Journal of
*Manufacturing and Materials Processing*

MDPI

*Article*

# Retrospective Evaluation of the In-Process Intelligent Compensation Method for Mass-Customized Manufacturing of Freeform Lenses Based on Ensemble Learning and Dynamic EWMA Fusion

**Xin Chen [1,2], Kai Cheng [2,*], Liang Zhao [3] and Yuanzheng Fu [1]**

[1] School of Intelligent Science and Information Engineering, Shenyang University, Shenyang 110044, China; zbcyyqx@163.com (X.C.); syu_star@126.com (Y.F.)
[2] Department of Mechanical and Aerospace Engineering, Brunel University of London, Kingston Lane, Uxbridge UB8 3PH, UK
[3] School of Mechanical-Electrical Engineering, Harbin Institute of Technology, Harbin 150006, China; zhaoliang309@hit.edu.cn
* Correspondence: kai.cheng@brunel.ac.uk

**Abstract**

Freeform ophthalmic lens manufacturing requires compensation for prescription-dependent deviations and process drift. This study retrospectively evaluates an order-to-order compensation method combining ensemble prediction, exponentially weighted moving average (EWMA), residual tracking, and a proposed material-specific three-level gate. The 2245 manufacturing records are partitioned by whole chronological days into training (1356), validation (484), and final-test (405) sets. Thirteen models are compared, with controller settings selected using validation data only. Under unit-gain replay, joint pass rates for spherical and cylindrical power deviations within the study-specific ±0.10 D band are 57.04% without added compensation, 68.40% with historical fixed-offset correction, 75.06% with static AI, 65.43% with pure EWMA, 74.81% with global AI + EWMA, and 76.54% with material-specific tracking and the gate. The material strategy gains 19.51 percentage points over the baseline (95% date-bootstrap interval: 14.57–25.70). Paired intervals comparing material tracking with static AI and global fusion include zero in the robustness analysis. Global fusion achieves higher joint pass rates than static AI in all five expanding-window evaluations under the robustness protocol. The gate remains FULL in the primary replay, producing identical commands to ungated tracking; one material performs below the baseline. These findings motivate prospective assessment of machine response, manufacturing yield, and gate protection.

**Keywords:** freeform lens manufacturing; in-process intelligent compensation; ensemble learning; dynamic EWMA tracking; high-precision manufacturing

Check for updates

Academic Editor: Yuki Shimizu

Received: 7 October 2026
Revised: 10 November 2026
Accepted: 12 November 2026
Published: 16 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

Mass-customized precision manufacturing of ophthalmic lenses is critically important for their optical quality and performance, while the manufacturing accuracy throughout the process directly affects the lens imaging quality, wearing comfort, and product consistency. In lens precision manufacturing, spherical power deviation (S-deviation) and cylindrical power deviation (C-deviation) are the core indicators for evaluating the lens processing quality and motivating accurate optical-power inspection [1]. This study uses

a common internal comparison band of ±0.10 D for both deviations; this band is not a complete product-conformity specification, such as the requirements for power-variation lenses in ISO 8980-2 [2]. However, the lens manufacturing process is jointly affected by multiple factors, including prescription-specific optical parameters, material properties, in-process blocker thermal management, machine status, tool wear, and variations in ambient temperature and humidity. Consequently, machining deviations may exhibit nonlinear, coupled, and time-varying behavior in line with the variation in lens curvatures associated with individual personalized prescription information and data. Traditional compensation methods for precision processing (freeform lens surface generation and surface polishing in particular) mostly rely on manually compiled statistical data, operator/practitioner experience, historical averages, or fixed rules for corrections. Although such traditional strategies can be effective when the manufacturing process state is relatively stable and lens curvature varies little between orders, they may be unable to cope with in-process dynamic drifts, particularly in high-throughput lens manufacturing. For instance, progressive tool wear may cause a gradual shift in the mean machining deviation; variations and differences among respective lenses and material batches may alter thermal shrinkage and stress-release behavior, and equipment thermal states and environmental disturbances may also affect the final surface-form accuracy. Therefore, achieving stable prediction and dynamic compensation of S-deviation and C-deviation under complex manufacturing conditions is a key issue for improving lens processing quality and high manufacturing yield. Figure 1 shows the proposed order-to-order loop: a controller combines pending-order inputs with measured S- and C-deviations from completed orders and issues a gate-scaled command to the primary target processes for the next eligible order.

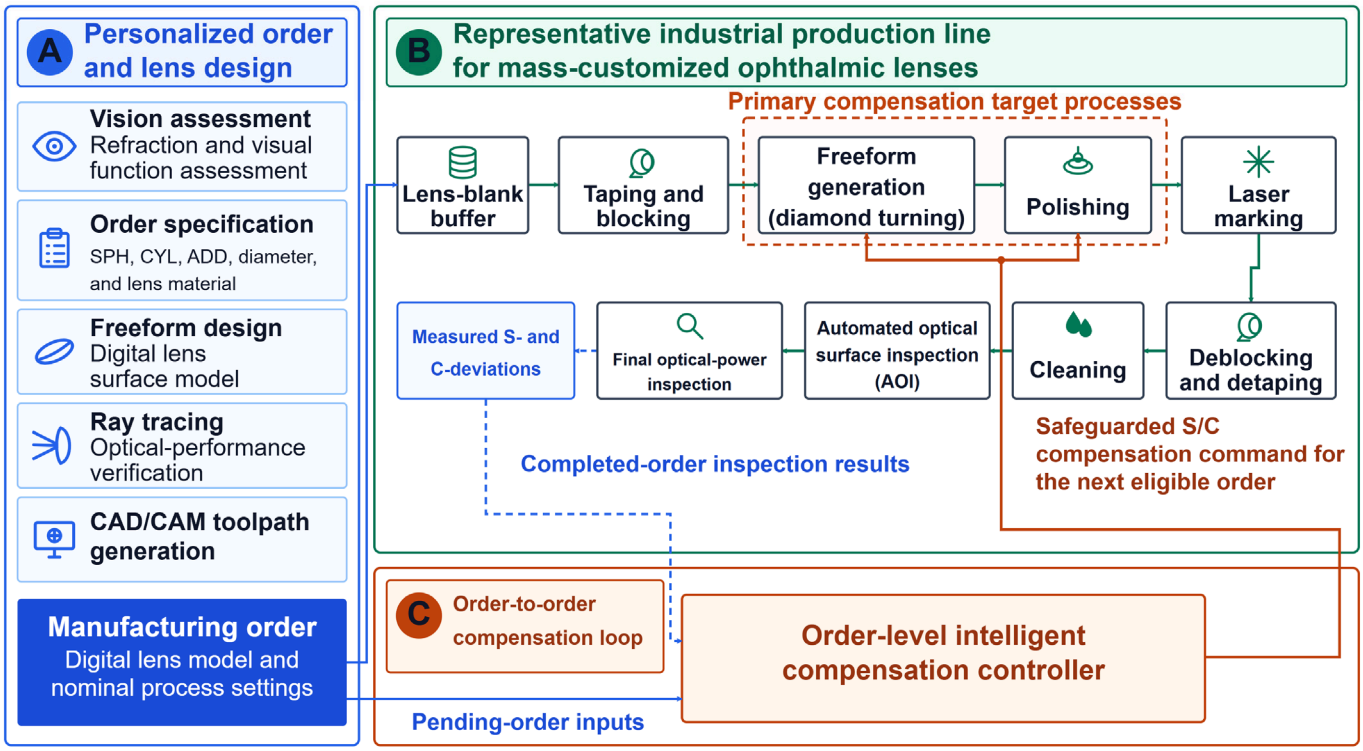



**Figure 1.** Ophthalmic lens manufacturing process and proposed order-to-order compensation loop. (**A**) Personalized order preparation and lens design; (**B**) Industrial production line with freeform generation and polishing as the primary compensation targets; (**C**) Order-to-order compensation using pending-order inputs and completed-order inspection feedback.

In recent years, data-driven machine learning methods have been widely applied to quality prediction, in-process monitoring, and process optimization in manufacturing [3,4]. Industrial dataspace research further addresses the integration of machining data and domain knowledge for operations control [5]. In complex manufacturing scenarios such as injection molding, previous studies have systematically reviewed the development of process quality control and applied deep learning, closed-loop control, and multi-objective optimization methods to fault detection, formulation optimization, and process parameter adjustment [6–9]. Data-driven approaches also support industrial process monitoring and fault diagnosis [10], while optical-manufacturing reviews address the process challenges of precision glass molding [11]. In related precision micromachining using diamond tools, Huo and Cheng showed that surface roughness and burr formation in oxygen-free, high-conductivity copper depend on microtool types and cutting conditions [12]. Meanwhile, statistical process control methods are of great value in dynamic process monitoring. EWMA control charts enhance the influence of recent observations on current state estimation through exponential weighting, making them suitable for monitoring small mean shifts in manufacturing processes. In recent years, the integration of machine learning and EWMA control charts has also been used to improve the sensitivity and robustness of process monitoring [13,14]. In addition, physics-informed machine learning (PIML) provides a new perspective for integrating physical laws with data-driven models. Related studies have shown that incorporating scientific knowledge or physical constraints into machine learning models can help improve model generalization, data efficiency, and physical consistency [15–19]. Research on explainable artificial intelligence in smart manufacturing and industrial cyber-physical systems has further emphasized the importance of model transparency, reliability, and deployability [20,21]. These studies motivate three requirements for order-to-order lens compensation [22–26]. First, static input–output prediction should be evaluated alongside adaptation to sequential residual changes. Second, the choice between global and material-specific tracking should account for heterogeneous order composition and available material history. Third, compensation requires explicit safeguards and tests of their protective effect under abnormal operating conditions. When systematic bias occurs in model prediction or compensation direction, it may reduce the yield of certain materials or even lead to batch-level quality risks.

To address order heterogeneity and evolving manufacturing states, this paper retrospectively evaluates an order-to-order compensation approach for mass-customized lens manufacturing, which integrates ensemble learning with dynamic EWMA residual tracking in the process. The primary analysis of this approach uses six recorded or directly derived prescription-order features and compares 13 model families on chronological partitions. Material-specific tracking and a three-level gate are evaluated as extensions of the global controller. Cooling time for the lens-fixing blocker thermal management is examined in a separate auxiliary cohort, while optical-prior neural models and rare-material transfer are exploratory comparisons. The evaluation is a retrospective replay of historical inspections, with no test-based parameter selection. The analysis quantifies compensation gains and failure modes across compensation strategies and lens materials. EWMA run-to-run control addresses metrology delay and plant-model mismatch [27]. This study integrates EWMA with prescription-dependent prediction and compares global and material-level residual adaptation chronologically in lens manufacturing.

## 2. Methodology

### *2.1. Overall Framework of the In-Process Intelligent Compensation Method*

The in-process intelligent compensation method proposed in this paper targets the problem of dynamic deviation control in precision lens processing. Its objective is to pre-

dict S-deviation and C-deviation before lens processing and calculate the corresponding in-process compensation values for the pending order. The basic idea of the method is as follows: first, machine learning models are used to learn the complex nonlinear relationships among order parameters, material properties, process parameters, and machining deviations; then, the EWMA method is employed to continuously track the temporal drift component in the model residuals; finally, the static prediction results and dynamic drift estimates are fused to obtain the in-process compensation values for the next eligible order. This method considers not only the individualized differences among orders but also the influence of time-varying manufacturing system states on machining deviations. As shown in Figure 2, the framework comprises three phases: (A) offline model development and freezing; (B) in-process compensation for a pending order; and (C) inspection-delayed material-state updating. Panel (B) of Figure 2 contains four compensation modules: the feature pipeline, static S/C prediction, AI–EWMA fusion, and the material-specific safety gate.

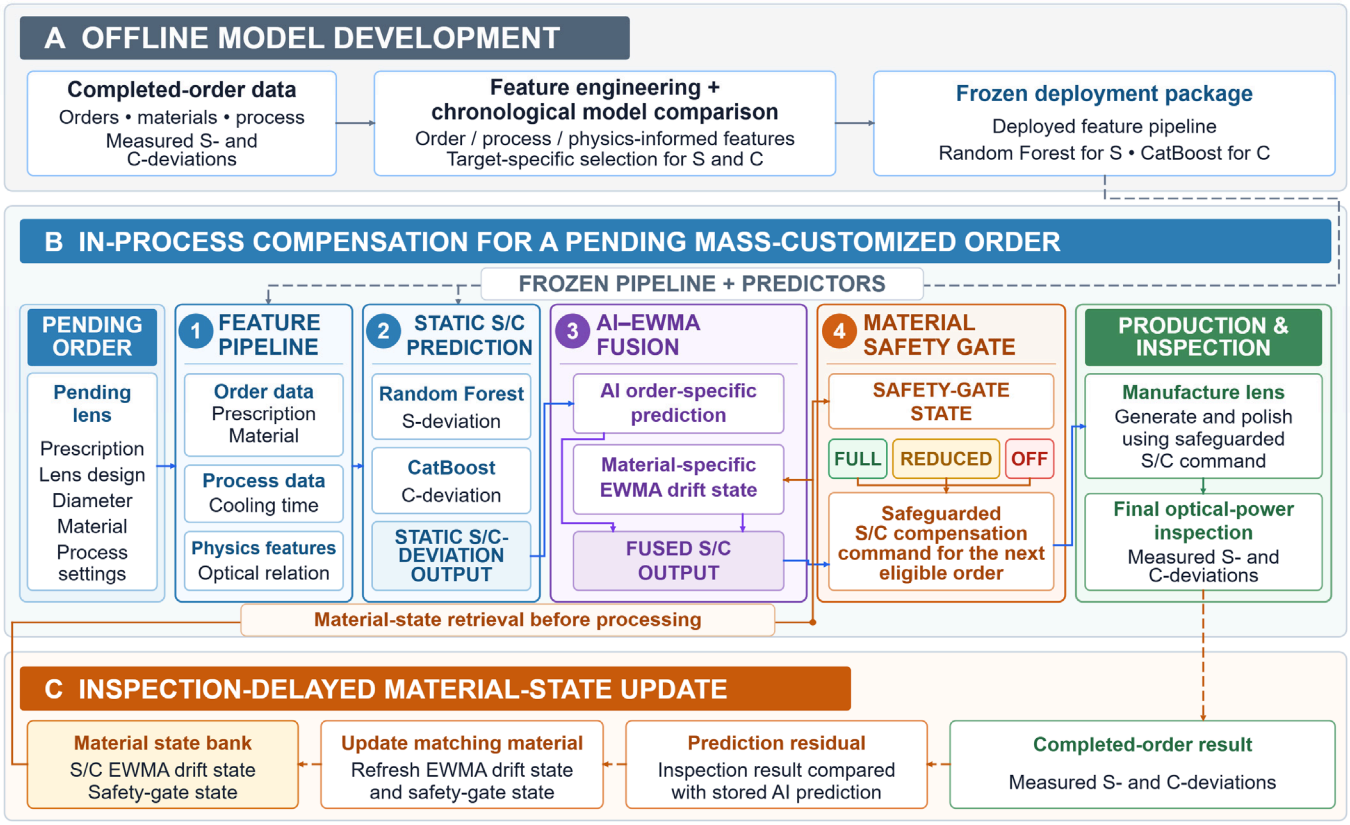



**Figure 2.** Framework of the proposed in-process intelligent compensation method. (**A**) Offline model development, chronological model comparison, and freezing of the feature pipeline and predictors; (**B**) Pending-order compensation through feature processing, static S/C prediction, AI–EWMA fusion, and a material-specific safety gate; (**C**) Inspection-delayed updating of material-specific EWMA and safety-gate states using completed-order feedback.

In this study, "in-process" denotes sequential order-to-order adaptation during the running manufacturing stream: compensation for a pending order is computed from its pre-processing features and the EWMA state available from completed and inspected orders, rather than by continuously adjusting the same lens during machining. "Mass-customized" denotes high-throughput order-level manufacturing in which each lens is defined by an individual customer's prescription-specific combination of SPH, CYL, ADD, diameter, material, and design requirements. Figure 2 shows the proposed deployment architecture. The

retrospective replay orders the recorded inputs by timestamp and releases each timestamp group's outcomes before the next group's decision. This release schedule is an evaluation assumption; elapsed time from manufacturing to inspection was not measured.

The primary predictor uses six inputs: front curve, normalized ADD, curve–power difference (front curve minus ordered SPH), ordered SPH, ordered CYL, and categorical material. Cooling time is evaluated in a separate auxiliary cohort, and 13 deterministic optical proxies are used in the exploratory models. Diameter is a candidate order field in the architecture but is excluded from the evaluated predictors. Before updating EWMA, the feedback path removes the applied command effect from the inspected deviation and subtracts the static prediction.

The static-prediction module compares 13 model families spanning linear, neighbor, kernel, ensemble, and neural methods. All transformations are fitted on historical training data. Random Forest for S and CatBoost for C attain the lowest target-specific validation MAEs among the tested configurations and form the core controller shown in Figure 2. Predictors are refitted on training plus validation records before final-test replay. The final-test labels do not determine model hyperparameters, EWMA settings, gate thresholds, or neural stopping epochs. Ensemble learning here refers to the trees combined within each target-specific Random Forest or CatBoost predictor. Validation MAE determines the choice for each target, and the common prediction interface permits tracking strategies to be compared with the predictors held fixed.

The third online module performs AI + EWMA fusion. The AI order-specific prediction estimates the pending order's machining deviation according to its order features, mainly reflecting individualized differences caused by order specifications, material properties, and process conditions. The EWMA module estimates recent systematic drift in the manufacturing system based on historical prediction residuals, which may reflect unrecorded factors such as equipment status, environmental changes, and material-batch fluctuations. By weighted fusion of these two types of information, the system can simultaneously consider order-specific differences and temporal drift during compensation. For a pending order, the fusion uses the most recent matching-material EWMA drift state available before machining; after final optical-power inspection, the prediction residual updates that material's EWMA state, while the compensated deviation updates its safety-gate state for a later eligible order.

The fourth module applies a matching-material safety-gate state. The gate can retain, reduce, or disable the S and C compensation components independently, using only previously available matching-material compensated deviations. FULL, REDUCED, and OFF weights are 1, 0.5, and 0, respectively. Gate activation and material-level performance are evaluated separately.

### 2.2. Order-Dependent Variation Modeling: Static Prediction Based on Ensemble Learning

Lens processing deviations are jointly affected by order-specific optical parameters, material properties, and process parameters, resulting in variation among orders. To characterize this order-dependent variation, this paper first constructs a static deviation prediction model. The objective of the static prediction model is to predict the possible S-deviation and C-deviation of the current order based on its input features, thereby providing a basis for subsequent compensation calculation. To systematically select base models suitable for lens processing deviation prediction, this paper compares 13 mainstream machine learning algorithms. According to their model structures and learning paradigms, these algorithms can be divided into four categories.

The first category consists of linear models, including linear regression, ridge regression, and Lasso regression. Linear regression assumes a linear relationship between the input features and the target variable. It has a simple model structure, high computational

efficiency, and strong interpretability. Ridge regression introduces an L2 regularization term on the basis of ordinary linear regression to improve model stability by suppressing excessively large model coefficients. Lasso regression introduces an L1 regularization term, which provides a certain degree of feature selection capability while controlling overfitting. Linear models are suitable as baseline methods for evaluating whether more complex models can significantly improve predictive performance. However, because lens processing deviations are often generated by nonlinear coupling among multiple factors, the expressive capability of linear models may be insufficient.

The second category includes instance- and kernel-based models. K-nearest neighbors regression (KNN) is an instance-based model. KNN selects the nearest K samples according to the distance between the sample to be predicted and the training samples in the feature space and computes a weighted average of their target values. This method can capture data patterns in local regions and has a certain ability to adapt to nonlinear relationships. However, when the feature dimensionality is high, or the scales of different features vary substantially, KNN is prone to the effects of dimensionality and noisy samples, and its computational cost during the prediction stage is relatively high. Support vector regression (SVR) was also included as a kernel-based nonlinear baseline.

The third category consists of ensemble learning models, including Random Forest, Extra Trees, Gradient Boosting, AdaBoost, XGBoost, LightGBM, and CatBoost [28–33]. Ensemble learning models improve overall predictive performance by combining multiple base learners and typically perform well in structured industrial data modeling. Among them, Random Forest and Extra Trees are randomized tree ensembles, which reduce model variance by constructing multiple decision trees in parallel and averaging their prediction results. Gradient Boosting, XGBoost, LightGBM, and CatBoost are boosting-based methods, which reduce bias by sequentially training multiple weak learners and progressively fitting the residuals. Since decision-tree-based models can naturally represent nonlinear relationships and high-order interactions among variables through node splitting, they are suitable for processing lens manufacturing data with complex coupling among material properties, order parameters, and process conditions.

Taking the Random Forest model as an example, suppose the model consists of $M$ decision trees, the input feature vector is $x_i$, and the prediction function of the $m$-th tree is $h_m(x_i; \Theta_m)$. The prediction output of the Random Forest can then be expressed as follows:

$$\hat{y}_i = \frac{1}{M} \sum_{m=1}^{M} h_m(x_i; \Theta_m) \quad (1)$$

where $\Theta_m$ denotes the set of random parameters generated by sample sampling and feature sampling during the training process of the $m$-th tree. Random Forest constructs different training subsets through bootstrap sampling and randomly selects a subset of candidate features during node splitting, thereby increasing the diversity among trees and reducing the model's sensitivity to individual training samples or local noise.

The fourth category consists of neural network models, namely the Multilayer Perceptron Regressor (MLP). The MLP establishes complex mapping relationships between input features and target variables through multiple layers of nonlinear transformations and, in theory, has strong function approximation capability. However, in industrial structured-data scenarios, the performance of neural network models often depends on sufficient data volume, a reasonable network architecture, and stable hyperparameter settings. When the samples exhibit distribution drift, noise, or category imbalance, neural networks may suffer from overfitting or unstable training.

All 2245 primary records are sorted by verified date/time, using worksheet row order only to break ties. Complete dates form approximately 60/20/20 chronological partitions: 1356 training records from 24 November 2025 to 9 April 2026, 484 validation records from

10 to 15 April 2026, and 405 final-test records from 16 to 25 April 2026. No recorded day crosses a partition. Numerical inputs are standardized, and the material is one-hot encoded using historical training data only; unknown categories are ignored by the encoder. Model and controller settings are frozen before final-test replay. Predictors are then refit on the earlier training and validation records, which form one development cohort of 1840 records. The nine final-test dates are the resampling units for paired uncertainty estimates, preserving all lens records within each sampled date.

Model performance is evaluated using the coefficient of determination $R^2$ and the root mean square error (RMSE) as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3}$$

where $y_i$ denotes the true deviation value of the $i$-th sample, $\hat{y}_i$ denotes the deviation value predicted by the model, $\overline{y}$ denotes the mean of the true deviation values in the test set, and $N$ denotes the number of test samples. $R^2$ reflects the model's ability to explain the variance of the target variable, with a value closer to 1 indicating stronger fitting capability. RMSE reflects the absolute level of prediction error, with a smaller value indicating that the model predictions are closer to the true values.

### *2.3. Temporal Variation Modeling: Dynamic Tracking Based on EWMA*

Static prediction can leave residual variation associated with unrecorded process conditions or changing order composition. The temporal predictability of these residuals is evaluated by chronological replay and expanding-window checks. EWMA tracks temporal changes in the prediction residuals of the AI model. The basic idea of EWMA is to apply exponentially decaying weights to historical observations, so that recent data have a greater influence on the current state estimate, while the influence of earlier historical data gradually diminishes [34]. This method can filter out high-frequency random noise while retaining low-frequency systematic drift information, making it suitable for monitoring slow drift in machining processes.

Let $e_t$ denote the prediction residual at time $t$, namely the difference between the recorded deviation before added compensation and the deviation predicted by the AI model as follows:

$$e_t = y_t - \hat{y}_t^{AI} \tag{4}$$

where $y_t$ denotes the recorded deviation before added compensation, and $\hat{y}_t^{AI}$ denotes the deviation predicted by the AI model. The recursive formula of the EWMA state variable $Z_t$ is given by the following:

$$Z_t = \alpha e_t + (1 - \alpha) Z_{t-1} \tag{5}$$

where $\alpha \in (0,1]$ is the smoothing coefficient. A larger $\alpha$ makes the system more sensitive to recent residual changes and can improve its response speed to rapid drift, but it may also amplify random noise. A smaller $\alpha$ produces a smoother estimate and enhances noise resistance but responds more slowly to sudden state changes. Therefore, the selection of $\alpha$ requires a trade-off between dynamic responsiveness and stability.

EWMA corrects the static prediction residual defined in Equation (4); pure EWMA instead tracks raw S/C deviations. For a span $L$, the smoothing coefficient is $\frac{2}{L+1}$. S/C spans are searched over {5, 10, 20}; fusion uses an AI weight of 1 and an EWMA weight in {0.5, 1}. Validation joint pass count is maximized, with summed S/C MAE breaking ties. Historical residuals initialize the states from zero; global tracking uses all materials, whereas

material tracking updates only the matching category. A material without an initialized state falls back to the current global state. All orders sharing a recorded timestamp are predicted before any result at that timestamp is revealed; their subsequent updates follow stable source-row order. No test outcome is used for model fitting or controller selection. This schedule assumes that each completed timestamp group becomes available before the next group; additional inspection latency is not simulated in the primary analysis. For the robustness replay, initialization uses residuals computed by predictors fitted to the full development cohort.

### *2.4. AI + EWMA Hybrid Compensation Mechanism*

After obtaining the static prediction result from the AI model and the dynamic drift estimate from EWMA, this paper fuses the two to form the ungated in-process compensation value. For the order to be processed at time $t + 1$, the system first uses the AI model to predict its potential machining deviation $\hat{y}_{t+1}^{AI}$ based on the order features and uses the EWMA state variable $Z_t$ up to time $t$ to characterize the current systematic drift of the manufacturing system. The ungated compensation value is defined as follows:

$$C_{t+1} = -\left(\gamma \hat{y}_{t+1}^{AI} + \beta Z_t\right) \tag{6}$$

where $C_{t+1}$ denotes the ungated compensation value for the $(t + 1)$-th order, $\gamma$ denotes the weight of the AI prediction term, and $\beta$ denotes the weight of the EWMA dynamic tracking term. The negative sign in the formula indicates that the compensation direction is opposite to the predicted deviation direction. That is, when the system predicts a positive deviation in a certain direction, the compensation value should be adjusted in the opposite direction to offset the expected error.

The static term represents order-dependent prediction and the EWMA term represents sequential residual correction. The fused controller is compared with static-AI and pure-EWMA baselines to evaluate the incremental value of residual adaptation.

Replay adds the applicable command to the recorded deviation, assuming a unit response between command and optical-power correction. For each target, let $r_t = y_t + C_{\text{final,t}}$ denote the compensated replay deviation, where $C_{\text{final,t}}$ is the command after any gate scaling; without a gate, it is the command in Equation (6). EWMA uses the prediction residual in Equation (4), whereas the gate uses previous compensated deviations $r_t$. In prospective deployment, the applied command effect must be removed from the inspected deviation before forming Equation (4); under unit gain, this gives $y_t = r_t - C_{\text{final,t}}$. The S/C commands specify optical-power increments in dioptres. Machine application requires a mapping from these increments to surface-design or process-parameter changes, calibrated for the relevant material and prescription range. A local response model is $\boldsymbol{r} = \boldsymbol{y} + \boldsymbol{Gc}$, where $\boldsymbol{y}$, $\boldsymbol{r}$ and $\boldsymbol{c}$ are two-component S/C vectors and $\boldsymbol{G}$ is a dimensionless response matrix. The primary replay uses the identity matrix; diagonal changes represent gain mismatch, and off-diagonal terms represent S/C coupling. Sensitivity calculations vary these matrix entries under assumed responses. The matrix entries are simulation parameters; applying this model to a machine requires experimental calibration of both the optical response and its relation to machine settings. The baseline is the recorded production outcome under existing practice, including any pre-existing correction. Replay evaluates the candidate additional correction relative to this baseline.

### *2.5. Material-Specific Tracking and Proposed Safety Gate*

In the global AI + EWMA strategy, all materials share the same EWMA state variable. This design captures pooled residual changes, while material-specific residual patterns may differ. Lens materials may differ substantially in refractive index, coefficient of thermal expansion, shrinkage rate, internal stress-release characteristics, and processing stability. If

all materials share the same EWMA state, residual changes in one material may interfere with the compensation of other materials, potentially reducing compensation accuracy for some materials.

To address this issue, this paper proposes a material-specific EWMA fine-grained tracking strategy. Let the material code be $k$. The system independently maintains an EWMA state variable $Z_{t,k}$ for each material, and its recursive formula is given by the following:

$$Z_{t,k} = \alpha e_{t,k} + (1 - \alpha) Z_{t-1,k} \tag{7}$$

where $e_{t,k}$ denotes the prediction residual of the order corresponding to material $k$ at time $t$. By establishing independent residual sequences and state variables for different materials, the system tracks material-level residual patterns separately rather than combining them in a single global state. States for other materials remain unchanged when a matching-material residual is updated.

The three-level gate monitors the absolute value of the mean of the most recent $W$ available compensated deviations for each material and each target. Validation compares $W \in \{5, 10, 20\}$, a first threshold in $\{0.02, 0.04, 0.06, 0.08\}$ D, and a second threshold equal to twice the first. The selected settings are $W = 20$, a first threshold of 0.04 D and a second threshold of 0.08 D. The primary replay starts gate histories empty at the evaluation boundary, and the gate remains FULL until $W$ matching-material observations have been revealed. EWMA states are initialized from historical records, whereas gate histories contain only compensated deviations generated during evaluation. This reset defines a monitoring episode that does not inherit earlier compensation commands. A continuous-state policy instead retains prior compensated deviations across predictor refitting. Window length and thresholds are tuned to the validation joint-pass objective. The signed mean responds to persistent bias, while opposite-signed errors can cancel even when individual errors are large. Evaluation of false-alarm behavior and anomaly protection requires a defined disturbance process. Initial attenuation and absolute-error or exceedance-based triggers are alternative monitoring designs for further investigation.

The gate weight $w_k$ of material $k$ is defined as follows:

$$w_k = \begin{cases} 1.0, & |\bar{r}_k| \le \tau_1 \\ 0.5, & \tau_1 < |\bar{r}_k| \le \tau_2 \\ 0.0, & |\bar{r}_k| > \tau_2 \end{cases} \tag{8}$$

In Equation (8), $\bar{r}_k$ denotes the mean of previously revealed compensated deviations, not the prediction residual in Equation (4), and is computed separately for S and C. Time and target indices are suppressed in Equations (8) and (9). FULL applies when its absolute value is at most 0.04 D; REDUCED applies above 0.04 D and at most 0.08 D; OFF applies above 0.08 D. The corresponding weights are 1, 0.5, and 0. The same timing rule used for EWMA also applies to the gate, including withholding all tied-timestamp outcomes until their predictions have been issued.

The final compensation value after safety-gate scaling is given by the following:

$$C_{final} = w_k C_{raw} \tag{9}$$

The gate scales the fused command as in Equation (9), with $C_{\text{raw}}$ given by Equation (6) and independent weights for S and C. Reducing the gate weight attenuates the applied correction, and OFF suppresses it.

### 2.6. Robustness Analysis

For robustness analysis, a separate implementation refits the core predictors and repeats sequential replay on the same cohort, using the six core inputs and Random Forest/CatBoost

settings of the primary analysis. Numerical scaling, one-hot encoding, and predictors are fitted on training data during validation and refitted on all 1840 development records before final testing. EWMA states start at zero and are warmed with residuals calculated from these refitted predictors. The warm-start residuals are in-sample because their records also contribute to predictor fitting. Pure EWMA instead uses historical measured deviations. Gate histories start empty, and all predictions within a timestamp group precede release of that group's outcomes. S/C spans are 10/20 for pure EWMA, 5/5 for global fusion, and 5/10 for material tracking; the corresponding EWMA weights are 1, 0.5, and 0.5. These settings remain fixed during the final-test sensitivity scenarios. The computations use seed 42, scikit-learn 1.7.2, and CatBoost 1.2.8. Paired intervals and response-sensitivity calculations use the trajectories generated by this protocol.

For sequential replay, preprocessing and candidate predictors are first fitted on training records. Predictors and controller settings are selected using validation data, after which preprocessing and the selected predictors are refitted on the combined development cohort. EWMA states are initialized from historical prediction residuals, or recorded deviations for pure EWMA. The nominal gate starts with empty histories. For unseen materials, the global state is used to generate commands until a material-specific state becomes available; the material-specific state is initialized to zero immediately before its first update. Before each timestamp group is processed, eligible feedback is released. EWMA states are updated using command-corrected deviations, with the saved prediction subtracted for AI-based tracking, while gate histories are updated using compensated deviations. All commands within a group are issued before any outcomes from that group are released, and outcomes are queued according to the specified feedback delay. Predictors and controller settings remain fixed throughout testing. Endpoint summaries and paired intervals are calculated from the completed trajectories.

State-continuity sensitivity compares four combinations of EWMA initialization and gate history at the final predictor refit. A training-fitted pipeline first predicts the 484 validation records; replay uses the fixed material-controller settings and a training-fitted residual warm start. Its last timestamp group's feedback is released before the first test command. Final testing then uses the common development-refitted predictor, with either development-fitted residual initialization or the EWMA states carried from validation, and with either empty gate buffers or retained validation buffers. Retained gate values are the compensated deviations generated by the training-fitted predictor during validation replay. All four combinations use unit response and the same feedback schedule, spans, weights, and thresholds; none is reselected on test outcomes. This comparison evaluates state transfer across predictor refitting while retaining the fitted-residual start of the training history.

Paired percentile intervals use 2000 bootstrap samples of the nine test dates, retaining all orders within each sampled date and scoring all strategies on the same sample. Predictors, commands, and replay trajectories remain fixed during resampling, so the intervals describe variation across scored dates conditional on the fitted models and realized controller paths. Leave-one-date-out checks omit one date from scoring while retaining those paths. Interval stability is examined at 1000 and 10,000 resamples with seed 42. Upper-tail error is the 95th percentile of $\max(|S|, |C|)$. Material-specific intervals resample the dates observed for each category. Intervals are pointwise estimates at the nominal 95% level for individual contrasts. Tolerance sensitivity scores the same frozen replay trajectories at six joint comparison bands; neither the predictors nor the controller settings are reselected for each band.

Response sensitivity uses $\boldsymbol{r} = \boldsymbol{y} + \boldsymbol{G}\boldsymbol{c}$ with common S/C gains of 0.50, 0.75, 1.00, 1.25, and 1.50, or unit diagonal gains with a symmetric off-diagonal coupling of $\pm 0.20$. The controller retains its unit-gain assumption and reconstructs uncompensated feedback as

$r - c$ before updating residuals. Perturbations begin at the start of the test interval after nominal historical initialization. Delay scenarios withhold outcomes for 1, 5, or 10 additional distinct timestamp groups, and each group is predicted before its tied outcomes are released. The gain and coupling values are specified simulation parameters, and delays are expressed in timestamp groups. These scenarios characterize the effect of assumed response and feedback schedules. Controller settings remain fixed across scenarios.

For the expanding-window robustness analysis, the 37 dates form six consecutive blocks of 7, 6, 6, 6, 6, and 6 dates. Each future block uses only preceding blocks. The earliest 80% of historical dates, rounded down to a whole number of dates, train validation-stage predictors; the remaining dates select spans and weights using the grids and objective in Section 2.3. Pure EWMA fixes its weight to 1. The gate window and threshold are selected at the retained material settings using Section 2.5. Preprocessing and predictors are then refit on the full available history; fitted residuals warm EWMA, gate histories start empty, and fixed offsets use the historical means. All strategies share the same dates and feedback schedule.

## 3. Experiments and Analysis

### *3.1. Dataset Construction and Feature Engineering*

The source dataset contains 2745 production rows from optical lens manufacturing company A. Inclusion requires complete front curve, ADD, ordered and measured SPH/CYL, and material, with both absolute recomputed deviations below 1 D. S and C are calculated as measured minus ordered power within the same row. Sixty-nine rows have incomplete required inputs (66 front-curve, two ADD and one material value), and seven complete rows have at least one absolute deviation of 1 D or more. The two exclusion groups do not overlap, leaving 2669 eligible records; no exact duplicates are removed. Of these, 424 have dates recorded without a year and are assigned to the auxiliary cohort. The primary chronological cohort contains the other 2245 records, spanning 24 November 2025 to 25 April 2026 and 18 material categories. The 1.60 PC blue-light-filtering product is encoded as a separate category alongside the numbered material codes. All 69 incomplete rows and four large-error rows are dated before validation; the other three large-error rows have dates recorded without a year. All dated observations in the validation and final-test periods satisfy the two inclusion rules. The 1 D cutoff is an outcome-based inclusion criterion. Across all raw rows, S/C MAEs are 0.0798/0.0744 D, compared with 0.0737/0.0684 D after these restrictions. Table 1a,b summarizes the exclusions and partitions. Mean ordered S/C powers change from −1.2893/−0.6794 D to −1.2794/−0.6824 D. Before and after filtering, ordered S has a median of −0.75 D and a 25th–75th percentile interval of −3.75 to +0.75 D; the corresponding C values are −0.52 D and −1.00 to 0 D. The largest material-share change is 0.32 pp (category 12). S/C 95th absolute-error quantiles are 0.1900/0.1700 D in both sets, although the excluded large deviations occupy the extreme tails.

All retained partitions have complete required predictors and targets. After terminal R/L suffixes are removed, no nonblank order group crosses a partition. This identifier-based check covers 2236 of the 2245 records; nine have blank order identifiers (four training, four validation and one final-test record). No exact duplicates are removed.

The six primary predictors are front curve, ADD, curve–power difference, ordered SPH, ordered CYL and material. ADD is recorded on two separate scales: values above 10 are divided by 100, whereas smaller values are retained in dioptres. This harmonization is inferred from the separated centidioptre and dioptre ranges; a no-ADD ablation is reported as a sensitivity analysis. Material identifiers are categorical and preserve the workbook codes. Numeric refractive indices for exploratory proxies are taken from the workbook material key. Diameter is not measured and is excluded from primary prediction. The

workbook material key spans nominal refractive indices 1.56–1.74. Codes distinguish coating and product variants as well as index, so categorical effects cannot be attributed solely to bulk material properties. Prescription composition, manufacturing period, and unrecorded equipment state may confound these associations. Recorded variants include clear, photochromic, polarizing, and blue-light-filtering products. Codes 1–5 and 17 have a nominal index of 1.56; codes 6–11 have a nominal index of 1.60; code 12 has a nominal index of 1.61; codes 13–15 have a nominal index of 1.67; and code 16 has a nominal index of 1.74. The separate 1.60 PC text category occurs in two training records.

**Table 1.** (**a**). Cohort construction and exclusion reasons. (**b**). Primary chronological partitions. (**c**). Baseline optical-power deviations in the final chronological test cohort.

(a)

| Stage or reason | Records | Details |
|---|---|---|
| Raw data-area rows | 2745 | First worksheet; header excluded |
| Excluded: incomplete required fields | 69 | Front curve 66; ADD 2; material 1 |
| Excluded: absolute S or C deviation ≥ 1 D | 7 | S only 3; C only 2; both 2 |
| Overlap of incomplete and deviation exclusions | 0 | No overlap |
| Eligible records | 2669 | Complete required fields and both deviations < 1 D in magnitude |
| Excluded from chronology: year absent | 424 | Month/day labels without a calendar year |
| Primary chronological cohort | 2245 | 37 recorded dates; 18 recorded material categories |

(b)

| Partition | Recorded date range | Records | Dates | Materials | Missing order IDs |
|---|---|---|---|---|---|
| Training | 24 Nov 2025–9 Apr 2026 | 1356 | 23 | 17 | 4 |
| Validation | 10–15 Apr 2026 | 484 | 5 | 17 | 4 |
| Final test | 16–25 Apr 2026 | 405 | 9 | 15 | 1 |

(c)

| Metric | S-deviation | C-deviation |
|---|---|---|
| Mean (D) | 0.0743 | −0.0379 |
| MAE (D) | 0.0876 | 0.0654 |
| RMSE (D) | 0.1107 | 0.0823 |
| Sample SD (D) | 0.0822 | 0.0732 |
| Marginal pass count/$n$ | 267/405 | 327/405 |
| Marginal pass rate | 65.93% | 80.74% |
| Joint S AND C pass | 231/405 (57.04%) | |

Prescription distributions and category frequencies varied across materials and chronological partitions. Across the full cohort, materials 1 and 13 had median ordered spherical powers of 0.00 and −3.75 D and median addition powers of 1.50 and 0.75 D, respectively. Material 13 accounted for 11.43% of training records and 3.70% of test records, while its median ordered spherical power shifted from −1.50 to −5.50 D. These prescription and temporal imbalances complicate attribution of category-dependent prediction performance to intrinsic material properties.

Cooling time is evaluated in the separate 424-record subset with dates recorded without a year. Excluding two records with ambiguous date labels leaves 422 auxiliary records. A grouped 80/20 holdout with seed 42 produces 336 training and 86 test records; order

identifiers with terminal R/L removed define groups, and no group crosses this split. This group-disjoint holdout evaluates the incremental predictive value of recorded cooling time.

The primary curve–power difference is front curve minus ordered SPH. The exploratory 19-input models add 13 explicitly defined optical proxies to the six base features. These models use the same primary chronological partitions.

Table 1c reports the recorded final-test deviations before additional compensation for the same 405 orders used in the primary strategy comparisons. Baseline S/C MAEs are 0.0876/0.0654 D, and their marginal pass rates are 65.93%/80.74%. Joint pass requires both $|S| \leq 0.10$ D and $|C| \leq 0.10$ D for the same lens and is 231/405 (57.04%). The common $\pm 0.10$ D band serves as an analysis criterion for comparing compensation strategies. Recorded focimeter SPH and CYL characterize spherical and cylindrical power deviations at the measurement location. For power-variation lenses, ISO 8980-2 [2] specifies distinct verification methods for back-vertex power, cylinder-axis direction, prismatic power, and variation power, including addition power. Comprehensive optical assessment additionally requires spatial power distributions. Measurement-system qualification requires instrument identification, specified lens reference locations and orientation, traceable calibration checks, repeated measurements with repositioning, and documented environmental conditions. A corresponding uncertainty budget is needed to quantify classification uncertainty relative to the selected tolerance. Between-lens deviations contain both manufacturing and measurement variation and cannot substitute for repeatability measurements. The present endpoint therefore quantifies agreement of the recorded S/C powers with their ordered values. Six comparison bands and continuous and upper-tail error summaries characterize compensation performance. For fixed-offset and robustness analyses, the pass criterion includes a numerical tolerance of $10^{-12}$ D to accommodate floating-point representation of values at the boundary.

Figure 3 shows the marginal distributions of S-deviation (a) and C-deviation (b) for the same 405 final-test records as Table 1, using 0.02 D bins. Gray dashed lines mark the study-specific $\pm 0.10$ D comparison limits, and black dotted lines mark the means. The mean S-deviation is +0.0743 D and the mean C-deviation is −0.0379 D. The opposite-signed offsets motivate target-specific prediction and correction.

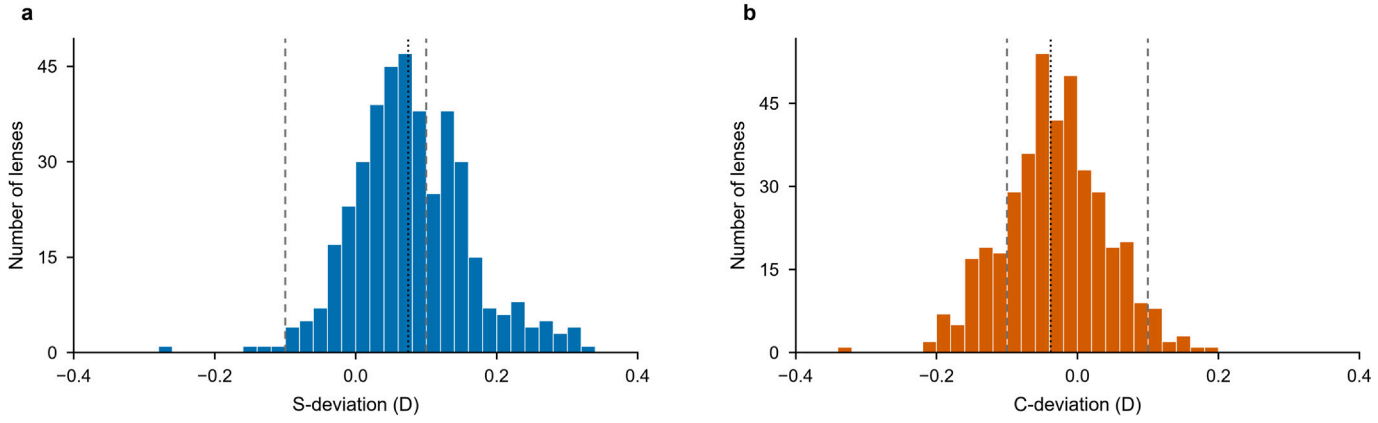



**Figure 3.** Final-test optical-power deviations. (**a**) Distribution of spherical power deviation (S-deviation); (**b**) Distribution of cylindrical power deviation (C-deviation). Gray dashed lines indicate the study-specific $\pm 0.10$ D comparison limits, and black dotted lines indicate the means.

### *3.2. Performance Evaluation of Base Models*

Thirteen static-model families share primary data partitions and preprocessing. Fixed-seed settings are: as follows Random Forest and Extra Trees, 200 trees, maximum depth 10 and minimum leaf size 3; CatBoost, 200 iterations, depth 6 and learning rate 0.05; XGBoost,

200 trees, depth 4 and learning rate 0.05; LightGBM, 200 trees, depth 5, 20 leaves and learning rate 0.05; Gradient Boosting, 200 trees, depth 3 and learning rate 0.05; AdaBoost, 100 estimators and learning rate 0.05; Ridge penalty 1; Lasso penalty 0.001; distance-weighted KNN with 15 neighbors; radial-basis SVR with $C = 1$ and $\varepsilon = 0.02$; and MLP with hidden widths 64/32, penalty 0.01 and 500 maximum iterations. Linear regression is unregularized. Seed 42 is used where applicable.

Table 2 lists final-test $R^2$ and RMSE for all 13 configurations, ordered by summed S/C validation MAE rather than by test performance. Random Forest and CatBoost achieve the lowest validation MAEs for S and C, respectively (0.0606 and 0.0556 D). Their corresponding final-test $R^2$ values are 0.0169 and 0.2828. S prediction is substantially less explanatory than C prediction on this future interval; negative $R^2$ values for several alternatives mean that their squared error exceeds the test-mean reference.

**Table 2.** Chronological final-test model comparison.

| Val. Order | Model | S $R^2$ | S RMSE (D) | C $R^2$ | C RMSE (D) |
|---|---|---|---|---|---|
| 1 | CatBoost | 0.0784 | 0.0788 | 0.2828 | 0.0619 |
| 2 | Random Forest | 0.0169 | 0.0814 | 0.2604 | 0.0629 |
| 3 | Gradient Boosting | 0.0790 | 0.0788 | 0.2639 | 0.0627 |
| 4 | LightGBM | 0.0555 | 0.0798 | 0.2662 | 0.0626 |
| 5 | XGBoost | 0.0850 | 0.0785 | 0.2461 | 0.0635 |
| 6 | Extra Trees | 0.0019 | 0.0820 | 0.2416 | 0.0636 |
| 7 | AdaBoost | −0.0865 | 0.0855 | 0.2639 | 0.0627 |
| 8 | KNN | −0.0445 | 0.0839 | 0.0802 | 0.0701 |
| 9 | SVR | −0.1224 | 0.0869 | −0.0070 | 0.0733 |
| 10 | Lasso | −0.0997 | 0.0861 | 0.0187 | 0.0724 |
| 11 | Ridge | −0.0850 | 0.0855 | 0.0390 | 0.0716 |
| 12 | Linear Regression | −0.0846 | 0.0855 | 0.0388 | 0.0716 |
| 13 | MLP | −0.0995 | 0.0861 | 0.0380 | 0.0717 |

Several tree ensembles perform better than the simpler baselines on this test interval, particularly for C-deviation. Predictive interactions among prescription, front curve, and material can be represented by tree models, while residual distribution change can still limit future performance. These comparisons evaluate the specified configurations under a common cohort, chronological split, preprocessing pipeline, and computational setting. The primary controller retains Random Forest for S and CatBoost for C, consistent with the target-specific validation results. These frozen choices provide a stable reference for comparing static AI, global residual EWMA, material residual EWMA, and the safety gate. Exploratory XGBoost and neural variants are evaluated separately from this validation-selected controller.

### *3.3. Feature Importance and Ablation Analysis*

Feature contributions are evaluated using permutation importance and single-feature removal with retraining. The basic idea of the ablation experiment is to remove one feature at a time from the complete feature set, retrain the model, and then observe the changes in prediction performance on the test set. If the model error increases after a certain feature is removed, it indicates that this feature contributes, under this fitted configuration, to the prediction task. In this paper, MAE is used as the primary evaluation metric for the ablation experiment, which is defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{10}$$

where $N$ denotes the number of test samples, $y_i$ denotes the true deviation value, and $\hat{y}_i$ denotes the deviation value predicted by the model. MAE directly reflects the average absolute deviation between the predicted values and the true values, making it suitable for evaluating prediction errors in machining deviations.

Figure 4 measures six-feature model reliance for Random Forest predicting S (a) and CatBoost predicting C (b) by the held-out MAE increase in D after permutation of one input, with 30 shuffles per feature on the same 405 test records. The bars show mean increases; feature order is shared, while horizontal scales differ between targets. Ordered CYL has the largest mean importance for both core targets.

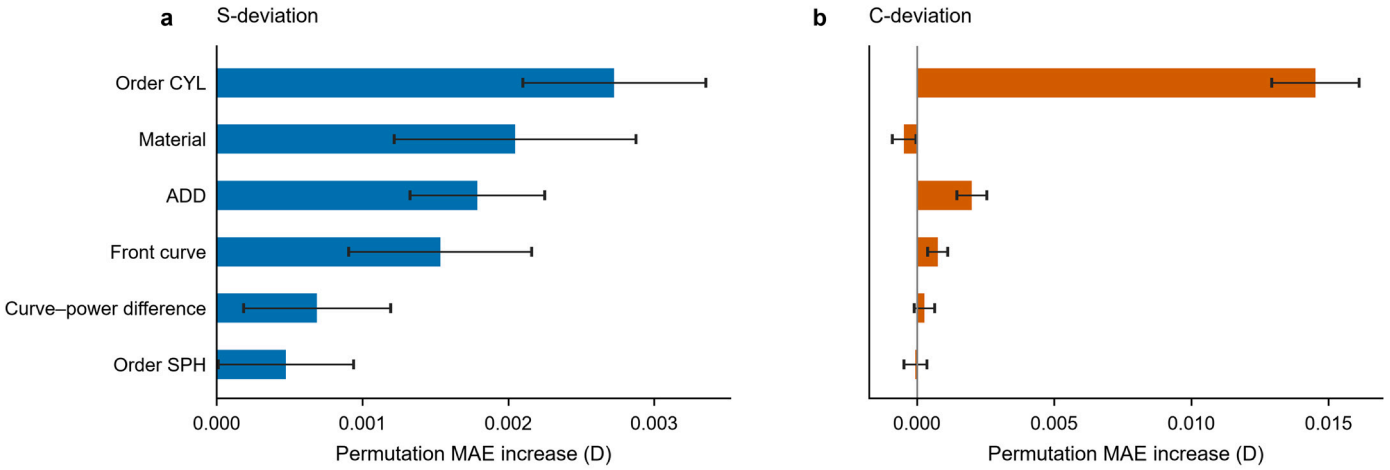



**Figure 4.** Permutation importance of the six core features. (**a**) Random Forest predicting spherical power deviation (S-deviation); (**b**) CatBoost predicting cylindrical power deviation (C-deviation).

Table 3 retrains each core predictor after removing one of its six inputs. Full-model S/C MAEs are 0.0612/0.0476 D. Removing ordered CYL increases C MAE to 0.0571 D, whereas removing material reduces S/C MAEs to 0.0593/0.0469 D on this test interval. The no-ADD sensitivity gives 0.0618/0.0487 D. Front curve, ordered SPH, and their difference are algebraically dependent: both pairs reconstruct the third. Removing one therefore does not remove the underlying optical information. The lower errors after material removal show that categorical material information adds no predictive benefit on this test interval; the encoded labels combine optical and product attributes.

**Table 3.** Single-feature removal from the six-feature core predictors on the same final test.

| Removed Feature | S MAE (D) | C MAE (D) |
|---|---|---|
| None (full six inputs) | 0.0612 | 0.0476 |
| Front curve | 0.0619 | 0.0479 |
| ADD | 0.0618 | 0.0487 |
| Curve–power difference | 0.0608 | 0.0471 |
| Order SPH | 0.0610 | 0.0476 |
| Order CYL | 0.0611 | 0.0571 |
| Material | 0.0593 | 0.0469 |

Table 4 compares six base inputs with the same inputs plus cooling time on the independent auxiliary holdout of 86 records. Cooling time increases $R^2$ for S in 12 of 13 model configurations and for C in 6 of 13; improvement is therefore not universal. For Random Forest, S $R^2$ changes from 0.0545 to 0.1282. For CatBoost, C $R^2$ changes from −0.0640 to 0.0026.

The auxiliary results suggest target- and model-dependent predictive information in cooling time. Cooling duration is observational and may covary with material, prescription, or processing route. Its causal effect on quality requires a controlled process intervention.

**Table 4.** Auxiliary cooling-time comparison.

| Model | S $R^2$ Without CT | S $R^2$ with CT | C $R^2$ Without CT | C $R^2$ with CT |
|---|---|---|---|---|
| Random Forest | 0.0545 | 0.1282 | −0.0253 | −0.0117 |
| CatBoost | 0.0372 | 0.1274 | −0.0640 | 0.0026 |
| Extra Trees | 0.0046 | 0.0836 | −0.1085 | −0.1663 |
| XGBoost | 0.0900 | 0.1162 | −0.1132 | −0.1084 |
| LightGBM | −0.0809 | 0.0186 | −0.1630 | −0.0928 |
| Gradient Boosting | 0.1061 | 0.0291 | −0.1501 | −0.0674 |
| KNN | −0.0037 | 0.0560 | 0.0146 | −0.0163 |
| MLP | −1.0328 | −0.7807 | −0.8321 | −0.2280 |
| AdaBoost | 0.0619 | 0.1012 | 0.1057 | 0.0915 |
| Ridge | 0.0365 | 0.0931 | −0.2017 | −0.2127 |
| Linear Regression | 0.0404 | 0.0978 | −0.2614 | −0.2702 |
| Lasso | −0.0190 | 0.0351 | −0.0203 | −0.0285 |
| SVR | −0.2351 | −0.1398 | −0.1323 | −0.2452 |

### *3.4. Time Series Evaluation of Compensation Strategies*

Compensation strategies are compared by joint S/C pass count, pass-rate change in percentage points (pp), and compensated-deviation MAE on the identical 405-record final test. The baseline means no additional intelligent compensation is applied to the recorded outcomes. All compensated results are obtained by retrospective replay under the unit-gain assumption. The validation-selected EWMA spans and weights are listed in Table 5, where $L_S/L_C$ denote the S/C spans and $\beta$ denotes the EWMA weight. The AI weight is 1 in AI-containing strategies and 0 for pure EWMA. Table 5 compares the five primary strategy types, the identical ungated material replay at the retained settings, and the historical fixed-offset baseline. Static AI gives 304/405 (75.06%), pure EWMA 265/405 (65.43%), global AI + EWMA 303/405 (74.81%), and material EWMA with the gate 310/405 (76.54%), against 231/405 (57.04%) at the baseline. Global fusion is 0.25 pp below static AI on this interval. Material tracking exceeds global tracking by 1.73 pp, corresponding to seven additional joint passes, but is not uniformly better at every material. Relative to static AI, material tracking has a net gain of six joint passes (1.48 pp), with S MAE lower by 0.0011 D and C MAE higher by 0.0005 D. The primary point estimates therefore show a small pass-count increase accompanied by opposite changes in the S and C errors. Gated and ungated material tracking use the same settings, isolating the effect of gate scaling. A fixed-offset command, estimated from the 1840 development records and frozen before testing, is −0.04889 D for S and +0.03319 D for C. It yields 277/405 joint passes (68.40%) and S/C MAEs of 0.0652/0.0573 D. Its gain over the baseline is 11.36 pp (95% paired date-bootstrap interval 7.89–14.29), showing the improvement obtainable from the historical mean correction alone.

At joint tolerances of 0.050, 0.075, 0.100, 0.125, 0.150, and 0.200 D, recorded outcomes yield 95, 161, 231, 271, 330, and 374 passes out of 405; fixed-offset compensation yields 122, 187, 277, 323, 356, and 386, respectively. The S/C 95th absolute-error quantiles decrease from 0.2300/0.1600 D to 0.1811/0.1419 D. Fixed-offset correction increases the pass count at all six tolerances. The fixed-offset gain is positive on eight test dates and zero on one; leave-one-date-out scored gains range from 10.44 to 12.16 pp.

**Table 5.** Primary strategy comparison.

| ID | Strategy | Settings | Joint Pass | S MAE (D) | C MAE (D) |
|---|---|---|---|---|---|
| 1 | No added compensation | No added command | 231/405 (57.04%) | 0.0876 | 0.0654 |
| 2 | Static AI | AI only | 304/405 (75.06%) | 0.0612 | 0.0476 |
| 3 | Pure EWMA | $L_S = 10$, $L_C = 20$; $\beta = 1$ | 265/405 (65.43%) | 0.0654 | 0.0592 |
| 4 | Global AI + EWMA | $L_S = 5$, $L_C = 5$; $\beta = 0.5$ | 303/405 (74.81%) | 0.0590 | 0.0487 |
| 5 | Material EWMA + gate | $L_S = 5$, $L_C = 10$; $\beta = 0.5$ | 310/405 (76.54%) | 0.0601 | 0.0481 |
| 6 | Material EWMA without gate | $L_S = 5$, $L_C = 10$; $\beta = 0.5$ | 310/405 (76.54%) | 0.0601 | 0.0481 |
| 7 | Fixed offset | Historical mean only | 277/405 (68.40%) | 0.0652 | 0.0573 |

Figure 5 shows absolute joint S/C pass rates (a) and paired changes from the baseline (b) for the five primary strategies in Table 5. Global fusion denotes global AI + EWMA, and material fusion denotes material-specific AI + EWMA with the safety gate. The global gain is 17.78 pp (95% date-bootstrap interval 13.57–23.85); the material-strategy gain is 19.51 pp (14.57–25.70). Whiskers in panel b show 95% intervals from 2000 paired bootstrap resamples of the nine test dates and preserve each sampled date's full set of orders.

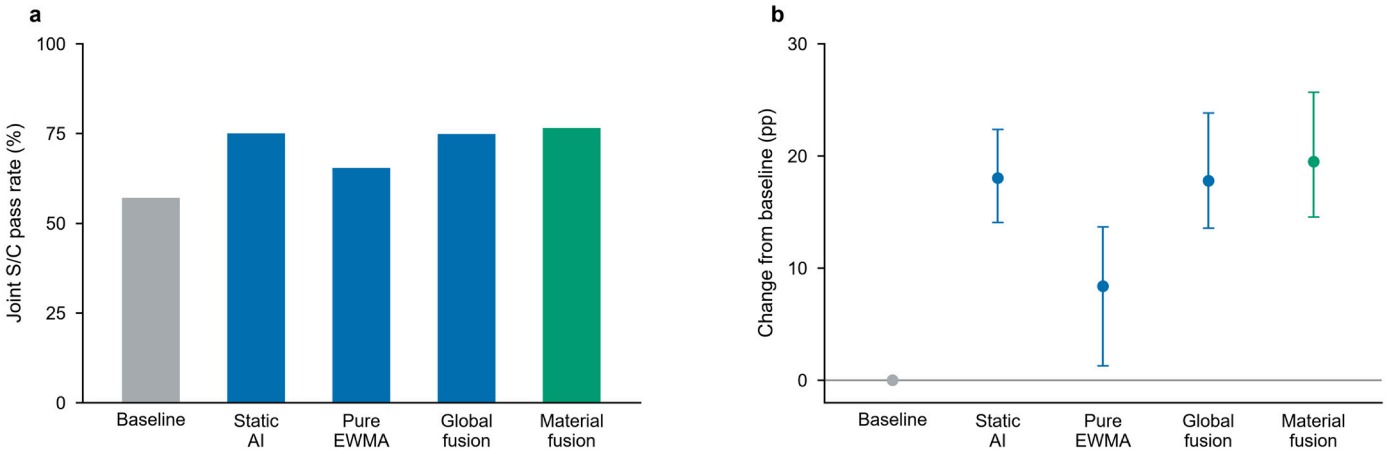



**Figure 5.** Comparison of the primary compensation strategies. (**a**) Joint S/C pass rates within the study-specific ±0.10 D band; (**b**) Changes in joint pass rate relative to baseline, expressed in percentage points (pp), with error bars indicating 95% intervals from 2000 paired bootstrap resamples of the nine test dates.

Table 6 provides five expanding-window temporal checks. The 37 recorded primary dates are divided into six consecutive date blocks; each evaluation block is predicted from earlier blocks only. An internal approximately 80/20 split of historical dates selects the controller settings in each fold, followed by refitting on all available history. Test-block sizes are 303, 211, 570, 493, and 233 records. Equal-fold mean pass rates are 56.68% at the baseline, 69.95% for global fusion, and 67.41% for material tracking with the gate. Equal-fold means assigning the same weight to each evaluation block. These checks assess internal temporal stability within the primary observation period. Weighting the five blocks by their sample sizes gives 1048/1810 (57.90%) for the baseline and 1265/1810 (69.89%) for global fusion, a gain of 11.99 pp.

Global fusion improves on the baseline in all five expanding-window checks, with gains from 7.19 to 20.17 pp. Material tracking with the gate also exceeds the baseline in every fold but is below global fusion in each fold. This contrasts with its advantage on the primary test and indicates that the relative performance of material-level adaptation varies across time series evaluation windows.

**Table 6.** Five expanding-window checks.

| Fold (Test $n$) | Baseline (%) | Global AI + EWMA (%) | Global Gain (pp) |
|---|---|---|---|
| 1 ($n = 303$) | 62.38 | 69.64 | +7.26 |
| 2 ($n = 211$) | 45.97 | 63.51 | +17.54 |
| 3 ($n = 570$) | 60.18 | 67.37 | +7.19 |
| 4 ($n = 493$) | 58.22 | 72.41 | +14.20 |
| 5 ($n = 233$) | 56.65 | 76.82 | +20.17 |
| Equal-fold mean | 56.68 | 69.95 | +13.27 |
| Sample-weighted mean | 57.90 | 69.89 | +11.99 |

### *3.5. Analysis of Material-Specific Fine-Grained Compensation Effects*

Material-level results distinguish aggregate benefit from local performance. All 15 final-test material categories are evaluated using the workbook identifiers and observed sample sizes. Material-level differences between strategies are observational associations subject to confounding by prescription composition, date, tool condition, and other unmeasured process factors.

Table 7 compares the same baseline, global, and material strategies as Table 5. The material strategy gives 310/405 (76.54%), with S/C MAEs of 0.0601/0.0481 D. Exactly 1 of 15 materials is below its own baseline under global fusion and 1 of 15 under material tracking with the gate. All 405 S weights and all 405 C weights remain FULL: REDUCED and OFF activate zero times. The material replay without gate scaling gives identical predictions, counts, and MAEs. Thus, the material-versus-global difference reflects the tracking strategy and its selected settings.

**Table 7.** Aggregate outcomes and gate activations for the baseline, global, and material strategies.

| Metric | Baseline | Global AI + EWMA | Material EWMA + Gate |
|---|---|---|---|
| Joint pass | 231/405 (57.04%) | 303/405 (74.81%) | 310/405 (76.54%) |
| Gain over baseline (pp) | 0.00 | +17.78 | +19.51 |
| S MAE (D) | 0.0876 | 0.0590 | 0.0601 |
| C MAE (D) | 0.0654 | 0.0487 | 0.0481 |
| Materials below own baseline | — | 1/15 | 1/15 |
| REDUCED/OFF activations (S; C) | Not applicable | Not applicable | 0/0; 0/0 |

Figure 6 compares global AI + EWMA and material-specific tracking with the gate using paired horizontal bars for all 15 final-test materials. Each bar gives the change from that material's own baseline in percentage points; row labels give the material identifier and sample size, and the zero line denotes no change. Material 3 changes from 10/15 at the baseline to 8/15 with global fusion and 10/15 with material tracking. Conversely, material 16 changes from 6/10 at the baseline and under global fusion to 4/10 under material tracking. The degraded category therefore changes from material 3 under global fusion to material 16 under material tracking. At these sample sizes, a single lens corresponds to 10 pp for material 16 and 6.67 pp for each 15-record category.

Figure 7 details materials 3, 16, and 13, which have the three lowest global-strategy gains in Figure 6. The grouped bars compare the baseline, global AI + EWMA, and material-specific tracking with the gate. In that order, pass counts are 10/15, 8/15, and 10/15 for material 3; 6/10, 6/10, and 4/10 for material 16; and 10/15, 11/15, and 12/15 for material 13. These counts illustrate differences among tracking strategies while the gate remains FULL.

Aggregate gains under material-specific adaptation coexist with material-level declines and changes in their ranking relative to global tracking across temporal folds.

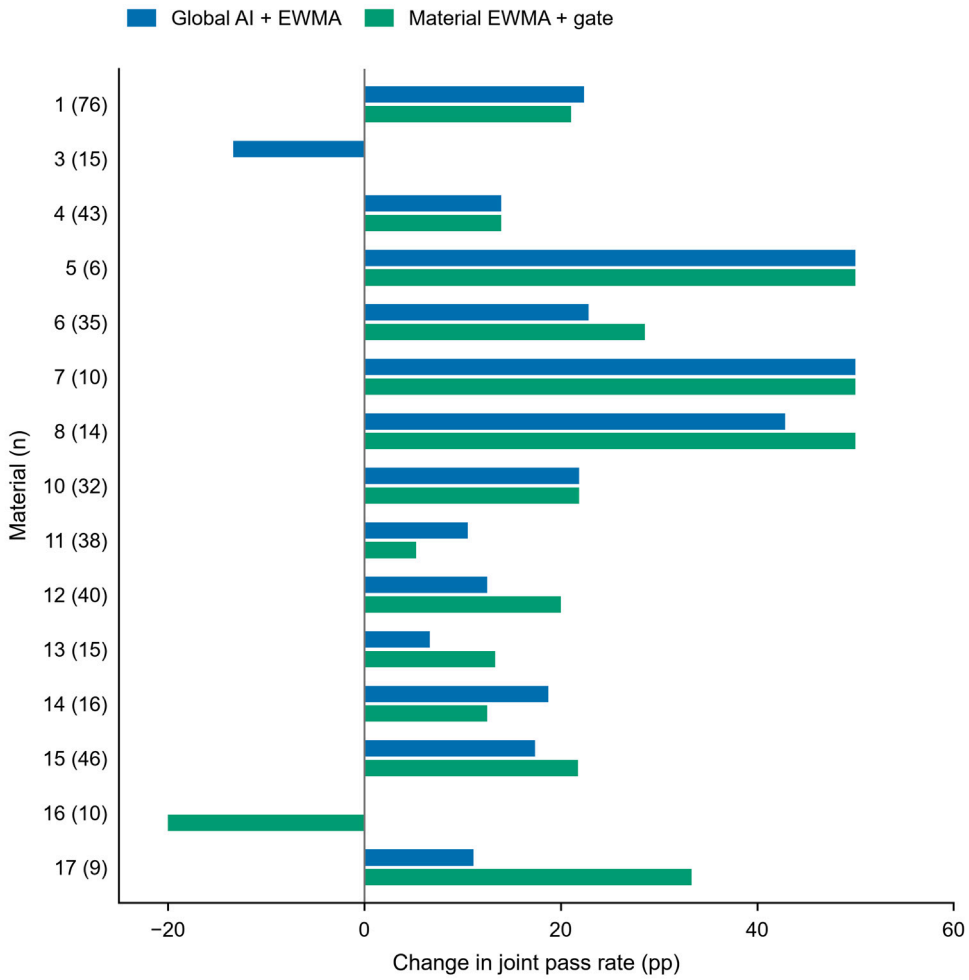



**Figure 6.** Material-level compensation gains.

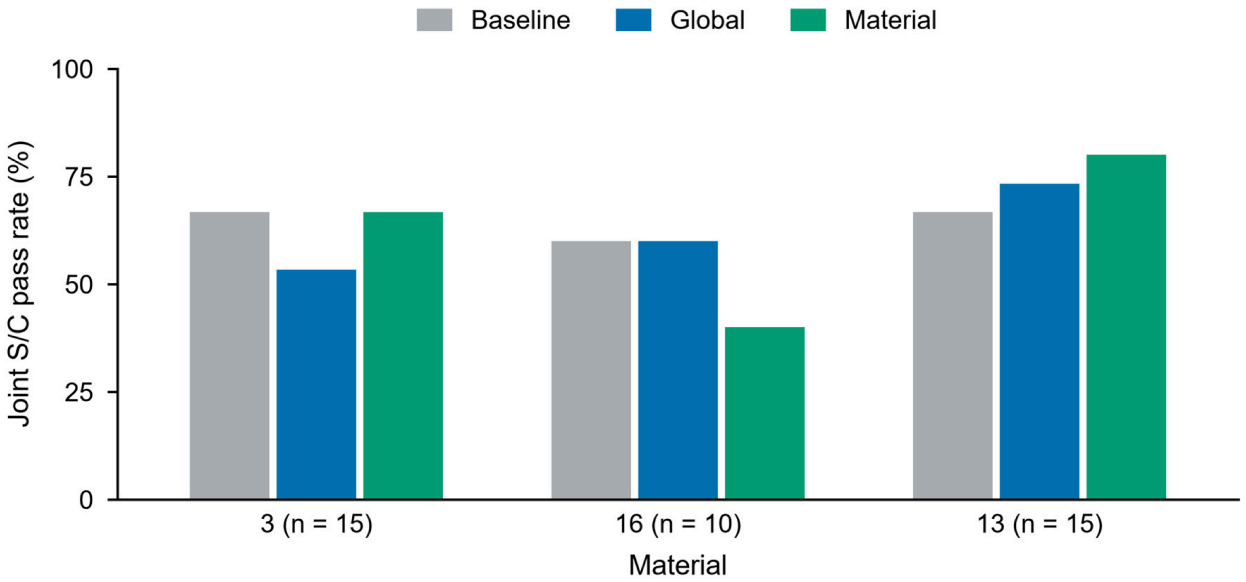



**Figure 7.** Compensation outcomes for three selected materials.

### *3.6. Robustness Results*

On the 405-record test, the robustness replay gives 303 static-AI passes, 265 pure-EWMA passes, 305 global-fusion passes, and 310 material-tracking passes; gated and ungated

material trajectories coincide. The corresponding primary static-AI and global-fusion counts in Table 5 are 304 and 303. These robustness trajectories form the basis of the paired interval and sensitivity analyses. The robustness-replay S/C MAEs are 0.0612/0.0476 D for static AI, 0.0591/0.0487 D for global fusion and 0.0601/0.0481 D for material tracking; corresponding 95th absolute-error quantiles are 0.1783/0.1221, 0.1674/0.1300 and 0.1617/0.1260 D.

Table 8 reports the first strategy minus the second; positive pass-rate differences and negative error differences favor the first strategy. All three pass-rate intervals comparing static, global and material strategies include zero. Across 1000–10,000 bootstrap resamples, their interval endpoints vary by at most 0.15 pp and continue to include zero. The material-minus-static daily difference ranges from −10.00 to +9.68 pp (five positive, three negative, and one zero date); leave-one-date-out differences range from +0.90 to +3.16 pp. Material-minus-global daily differences range from −10.00 to +3.37 pp, with leave-one-date-out differences of +0.63 to +1.64 pp. These date-level differences show temporal variation in relative ranking within the nine-date test interval.

**Table 8.** Paired differences and 95% intervals in the robustness replay.

| Contrast | Pass Rate (pp) | S MAE (D) | C MAE (D) | 95th max(\|S\|, \|C\|) (D) |
|---|---|---|---|---|
| Material tracking—static AI | 1.73 [−1.80, 4.89] | −0.0011 [−0.0030, 0.0008] | 0.0005 [−0.0006, 0.0014] | −0.0167 [−0.0248, 0.0062] |
| Material tracking—global fusion | 1.23 [−1.00, 2.50] | 0.0011 [−0.0014, 0.0030] | −0.0006 [−0.0021, 0.0006] | 0.0012 [−0.0089, 0.0195] |
| Global fusion—static AI | 0.49 [−3.07, 3.54] | −0.0022 [−0.0042, 0.0004] | 0.0011 [0.0004, 0.0020] | −0.0179 [−0.0251, 0.0068] |
| Static AI—fixed offset | 6.42 [4.07, 10.34] | −0.0040 [−0.0066, −0.0019] | −0.0097 [−0.0140, −0.0057] | −0.0020 [−0.0219, 0.0066] |

The 15 test material categories contain 6–76 records over 4–8 dates each. For material tracking in the robustness replay, date-bootstrap pass-rate intervals have widths of 10.28–75.00 pp. Material 16 has 4/10 passes over four dates, with a 95% interval of 12.5–50.0%; material 5 has 4/6 passes over four dates, with an interval of 25.0–100.0%.

The strategy ranking changes with assumed gain and feedback delay (Table 9). At gain 0.50, ungated and gated material tracking yield 280 and 276 passes, respectively, with 38 non-FULL target-order weights across S and C. At this gain, gate attenuation reduces the joint pass count by four relative to ungated tracking. With delays of 1, 5, and 10 groups, global fusion yields 301, 300, and 297 passes, compared with 305 without additional delay. These scenarios quantify sensitivity to assumed feedback delays and command responses.

**Table 9.** Gain, coupling, and delay sensitivity: joint passes out of 405.

| Scenario | Static AI | Global Fusion | Material Ungated | Material Gated |
|---|---|---|---|---|
| Nominal | 303 | 305 | 310 | 310 |
| Gain 0.50 | 264 | 286 | 280 | 276 |
| Gain 0.75 | 290 | 302 | 301 | 300 |
| Gain 1.25 | 308 | 315 | 312 | 312 |
| Gain 1.50 | 295 | 311 | 299 | 299 |
| Coupling −0.20 | 308 | 311 | 308 | 308 |
| Coupling +0.20 | 296 | 304 | 302 | 302 |
| Delay 1 group | 303 | 301 | 300 | 300 |
| Delay 5 groups | 303 | 300 | 305 | 305 |
| Delay 10 groups | 303 | 297 | 303 | 303 |

In the robustness replay (Table 10), global fusion exceeds static AI in all five blocks, with equal-fold means of 70.77% and 68.26% and sample-weighted rates of 70.99% and 68.90%, respectively. Gated material tracking has an equal-fold mean of 68.21% and a weighted rate of 69.23%, below ungated material tracking (70.10% and 70.28%). These results support a time-dependent role for residual adaptation while showing that the proposed gate can

reduce performance. Both expanding-window analyses show positive global-fusion gains over the baseline in every block.

**Table 10.** Expanding-window joint pass rates in the robustness replay (%).

| Fold ($n$) | Baseline | Fixed | Static AI | Pure EWMA | Global | Material Ungated | Material Gated |
|---|---|---|---|---|---|---|---|
| 1 (303) | 62.38 | 63.37 | 68.32 | 66.01 | 70.96 | 68.65 | 68.98 |
| 2 (211) | 45.97 | 54.98 | 58.29 | 58.29 | 63.51 | 63.98 | 54.03 |
| 3 (570) | 60.18 | 63.51 | 69.82 | 64.56 | 70.35 | 69.65 | 69.82 |
| 4 (493) | 58.22 | 63.69 | 69.78 | 64.71 | 72.21 | 71.81 | 71.81 |
| 5 (233) | 56.65 | 71.67 | 75.11 | 66.52 | 76.82 | 76.39 | 76.39 |
| Equal-fold mean | 56.68 | 63.44 | 68.26 | 64.02 | 70.77 | 70.10 | 68.21 |
| Sample-weighted | 57.90 | 63.59 | 68.90 | 64.36 | 70.99 | 70.28 | 69.23 |

Table 11 evaluates the frozen robustness trajectories at six joint tolerance bands and summarizes the distribution of the larger absolute S/C deviation. All compensated strategies have higher joint pass rates than the baseline at each tested band. The highest observed rate changes from a tie among static AI, global fusion, and material tracking at 0.050 D, to global fusion at 0.075 D and material tracking at 0.100 D; global fusion has the highest rate at 0.125, 0.150, and 0.200 D.

**Table 11.** Joint pass rates (%) and tail errors (D) in the robustness replay.

| Tolerance/Metric | Baseline | Fixed Offset | Static AI | Pure EWMA | Global Fusion | Material Tracking |
|---|---|---|---|---|---|---|
| ±0.050 D | 23.46 | 30.12 | 34.57 | 25.93 | 34.57 | 34.57 |
| ±0.075 D | 39.75 | 46.17 | 59.75 | 47.41 | 60.99 | 58.52 |
| ±0.100 D | 57.04 | 68.40 | 74.81 | 65.43 | 75.31 | 76.54 |
| ±0.125 D | 66.91 | 79.75 | 86.17 | 80.49 | 86.91 | 85.43 |
| ±0.150 D | 81.48 | 87.90 | 90.86 | 88.89 | 91.36 | 91.11 |
| ±0.200 D | 92.35 | 95.31 | 96.30 | 97.53 | 97.78 | 96.79 |
| 95th percentile (D) | 0.2300 | 0.1907 | 0.1887 | 0.1759 | 0.1707 | 0.1719 |
| Observed maximum (D) | 0.3300 | 0.3189 | 0.2955 | 0.3407 | 0.2884 | 0.2939 |

Tolerance rows report joint pass rates (%); the final two rows report the 95th percentile and observed maximum of the larger absolute S/C deviation (D). Material tracking has identical gated and ungated values in this replay.

The tail summaries complement the threshold-based comparisons. Global fusion reduces the 95th percentile of the larger absolute S/C deviation from 0.2300 to 0.1707 D and its observed maximum from 0.3300 to 0.2884 D. Pure EWMA attains 97.53% at the 0.200 D band, while its observed maximum reaches 0.3407 D. Thus, a higher pass rate at a permissive band can coexist with a larger extreme error. These comparison bands assess numerical sensitivity; optical acceptance additionally depends on applicable characteristic-specific requirements and verification methods, including ISO 8980-2 for power-variation lenses.

State inheritance changes monitoring coverage and compensation outcomes. Retaining validation gate histories increases the number of test orders with at least 20 available matching-material observations from 170/405 (41.98%) to 353/405 (87.16%). With development-fitted EWMA initialization, reset and retained gate histories yield 310/405 (76.54%) and 308/405 (76.05%) joint passes, with zero and 28 REDUCED S weights, respectively. With validation-carried EWMA states, both gate policies yield 309/405 (76.30%), with zero and three REDUCED S weights. No C weight changes and no OFF states occur. Material 16 inherits only two validation observations and remains below the monitoring window throughout its ten test orders; all four combinations give four joint passes. Increased historical coverage therefore does not imply increased compensation accuracy at the refitting boundary.

## 4. Further Discussion

### 4.1. Conditional Value of Static Prediction and Dynamic Tracking

The primary replay shows that the benefits of static prediction and residual adaptation depend on the model, time interval, and material. Static AI alone gives a higher joint pass rate than global fusion on the final test, despite the latter's lower S MAE. Joint pass requires both deviations to satisfy their limits on the same lens, so a reduction in average error need not increase the joint pass count. Continuous prediction error and joint threshold performance therefore capture different aspects of compensation.

The core ensembles use six order-derived and material inputs to estimate order-specific deviations. Their held-out explanatory power is limited, particularly for S. EWMA tracks sequential residual offsets beyond the frozen predictor; these offsets can reflect equipment state, changing order mix, or static-model error.

Pure EWMA improves the primary baseline while remaining below static AI and global fusion. Residual EWMA acts as an empirical correction mechanism whose smoothing span controls the balance between responsiveness and stability. Its initialization, material support, and observation delay influence its deployment behavior.

Material tracking improves the primary aggregate rate, while global tracking performs better in all five primary temporal checks. Material-specific states use separate observation histories and are updated only when matching-material inspections become available. The variation across windows therefore supports selecting adaptation granularity using historical validation data while considering the available material-specific history.

### 4.2. Exploratory Optical Proxies and Heuristic Regularization

Following the core controller analysis, exploratory models test whether explicit optical features or prior-based regularization improve the same prediction and replay tasks. These comparisons retain the primary chronological split and distinguish optical-feature expansion from the addition of an optical-prior penalty.

Physics-informed learning can incorporate scientific structure through features or regularization [35]. The exploratory models here use algebraic optical proxies and heuristic penalties. Their physical content is confined to the specified feature relationships and regularization terms; governing thermal, mechanical, and optical-field equations are not included.

Thirteen optical proxies augment the six base inputs. Let ordered sphere, ordered cylinder, front curve, normalized addition, and nominal material refractive index be denoted by $P$, $Q$, $F$, $A$ and $n$, respectively. The power- and radius-related proxies are spherical equivalent $\mathrm{SE} = P + \frac{Q}{2}$; front-radius proxy $R_1 = \frac{1000(n-1)}{F}$; back-power proxy $B = F - \mathrm{SE}$; back-radius proxy $R_2 = \frac{1000(n-1)}{B}$; and curvature contrast $\left|\frac{1}{\mathrm{clip}(R_1)} - \frac{1}{\mathrm{clip}(R_2)}\right|$.

The remaining proxies are the thickness proxy $\frac{\mathrm{SE}\times 35^2}{8000(n-1)}$; absolute power sum $|P| + |Q|$; CYL–curve interaction $|Q|F$; ADD–index interaction $An$; index–curve interaction $(n-1.5)F$; front-curve offset $F - \left(\frac{\mathrm{SE}}{2} + 6\right)$; squared-offset proxy $n\left[F - \left(\frac{\mathrm{SE}}{2} + 6\right)\right]^2$; and nominal refractive index $n$.

For the radius proxies, a denominator magnitude $\leq 0.01$ results in a finite sentinel value of 9999 for the corresponding radius proxy; clip limits are 1 and 9999 for the reciprocal contrast. The thickness proxy uses a fixed 35 mm reference length rather than measured lens geometry. This signed proxy is expressed in millimeters and follows a paraxial curvature approximation at the fixed reference scale; its sign follows the spherical equivalent. These algebraic proxies describe order-based optical relationships and are not measurements of stress, thermal shrinkage, or thickness. Feature counts refer to original fields before one-hot expansion of material. Neural comparisons include six- and 19-input MLPs without the

optical-prior penalty, and three optical-prior variants (ordinary, strong and deep). $P$, $Q$, $F$ and $A$ are expressed in dioptres, whereas $n$ is dimensionless. The factor 1000 expresses the radius proxies in millimeters, and their reciprocal contrast in inverse millimeters; the CYL–curve interaction and the squared-offset proxy have units of D squared. The additive constant 6 in the offset is on the dioptre scale. Together, the fixed reference scale, finite sentinels, and clipping define an algebraic feature construction whose relationship to lens geometry and thermomechanical response requires physical calibration. The training objective is as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}}(\mathrm{y}, \hat{\mathrm{y}}) + \lambda \mathcal{L}_{\text{Physics}}\left(\hat{\mathrm{y}}, \mathrm{X}_{\text{physics}}\right) \tag{11}$$

In Equation (11), $\mathcal{L}_{\text{MSE}}$ is the mean squared error between recorded deviations $y$ and predictions $\hat{y}$; $\mathcal{L}_{\text{Physics}}$ is the heuristic prior term based on predictions and proxy features $X_{\text{physics}}$, and $\lambda$ controls its weight. The heuristic prior term equals 0.5 times the squared batch-mean prediction, minus 0.1 times the correlation between absolute prediction and the absolute standardized power-sum proxy, minus 0.1 times its correlation with the absolute standardized front-curve offset for S or the absolute standardized CYL–curve interaction for C, plus 0.01 times the mean adjacent-batch prediction-difference square divided by squared feature distance plus $10^{-6}$. Correlations use a $10^{-10}$ variance-product stabilizer. These terms encode preferences for prediction mean, feature association, and feature-space smoothness. When manufacturing deviations have a systematic offset, the mean penalty can suppress the bias that compensation needs to learn. The correlation and smoothness terms are empirical regularizers with numerical weights. The penalty coefficients are empirical scale weights for deviations expressed in dioptres and standardized proxy features.

Prior weights are 0.1, 0.5, and 0.1 for ordinary, strong, and deep variants, respectively. Hidden widths are 128/64/32, or 256/128/64/32 for the deep variant, with batch normalization, ReLU, and dropout at 0.2. Adam uses a learning rate of 0.001, a weight decay of 0.0001, a batch size of 64, and a 200-epoch cap. Validation MAE controls learning-rate reduction after 10 non-improving epochs and stops after 30 non-improving epochs. Final development refits use the selected epoch count with a fixed learning rate of 0.001. Seed 42 is shared. The regularized variant for material analysis is selected by summed S/C validation MAE.

Figure 8 reports permutation importance for a separate 19-feature XGBoost predictor (300 trees, depth 6, learning rate 0.05). Panels a and b show S and C, respectively; points and whiskers give the mean MAE increase in D and $\pm 1$ standard deviation across 30 feature permutations on the same 405 test records. Features are ranked independently by target, with different horizontal scales. The index–curve interaction has the largest mean S importance, and the CYL–curve interaction has the largest C importance. Some proxy importances are near zero or negative; the rankings describe model reliance on correlated algebraic inputs.

The validation-selected regularized variant is the Optical-prior MLP, whose global-EWMA replay gives 275/405 (67.90%). Figure 9 shows all 15 test materials, including negative changes for materials 3, 13, and 16. Equal-sized open and filled markers show the baseline and compensated joint S/C pass rates, respectively, with a thin line connecting each pair. Row labels give material identifiers and sample sizes. Its aggregate pass rate is below that of both neural alternatives without the optical-prior penalty. The heterogeneous material results motivate stronger mechanistic constraints and independent validation. Related precision-engineering work on the geometrically nonlinear design of compliant amplification mechanisms illustrates geometry-dependent mechanical modeling [36]; comparable mechanistic detail in lens compensation would require optical and material-specific formulations. Separating the contribution of regularization from architecture and training variability requires matched architectures and tuning budgets evaluated across repeated seeds.

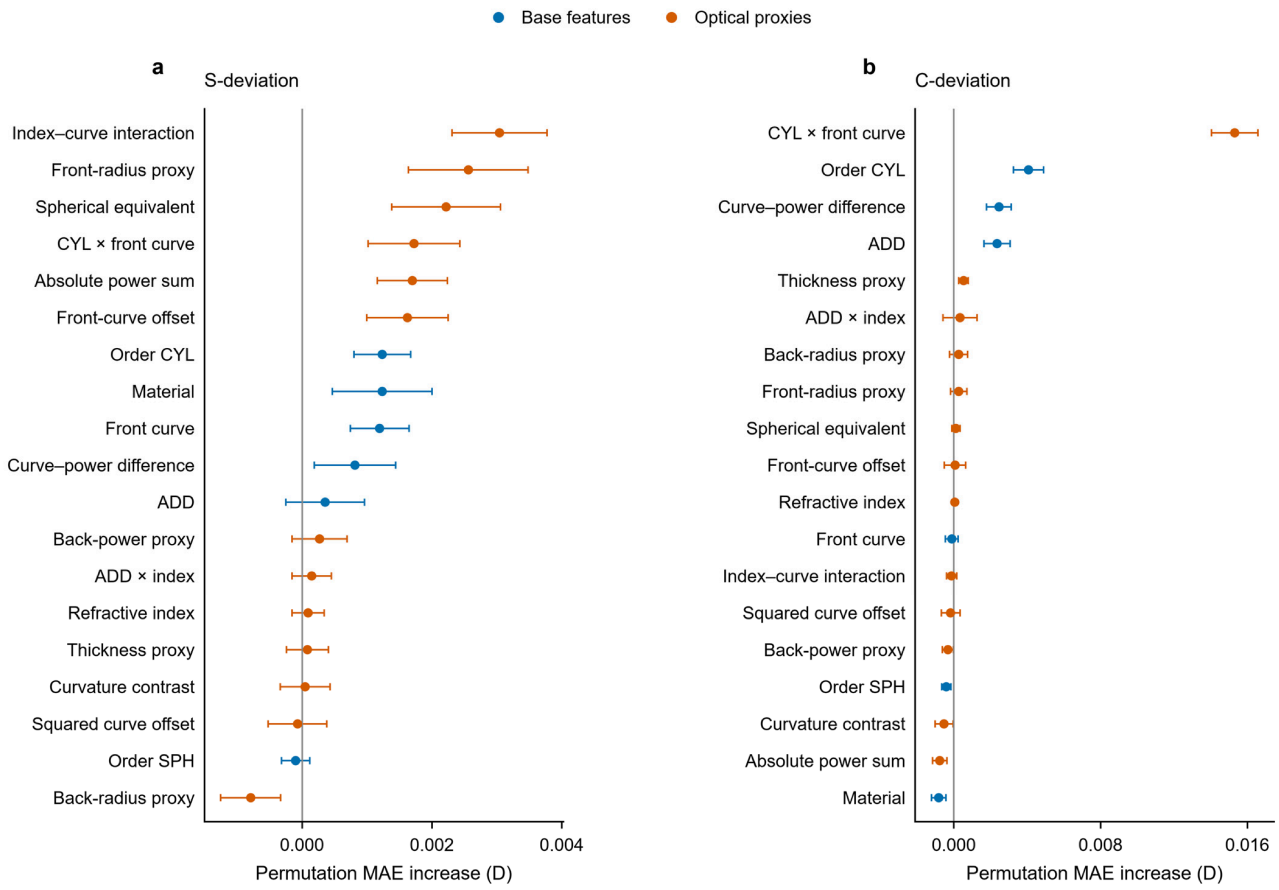



**Figure 8.** Permutation importance in the 19-feature XGBoost models. (**a**) Prediction of spherical power deviation (S-deviation); (**b**) Prediction of cylindrical power deviation (C-deviation).

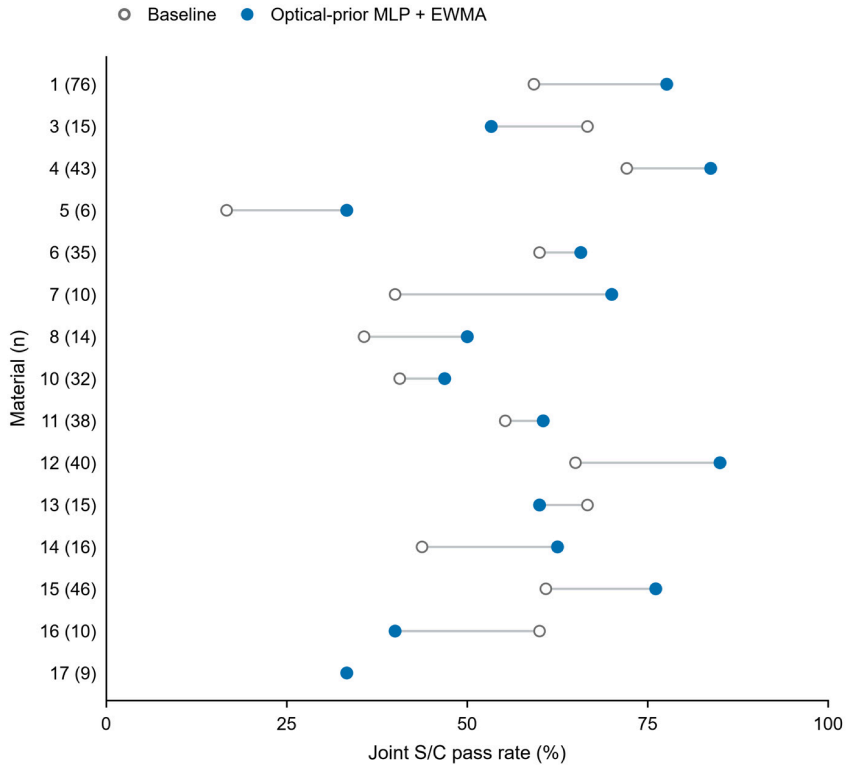



**Figure 9.** Material-level performance of the optical-prior MLP with EWMA.

### 4.3. Exploratory Transfer Learning for Rare Materials

Limited historical observations for newly introduced or low-frequency materials can lead to unstable prediction and overfitting, constraining the use of intelligent compensation during the manufacturing launch of new materials.

The transfer experiment defines rare materials as categories present in the training set with fewer than 30 records. The rare-material final test contains 25 orders across eight dates: six from material 5, ten from material 7, and nine from material 17. A source XGBoost is fitted with 150 trees on 1736 remaining development records, then continued for 50 trees using 104 historical rare-material records; depth is 4 and learning rate is 0.05. The comparator is a 200-tree global XGBoost trained on all development records. Both use the same six inputs, preprocessing, and rare-material test orders. For full-cohort replay, non-rare orders retain global-model predictions. This is a pooled rare-category transfer experiment.

Figure 10 shows lower predictive performance of the transferred model than the pooled global model on the same rare-material test subset ($n = 25$). Panels a and b report S $R^2$ and MAE, and panels c and d report C $R^2$ and MAE. Bars start at zero; $R^2$ axes are target-specific, whereas MAE axes share the same scale in D. Higher $R^2$ and lower MAE indicate better prediction. S $R^2$ changes from −0.2875 to −1.5137, with MAE increasing from 0.0664 to 0.0807 D. C $R^2$ changes from 0.4069 to 0.0642, with MAE increasing from 0.0421 to 0.0489 D. The small test subset limits the precision of the transfer-performance estimates.

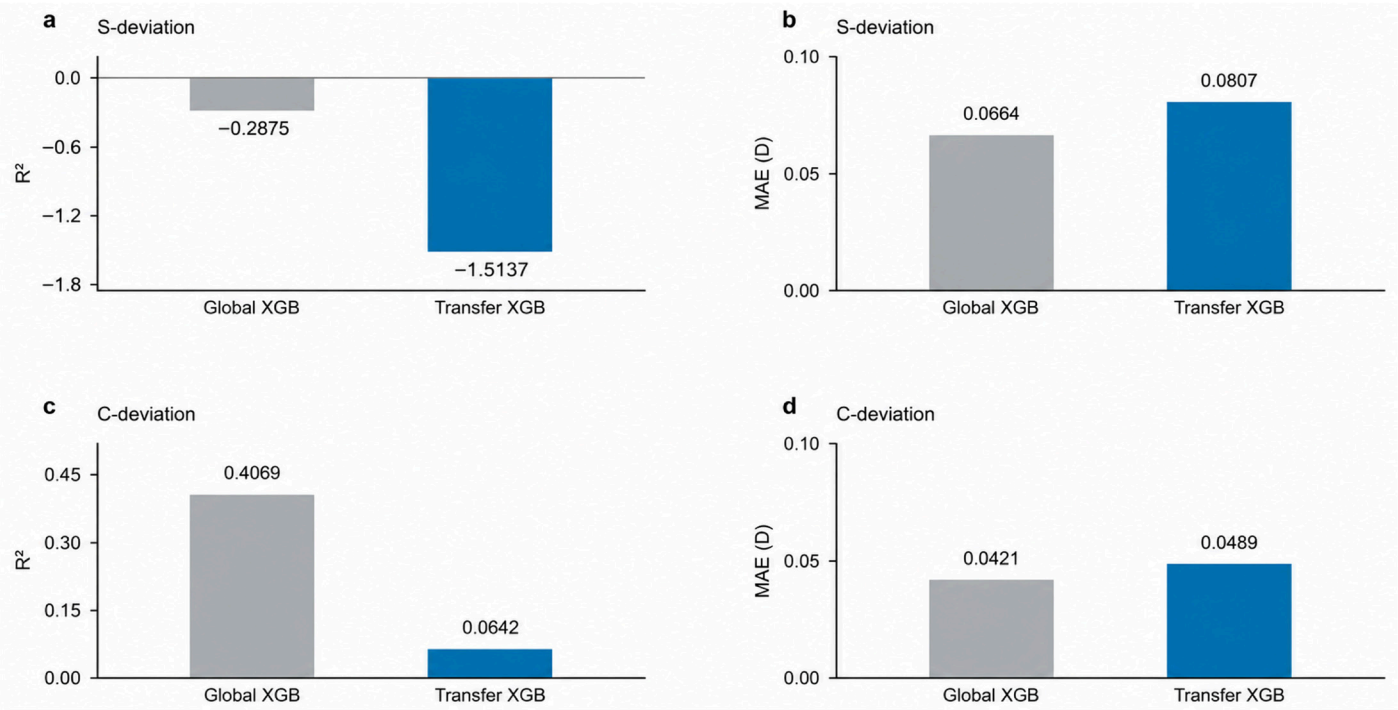



**Figure 10.** Global and transferred XGBoost on rare-material orders. (**a**) Coefficient of determination ($R^2$) for spherical power deviation (S-deviation); (**b**) Mean absolute error (MAE) for S-deviation; (**c**) $R^2$ for cylindrical power deviation (C-deviation); (**d**) MAE for C-deviation.

A separate prediction benchmark compares training-data allocation and adaptation depth under a common XGBoost 3.1.1 implementation. The six inputs share a standard scaler and one-hot category vocabulary fitted on the 1840 development records. Models use histogram tree construction, depth 4, learning rate 0.05, squared-error loss, full row and column sampling, L2 leaf regularization of 1, and seed 42. Pooled-global, source-only, and rare-only models each use 200 trees; source-to-rare continuation allocates these 200 trees as

150 plus 50, 100 plus 100, or 175 plus 25. The unadapted 150-tree source is also evaluated to distinguish continuation from direct application of its starting model. All settings are fixed for this benchmark without test-based selection. Table 12 reports prediction errors on the same 25 rare-material orders.

**Table 12.** Rare-material prediction under the common XGBoost benchmark.

| Training Allocation | S MAE | C MAE | S MAE Difference (95% CI) | C MAE Difference (95% CI) |
|---|---|---|---|---|
| Pooled global | 0.0629 | 0.0419 | Reference | Reference |
| Source only (150) | 0.0759 | 0.0421 | 0.0131 (−0.0069, 0.0444) | 0.0003 (−0.0038, 0.0058) |
| Source only (200) | 0.0746 | 0.0423 | 0.0117 (−0.0089, 0.0431) | 0.0004 (−0.0039, 0.0061) |
| Rare only | 0.0952 | 0.0545 | 0.0324 (0.0024, 0.0846) | 0.0127 (−0.0080, 0.0260) |
| Continuation 150 + 50 | 0.0826 | 0.0505 | 0.0197 (0.0027, 0.0568) | 0.0086 (−0.0024, 0.0191) |
| Continuation 100 + 100 | 0.0917 | 0.0457 | 0.0289 (0.0009, 0.0825) | 0.0038 (−0.0146, 0.0187) |
| Continuation 175 + 25 | 0.0703 | 0.0490 | 0.0074 (−0.0006, 0.0227) | 0.0071 (−0.0001, 0.0136) |

MAE differences are relative to the pooled-global benchmark, with positive values denoting larger errors. Pointwise 95% percentile intervals use 2000 paired resamples of the eight test dates, retaining all orders from each sampled date and keeping fitted models fixed. The intervals quantify test-date variation for this benchmark.

In the common benchmark, pooled-global training gives S and C MAEs of 0.0629 and 0.0419 D, respectively. The 150-plus-50 continuation increases S MAE by 0.0197 D (95% interval: 0.0027–0.0568 D) and C MAE by 0.0086 D (−0.0024–0.0191 D). Reducing adaptation to 25 trees lowers S MAE to 0.0703 D; its S and C difference intervals both include zero. Rare-only training gives the largest S and C point-estimate errors. Thus, the observed deterioration depends on data allocation, target, and adaptation depth. These results concern continuation of pooled rare-category records; evaluation of genuinely new materials requires a separate target cohort.

### *4.4. Manufacturing Deployment Considerations and Limitations*

The proposed architecture exposes the static prediction, dynamic correction, and gate state as separate components that can be logged and audited. Its potential integration into manufacturing control remains to be tested with measured command–response calibration, actual inspection delays, throughput constraints, and fail-safe operating procedures. For an illustrative manufacturing throughput (units per hour, UPH) at the rate of 60 lenses per hour or higher, command generation would need to satisfy the order-release schedule while using only the inspection results already available. Inspection feedback must also account for the applied command before updating the residual prediction. Manufacturing throughput, economic benefit, and protective performance require prospective measurement.

Figure 11 compares 13 strategy configurations on the same 405 test orders. Points show changes in joint S/C pass rate from the common baseline of 231/405 (57.04%); whiskers are 95% paired date-bootstrap intervals from 2000 resamples of the nine test dates. XGBoost with global EWMA has the highest observed rate, 319/405 (78.77%), followed by transferred XGBoost with global EWMA at 78.52%. The 19-input MLP without the optical-prior penalty gives 311/405 (76.79%), while the core material strategy gives 310/405 (76.54%). The intervals quantify gains over the baseline rather than pairwise differences between configurations. The primary static-AI comparison is given in Table 5. Neural and XGBoost variants reuse the validation-selected global EWMA settings; the transferred gated variant reuses the core material settings. Each configuration's gain over the baseline is summarized by a pointwise interval at the nominal 95% level. The core controller was selected using validation data before the test comparison.

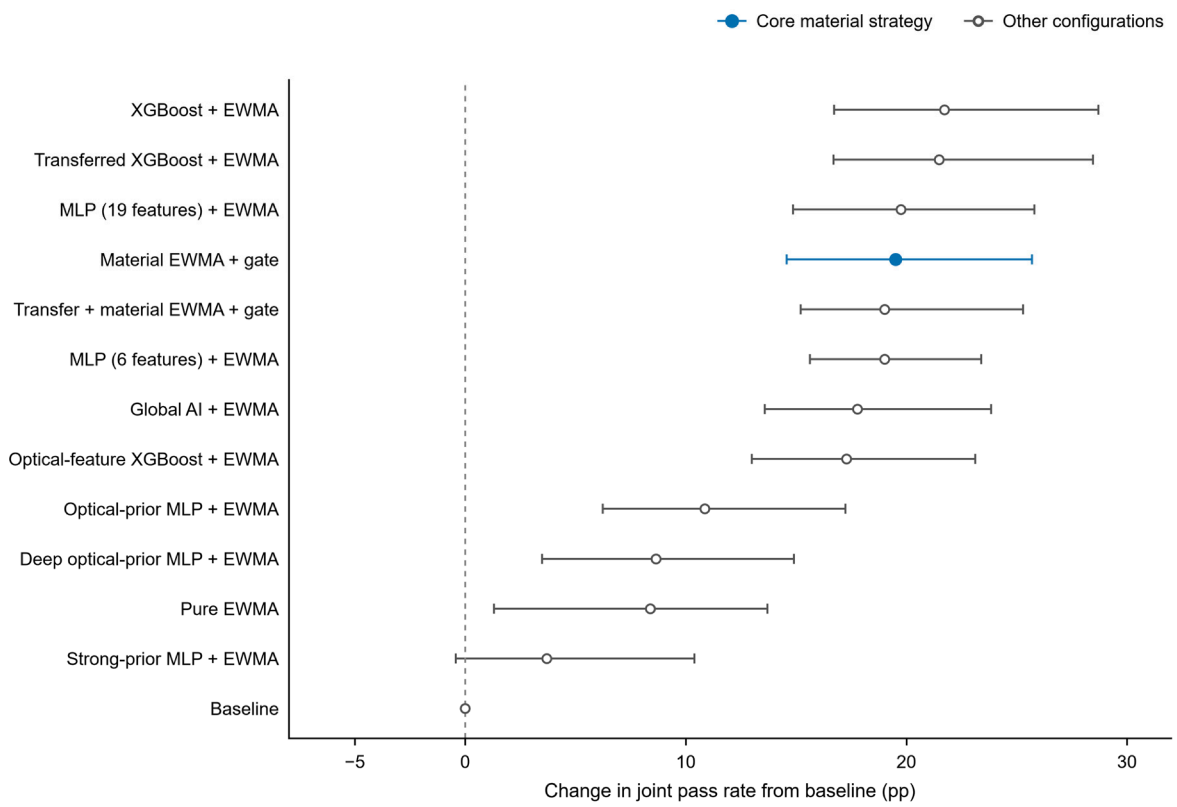



**Figure 11.** Retrospective comparison of 13 compensation configurations.

The historical fixed-offset baseline attains 68.40% on the primary test, compared with 75.06% for the primary static AI strategy. On the same 405 test orders, this corresponds to 27 additional passes with static AI. The data represent a single enterprise and nine test dates, and the date-bootstrap intervals characterize variation within this setting. Generalization across manufacturing systems requires replication.

Chronological analyses use records with complete calendar dates, harmonize ADD units from the observed scale separation, and apply an outcome-dependent deviation cutoff below 1 D. Related-order separation is assessed for records with identifiers. Timestamp groups define the assumed feedback schedule, while physical command response and actual inspection timing require prospective measurements. Optical proxies and regularization are heuristic, and the transfer and neural comparisons use a small test subset and a single training seed, respectively. For the primary EWMA estimates, historical residual construction is a source of methodological uncertainty because initialization residuals may derive from in-sample fitted values or from predictions based only on preceding observations. The robustness analyses specify fitted-residual initialization, compare validation-carried states at predictor refitting, and vary command response and feedback delay. Primary comparisons among static AI, global fusion, and material tracking are expressed as point estimates; paired intervals for these contrasts are calculated from the robustness trajectories.

The gate remains FULL throughout the primary test and produces the same commands as ungated material tracking. With gate histories initialized at the evaluation boundary and a window of $W = 20$, materials with fewer than 20 test observations remain in the warm-up state. Material 16, with 10 test orders, is one such category. Its deterioration highlights the limited monitoring coverage of this warm-up rule for infrequently inspected materials. Controlled gate evaluations should include nominal sequences and sequences with abrupt or gradual drift, anomalous feedback, and sparse material histories, with predefined event onset and recovery criteria. Outcomes should include activation frequency,

detection delay, false-alarm rate, recovery time, and maximum absolute S/C deviation. Cold-start gate policies should compare initial attenuation, fallback to global monitoring histories, and hierarchical state sharing; signed-mean triggers should be compared with absolute-error, tail-error, and exceedance-probability triggers. Sensitivity to window length and thresholds should be assessed on evaluation sequences distinct from those used for calibration. Prospective manufacturing studies should also assess calibrated machine response and inspection delays across materials, equipment, and manufacturing periods.

## 5. Conclusions

This study retrospectively evaluates an order-to-order compensation framework for mass-customized lens manufacturing that combines static ensemble prediction with global or material-specific EWMA residual tracking and a proposed material-specific gate. The system framework integrates order-dependent deviation prediction with sequential updating from completed-order inspections.

On the 405-record primary test, retrospective unit-gain replay within the study-specific $\pm 0.10$ D band yields joint S/C pass rates of 57.04% at the baseline, 75.06% with static AI, 65.43% with pure EWMA, 74.81% with global fusion, and 76.54% with material-specific tracking and the gate. The material-specific strategy achieves a gain of 19.51 percentage points over the baseline, although one of the 15 materials has a lower pass rate than its own baseline. The gate remains FULL throughout this test, leaving the compensation commands unchanged. Global fusion exceeds both the baseline and material-specific tracking in all five primary expanding-window evaluations, whereas static AI has a slightly higher observed joint pass rate than global fusion on the primary test. Material tracking adds six joint passes over static AI, with all gate weights remaining FULL. The historical fixed-offset baseline reaches 68.40%, indicating that systematic bias correction accounts for a substantial part of the improvement over recorded outcomes. In the robustness analysis, paired pass-rate intervals for material versus static or global tracking include zero.

Cooling-time associations in the auxiliary cohort are target-dependent, heuristic optical priors are not consistently beneficial, and the tested transfer procedure yields lower rare-material prediction accuracy than the pooled global comparator. Overall, compensation performance depends on the target, material, and evaluation interval, supporting validation-based selection of the residual-tracking strategy.

The results quantify changes in recorded S/C deviations under an assumed unit command response. Prospective manufacturing trials should calibrate this response, incorporate measured inspection delays and assess gate protection under defined abnormal conditions, with release yield and full optical quality as additional outcomes.

**Author Contributions:** Conceptualization, X.C. and K.C.; methodology, X.C. and K.C.; software, X.C., K.C., L.Z. and Y.F.; validation, X.C., K.C., L.Z. and Y.F.; formal analysis, X.C., K.C., L.Z. and Y.F.; investigation, X.C., K.C., L.Z. and Y.F.; resources, X.C., K.C., L.Z. and Y.F.; data curation, X.C., K.C. L.Z. and Y.F.; writing—original draft preparation, X.C.; writing—review and editing, X.C. and K.C.; visualization, X.C. and K.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study are available on request from the authors.

**Acknowledgments:** The authors are grateful for the technical support provided by Jiangsu Gongda Jinkai High Value Manufacturing Ltd., throughout this research, including valuable assistance with the research activities, data provision, and experimental support. The authors also gratefully ac-

knowledge the support provided by the Liaoning Provincial Project for Enterprise Science and Technology Commissioners.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References


1. Alonso, J.; Gómez-Pedrero, J.A.; Quiroga, J.A. *Modern Ophthalmic Optics*; Cambridge University Press: Cambridge, UK, 2019.
2. *ISO 8980-2*; Ophthalmic Optics—Uncut Finished Spectacle Lenses—Part 2: Specifications for Power-Variation Lenses, 3rd ed. ISO: Geneva, Switzerland, 2017.
3. Arora, S.; Rani, R.; Saxena, N. A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques. *WIREs Data Min. Knowl. Discov.* **2024**, *14*, e1536. [CrossRef]
4. Perno, M.; Hvam, L.; Haug, A. A machine learning digital twin approach for critical process parameter prediction in a catalyst manufacturing line. *Comput. Ind.* **2023**, *151*, 103987. [CrossRef]
5. Li, P.; Cheng, K.; Jiang, P.; Katchasuwanmanee, K. Investigation on industrial dataspace for advanced machining workshops: Enabling machining operations control with domain knowledge and application case studies. *J. Intell. Manuf.* **2022**, *33*, 103–119. [CrossRef]
6. Chen, Z.; Turng, L.S. A review of current developments in process and quality control for injection molding. *Adv. Polym. Technol.* **2005**, *24*, 165–182. [CrossRef]
7. Ko, J.U.; Lee, J.; Kim, T.; Kim, Y.C.; Youn, B.D. Deep-learning-based fault detection and recipe optimization for a plastic injection molding process under the class-imbalance problem. *J. Comput. Des. Eng.* **2023**, *10*, 694–710. [CrossRef]
8. Krantz, J.; Licata, J.; Raju, M.A.; Gao, P.; Ma, R.; Masato, D. Machine learning-based process control for injection molding of recycled polypropylene. *Polymers* **2025**, *17*, 940. [CrossRef] [PubMed]
9. Moayyedian, M.; Abhary, K.; Marian, R. Optimization of injection molding process based on fuzzy quality evaluation and Taguchi experimental design. *CIRP J. Manuf. Sci. Technol.* **2018**, *21*, 150–160. [CrossRef]
10. Yin, S.; Ding, S.X.; Xie, X.; Luo, H. A review on basic data-driven approaches for industrial process monitoring. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6418–6428. [CrossRef]
11. Zhou, T.; Zhu, Z.; Liu, X.; Liang, Z.; Wang, X. A review of the precision glass molding of chalcogenide glass (ChG) for infrared optics. *Micromachines* **2018**, *9*, 337. [CrossRef] [PubMed]
12. Huo, D.; Cheng, K. Experimental investigation on micromilling of oxygen-free, high-conductivity copper using tungsten carbide, chemistry vapour deposition, and single-crystal diamond micro tools. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2010**, *224*, 995–1003. [CrossRef]
13. Zaagan, A.A.; Aljohani, H.M.; Zogan, H.; Nabi, M. A machine learning assisted truncated adaptive EWMA chart for robust process monitoring. *Sci. Rep.* **2026**, *16*, 3196. [CrossRef] [PubMed]
14. Noor-ul-Amin, M.; Kazmi, M.W.; Alkhalaf, S.; Abdel-Khalek, S.; Nabi, M. Machine learning based parameter-free adaptive EWMA control chart to monitor process dispersion. *Sci. Rep.* **2024**, *14*, 31271. [CrossRef] [PubMed]
15. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
16. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
17. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv. (CSUR)* **2022**, *55*, 1–37. [CrossRef]
18. Sharma, R.; Raissi, M.; Guo, Y.B. Physics-informed machine learning for smart additive manufacturing. *arXiv* **2024**, arXiv:2407.10761. [CrossRef]
19. Ra, D.; Lee, J.; Lee, M.; Kwak, S.; Lee, S.; Lee, S. Physics-informed machine learning across manufacturing processes: Recent advances, challenges, and directions. *J. Manuf. Syst.* **2026**, *85*, 72–95. [CrossRef]
20. Moosavi, S.; Farajzadeh-Zanjani, M.; Razavi-Far, R.; Palade, V.; Saif, M. Explainable AI in Manufacturing and Industrial Cyber–Physical Systems: A Survey. *Electronics* **2024**, *13*, 3497. [CrossRef]
21. Puthanveettil Madathil, A.; Luo, X.; Liu, Q.; Walker, C.; Madarkar, R.; Qin, Y. A review of explainable artificial intelligence in smart manufacturing. *Int. J. Prod. Res.* **2025**, *63*, 8654–8697. [CrossRef]
22. Wang, J.; Ma, Y.; Zhang, L.; Gao, R.X.; Wu, D. Deep learning for smart manufacturing: Methods and applications. *J. Manuf. Syst.* **2018**, *48*, 144–156. [CrossRef]
23. Flores-García, E.; Kwak, D.H.; Jeong, Y.; Wiktorsson, M. Machine learning in smart production logistics: A review of technological capabilities. *Int. J. Prod. Res.* **2025**, *63*, 1898–1932. [CrossRef]
24. Montgomery, D.C. *Introduction to Statistical Quality Control*, 8th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2020; ISBN 978-1-119-72309-7.

25. Box, G.E.P.; Luceño, A.; Paniagua-Quiñones, M.D.C. *Statistical Control by Monitoring and Adjustment*; Wiley: Hoboken, NJ, USA, 2009. [CrossRef]
26. Arinez, J.F.; Chang, Q.; Gao, R.X.; Xu, C.; Zhang, J. Artificial intelligence in advanced manufacturing: Current status and future outlook. *J. Manuf. Sci. Eng.* **2020**, *142*, 110804. [CrossRef]
27. Good, R.; Qin, S.J. Performance synthesis of multiple input-multiple output (MIMO) exponentially weighted moving average (EWMA) run-to-run controllers with metrology delay. *Ind. Eng. Chem. Res.* **2011**, *50*, 1400–1409. [CrossRef]
28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
29. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
30. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*; ACM: New York, NY, USA, 2017; pp. 3149–3157. Available online: https://dl.acm.org/doi/10.5555/3294996.3295074 (accessed on 10 June 2025).
31. Zheng, H.; Gao, X.; Yang, M.; Yang, X.; Li, Y.; Ding, Y. Complex product quality prediction method based on an improved light gradient boosting machine. *Appl. Intell.* **2025**, *55*, 248. [CrossRef]
32. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31*; NeurIPS: Vancouver, BC, Canada, 2018. Available online: https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html (accessed on 22 May 2025).
33. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.
34. Roberts, S.W. Control chart tests based on geometric moving averages. *Technometrics* **1959**, *1*, 239–250. [CrossRef]
35. Zhu, Q.; Liu, Z.; Yan, J. Machine learning for metal additive manufacturing: Predicting temperature and melt pool fluid dynamics using physics-informed neural networks. *Comput. Mech.* **2021**, *67*, 619–635. [CrossRef]
36. Gao, H.; Liu, J.; Ling, M.; Chen, T. Geometrically nonlinear design of a rhombus-nested compliant amplification mechanism for use in precision actuators and sensors. *Precis. Eng.* **2024**, *90*, 164–175. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.