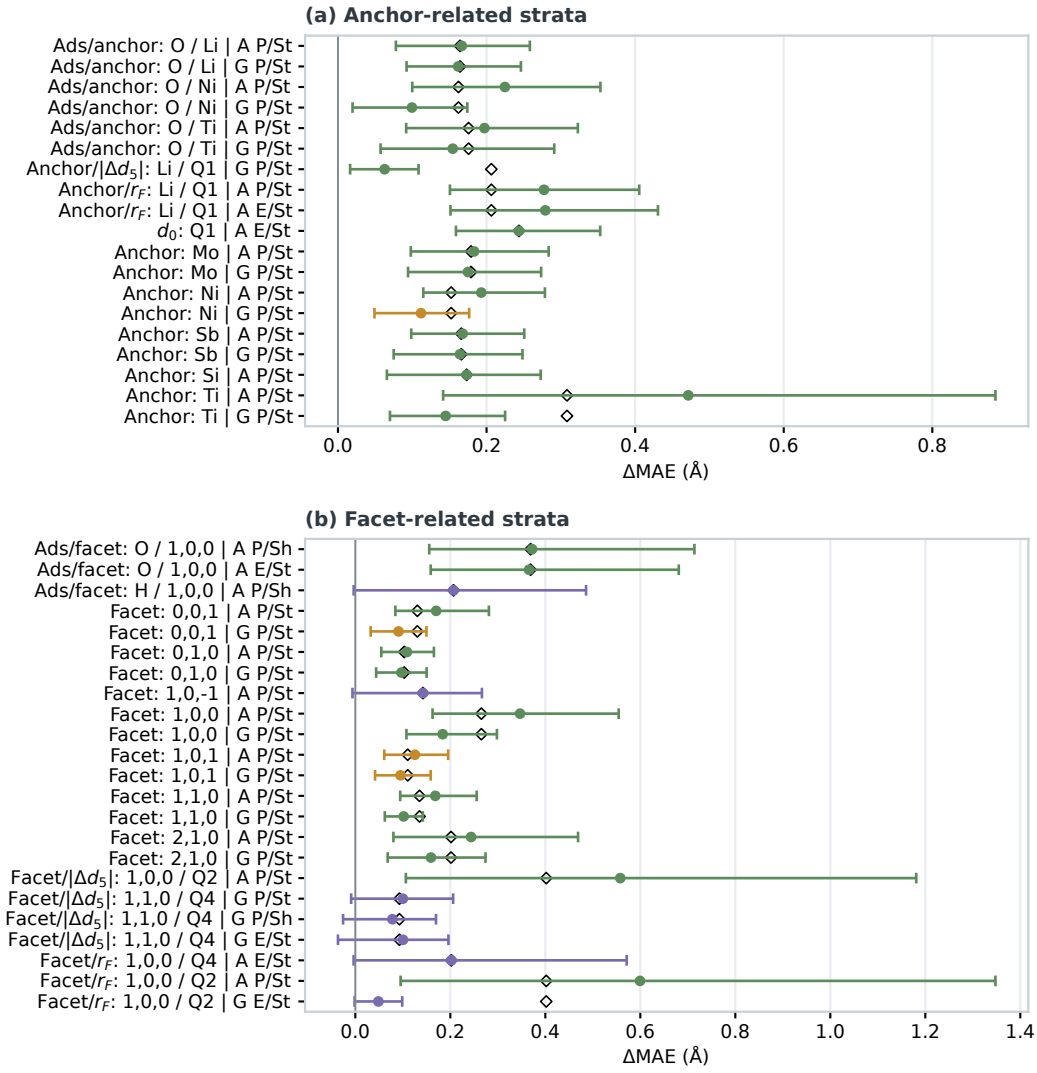

Figure A.1: Anchor and facet chemistry-strata comparisons under the original available-frame protocol, complementing Figure 5. (a–b) All 19/23 contributing comparisons for the ten/twelve anchor/facet stratum-value groups with the largest unweighted mean gains. Points show gains and 95% bootstrap intervals; open diamonds repeat each group mean. Selection and means use stored stratum values, which can span grouping fields. Labels give field, value, subset and model/control: A/G = Anch/Geom, P/E = Path/EndM ridge, St/Sh = Static/Shuffle. All rows use Surf and A-M. Ads denotes adsorbate; d_0 initial distance; r_F and $|\Delta d_5|$ frame-5 force ratio and absolute distance change; Q1–Q4 ranked quartiles. Facets retain stored index triples. Role colors follow Figure 3.

C.4 Baseline Pool and Raw-Error Scale

Table 6 retains all 57 baseline rows: 20 core rows and 37 physical-reference rows. The cross-validation references ask how accurately simple predictors estimate the same distance-change endpoint. In the larger-sample protocol, signed-distance ridge gives MAE 0.273 and $R^2 = 0.639$, providing a strong reference from initial distance, signed frame-5 and frame-10 distance changes, and adsorbate/anchor categories. Its performance makes local distance change an informative component to examine alongside the combined representation.

The seven pilot-CV references form a separate comparison on 489 trajectories. Initial-geometry ridge has lower MAE than the early-response and endpoint-matched ridge variants there, and their negative R^2 values remain in Table 6. For the three response-based pilot ridge variants, the same HfO_2 trajectory accounts for 94.4–98.9% of the total squared error. The 37 larger-sample physical references use 11941 trajectories.

Table 6: All 57 baseline rows for the primary adsorbate–metal endpoint, grouped by evaluation protocol. MAE is in Å. PM denotes physical group means, BM binned means, PR physical-coordinate ridge, Atom atom-set networks, and GNN the implemented local graph networks. Neural metrics are means over seeds 101, 202, and 303. Comparisons retain their original observation convention.

Red/blue bold values rank best/second within each protocol, using unrounded metrics (neural seed means); colors do not indicate significance. Light red identifies ERPA early-response implementations. Light blue marks the non-ERPA row with most second-best cells per protocol; ties use original row order.

Baseline	Class	Protocol	n	MAE (Å)↓	R^2 ↑
Held-out Surface Validation (13 rows)					
init geom MLP	MLP	Surf	1035	0.307	0.251
E/F MLP	MLP	Surf	1035	0.301	0.234
early response MLP	MLP	Surf	1035	**0.228**	**0.635**
endpoint-matched MLP	MLP	Surf	1035	**0.233**	**0.622**
init atomset	Atom	Surf	1035	0.395	-0.015
F5 atomset	Atom	Surf	1035	0.351	0.032
F10 atomset	Atom	Surf	1035	0.347	0.033
init graph	GNN	Surf	1035	0.382	-0.007
F5 graph	GNN	Surf	1035	0.370	-0.021
F10 graph	GNN	Surf	1035	0.375	-0.021
tuned init graph	GNN	Surf	1035	0.387	-0.005
tuned F5 graph	GNN	Surf	1035	0.363	-0.023
tuned F10 graph	GNN	Surf	1035	0.371	-0.005
Pilot Cross-Validation (7 rows)					
global mean	mean	Pilot CV	489	0.370	-0.002
ads mean	mean	Pilot CV	489	**0.363**	**-0.000**
ads-anchor mean	mean	Pilot CV	489	0.366	-0.017
init geom ridge	linear	Pilot CV	489	**0.335**	**0.083**
E/F ridge	linear	Pilot CV	489	0.466	-17.247
early response ridge	linear	Pilot CV	489	0.403	-15.937
endpoint-matched ridge	linear	Pilot CV	489	0.495	-54.122
Larger-Sample Five-Fold Trajectory-Hash CV (37 rows)					
F10 ads motion bin	BM	Hash CV	11941	0.389	0.076
F5 anchor motion bin	BM	Hash CV	11941	0.377	0.106
F10 anchor motion bin	BM	Hash CV	11941	0.380	0.104
F5 surface-O motion bin	BM	Hash CV	11941	0.417	0.015
F10 surface-O motion bin	BM	Hash CV	11941	0.417	0.010
ads response-proxy bin	BM	Hash CV	11941	0.414	0.008
anchor response-proxy bin	BM	Hash CV	11941	0.413	0.018
shell response-proxy bin	BM	Hash CV	11941	0.415	0.011
init chemistry ridge	PR	Hash CV	11941	0.386	**0.178**

Table 6 continued

Baseline	Class	Protocol	n	MAE (Å)↓	R^2 ↑
energy-only ridge	PR	Hash CV	11941	0.413	-0.013
force-only ridge	PR	Hash CV	11941	0.391	0.144
ads motion ridge	PR	Hash CV	11941	0.393	0.083
anchor motion ridge	PR	Hash CV	11941	0.384	0.145
surface-O motion ridge	PR	Hash CV	11941	0.404	0.017
response-proxy ridge	PR	Hash CV	11941	0.405	0.009
signed-distance ridge	PR	Hash CV	11941	**0.273**	**0.639**
E/F/proxy ridge	PR	Hash CV	11941	0.410	-0.151
geom/chem ridge	PR	Hash CV	11941	**0.375**	0.124
anchor elem mean	PM	Hash CV	11941	0.411	0.004
facet mean	PM	Hash CV	11941	0.421	-0.010
coord mean	PM	Hash CV	11941	0.412	0.004
O-neighbor mean	PM	Hash CV	11941	0.412	0.004
ads-facet mean	PM	Hash CV	11941	0.417	-0.012
ads-coord mean	PM	Hash CV	11941	0.403	0.010
ads-O-shell mean	PM	Hash CV	11941	0.403	0.011
anchor-coord mean	PM	Hash CV	11941	0.405	-0.005
anchor-O-shell mean	PM	Hash CV	11941	0.408	-0.010
facet-coord mean	PM	Hash CV	11941	0.419	-0.020
ads-anchor-coord mean	PM	Hash CV	11941	0.405	-0.039
bulk-comp mean	PM	Hash CV	11941	0.419	-0.011
init force bin	BM	Hash CV	11941	0.405	0.059
init A-M distance bin	BM	Hash CV	11941	0.390	0.066
F5 energy bin	BM	Hash CV	11941	0.419	0.003
F10 energy bin	BM	Hash CV	11941	0.418	0.004
F5 force-ratio bin	BM	Hash CV	11941	0.415	0.014
F10 force-ratio bin	BM	Hash CV	11941	0.415	0.015
F5 ads motion bin	BM	Hash CV	11941	0.393	0.070

D Further Discussion

Each comparison addresses a different part of early distance-change prediction. The static reference establishes performance available from chemical and initial-state descriptors. Geometry scrambling tests whether prediction improves when the increments remain paired with the recipient's retained response. Physical references examine simpler summaries, and neural baselines show how several representations perform on the endpoint task. Their joint reading separates the role of the chosen information from the score of a single predictor.

For scientific data mining, we link each prediction gain to its endpoint, subset, split, stratum, model, control, sample count, interval, and qualification statistics. These fields preserve the conditions under which a score was obtained and provide a common structure for retrieving comparisons by chemistry, response regime, or endpoint. The resulting organization connects a numerical gain to the physical question it answers.

The same comparison structure can be applied to other trajectory archives by specifying the observation rule, endpoint, static reference, and geometric reassignment. Predictor architectures, adsorbate families, and surface contexts can then be varied within that explicit design.

The absolute-error comparisons clarify the practical meaning of the representation gain. Combining observed geometry with force and energy descriptors improves Surf neural prediction over both initial geometry and energy–force inputs (Table 6).

E Comparison Details and Boundary Procedures

This appendix details how ERPA evaluates the predictive value of the early response and interprets the resulting distance-change gains. Module A defines how trajectories are decomposed and compared,

while Module B organizes the resulting evidence by chemistry, local response coordinates, surface context, and evidence strength. Figures A.2 and A.3 show the two modules separately.

E.1 Module A: Early-Response Prediction and Controls

Module A separates static context S_i, early trajectory block $X_{i,0:k}$, and endpoint $Y_i = |\Delta d_{A-M,i}|$. Static, paired-path, and shuffled-path ridge models are evaluated under the same split protocol, with endpoint labels, validation statistics, and row-role labels excluded from the feature block; the available-frame rule applies to short trajectories.

Mean gain, trajectory-level consistency, and bootstrap stability. For a stratum $\mathcal{I}$, the comparison uses paired error differences $e_i^{\text{control}} - e_i^{\text{path}}$, where $e_i^m = |Y_i - \hat{Y}_i^m|$. The reported ΔMAE is the mean paired reduction; paired-positive fraction is the fraction of trajectories with smaller path error. Bootstrap intervals resample these fixed paired errors within $\mathcal{I}$, with replacement at the trajectory level, and take the 2.5th and 97.5th percentiles of the mean gains (Efron, 1979; Efron and Tibshirani, 1993). The original global comparisons use 1000 resamples, the original stratified and sensitivity comparisons 500, and the expanded comparisons 400. Each resampled unit carries the model and control errors for the same trajectory together. The interval describes variation in the mean difference, while the paired-positive fraction measures how widely the improvement is shared across trajectories. Reading these quantities together distinguishes a large average gain from an improvement shared across many systems.

Geometry and path controls. Initial, observed and final local distances all call ASE minimum-image distance; an exception triggers direct Cartesian distance. Adsorbate, anchor and surface-oxygen centroid displacements use Cartesian centroid differences. The raw-trajectory check found no distance fallback among its 18 sampled trajectories; full-cache fallback incidence was not recorded. The shuffled-path control replaces X_i^{inc} with $X_{\pi(i)}^{\text{inc}}$ inside the same split, adsorbate identity, and surface-anchor element group, retaining each recipient's initial, force, and energy fields; initial-distance bins provide a descriptive sensitivity view. Prediction features use training-only normalization and are constructed separately from target values and evidence-role labels. Descriptive quartiles are formed across the input cache using ranks with ties ordered by first occurrence and four equal-count bins; missing values form a separate missing category, and failed binning receives an unbinned label.

Baseline implementation details. Tabular processing and ridge calculations use pandas and NumPy (McKinney, 2010; Harris et al., 2020); statistical plots use Matplotlib (Hunter, 2007). The tabular MLP uses learning rate 10^{-3}, batch size 512, at most 100 epochs and patience 14. The atom-set model uses 16-dimensional element embeddings and two width-96 atom layers followed by masked mean pooling; it uses learning rate 10^{-3}, batch size 128, at most 120 epochs and patience 18. Atom-set and graph inputs contain centered initial coordinates for all atoms, adding frame-5 displacement for F5 and both frame-5 and frame-10 displacements for F10. Padding to 220 atoms is masked. The local-graph model constructs its 12-neighbor graph from initial Cartesian distances, passes relative coordinates and distance with the node embeddings, and uses two width-96 message-passing layers with residual updates, layer normalization, and masked mean pooling. Its learning rate is 8×10^{-4}, batch size 64, maximum epoch count 80 and patience 12; the tuned runs increase the training batch size to 96 and evaluation batch size from 96 to 128. All three neural families use AdamW with weight decay 10^{-3}, dropout 0.1, squared-error training loss on training-standardized targets, and checkpoint selection by ID MAE after restoring the target length scale. Table 6 reports the arithmetic mean of each neural metric over seeds 101, 202, and 303.

Physical-reference implementation details appear in Appendix B.4.

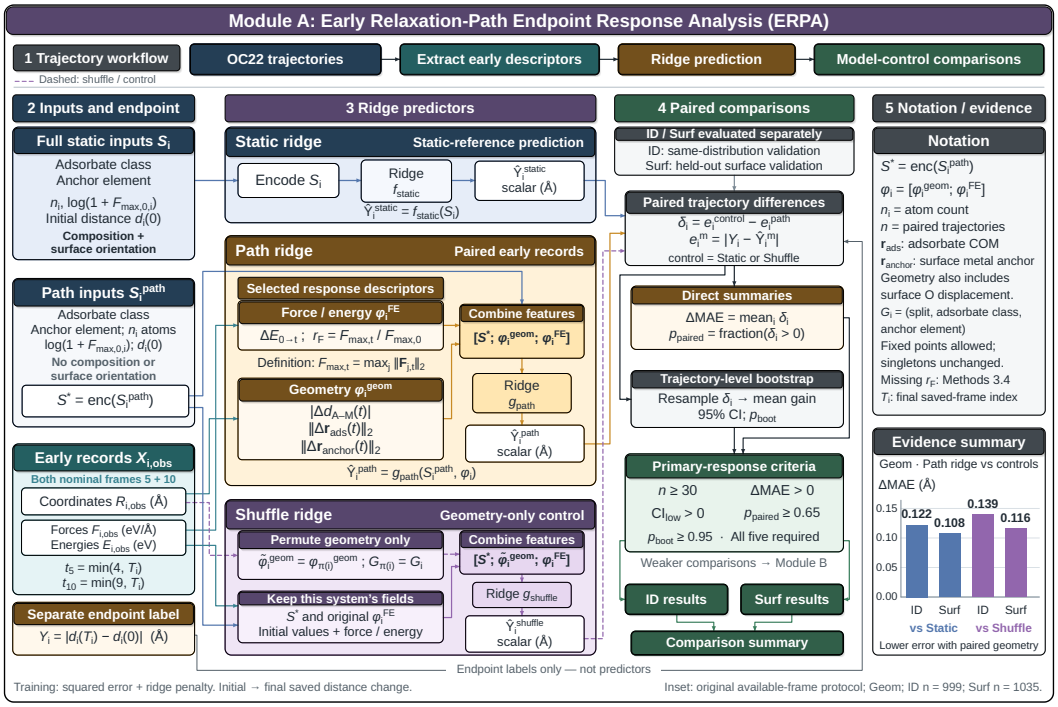

Figure A.2: Module A: controlled local-response comparison under the original available-frame protocol. Static uses S_i; Path and Shuffle use S_i^{path} and the observed descriptors. In the Combine features label, $S^* = \mathrm{enc}(S_i^{\mathrm{path}})$. Both indices $t_5 = \min(4, T_i)$ and $t_{10} = \min(9, T_i)$ contribute; nominal frame 10 uses the final stored frame for $T_i \leq 9$. Shuffle reassigns geometric increments within split/adsorbate/anchor groups and retains the recipient's initial, force and energy fields. Fixed points are allowed and singleton groups remain unchanged. The gain is the mean paired absolute-error reduction; p_{paired} and p_{boot} measure trajectory consistency and bootstrap positivity. The right inset retains the Geom Path gains against Static and Shuffle for ID and Surf. Ads denotes adsorbate.

E.2 Module B: Chemistry and Response Stratification

Module B uses the matched comparisons from Module A to examine how prediction gains vary with adsorbate chemistry for the same local adsorbate–metal endpoint. Oxygenated adsorbates resolve the response by adsorbate identity; early force response, energy drop, and adsorbate–metal distance change provide effective response coordinates; anchor, facet, and adsorbate–anchor strata locate the signal in its local surface context. The evidence-role layer retains primary, supporting, secondary-endpoint, wide-interval, below-stability, and weak rows in one comparison table.

Local distance as a predictive reference. The physical references in Table 6 add a direct geometric interpretation: signed early distance change preserves whether the local distance initially increases or decreases, whereas the target records the absolute initial-to-final change. Its strong performance within the physical-reference cross-validation protocol shows the predictive value of this simple local summary. The combined-response comparisons place that geometric information alongside the system's force and energy response.

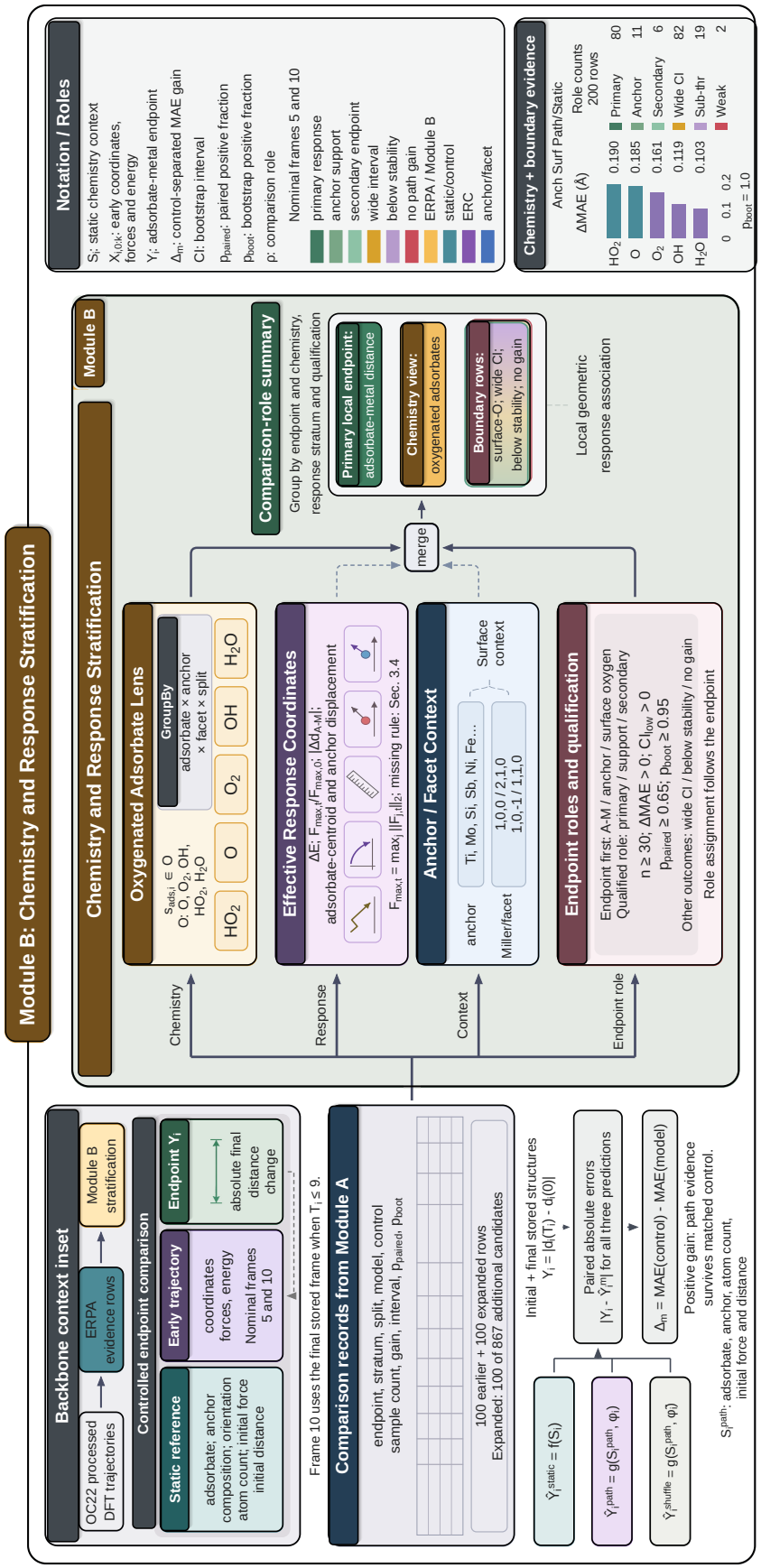

Figure A.3: Module B: chemistry and response stratification under the original available-frame protocol. Inputs and paired errors feed adsorbate, response-coordinate and surface-context views; endpoint and numerical qualification determine roles. The 200 rows comprise 100 earlier rows and 100 expanded rows selected from 867 additional candidates. Role counts span three endpoint families; the 19 below-stability rows comprise 12 adsorbate–metal, one anchor and six surface-oxygen rows (Table 5). The chemistry inset shows Anch Surf Path/Static gains for five oxygenated adsorbates, each with $p_{\text{boot}} = 1$. Gains are in Å; p_{paired} and p_{boot} denote trajectory consistency and bootstrap positivity. The grid represents comparison records with the listed fields.

E.3 Strict Pre-Endpoint Refit Details

The 24 prespecified comparisons in Table 7 evaluate the early-response gain against Static and Shuffle using observations strictly before the final stored frame. All cohorts preserve the processed source split identities and exclude short trajectories before within-cohort geometry reassignment. The endpoint remains the final stored frame.

Each cohort uses the original regularization candidates, ID-RMSE selection, and minimum training category count 20; geometry reassignment occurs only within its split/adsorbate/anchor groups, with seed 909. Paired-trajectory percentile intervals use 1000 resamples of fixed fitted prediction errors, also seeded at 909 (Efron, 1979; Efron and Tibshirani, 1993). This adds 42 fits while retaining the original 57-baseline protocols.

Input-force sensitivity. We additionally require finite maximum forces no greater than 50 eV/Å at indices 0, 4, and 9 in each $L_i \geq 11$ split before normalization, shuffling, and refitting Static, Path, and Shuffle. Geom retains 9341/925/985 train/ID/Surf trajectories and Anch 10676/1043/1101; model selection, seed 909, and 1000 paired resamples are unchanged. Surf Path gains against Static and Shuffle are respectively 0.058361 [0.046438, 0.071029] and 0.058749 [0.048329, 0.070129] Å for Geom, and 0.060480 [0.049278, 0.071170] and 0.060837 [0.049922, 0.071867] Å for Anch. All four ID intervals are also positive. Gains are smaller than in the length-only cohorts; Surf Path R^2 is 0.286 for Geom and 0.112 for Anch. This screen tests sensitivity to large input forces; retained training trajectories still include an absolute frame-10 energy change of 8105.608634 eV. Electronic-convergence metadata are absent from the analysis cache.

Table 7: Strict pre-endpoint Path ridge comparisons against Static and Shuffle. Gains and intervals are in Å. The length column is the minimum frame count. All fits use the common complete cohort for that subset and length cutoff; ID is development validation. Two Static rows fall below the 0.65 trajectory-consistency threshold: Geom $L \geq 21$ Surf and Anch $L \geq 11$ Surf.

Subset	$L_{\min}$	Split	Control	n	ΔMAE	95% CI	p_{paired}	p_{boot}
Geom	11	ID	Static	979	0.109809	[0.082294, 0.139806]	0.667	1.000
Geom	11	ID	Shuffle	979	0.129453	[0.103954, 0.163914]	0.720	1.000
Geom	11	Surf	Static	1022	0.094553	[0.076246, 0.113587]	0.653	1.000
Geom	11	Surf	Shuffle	1022	0.097738	[0.080517, 0.115895]	0.686	1.000
Geom	21	ID	Static	957	0.097435	[0.079162, 0.119729]	0.656	1.000
Geom	21	ID	Shuffle	957	0.112012	[0.094767, 0.133205]	0.718	1.000
Geom	21	Surf	Static	1009	0.087424	[0.069596, 0.106241]	0.638	1.000
Geom	21	Surf	Shuffle	1009	0.093569	[0.076663, 0.111375]	0.690	1.000
Geom	51	ID	Static	916	0.106767	[0.086919, 0.129578]	0.688	1.000
Geom	51	ID	Shuffle	916	0.121790	[0.102271, 0.144850]	0.721	1.000
Geom	51	Surf	Static	964	0.092497	[0.073636, 0.111466]	0.659	1.000
Geom	51	Surf	Shuffle	964	0.099118	[0.079485, 0.118838]	0.687	1.000
Anch	11	ID	Static	1106	0.122564	[0.093446, 0.153495]	0.662	1.000
Anch	11	ID	Shuffle	1106	0.138956	[0.115272, 0.168065]	0.679	1.000
Anch	11	Surf	Static	1151	0.111255	[0.091262, 0.132560]	0.639	1.000
Anch	11	Surf	Shuffle	1151	0.111055	[0.091964, 0.130603]	0.659	1.000
Anch	21	ID	Static	1082	0.127575	[0.107113, 0.149380]	0.662	1.000
Anch	21	ID	Shuffle	1082	0.130231	[0.107211, 0.155588]	0.661	1.000
Anch	21	Surf	Static	1135	0.118458	[0.098100, 0.141305]	0.654	1.000
Anch	21	Surf	Shuffle	1135	0.115825	[0.093868, 0.137973]	0.661	1.000
Anch	51	ID	Static	1032	0.120950	[0.100539, 0.141951]	0.653	1.000
Anch	51	ID	Shuffle	1032	0.118486	[0.100025, 0.141003]	0.661	1.000
Anch	51	Surf	Static	1085	0.111989	[0.089367, 0.133398]	0.666	1.000
Anch	51	Surf	Shuffle	1085	0.108116	[0.087185, 0.129809]	0.666	1.000

Tables 8 and 9 extend the analysis to all seven representations and 48 additional paired comparisons, giving 72 comparisons in total, including material-extended Path and endpoint-matched references.

Table 8: All 42 strict pre-endpoint ridge fits. MAE is in Å; the final columns give the selected regularization and encoded input count. Input definitions follow Section 3.5 and Appendix B.1; each row reports ID development validation and Surf on its own cohort.

Red/blue bold values rank best/second independently in each performance column within a subset/length cohort, using unrounded values; colors do not indicate significance. Light red identifies Path (ERPA) and its Path+material variant. Light blue marks the non-ERPA row with most second-best cells in a cohort, when any exist. EndM is a reference.

Model	ID MAE↓	ID R^2 ↑	Surf MAE↓	Surf R^2 ↑	α	Inputs
Geom, $L \geq 11$: train/ID/Surf = 9758/979/1022						
Static-AA	0.376605	0.153975	0.352717	0.162274	10	58
Static	0.375656	0.156527	0.354468	0.155115	10	118
E/F	0.399193	0.025395	0.360251	0.132499	1000	62
Path	**0.265847**	**0.599042**	0.259915	0.486761	1000	70
Path+material	**0.265682**	0.598286	**0.259762**	**0.486900**	1000	130
EndM	0.266094	**0.600264**	**0.259426**	**0.487163**	1000	64
Shuffle	0.395300	0.046529	0.357652	0.144966	1000	70
Geom, $L \geq 21$: train/ID/Surf = 9631/957/1009						
Static-AA	0.350930	0.167195	0.346424	0.166830	10	58
Static	0.349910	0.170750	0.348085	0.160371	10	118
E/F	0.366936	0.132327	0.356928	0.127962	1000	62
Path	**0.252474**	**0.532843**	0.260660	0.474985	1000	70
Path+material	**0.252207**	**0.533092**	**0.260540**	**0.475120**	1000	130
EndM	0.252535	0.532737	**0.260355**	**0.475017**	1000	64
Shuffle	0.364486	0.138692	0.354229	0.139890	1000	70
Geom, $L \geq 51$: train/ID/Surf = 9090/916/964						
Static-AA	0.354495	0.135596	0.346354	0.139806	10	57
Static	0.354120	0.140492	0.348477	0.131528	10	114
E/F	0.369375	0.110770	0.355722	0.104885	1000	61
Path	**0.247353**	**0.500479**	**0.255980**	**0.424518**	100	69
Path+material	0.247961	0.498526	0.256943	0.423700	100	126
EndM	**0.245978**	**0.502913**	**0.254132**	**0.429499**	100	63
Shuffle	0.369143	0.110999	0.355098	0.106866	1000	69
Anch, $L \geq 11$: train/ID/Surf = 11211/1106/1151						
Static-AA	0.467038	0.123606	0.425664	0.075221	10	59
Static	0.467985	0.122650	0.423623	0.074499	100	127
E/F	0.486251	0.048486	0.426086	0.070785	1000	63
Path	**0.345421**	0.313493	**0.312368**	**0.310130**	1000	71
Path+material	0.345600	**0.313913**	0.312428	0.309864	1000	139
EndM	**0.328848**	**0.327105**	**0.302156**	**0.321610**	10	65
Shuffle	0.484376	0.051993	0.423423	0.076082	1000	71
Anch, $L \geq 21$: train/ID/Surf = 11051/1082/1135						
Static-AA	0.430788	0.106804	0.415144	0.083205	100	58
Static	0.430206	0.106306	0.415172	0.083297	100	125
E/F	0.432545	0.041651	0.412405	0.087805	10	62
Path	**0.302631**	0.319604	**0.296713**	**0.335082**	1	70
Path+material	0.303572	**0.321270**	0.299991	0.332147	10	137
EndM	**0.296042**	**0.338479**	**0.291721**	**0.343032**	0.01	64
Shuffle	0.432862	0.041455	0.412538	0.085601	10	70
Anch, $L \geq 51$: train/ID/Surf = 10447/1032/1085						
Static-AA	0.390688	0.172040	0.387656	0.111735	100	58
Static	0.390109	0.170713	0.387216	0.113771	100	121
E/F	0.389095	0.180689	0.385528	0.116962	100	62
Path	**0.269159**	**0.525333**	**0.275227**	**0.485260**	100	70
Path+material	0.269871	0.524030	0.276218	0.484522	100	133
EndM	**0.262463**	**0.555766**	**0.269571**	**0.505975**	10	64
Shuffle	0.387645	0.186137	0.383343	0.127059	100	70

The $L_i \geq 21$ and $L_i \geq 51$ cohorts test the early-response gains when the nominal tenth observation lies further from the stored endpoint. Table 9 gives the additional paired comparisons.

Table 9: Additional strict pre-endpoint Path ridge comparisons against all other fitted representations. Gains and intervals are in Å. The length column is the minimum frame count. All fits use the common complete cohort for that subset and length cutoff; ID is development validation. Static-AA retains adsorbate/anchor categories and initial numerical fields; Path+material adds composition and orientation. Negative gains favor the reference. E/F and EndM retain the input definitions in Section 3.5.

Subset	$L_{\min}$	Split	Control	n	ΔMAE	95% CI	p_{paired}	p_{boot}
Geom	11	ID	Static-AA	979	0.110758	[0.083915, 0.143252]	0.659	1.000
Geom	11	ID	E/F	979	0.133346	[0.107639, 0.163244]	0.725	1.000
Geom	11	ID	Path+material	979	-0.000165	[-0.000631, 0.000274]	0.527	0.230
Geom	11	ID	EndM	979	0.000247	[-0.000575, 0.001002]	0.633	0.722
Geom	11	Surf	Static-AA	1022	0.092802	[0.075472, 0.111932]	0.655	1.000
Geom	11	Surf	E/F	1022	0.100336	[0.082867, 0.117047]	0.693	1.000
Geom	11	Surf	Path+material	1022	-0.000153	[-0.000556, 0.000278]	0.526	0.226
Geom	11	Surf	EndM	1022	-0.000489	[-0.001212, 0.000160]	0.658	0.081
Geom	21	ID	Static-AA	957	0.098456	[0.079857, 0.123200]	0.654	1.000
Geom	21	ID	E/F	957	0.114461	[0.095679, 0.136315]	0.711	1.000
Geom	21	ID	Path+material	957	-0.000267	[-0.000652, 0.000119]	0.561	0.090
Geom	21	ID	EndM	957	0.000061	[-0.000574, 0.000732]	0.436	0.569
Geom	21	Surf	Static-AA	1009	0.085764	[0.067376, 0.104155]	0.637	1.000
Geom	21	Surf	E/F	1009	0.096268	[0.079136, 0.115221]	0.700	1.000
Geom	21	Surf	Path+material	1009	-0.000120	[-0.000493, 0.000225]	0.602	0.233
Geom	21	Surf	EndM	1009	-0.000306	[-0.000863, 0.000210]	0.450	0.136
Geom	51	ID	Static-AA	916	0.107141	[0.085867, 0.126988]	0.670	1.000
Geom	51	ID	E/F	916	0.122022	[0.101133, 0.144871]	0.714	1.000
Geom	51	ID	Path+material	916	0.000608	[-0.000629, 0.001785]	0.571	0.835
Geom	51	ID	EndM	916	-0.001376	[-0.002842, -0.000013]	0.404	0.023
Geom	51	Surf	Static-AA	964	0.090374	[0.071200, 0.110987]	0.652	1.000
Geom	51	Surf	E/F	964	0.099742	[0.082535, 0.119625]	0.696	1.000
Geom	51	Surf	Path+material	964	0.000963	[-0.000193, 0.002160]	0.579	0.943
Geom	51	Surf	EndM	964	-0.001848	[-0.003331, -0.000306]	0.436	0.005
Anch	11	ID	Static-AA	1106	0.121617	[0.092304, 0.153635]	0.652	1.000
Anch	11	ID	E/F	1106	0.140830	[0.115168, 0.167750]	0.682	1.000
Anch	11	ID	Path+material	1106	0.000179	[-0.000240, 0.000609]	0.486	0.793
Anch	11	ID	EndM	1106	-0.016573	[-0.022876, -0.010733]	0.429	0.000
Anch	11	Surf	Static-AA	1151	0.113296	[0.092957, 0.133930]	0.645	1.000
Anch	11	Surf	E/F	1151	0.113718	[0.094543, 0.134639]	0.666	1.000
Anch	11	Surf	Path+material	1151	0.000060	[-0.000376, 0.000497]	0.471	0.629
Anch	11	Surf	EndM	1151	-0.010212	[-0.015560, -0.005530]	0.459	0.000
Anch	21	ID	Static-AA	1082	0.128157	[0.107317, 0.149277]	0.664	1.000
Anch	21	ID	E/F	1082	0.129915	[0.108374, 0.155731]	0.657	1.000
Anch	21	ID	Path+material	1082	0.000941	[-0.001098, 0.003016]	0.523	0.810
Anch	21	ID	EndM	1082	-0.006589	[-0.011122, -0.002760]	0.416	0.000
Anch	21	Surf	Static-AA	1135	0.118430	[0.097354, 0.139493]	0.656	1.000
Anch	21	Surf	E/F	1135	0.115692	[0.094165, 0.137609]	0.654	1.000
Anch	21	Surf	Path+material	1135	0.003278	[0.000941, 0.005657]	0.529	0.996
Anch	21	Surf	EndM	1135	-0.004992	[-0.009265, -0.001432]	0.443	0.004
Anch	51	ID	Static-AA	1032	0.121530	[0.101129, 0.143395]	0.650	1.000
Anch	51	ID	E/F	1032	0.119937	[0.100371, 0.141661]	0.657	1.000
Anch	51	ID	Path+material	1032	0.000713	[-0.000701, 0.002225]	0.556	0.833
Anch	51	ID	EndM	1032	-0.006695	[-0.012069, -0.002682]	0.444	0.000
Anch	51	Surf	Static-AA	1085	0.112429	[0.091704, 0.134476]	0.672	1.000
Anch	51	Surf	E/F	1085	0.110301	[0.088634, 0.132867]	0.669	1.000
Anch	51	Surf	Path+material	1085	0.000991	[-0.000562, 0.002506]	0.551	0.916
Anch	51	Surf	EndM	1085	-0.005657	[-0.009924, -0.001948]	0.454	0.000