analytica

MDPI

Article

Uncertainty-Aware Adaptive Control Charts for Detecting Analytical Drift in Spectroscopic Measurements

Manuel J. Campuzano *, Rick Acosta-Vega and Enrique Delahoz-Domínguez

Facultad de Ingeniería, Universidad del Magdalena, Santa Marta 470004, Colombia; racosta@unimagdalena.edu.co (R.A.-V.); enriquedelahoz@unimagdalena.edu.co (E.D.-D.)
* Correspondence: mcampuzano@unimagdalena.edu.co

Abstract

We tested whether weighting spectroscopic prediction errors by their estimated uncertainty improves drift detection. Using public spectra from corn and pharmaceutical tablets, we compared an uncertainty-weighted chart (FR-U) with its fixed mean-and-energy counterpart (Fixed-GV) and established control charts. We simulated changes under three fitted designs per dataset, with a nominal 5% probability of any false alarm over 200 inspections under four specified normal regimes. At a positive bias of half the training out-of-fold root mean square error in the first corn design, FR-U detected 21.2% of changes by the horizon, compared with 10.3% for Fixed-GV. Fewer missed changes accounted for most of the seven-inspection reduction in restricted mean delay. This gain was not reproduced in the other two corn designs, where FR-U was slower by about two and one inspections, respectively. At the reference settings, EWMA and CUSUM were faster than FR-U for the positive unit bias and masking ramp in both primary designs. A published joint chart was faster for increased noise. Uncertainty weighting therefore merits application-specific testing on representative control materials before prospective use.

Keywords: analytical drift; near-infrared spectroscopy; statistical process control; predictive uncertainty; partial least squares; conformal calibration; false-alarm control

1. Introduction

Consider a laboratory that checks a corn moisture calibration with sealed reference materials at the start of each batch. A gradual change in sample presentation can affect the predicted moisture even when the assigned moisture remains stable. The operational question is when to investigate the measurement system. Spectroscopic calibration models are often developed from a finite collection of samples and subsequently deployed under measurement conditions that differ from those represented during calibration. Changes in instrument response, sample presentation, background absorption and the population of samples can alter both the spectra and the reliability of the resulting analytical predictions. Preprocessing and model selection are therefore important parts of the analytical procedure, rather than interchangeable preliminary steps [1]. Equally important is the distinction between a change in the spectral distribution and a change that materially degrades the predicted analytical response [2].

Recent calibration-transfer research illustrates several ways of addressing instrument differences. Adaptive artificial neural networks update a previously trained calibration model using information from a secondary instrument [3]. Likelihood-based transfer propagates uncertainty in the relationship between spectral domains, including domains with

Check for updates

Academic Editor: Marcello Locatelli

Received: 15 October 2026
Revised: 11 November 2026
Accepted: 20 November 2026
Published: 26 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.

non-overlapping wavelength ranges [4]. Low-rank adaptation has also been investigated for transferring deep learning models between Raman spectrometers [5]. These methods address model maintenance, but deciding when a model requires intervention remains a separate problem. Long-term transmission Raman studies show that the benefit of a correction can depend on the instrument condition and validation set [6], while recent work on online near-infrared (NIR) measurements explicitly connects model evaluation with maintenance [7].

Uncertainty quantification offers a natural way to distinguish ordinary prediction variability from deterioration. Recent spectroscopic applications include quantile regression forests [8], conformal predictors for food adulteration [9], Monte Carlo conformal prediction for soil spectra [10], and uncertainty calibration for latent-variable regression [11]. Nevertheless, widening an uncertainty interval serves a different purpose from detecting a harmful change. A scale estimate that increases for a spectrum affected by drift may reduce a standardized residual even when the error in analytical units is increasing. Conversely, a spectrally unusual observation can retain an acceptable prediction. Neither a large uncertainty estimate nor a measure of spectral distance alone resolves this distinction.

Conformal prediction provides finite-sample marginal coverage under exchangeability [12]. Extensions address violations of exchangeability [13], adaptive coverage under distribution shift [14,15], and dependent time series [16,17]. Conformal proportional–integral–derivative (PID) control and state-aware forecasting further illustrate how predictive intervals can adapt to changing conditions [18,19]. These results do not, without an additional argument, control the probability that a continuously inspected control chart produces an alarm. Shared calibration data also induce dependence between conformal p-values, a point that matters when they are combined across observations [20].

Sequential change detection has developed complementary approaches to calibration and adaptation. Adaptive cumulative sum (CUSUM) charts address gradual drift [21], and data-adaptive symmetric CUSUM addresses changes involving both location and scale [22]. Simulation-based calibration can target a time-dependent false-detection rate [23]. Conformal CUSUM [24], e-detectors [25], and reductions based on confidence sequences [26] provide other routes to sequential validity under their respective assumptions. These procedures address different operating criteria; finite-horizon risk, average run length (ARL) and time-uniform error control are distinct concepts [27]. The exponentially weighted moving average (EWMA) chart based on conformal p-values proposed by Li et al. [28] is particularly relevant to this study.

A relevant chemometric precedent is the study by Nikzad-Langerodi et al., which combined disagreement among partial least squares (PLS) models, an adaptive Page–Hinkley (PH) detector and calibration maintenance in melamine resin production [29]. Joint surveillance of location and dispersion also predates the present work. Max-EWMAMS combines transformed moving mean and squared-deviation statistics for individual observations [30]; Max-mixed EWMA provides a more recent subgroup-based construction [31]. Both the use of maxima to combine branches and the integration of uncertainty with drift detection have precedents. Our question is whether an input-dependent error scale contributes useful sensitivity while fixed analytical evidence remains available. The need to preserve relevant signal is also illustrated by a spectral correction study in which an overly flexible correction removed genuine spatial segregation structure [32]. Controlled perturbations help study such mechanisms, while independent measurements remain necessary for instrumental validation [33].

We use a fixed-reference chart for the mean and energy of the prediction error (Fixed-GV) as a practical baseline. FR-U adds a third branch: it gives more weight to errors for spectra judged easier to predict. A laboratory already running Fixed-GV would need a

frozen uncertainty model and a new joint alarm limit to add this branch. The benefit is an earlier warning when weighting emphasizes a relevant bias; the cost is that the joint limit can delay a warning already visible to the fixed branches. Twelve reference procedures, signed perturbations and repeated fitted designs let us examine both outcomes. The analyses use public spectra and constructed monitoring episodes, so their conclusions concern the specified experiments rather than observed longitudinal drift.

2. Monitoring Method and Calibration

2.1. Analytical Reference and Predictive Uncertainty

Let $\mathbf{x}_t \in \mathbb{R}^p$ denote a measured spectrum and y_t^{ref} its independently available reference response. A predictive model $\hat{f}$ is fitted before monitoring. The centered residual is

$$r_t = y_t^{\mathrm{ref}} - \hat{f}(\mathbf{x}_t) - b_0, \qquad b_0 = \frac{1}{n_Q} \sum_{i \in Q} \{y_i^{\mathrm{ref}} - \hat{f}(\mathbf{x}_i)\}, \tag{1}$$

where Q is the designated Phase-I control panel. The panel bias b_0, predictor, preprocessing and uncertainty model remain fixed during monitoring. This construction monitors departures from an established analytical baseline; it does not certify that the original baseline has zero bias. Reference-assay error is part of the residual and is not separately identifiable from these data.

The prediction model uses PLS regression [34] fitted to mean-centered spectra. The number of components is selected by fivefold cross-validation within the training set from $\{2, 4, 6, 8, 10, 12\}$. Let e_i^{OOF} be the out-of-fold residuals at the selected component count and s_0 their root mean square. Five whitened principal component analysis (PCA) scores from the training data define the nearest-neighbor space. With $K_{\mathrm{NN}} = \min(10, \lfloor n_{\mathrm{tr}}/4 \rfloor)$ and $B_f = 20$ bootstrap PLS fits, define

$$v_{\mathrm{loc}}(\mathbf{x}) = \frac{1}{K_{\mathrm{NN}}} \sum_{i \in \mathcal{N}_{K_{\mathrm{NN}}}(\mathbf{x})} (e_i^{\mathrm{OOF}})^2, \tag{2}$$

$$v_{\mathrm{ens}}(\mathbf{x}) = \frac{1}{B_f} \sum_{b=1}^{B_f} \left\{ \hat{f}_b(\mathbf{x}) - \bar{f}_{B_f}(\mathbf{x}) \right\}^2, \tag{3}$$

$$\tilde{s}(\mathbf{x}) = \mathrm{clip}\left(\sqrt{v_{\mathrm{loc}}(\mathbf{x}) + v_{\mathrm{ens}}(\mathbf{x})},\ 0.25 s_0,\ 4 s_0 \right). \tag{4}$$

The bootstrap models use the selected PLS dimension; hyperparameters are not reselected using monitoring data. Equation (4) defines a positive, input-dependent scale; it does not identify separate aleatoric and epistemic uncertainty components. Cross-validation selection can make s_0 optimistic, but the sequential threshold is calibrated separately after this scale has been frozen.

The five scores describe broad spectral variation without defining neighborhoods in hundreds of channels. Ten neighbors smooth noisy squared errors; the quarter-sample cap prevents a neighborhood from dominating a small training set. Twenty bootstrap fits provide a finite committee, and the bounds limit residual weights to between 1/4 and 4 times the fixed weight. These are working design choices, without a claim of optimality or a documented preliminary parameter search. Fewer neighbors can make individual errors influential; more can average away local differences. Whitening can also amplify weak score directions. The supplement varies the score count, neighborhood size and committee size separately.

For a separate diagnostic of predictive uncertainty, normalized absolute prediction errors on Q yield a split-conformal 90% interval

$$\hat{f}(\mathbf{x}) \pm q_{.9} \tilde{s}(\mathbf{x}), \qquad q_{.9} = a_{(\lceil .9(n_Q+1) \rceil)}, \quad a_i = \frac{|y_i^{\mathrm{ref}} - \hat{f}(\mathbf{x}_i)|}{\tilde{s}(\mathbf{x}_i)}. \tag{5}$$

An unavailable order statistic is interpreted as $+\infty$. These intervals use the uncentered prediction error; the monitoring correction b_0 is not fitted to their calibration residuals. Their coverage on external records is reported descriptively. A coverage guarantee requires exchangeability of the interval-calibration and future records; it is not asserted for the provider-defined tablet blocks. The interval quantile is not used as the alarm threshold.

Error ordering and monitoring utility are different properties. Write $R = r/s_0$ and $W = s_0/\tilde{s}$. A rank correlation between $\tilde{s}$ and $|r|$ describes ordering, whereas the adaptive mean branch depends on the signed product WR. If the weights are held fixed while a small constant shift Δ is added to R, its mean changes by $\Delta\mathbb{E}(W)$. Relative to unweighted monitoring, the corresponding first-order innovation signal-to-noise ratio is

$$\mathcal{E}_W = \frac{\mathbb{E}(W)\,\mathrm{sd}(R)}{\mathrm{sd}(WR)}. \tag{6}$$

This diagnostic ignores change-induced variance, serial dependence, changes in the weights themselves and the joint decision threshold. It is not a detection-power formula. It explains why modest rank correlation need not preclude a benefit, and why rank correlation or interval coverage alone cannot establish one. We therefore assess ordering, signed weighting and actual paired detection outcomes separately.

2.2. Adaptive and Fixed-Reference Branches

With $\lambda \in (0, 1]$ and $q = 1 - \lambda$, the three branch states satisfy

$$A_t = qA_{t-1} + \lambda r_t/\tilde{s}_t, \tag{7}$$

$$G_t = qG_{t-1} + \lambda r_t/s_0, \tag{8}$$

$$V_t = qV_{t-1} + \lambda\{(r_t/s_0)^2 - m_{2,0}\}. \tag{9}$$

Here $\tilde{s}_t = \tilde{s}(\mathbf{x}_t)$, and $m_{2,0}$ is a frozen second-moment reference estimated from independent design simulations. The EWMA structure follows the classical construction [35]. A_t responds to uncertainty-scaled signed error, G_t retains signed error on a fixed scale, and V_t responds to increased error energy. Only A_t uses the input-dependent denominator. All states start at zero before a prescribed warm-up; every calibration and evaluation episode applies the same warm-up rule.

A sequence of predictions below their references pushes both signed branches upwards; the adaptive branch moves furthest when the assigned uncertainty is small. Large errors of alternating sign can cancel in either mean branch while accumulating in the energy branch.

For fixed positive branch constants c_A, c_G, c_V, define

$$M_t = \max\{|A_t|/c_A,\ |G_t|/c_G,\ (V_t)_+/c_V\}, \qquad T = \inf\{t \geq 1 : M_t > h\}. \tag{10}$$

The monitored index begins after warm-up, and the episode ends at its first alarm. Updating a branch state does not update the predictive model or the reference distribution. In particular, no monitoring residual is used to redefine b_0, s_0, $m_{2,0}$ or the control limits.

The adaptive branch can be expressed in the native response units. Given A_{t-1}, its no-alarm interval for the residual is

$$\frac{\tilde{s}_t}{\lambda}\{-hc_A - qA_{t-1}\} \leq r_t \leq \frac{\tilde{s}_t}{\lambda}\{hc_A - qA_{t-1}\}. \tag{11}$$

These limits vary with the current spectrum. The fixed branches remain available if this interval widens. FR-U is therefore adaptive to the input uncertainty, while its learned mapping and analytical reference are frozen.

Figure 1 follows a control measurement from the frozen prediction model through the three branches to a single alarm decision.

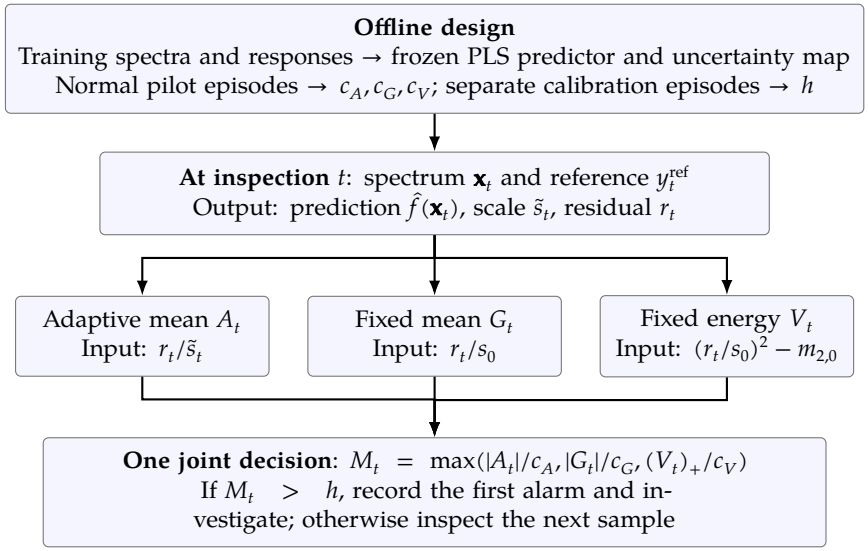

Figure 1. Information flow of FR-U. The fitted model maps spectra to predictions and uncertainty scales. References are required to form the residual. Recursive branch updates do not refit the model or the baseline. Pilot normalization and final threshold calibration use separate normal episodes.

2.3. Pilot Balancing and Episode Calibration

Let $\mathcal{P}_0 = \{P_{0,1}, \ldots, P_{0,J}\}$ be a declared finite set of acceptable episode laws. Each law specifies the control-panel sampling distribution, noise level, dependence and warm-up. For branch $j \in \{A, G, V\}$, let $Z_{A,t} = |A_t|$, $Z_{G,t} = |G_t|$ and $Z_{V,t} = (V_t)_+$. On independent pilot episodes of length H, set

$$c_j = \max_{1 \le \ell \le J} \widehat{Q}_{.95}^{\mathrm{pilot},\ell}\Big(\max_{1 \le t \le H} Z_{j,t}\Big), \tag{12}$$

with a positive numerical floor. This pilot step balances branch scales against the accepted operating envelope. It does not assign a separate 5% risk allowance to each branch and does not establish the final false-alarm guarantee.

For example, dividing a mean branch and an energy branch by their respective pilot 95th percentiles of episode maxima puts their peaks on comparable scales. The chance that either branch crosses still depends on their joint behavior. The separate calibration sample supplies one limit for that joint event.

After all design choices are frozen, independently simulate B normal calibration episodes from each $P_{0,\ell}$ and compute

$$W_{b\ell} = \max_{1 \le t \le H} M_t^{(b,\ell)}, \qquad k = \lceil (B+1)(1-\alpha) \rceil, \qquad h = \max_{\ell} W_{(k),\ell}. \tag{13}$$

The same rule is used for every comparator, with its own episode statistic. Thus, all procedures target the same false-alarm budget over the monitoring horizon, although each has its own numerical threshold. We use strict exceedance and $h = +\infty$ if $k > B$.

Proposition 1 (Episode-level false-alarm control)**.** *Conditional on the frozen design, suppose that, for a specified ℓ, its B calibration episodes and a new normal episode are exchangeable. Then Equation (13) gives*

$$\Pr_{0,\ell}(T \le H) \le \alpha, \tag{14}$$

marginally over calibration and the new episode. The result holds separately for every law satisfying this assumption, without requiring independence within an episode.

A proof is given in the Supplementary Materials on calibration and fixed-reference retention. This guarantee follows from established rank calibration. The multiplicity of

inspections and branches is included in the episode maximum. The result is not conditional validity for every realized calibration set, an average-run-length identity, or a time-uniform guarantee for indefinite monitoring. Arbitrary switching between accepted regimes is also outside the statement unless switching episodes are included in $\mathcal{D}_0$.

For K planned restarts with valid episode risks $\alpha_1, \ldots, \alpha_K$, a union bound gives lifetime risk at most $\min(1, \sum_{k=1}^{K} \alpha_k)$; independence is unnecessary. If complete episodes are independent conditional on one realized threshold with episode risk p_h, their cumulative risk is $1 - (1 - p_h)^K$. For illustration, $p_h = 0.05$ yields 64.2% over 20 restarts. The marginal guarantee in Equation (14) does not imply $p_h = 0.05$ for each realized calibration set. A planned campaign therefore requires calibration of its full horizon or an explicit allocation of its total risk across episodes.

A complementary tolerance-limit calibration [36] controls variability in the realized threshold. For independent episode maxima, select the smallest rank k_{tol} such that

$$\Pr\{\text{Bin}(B, 1 - \alpha) \geq k_{\text{tol}}\} \leq \delta/(JL), \tag{15}$$

where L is the number of procedures for which simultaneous calibration confidence is claimed. Taking the maximum of the corresponding regime-specific order statistics gives, with calibration probability at least $1 - \delta$, a conditional episode risk at most α for all JL procedure–regime combinations. The binomial tail bounds the probability that an order statistic lies below the population $(1 - \alpha)$ quantile; a union bound supplies simultaneity. Here $L = 12$ and $J = 4$, per fitted design. This simultaneous claim covers only the twelve reference procedures; the centered diagnostic below uses rank calibration and is excluded from this tolerance family. This stronger calibration-confidence statement requires a higher threshold.

In repeated calibrations the rank rule controls the average risk for a new matching episode. The tolerance rule asks a different question: how confident can we be that the particular limits installed after one calibration meet the risk target across the specified family? It spends calibration precision on that stronger statement.

2.4. Retention of Evidence and Its Calibration Cost

Let $T_G(h) = \inf\{t : |G_t|/c_G > h\}$ and define $T_V(h)$ analogously.

Proposition 2 (Fixed-reference retention)**.** *For every residual and uncertainty sequence, $T \leq \min\{T_G(h), T_V(h)\}$. Holding residuals and frozen parameters fixed, any change in the uncertainty values leaves G_t and V_t unchanged.*

A proof is given in the Supplementary Materials on calibration and fixed-reference retention. This retention property applies at a common threshold. A standalone EWMA receives its own threshold h_G, which can be smaller than h. With shared calibration episodes, monotonicity of maxima and order statistics gives $h \geq h_G$. FR-U therefore preserves fixed-reference evidence at its deployed threshold but does not dominate a separately calibrated EWMA in detection delay.

The masking mechanism can be expressed directly. If an unrestricted scale follows a positive deterministic ramp $r_t = dt$ through $s_t = dt/a$, then $r_t/s_t = a$. Starting from zero, the adaptive state is $A_t = a(1 - q^t)$ and never crosses a raw-state threshold at least a, even though r_t diverges. In contrast,

$$G_t = \frac{d}{s_0}\left[t - \frac{q(1 - q^t)}{\lambda}\right], \tag{16}$$

which diverges and eventually crosses every finite threshold. A nonzero initial state adds $q^t G_0$. Bounded uncertainty already prevents this particular asymptotic counterexample;

the fixed branches additionally retain evidence independently of the size of the uncertainty cap. Equation (16) describes a deterministic mechanism, not a guaranteed finite-horizon detection probability under unbounded random noise.

2.5. Centered Adaptive Innovation as a Diagnostic

Mean centering of the analytical residual does not generally center its uncertainty-weighted counterpart:

$$\mathbb{E}_0(r_t) = 0 \quad \not\Rightarrow \quad \mathbb{E}_0(r_t/\tilde{s}_t) = 0. \tag{17}$$

Indeed, when $\mathbb{E}_0(r_t) = 0$, the latter mean equals $\mathrm{Cov}_0(r_t, 1/\tilde{s}_t)$. A nonzero value can move the adaptive state towards one side of its two-sided boundary before any change occurs. To diagnose this effect, FR-U-c replaces only Equation (7) by

$$A_t^c = qA_{t-1}^c + \lambda\{r_t/\tilde{s}_t - \widehat{m}_{A,0}\}, \qquad A_0^c = 0, \tag{18}$$

where $\widehat{m}_{A,0}$ is estimated from independent normal design episodes and then frozen. The predictor, uncertainty map, residual centering and both fixed branches are unchanged. The modified adaptive branch receives a new pilot constant c_A^c and the joint statistic a new rank-calibrated limit h^c, using separate normal samples.

For the same input sequence and age j from initialization, the unnormalized states satisfy $A_j^c = A_j - \widehat{m}_{A,0}(1 - q^j)$. This identity isolates the baseline translation before normalization; differences in calibrated detection additionally reflect the new pilot scale and joint threshold. A center estimated under one declared regime need not center every other regime, and neither centered residuals nor centered innovations guarantee sign-symmetric detection. Residual distribution shape, residual–scale dependence and changes in the scale under the alternative can still matter. FR-U-c is retained as a diagnostic, not selected as a replacement from abnormal evaluation outcomes.

3. Computational Experiments

3.1. Public Data and Separation of Roles

The corn and pharmaceutical-tablet archives were obtained from Eigenvector Research [37]. Exact download URLs and SHA-256 checksums are included in the supplementary manifest. Corn contains 80 samples measured on instruments m5, mp5 and mp6, with 700 channels from 1100 to 2498 nm. Moisture is the response, and m5 is the primary instrument. A seeded sample-level permutation assigns 40 sample identifiers to model training, 20 to the control panel and 20 to external predictive evaluation. The same identifiers are retained across instruments; paired spectra are not divided between model training and evaluation.

The tablet archive contains 155 calibration, 40 validation and 460 test rows per instrument, with 650 channels from 600 to 1898 nm. These original blocks are retained as training, control-panel and external-evaluation blocks, respectively. The response is the column labeled `assay`. The provider describes 654 tablets, whereas the downloaded blocks contain 655 rows per instrument. We retain the archive without deleting a record to force agreement with the description. No physical-specimen ledger is available to resolve this discrepancy, so the row counts cannot be verified as counts of distinct tablets. All responses and prediction errors are reported in the native response units because the distributed response labels do not themselves specify units. No observations are excluded as outliers.

Only m5 contributes to the 20-row corn monitoring panel and only instrument 1 to the 40-row tablet panel. Measurements on two further corn instruments and one further tablet instrument are paired transfer challenges, not new panel materials. The archives do not identify the panels by manufacturing batch, collection date or laboratory, or establish

independent tablet identities. File creation and modification dates are not acquisition dates. We use the spectra as distributed, with column centering inside the PLS fits and no autoscaling, SNV, MSC, differentiation or smoothing in the reference analysis. The stored file histories do not establish all processing performed before distribution.

Designs 14, 15 and 16 denote seeds 20261014, 20261015 and 20261016. Design 14 is the first entry in the executable protocol and supplies the detailed illustrations. There is no dated preregistration establishing it as an inferential priority. We report all three designs side by side, without selecting a seed or pooling them as independent experiments. In corn the train, panel and external partitions change and can overlap across designs. Tablets retain exactly the same three blocks; only the fitting and simulation draws change.

The operational analogue is a scheduled sequence of stable control materials with reference values assigned before monitoring. The stored response is reused when a material is inspected; the benchmark does not assume a new assay for every spectrum. If routine samples require a fresh assay, an inspection is complete only when that reference becomes available, and delays below are counted in completed inspections rather than elapsed acquisition time. The response column is not a certified material specification. Accordingly, s_0 defines statistical effect sizes only: neither a moisture acceptance tolerance nor a pharmaceutical release limit can be inferred from these archives.

Operational use would additionally require evidence that the control materials and their assigned values remain stable. Reusing a stored value does not eliminate its assignment uncertainty or protect against material degradation. The present data cannot distinguish changes in the reference material, the reference assay and the spectroscopic measurement from residuals alone. An alarm therefore initiates an investigation; it does not identify the instrument as the cause. No throughput or cost advantage over methods that require no reference assay is claimed.

There are two distinct uses of held-out data. External records assess prediction and panel transportability. The control panel defines a fixed set of reference materials from which normal and perturbed episodes are simulated. Calibration and evaluation episodes use independent random draws from this same panel. Consequently, their conditional episode law can be matched without claiming that thousands of simulated inspections constitute thousands of new physical samples.

3.2. Normal Episode Generator and Calibration Budget

Write the fitted linear PLS predictor as $\hat{f}(\mathbf{x}) = a + \mathbf{b}^{\mathsf{T}}\mathbf{x}$. An episode selects panel indices I_t independently and uniformly. Eight cosine functions over the wavelength grid, at frequencies $\{0, 1, 2, 4, 8, 16, 32, 64\}$, form rows of a matrix N. The rows are first normalized in Euclidean norm and then jointly scaled so that $\|N\mathbf{b}\| = s_0$. The generated spectrum is

$$\mathbf{x}_t^* = \mathbf{x}_{I_t} + \eta N^{\mathsf{T}}\mathbf{z}_t, \qquad \mathbf{z}_t = \rho\mathbf{z}_{t-1} + \sqrt{1-\rho^2}\,\boldsymbol{\epsilon}_t, \tag{19}$$

where $\mathbf{z}_0$ and $\boldsymbol{\epsilon}_t$ have independent standard Gaussian coordinates. The independent reference response of the selected panel record is retained. This normalization gives the added prediction perturbation a marginal standard deviation ηs_0. The cosine basis represents a reproducible family of smooth spectral perturbations, not an estimated instrument-noise covariance.

The four regimes bracket intensity and serial dependence in the simulation. They are test conditions, not measurements of routine instrument noise. For daily controls $H = 200$ means 200 measurement days; weekly controls span 200 weeks, and one control per batch means 200 batches. A laboratory should choose the horizon and alarm budget from its investigation capacity and the length of a planned campaign. With $\lambda = 0.1$, the half-life of a signed observation is about 6.6 inspections and only $0.9^{50} \approx 0.005$ of an initial

state remains after warm-up. A different sampling schedule or smoothing choice requires independent recalibration.

Four normal regimes are declared: $(\eta, \rho) = (0, 0), (0.5, 0), (1, 0)$ and $(0.5, 0.5)$. Each episode has 50 warm-up observations followed by $H = 200$ monitored inspections. The nominal probability of at least one alarm is $\alpha = 0.05$. All procedures use $\lambda = 0.1$ where applicable. A separate design sample of 160 episodes at $(0.5, 0)$ supplies second-moment, covariance and comparator-score references. Each accepted regime then supplies 399 pilot episodes for Equation (12) and a separate set of 1999 final calibration episodes for Equation (13). Thus the final rank is 1900. The tolerance alternative uses $\delta = 0.05$, allocated over 48 method–regime combinations; its rank is 1929. Calibration confidence applies to each fitted design, rather than simultaneously to all fitted designs in the study.

Each evaluation condition uses 1000 new episodes. Common random numbers pair competing methods and perturbation conditions within a fitted design. Different noise levels are treated as acceptable only when explicitly included in the normal-regime registry. Larger noise and stronger dependence are reported as out-of-envelope challenges. Changing the panel is evaluated separately and is not covered by the matching-law guarantee.

3.3. Perturbations and Analytical Impact

Changes begin at monitored inspection $\tau = 51$, after the 50 pre-change monitored inspections. Unless specified otherwise, noise remains at $(\eta, \rho) = (0.5, 0)$ and an additive perturbation $g_t \mathbf{d}_{I_t}$ is applied to the original spectrum before projection. Let σ_X be the square root of the mean training channel variance. Table 1 relates the perturbations to possible changes in measurement conditions.

Two diagnostic perturbations separate analytical impact from spectral novelty. For a Gaussian-shaped band vector $\mathbf{u}$, define

$$\mathbf{v}_\perp = \mathbf{u} - \frac{\mathbf{b}^\top \mathbf{u}}{\|\mathbf{b}\|^2} \mathbf{b}. \tag{20}$$

After rescaling to a root mean square (RMS) magnitude of σ_X, $0.3\mathbf{v}_\perp$ is a prediction-preserving negative control: $\hat{f}(\mathbf{x} + 0.3\mathbf{v}_\perp) = \hat{f}(\mathbf{x})$ up to numerical precision. This is a null direction of the fitted predictor, not proof that a physical perturbation would be chemically harmless.

For the masking stress, the same null direction is rescaled so that the standard deviation of its bootstrap prediction increments is $6s_0$, and then added to $\mathbf{d}_{\text{pred}}$. A ramp along this combined direction increases the primary prediction error while increasing committee disagreement. The construction deliberately challenges uncertainty normalization and is not presented as a realistic instrument-failure frequency model. Its spectral magnitude and induced error are recorded in the supplement. A second diagnostic holds each residual trajectory fixed and multiplies its uncertainty scale by 1, 2, 4 or 8 before the cap is applied.

The deliberate prediction impact is defined by a paired counterfactual:

$$\Delta_t = |\hat{f}(\mathbf{x}_t^{\text{pert}}) - \hat{f}(\mathbf{x}_t^{\text{normal}})|, \quad \mathbf{x}_t^{\text{normal}} = \mathbf{x}_{I_t} + \eta N^\top \mathbf{z}_t, \quad \mathbf{x}_t^{\text{pert}} = \mathbf{x}_t^{\text{normal}} + g_t \mathbf{d}_{I_t}. \tag{21}$$

The two spectra share their panel record and noise draw. For an additive change and the affine predictor, $\Delta_t = |g_t \mathbf{b}^\top \mathbf{d}_{I_t}|$. The code stores the signed residual increment, whose absolute value is Δ_t. For a noise increase, the paired increment also includes $(\eta_t - \eta) N^\top \mathbf{z}_t$. This induced effect is reported separately from total error against the supplied reference. A change of s_0 is used as a model-relative impact threshold. It is a reproducible statistical reference, not an application-certified acceptance limit. The physical perturbation families include both small and large induced effects; their full results are retained rather than selecting only detectable cases.

Table 1. Constructed changes and possible laboratory analogues. These challenges do not identify a physical cause or represent dated instrument trajectories.

Input Change	Possible Interpretation and Limits
Offset $(0.1$ or $0.3\sigma_X)$	A baseline or background displacement after imperfect referencing.
Gain (0.2% or 1%)	An intensity change associated with optical contact, path length or sample presentation.
Gaussian band	A localized change of height 0.1 or $0.3\sigma_X$, centered at 60% of the grid with width 3%. Temperature or matrix changes may affect bands; no particular cause is identified.
Wavelength shift	Interpolation by 0.2 or 1 nm mimics registration error; endpoints are held constant.
Increased noise	Raising the multiplier from 0.5 to 2 or 3 stresses repeatability. The covariance is constructed.
Instrument replacement	Paired readings anchor the endpoints; interpolated differences are not observed aging.
Signed step or ramp	$\mathbf{d}_{\text{pred}} = -s_0\mathbf{b}/\|\mathbf{b}\|^2$ induces a known response bias. Step magnitudes are 0.25, 0.5, 1 and 2; ramps reach 1 or 2 in 100 inspections.

3.4. Uncertainty Diagnostics and Independently Defined Changes

Uncertainty ordering is assessed on external records using the Spearman correlation between $\tilde{s}(\mathbf{x})$ and absolute prediction error. Root mean square error (RMSE), interval width and coverage are summarized descriptively for four groups ordered by the scale. Equal-sized groups are formed separately within each dataset; the five corn records per group preclude a strong population-level interpretation.

Two heteroscedastic profiles assign larger noise either to materials above or below the training median of mean spectral intensity. The factors are 1.5 and 0.5, divided by their root mean square over the control panel, and multiply the noise term in Equation (19). This grouping does not use the fitted prediction, bootstrap disagreement or detector outcome. Both orientations, all three fitted designs, bias steps of $0.25s_0$, $0.5s_0$ and s_0, and the noise increase to $\eta = 2$ are retained. Each profile receives independent design, pilot, calibration and evaluation episodes of the same sizes as the main experiment. These are controlled heterogeneity experiments; the intensity–noise relation is not asserted as a measured instrument law.

An empirically anchored perturbation uses $\mathbf{d}_i = \mathbf{x}_i^{\text{secondary}} - \mathbf{x}_i^{\text{primary}}$ from paired instrument measurements. The fractions $g \in \{0.01, 0.05, 0.10, 0.25, 0.50, 1\}$ are fixed across datasets and applied after τ. The full difference is physically observed; intermediate fractions interpolate that difference and do not represent dated trajectories or certified drift tolerances. Unlike the masking direction, these spectral shapes are not constructed from model coefficients. Each fraction is accompanied by its spectral RMS relative to σ_X, induced prediction RMS relative to s_0, and detection estimates. Representative spectra make the much larger masking stress visible.

Panel transport is examined through reference-response ranges and residual bias and dispersion. For the primary design, two descriptive counterfactuals subtract the external-panel residual mean and then also multiply centered residuals by the panel-to-external residual standard-deviation ratio. They retain the original thresholds and uncertainty values. These transformations use the complete external panel for diagnosis and are not proposed as deployable corrections or as independent validation.

3.5. Comparators, Ablations and Performance Criteria

The twelve reference procedures and the additional centered diagnostic are specified in the supplementary procedure inventory. EWMA, CUSUM, PCA residual monitoring and multivariate EWMA (MEWMA) provide classical comparators [35,38–40]. The fixed mean-plus-energy ablation is essential: it distinguishes the value of uncertainty adaptation from the benefit of simply monitoring two error features. The CPE-h and PH-CD-h implementations harmonize recent conformal EWMA and the earlier chemometric committee detector with the present episode design [28,29]. They are not claimed to reproduce the authors' original datasets or numerical results.

Max-EWMAMS-h retains the published individual-observation statistic [30], using the fixed reference scale $s_0\sqrt{m_{2,0}}$ and $\lambda = 0.1$. It receives the same warm-up and episode calibration as the other methods. Its dispersion component is two-sided, whereas FR-U monitors increased energy. Max-mixed EWMA [31] requires subgroup information, so assigning it groups of repeated spectra would change the inspection unit and reference budget; it is compared conceptually rather than numerically. The supplement gives a source-to-implementation correspondence and an independent reference-setting check for Max-EWMAMS-h.

Online-scale updates its squared denominator as $v_t = 0.95v_{t-1} + 0.05(r_t/s_0)^2$ after using v_{t-1}, with a lower scale bound of 0.25 and no upper bound. CPE-h uses a 29-observation weighted window and a fixed score-reference sample from the independent design generator. PH-CD-h uses a fading factor of 0.99, log-transformed disagreement and Welford moment updates to specify an unambiguous implementation of the moment recursions. Full recursions are provided in the supplement. Spectral-only methods receive less contemporaneous reference information and address a broader spectral-change endpoint; their speed is therefore not interpreted as a like-for-like demonstration of analytical-error detection.

For a normal condition, the primary estimate is $\widehat{\Pr}(T \le H)$. We report exact binomial intervals with Bonferroni allocation over the four regimes when summarizing the worst regime. These intervals describe Monte Carlo error conditional on the fitted model, panel and calibrated threshold. They do not quantify uncertainty from unobserved laboratories or instruments.

For a changed condition, pre-change alarms are recorded separately. Among episodes surviving to τ, we report detection probability by H, mean delay among detected episodes, and the restricted delay

$$D_H = \mathbb{E}[\min\{T - \tau + 1, H - \tau + 2\} \mid T \ge \tau]. \tag{22}$$

No detection is coded as $T = H + 1$, giving a penalty of 151 inspections. By retaining nondetections, this metric prevents a detector from appearing fast merely because it detects only a few easily detectable changes. It is not an uncensored ARL. Smoothing values 0.05 and 0.2 and a reduced uncertainty cap of $2s_0$ are assessed with fresh pilot and final calibration episodes and the same evaluation size. These sensitivity configurations are reported in full rather than used to replace the main specification.

Detection probabilities have Wilson 95% intervals and an explicit eligible-episode denominator. Restricted delays and method differences use 3000 bootstrap resamples of the

shared episode indices. Each resample recalculates each method's own survival condition and retains the nondetection penalty, rather than restricting the comparison to a common subset of surviving episodes. Intervals are pointwise and conditional on the fitted design and deployed thresholds. If the bootstrap interval for a difference in detection probabilities is degenerate, conservative bounds are formed from two 97.5% exact binomial intervals; a bounded-mean Hoeffding construction similarly handles degenerate delay distributions. Thus observed 0% or 100% rates do not imply exact population probabilities. These Monte Carlo intervals are separate from the tolerance calibration and from the descriptive variation across fitted designs.

For descriptive comparison with Fixed-GV, a delay difference is classified as a benefit when its entire 95% interval is below zero, as harm when the interval is entirely above zero, and as inconclusive otherwise. Empty conditional populations remain undefined. Detection-probability differences are classified separately, with the signs reversed. These labels summarize the existing paired estimates; they neither establish practical importance nor provide simultaneous error control across scenarios. An inconclusive result, including a tie caused by nondetection, is not evidence of equivalence. All classifications are retained in the supplementary outputs.

First-crossing attribution uses the deployed FR-U threshold for each branch. Simultaneous crossings are retained as separate categories. We report first crossings by the adaptive branch alone and detections contributed by A_t when neither fixed branch crosses by H. A further counterfactual evaluates Fixed-GV at the FR-U threshold, separating the adaptive branch's contribution at a common limit from the cost of raising the fixed-only limit. These comparisons characterize alarm generation; they do not establish the physical cause of a fault.

3.6. Additional Experiments

The supplementary protocols specify the wavelength-defined shapes, independent residual–scale coupling, replacement-panel calibration and signed-bias experiments. The full signed grid uses $\Delta/s_0 = \pm 0.25, \pm 0.5, \pm 1, \pm 2$, paired across signs and methods, with the original fitted models and limits frozen. The centered diagnostic estimates its offset from 1000 independent normal design episodes at $(\eta, \rho) = (0.5, 0)$ and uses fresh pilot scales and joint calibration. It is outside the simultaneous tolerance family.

Three exploratory checks address different sources of variation. First, we resample entire records within each fixed panel and evaluate the deployed FR-U and Fixed-GV charts. Second, we compare five chart families using independent tuning episodes, the same objective across five changes, fresh final calibration and held-out evaluation. Third, we repeat PLS selection by nested cross-validation and compare four preprocessing alternatives with the archived-input pipeline. Each experiment has a distinct estimand; none selects a replacement for the reference design. Protocols, all settings and machine-readable results are in the supplementary methods on design sensitivity.

3.7. Reproducibility and Computational Checks

The supplementary package contains the source code, protocol, all aggregate results, random seeds, archive checksums and figure-generation scripts. Experiments used Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0 and scikit-learn 1.8.0 with one numerical-library thread. Linear spectral features are propagated algebraically for efficiency. Independent reconstruction of perturbed spectra verifies the residual and PCA residual calculations to maximum absolute errors of 4.1×10^{-14} and 1.4×10^{-17}, respectively. Additional checks verify fixed-branch invariance, the ramp identity and the selected binomial tolerance rank.

The reference pipeline was rerun in a new virtual environment with pinned dependencies and an empty results directory. Recomputed numerical values matched the recorded values exactly across 99 result files, including all saved integer stopping times. Generated tables, figure renders and decision classifications were also checked against their source outputs. Machine-readable execution records, a comparison script and a result-to-source index are included in the supplement. The signed-bias extension separately reconstructs all original thresholds, verifies positive-case stopping times and intervals, and checks the centered recurrence and gain–cost identities. Its execution and reproduction evidence are supplied separately. These audits establish computational reproducibility within the recorded software environment, rather than validation of the assumed laboratory process.

4. Results

4.1. Prediction, Uncertainty Ordering and False-Alarm Calibration

Table 2 reports predictive performance for the primary fitted designs. PLS selected 12 components for corn and six for tablets. External RMSE was 0.0163 and 5.1711 in the respective native response units. Nominal 90% intervals covered 19 of 20 corn records and 441 of 460 tablet records. These predictive diagnostics do not establish sequential false-alarm validity.

The corn design-14 external response range (9.673–10.826) is narrower than the training range (9.407–10.993). Twelve components were selected at the upper grid boundary in every corn design. Nested fivefold validation gave RMSE 0.0220 in design 14 versus 0.0205 after single-loop selection; the external RMSE of 0.0163 alone does not establish stability. Five of 20 external corn spectra exceeded the empirical 95th percentile of training leverage and eight exceeded the corresponding spectral residual threshold. There was no outlier deletion. Complete response ranges, CV curves, slopes and performance measures are in the supplement.

On the external records, the Spearman correlations between uncertainty scale and absolute prediction error were 0.272 for corn and 0.153 for tablets. Grouped errors were not monotonically ordered by uncertainty. For tablets, mean interval width increased from 12.42 to 32.74 between the lowest- and highest-uncertainty groups, while coverage increased from 91.3% to 98.3%. Aggregate coverage therefore did not reflect differences in uncertainty calibration across groups. The corn estimates are particularly uncertain because each group contains only five records.

Coverage counts how often an interval contains its reference; ranking asks whether the largest predicted uncertainties identify the largest individual errors. A broad interval can cover most observations even with poor ranking. Changing the number of PCA scores from five to three changed the primary corn rank correlation from 0.27 to 0.61 without changing coverage (19/20). That difference on 20 records is not evidence for choosing the three-score map.

Table 2. PLS calibration in design 14. Nested and external columns give RMSE in response units; s_0 is the selected training out-of-fold RMSE. The final four columns describe external predictions: R^2, reference SD divided by RMSE (RPD), mean prediction minus reference, and slope of prediction on reference. Panel and training ranges and all six fits are reported in the supplement.

Data	LV	s_0	Nested	External	R^2	RPD	Bias	Slope
Corn	12	0.0205	0.0220	0.0163	0.998	21.09	0.0027	0.989
Tablets	6	5.15	5.38	5.17	0.892	3.05	−0.13	0.936

Across the four accepted normal regimes, the largest observed FR-U alarm proportions were 4.7% for corn and 5.3% for tablets. For the other procedures, the corresponding worst-

regime alarm proportions ranged from 4.1% to 5.4% and from 3.9% to 5.5%, respectively. These estimates indicate comparable false-alarm levels, but do not establish identical true risks. Taking the maximum threshold across regimes makes lower-noise conditions conservative. The tolerance rank of 1929 reduced the worst-regime FR-U estimates to 3.2% and 3.6%. Across the three fitted designs, worst-regime rank-calibration estimates ranged from 4.7% to 5.9%; this descriptive variation is separate from conditional Monte Carlo uncertainty.

The supplement reports FR-U and Fixed-GV separately for every accepted regime and fitted design under both calibration rules. A nominal 5% episode target does not require each empirical estimate from 1000 evaluation episodes to equal 5%, nor does a point estimate slightly above 5% alone contradict the marginal calibration statement.

4.2. Detection Probability and Paired Differences

Table 3 compares the restricted delays directly. All twelve procedures and their eligible counts are reported in the supplement. For a positive unit-s_0 bias, FR-U detected all primary episodes, with delays of 27.1 inspections for corn and 31.0 for tablets. Relative to EWMA, it was slower by 1.8 [1.3, 2.5] and 6.8 [6.1, 7.5] inspections; relative to CUSUM, it was slower by 5.2 [4.5, 5.9] and 9.6 [8.7, 10.4]. Brackets throughout this subsection are pointwise 95% Monte Carlo intervals.

Table 3. Restricted delay in inspections for the reference settings in design 14. The step is a positive unit bias, the masking ramp ends at two units, and the noise multiplier increases to two. Lower values are faster.

Method	Corn Bias	Mask	Noise	Tablet Bias	Mask	Noise
FR-U	27.1	60.1	17.3	31.0	58.2	17.0
A-only	20.0	151.0	82.7	118.1	151.0	105.6
Fixed-GV	33.8	59.2	16.9	28.7	56.6	16.4
EWMA	25.2	54.6	71.3	24.2	53.7	65.3
CUSUM	21.9	53.4	64.3	21.5	53.4	60.4
Max-EWMAMS-h	102.9	73.6	15.6	93.3	71.7	16.0

Under increased noise, the published joint comparator, Max-EWMAMS-h, reduced delay relative to FR-U by 1.7 [1.3, 2.2] inspections for corn and 1.0 [0.7, 1.3] for tablets. Its unit-bias detection probabilities were lower, 64.2% and 75.3%, with restricted delays of 102.9 and 93.3. These outcomes reflect the specific statistic and common episode calibration, including the cost of two-sided dispersion surveillance; they do not rank families of joint charts as a whole.

Paired comparisons with Fixed-GV isolate the incremental value of the uncertainty branch. In corn design 14 at positive half-unit bias, FR-U detected 212 of 1000 eligible episodes by the horizon, compared with 103 of 1000 for Fixed-GV, a gain of 10.9 [9.0, 12.8] percentage points. Restricted mean delay was 136.79 versus 144.13 inspections, a reduction of 7.3 [5.9, 8.8]. Missed changes receive a delay of 151, so the difference mainly reflects the larger fraction detected. Among the episodes each chart detected, mean delays were 83.98 and 84.31 inspections, respectively. Those means refer to different subsets and do not measure how much earlier one chart warned on the same episodes. At a unit bias the reduction was 6.7 [5.9, 7.6]. On tablets, adding the uncertainty branch increased delay in the four positive-bias, masking and noise contrasts in the supplementary paired table. Both datasets also showed a small cost on the masking and noise alternatives. The supplementary tables give probabilities, eligible denominators and paired intervals.

4.3. Sign Reversal and Centered Adaptive Innovations

Positive and negative residuals represent underprediction and overprediction, respectively, relative to the established bias. If an assay is used near a release limit, either sign may have the greater cost, depending on the decision. We therefore report both directions for every fitted design. The gain at positive half-unit bias in corn design 14 was not reproduced in designs 15 and 16; either would have reversed the conclusion from a single illustrative design.

Reversing the bias changes the incremental value of uncertainty (Table 4, Figure 2). For the primary half-unit corn bias, FR-U reduced delay by 7.34 inspections for the positive shift but increased it by 1.40 [0.70, 2.17] for the negative shift. In tablets, the positive shift increased delay by 3.19 [2.27, 4.24], whereas the negative shift reduced it by 5.11 [3.37, 6.79] and increased detection by 6.8 [4.7, 8.8] percentage points. Thus the positive-bias findings do not establish an absence of adaptive benefit in tablets. Across their twelve negative-bias design–magnitude combinations, four delay contrasts favor FR-U and eight are inconclusive; across the twelve positive combinations, nine favor Fixed-GV and three are inconclusive. Corn has seven benefits, one harm and four inconclusive negative-bias contrasts, versus four of each classification for positive biases. These counts describe pointwise evidence, not simultaneous superiority or practical importance.

Table 4. All three fitted designs: FR-U minus Fixed-GV restricted delay (inspections) for negative and positive half-unit bias. Brackets are pointwise Monte Carlo intervals conditional on each model, panel and realized limits. They do not cover material or instrument variation.

Data	Design	$-0.5s_0$	$+0.5s_0$
Corn	14	1.4 [0.7, 2.2]	−7.3 [−8.8, −5.9]
Corn	15	−2.6 [−3.8, −1.6]	2.3 [1.5, 3.3]
Corn	16	−2.9 [−4.2, −1.7]	0.8 [0.2, 1.5]
Tablets	14	−5.1 [−6.8, −3.4]	3.2 [2.3, 4.2]
Tablets	15	−0.1 [−0.9, 0.8]	1.9 [1.2, 2.7]
Tablets	16	0.2 [−0.7, 1.2]	2.7 [1.8, 3.6]

The estimated N05 centers were 0.05646 for primary corn and −0.13617 for primary tablets, with design Monte Carlo standard errors 0.00171 and 0.00246. In independent N05 evaluation episodes, the original weighted innovation means were 0.05581 and −0.13451; after centering they were −0.00064 and 0.00166. Centering did not remove every regime-specific offset: the primary tablet means became 0.01117 in N0 and −0.02770 in N1. Across all six designs, the largest normal-regime FR-U-c alarm estimates ranged from 4.5% to 5.9%. The supplement reports all regime-specific means, alarm intervals, pilot constants and limits.

Centering redistributed the corn advantage. In its primary design, FR-U-c reduced delay relative to Fixed-GV by 2.11 inspections at $-0.5s_0$ and 4.13 at $+0.5s_0$. The paired positive-minus-negative contrast changed from −8.74 [−10.36, −7.11] for FR-U to −2.02 [−3.54, −0.38] for FR-U-c. At unit magnitude, that contrast changed from −6.42 [−7.44, −5.49] to 0.78 [−0.33, 1.82]. Across the full corn grid, FR-U-c has fourteen delay benefits and ten inconclusive comparisons with Fixed-GV. This does not establish equivalence in the inconclusive cases, nor performance outside this grid.

Tablet asymmetry persisted after centering: the half-unit sign contrast was 8.30 [6.35, 10.20] for FR-U and 9.01 [7.04, 11.04] for FR-U-c. At the negative half-unit bias, the original common-limit adaptive gain was 8.59 inspections and the joint-limit cost was 3.48, leaving a net gain of 5.11. With centering, these became 9.45 and 3.81, leaving 5.65. At the positive half-unit bias, neither variant added a detection advantage through its adaptive branch at the common limit; the net delay increase equaled the limit cost, 3.19 or 3.36. Baseline

displacement therefore contributes to some directional differences but does not explain all of them. The uncertainty map, signed residual distribution and calibration cost remain relevant. The original procedure is retained as the reference; these outcomes do not select the centered variant for deployment.

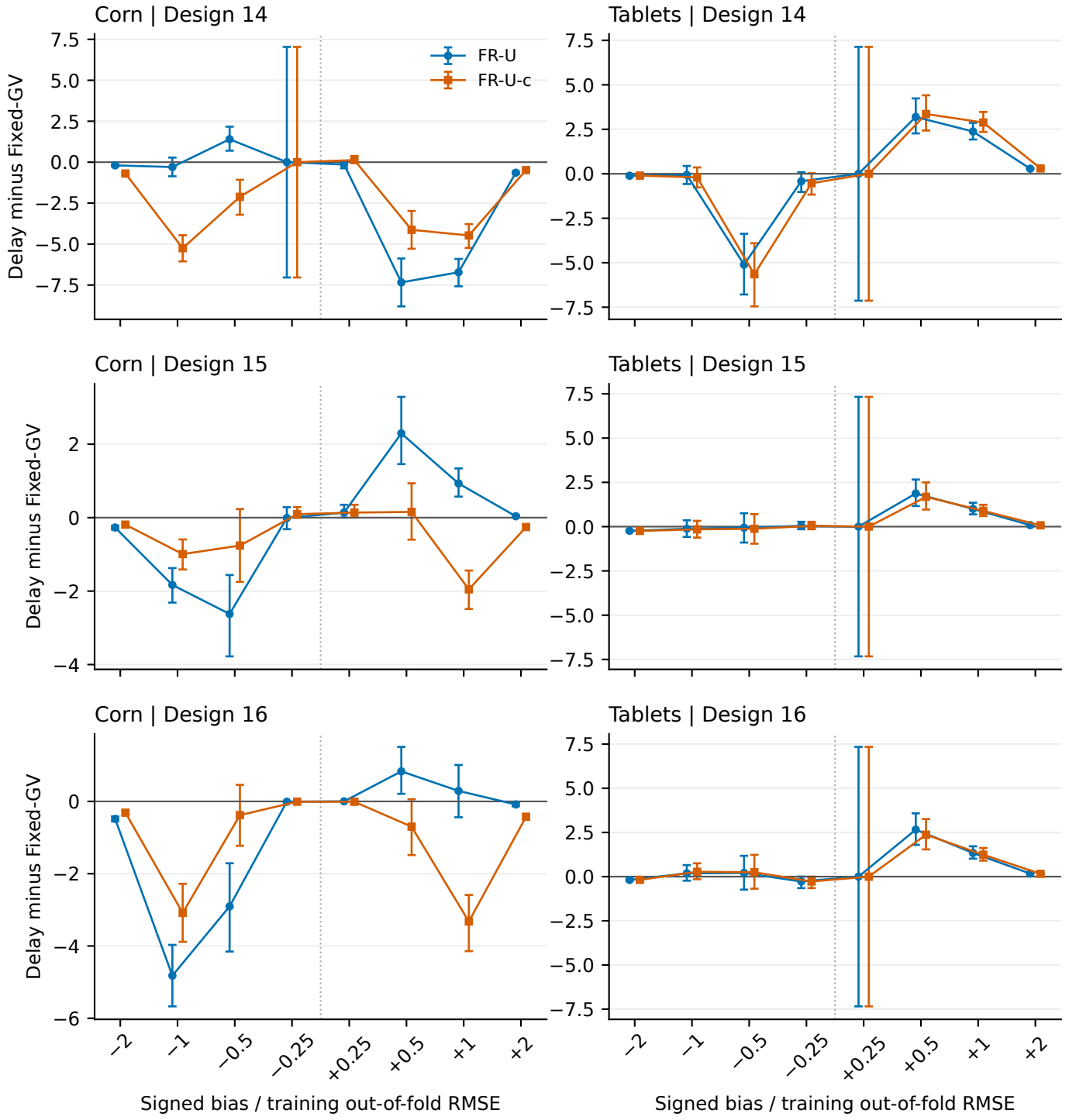

Figure 2. Complete symmetric-bias grid for the three fitted designs in each dataset. Points show restricted-delay differences from Fixed-GV; bars are paired pointwise 95% intervals. The horizontal line denotes zero difference, not an equivalence margin. Original limits remain frozen; only the centered diagnostic receives a new adaptive pilot scale and joint limit.

4.4. What the Adaptive Branch Contributes

At $+0.5s_0$ in corn design 14, all 1000 episodes survived to the change. Of the 212 detected by FR-U, 160 first crossed through A alone, four through G alone and 48 through both simultaneously. In 127 of those 212 episodes, neither fixed branch reached the deployed FR-U limit by inspection 200; in the other 33 adaptive-first episodes a fixed branch crossed later. At the common limit, the adaptive branch saved 8.57 inspections in restricted delay. The higher joint limit cost 1.23, leaving a net reduction of 7.34 relative to Fixed-GV. At $+1s_0$, all 1000 eligible episodes signaled, but only one had no fixed-branch crossing by the horizon; the corresponding gain, cost and net reduction were 8.21, 1.49 and 6.72 inspections. For the two positive tablet biases in design 14, A did not contribute a first crossing. The masking ramp first triggered G in all 1000 episodes of both datasets, while increased noise first triggered V alone in 926 corn and 910 tablet episodes.

The independently defined heteroscedastic profiles did not establish a general adaptive advantage. At a half-unit bias, both orientations favored FR-U over Fixed-GV in the primary corn design, but neither favored FR-U in either of the other corn designs. All six tablet profile–design combinations favored Fixed-GV. Worst-regime FR-U false-alarm estimates across these profiles ranged from 4.1% to 6.3%. All profile orientations, seeds and perturbation magnitudes are retained in the supplementary outputs. Heterogeneous noise alone therefore does not justify adding the uncertainty branch.

Table 5 relates uncertainty diagnostics to the gain–cost decomposition across all six fitted designs. The first two corn designs had positive external rank correlations of 0.272 and 0.220, yet their half-unit delay differences had opposite signs. All three corn values of $\mathcal{E}_W$ exceeded one, but only the primary design benefited for this alternative. These diagnostics do not account for changes in the uncertainty scale under the alternative or for the joint threshold. Consequently, neither positive error ordering nor the local efficiency approximation is a sufficient selection criterion.

Table 5. Uncertainty diagnostics and half-unit-bias performance. Seed* denotes the last two digits of the fitted-design seed. ρ_E is the external scale–absolute-error Spearman correlation; $\mathcal{E}_W$ uses an independent normal-episode draw. Gain and cost use the common-limit decomposition, in inspections. ΔD_H is FR-U minus Fixed-GV, with the paired 95% confidence interval.

Data	Seed*	ρ_E	$\mathcal{E}_W$	Gain	Cost	ΔD_H [95% CI]
Corn	14	0.272	1.079	8.57	1.23	−7.3 [−8.8, −5.9]
Corn	15	0.220	1.066	0.00	2.29	2.3 [1.5, 3.3]
Corn	16	−0.132	1.043	0.24	1.07	0.8 [0.2, 1.5]
Tablets	14	0.153	0.951	0.00	3.19	3.2 [2.3, 4.2]
Tablets	15	0.129	1.019	0.00	1.87	1.9 [1.2, 2.7]
Tablets	16	0.124	1.015	0.00	2.66	2.7 [1.8, 3.6]

The independent-scale diagnostic further distinguishes scale variability from its association with analytical residuals. For the primary half-unit corn bias, removing this association reduced common-limit gain from 8.57 to 4.05 inspections and increased calibration cost from 1.23 to 1.88. The resulting FR-U delay increased by 5.17 [3.64, 6.75] inspections relative to the original coupling. In the other two corn designs, the same diagnostic instead reduced delay by 2.81 [1.84, 3.84] and 1.34 [0.60, 2.12]. It also reduced delay in all three tablet designs, although their independently coupled FR-U point estimates remained slower than Fixed-GV. The original association can therefore help or hinder monitoring; its mere existence is not evidence of useful uncertainty. Worst-regime alarm estimates for this diagnostic ranged from 4.8% to 6.6%, with uncertainty intervals reported in the supplementary outputs. Separate recalibration is part of the diagnostic, so the contrast measures both the change in information and its threshold cost.

4.5. Spectral Magnitude, Prediction Impact and Diagnostic Stresses

The supplementary impact table relates the spectral perturbations to prediction error. Identical perturbation labels do not imply equal analytical difficulty: a band of height $0.3\sigma_X$ induced a prediction impact of approximately $40.4s_0$ in corn but only $0.10s_0$ in tablets. A 0.2 nm wavelength change induced approximately $6.90s_0$ and $0.37s_0$, respectively. The corresponding detection proportions must be interpreted against these effects, rather than as evidence that every numerically small spectral perturbation is a small analytical fault. Figure 3 shows the full grid of ordinary perturbations.

The masking direction has spectral RMS $0.1388\sigma_X$ per unit ramp amplitude for corn and $14.2362\sigma_X$ for tablets. Consequently, the endpoint amplitude of two reaches $0.2777\sigma_X$ and $28.4724\sigma_X$. The tablet stress is extreme and does not represent plausible natural drift. The supplementary spectra illustrate this difference in scale. A-only detected no changes in any of the 1000 episodes per dataset, whereas FR-U and Fixed-GV detected changes in all of them. This result illustrates how uncertainty normalization can mask a change and how fixed-reference monitoring retains evidence of that change; it does not estimate the prevalence of such failures in laboratory use. The first generated trajectories and isolated uncertainty-inflation experiment are provided in the supplement.

The prediction-preserving band caused no primary FR-U alarms in either dataset. In corn, PCA-T2Q and MEWMA signaled in all episodes, illustrating that spectral novelty and deterioration of the fitted prediction are different endpoints. This constructed negative control does not establish chemical harmlessness.

The measured-direction grid supplies an alternative independent of the uncertainty-model construction. At 1% of the corn instrument difference, spectral RMS was $0.012\sigma_X$ and prediction impact was $0.708s_0$. FR-U detection was 75.1% [72.3, 77.7], compared with 55.5% [52.4, 58.6] for Fixed-GV. This advantage did not replicate: FR-U minus Fixed-GV differences at the same fraction were −0.2 [−0.6, 0.2] and −2.7 [−3.7, −1.7] percentage points in the additional corn designs. For tablets, 50% of the measured difference induced $0.508s_0$ with 5.2% FR-U detection, whereas the full difference induced $1.015s_0$ with 98.2% detection. These estimates use fresh evaluation episodes under the original thresholds; intermediate fractions represent interpolated changes rather than observed temporal drift.

4.6. Independent Spectral Shapes

The additional wavelength-defined shapes yielded a different pattern from the prediction-aligned bias grid (Table 6). The paired delay intervals supported FR-U benefits in 8, 4 and 4 of the twelve corn conditions, harm in 2, 6 and 5, and inconclusive differences in 2, 2 and 3, respectively. Point-estimated differences ranged from −2.21 to 0.03, −2.21 to 2.87, and −8.25 to 0.56 inspections. Thus even the third corn design, which did not benefit on the positive bias steps, benefited for selected negative prediction shifts induced by the paired-band direction. Conversely, a favorable point estimate in the primary design did not always have an interval excluding zero. These counts describe the complete grid using pointwise intervals; they are not independent replications or simultaneous significance tests.

There was no point-estimated FR-U advantage in any of the 36 tablet shape–design conditions. Each tablet design had two conditions classified as harm and ten as inconclusive. Many inconclusive cases were ties caused by nondetection: tablet endpoint prediction impacts ranged from approximately $0.00016s_0$ to $0.681s_0$, compared with $0.371s_0$ to $23.17s_0$ in corn. Equal spectral RMS therefore did not imply equal analytical severity. Across all 72 conditions, the delay intervals supported 16 benefits and 19 harms, with 37 inconclusive comparisons. These totals do not weight application relevance or effect size. The broader grid supports direction-dependent adaptive value, while providing no evidence of general

improvement in tablets. The supplement gives the full primary grid and machine-readable results for all designs and procedures.

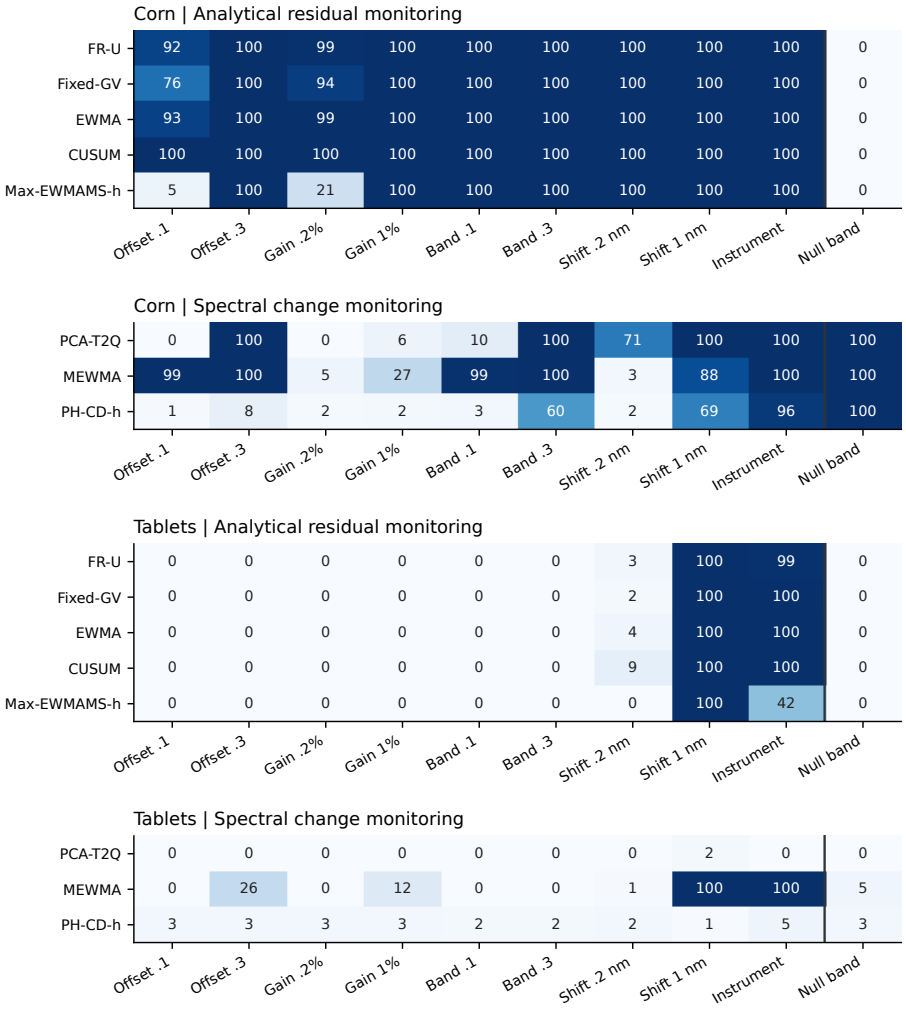

Figure 3. Detection percentages under primary calibration, grouped by monitoring endpoint. Darker blue indicates a higher detection percentage (0–100%). The number in each cell gives the corresponding percentage. Analytical-residual charts and spectral-change charts are displayed separately. The last column, separated by a vertical line, is the prediction-preserving band; preceding columns are ordinary perturbations with effects quantified in the supplementary impact table. Offset and band height use σ_X; gain and wavelength use the displayed units. Instrument is m5-to-mp5 for corn and instrument 1-to-2 for tablets. Full twelve-procedure results, intervals and eligible counts remain in the supplement.

Table 6. All twelve independent-shape alternatives per design. Impact is the absolute endpoint prediction change relative to s_0. Benefit/harm means that the pointwise 95% interval for FR-U minus Fixed-GV restricted delay lies entirely below/above zero; unclear means inconclusive, not equivalent. Ranges give point estimates in inspections. Counts are descriptive and do not establish simultaneous or practically important superiority. Seed* denotes the last two digits of the fitted-design seed.

Data	Seed*	Impact/s_0	Benefit	Harm	Unclear	ΔD_H Range
Corn	14	0.411–17.1	8	2	2	$[-2.2, 0.029]$
Corn	15	0.53–23.2	4	6	2	$[-2.2, 2.9]$
Corn	16	0.371–14	4	5	3	$[-8.3, 0.6]$
Tablets	14	0.000157–0.512	0	2	10	$[0.0, 3.5]$
Tablets	15	0.00513–0.681	0	2	10	$[0.0, 3.5]$
Tablets	16	0.0049–0.651	0	2	10	$[0.0, 4.8]$

4.7. Panel Transport and Design Sensitivity

Replacing the control panel without recalibration produced primary FR-U alarm proportions of 0.1% for corn and 59.7% for tablets without an injected instrumental fault. The range of tablet reference responses widened from 188.2–199.5 in the control panel to 154.3–237.7 in the external panel. Residual standard deviation increased from $0.743s_0$ to $1.003s_0$, with an external mean residual of $0.340s_0$. Subtracting this mean alone yielded 61.4% alarms; additionally rescaling dispersion yielded 33.6%, still far above 5%. Thus mean bias alone did not explain the transport failure. The external panel has a wider response range, but these diagnostics cannot separate the causal contributions of residual tails, residual–scale dependence and unrecorded sample characteristics.

Evaluation of all four noise/dependence settings on the replacement panel makes the transfer limitation more explicit (Table 7). With transported thresholds, worst-setting alarm proportions ranged from 6.6% to 47.3% in corn and from 77.8% to 85.1% in tablets. These are alarms without an injected fault under a changed panel law, not false-alarm estimates under the original calibrated law. Recalibrating only the threshold on independent replacement-panel episodes yielded worst-setting estimates of 4.0–5.9% and 4.4–5.7%, respectively, with intervals for the maximum regime risk that included 5%.

This recovery of the intended alarm budget had a substantial sensitivity cost. Tablet thresholds increased by factors of 3.66–4.05; unit-s_0 bias detection under the new panel and its recalibrated limits was only 14.9–18.0%. Corn responses also depended strongly on the partition, including 6.6% unit-bias detection in the second design. Accepting a broader panel as normal can therefore make a previously detectable statistical bias difficult to identify. Threshold recalibration restores a matching-law comparison; it does not restore the original analytical performance or establish acceptability of the new panel.

Table 7. FR-U under replacement-panel laws. Alarm columns are the largest percentage over four settings, with exact binomial Bonferroni bounds for that maximum. h_E/h_Q is the recalibrated-to-original threshold ratio. Detection and restricted delay concern a unit-s_0 bias on the replacement panel after recalibration. These simulations assume the replacement law has been accepted as normal; they do not justify accepting an unexplained process change. Seed* denotes the last two digits of the fitted-design seed.

Data	Seed*	Transported (%)	Recalibrated (%)	h_E/h_Q	Det. (%)	Delay
Corn	14	6.6 [4.8, 8.8]	5.9 [4.2, 8.0]	1.03	100.0	25.0
Corn	15	47.3 [43.3, 51.3]	4.9 [3.4, 6.9]	1.25	6.6	146.6
Corn	16	13.7 [11.1, 16.6]	4.0 [2.6, 5.8]	1.11	99.4	43.8
Tablets	14	85.1 [82.1, 87.8]	5.7 [4.0, 7.8]	3.66	15.7	139.6
Tablets	15	80.9 [77.6, 83.9]	4.4 [2.9, 6.3]	4.05	18.0	137.7
Tablets	16	77.8 [74.3, 81.0]	5.7 [4.0, 7.8]	3.87	14.9	140.0

Changing λ to 0.05 reduced the primary unit-bias and masking delays, whereas 0.2 increased them; noise detection followed a different trade-off. Reducing the scale cap to $2s_0$ changed some delays without removing the fixed branches. Holding residuals fixed while inflating uncertainty left G and V exactly unchanged. The complete sensitivity results, including normal-episode performance, are reported in the supplement; no fault-specific setting replaces the primary specification.

4.8. Sensitivity to Panel Composition, Tuning and Preprocessing

In 100 bootstrap panels of the original size, the estimated between-composition standard deviations of the FR-U minus Fixed-GV delay at positive half-unit bias were 9.6, 2.1 and 2.3 inspections across the corn designs; the tablet values were 3.3, 1.1 and 1.7. For corn design 14 the central 95% of the resampled estimates ranged from approximately −34 to

+0.2 inspections. This spread concerns repeated draws of the archived control records at frozen deployment settings. It is distinct from the much narrower Monte Carlo interval at the original panel and does not estimate variation across laboratories.

We tuned five chart families: FR-U, Fixed-GV, EWMA, CUSUM and Max-EWMAMS-h. For each family, we selected the setting with the smallest mean restricted delay over five equally weighted changes, then recalibrated it and evaluated it on fresh episodes. All 30 selected method–design pairs targeted the same 5% episode risk. The largest observed alarm rate across the four normal regimes ranged from 3.5% to 7.3% across those pairs. FR-U had the lowest held-out mean delay in corn designs 14 and 15; CUSUM had the lowest in corn 16 and tablet designs 15 and 16. In tablet design 14, the point estimates for Fixed-GV and CUSUM were nearly identical (62.50 and 62.59 inspections). This numerical ordering does not establish a meaningful difference for the assay. Table 3 gives the separate comparison at the fixed reference settings.

Preprocessing also changed the comparison. For the same positive half-unit input perturbation in primary corn, fresh calibration gave FR-U minus Fixed-GV delays of −12.8 inspections using archived spectra, −0.6 with standard normal variate correction and +2.4 with a first derivative. Each pipeline saw the same perturbed spectrum, so the induced error in its own response units can differ. The fresh archived-input result is a separate calibration realization, not a replacement for the −7.3 from the reference experiment. The complete grid includes changes of direction in tablets as well.

5. Discussion

Fixed-GV is the natural starting point for a laboratory that needs to monitor bias and increased error magnitude. It requires a frozen predictor, reference values for the controls and a calibrated joint limit. FR-U uses the same reference measurements; its additional requirements are an uncertainty map, a third pilot scale and calibration of the three-branch limit. At a common limit it cannot lose a fixed-branch crossing. Under the same nominal false-alarm budget its own limit can be higher, however. The practical question is whether the extra branch saves enough completed inspections to justify that cost.

At the reference settings, EWMA and CUSUM signaled the positive unit shift and the masking ramp sooner than FR-U in both primary designs. Fixed-GV and Max-EWMAMS-h were faster for increased noise. These are comparisons of specified implementations and alternatives. Fair tuning changes some rankings but yields no consistent best chart. Methods that use only spectral scores or model disagreement can operate without a reference value; they can also signal a change that leaves the fitted response intact. A laboratory interested in prediction error must distinguish this endpoint from a spectral warning.

The sign of an error matters near an analytical decision limit. With the residual convention used here, overprediction produces a negative residual; at a lower release limit, it can conceal a below-limit sample. At an upper limit, underprediction can conceal an above-limit sample. Before comparing charts, an analyst can specify the changes that threaten that decision, a useful improvement in delay, and the largest tolerable loss of sensitivity in the opposite direction. An equal-weight average is appropriate only when it represents those priorities. The centered diagnostic subtracts the average weighted error under a normal design law. This gives the adaptive branch a more neutral starting level in that law; it reduced some corn asymmetry, but the tablet difference persisted. It does not establish a preferred correction or justify one-sided monitoring.

A proposed adoption criterion is that the upper confidence bound on FR-U minus Fixed-GV delay should fall below a prespecified negative margin in the target scenarios and stay below an acceptable deterioration margin elsewhere. Detection must also meet a separate minimum. With zero permissible deterioration across the twelve independent

spectral shapes, none of the six fitted designs met even the pointwise delay screen. These margins require assay-specific costs and tolerances. A formal selection rule would need independent confirmation and simultaneous inference across its declared scenarios.

Panel composition is a separate source of uncertainty. Repeating a conditional episode a thousand times estimates the behavior of its generator precisely; it adds no new corn materials, manufacturing lots or instruments. The panel bootstrap shows how much results can move when only the frequencies of the available records change. It cannot supply uncertainty for unobserved materials or populations. Likewise, the two additional tablet fits reuse the same panel. Neither the narrow episode intervals nor the spread of three fits is a population interval.

Replacing a control panel is an analytical decision before it is a calibration exercise. A laboratory could run the old and proposed panels in parallel, match their response ranges, matrix types and handling conditions, and check predictions against independent reference assignments. Inspect bias, slope, residual spread, spectral leverage and reconstruction error against tolerances agreed for the assay. If performance is unacceptable, investigate the reference, instrument and prediction model before setting a new control limit. In tablets, recalibration brought no-fault alarms near the target but left detection of a unit bias at only 15–18%. A higher threshold cannot repair a biased or poorly transported predictor. This distinction complements work on calibration transfer and predictive coverage [2–5,15,18,19].

A feasible prospective study would rotate stable control materials spanning the working concentration range, with replicate preparations across days, batches and preferably instruments. For example, assay and seal several corn control lots, inspect one or more per batch, log storage, temperature, operator and instrument servicing, and periodically reassay retained aliquots against the reference method. Quarantine a control if its assignment changes. Use an initial stable period to choose settings and calibrate; freeze the procedure before a later, independently adjudicated monitoring period. Record both completed inspections and calendar time, missed controls and alarms without a confirmed analytical change. Report performance by lot and instrument as well as overall.

The error unit s_0 is a training prediction statistic. For prospective validation, specify acceptable positive and negative errors in assay units, account for uncertainty in the reference method, and test whether an alarm precedes a confirmed tolerance breach within a useful time. Stable-period specificity, missed breaches and the interval between a confirmed change and the first warning are distinct outcomes. No certified tolerance, deterioration rate or longitudinal change time can be obtained from the present archives.

6. Conclusions

Uncertainty weighting helped for particular directions and fitted designs. The corn gain at positive half-unit bias did not recur in the other two fits, and some negative tablet biases favored FR-U despite poorer results for positive shifts. The simple mean and joint comparators were faster for several reference alternatives. Independent tuning, preprocessing and panel resampling show why a favorable result on one fitted panel is insufficient for choosing a chart. The practical case for FR-U requires an improvement of useful size for the errors that matter to the assay, verified on stable reference materials and prospectively collected measurements.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/analytica7040108/s1. The Supplementary Materials provide source-to-implementation comparisons, comparator verification, uncertainty estimates for paired comparisons, branch attribution, diagnostics of spectral impact and panel differences, signed-bias and centered-innovation experiments, sensitivity analyses, justifications for historical references, a data-download

manifest, executable Python code, seeds and episode-level stopping times. The supplementary document is supplied with editable LaTeX source.

Author Contributions: Conceptualization, M.J.C.; methodology, M.J.C., R.A.-V. and E.D.-D.; software, M.J.C.; validation, M.J.C. and R.A.-V.; formal analysis, M.J.C., R.A.-V. and E.D.-D.; investigation, M.J.C., R.A.-V. and E.D.-D.; data curation, M.J.C. and R.A.-V.; writing—original draft preparation, M.J.C. and R.A.-V.; writing—review and editing, M.J.C., R.A.-V. and E.D.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was paid by the authors.

Institutional Review Board Statement: Not applicable. This computational study uses public, non-human spectroscopic benchmark data.

Informed Consent Statement: Not applicable.

Data Availability Statement: Original spectral archives are publicly available from Eigenvector Research at https://eigenvector.com/resources/data-sets/ (accessed on 3 October 2026). Exact archive URLs and SHA-256 checksums are provided in the Supplementary Data manifest. Code, seeds, perturbation definitions and aggregate simulation outputs are included in the Supplementary Materials. Third-party raw archives are retrieved from the original provider by the supplied download script and are not redistributed in the package.

Acknowledgments: The authors thank the Universidad del Magdalena, Santa Marta, Colombia, for its institutional support of this research.

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

References

1. Mokari, A.; Guo, S.; Bocklitz, T. Exploring the Steps of Infrared (IR) Spectral Analysis: Pre-Processing, (Classical) Data Modelling, and Deep Learning. *Molecules* **2023**, *28*, 6886. [CrossRef] [PubMed]
2. Roger, J.M.; Diaz, V.F.; Nikzad-Langerodi, R. Cloning instruments, model maintenance and calibration transfer. *TrAC Trends Anal. Chem.* **2025**, *191*, 118319. [CrossRef]
3. Yap, X.Y.; Chia, K.S.; Suarin, N.A.S. Adaptive Artificial Neural Network in near infrared spectroscopy for standard-free calibration transfer. *Chemom. Intell. Lab. Syst.* **2022**, *230*, 104674. [CrossRef]
4. Lavoie, F.B.; Robert, G.; Langlet, A.; Gosselin, R. Calibration transfer by likelihood maximization: A standard-free approach capable of handling non-overlapping wavelength ranges. *Chemom. Intell. Lab. Syst.* **2023**, *234*, 104766. [CrossRef]
5. Lai, J.; Li, M.; Chen, S.; Long, J.; Chen, Y.; Lu, H.; Zou, C.; Zhang, Z. Calibration Transfer of Deep Learning Models among Multiple Raman Spectrometers via Low-Rank Adaptation. *Anal. Chem.* **2025**, *97*, 19009–19018. [CrossRef] [PubMed]
6. Pedge, N.I.; Papillaud, M.; Roger, J.M. Investigation of long-term stability of a transmission Raman calibration model using orthogonal projection methods. *Chemom. Intell. Lab. Syst.* **2024**, *248*, 105115. [CrossRef]
7. Zhang, Z.; Wang, L.; Wen, Y.; Ma, X.; Zhen, B.; Han, L.; Liu, H.; Li, J.; Jing, Y.; Li, J. Real-Time Evaluation and Maintenance Method of Online Near-Infrared Detection Model. *J. Spectrosc.* **2026**, *2026*, 8525354. [CrossRef]
8. Wadoux, A.M.C.; Ramirez-Lopez, L. Uncertainty of predictions in absorption spectroscopy: Modelling with quantile regression forest. *Chemom. Intell. Lab. Syst.* **2025**, *265*, 105473. [CrossRef]
9. Jovic, O. Conformal predictors in chemometric study of mid-infrared food adulteration: Quantification of prediction uncertainty. *Food Chem.* **2025**, *492*, 145387. [CrossRef] [PubMed]
10. Huang, Y.C.; Padarian, J.; Minasny, B.; McBratney, A.B. Using Monte Carlo conformal prediction to evaluate the uncertainty of deep-learning soil spectral models. *SOIL* **2025**, *11*, 553–563. [CrossRef]
11. Duma, Z.S.; Lamminpää, O.; Susiluoto, J.; Haario, H.; Zheng, T.; Sihvonen, T.; Braverman, A.; Townsend, P.A.; Reinikainen, S.P. Uncertainty calibration for latent-variable regression models. *Chemom. Intell. Lab. Syst.* **2026**, *273*, 105711. [CrossRef]
12. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [CrossRef]
13. Barber, R.F.; Candès, E.J.; Ramdas, A.; Tibshirani, R.J. Conformal prediction beyond exchangeability. *Ann. Stat.* **2023**, *51*, 816–845. [CrossRef]

14. Gibbs, I.; Candès, E. Adaptive Conformal Inference Under Distribution Shift. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 1660–1672.
15. Gibbs, I.; Candès, E.J. Conformal Inference for Online Prediction with Arbitrary Distribution Shifts. *J. Mach. Learn. Res.* **2024**, *25*, 1–36.
16. Zaffran, M.; Feron, O.; Goude, Y.; Josse, J.; Dieuleveut, A. Adaptive Conformal Predictions for Time Series. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Volume 162, pp. 25834–25866.
17. Xu, C.; Xie, Y. Sequential Predictive Conformal Inference for Time Series. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; Volume 202, pp. 38707–38727.
18. Angelopoulos, A.N.; Candès, E.J.; Tibshirani, R.J. Conformal PID Control for Time Series Prediction. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2023; Volume 36, pp. 23047–23074. [CrossRef]
19. Sun, S.; Yu, R. Conformal Prediction for Time-series Forecasting with Change Points. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2025; Volume 38, pp. 12243–12272. [CrossRef]
20. Bates, S.; Candès, E.; Lei, L.; Romano, Y.; Sesia, M. Testing for outliers with conformal p-values. *Ann. Stat.* **2023**, *51*, 149–178. [CrossRef]
21. Yi, F.; Qiu, P. An adaptive CUSUM chart for drift detection. *Qual. Reliab. Eng. Int.* **2022**, *38*, 887–894. [CrossRef]
22. Ahad, N.; Davenport, M.A.; Xie, Y. Data-adaptive symmetric CUSUM for sequential change detection. *Seq. Anal.* **2024**, *43*, 1–27. [CrossRef] [PubMed]
23. Cobb, O.; Van Looveren, A.; Klaise, J. Sequential Multivariate Change Detection with Calibrated and Memoryless False Detection Rates. In Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, Online, 28–30 March 2022; Volume 151, pp. 226–239.
24. Vovk, V.; Nouretdinov, I.; Gammerman, A. Validity and efficiency of the conformal CUSUM procedure. In Proceedings of the Fourteenth Symposium on Conformal and Probabilistic Prediction with Applications, London, UK, 10–12 September 2025; Volume 266, pp. 576–594.
25. Shin, J.; Ramdas, A.; Rinaldo, A. E-detectors: A Nonparametric Framework for Sequential Change Detection. *N. Engl. J. Stat. Data Sci.* **2024**, *2*, 229–260. [CrossRef]
26. Shekhar, S.; Ramdas, A. Reducing sequential change detection to sequential estimation. In Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024; Volume 235, pp. 44628–44642.
27. Wang, H.; Xie, Y. Sequential change-point detection: Computation versus statistical performance. *WIREs Comput. Stat.* **2024**, *16*, e1628. [CrossRef]
28. Li, G.; Deng, Z.; Tian, M.; Ju, Y. A distribution-free EWMA control chart based on conformal p-values. *Stat. Pap.* **2026**, *67*, 109. [CrossRef]
29. Nikzad-Langerodi, R.; Lughofer, E.; Cernuda, C.; Reischer, T.; Kantner, W.; Pawliczek, M.; Brandstetter, M. Calibration model maintenance in melamine resin production: Integrating drift detection, smart sample selection and model adaptation. *Anal. Chim. Acta* **2018**, *1013*, 1–12. [CrossRef] [PubMed]
30. Ostadsharif Memar, A.; Niaki, S.T.A. The Max EWMAMS control chart for joint monitoring of process mean and variance with individual observations. *Qual. Reliab. Eng. Int.* **2011**, *27*, 499–514. [CrossRef]
31. Malik, S.; Hanif, M.; Noor-ul Amin, M.; Khan, I.; Ahmad, B.; Abaker, A.O.I.; Ahmed Darwish, J. Max-mixed EWMA control chart for joint monitoring of mean and variance: An application to yogurt packing process. *Sci. Rep.* **2024**, *14*, 10372. [CrossRef] [PubMed]
32. Zhang, X.; Jia, Y.; Yang, Z.; Sheng, L.; Yuan, L.; Zhang, Q.; Yang, D. Spectral intensity drift correction of Spark Mapping Analysis for large-size metal materials. *Anal. Chim. Acta* **2024**, *1322*, 343075. [CrossRef] [PubMed]
33. Schiemer, R.; Rüdt, M.; Hubbuch, J. Generative data augmentation and automated optimization of convolutional neural networks for process monitoring. *Front. Bioeng. Biotechnol.* **2024**, *12*, 1228846. [CrossRef] [PubMed]
34. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
35. Roberts, S.W. Control Chart Tests Based on Geometric Moving Averages. *Technometrics* **1959**, *1*, 239–250. [CrossRef]
36. Wilks, S.S. Determination of Sample Sizes for Setting Tolerance Limits. *Ann. Math. Stat.* **1941**, *12*, 91–96. [CrossRef]
37. Eigenvector Research. Data Sets: Corn and Pharmaceutical Tablet Spectra. Source Archive Checksums Are Provided in the Supplementary Materials. 2026. Available online: https://eigenvector.com/resources/data-sets/ (accessed on 3 October 2026).
38. Page, E.S. Continuous Inspection Schemes. *Biometrika* **1954**, *41*, 100–115. [CrossRef]

39. Jackson, J.E.; Mudholkar, G.S. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics* **1979**, *21*, 341–349. [CrossRef]
40. Lowry, C.A.; Woodall, W.H.; Champ, C.W.; Rigdon, S.E. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* **1992**, *34*, 46–53. [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.