



Certified proximal majorization for robust parameter estimation in differential equations

Alfred Daci¹, Tatyana Vokhmintceva², Klodian Dhoska³, Marek Dudek⁴, Aleksander Vokhmintcev⁵, Mostafa Abotaleb^{2,*}, and Nigar Nagiyeva^{6,*}

¹Department of Mathematical Engineering, Polytechnic University of Tirana, Albania.

²Engineering School of Digital Technologies, Yugra State University, Khanty-Mansiysk, Russia.

³Department of Mechanics, Polytechnic University of Tirana, Albania.

⁴AGH University of Krakow, Poland.

⁵Chelyabinsk State University, Chelyabinsk, Russia.

⁶Faculty of Physics and Mathematics, Department of Informatics, Nakhchivan State University, Nakhchivan, Azerbaijan.

Abstract

Stationarity computed from a numerical ODE solution need not persist for the continuous observation map. Theorem 8.1 transfers a feasible stationarity witness between these maps at the same parameter. For arctangent losses, explicit observation-map and sensitivity-error terms inflate the bound. Enlarging the residual tube preserves admissible slopes when discretization moves a residual across the nonsmooth cusp. The resulting decision rule separates further optimization from mesh refinement and records unresolved computations. Twelve scalar inequalities are verified pointwise by exact-rational enclosures. Numerical verification comprises 160 oscillator checks and a Khanty-Mansiysk January NDVI-related ODE calibration on three RK4 meshes; these use floating-point reference discrepancies, not validated integration bounds. The guarantee concerns local stationarity for the enlarged tube. It transfers neither an optimizer nor parameter accuracy, and a smaller stationarity bound does not establish better prediction.

Keywords. ODE calibration, Stationarity-certificate transfer, Residual tube, Nonsmooth optimization, Sensitivity error, Arctangent loss.

2010 Mathematics Subject Classification. 65L09, 65L06, 65K10, 49J52, 90C26.

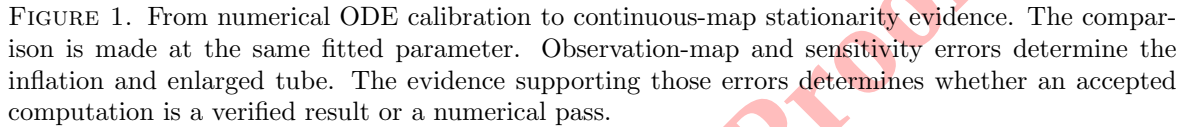
1. INTRODUCTION

An ODE calibration algorithm fits parameters using a numerical solution of the differential equation. It therefore optimizes an approximate observation map. A small stationarity residual for that map does not automatically imply a small stationarity residual for the continuous ODE model. Errors in both predicted observations and their parameter sensitivities can alter the first-order condition.

The principal contribution is Theorem 8.1: a cusp-aware transfer of a feasible stationarity witness from a discretized ODE observation map to its continuous counterpart at the same parameter. Explicit observation-map and sensitivity errors inflate the bound. The response tube is enlarged to accommodate residuals that may cross the cusp at zero. This enlargement is essential: an arbitrarily small residual shift can change its sign and admissible slope set. Here, “certified” refers to the stated stationarity bound under justified mathematical and arithmetic assumptions, not to the accuracy of an estimated parameter.

Received: 14 September 2026; Accepted: 08 October 2026.

* Corresponding authors. Emails: abotalebmostafa@bk.ru, abotalebmostafa@bk.ru, nigar.nagiyeva4@gmail.com, nigar.nagiyeva4@gmail.com.


$$\underbrace{R_{\tau+e_h,0}^{q,\lambda}}_{\text{continuous-map residual}} \leq \underbrace{\hat{R}_{\tau,h}^{q,\lambda}}_{\text{discrete stationarity bound}} + \underbrace{\sum_i a_i (E_{i,h}/c + Q_i L_q e_{i,h})}_{B_{q,h}: \text{ sensitivity and observation-map errors}}. \quad (1.1)$$

Terminology and evidence levels. A *certificate* is an upper bound justified by a feasible witness and stated assumptions. A *residual tube* allows componentwise response changes $|\Delta y_i| \leq \tau_i$. Its full cusp-slope interval gives a *backward-data certificate* for admissible nearby responses. Four evidence labels are used: *Proved* for mathematical deductions; *Validated computation* for justified arithmetic enclosures; *Numerical verification* for floating-point checks; and *Empirical observation* for measured fitting or forecasting behavior. These labels describe evidence. The algorithm’s outcomes are VERIFIED, NUMERICAL PASS and UNRESOLVED. The first requires justified bounds and witness evaluation; the second uses measured reference discrepancies. Table 9 collects the scope restrictions.

Robust M-estimation distinguishes bounded influence, leverage resistance, breakdown and efficiency [15, 25, 27, 40, 50, 51]. In dynamical calibration, heavy-tailed observation models address outliers [32], weak-form methods avoid differentiating noisy observations [10, 42], and sensitivity-enabled methods evaluate observation Jacobians [21, 29, 43]. Probabilistic integration provides another treatment of numerical uncertainty [6]. Related forward and inverse methods include fractional logistic, Galerkin–homotopy, alternating-direction, inverse spectral and fading-memory approaches [23, 33, 34, 39, 45]. The question here is whether local stationarity evidence survives a change of numerical observation map.



GLDM has been used for recurrent identification and forecasting [1–3, 36, 37]. Its loss applies arctangent to the absolute residual. This differs from Ceres’ squared-residual arctangent loss [12] and the unbounded residual contribution of ArcTanLoss [52]. Generalized correntropy already includes bounded similarities and a Laplace member with weighted-LAD majorization [13]. Proximal majorization and composite optimization also have established general theories [18, 26, 30, 31, 35, 47].

The specific contribution is the cusp-aware transfer of a computable feasible witness between numerical and continuous ODE maps, with explicit sensitivity-error inflation and response-tube enlargement. Weighted-LAD majorization and generic perturbation analysis are supporting tools, not the novelty claimed. Table 1 compares assumptions and conclusions at theorem level. The reciprocal GLDM representation is an exact loss–gain identity and an implementation check. Inverse-power GLDM instead changes the criterion; it is a separate loss-shape choice within the same transfer analysis.

TABLE 1. Theorem-level scope relative to existing optimization frameworks. The cited theorem numbers refer to the corresponding author preprints.

Result	Assumptions and conclusion	Relation to the present result
Prox-linear near-stationarity, Theorem 4.5 in [18]	Convex Lipschitz outer loss, smooth inner map with Lipschitz Jacobian; prox-gradient/Moreau-envelope comparison	The redescending outer loss is nonconvex. Theorem 5.3 uses a concave-in-magnitude tangent and an explicit response tube.
Inexact prox-linear convergence, Theorem 5.2 in [18]	Convex composite subproblems and objective-gap errors; step-residual complexity	Corollary 6.4 specializes strong-convexity gap control to the weighted-LAD model. This is supporting convergence analysis, not a new general gap principle.
Model BPG, Theorem 28 in [35]	Model-approximation property and its KL/Lyapunov hypotheses; whole-sequence convergence to a critical point	The nonlinear theorem gives stationary accumulation points and shrinking tubes, without asserting the stronger whole-sequence conclusion.
Generic smooth score/Jacobian comparison	A single-valued Lipschitz score permits a triangle inequality at the same parameter	Does not handle a sign jump at zero without a non-crossing assumption. The enlarged cusp interval in Theorem 8.1 preserves a feasible witness across such crossings.
Theorem 8.1 here	Two differentiable observation maps at one point; score regularity, common constraints/penalty and map/Jacobian error bounds	Transfers a stationarity upper bound at that same point, enlarging the response tube by e_h and adding $B_{q,h}$. No optimization sequence or KL assumption is needed.

The validation has distinct roles. The scalar benchmark supplies twelve pointwise exact-rational checks of the transfer inequality. The oscillator study tests mesh-dependent inflation, sensitivity control and cusp crossings numerically. The Khanty-Mansiysk share-relaxation ODE applies the same numerical verification to supplied January class tabulations. STAT5 illustrates why improved optimization need not improve chronological prediction. The separate polynomial January study shows that regularization explains most of its reported degree-five forecasting improvement.

Sections 2–7 develop the supporting loss calculus and optimization machinery. Section 8 gives the self-contained transfer theorem, witness evaluator and decision rule. Numerical evidence follows in Section 9; Section 10 summarizes polynomial forecasting. Detailed geometry, loss-shape calculations, maps, class-specific tables, sensitivity analyses and secondary timing records remain in the internal appendices. The affine consistency theorem is a conditional statistical reference; its fixed-scale, bounded-design and near-global-solution assumptions do not cover the implemented pilot-scale local procedures.

2. OBSERVATION MAPS AND NONSMOOTH GEOMETRY

The observation map takes a proposed parameter vector and returns the quantities that would be measured. An observation residual measures disagreement with data. A stationarity residual instead measures how closely the parameter satisfies a first-order optimization condition. These two residuals need not be small together. A good stationary fit can leave substantial data error, and small data errors alone do not prove stationarity.

The notation in Table 2 separates model dimension, loss exponent, polynomial degree and numerical tolerances. All observations and their parameter sensitivities use the same fixed standardization.

TABLE 2. Notation and units. All scales are frozen within a fit.

Symbol	Meaning
p, θ, Θ	Parameter dimension, parameter vector and nonempty closed convex feasible set. Compactness is invoked explicitly when needed.
g_i, J_i, r_i	Observation map, its row Jacobian and standardized residual $y_i - g_i(\theta)$.
a_i, W, c	Positive fixed weights, their sum and positive loss scale; experiments generally use $a_i = 1/n$.
F, I	Ordinary arctangent loss and its reciprocal GLDM gain representation. Minimize F and maximize I .
q, A_q, G_c	Gain exponent, normalization $\pi/(2q)$ and normalized reciprocal gain. $q = 1$ reproduces ordinary GLDM.
$J_{q,\lambda}, K_{q,\lambda}$	Penalized inverse-power loss and complementary gain. Here λ is a penalty coefficient.
w_q, Ξ_q, L_q	Positive absolute-residual derivative, even-loss subdifferential and scalar weak-convexity bound.
μ, μ_k	Positive proximal coefficient, constant or iteration-dependent, in every MM and PMM model; distinct from the statistical penalty coefficient λ .
$P, \mathcal{G}, d$	Criterion penalty, feasible inner primal-dual gap and parameter step. The local model $P_{\theta,\mu}$ has its full subscript.
$\delta, \mathcal{G}_\alpha$	Smoothing radius in standardized residual units and projected-gradient mapping.
$\tau, R_\tau, R_\tau^{q,\lambda}$	Response tube width, ordinary GLDM certificate and penalized inverse-power certificate; τ can be observation-specific.
$h, e_{i,h}, E_{i,h}, B_h, B_{q,h}$	ODE step size, state-map error, Jacobian error and ordinary/inverse-power certificate inflation.
H_q, f, IF_q	Population score curvature, error density and local location influence function. H_q is not A_q .

Consider an ODE and scalar observation functionals

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), \theta), & x(0) &= x_0(\theta), \\ g_i(\theta) &= h_i(x(t_i; \theta), \theta), & y_i &= g_i(\theta_\star) + e_i. \end{aligned} \quad (2.1)$$

Multiple state coordinates can be stacked as separate observations. Any fixed measurement standardization is absorbed into y_i and g_i . On a nonempty closed convex parameter set $\Theta \subset \mathbb{R}^p$, define

$$\begin{aligned} F(\theta) &= \sum_{i=1}^n a_i \rho(r_i(\theta)), & r_i(\theta) &= y_i - g_i(\theta), \\ \rho(r) &= \phi(|r|), & \phi(u) &= \arctan(u/c), & a_i > 0, & c > 0. \end{aligned} \quad (2.2)$$

The constants c and a_i are fixed during each fit. The experiments use $a_i = 1/n$. The original unnormalized regression objective is recovered with $a_i = 1$; its proximal parameter and certificate must then be rescaled consistently. Put $J_i(\theta) = \nabla g_i(\theta)^\top$, and stack the rows into $J(\theta)$. All vector norms without a subscript are Euclidean, all matrix norms are their induced operator norms, and dist_∞ uses the infinity norm. Write

$$N_\Theta(\theta) = \{v : \langle v, z - \theta \rangle \leq 0 \text{ for every } z \in \Theta\}. \quad (2.3)$$



TABLE 2. Notation and units (continued). All scales are frozen within a fit.

Symbol	Meaning
$\Gamma, \Gamma_q, \kappa_{\text{sh}}$	Observation-map curvature aggregates and the local affine sharpness constant; these are not derivative orders or polynomial degrees.
$\nu, n_{\mathcal{T}}$	Integer derivative order and number of training years.
ν_A, ν, κ_g	Nonnegative active-inequality multipliers, free equality multipliers, and a positive local quadratic-growth constant.
κ_d, ω	Oscillator damping ratio and frequency; κ_d is distinct from normalized calendar time.
M_s, D_s, K_s	Constant mechanical mass, viscous damping and stiffness; unrelated to polynomial degree or the GLDM gain.
$G_\theta, E_\theta, \gamma_\theta, \varrho_\theta$	Oscillator energy matrix, mass-normalized energy, squared energy-norm amplification and spectral amplification of RK4.
m, j, k	Polynomial degree, NDVI-class index and Bernstein coefficient index. m is not parameter dimension p or exponent q .
$\zeta(t), T, t_0$	Normalized calendar time, chosen final forecast year and initial year. $\zeta(t)$ is unrelated to tube width τ .
s_j, β_{jk}	Frozen class scale in percentage points and bounded Bernstein coefficients.
κ, ξ, η, β (January ODE)	Relaxation rate, equilibrium share and their fitted log-rate/logit coordinates; β here is not a Bernstein coefficient.
$\widehat{e}_{i,h}, \widehat{E}_{i,h}, \widehat{Q}_i$	Floating-point reference discrepancies and evaluated Jacobian-row norms in the January ODE study, distinguished from rigorously justified bounds.
$C(v), \Delta_{t,j}$	Compositional closure and paired absolute-error improvement. Neither is an independent statistical replication.

At an interior parameter point the normal cone is $\{0\}$; on a parameter boundary it contains the outward directions that account for the constraint. Thus constrained stationarity permits the data-fit slopes to be balanced by a boundary reaction. The Clarke subdifferential used below is the appropriate set of slopes at a corner. For the scalar loss it is a single derivative at a nonzero residual and the interval $[-1/c, 1/c]$ at zero. The interval is part of the loss definition; it is not an uncertainty interval for the parameter.

Definition 2.1 (Local regularity). On a convex neighborhood of the iterates, assume g_i is continuously differentiable and

$$\|J_i(\theta)\| \leq Q_i, \quad \|J_i(z) - J_i(\theta)\| \leq M_i \|z - \theta\|. \quad (2.4)$$

For the global subsequence results, Θ is compact and these bounds hold on a neighborhood of Θ . In an affine model $g_i(\theta) = x_i^T \theta$, take $Q_i = \|x_i\|$ and $M_i = 0$; boundedness of the iterates then suffices.

Lemma 2.2 (Scalar curvature and the cusp). For $u \geq 0$,

$$\phi'(u) = \frac{c}{c^2 + u^2}, \quad \phi''(u) = -\frac{2cu}{(c^2 + u^2)^2}, \quad L_\phi = \sup_{u \geq 0} |\phi''(u)| = \frac{9}{8\sqrt{3}c^2}. \quad (2.5)$$

The function ρ is $1/c$ -Lipschitz, Clarke regular, and L_ϕ -weakly convex. Its exact scalar Clarke subdifferential is

$$\Xi(r) = \partial_C \rho(r) = \begin{cases} \{c \operatorname{sgn}(r)/(c^2 + r^2)\}, & r \neq 0, \\ [-1/c, 1/c], & r = 0. \end{cases} \quad (2.6)$$

Proof. Differentiation gives the first two expressions. Maximizing $2cu/(c^2 + u^2)^2$ gives $u = c/\sqrt{3}$, and hence the stated constant. The one-sided derivatives at zero are $-1/c$ and $1/c$. Away from zero the second derivative of ρ is at least $-L_\phi$; at zero its derivative has an upward jump. Therefore the derivative of $\rho(r) + L_\phi r^2/2$ is monotone, so this function is convex. Subtracting a smooth quadratic from a finite convex function proves regularity and Eq. (2.6). The derivative bound also proves Lipschitz continuity. $\square$



Proposition 2.3 (Exact calculus and weak convexity). *Assume Definition 2.1 and a nonempty closed convex Θ .*

$$\partial_C F(\theta) = -J(\theta)^\top \{(a_i \xi_i)_{i=1}^n : \xi_i \in \Xi(r_i(\theta))\}. \quad (2.7)$$

The function is regular and Γ -weakly convex on the stated convex region, where

$$\Gamma = \sum_i a_i (L_\phi Q_i^2 + M_i/c). \quad (2.8)$$

Locally on Θ , the limiting subdifferential of $G = F + \iota_\Theta$ satisfies $\partial G = \partial_C F + N_\Theta$.

Proof. The regular scalar chain rule applies to each $\rho(y_i - g_i(\theta))$. Its directional derivative is the support function of $-J_i(\theta)^\top \Xi(r_i)$. The sum of these support functions is the support function of their Minkowski sum; this proves the exact, rather than inclusion-only, formula in Eq. (2.7) [14, 41]. For any $\xi_i \in \Xi(r_i(\theta))$, scalar weak convexity gives

$$\begin{aligned} \rho(r_i(z)) &\geq \rho(r_i(\theta)) + \xi_i(r_i(z) - r_i(\theta)) - \frac{1}{2} L_\phi |r_i(z) - r_i(\theta)|^2, \\ &\geq \rho(r_i(\theta)) - \xi_i J_i(\theta)(z - \theta) - \frac{1}{2} (M_i/c + L_\phi Q_i^2) \|z - \theta\|^2. \end{aligned} \quad (2.9)$$

Here $|\xi_i| \leq 1/c$, the Taylor remainder is at most $M_i \|z - \theta\|^2 / 2$, and $|r_i(z) - r_i(\theta)| \leq Q_i \|z - \theta\|$. Summing produces supporting hyperplanes for $F + \Gamma \|\cdot\|^2 / 2$, proving its convexity. Applying the convex subdifferential sum rule to this continuous convexification and ι_Θ , then subtracting the smooth quadratic, yields the final equality. $\square$

Figure 2 displays the bounded loss and the interval at its cusp. For large $|r|$, the arctangent score decays as $c/|r|^2$; it never becomes identically zero at a finite nonzero residual. Its behavior near zero remains LAD-like. Thus boundedness does not, by itself, distinguish GLDM from all existing redescending estimators; the specific computational advantage here is concavity in $|r|$ and the resulting exact weighted-LAD model.

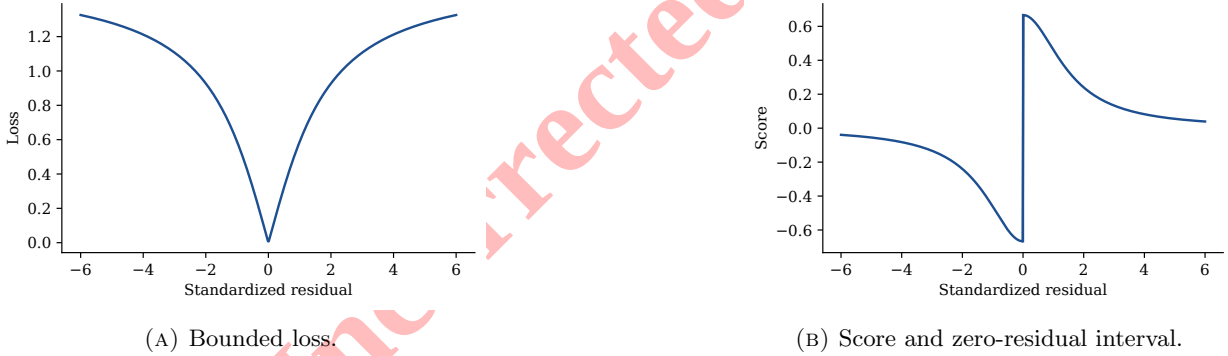


FIGURE 2. The unsmoothed arctangent loss for $c = 1.5$. The vertical segment is the subdifferential at zero, not a choice of a single score.

2.1. The location and sharpness of affine minima. Let $g_i(\theta) = x_i^\top \theta$, let X have rows $x_i^\top$, and put $A = \text{diag}(a_i)$. On a region where all residual signs are fixed and nonzero,

$$\nabla^2 F(\theta) = X^\top \text{diag} \left(-\frac{2a_i c |r_i|}{(c^2 + r_i^2)^2} \right) X, \quad -L_\phi X^\top A X \preceq \nabla^2 F \preceq 0. \quad (2.10)$$

Thus each smooth residual cell is concave. At a point $\bar{\theta}$, let $I_0 = \{i : r_i(\bar{\theta}) = 0\}$ and let X_0 collect its zero-residual rows.

Theorem 2.4 (Interpolation rank). *Suppose $\text{rank}(X) = p$. Every interior local minimum $\bar{\theta}$ of the affine objective satisfies $\text{rank}(X_0) = p$. In particular, at least p linearly independent observations are interpolated exactly in exact arithmetic.*



Proof. If $\text{rank}(X_0) < p$, choose a nonzero v in its nullspace. Along $\bar{\theta} + tv$, all zero residuals remain zero. For sufficiently small $|t|$, every other residual keeps its sign. Full column rank of X implies $x_i^\top v \neq 0$ for at least one nonzero-residual row. Eq. (2.10) then shows that $t \mapsto F(\bar{\theta} + tv)$ has strictly negative second derivative near zero. Strict concavity implies $F(\bar{\theta}) > [F(\bar{\theta} + tv) + F(\bar{\theta} - tv)]/2$ for small $t \neq 0$, contradicting local minimality. Both directions are feasible because the point is interior. $\square$

Proposition 2.5 (Strict dual interiority implies a sharp minimum). *In the unconstrained affine problem, suppose X_0 has full column rank and there is $\alpha^* \in \mathbb{R}^n$ with*

$$X^\top \alpha^* = 0, \quad \alpha_i^* = a_i \phi'(|r_i(\bar{\theta})|) \text{sgn } r_i(\bar{\theta}) \quad (i \notin I_0), \quad |\alpha_i^*| < a_i/c \quad (i \in I_0). \quad (2.11)$$

Then $F(\bar{\theta} + d) \geq F(\bar{\theta}) + \kappa_{\text{sh}} \|d\|/2$ for all sufficiently small d , where $\kappa_{\text{sh}} = \eta \sigma_{\min}(X_0) > 0$ and $\eta = \min_{i \in I_0} (a_i/c - |\alpha_i^*|)$.

Proof. The directional derivative is

$$F'(\bar{\theta}; d) = \sum_{i \in I_0} [(a_i/c) |x_i^\top d| + \alpha_i^* x_i^\top d] \geq \eta \|X_0 d\|_1 \geq \kappa_{\text{sh}} \|d\|. \quad (2.12)$$

The first equality uses $X^\top \alpha^* = 0$ to eliminate the nonzero-residual scores. Weak convexity implies $F(\bar{\theta} + d) \geq F(\bar{\theta}) + F'(\bar{\theta}; d) - \Gamma \|d\|^2/2$. Taking $\|d\| \leq \kappa_{\text{sh}}/\Gamma$ proves the claim, with no restriction needed if $\Gamma = 0$. $\square$

Example 2.6 (Stationarity can identify a local maximum). Take $p = 1$, $x_1 = x_2 = 1$, $y_1 = -2$, $y_2 = 2$, $a_1 = a_2 = 1/2$, and $c = 1$. At $\theta = 0$ the two scores cancel, but Eq. (2.10) is strictly negative. Moreover,

$$F(0) = \arctan 2 > \frac{1}{2} \arctan 4 = F(2) = F(-2). \quad (2.13)$$

Consequently, neither a zero Clarke residual nor the fixed-point condition proved below certifies global minimization.

3. RECIPROCAL GLDM REPRESENTATION AND INVERSE-POWER LOSSES

The reciprocal GLDM representation rewrites the original loss as a complementary gain. It is an identity, not a second estimator. Inverse-power GLDM raises the normalized gain to a power and changes the criterion when $q > 1$. The first construction checks implementation consistency; the second is a modelling choice. Both supply explicit score constants for the transfer theorem.

The reciprocal GLDM representation is defined as follows. It does not denote inversion of the ODE observation map or replacement of the response by its reciprocal. For the same fixed $c > 0$ and $a_i > 0$ as in Eq. (2.2), put

$$\begin{aligned} \psi_c(r) &= \begin{cases} \arctan(c/|r|), & r \neq 0, \\ \pi/2, & r = 0, \end{cases} & W &= \sum_{i=1}^n a_i, \\ I(\theta) &= \sum_{i=1}^n a_i \psi_c(r_i(\theta)), & \hat{\theta}_1 &\in \arg \max_{\theta \in \Theta} I(\theta). \end{aligned} \quad (3.1)$$

Large values of I indicate agreement with the observations, so the reciprocal GLDM representation is a *maximization* problem. Minimizing Eq. (3.1) would instead reward increasing residual magnitudes. The continuous extension at zero avoids a division-by-zero singularity. This definition is also different from $\sum_i a_i/\rho(r_i)$, which becomes unbounded at any interpolated observation.

Theorem 3.1 (Loss–gain equivalence). *For every $\theta \in \Theta$,*

$$F(\theta) + I(\theta) = \frac{\pi W}{2}. \quad (3.2)$$

Consequently, the global minimizers of F and global maximizers of I coincide, as do local minimizers and local maximizers, respectively. Whenever these extrema are attained,

$$F(\theta) - \min_{\Theta} F = \max_{\Theta} I - I(\theta). \quad (3.3)$$

The statement remains true with a common penalty $P(\theta)$ when the two criteria are $F + P$ and $I - P$.



Proof. For $u > 0$, differentiating $\arctan(u/c) + \arctan(c/u)$ gives zero, and evaluation at $u = c$ gives $\pi/2$. Continuity gives the same identity at $u = 0$. Multiplication by a_i and summation prove Eq. (3.2). Subtracting the identity at two parameter points reverses their objective ordering. This proves both the global and local assertions and, on taking extrema, Eq. (3.3). The penalties cancel in the same calculation. $\square$

This theorem concerns identical data, observation maps, weights, scales, parameter constraints, and optimization accuracy. Under these conditions reciprocal GLDM is a gain representation of the same estimator, not an additional statistical estimator with independently improved robustness. Different starts, tolerances, scales, or penalties can produce different numerical fits, but such differences cannot be attributed to the reciprocal identity itself. A residual-dependent re-estimation of weights or scale must be included in the objective before any corresponding equivalence is asserted.

A three-residual calculation. Take equal weights $a_i = 1/3$ and residual magnitudes $(0, c, 3c)$. Direct evaluation gives

$$F = \frac{\pi/4 + \arctan 3}{3}, \quad I = \frac{\pi/2 + \pi/4 + \arctan(1/3)}{3} = \frac{\pi}{2} - F. \quad (3.4)$$

The zero residual contributes nothing to the loss and the largest possible contribution to the gain. The residual c contributes $\pi/4$ to either unweighted scalar term. The residual $3c$ has a larger loss and a smaller gain. If another parameter lowers F by d , it raises I by exactly d ; an apparent difference in prediction cannot come from changing between these two formulas alone.

The corresponding weighted-LAD coefficients are

$$(w_1^k, w_2^k, w_3^k) = \frac{1}{3c}(1, 1/2, 1/10). \quad (3.5)$$

These weights are identical for GLDM and its reciprocal representation. The exact-zero observation retains the full slope interval, while the larger residual is downweighted. The weights multiply absolute residuals in the next subproblem, so substituting a weighted least-squares solve would implement a different update.

For a penalized criterion, the same interpretation requires opposite penalty signs: minimize $F + \lambda P$ or maximize $I - \lambda P$. Maximizing $I + \lambda P$ would reward a large penalty and would break the equivalence. When presenting objective values from separate runs, compare a decrease in loss with the corresponding increase in gain after matching this constant offset and penalty sign.

The complete reciprocal calculus, smoothing derivations, algorithmic equivalence and three-dimensional geometry are retained in Appendix A. The main construction now proceeds to the explicit score constants needed by Theorem 8.1.

3.1. A distinct inverse-power family. The reciprocal transformation alone cannot improve an estimator because Theorem 3.1 preserves every objective ordering. The inverse-power family changes the criterion. Bounded-similarity regression already has an established statistical literature [19, 20], and the generalized-correntropy family permits both smooth and nonsmooth residual similarities [13]. The construction below fixes the cusp slope while varying near-zero curvature, saturation and algebraic tail decay. Its role is to furnish explicit constants for the combined majorization, residual-tube and solver-transfer analysis; neither powering nor penalization is a new general optimization principle.

Definition 3.2 (Inverse-power GLDM). Use the exponent set $\mathcal{Q} = \{1\} \cup [2, \infty)$. For $u \geq 0$, $c > 0$ and $q \in \mathcal{Q}$, define

$$\begin{aligned} G_c(u) &= \frac{2}{\pi} \operatorname{atan2}(c, u), & A_q &= \frac{\pi}{2q}, \\ \ell_{q,c}(u) &= A_q \{1 - G_c(u)^q\}, & \Psi_{q,c}(r) &= A_q G_c(|r|)^q. \end{aligned} \quad (3.6)$$

For standardized residuals r_i and a nonnegative convex continuously differentiable penalty P , minimize, or equivalently maximize,

$$\begin{aligned} J_{q,\lambda}(\theta) &= \sum_i a_i \ell_{q,c}(|r_i(\theta)|) + \lambda P(\theta), \\ K_{q,\lambda}(\theta) &= \sum_i a_i \Psi_{q,c}(r_i(\theta)) - \lambda P(\theta). \end{aligned} \quad (3.7)$$



Here $\lambda \geq 0$ is fixed within a fit. The model degree m and gain exponent q are different quantities. The order study uses $m = 1, \dots, 5$ and fixes $q = 2$ for the modified estimator.

The normalization makes $\ell'_{q,c}(0+) = 1/c$ for every q . Thus changes in q do not merely rescale the slope at the origin. GLDM and its reciprocal representation both have $q = 1$, with $J_{1,0} = F$ and $K_{1,0} = I$. The distinct extension evaluated below, termed inverse-power GLDM, uses $q = 2$ and validation-selected λ . It remains a redescending M-estimator. The case $q = 2$ is the smallest nontrivial integer power: it preserves the slope $1/c$ at zero and changes quadratic score decay to cubic decay. Exponents $q = 3, 4$ are secondary sensitivity checks. Values $1 < q < 2$ are outside the proposed family and all convergence assertions in this paper; no claim is needed about them to justify the implemented estimators.

Theorem 3.3 (Shape, derivatives, and the distinction between estimators). *For $q \in \mathcal{Q}$, $\ell_{q,c}$ is increasing, strictly concave on $(0, \infty)$, and takes values in $[0, A_q)$. Its first two derivatives are*

$$w_q(u) := \ell'_{q,c}(u) = \frac{cG_c(u)^{q-1}}{c^2 + u^2}, \quad (3.8)$$

$$\ell''_{q,c}(u) = -\frac{2c}{(c^2 + u^2)^2} \left[uG_c(u)^{q-1} + \frac{(q-1)c}{\pi} G_c(u)^{q-2} \right]. \quad (3.9)$$

For $q = 1$ the second summand in brackets is zero. Moreover,

$$0 < w_q(u) \leq 1/c, \quad w_q(u) \sim c(2c/\pi)^{q-1} u^{-(q+1)} \quad (u \rightarrow \infty), \quad (3.10)$$

$$J_{q,\lambda} + K_{q,\lambda} = A_q W. \quad \ell_{2,c}(u) = \ell_{1,c}(u) - \ell_{1,c}(u)^2/\pi. \quad (3.11)$$

For fixed q the two objectives have identical extrema with reversed direction. For different q , the sample minimizers need not agree.

Proof. Since $G'_c(u) = -2c/\{\pi(c^2 + u^2)\}$, differentiating Eq. (3.6) gives Eqs. (3.8)–(3.9). Both terms in brackets are nonnegative and the first is positive for $u > 0$. The limits $G_c(0) = 1$, $G_c(\infty) = 0$ prove boundedness. Also $G_c(u) \sim 2c/(\pi u)$, proving the tail expression. The complement identity follows term by term, with cancellation of the penalty. Substituting $G_c = 1 - 2\ell_{1,c}/\pi$ proves the quadratic relation. To show that the estimators can differ, take two feasible residual vectors whose ordinary losses are $(0, 1.4)$ and $(0.6, 0.6)$; these correspond to residual magnitudes $c \tan(0)$, $c \tan(1.4)$ and $c \tan(0.6)$. Their ordinary loss sums are 1.4 and 1.2, respectively. Their $q = 2$ loss sums are $1.4 - 1.96/\pi$ and $1.2 - 0.72/\pi$. The latter ordering is reversed, because $0.2 - 1.24/\pi < 0$. $\square$

For $q = 2$ the tail score decays as u^{-3} instead of u^{-2} . This is a precise reduction of tail sensitivity, not a theorem of lower prediction error. Downweighting a large but genuine observation can increase bias. Increasing q also lowers the saturation level and can create different local minima. Since the loss is bounded, $\exp(-\ell_{q,c}(|r|))$ does not integrate to a finite constant on the real line and is not by itself a probability density.

Figure 3 illustrates the different saturation levels and score decay. The change in scalar shape does not order forecast risks.

For direct comparison, ignoring positive normalizations, Ceres uses $\rho_{\text{sq}}(r) = \arctan(r^2)$, with $\rho'_{\text{sq}}(0) = 0$ and $\rho'_{\text{sq}}(r) = 2r/(1+r^4)$ [12]. GLDM instead has one-sided derivatives $\pm 1/c$ at zero. The arctangent-score component of ArcTanLoss is $ar \arctan(kr) - a \log(1 + k^2 r^2)/(2k)$, with derivative $a \arctan(kr)$; the complete loss also includes absolute-error and squared-prediction terms [52]. The distinction between a bounded loss and a bounded non-redescending score is therefore substantive. For the present family, Eq. (3.9) gives

$$\ell'_{q,c}(0+) = \frac{1}{c}, \quad \ell''_{q,c}(0+) = -\frac{2(q-1)}{\pi c^2}, \quad w_q(u) \asymp u^{-(q+1)} \quad (u \rightarrow \infty). \quad (3.12)$$

The first identity prevents equality with a positive rescaling of a smooth even squared-residual loss. For a fixed cusp slope, the second separates different q values. Algebraic rather than exponential score decay separates this family from Laplace correntropy. These identities distinguish functions; they do not establish historical priority or universal predictive superiority.



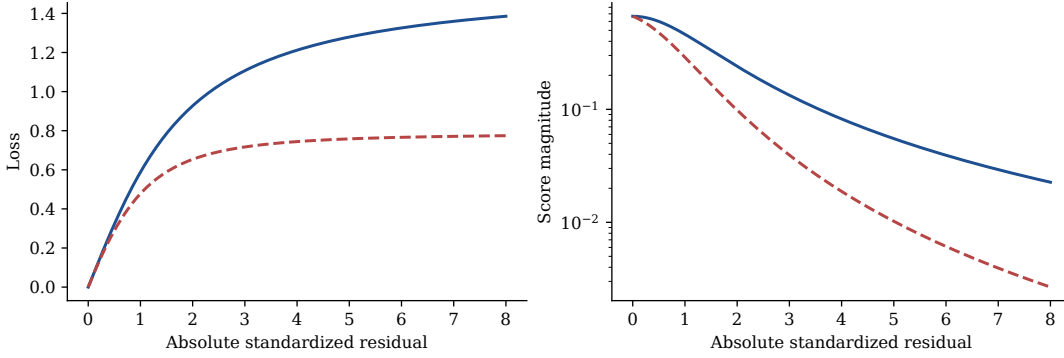


FIGURE 3. Ordinary arctangent loss ($q = 1$, blue solid) and the inverse-power loss ($q = 2$, red dashed), at $c = 1.5$. The left panel shows loss and the right panel shows score magnitude on a logarithmic scale. The normalization preserves the cusp slope while allowing near-zero curvature, saturation level and tail decay to change. The comparison does not isolate tail behavior.

TABLE 3. Scalar comparison with established redescending and bounded-similarity criteria. Tail expressions omit positive constants; $u = |r|$ and $b > 0$ is the comparison bandwidth. Smoothness concerns the even loss at zero. The majorizers shown use the stated residual variable.

Loss	Bounded	Smooth	Tail score	Tangent model
GLDM	Yes	No	u^{-2}	Weighted LAD in u
Inverse-power, $q \geq 2$	Yes	No	$u^{-(q+1)}$	Weighted LAD in u
Gaussian correntropy [20]	Yes	Yes	$ue^{-u^2/(2b^2)}$	Weighted squares in u^2
Laplace correntropy [13]	Yes	No	$e^{-u/b}$	Weighted LAD in u
Cauchy	No	Yes	u^{-1}	Weighted squares in u^2
Tukey bisquare	Yes	Yes	Zero beyond cutoff	Weighted squares in u^2
Huber	No	Yes	Constant	Convex/IRLS

Generalized correntropy has normalized loss $1 - \exp[-(u/b)^a]$, $a > 0$ [13]. At $a = 1$ it is concave in u , has an even-loss cusp and admits weighted-LAD majorization; at $a = 2$ it is smooth at zero and concave in u^2 , giving weighted squares. These structures are therefore shared. Arctangent-specific features are its algebraic score tail, slope $1/c$ and constants L_q . Section 9.5 compares the Laplace member under matched starts without inferring dominance. Weights, scales, starts and penalties also matter; bounded residual scores do not control unbounded predictors.

Proposition 3.4 (Existence and exact nonsmooth calculus). *Suppose Θ is nonempty, compact and convex, $q \in \mathcal{Q}$, and the observation maps and P are continuous. Then Eq. (3.7) attains its extrema. For $q = 1$ or $q \geq 2$, define*

$$L_q = \frac{9}{8\sqrt{3}c^2} + \frac{2(q-1)}{\pi c^2}, \quad \Xi_q(r) = \begin{cases} \{\text{sgn}(r)w_q(|r|)\}, & r \neq 0, \\ [-1/c, 1/c], & r = 0. \end{cases} \quad (3.13)$$

The even loss is regular and L_q -weakly convex. For the following calculus and local-optimality assertions, also assume Definition 2.1 and that P is continuously differentiable on a neighborhood of Θ :

$$\partial_C J_{q,\lambda} = -J(\theta)^\top \{(a_i \xi_i)_i : \xi_i \in \Xi_q(r_i)\} + \lambda \nabla P(\theta). \quad (3.14)$$

A constrained local minimizer satisfies $0 \in \partial_C J_{q,\lambda} + N_\Theta$; this is necessary and generally not sufficient.

Proof. Existence follows from continuity on a compact set. For $q \geq 2$, the powers G_c^{q-1} and G_c^{q-2} are at most one. Eq. (3.9) then bounds the two terms by $9/(8\sqrt{3}c^2)$ and $2(q-1)/(\pi c^2)$. The case $q = 1$ is Lemma 2.2. At zero the

one-sided derivatives have the upward jump from $-1/c$ to $1/c$. Addition of $L_q r^2/2$ therefore gives a convex function. Regularity and the exact chain and sum rules follow as in Proposition 2.3, now with L_q and the smooth penalty. The constrained Fermat rule gives the last inclusion. $\square$

Compact constraints are substantive. A difference penalty vanishes along constant coefficient vectors, so it does not alone make a bounded-loss objective coercive. The numerical model below imposes a box. A small stationarity residual does not certify global optimality.

Theorem 3.5 (Affine inverse-power minorization). *Let $q \in \mathcal{Q}$, g be affine, P convex and continuously differentiable, and Θ nonempty, compact and convex. Fix $\lambda \geq 0$, assume P is defined on a neighborhood of Θ , and start at $\theta^0 \in \Theta$. With $0 < \mu_{\min} \leq \mu_k \leq \mu_{\max} < \infty$, define*

$$\theta^{k+1} = \arg \min_{z \in \Theta} \left\{ \sum_i a_i w_q(|r_i^k|) |r_i(z)| + \lambda P(z) + \frac{\mu_k}{2} \|z - \theta^k\|^2 \right\}. \quad (3.15)$$

The subproblem has a unique minimizer and

$$J_{q,\lambda}(\theta^k) - J_{q,\lambda}(\theta^{k+1}) = K_{q,\lambda}(\theta^{k+1}) - K_{q,\lambda}(\theta^k) \geq \mu_k \|\theta^{k+1} - \theta^k\|^2. \quad (3.16)$$

The steps are square summable and every accumulation point is constrained Clarke stationary for $J_{q,\lambda}$.

Proof. Concavity gives $\ell(v) \leq \ell(u) + w_q(u)(v - u)$ for $u, v \geq 0$. Adding the penalty and proximal term yields a tight majorizer. The subproblem is μ_k -strongly convex, so its objective drop from θ^k to its minimizer is at least $\mu_k \|\theta^{k+1} - \theta^k\|^2 / 2$. Removing the proximal term provides the other half and proves Eq. (3.16). Summation and a lower bound on $J_{q,\lambda}$ give square summability. The subproblem optimality condition contains vectors

$$a_i w_q(|r_i^k|) \sigma_i^{k+1}, \quad \sigma_i^{k+1} \in \text{Sign}(r_i(\theta^{k+1})). \quad (3.17)$$

They lie in a fixed compact interval. Along a convergent subsequence, the vanishing steps give the same limit for θ^k and θ^{k+1} . Continuity of w_q , closedness of the sign graph, continuity of ∇P , and the closed normal-cone graph pass the optimality condition to Eq. (3.14). The normals are bounded because all other terms in that condition are bounded. $\square$

The nonlinear extension is stated in Theorem 5.3, its inexact version in Corollary 6.4, and its discretization transfer in Theorem 8.1. These results retain the convex penalty exactly and use L_q for the scalar score bound. They do not extend the affine interpolation-rank or finite-termination conclusions to arbitrary nonlinear maps or positive penalties.

Proposition 3.6 (Smoothing, stationarity, and a computable residual). *Assume Definition 2.1, a nonempty closed convex Θ , a penalty continuously differentiable on a neighborhood of Θ , fixed $\lambda \geq 0$, and $q \in \mathcal{Q}$. Let $v_\delta(r) = \sqrt{r^2 + \delta^2}$ and replace $|r|$ by $v_\delta(r)$ in $J_{q,\lambda}$, keeping the penalty fixed, to obtain $J_{q,\lambda}^\delta$. Then*

$$0 \leq J_{q,\lambda}^\delta - J_{q,\lambda} \leq W\delta/c, \quad (3.18)$$

$$\xi_{q,\delta}(r) = w_q(v_\delta(r)) r / v_\delta(r), \quad |\xi'_{q,\delta}(r)| \leq L_q + 1/(c\delta). \quad (3.19)$$

If the unsmoothed infimum is finite, an ϵ -optimal global solution of the smooth problem has objective gap above $\inf_\Theta J_{q,\lambda}$ at most $\epsilon + W\delta/c$. If $\delta_k \downarrow 0$, $\theta_k \rightarrow \bar{\theta}$, and

$$\text{dist}\left(0, \nabla J_{q,\lambda}^{\delta_k}(\theta_k) + N_\Theta(\theta_k)\right) \rightarrow 0, \quad (3.20)$$

then $\bar{\theta}$ is constrained Clarke stationary for $J_{q,\lambda}$.

Proof. Use $0 \leq v_\delta - |r| \leq \delta$ and $0 < w_q \leq 1/c$ to prove the uniform bound. Differentiating the smooth score gives

$$\xi'_{q,\delta}(r) = \ell''_{q,c}(v_\delta)(r/v_\delta)^2 + w_q(v_\delta)\delta^2/v_\delta^3. \quad (3.21)$$

The stated bound follows. For any $\eta > 0$, choose $z_\eta \in \Theta$ with $J_{q,\lambda}(z_\eta) \leq \inf_\Theta J_{q,\lambda} + \eta$. Comparing the smooth approximate solution with z_η gives a gap at most $\epsilon + W\delta/c + \eta$; letting $\eta \downarrow 0$ proves the assertion without assuming



attainment on a noncompact set. For Eq. (3.20), select approximate stationarity normals. Bounded scores, bounded Jacobians, and bounded penalty gradient give bounded normals. At nonzero limiting residuals the scores converge to their singletons. At zero their subsequential limits lie in $[-1/c, 1/c]$. The closed normal graph and Eq. (3.14) finish the proof. $\square$

The numerical continuation ends at $\delta = 10^{-6}$ in standardized residual units. With $W = 1$ and $c = 1.5$, its objective perturbation is at most 6.67×10^{-7} . This is an objective bound, not a parameter-error bound or a certificate that a local optimizer found a global solution.

Proposition 3.7 (Projected-gradient descent for a fixed smoothing level). *Let Θ be nonempty, compact and convex, let $\delta > 0$, and let $\theta^0 \in \Theta$. Suppose $J_{q,\lambda}^\delta$ is continuously differentiable on a neighborhood of Θ and its gradient is L -Lipschitz with $L > 0$. For $0 < \alpha \leq 1/L$, put*

$$\theta^{k+1} = \text{proj}_\Theta(\theta^k - \alpha \nabla J_{q,\lambda}^\delta(\theta^k)), \quad \mathcal{G}_\alpha(\theta^k) = (\theta^k - \theta^{k+1})/\alpha. \quad (3.22)$$

Then $J^\delta(\theta^k) - J^\delta(\theta^{k+1}) \geq \alpha \|\mathcal{G}_\alpha(\theta^k)\|^2 / 2$ and

$$\min_{0 \leq k < K} \|\mathcal{G}_\alpha(\theta^k)\|^2 \leq \frac{2\{J^\delta(\theta^0) - \inf_\Theta J^\delta\}}{\alpha K}. \quad (3.23)$$

Proof. Projection optimality gives $\langle \nabla J^\delta(\theta^k), d_k \rangle \leq -\|d_k\|^2 / \alpha$. The descent lemma then gives a decrease of at least $(1/\alpha - L/2) \|d_k\|^2 \geq \|d_k\|^2 / (2\alpha)$. Sum over k and use the smallest squared mapping norm to obtain Eq. (3.23). $\square$

This proposition describes a guaranteed first-order option. The reported implementation uses multistart L-BFGS-B and records its projected-gradient residual independently; a solver success flag is not substituted for Eq. (3.20).

Proposition 3.8 (A backward-data certificate for the inverse-power fit). *For $q \in \mathcal{Q}$, fixed $\lambda \geq 0$, an affine predictor and a penalty P continuously differentiable near a nonempty closed convex Θ , fix $\theta \in \Theta$ and $\tau \geq 0$. Replace every residual with $|r_i| \leq \tau$ by zero and retain all other residuals. Let $\tilde{y}$ be the corresponding standardized response vector. Choose $\alpha_i \in a_i[-1/c, 1/c]$ for the zeroed residuals, and $\alpha_i = a_i \text{sgn}(r_i) w_q(|r_i|)$ otherwise. For any $v \in N_\Theta(\theta)$,*

$$\|\tilde{y} - y\|_\infty \leq \tau, \quad \text{dist}_\infty(0, \partial_C J_{q,\lambda}(\theta; \tilde{y}) + N_\Theta(\theta)) \leq \|-X^\top \alpha + \lambda \nabla P(\theta) + v\|_\infty. \quad (3.24)$$

For a box, minimizing the right-hand side over α, v is a linear program.

Proof. Set $\tilde{y}_i = g_i(\theta)$ at the selected indices and leave the other responses unchanged. The perturbation bound follows immediately. The chosen scores belong to the exact scalar subdifferentials for these modified data. Equation (3.14) therefore places the vector on the right inside the stationary set on the left. Score intervals, box normal-cone sign conditions at a fixed point, and the epigraph of the infinity norm are linear constraints. $\square$

In physical units the response perturbation for class j is at most $s_j \tau$. The certificate concerns nearby data and local stationarity, not a global optimum or sampling uncertainty. Floating-point LP witnesses are projected onto their scalar bounds before their stationarity residual is evaluated.

3.2. Conditional statistical interpretation.

Proposition 3.9 (Location influence under a curvature condition). *Let $Y = \theta_0 + \varepsilon$, with $q \in \mathcal{Q}$, and consider an interior unpenalized population location solution. Sufficient density conditions are: f is symmetric and continuously differentiable, f is bounded and tends to zero at infinity, and for some $\eta > 0$ there is an integrable D with $|f'(u+h)| \leq D(u)$ for every $|h| \leq \eta$. Assume*

$$H_q = \frac{2f(0)}{c} + \int_{\mathbb{R}} \ell''_{q,c}(|u|) f(u) du > 0. \quad (3.25)$$

For the local differentiable solution branch under point contamination at z ,

$$\text{IF}_q(z) = \frac{\text{sgn}(z - \theta_0) w_q(|z - \theta_0|)}{H_q}, \quad |\text{IF}_q(z)| \leq \frac{1}{cH_q}. \quad (3.26)$$



Away from the zero-residual convention this influence redescends at order $|z|^{-(q+1)}$.

Proof. The score is bounded by $1/c$. Writing its population expectation as an integral against $f(u + \theta - \theta_0)$ permits differentiation by the stated integrable domination. Integration by parts on the two half-lines is justified by bounded scores and $f(u) \rightarrow 0$; the boundary contributions at infinity vanish. The score $\text{sgn}(u)w_q(|u|)$ has an upward jump $2/c$ at zero and derivative $\ell''_{q,c}(|u|)$ elsewhere. Its distributional derivative therefore contributes both the integral and the density term in Eq. (3.25). For $z \neq \theta_0$, the point-contamination score is smooth in a neighborhood of θ_0 , and $H_q > 0$ permits the local implicit-function argument. Differentiating the contaminated score equation at contamination fraction zero gives $-H_q \text{IF}_q(z) + \text{sgn}(z - \theta_0)w_q(|z - \theta_0|) = 0$. The bound and decay follow from Theorem 3.3. At $z = \theta_0$, symmetry admits the zero directional shift. $\square$

The positive-curvature and local-branch assumptions matter. The formula is not a finite-sample breakdown result. In regression, the score also multiplies the design vector, so unbounded leverage is not controlled by bounded residual scores alone. Moreover, changing q changes H_q ; faster tail decay alone does not order the full influence functions or their clean-data variances.

TABLE 4. Gaussian location influence at $c = 1.5$. The last column is the squared L^2 norm of the influence function, not an empirical January variance.

q	H_q	$\sup \text{IF}_q $	$E(\text{IF}_q^2)$
1	0.329208	2.0251	2.6304
2	0.211496	3.1522	4.1720
3	0.143437	4.6478	6.5715

For a standard Gaussian error, integration by parts gives

$$H_q = 2 \int_0^\infty u w_q(u) \varphi(u) du, \quad E(\text{IF}_q^2) = \frac{2}{H_q^2} \int_0^\infty w_q(u)^2 \varphi(u) du. \quad (3.27)$$

Table 4 shows that $\sup_z |\text{IF}_q(z)|$ increases from 2.0251 at $q = 1$ to 3.1522 at $q = 2$ and 4.6478 at $q = 3$, although the tail decays faster. The curvature denominator H_q decreases enough to produce this increase. Faster redescending therefore does not imply smaller maximum influence; the squared L^2 norm also increases in this Gaussian example. These are properties of the local functional at a specified density. They do not assert asymptotic normality for the January series. For the scalar sample $(0, 0.3, 6.3, 7.1, 9.2)$ and $c = 1.5$, global location minimizers can be found among the five observations: between consecutive observations the objective is concave, and outside their range it improves when moving toward them. The $q = 1$ minimizer is 7.1 and the $q = 2$ minimizer is 0.3. This is a location-estimation example of different minima, beyond the two-vector ordering example in Theorem 3.3.

Theorem 3.10 (Conditional consistency for affine regression). *Let (X_i, Y_i) be stationary and ergodic, with $\|X_i\| \leq B$ almost surely. Let Θ be nonempty and compact, $s > 0$ fixed, and P continuous. Use residuals $(Y_i - X_i^\top \theta)/s$, $a_i = 1/n$, fixed $q \in \mathcal{Q}$, $c > 0$, and $0 \leq \lambda_n \rightarrow \lambda$. Suppose the population criterion*

$$\mathcal{J}(\theta) = E \ell_{q,c}(|Y - X^\top \theta|/s) + \lambda P(\theta) \quad (3.28)$$

has a unique minimizer $\theta^\dagger$. If empirical solutions have global objective errors tending to zero almost surely, then they converge almost surely to $\theta^\dagger$.

Proof. The bounded scalar loss is integrable and is $B/(sc)$ -Lipschitz in θ . Cover the compact set by a finite η -net. The ergodic theorem makes the empirical-minus-population difference vanish at every net point. The supremum difference is bounded by this finite maximum plus $2B\eta/(sc)$. Applying the argument along a sequence $\eta \downarrow 0$ proves a uniform law almost surely. The continuous penalty is bounded on Θ , so $(\lambda_n - \lambda)P$ vanishes uniformly. For each neighborhood of the unique minimizer, continuity and compactness give a positive population objective gap on its complement. Uniform convergence and the vanishing optimization error eventually exclude that complement. $\square$



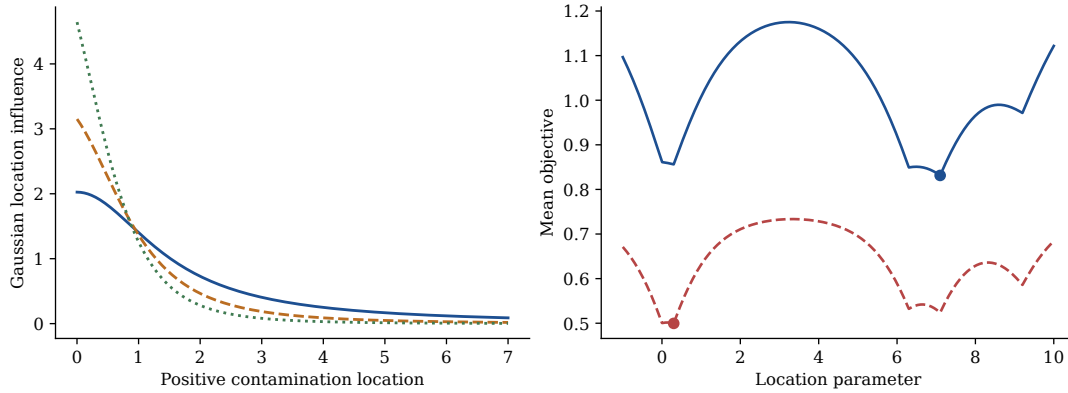


FIGURE 4. Left: Gaussian location influence for $q = 1, 2, 3$ (blue solid, orange dashed, green dotted). Right: scalar sample objectives for $q = 1$ (blue) and $q = 2$ (red dashed), with their global minima marked.

The theorem provides a conditional statistical reference result rather than a consistency guarantee for the implemented numerical procedure. Section 9 uses unbounded Gaussian predictors, a LAD-pilot scale and local multistart fits without a vanishing global optimization-error bound. Freezing the pilot scale within a fit does not make it deterministic asymptotically. Extension would require a suitable random Lipschitz envelope, control of the pilot scale, identification of the population target and an adequate empirical-solution condition. For nonlinear ODEs, deterministic stationarity and transfer results do not identify a physical parameter. Table 9 separates the statistical, optimization and verification scopes.

TABLE 5. Assumptions of the statistical statements are separate from optimization assumptions.

Result	Sampling/target assumptions	Optimization requirement
Proposition 3.9	Symmetric C^1 density with the stated domination; positive H_q ; local location branch	Interior unpenalized population solution
Theorem 3.10	Affine regression; stationary ergodic sample, bounded predictors, fixed scale, compact set, unique population minimizer	Global empirical objective error tends to zero
Nonlinear ODE results	Deterministic map regularity; no statistical consistency or physical identification theorem	Local optimization and solver-transfer guarantees only
January forecasts	One annual record; no asserted stationary sampling law or validated physical target	Finite-dimensional numerical fits only

3.3. Role of the inverse-power option. The exponent is a loss-shape choice within the certificate framework. At the same residual magnitude c , its weighted-LAD coefficient is 2^{1-q} times the ordinary-GLDM coefficient. Increasing q also lowers the saturation level and changes curvature near zero. These effects cannot be interpreted as a change of tail behavior alone.

The practical reference is $q = 1$, with its equivalent reciprocal GLDM representation. Use a different exponent only under a declared scale, penalty convention and training-only selection rule. The controlled forecasts do not support a general preference for $q > 1$. Section B gives the detailed scalar comparisons and their proofs. They support interpretation of the constants in Theorem 8.1, rather than replacing certificate transfer as the main result.



4. AFFINE PROXIMAL MAJORIZATION AND FINITE TERMINATION

Majorization–minimization replaces a difficult objective by an easier function that touches it at the current parameter and lies above it. Here the concavity of the loss in residual magnitude produces a weighted sum of absolute residuals. The quadratic proximal term discourages an excessively large parameter step and makes the subproblem uniquely solvable. It is an algorithmic term, controlled by μ , rather than the statistical penalty λP in the fitted criterion. An exact update minimizes this convex subproblem; it does not globally minimize the original nonconvex objective.

For $\theta \in \Theta$, set $w_i(\theta) = a_i \phi'(|r_i(\theta)|)$ and define the proximal weighted-LAD map

$$T_\mu(\theta) = \arg \min_{z \in \Theta} \left\{ \sum_i w_i(\theta) |y_i - x_i^\top z| + \frac{\mu}{2} \|z - \theta\|^2 \right\}, \quad \mu > 0. \quad (4.1)$$

The objective is strongly convex and coercive on Θ , so the update exists and is unique, even when X is rank deficient. The weights multiply absolute residuals, not squared residuals; this is the essential distinction from an iteratively reweighted least-squares implementation.

Theorem 4.1 (Descent, fixed points, and a relative error). *Let Θ be nonempty, closed and convex, let $g_i(\theta) = x_i^\top \theta$, and let $\theta \in \Theta$ and $\mu > 0$. With fixed $a_i, c > 0$, put $\theta^+ = T_\mu(\theta)$ and $d = \theta^+ - \theta$. Then*

$$F(\theta) - F(\theta^+) \geq \mu \|d\|^2, \quad (4.2)$$

$$T_\mu(\theta) = \theta \iff 0 \in \partial_C F(\theta) + N_\Theta(\theta). \quad (4.3)$$

If $\mu \leq \mu_+$, there exists $v^+ \in \partial_C F(\theta^+) + N_\Theta(\theta^+)$ such that

$$\|v^+\| \leq C_A \|d\|, \quad C_A = \mu_+ + L_\phi \|A^{1/2} X\|^2. \quad (4.4)$$

Proof. Concavity of ϕ gives $\phi(v) \leq \phi(u) + \phi'(u)(v - u)$. If $Q_\theta(z)$ denotes the objective in Eq. (4.1), this implies

$$F(z) \leq F(\theta) + Q_\theta(z) - Q_\theta(\theta) - \frac{\mu}{2} \|z - \theta\|^2. \quad (4.5)$$

Strong convexity and constrained optimality give $Q_\theta(\theta) - Q_\theta(\theta^+) \geq \mu \|d\|^2/2$, proving Eq. (4.2). At $z = \theta$, the convex optimality inclusion for Q_θ is exactly Eq. (2.7) plus the normal cone; uniqueness proves Eq. (4.3).

For the last claim, choose $s_i \in \text{Sign}(r_i(\theta^+))$ and $n^+ \in N_\Theta(\theta^+)$ satisfying

$$-X^\top (w_i(\theta) s_i)_i + \mu d + n^+ = 0. \quad (4.6)$$

Replacing $w_i(\theta)$ by $w_i(\theta^+)$ produces an admissible vector v^+ for the original objective. Since $|s_i| \leq 1$ and $|\phi'(|r_i(\theta^+)|) - \phi'(|r_i(\theta)|)| \leq L_\phi |x_i^\top d|$, the change has norm at most $L_\phi \|A^{1/2} X\|^2 \|d\|$. Eq. (4.6) then gives Eq. (4.4). $\square$

Theorem 4.2 (Affine convergence and complexity). *Under the affine and closed-convex-set assumptions of Theorem 4.1, suppose the exact iterates start at $\theta^0 \in \Theta$, are bounded, and satisfy $0 < \mu_- \leq \mu_k \leq \mu_+ < \infty$. Every accumulation point is constrained Clarke stationary. If $F_* = \lim_k F(\theta^k)$, then*

$$\min_{0 \leq k < K} \text{dist}_\infty(0, \partial_C F(\theta^{k+1}) + N_\Theta(\theta^{k+1})) \leq C_A \sqrt{\frac{F(\theta^0) - F_*}{\mu_- K}}. \quad (4.7)$$

For polyhedral Θ , the whole bounded sequence converges and has finite length.

Proof. Summing Eq. (4.2) yields $\sum_k \|\theta^{k+1} - \theta^k\|^2 \leq (F(\theta^0) - F_*)/\mu_-$. Hence the steps vanish and Eq. (4.7) follows from Eq. (4.4). For any convergent subsequence, the scalar subgradients in Eq. (4.6) are bounded. Its normals are bounded as well, because they equal a bounded score vector minus $\mu_k d_k$. Passing to subsequences and using the separate closed graphs of the scalar subdifferential and the convex normal cone proves stationarity. No closed-graph assertion for an arbitrary sum of set-valued maps is needed.

For completeness, on a bounded region the restriction of arctangent is real analytic and definable in the restricted-analytic structure. Absolute values, affine maps, and polyhedral indicators preserve definability. Thus $G = F + \iota_\Theta$



has the Kurdyka–Łojasiewicz (KL) property near the compact accumulation set [4, 9]. Proposition 2.3 supplies the required limiting-subdifferential equality. Uniformization yields a concave desingularizing function φ with $\varphi'(F(\theta^k) - F_*) \text{dist}(0, \partial G(\theta^k)) \geq 1$ for all sufficiently large nonstationary iterates. Setting $d_k = \theta^{k+1} - \theta^k$ gives

$$2 \|d_k\| \leq \|d_{k-1}\| + \frac{C_A}{\mu_-} [\varphi(F(\theta^k) - F_*) - \varphi(F(\theta^{k+1}) - F_*)]. \quad (4.8)$$

This follows from concavity, sufficient decrease, relative error, and $2\sqrt{ab} \leq a + b$. Summation proves finite length and hence convergence. If an objective gap is zero, sufficient decrease makes all subsequent steps zero. $\square$

Theorem 4.3 (Finite termination near a nondegenerate interpolating minimum). *Under the unconstrained affine assumptions of Proposition 2.5, and for fixed $0 < \mu_- \leq \mu_+ < \infty$, there is a neighborhood U of $\bar{\theta}$ such that $T_\mu(\theta) = \bar{\theta}$ for every $\theta \in U$ and every $\mu \in [\mu_-, \mu_+]$. Consequently, an exact affine PMM sequence converging to $\bar{\theta}$ terminates after finitely many iterations.*

Proof. For $i \notin I_0$, prescribe $\alpha_i(\theta) = w_i(\theta) \text{sgn } r_i(\bar{\theta})$, which tends to α_i^* . At $z = \bar{\theta}$, subproblem optimality requires

$$X_0^\top \alpha_0(\theta) = \mu(\bar{\theta} - \theta) - X_1^\top \alpha_1(\theta). \quad (4.9)$$

The right side tends uniformly in μ to $X_0^\top \alpha_0^*$. Full column rank of X_0 gives a bounded right inverse of $X_0^\top$, so Eq. (4.9) has a solution tending uniformly to α_0^* . Strict interiority in Eq. (2.11), together with $w_i(\theta) \rightarrow a_i/c$ on I_0 , ensures $|\alpha_i(\theta)| \leq w_i(\theta)$. These scores satisfy the complete convex optimality condition for the subproblem at $\bar{\theta}$. Uniqueness then proves the assertion. Once a convergent sequence enters U , one more exact update reaches its limit. $\square$

The strict-interiority and rank hypotheses are substantive. They are not implied by stationarity, and fixed inner tolerances only support approximate termination. Theorem 4.3 explains why accurate affine weighted-LAD updates may settle rapidly on an interpolating set without asserting that every run does so.

Algorithm 1 Affine proximal MM with an explicit accuracy record.

Require: X, y, Θ , a feasible θ^0 , fixed a_i, c , proximal bounds, and tolerances.

- 1: **for** $k = 0, 1, \dots$ **do**
 - 2: Compute $r^k = y - X\theta^k$ and $w_i^k = a_i c / (c^2 + (r_i^k)^2)$.
 - 3: Solve Eq. (4.1); record a feasible primal–dual gap using Algorithm 3.
 - 4: Set θ^{k+1} to the candidate and record the actual objective difference and step norm.
 - 5: Evaluate the residual-tube LP in Eq. (7.1); record its tube width separately.
 - 6: **if** the prescribed step and inner-gap tolerances are satisfied **then**
 - 7: **return** θ^{k+1} , its objective, gap, tube width, and LP residual.
 - 8: **end if**
 - 9: **end for**
-

5. NONLINEAR PROXIMAL LINEARIZATION

For an affine observation map, its first-order expansion is exact. For a nonlinear ODE map, the algorithm uses a first-order approximation only inside the current subproblem, then evaluates the true numerical criterion before accepting the step. Backtracking increases the proximal coefficient until this check passes. The extra proof terms below measure the mismatch between actual and linearized residuals and between the old and new sensitivities. This is the reason nonlinear convergence needs a response tube as well as a small parameter step.

For a nonlinear predictor, define $r = r(\theta)$, $J = J(\theta)$, $D_\theta = \Theta - \theta$, and

$$P_{\theta, \mu}(d) = \sum_i w_i(\theta) |r_i - J_i d| + \frac{\mu}{2} \|d\|^2, \quad d \in D_\theta. \quad (5.1)$$



The model remains a strongly convex weighted-LAD problem. It is not the original nonlinear objective, and its minimizer need not interpolate any actual observation. For a fixed $\sigma > 0$, accept $z = \theta + d$ when

$$F(z) \leq F(\theta) + \sum_i w_i(\theta)(|r_i - J_i d| - |r_i|) + \frac{\mu - \sigma}{2} \|d\|^2. \quad (5.2)$$

Lemma 5.1 (Finite backtracking and nonlinear decrease). *Under Definition 2.1, Eq. (5.2) holds for every feasible d whenever*

$$\mu \geq \sigma + \sum_i w_i(\theta) M_i \leq \sigma + \sum_i a_i M_i / c. \quad (5.3)$$

If d minimizes Eq. (5.1) exactly, an accepted step satisfies

$$F(\theta) - F(\theta + d) \geq \frac{\mu + \sigma}{2} \|d\|^2. \quad (5.4)$$

Proof. Taylor's theorem gives $|r_i(\theta + d) - (r_i - J_i d)| \leq M_i \|d\|^2 / 2$. Hence $|r_i(\theta + d)| \leq |r_i - J_i d| + M_i \|d\|^2 / 2$. Combining this inequality with monotonicity and the tangent inequality for ϕ proves Eq. (5.2) at the threshold. Doubling μ therefore terminates in finitely many trials. For an exact step, strong convexity gives $P_{\theta, \mu}(0) - P_{\theta, \mu}(d) \geq \mu \|d\|^2 / 2$. Rewriting Eq. (5.2) as $F(\theta + d) \leq F(\theta) + P(d) - P(0) - \sigma \|d\|^2 / 2$ proves Eq. (5.4). $\square$

5.1. A certificate appropriate to linearization. For a nonnegative vector $\tau = (\tau_i)$, introduce the enlarged scalar score sets

$$\Xi_{\tau_i}(r) = \begin{cases} \{c \operatorname{sgn}(r)/(c^2 + r^2)\}, & |r| > \tau_i, \\ [-1/c, 1/c], & |r| \leq \tau_i, \end{cases} \quad (5.5)$$

and the residual-tube stationarity measure

$$R_{\tau}(\theta) = \operatorname{dist}_{\infty}(0, -J(\theta)^{\top} \{(a_i \xi_i)_i : \xi_i \in \Xi_{\tau_i}(r_i(\theta))\} + N_{\Theta}(\theta)). \quad (5.6)$$

At $\tau = 0$ this is the exact constrained Clarke residual. For $\tau > 0$ it is a different object. In particular, small R_{τ} does not bound R_0 from above.

Theorem 5.2 (Nonlinear exact steps and shrinking tubes). *Let d_k be exact minimizers of Eq. (5.1), accepted by Eq. (5.2), with $0 < \mu_- \leq \mu_k \leq \mu_+$. Assume Θ is nonempty, compact and convex, $\theta^0 \in \Theta$, $\sigma > 0$, and Definition 2.1 holds on a neighborhood of Θ . Let $F_* = \lim_k F(\theta^k)$. Put*

$$\varepsilon_{i,k} = \frac{M_i}{2} \|d_k\|^2, \quad C_N = \mu_+ + \Gamma, \quad a_0 = (\mu_- + \sigma)/2. \quad (5.7)$$

Then $R_{\varepsilon_k}(\theta^{k+1}) \leq C_N \|d_k\|$, and every accumulation point is constrained Clarke stationary. If k_K minimizes $\|d_k\|$ over $0 \leq k < K$, then

$$\begin{aligned} R_{\varepsilon_{k_K}}(\theta^{k_K+1}) &\leq C_N \sqrt{\frac{F(\theta^0) - F_*}{a_0 K}}, \\ \varepsilon_{i,k_K} &\leq \frac{M_i [F(\theta^0) - F_*]}{2a_0 K}. \end{aligned} \quad (5.8)$$

Proof. Subproblem optimality supplies $s_i \in \operatorname{Sign}(r_i(\theta^k) - J_i(\theta^k)d_k)$ and $n_k \in N_{\Theta}(\theta^{k+1})$ with

$$-J(\theta^k)^{\top} (w_i(\theta^k)s_i)_i + \mu_k d_k + n_k = 0. \quad (5.9)$$

If $|r_i(\theta^{k+1})| > \varepsilon_{i,k}$, the actual and linearized residuals have the same sign; otherwise the enlarged score interval permits every $s_i \in [-1, 1]$. Thus $a_i \phi'(|r_i(\theta^{k+1})|)s_i$ is admissible in Eq. (5.6). Replacing the Jacobian and weights in Eq. (5.9) by their values at θ^{k+1} changes the vector by at most $\Gamma \|d_k\|$: the weight contribution is bounded by $\sum_i a_i L_{\phi} Q_i^2 \|d_k\|$ and the Jacobian contribution by $\sum_i a_i M_i \|d_k\| / c$. This proves the certificate bound.

Eq. (5.4) gives square summability and Eq. (5.8). At a cluster point, $d_k \rightarrow 0$ and $\varepsilon_k \rightarrow 0$. Bounded selections s_i have convergent subsequences. If a limiting residual is nonzero, its limiting sign is its actual sign; at a zero residual,



every limiting selection in $[-1, 1]$ is admissible. The normals in Eq. (5.9) are bounded. The closed graph of the normal cone then gives the exact limiting stationarity inclusion. $\square$

Theorem 5.3 (Nonlinear inverse-power descent and tube certificates). *Let $\Theta \subset \mathbb{R}^p$ be nonempty, compact and convex, let $\theta^0 \in \Theta$, and assume Definition 2.1 on a convex neighborhood of Θ . Fix $q \in \mathcal{Q}$, $c > 0$, weights $a_i > 0$ and $\lambda \geq 0$. Let P be convex and continuously differentiable on that neighborhood. For $\tau_i \geq 0$, define*

$$\Xi_{\tau_i}^q(r) = \begin{cases} \{\operatorname{sgn}(r)w_q(|r|)\}, & |r| > \tau_i, \\ [-1/c, 1/c], & |r| \leq \tau_i, \end{cases} \quad (5.10)$$

$$R_{\tau_i}^{q,\lambda}(\theta) = \operatorname{dist}_{\infty}(0, -J(\theta)^T \{(a_i \xi_i)_i : \xi_i \in \Xi_{\tau_i}^q(r_i(\theta))\} + \lambda \nabla P(\theta) + N_{\Theta}(\theta)). \quad (5.11)$$

At iteration k , write $r_i^k = r_i(\theta^k)$, $J_i^k = J_i(\theta^k)$ and let

$$\mathcal{M}_{q,k}(d) = \sum_i a_i [\ell_{q,c}(|r_i^k|) + w_q(|r_i^k|)(|r_i^k - J_i^k d| - |r_i^k|)] + \lambda P(\theta^k + d), \quad (5.12)$$

$$\mathcal{S}_{q,k}(d) = \mathcal{M}_{q,k}(d) + \frac{\mu_k}{2} \|d\|^2, \quad d \in \Theta - \theta^k. \quad (5.13)$$

For $\sigma > 0$, accept an exact minimizer d_k of $\mathcal{S}_{q,k}$ when

$$J_{q,\lambda}(\theta^k + d_k) \leq \mathcal{M}_{q,k}(d_k) + \frac{\mu_k - \sigma}{2} \|d_k\|^2. \quad (5.14)$$

This test holds for every feasible d whenever

$$\mu_k \geq \sigma + \sum_i a_i w_q(|r_i^k|) M_i; \quad \sigma + \sum_i a_i M_i / c \text{ is a uniform sufficient threshold.} \quad (5.15)$$

Thus a doubling search starting at a fixed $\mu_{\min} > 0$ terminates. Suppose the accepted coefficients satisfy $0 < \mu_- \leq \mu_k \leq \mu_+ < \infty$ and put

$$\Gamma_q = \sum_i a_i (L_q Q_i^2 + M_i / c), \quad C_q = \mu_+ + \Gamma_q, \quad a_q = (\mu_- + \sigma) / 2, \quad \varepsilon_{i,k} = M_i \|d_k\|^2 / 2. \quad (5.16)$$

Then $\theta^{k+1} = \theta^k + d_k$ satisfies

$$J_{q,\lambda}(\theta^k) - J_{q,\lambda}(\theta^{k+1}) \geq \frac{\mu_k + \sigma}{2} \|d_k\|^2, \quad (5.17)$$

$$R_{\varepsilon_k}^{q,\lambda}(\theta^{k+1}) \leq C_q \|d_k\|. \quad (5.18)$$

The steps are square summable and every accumulation point is constrained Clarke stationary for $J_{q,\lambda}$. Let $J_{\infty} = \lim_k J_{q,\lambda}(\theta^k)$ and let k_K minimize $\|d_k\|$ over $0 \leq k < K$. Then

$$\begin{aligned} R_{\varepsilon_{k_K}}^{q,\lambda}(\theta^{k_K+1}) &\leq C_q \sqrt{\frac{J_{q,\lambda}(\theta^0) - J_{\infty}}{a_q K}}, \\ \varepsilon_{i,k_K} &\leq \frac{M_i [J_{q,\lambda}(\theta^0) - J_{\infty}]}{2a_q K}. \end{aligned} \quad (5.19)$$

Proof. The function $\mathcal{M}_{q,k}$ is convex because its residual terms are positive multiples of absolute affine functions and its penalty is retained exactly. Compactness and $\mu_k > 0$ give a unique inner minimizer. Taylor's remainder and the increasing concave scalar loss imply

$$J_{q,\lambda}(\theta^k + d) \leq \mathcal{M}_{q,k}(d) + \frac{1}{2} \sum_i a_i w_q(|r_i^k|) M_i \|d\|^2. \quad (5.20)$$

The penalty appears identically on both sides. This proves the acceptance threshold; $w_q \leq 1/c$ gives its uniform bound. A doubling search from $\mu_{\min}$ accepts a coefficient no larger than $\max\{\mu_{\min}, 2(\sigma + \sum_i a_i M_i / c)\}$. For an exact minimizer, strong convexity gives $\mathcal{S}_{q,k}(0) - \mathcal{S}_{q,k}(d_k) \geq \mu_k \|d_k\|^2 / 2$. Since $\mathcal{M}_{q,k}(0) = J_{q,\lambda}(\theta^k)$, acceptance yields Eq. (5.17).



Optimality supplies $s_i \in \text{Sign}(r_i^k - J_i^k d_k)$ and $n_k \in N_\Theta(\theta^{k+1})$ such that

$$-J(\theta^k)^\top (a_i w_q(|r_i^k|) s_i)_i + \lambda \nabla P(\theta^{k+1}) + \mu_k d_k + n_k = 0. \quad (5.21)$$

Outside $\varepsilon_{i,k}$ the actual and linearized residuals have the same sign. Inside the tube, $w_q(|r_i(\theta^{k+1})|) s_i$ lies in $[-1/c, 1/c]$. These choices are therefore admissible in Eq. (5.11). Replacing the old weights and Jacobian by their new values changes their product by at most

$$\sum_i a_i (L_q Q_i^2 + M_i/c) \|d_k\| = \Gamma_q \|d_k\|. \quad (5.22)$$

The gradient of the exact penalty is already evaluated at the new point and contributes no gradient-Lipschitz remainder. The infinity norm is bounded by the Euclidean norm, giving Eq. (5.18).

Summing descent proves square summability and Eq. (5.19). Along an accumulation subsequence, $d_k \rightarrow 0$ and $\varepsilon_k \rightarrow 0$. Bounded scalar selections have convergent subsequences. At a nonzero limiting residual their signs become the actual signs; at zero their limits belong to the full scalar interval. Continuity of ∇P and boundedness of J on the compact set bound the normals. The closed normal-cone graph gives constrained Clarke stationarity. No global-minimum or whole-sequence convergence assertion follows from this argument. $\square$

Remark 5.4 (Scope of nonlinear tube rates). A linearized residual may vanish while the true residual is nonzero of order $\|d\|^2$. At that nearby nonzero residual the scalar Clarke set is a singleton, even if the model used an interior value of $[-1/c, 1/c]$. The resulting change need not be of order $\|d\|$. Eq. (5.8) controls both the certificate and the vanishing data tube. It does not claim $R_0(\theta^{k+1}) = O(\|d_k\|)$ for nonlinear g_i . No whole-sequence nonlinear KL theorem is asserted here from this weaker certificate alone.

Algorithm 2 Backtracking proximal linearization for the inverse-power family.

Require: (g, J) , Θ , feasible θ^0 , fixed a_i, c, q, λ, P , $\sigma > 0$, $\mu_{\min} > 0$, and inner error targets η_k .

- 1: Set $\mu_{-1} \leftarrow \mu_{\min}$.
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: Evaluate r^k, J_k and $a_i w_q(|r_i^k|)$; retain λP exactly in the model.
 - 4: Start $\mu \leftarrow \max(\mu_{\min}, \mu_{k-1}/2)$.
 - 5: **repeat**
 - 6: Form $\mathcal{S}_{q,k}$ in Eq. (5.13) with the current trial coefficient μ .
 - 7: Find feasible d with $\mathcal{S}_{q,k}(d) \leq \mathcal{S}_{q,k}(0)$ and certified inner error $\mathcal{G}_k \leq \eta_k$.
 - 8: Test Eq. (5.14) using the actual nonlinear objective.
 - 9: **if** the acceptance test fails **then**
 - 10: $\mu \leftarrow 2\mu$.
 - 11: **end if**
 - 12: **until** accepted at the requested inner accuracy
 - 13: Set $\mu_k \leftarrow \mu$, $\theta^{k+1} \leftarrow \theta^k + d$; record step, gap and objective decrease.
 - 14: Evaluate the chosen tube certificate independently of the solver status.
 - 15: Stop at the reported finite tolerances, or continue with vanishing gaps as in Corollary 6.4.
 - 16: **end for**
-

6. DUAL GAPS AND INEXACT SUBPROBLEMS

An inner solver generally stops before finding the exact subproblem minimizer. A feasible primal point supplies an upper bound on that minimum, and a feasible dual point supplies a lower bound. Their difference, $\mathcal{G}$, bounds the remaining inner optimization error independently of the solver's success flag. Strong convexity also turns it into a distance bound: $\|d - \bar{d}\| \leq \sqrt{2\mathcal{G}/\mu}$ for the exact subproblem solution $\bar{d}$. For illustration, $\mathcal{G} = 10^{-8}$ and $\mu = 2$ imply a distance of at most 10^{-4} in the chosen parameter coordinates. This is an arithmetic illustration of the gap bound, not an additional numerical experiment.



Consider Eq. (5.1) at a fixed outer iterate; abbreviate $D = \Theta - \theta$. A closed box Θ makes every projection below componentwise clipping, but the theory allows any nonempty closed convex D .

Proposition 6.1 (A smooth dual for a constrained weighted-LAD model). *Let $\mu > 0$, $w_i \geq 0$ and $D \subset \mathbb{R}^p$ be nonempty, closed and convex. For fixed $r \in \mathbb{R}^n$, $J \in \mathbb{R}^{n \times p}$ and $|\alpha_i| \leq w_i$, set*

$$\begin{aligned} d_\alpha &= \text{proj}_D(J^\top \alpha / \mu), \\ q(\alpha) &= r^\top \alpha + \frac{\mu}{2} \|d_\alpha\|^2 - (J^\top \alpha)^\top d_\alpha. \end{aligned} \quad (6.1)$$

The primal minimum equals $\max_{|\alpha_i| \leq w_i} q(\alpha)$, and the maximum is attained. The dual is concave and continuously differentiable, with

$$\nabla q(\alpha) = r - Jd_\alpha, \quad \|\nabla q(\alpha) - \nabla q(\beta)\| \leq \frac{\|J\|^2}{\mu} \|\alpha - \beta\|. \quad (6.2)$$

For every feasible primal–dual pair (d, α) , the gap $\mathcal{G} = P(d) - q(\alpha)$ satisfies

$$0 \leq P(d) - P(d^*) \leq \mathcal{G}, \quad \|d - d^*\| \leq \sqrt{2\mathcal{G}/\mu}, \quad (6.3)$$

where d^* is the unique primal minimizer.

Proof. Use $\sum_i w_i |r_i - J_i d| = \max_{|\alpha_i| \leq w_i} \alpha^\top (r - Jd)$. The remaining minimization is a strongly convex quadratic over D ; completing the square gives d_α and Eq. (6.1). Fenchel duality applies because the weighted absolute-value function is finite and continuous everywhere and D is nonempty. Compactness of the dual box gives attainment. Uniqueness of the inner quadratic minimizer allows differentiation by Danskin’s theorem. Nonexpansiveness of projection proves Eq. (6.2). Weak duality and strong convexity prove Eq. (6.3). $\square$

For $D = \mathbb{R}^p$, maximizing q is equivalent to minimizing $\|J^\top \alpha\|^2 / (2\mu) - r^\top \alpha$ over the dual box. An accelerated projected-gradient method has the usual $O(j^{-2})$ dual objective rate [5]; this alone is not a stated rate for its computable primal–dual gap. The numerical implementation solves the sparse epigraph quadratic program using OSQP [46] and independently evaluates Eq. (6.1). Solver status alone is not used as a gap certificate.

Algorithm 3 Feasible dual iteration and an independently evaluated inner gap.

Require: J, r , weights $w_i \geq 0$, $\mu > 0$, nonempty closed convex D , $|\alpha_i^0| \leq w_i$, and $\eta \geq 0$.

```

1: if  $J = 0$  then
2:    $d^* \leftarrow \text{proj}_D(0)$ ;  $\alpha_i^* \leftarrow w_i \text{sgn}(r_i)$  for every  $i$ , with  $\text{sgn}(0) = 0$ .
3:   return  $(d^*, \alpha^*, 0)$ . ▷ Zero gap in exact arithmetic
4: end if
5:  $L \leftarrow \|J\|^2 / \mu > 0$ .
6: for  $j = 0, 1, \dots$  do
7:    $d_j \leftarrow \text{proj}_D(J^\top \alpha^j / \mu)$ .
8:    $\mathcal{G}_j \leftarrow \sum_i w_i |r_i - J_i d_j| + \mu \|d_j\|^2 / 2 - q(\alpha^j)$ .
9:   if  $\mathcal{G}_j \leq \eta$  then
10:    return  $(d_j, \alpha^j, \mathcal{G}_j)$ .
11:   end if
12:    $\alpha^{j+1} \leftarrow \text{proj}_{[-w, w]}(\alpha^j + (r - Jd_j)/L)$ .
13: end for

```

For $J = 0$, the primal minimizer is $d^* = \text{proj}_D(0)$ and the feasible dual vector $\alpha_i^* = w_i \text{sgn}(r_i)$ maximizes $r^\top \alpha$. Both objective values equal $\sum_i w_i |r_i| + \mu \|d^*\|^2 / 2$. At a feasible outer iterate, $0 \in D$ and hence $d^* = 0$. The immediate return in Algorithm 3 therefore has zero gap in exact arithmetic and never evaluates a division by $L = 0$.



Proposition 6.2 (A feasible inner-gap bound with a quadratic penalty). *For $\mu > 0$, a symmetric positive semidefinite Q_P , a nonempty closed box Θ , and fixed weights $w_i \geq 0$, with $z_0 \in \mathbb{R}^p$, consider*

$$M(z) = \sum_i w_i |y_i - x_i^T z| + \frac{1}{2} z^T Q_P z + \frac{\mu}{2} \|z - z_0\|^2. \quad (6.4)$$

For any feasible z , residuals $r_i = y_i - x_i^T z$, any $|\alpha_i| \leq w_i$ and any $v \in N_\Theta(z)$, define

$$\begin{aligned} e &= (Q_P + \mu I)z - \mu z_0 - X^T \alpha + v, \\ \bar{\mathcal{G}} &= \sum_i \{w_i |r_i| - \alpha_i r_i\} + \|e\|^2 / (2\mu). \end{aligned} \quad (6.5)$$

Then $0 \leq M(z) - \min_\Theta M \leq \bar{\mathcal{G}}$ and $\|z - z^*\| \leq \sqrt{2\bar{\mathcal{G}}/\mu}$, where z^* is the unique minimizer of M on Θ .

Proof. The quadratic $H(u) = u^T Q_P u / 2 + \mu \|u - z_0\|^2 / 2 - \alpha^T X u$ is μ -strongly convex on the box. Its supporting inequality at z gives $\inf_\Theta H \geq H(z) - \|e\|^2 / (2\mu)$ by completing the square; the feasible normal has nonpositive inner product with every feasible displacement. Thus $\alpha^T y + H(z) - \|e\|^2 / (2\mu)$ is a dual lower bound. Subtracting it from $M(z)$ gives the displayed upper gap. Strong convexity of M gives the distance bound. $\square$

This bound supplies the quadratic-penalty inner error used below and in the affine MM cross-check of Section E. The matrix Q_P includes the criterion penalty coefficient λ ; $\mu > 0$ remains the proximal coefficient. Feasible scores and actual box activity are checked independently of solver status.

Theorem 6.3 (Inexact cluster-point convergence). *Assume nonempty compact convex Θ , $\theta^0 \in \Theta$, Definition 2.1 on a neighborhood of Θ , $\sigma > 0$, and $0 < \mu_- \leq \mu_k \leq \mu_+ < \infty$. Put $F_* = \lim_k F(\theta^k)$. At iteration k , let d_k be feasible and satisfy*

$$P_k(d_k) \leq P_k(0), \quad \mathcal{G}_k \leq \eta_k, \quad \eta_k \rightarrow 0, \quad (6.6)$$

together with Eq. (5.2). Then $\sum_k \|d_k\|^2 < \infty$, and every accumulation point is constrained Clarke stationary. If, additionally, $\mathcal{G}_k \leq \gamma \|d_k\|^2$ for a fixed $\gamma \geq 0$, the exact model mapping obeys

$$\min_{k < K} \mu_k \|d_k^*\| \leq \mu_+ \left(1 + \sqrt{2\gamma/\mu_-}\right) \sqrt{\frac{2[F(\theta^0) - F_*]}{\sigma K}}. \quad (6.7)$$

Proof. Acceptance and model decrease yield $F(\theta^k) - F(\theta^{k+1}) \geq \sigma \|d_k\|^2 / 2$. Hence the computed steps vanish. Eq. (6.3) gives $\|d_k - d_k^*\| \rightarrow 0$, so the exact model minimizers d_k^* vanish too. Their optimality inclusions have bounded score selections and bounded normals at $\theta^k + d_k^*$. These points and θ^{k+1} share each cluster limit. The same sign and normal-cone closed-graph argument as in Theorem 5.2 proves stationary accumulation. Under the relative-gap condition, $\|d_k^*\| \leq (1 + \sqrt{2\gamma/\mu_-}) \|d_k\|$. Summing sufficient decrease and selecting the smallest computed step proves Eq. (6.7). $\square$

Corollary 6.4 (Inexact nonlinear inverse-power steps). *Assume the domain, map, penalty and fixed-parameter hypotheses of Theorem 5.3, with $0 < \mu_- \leq \mu_k \leq \mu_+ < \infty$. Let d_k^* minimize $\mathcal{S}_{q,k}$ exactly. Suppose feasible computed steps satisfy Eq. (5.14) and*

$$\mathcal{S}_{q,k}(d_k) \leq \mathcal{S}_{q,k}(0), \quad 0 \leq \mathcal{S}_{q,k}(d_k) - \mathcal{S}_{q,k}(d_k^*) \leq \mathcal{G}_k \leq \eta_k, \quad \eta_k \rightarrow 0. \quad (6.8)$$

Here $\mathcal{G}_k$ is a certified upper bound on inner objective error. Then the computed steps are square summable and every accumulation point is constrained Clarke stationary for $J_{q,\lambda}$. If also $\mathcal{G}_k \leq \gamma \|d_k\|^2$ for a fixed $\gamma \geq 0$, then, with $J_\infty = \lim_k J_{q,\lambda}(\theta^k)$,

$$\min_{0 \leq k < K} \mu_k \|d_k^*\| \leq \mu_+ \left(1 + \sqrt{2\gamma/\mu_-}\right) \sqrt{\frac{2[J_{q,\lambda}(\theta^0) - J_\infty]}{\sigma K}}. \quad (6.9)$$



Proof. Acceptance and model decrease give objective decrease at least $\sigma \|d_k\|^2/2$, so the computed steps vanish and are square summable. Strong convexity gives $\|d_k - d_k^*\| \leq \sqrt{2\mathcal{G}_k/\mu_k}$; consequently $d_k^* \rightarrow 0$. Apply the optimality inclusion in Theorem 5.3 to d_k^* at $\theta^k + d_k^*$. These points and the computed iterates have the same accumulation limits. Bounded scores, continuous ∇P , and the closed sign and normal graphs give stationary limits. Under the relative-gap hypothesis, $\|d_k^*\| \leq (1 + \sqrt{2\gamma/\mu_-}) \|d_k\|$. Summing the computed decrease and selecting the smallest step proves Eq. (6.9). $\square$

For $\lambda = 0$, the dual construction in Proposition 6.1 applies with weights $a_i w_q(|r_i^k|)$ and proximal coefficient μ_k . For a quadratic penalty, Proposition 6.2 supplies an independently evaluated upper gap. A fixed finite tolerance reports an approximate solve; it does not satisfy the vanishing-gap hypothesis by itself. The inexact rate concerns the exact model mapping, not the unchanged-data Clarke residual at a computed nonlinear point.

An observable bound for the model mapping is

$$\mu_k \|d_k^*\| \leq \mu_k \left(\|d_k\| + \sqrt{2\mathcal{G}_k/\mu_k} \right). \quad (6.10)$$

It should not be relabeled as an exact Clarke residual at the computed nonlinear point. A vanishing primal gap alone cannot justify that substitution: for $P(d) = |d| + \mu d^2/2$, feasible positive $d \downarrow 0$ have vanishing objective gap but subgradient magnitude $1 + \mu d$, which tends to one. Finite experimental tolerances are reported in Table 32.

7. EXACT AND BACKWARD-DATA STATIONARITY CERTIFICATES

The certificate asks whether admissible residual slopes and constraint reactions can balance at the reported parameter. The smallest remaining imbalance is the stationarity residual. The linear program below chooses these slopes explicitly rather than assuming that an optimizer's stopping message establishes the balance. A response tube allows observations already within a declared tolerance of the fitted map to be matched exactly in a nearby dataset. Both the tube width and the remaining imbalance must be reported: a zero imbalance obtained from a large tube is a different statement from exact stationarity for the original data.

Let $\Theta = \{\theta : B\theta \leq b, E\theta = e\}$ be polyhedral, with active inequality rows B_A at the point being tested. Split residuals into $I_0 = \{i : r_i = 0\}$ and $I_1 = I_0^c$, and let $\bar{\alpha}_i = a_i c \operatorname{sgn}(r_i)/(c^2 + r_i^2)$ on I_1 .

Proposition 7.1 (A linear program for the exact Clarke residual). *The exact value $R_0(\theta)$ is the optimal value of*

$$\begin{aligned} \min_{t, \alpha_0, \nu_A, v} \quad & t, \\ -t\mathbf{1} \leq \quad & -J_0^\top \alpha_0 - J_1^\top \bar{\alpha}_1 + B_A^\top \nu_A + E^\top v \leq t\mathbf{1}, \\ -a_i/c \leq \alpha_i \leq a_i/c \quad & (i \in I_0), \quad \nu_A \geq 0, \quad t \geq 0. \end{aligned} \quad (7.1)$$

Replacing I_0 by $I_\tau = \{i : |r_i| \leq \tau_i\}$ yields $R_\tau(\theta)$.

Proof. Eq. (2.7) gives the free score intervals on zero residuals and the fixed scores on the complement. The polyhedral normal cone is $\{B_A^\top \nu_A + E^\top v : \nu_A \geq 0\}$. The two componentwise inequalities represent an infinity-norm epigraph. Minimization therefore computes precisely Eq. (5.6). The image of the score box plus the finitely generated normal cone is closed, so the distance is attained. $\square$

Theorem 7.2 (Backward-data meaning of a numerical zero test). *For a fixed θ and tube τ , define a perturbed dataset*

$$\tilde{y}_i = \begin{cases} g_i(\theta), & |y_i - g_i(\theta)| \leq \tau_i, \\ y_i, & \text{otherwise.} \end{cases} \quad (7.2)$$

Then $|\tilde{y}_i - y_i| \leq \tau_i$ and $R_\tau(\theta)$ equals the exact constrained Clarke residual for the objective with data $\tilde{y}$, evaluated at θ . The weights, scale, parameter constraints, and observation map are unchanged.

Proof. Every selected residual is exactly zero for $\tilde{y}$, so its scalar subdifferential is $[-1/c, 1/c]$. All unselected residuals remain unchanged. Eq. (2.7) and the normal-cone formula then coincide exactly with Eq. (5.6). The componentwise perturbation bound follows from the selection rule. $\square$



Example 7.3 (A small residual does not imply a small exact certificate). With $g_1(\theta) = g_2(\theta) = \theta$, $y = (0, 2)$, $a_i = 1/2$, $c = 1$, and $\theta = 10^{-7}$, the exact residual R_0 is approximately 0.4. A tube of width 10^{-6} gives $R_\tau = 0$, because the first response can be moved to 10^{-7} and its zero-residual interval can cancel the second score. This is a valid backward-data certificate and an invalid claim of exact stationarity for the unchanged responses.

The radius of a selected interval in Eq. (7.1) must be a_i/c , not $a_i\phi'(|r_i|)$. The latter would describe a different relaxation. Similarly, a nearly active parameter constraint must not be treated as active without a separate bound-perturbation interpretation. The implementation clips feasible box iterates and treats only stored coordinates equal to the bound as active. Numerical LP solutions are still subject to the LP solver's feasibility tolerances; algebraic exactness of the formulation is not an assertion of interval-verified floating-point arithmetic.

8. ODE SENSITIVITIES AND SOLVER-AWARE ERROR CONTROL

The fit depends on two numerical outputs: predicted observations and their derivatives with respect to the parameters. Accurate trajectories alone do not ensure accurate derivatives. Theorem 8.1 therefore charges separately for observation-map error and sensitivity error. It transfers a first-order stationarity statement at the fitted parameter, rather than a bound on its distance to the true physical parameter. This distinction makes the stopping decision useful without requiring identification or global optimization to have been established first.

For Eq. (2.1), the state sensitivity $S = \partial_\theta x$ satisfies

$$\dot{S} = f_x(t, x, \theta)S + f_\theta(t, x, \theta), \quad S(0) = \partial_\theta x_0(\theta), \quad J_i = h_{i,x}S(t_i) + h_{i,\theta}. \quad (8.1)$$

8.1. Transfer of a discrete certificate.

Theorem 8.1 (Pointwise residual-tube certificate transfer). *Let $N, p \geq 1$, $\Theta \subset \mathbb{R}^p$ be nonempty, closed and convex, and fix $\theta \in \Theta$. For $b \in \{0, h\}$, let $g_b : \mathbb{R}^p \rightarrow \mathbb{R}^N$ be differentiable at θ , with row Jacobians $J_{i,b} = \nabla g_{i,b}(\theta)^\top$ and residuals $r_{i,b} = y_i - g_{i,b}(\theta)$. Index 0 denotes the continuous observation map and h its numerical approximation. All quantities below are evaluated at this same parameter. Both criteria use identical responses, fixed standardization, positive weights a_i , scale $c > 0$, penalty coefficient $\lambda \geq 0$, penalty P continuously differentiable near θ , and constraints Θ . Convexity of P and compactness of Θ are not needed for this transfer statement. A pilot-estimated scale may be frozen and conditioned on here; this is not a fixed-scale statistical consistency claim.*

For $q \in \{1\} \cup [2, \infty)$ and $u \geq 0$, define

$$G_c(u) = \frac{2}{\pi} \operatorname{atan2}(c, u), \quad w_q(u) = \frac{c G_c(u)^{q-1}}{c^2 + u^2},$$

$$L_q = \frac{9}{8\sqrt{3}c^2} + \frac{2(q-1)}{\pi c^2}. \quad (8.2)$$

Here $\operatorname{atan2}(c, 0) = \pi/2$. The positive-side weight satisfies $0 \leq w_q \leq 1/c$ and $|w_q(u) - w_q(v)| \leq L_q |u - v|$. For any nonnegative radius vector $\rho \in \mathbb{R}^N$, set

$$\begin{aligned} \mathcal{A}_{i,b}(\rho_i) &= \begin{cases} \{a_i \operatorname{sgn}(r_{i,b}) w_q(|r_{i,b}|)\}, & |r_{i,b}| > \rho_i, \\ [-a_i/c, a_i/c], & |r_{i,b}| \leq \rho_i, \end{cases} \\ N_\Theta(\theta) &= \{n \in \mathbb{R}^p : \langle n, v - \theta \rangle \leq 0 \text{ for every } v \in \Theta\}, \\ R_{\rho,b}^{q,\lambda}(\theta) &= \inf_{\substack{\alpha_i \in \mathcal{A}_{i,b}(\rho_i) \\ n \in N_\Theta(\theta)}} \|-J_b^\top \alpha + \lambda \nabla P(\theta) + n\|_\infty. \end{aligned} \quad (8.3)$$

The vector norm $\|v\|_\infty = \max_j |v_j|$ measures stationarity in the stated parameter coordinates; $\|v\|_2 = (\sum_j v_j^2)^{1/2}$ is used for each Jacobian row. Suppose finite nonnegative componentwise bounds satisfy

$$|g_{i,h} - g_{i,0}| \leq e_{i,h}, \quad \|J_{i,h} - J_{i,0}\|_2 \leq E_{i,h}, \quad \|J_{i,0}\|_2 \leq Q_i \quad (1 \leq i \leq N). \quad (8.4)$$

For any $\tau_i \geq 0$, let $e_h = (e_{i,h})_{i=1}^N$ and define

$$R_{\tau+e_h,0}^{q,\lambda}(\theta) \leq R_{\tau,h}^{q,\lambda}(\theta) + B_{q,h}, \quad B_{q,h} = \sum_{i=1}^N a_i (E_{i,h}/c + Q_i L_q e_{i,h}). \quad (8.5)$$

Moreover, every feasible discrete pair (α_h, n_h) gives

$$R_{\tau+e_h,0}^{q,\lambda}(\theta) \leq \hat{R}_{\tau,h}^{q,\lambda} + B_{q,h}, \quad \hat{R}_{\tau,h}^{q,\lambda} = \|-J_h^\top \alpha_h + \lambda \nabla P(\theta) + n_h\|_\infty. \quad (8.6)$$

Any larger justified Lipschitz constant may replace L_q . No optimizer, optimization sequence, statistical identification, unchanged-tube stationarity or parameter-error bound is transferred.

Proof. Choose any discrete admissible weighted score α_h and normal $n \in N_\Theta(\theta)$. If the continuous residual satisfies $|r_i(\theta)| > \tau_i + e_{i,h}$, the discrete residual exceeds τ_i and has the same sign. Replacing its score by the continuous score changes that component by at most $a_i L_q e_{i,h}$. Otherwise the enlarged continuous score set is $[-a_i/c, a_i/c]$, so it admits the unchanged discrete score. The resulting vector α_0 is feasible for the continuous certificate, and

$$\begin{aligned} \|J_0^\top \alpha_0 - J_h^\top \alpha_h\|_2 &\leq \sum_i Q_i |\alpha_{i,0} - \alpha_{i,h}| + \sum_i E_{i,h} |\alpha_{i,h}| \\ &\leq \sum_i a_i (Q_i L_q e_{i,h} + E_{i,h}/c) = B_{q,h}. \end{aligned} \quad (8.7)$$

The fixed penalty gradient and the normal n cancel in the vector comparison. The infinity norm is no larger than the Euclidean norm. The triangle inequality therefore gives the claimed bound for each discrete feasible pair. Taking the infimum over those pairs proves Eq. (8.5); attainment of that infimum is unnecessary. $\square$



Sufficient smoothness of f, h_i, x_0 on a solution tube, existence over the observation horizon, and a compact parameter set justify Definition 2.1. Blow-up, changing event sequences, or nonsmooth observation operators fall outside these assumptions. Differentiating every stage of a fixed-step Runge–Kutta map produces the Jacobian of that *discrete* map. Differentiating a different solver, or ignoring parameter-dependent adaptive step decisions, need not do so [21, 24].

For the ordinary-GLDM specialization, let $g_{i,h}$ and $J_{i,h}$ be a solver map and its Jacobian at the point being tested. Suppose bounds are available at that point:

$$|g_{i,h} - g_i| \leq e_{i,h}, \quad \|J_{i,h} - J_i\| \leq E_{i,h}. \quad (8.8)$$

Use $R_{\tau,h}$ for the tube certificate formed from the discrete map, and $R_{\tau,0}$ for the continuous observation map.

Corollary 8.2 (Ordinary-GLDM specialization). *Under Theorem 8.1 with $q = 1$ and $\lambda = 0$,*

$$R_{\tau+e_h,0}(\theta) \leq R_{\tau,h}(\theta) + B_h, \quad B_h = \sum_i a_i (E_{i,h}/c + Q_i L_\phi e_{i,h}). \quad (8.9)$$

Proof. Use $w_1 = \phi'$ and $L_1 = L_\phi$ in Eq. (8.5). □

Theorem 8.1 transports a feasible stationarity bound while quantifying the required response perturbation. Its proof also applies to signed scores bounded by K with L -Lipschitz positive-side weights, replacing $1/c, L_q$ by K, L . Arctangent specificity lies in the explicit weights, constants, weighted-LAD construction and nonlinear tube analysis, not an exclusive perturbation principle or a claim of exhaustive historical priority.

The proof has two cases that explain the tube. If a continuous residual lies farther than $\tau_i + e_{i,h}$ from zero, its discrete counterpart lies outside the old tube and has the same sign. The score then changes by a Lipschitz-controlled amount. If it lies inside the enlarged tube, the full zero-residual slope interval is available, so the old discrete slope is still admissible. The remaining imbalance is just the effect of changing slopes and Jacobians, bounded by $B_{q,h}$. No change of parameter or new optimization is used in this comparison.

Every term in a two-component example. Take $\Theta = [-1, 1]$, $\theta = 0$, $P = 0$, $\lambda = 0$, $q = c = 1$, $a_1 = a_2 = 1/2$, and the affine maps

$$g_0(\theta) = (\theta, 2\theta), \quad g_h(\theta) = (1.02\theta + 0.01, 1.98\theta - 0.02). \quad (8.10)$$

Choose $y = (0.025, 1)$ and $\tau = (0.02, 0.1)$. At zero, the residuals are $r_h = (0.015, 1.02)$ and $r_0 = (0.025, 1)$. Exact pointwise bounds are $e_h = (0.01, 0.02)$, $E_h = (0.02, 0.02)$ and $Q = (1, 2)$. Use the conservative $L = 1 \geq L_1 = 9/(8\sqrt{3})$. The first discrete residual is inside its tube; the second is outside. The feasible choice

$$\alpha_h = \begin{pmatrix} -41250 & 1250 \\ 86717 & 5101 \end{pmatrix}, \quad n_h = 0, \quad \hat{R}_h = 0 \quad (8.11)$$

cancels the two discrete Jacobian contributions exactly. The enlarged continuous tube is $(0.03, 0.12)$. Retain its first score and replace its second by $1/4$, obtaining $\alpha_0 = (-41250/86717, 1/4)$. Every inflation contribution is now explicit:

$$\begin{aligned} B_1 &= \frac{1}{2}(0.02 + 1 \cdot 1 \cdot 0.01) = 0.015, \\ B_2 &= \frac{1}{2}(0.02 + 2 \cdot 1 \cdot 0.02) = 0.030, \\ B &= 0.045, \quad \hat{R}_0 = \frac{4217}{173434} \simeq 0.024315, \\ \hat{R}_h + B - \hat{R}_0 &= \frac{358753}{17343400} > 0. \end{aligned} \quad (8.12)$$

The optimal enlarged-tube continuous residual is actually zero, using $\alpha_0^* = (-1/2, 1/4)$. Thus the transferred witness supplies an upper bound without solving a second optimization problem. In contrast, the unchanged continuous tube fixes both scores and gives $R_{\tau,0} = 3201/3202 > 0.045$. This example shows both the witness construction and the necessity of enlargement. Its rational arithmetic is independently checkable; it is an algebraic illustration, not an additional validated ODE integration experiment.

For nonlinear ODEs, rigorous state and sensitivity enclosures require validated integration on a solution tube, regular observation operators and controlled rounding. The present nonlinear experiments do not provide such enclosures. The scalar linear construction in Section 9.8 does. An approximate normal at a nearly active box face is not automatically



TABLE 6. Inputs required by Theorem 8.1. All errors use the same fixed observation standardization. Numerical solver tolerances alone are not bounds in the theorem.

Input	Justified construction	Measured diagnostic only
$e_{i,h}$	Enclose the exact observation map and subtract the discrete map, including observation-operator error and rounding	Difference from an analytic floating-point reference, tighter solver or mesh
$E_{i,h}$	Enclose continuous sensitivities and the consistently differentiated numerical map; subtract row-wise	Two-solver or mesh Jacobian differences
Q_i	Upper bound of a continuous Jacobian enclosure; alternatively $\ J_{i,h}\ _2 + E_{i,h}$ with rigorous norm rounding	Computed reference row norm
$L_q, 1/c$	Analytic bounds, enclosed upward when evaluated on a computer	Rounded constants without enclosure
Witness	Check score intervals and normal-cone signs; evaluate its stationarity expression with justified arithmetic	Solver-reported optimality, tolerance-based activity or floating-point LP residual

feasible: normal membership uses the actual parameter and actual constraint activity. If an interval comparison cannot decide a residual's tube membership, enclosure refinement or an unresolved return is required. Bounds need only hold at the tested parameter; a uniform-domain claim would require additional uniform bounds.

For fixed error bounds, scale and weights, the chosen analytic constants also give

$$B_{q_2,h} - B_{q_1,h} = \frac{2(q_2 - q_1)}{\pi c^2} \sum_i a_i Q_i e_{i,h} \geq 0 \quad (q_2 \geq q_1, q_1, q_2 \in \mathcal{Q}). \quad (8.13)$$

Thus stronger pointwise score attenuation can require a larger conservative transfer allowance. This comparison concerns the declared upper bound L_q , not the smallest possible error at a particular fit. It follows directly from Eq. (3.13) and does not rank statistical robustness.

The inequalities above are rigorous conditional mathematical statements. A rigorous computer-assisted application additionally requires justified upper bounds for the state and sensitivity errors, for Q_i , and for floating-point evaluation of feasible LP witnesses. Subtracting floating-point analytic and discrete evaluations produces measured discrepancies, not validated error enclosures. Section 9.7 therefore reports a numerical certificate-transfer verification; it does not claim a computer-assisted proof of the continuous-map certificate.

Proposition 8.3 (Objective and local minimizer perturbation). *For a nonempty Θ and fixed $a_{i,c} > 0$, suppose Eq. (8.8) holds uniformly on Θ . Define $\mathcal{E}_h = \sum_i a_i e_{i,h}/c$. Then $\sup_{\Theta} |F_h - F| \leq \mathcal{E}_h$. If $\theta^\dagger \in \Theta$ is a constrained minimizer on $U \cap \Theta$ for a neighborhood U , $F(\theta) \geq F(\theta^\dagger) + \kappa_g \|\theta - \theta^\dagger\|^2/2$ on $U \cap \Theta$ for some $\kappa_g > 0$, and a computed $\hat{\theta}_h \in U \cap \Theta$ satisfies $F_h(\hat{\theta}_h) \leq F_h(\theta^\dagger) + \epsilon_{\text{opt}}$ with $\epsilon_{\text{opt}} \geq 0$, then*

$$\|\hat{\theta}_h - \theta^\dagger\| \leq \sqrt{\frac{2(\epsilon_{\text{opt}} + 2\mathcal{E}_h)}{\kappa_g}}. \quad (8.14)$$

Proof. The Lipschitz constant $1/c$ from Lemma 2.2 gives the uniform objective bound. Consequently $F(\hat{\theta}_h) \leq F(\theta^\dagger) + \epsilon_{\text{opt}} + 2\mathcal{E}_h$. Combining this inequality with the assumed quadratic growth proves Eq. (8.14). $\square$

The proposition requires a value comparison within one basin, not stationarity alone. Consistently differentiated RK4 has $e_{i,h}, E_{i,h} = O(h^4)$ under sufficient smoothness on a fixed horizon [24], hence $B_h = O(h^4)$. Horizon and conditioning affect its constants; the order alone is not a computable error bound.



Corollary 8.4 (A guarded continuous-objective descent test). *Suppose $|F_h(v) - F(v)| \leq \mathcal{E}_h(v)$ at the two tested points. If*

$$F_h(\theta + d) + \mathcal{E}_h(\theta + d) \leq F_h(\theta) - \mathcal{E}_h(\theta) - \frac{\sigma}{2} \|d\|^2, \quad (8.15)$$

then $F(\theta + d) \leq F(\theta) - \sigma \|d\|^2 / 2$.

Proof. Bound the new continuous value above by $F_h(\theta + d) + \mathcal{E}_h(\theta + d)$ and the old value below by $F_h(\theta) - \mathcal{E}_h(\theta)$. $\square$

Mesh refinement should therefore be tied to a certificate budget, not to a fixed decimal precision alone. A practical target is $B_h \leq \epsilon_R/2$ and $\max_i e_{i,h} \leq \epsilon_y/2$, leaving the remaining budgets for the discrete LP residual and its chosen tube. Step doubling is useful for detecting insufficient resolution, but without a justified remainder estimate it is a diagnostic rather than a rigorous upper bound. The analytic oscillator in Section 9.7 permits numerical checks of observation-map and sensitivity errors.

Algorithm 4 Calibration with finite budgets and explicit outcome status.

Require: Fixed data, scales, box, q, λ, P ; initial mesh/tube; budgets $\epsilon_R, \epsilon_y > 0$; maximum refinements K ; finite inner limits; stagnation window $s \geq 1$ and threshold $\eta > 0$.

```

1: Initialize a feasible fit, history, and best budget ratio  $b \leftarrow \infty$ .
2: for  $j = 0, \dots, K$  do
3:   Fit or refine using the same solver map and Jacobian, subject to the finite inner limits.
4:   if integration or fitting fails then
5:     return UNRESOLVED, reason “solver failure”, last feasible fit and unmet/unknown conditions.
6:   end if
7:   Evaluate a feasible discrete residual upper bound  $\hat{R}_h \geq R_{\tau,h}^{q,\lambda}$  and  $(e_h, E_h, Q, B_{q,h})$  when available.
8:   if a required quantity is unavailable or nonfinite then
9:     return UNRESOLVED, reason “missing error or witness bound”, fit and missing quantities.
10:  end if
11:  Set  $v_R = \hat{R}_h + B_{q,h}$ ,  $v_y = \max_i(\tau_i + e_{i,h})$ .
12:  Record  $U = \{R : v_R > \epsilon_R\} \cup \{y : v_y > \epsilon_y\}$  and the numerical values.
13:  if  $U = \emptyset$  then
14:    if all map-error and witness bounds are justified then
15:      return VERIFIED, fit, tube and the bound in Theorem 8.1.
16:    else
17:      return NUMERICAL PASS, fit and measured checks; no rigorous continuous-map certificate.
18:    end if
19:  end if
20:  Update  $b = \min\{b, \max(v_R/\epsilon_R, v_y/\epsilon_y)\}$  and its history.
21:  if  $j = K$  or  $b$  improves by less than  $\eta$  over the last  $s$  refinements then
22:    return UNRESOLVED, reason “limit” or “stagnation”, fit,  $U$ ,  $v_R$ ,  $v_y$  and history.
23:  end if
24:  If  $B_{q,h} > \epsilon_R/2$  or  $\max_i e_{i,h} > \epsilon_y/2$ , refine the mesh.
25:  Otherwise, if  $\hat{R}_h > \epsilon_R - B_{q,h}$ , improve the fit and inner solve.
26:  Otherwise, reduce the tube if feasible; recompute the witness. Warm start the next finite fit.
27: end for

```

The stagnation test is applied only once s refinements have been recorded. Algorithm 4 always returns after at most $K + 1$ finite fitting attempts, but need not meet either accuracy budget. An unresolved return records which inequality failed, or which quantities could not be assessed. A numerical pass uses measured reference discrepancies and is distinct from a verified return with justified bounds, including feasible-witness evaluation. The fixed-mesh experiments



implement numerical checks of the transfer inequality, not validated adaptive integration. The accompanying control-flow implementation checks successful, missing-bound, failed-solver, stagnating and iteration-limited exits separately.

Algorithm 5 Independent evaluation of a feasible transfer witness.

Require: The data in Theorem 8.1; proposed discrete scores α_h and normal n_h ; evidence type for each bound.

- 1: Check $\theta \in \Theta$, $\alpha_{i,h} \in \mathcal{A}_{i,h}(\tau_i)$, $n_h \in N_\Theta(\theta)$, nonnegativity of the error bounds and common scaling.
 - 2: **if** feasibility or a required membership test is unresolved **then**
 - 3: **return** UNRESOLVED and the failed or undecided test.
 - 4: **end if**
 - 5: Evaluate $\widehat{R}_h = \|-J_h^\top \alpha_h + \lambda \nabla P(\theta) + n_h\|_\infty$ independently of the optimizer.
 - 6: **for** $i = 1, \dots, N$ **do**
 - 7: Compute $B_i = a_i(E_{i,h}/c + Q_i L_q e_{i,h})$ and $\rho_i = \tau_i + e_{i,h}$.
 - 8: **if** the continuous map is available and $|y_i - g_{i,0}| > \rho_i$ **then**
 - 9: Set $\alpha_{i,0} = a_i \operatorname{sgn}(y_i - g_{i,0}) w_q(|y_i - g_{i,0}|)$.
 - 10: **else**
 - 11: Retain $\alpha_{i,0} = \alpha_{i,h}$ when membership in the enlarged tube is resolved; otherwise record the continuous witness as unevaluated.
 - 12: **end if**
 - 13: **end for**
 - 14: Set $U = \widehat{R}_h + \sum_i B_i$; this bounds the continuous residual when all inputs are justified.
 - 15: If the continuous witness is evaluated, compute $\widehat{R}_0 = \|-J_0^\top \alpha_0 + \lambda \nabla P(\theta) + n_h\|_\infty$ and $S = U - \widehat{R}_0$.
 - 16: **return** U , componentwise inflations, tube, witness, any available slack and evidence type; use Algorithm 4 for the final budget status.
-

The evaluator needs no optimal LP solution: any feasible witness suffices. The theorem supplies U even if the exact continuous map is unavailable for direct evaluation. When both witnesses are enclosed, $S \geq 0$ is a sufficient independent check; a negative slack for the specifically constructed witness flags inconsistent bounds, feasibility or arithmetic and requires investigation. An arbitrary nonoptimal continuous witness must not be confused with the optimal residual. The reference implementation returns ordinary floating-point diagnostics explicitly as such, and separately verifies the exact rational illustration in Eq. (8.12).

Worked refinement sequence. The exact scalar benchmark in Section 9.8 provides a concrete use of Algorithm 4. Keep $q = 2$, $c = 1.5$, $\lambda = 0$, the standardized tube $\tau = 10^{-4}$ and the same ten observations. Prescribe $\epsilon_R = 10^{-4}$ and $\epsilon_y = 1.1 \times 10^{-4}$. At each mesh use its recorded rational fitted point and recompute the observation, Jacobian and scalar-LP enclosures by the arithmetic in Proposition 9.1. Table 7 replays these three recorded candidates; it does not assert a new parameter-identification experiment. The feasible discrete upper bound is zero in each case. The decisions compare exact rational endpoints, not the displayed rounded values.

TABLE 7. Algorithm 4 applied to the scalar enclosure records. $v_R = \widehat{R}_h + B_{2,h}^+$ and $v_y = \tau + e_h^{\max}$. Displayed positive bounds are rounded upward; R and y identify the unmet stationarity and response budgets.

h	$e_h^{\max}$	$\widehat{R}_h$	$B_{2,h}^+$	v_R	v_y	Decision
1/2	1.164×10^{-3}	0	6.514×10^{-3}	6.514×10^{-3}	1.264×10^{-3}	Refine: R, y
1/4	6.287×10^{-5}	0	3.461×10^{-4}	3.461×10^{-4}	1.629×10^{-4}	Refine: R, y
1/8	3.656×10^{-6}	0	1.995×10^{-5}	1.995×10^{-5}	1.037×10^{-4}	Verified

Both budgets fail at $h = 1/2$ and $h = 1/4$. At $h = 1/8$, $v_R < \epsilon_R$ and $v_y < \epsilon_y$; the justified enclosures therefore permit a VERIFIED return. The same numerical values supported only by floating-point discrepancies would yield



NUMERICAL PASS. Limiting the run to one refinement would instead return UNRESOLVED at $h = 1/4$ and record both unmet budgets. This distinction is about the evidence supporting the bounds, not the apparent accuracy of a trajectory plot.

Supporting oscillator momentum and energy calculations are retained in Appendix A; they are not hypotheses of the transfer theorem.

8.2. Assumptions and scope of the results. Table 8 records the setting and essential hypotheses of the main results. The complete statements govern each claim; local stationarity and statistical consistency require different assumptions. Table 9 separates deterministic conclusions from empirical findings.

TABLE 8. Map of deterministic and statistical results. “Penalty” refers to the additional criterion penalty, not the proximal term.

Result	Setting	Essential assumptions	Conclusion
Theorem 3.1	Affine/nonlinear; nonsmooth	Same residuals, fixed scales, weights and feasible set	Exact reversed loss–gain ordering
Theorem 3.3	Scalar; nonsmooth	$c > 0$, $q \in \{1\} \cup [2, \infty)$	Derivatives, bounded loss, score decay, distinct criteria
Theorem 3.5	Affine; convex smooth penalty	Compact convex set; exact subproblems; bounded positive proximal parameters	Descent and stationary cluster points
Theorem 4.2	Affine; no additional penalty	Bounded exact iterates; proximal bounds; polyhedral set for finite length	Stationarity rate; finite-length convergence in the stated case
Theorem 4.3	Unconstrained affine	Full interpolation rank, strict dual interiority, convergence to that minimum	Local exact finite termination
Theorem 5.3	Nonlinear inverse power; exact convex smooth penalty	Compact convex set; Lipschitz Jacobians; exact accepted models	Finite backtracking, tube rate and stationary cluster points
Theorem 5.2	Nonlinear; nonsmooth	Compact convex set, Lipschitz Jacobians, exact accepted models	Shrinking-tube rate and stationary cluster points
Corollary 6.4	Nonlinear inverse power; inexact models	Vanishing certified inner errors, model decrease and acceptance	Stationary cluster points; relative-gap model rate
Theorem 6.3	Nonlinear; inexact models	Vanishing feasible gaps, model decrease, acceptance, positive proximal bounds	Cluster-point stationarity; conditional model rate
Theorem 7.2	Affine/nonlinear; fixed point	Fixed map, scales, weights and constraints	Exact interpretation for perturbed responses
Theorem 8.1	Discrete/continuous inverse power; common smooth penalty	Pointwise map and Jacobian error bounds; fixed scale and constraints	Enlarged-tube transfer with L_q
Corollary 8.2	Discrete/continuous ODE maps	Pointwise state and Jacobian error bounds	Enlarged-tube certificate transfer
Proposition 9.1	Scalar linear ODE at rational interior points	Rational input, enclosing series arithmetic, resolved tube membership	Pointwise exact-rational transfer check; no global optimization claim
Proposition 9.2	Experimental eight-state model	Positive bounded rates, finite time, positive initial monomer fractions and specificity	Nonnegativity, conservation and a smooth observation map; not identifiability
Theorem A.2	Underdamped oscillator	Common initial state; known parameter pair and mesh	Parameter plus RK4 error envelope
Theorem 3.10	Statistical affine regression	Ergodicity, bounded design, compact set, fixed scale, unique target, near-global empirical solutions	Consistency for that population target

TABLE 9. Evidence levels and their scope. Algorithmic outcomes are defined separately in Section 1.

Evidence level	Result or observation	Scope
Proved	Theorem 8.1: feasible witness transfer with inflation and tube enlargement	Same parameter, common criterion settings and stated pointwise error bounds
Proved	Loss–gain identity, loss calculus and local convergence	Each result’s hypotheses in Table 8; no general parameter-accuracy conclusion
Proved	Affine statistical consistency	Fixed scale, bounded design and near-global solutions; not the implemented full procedure
Validated computation	Twelve scalar transfer inequalities	Exact-rational enclosures at twelve points; not a uniform parameter-domain result
Numerical verification	Oscillator mesh checks and January ODE transfer	Floating-point reference discrepancies and witness evaluations; not validated ODE integration
Empirical observation	Synthetic comparisons, STAT5 and SIR forecasts	Declared generators or limited time courses, local constrained fits and specified search protocols
Empirical observation	January polynomial forecasts and regularization	One supplied annual composition series; overlapping origins and one holdout

These conclusions do not jointly establish global optimality, parameter identification or accuracy, universal prediction superiority, or environmental trends. The transfer theorem concerns local stationarity for an enlarged response tube. Each computational claim retains its listed evidence level.

9. REPRODUCIBLE NUMERICAL VALIDATION

9.1. Experimental design and computational environment. The experiments separate three questions: whether a feasible stationarity bound transfers between maps, whether numerical stopping criteria are met, and whether a fitted model forecasts well. These questions are evaluated separately; predictive error is not a stationarity certificate.

The numerical study evaluates parameter estimation, forecast error and optimization certificates under specified data-generating mechanisms. Affine regression and nonlinear ODE trajectories provide known parameters and reference solutions; paired comparisons use the same observations and starting points. The computational package contains every generated observation, selected parameter estimate, multistart diagnostic, forecast record and figure-production script. Section 9.11 adds published experimental measurements of a nonlinear STAT5 system. Section 9.14 adds a January ODE calibration and mesh-transfer study using observed class compositions. The polynomial January study in Sections C and D, with sensitivity analyses in Section E, evaluates forecasting separately.

The primary design contains 600 affine datasets and 480 nonlinear state-observation datasets, the latter comprising 160 each for a two-parameter logistic model, a two-state oscillator, and logistic growth with an unknown initial condition. A separate contamination grid contains 780 affine and 260 three-parameter logistic datasets. Twenty paired oscillator datasets examine four meshes at $q = 1, 2$, and twenty additional logistic datasets examine scale and QP tolerances. All replications have independently initialized random generators; methods within one replication share the same data. The reproducibility package records all random seeds and their derivation from base integer 2026092701.

The controlled simulations were executed serially on an AMD EPYC 9V74 80-Core Processor host, using Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0 and OSQP 1.1.3 on Linux. BLAS/OpenMP used one thread. Timing includes the stated optimizer starts and excludes data generation, final tube LPs and figure rendering. It is an implementation measurement, not a hardware-independent complexity comparison. Objective and Jacobian call counts accompany wall time in the raw records.

For independent Monte Carlo replications $V_1, \dots, V_R$, the reported mean and Monte Carlo standard error are

$$\bar{V} = R^{-1} \sum_b V_b, \quad \text{MCSE}(\bar{V}) = \sqrt{\frac{\sum_b (V_b - \bar{V})^2}{R(R-1)}}. \quad (9.1)$$



Serial errors within a trajectory are compatible with this calculation across independent replications. Figure bars are 1.96 MCSE, not parameter confidence intervals. The single January record and the experimental STAT5 time course receive no Monte Carlo error bars.

9.2. Affine contamination and clean efficiency. This empirical comparison measures clean-data efficiency and response to contamination under common pilot scales. It tests loss behavior, not Theorem 8.1, because the observation map is affine and has no ODE discretization error.

Each sample has $n = 160$, parameter dimension $p = 4$, generating vector $(1.25, -0.9, 0.55, -0.25)^T$ and Gaussian predictor covariance $\Sigma_{jk} = 0.25^{|j-k|}$. The six regimes have 100 replications each. Gaussian errors have standard deviation 0.5; Student errors are $0.5t_3/\sqrt{3}$; stationary AR(1) errors have coefficient 0.65 and marginal standard deviation 0.5. Vertical contamination adds random signs times 10 to ten percent of responses. The leverage regime replaces eight predictor rows by $N(0, 64I_4)$ and adds random signs times 16 to their errors. The asymmetric regime adds exponential errors of mean 10 to ten percent of responses. Locations are sampled without replacement.

LAD [28] is solved by linear programming. Its residual MAD, multiplied by 1.4826022185 and floored at 10^{-4} , is shared by all standardized fits. Least squares, LAD, Huber (1.345), GLDM and inverse-power GLDM ($q = 2$) are compared. Both arctangent losses use $c = 1.5$ and the same LAD start. The numerical parameter box is $[-10, 10]^4$. The clean-design forecast RMSE is $[(\hat{\beta} - \beta_*)^T \Sigma (\hat{\beta} - \beta_*)]^{1/2}$.

TABLE 10. Affine parameter error, mean and MCSE over 100 independent replications per regime. $q = 1$ is GLDM and $q = 2$ is inverse-power GLDM.

Regime	LS	LAD	Huber	$q = 1$	$q = 2$
Gaussian	0.0779	0.0930	0.0782	0.1128	0.1232
	(0.0033)	(0.0038)	(0.0034)	(0.0046)	(0.0050)
Student	0.0751	0.0656	0.0591	0.0729	0.0792
	(0.0032)	(0.0026)	(0.0023)	(0.0025)	(0.0027)
Serial	0.0792	0.0977	0.0822	0.1140	0.1235
	(0.0026)	(0.0037)	(0.0027)	(0.0044)	(0.0053)
Vertical	0.4948	0.1055	0.0929	0.1154	0.1235
	(0.0218)	(0.0040)	(0.0036)	(0.0046)	(0.0050)
Leverage	1.1205	0.2719	0.3459	0.1227	0.1485
	(0.0314)	(0.0124)	(0.0162)	(0.0048)	(0.0072)
Asymmetric	0.6225	0.1054	0.0944	0.1173	0.1259
	(0.0304)	(0.0044)	(0.0039)	(0.0050)	(0.0054)

Table 10 is a comparison of complete fixed procedures, including the pilot. Clean-data efficiency and contaminated-data behavior must be assessed together. The favorable performance of a bounded residual score in a particular leverage simulation is not a theorem of protection against arbitrary leverage; the regression score still multiplies the predictor vector.

9.3. Nonlinear state observations and an unknown initial state. This experiment examines local calibration from noisy states, including serial errors and an unknown initial condition. Its forecast comparisons are empirical observations. The separate oscillator mesh study in Section 9.7 tests transfer and sensitivity-error control.

The logistic models solve

$$\dot{x} = rx(1 - x/K), \quad x(0) = x_0, \quad x(t) = \frac{K}{1 + (K/x_0 - 1)e^{-rt}}. \quad (9.2)$$

The two-parameter version fixes $x_0 = 0.3$ and estimates $(r, K) = (0.7, 3)$ within $[0.2, 1.3] \times [1.5, 5]$. The three-parameter version also estimates $x_0 \in [0.1, 0.6]$, with generating value 0.3. This adds an unknown initial condition and a third unknown, but is still a modest nonstiff problem. The two-state oscillator in Eq. (A.2) estimates $(\kappa_d, \omega) = (0.18, 1.4)$ within $[0.04, 0.6] \times [0.7, 2.1]$ and observes its first state only.

All models have sixty state observations at $t_i = 0.1i$ and forty held-out forecast times from 6.1 to 10. Noise standard deviations are 0.045 for logistic growth and 0.035 for the oscillator. Each model has forty replications of Gaussian



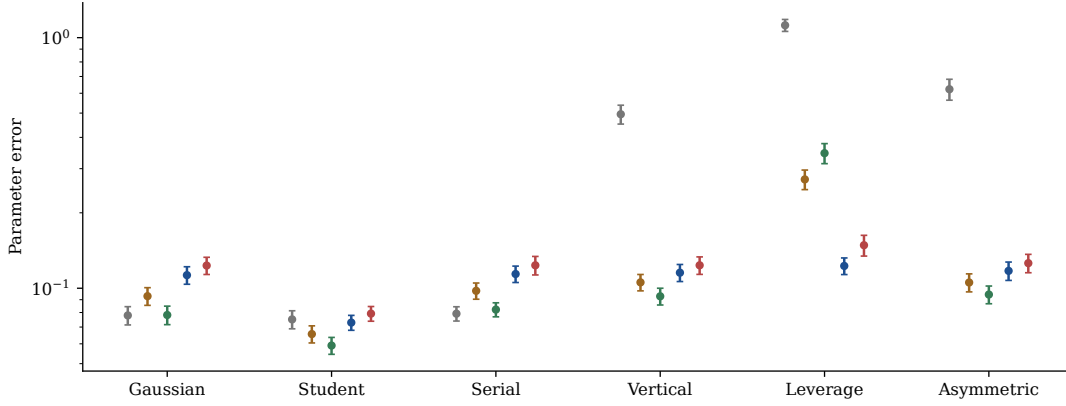


FIGURE 5. Affine parameter errors with 1.96 MCSE bars. Left-to-right points in each regime correspond to LS (gray), LAD (brown), Huber (green), GLDM (blue) and inverse-power GLDM (red).

errors, ten percent signed outliers of size 20σ , stationary AR(1) errors with coefficient 0.65, and AR(1) errors with those outliers. No noisy derivative observations are used. A common, training-only normalization is

$$s = \max\{10^{-4}, (1.4826022185/\sqrt{6}) \text{MAD}(\Delta^2 y)\}. \quad (9.3)$$

The known generating noise scale is not supplied to the estimators. Physical parameters are transformed to the unit box. The three starts are the center and alternating coordinates 0.25, 0.75 in opposite orders. Every method retains its smallest own objective among these starts.

Comparators are least squares, Huber, Cauchy (scale 2.385) and Tukey bisquare (cutoff 4.685), optimized by bounded L-BFGS-B with analytic/complex-step Jacobians, maximum 600 iterations, function tolerance 10^{-13} and gradient tolerance 10^{-8} . GLDM and inverse-power GLDM use the unsmoothed proximal weighted-LAD solver, maximum 120 accepted steps, step tolerance 2×10^{-6} , independently evaluated gap tolerance 2×10^{-7} , acceptance parameter 10^{-3} and minimum proximal coefficient 0.025. OSQP uses absolute/relative tolerances 10^{-9} with polishing; acceptance also checks the model decrease and gap. Nontermination is recorded rather than removed.

Logistic maps use the analytic solution. Oscillator fitting uses fixed-step RK4 with $h = 0.025$; its Jacobian differentiates the same discrete matrix polynomial. Complex-step differentiation avoids a finite-difference subtraction here because the matrix-power evaluation is analytic in the parameters. Analytic continuous trajectories generate the data and score primary forecasts. Therefore the primary RMSE isolates estimation error; the paired-mesh experiment below also includes numerical propagation error.

9.4. Controlled inverse-power contamination experiment. This empirical experiment asks whether changing q improves estimation or forecasting under a matched scale and zero penalty. It compares the full loss-shape change, including curvature and saturation, rather than isolating tail decay.

With the scale and penalty fixed, a separate grid crosses contamination fractions 0, 0.05, 0.10, 0.20, 0.30 with signed outlier magnitudes 5σ , 10σ , 20σ . The uncontaminated row is generated once, not counted three times. There are sixty affine and twenty three-parameter logistic replications at each of the thirteen conditions. Methods share their data, scales, constraints and starts; no regularization is used. All thirteen conditions and every fit are in the CSV records. Parameter error, forecast RMSE, vector bias and squared parameter error are stored. Bias means the norm of the mean parameter-error vector, not the mean norm.

Changing the exponent changes curvature, saturation and score-tail decay together; this experiment cannot assign the risk difference to any one of these properties. The clean row measures an explicit efficiency cost, and contamination results depend on magnitude, fraction, pilot scale and the selected nonconvex basin. The prescribed equal-scale comparison isolates the loss change; retuning each estimator for maximum Gaussian efficiency would answer a different question. Neither this experiment nor the influence calculation establishes finite-sample breakdown protection.



TABLE 11. Held-out state forecast RMSE, mean and MCSE over forty replications. Logistic (3) includes the unknown initial condition. All fits use three common starts; these are local procedures.

Model / regime	LS	Huber	Cauchy	Tukey	$q = 1$	$q = 2$
Logistic (2) / clean	0.0189	0.0201	0.0199	0.0202	0.0282	0.0319
	(0.0025)	(0.0024)	(0.0024)	(0.0024)	(0.0029)	(0.0039)
Logistic (2) / outlier	0.1209	0.0331	0.0275	0.0256	0.0341	0.0358
	(0.0124)	(0.0041)	(0.0040)	(0.0039)	(0.0043)	(0.0046)
Logistic (2) / serial	0.0463	0.0446	0.0448	0.0473	0.0526	0.0718
	(0.0050)	(0.0050)	(0.0050)	(0.0051)	(0.0057)	(0.0096)
Logistic (2) / both	0.1069	0.0454	0.0460	0.0578	0.0640	0.0738
	(0.0146)	(0.0041)	(0.0044)	(0.0109)	(0.0120)	(0.0133)
Oscillator / clean	0.0057	0.0059	0.0059	0.0059	0.0079	0.0119
	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0007)	(0.0016)
Oscillator / outlier	0.0329	0.0062	0.0054	0.0055	0.0081	0.0092
	(0.0024)	(0.0005)	(0.0004)	(0.0004)	(0.0007)	(0.0008)
Oscillator / serial	0.0102	0.0107	0.0108	0.0118	0.0133	0.0318
	(0.0008)	(0.0007)	(0.0007)	(0.0008)	(0.0011)	(0.0043)
Oscillator / both	0.0339	0.0100	0.0098	0.0098	0.0117	0.0195
	(0.0027)	(0.0008)	(0.0009)	(0.0009)	(0.0013)	(0.0029)
Logistic (3) / clean	0.0273	0.0285	0.0279	0.0287	0.0401	0.0443
	(0.0033)	(0.0034)	(0.0033)	(0.0034)	(0.0041)	(0.0048)
Logistic (3) / outlier	0.1776	0.0341	0.0296	0.0297	0.0493	0.0513
	(0.0207)	(0.0035)	(0.0034)	(0.0035)	(0.0055)	(0.0057)
Logistic (3) / serial	0.0543	0.0526	0.0527	0.0535	0.0577	0.0824
	(0.0054)	(0.0058)	(0.0058)	(0.0060)	(0.0070)	(0.0168)
Logistic (3) / both	0.1804	0.0536	0.0521	0.0528	0.0644	0.0663
	(0.0258)	(0.0074)	(0.0075)	(0.0077)	(0.0084)	(0.0082)

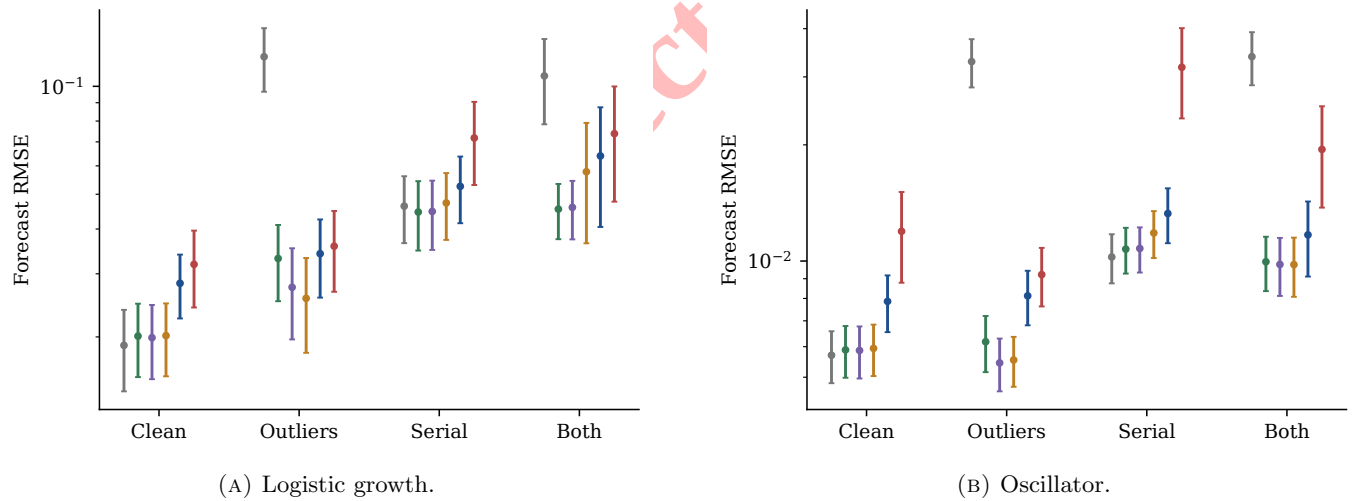


FIGURE 6. Logistic (left) and oscillator (right) forecast RMSE. Six points in each regime are LS, Huber, Cauchy, Tukey, GLDM and inverse-power GLDM, in gray, green, purple, orange, blue and red. Bars are 1.96 MCSE.

9.5. Matched comparison with Laplace correntropy. This empirical comparison asks whether the arctangent family differs predictively from another bounded cusp loss with the same weighted-LAD structure. No penalty is used, so the comparison does not mix a change in loss with a change in regularization.



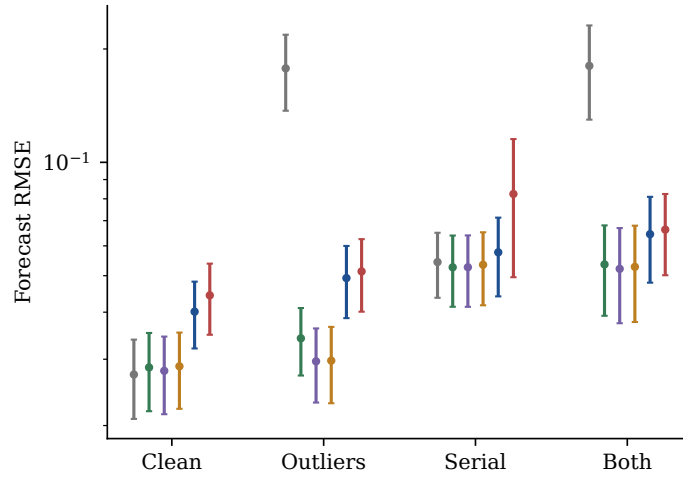


FIGURE 7. Three-parameter logistic calibration with an unknown initial condition. Colors and point order match Figure 6; each regime has forty replications.

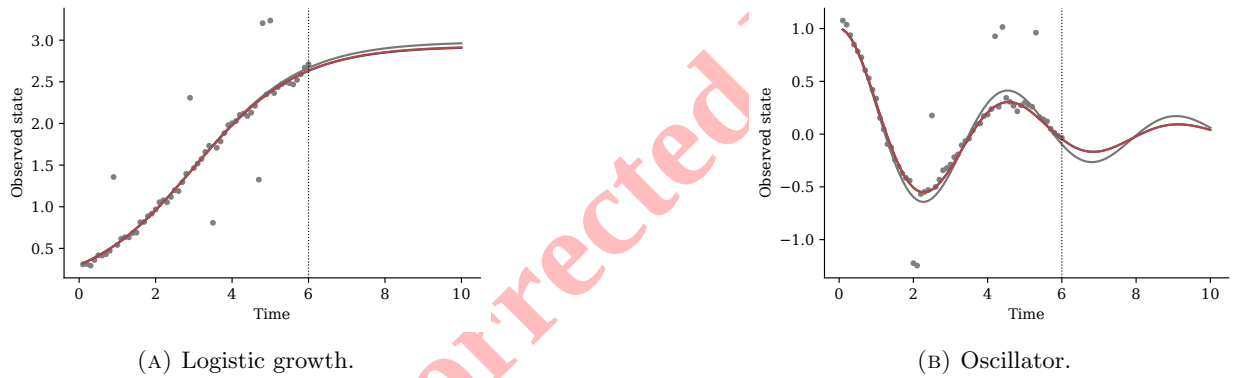


FIGURE 8. First predetermined serial-plus-outlier replication for logistic growth (left) and the oscillator (right). Gray points are noisy training states. Gray, blue and red curves are LS, GLDM and inverse-power GLDM. The dotted line separates fitting and forecasting.

The Laplace member of generalized correntropy [13] is a particularly close structural comparator. At the common scale $c = 1.5$, write

$$\ell_L(u) = 1 - e^{-u/c}, \quad \ell'_L(u) = c^{-1}e^{-u/c}, \quad \ell_L(v) \leq \ell_L(u) + \ell'_L(u)(v - u), \quad u, v \geq 0. \quad (9.4)$$

It shares the arctangent family's slope $1/c$ at the cusp and its weighted-LAD tangent structure, but has exponential rather than algebraic tail-score decay. This comparison tests a relevant alternative without presenting weighted-LAD majorization as unique to arctangent losses.

The matched study reuses 40 clean and 40 serial-outlier replications per model: 160 datasets, 480 selected fits and 1,440 starts. Analytic maps, frozen second-difference MAD scales, no penalty and three unit-box starts are common. Each method uses L-BFGS-B continuation at $10^{-2}, 10^{-4}, 10^{-6}$, up to 600 iterations per radius, selecting its final training objective. This compares complete local procedures separately from RK4/MM timing. Forecast RMSE uses the known noise-free trajectory; paired MCSE uses the 40 within-replication differences, not independent-method errors.



TABLE 12. Contamination grid at magnitude 20σ , with the shared clean row. Forecast RMSE is mean (MCSE); the last two columns are norms of vector bias. All magnitudes are shown in Figure 9 and the supplied records.

Model	ϵ	LS RMSE	$q = 1$ RMSE	$q = 2$ RMSE	$q = 1$ bias	$q = 2$ bias
Affine	0.00	0.0768 (0.0040)	0.1168 (0.0061)	0.1259 (0.0066)	0.0158	0.0162
Affine	0.05	0.3665 (0.0178)	0.1239 (0.0058)	0.1371 (0.0069)	0.0236	0.0253
Affine	0.10	0.5082 (0.0259)	0.1237 (0.0059)	0.1366 (0.0068)	0.0127	0.0175
Affine	0.20	0.6557 (0.0286)	0.1116 (0.0054)	0.1186 (0.0056)	0.0120	0.0160
Affine	0.30	0.8293 (0.0379)	0.1157 (0.0065)	0.1288 (0.0071)	0.0097	0.0111
Logistic (3)	0.00	0.0267 (0.0049)	0.0392 (0.0062)	0.0423 (0.0079)	0.0085	0.0092
Logistic (3)	0.05	0.1149 (0.0200)	0.0444 (0.0092)	0.0476 (0.0098)	0.0396	0.0532
Logistic (3)	0.10	0.2383 (0.0520)	0.0465 (0.0071)	0.0536 (0.0074)	0.0302	0.0288
Logistic (3)	0.20	0.3266 (0.0679)	0.0484 (0.0071)	0.0503 (0.0077)	0.0142	0.0151
Logistic (3)	0.30	0.2516 (0.0312)	0.0525 (0.0076)	0.0474 (0.0076)	0.0130	0.0159

TABLE 13. Clean-data parameter MSE. Relative efficiency is the MSE for $q = 1$ divided by that for $q = 2$. Values below one indicate an efficiency cost for the inverse power.

Model	LS MSE	$q = 1$ MSE	$q = 2$ MSE	Relative efficiency
affine	0.007484	0.017236	0.019872	0.867
logistic3	0.002662	0.004940	0.007825	0.631

Ordinary GLDM has the smallest mean RMSE in all four blocks. The smallest difference occurs between Laplace and ordinary GLDM for the contaminated oscillator, 0.000181 with paired MCSE 0.000254. The clean logistic comparison gives a larger inverse-power increase, 0.01226 with paired MCSE 0.00411. These results do not support a general ordering by tail decay: discarding more large residuals can remove useful trajectory information. This experiment supplements, rather than replaces, the Huber/Cauchy comparisons and the contamination grid. It does not compare every possible bandwidth or a separately tuned version of each loss.

9.6. Numerical stopping checks. This numerical verification checks the finite step, inner-gap and tube-residual targets. It assesses the optimizer output independently of predictive performance. Secondary wall-time and evaluation-count comparisons appear in Appendix E.9.

For $q = 1$, 2120 of 2120 selected primary-plus-stress fits satisfied both finite stopping tolerances. The largest final recorded gap was $1.19e - 08$ and the largest 10^{-4} -tube LP residual was $3.11e - 15$. For $q = 2$, 2120 of 2120 selected primary-plus-stress fits satisfied both finite stopping tolerances. The largest final recorded gap was $3.4e - 08$ and the largest 10^{-4} -tube LP residual was $2.22e - 15$. These are measured diagnostics, not a claim of exact arithmetic or global optimization. Raw start records include iteration limits and every success indicator. An LP residual must be interpreted with its stated response tube; it is not the unchanged-data Clarke residual.

9.7. Numerical certificate-transfer verification. This oscillator experiment addresses Theorem 8.1 directly: does mesh refinement reduce observation-map and sensitivity errors, and is tube enlargement needed to transfer a feasible witness? Its evidence level is numerical verification, not validated computation or an independent estimator-ranking experiment.



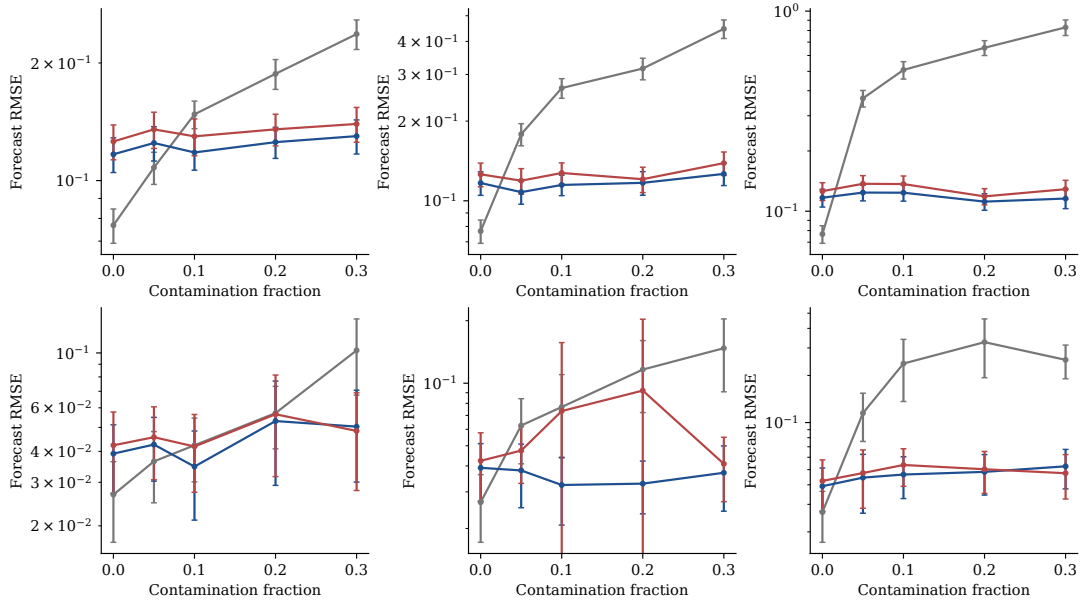


FIGURE 9. Contamination-strength study. Top: affine model, sixty replications per condition. Bottom: logistic model with unknown initial condition, twenty replications. Columns use magnitudes 5, 10, 20 noise standard deviations. LS is gray, GLDM blue and inverse-power GLDM red; bars are 1.96 MCSE.

TABLE 14. Matched arctangent-Laplace comparison: mean forecast RMSE (MCSE), and paired RMSE difference relative to ordinary GLDM. A positive difference favors ordinary GLDM. Each block contains 40 independent trajectory replications.

Method	RMSE (MCSE)	Paired difference	Paired MCSE
<i>Logistic, unknown initial state, clean</i>			
GLDM ($q = 1$)	0.03440 (0.00369)	+0.00000	0.00000
Inverse power ($q = 2$)	0.04665 (0.00505)	+0.01226	0.00411
Laplace correntropy	0.04160 (0.00428)	+0.00721	0.00294
<i>Logistic, unknown initial state, serial errors + outliers</i>			
GLDM ($q = 1$)	0.06035 (0.00823)	+0.00000	0.00000
Inverse power ($q = 2$)	0.07069 (0.00886)	+0.01034	0.00615
Laplace correntropy	0.06705 (0.00897)	+0.00670	0.00494
<i>Two-state oscillator, clean</i>			
GLDM ($q = 1$)	0.00819 (0.00074)	+0.00000	0.00000
Inverse power ($q = 2$)	0.01014 (0.00110)	+0.00195	0.00069
Laplace correntropy	0.00879 (0.00075)	+0.00059	0.00019
<i>Two-state oscillator, serial errors + outliers</i>			
GLDM ($q = 1$)	0.01068 (0.00091)	+0.00000	0.00000
Inverse power ($q = 2$)	0.01248 (0.00118)	+0.00180	0.00081
Laplace correntropy	0.01086 (0.00095)	+0.00018	0.00025

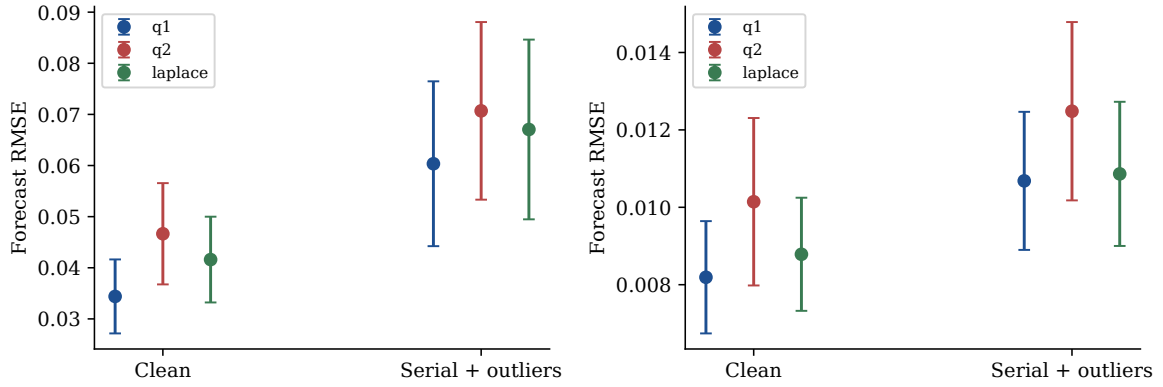


FIGURE 10. Matched loss comparison for the logistic model with unknown initial state (left) and two-state oscillator (right). Points show mean forecast RMSE; bars show 1.96 MCSE across 40 replications. Blue denotes ordinary GLDM, red inverse-power GLDM and green Laplace correntropy. The figure compares statistical error, not optimization certificate size.

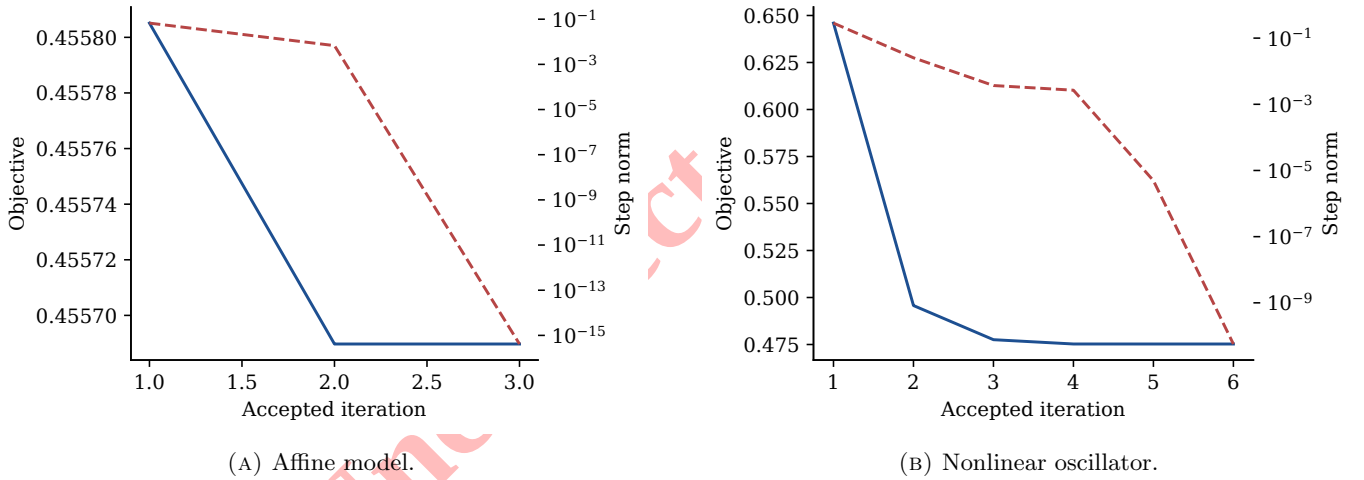


FIGURE 11. Predetermined $q = 1$ histories for the first affine leverage-contaminated dataset (left) and first clean oscillator dataset, first start (right). Blue: objective. Red dashed: step norm on the right logarithmic axis.

Twenty oscillator datasets with serial errors and outliers are fitted on four paired meshes $h = 0.1, 0.05, 0.025, 0.0125$ using GLDM ($q = 1$) and inverse-power GLDM ($q = 2$). The procedures share observations, frozen scales, $c = 1.5$, three starts and finite stopping tolerances. Each exponent supplies eighty selected fits, with no criterion penalty ($\lambda = 0$). Each fit is propagated by the same RK4 solver over the forecast interval. Analytic continuous solutions and sensitivities provide pointwise numerical references at the fitted parameters. These experiments evaluate the ODE solver and transfer machinery; they are separate from the analytically propagated January forecasting example.

Hats in this subsection denote floating-point evaluations. At each fitted point, define $\hat{e}_{i,h} = |\hat{g}_{i,h} - \hat{g}_{i,0}|$, $\hat{E}_{i,h} = \|\hat{J}_{i,h} - \hat{J}_{i,0}\|$, $\hat{Q}_i = \|\hat{J}_{i,0}\|$ and $\hat{e}_h = \max_i \hat{e}_{i,h}$. The common response tube is $\tau = 10^{-4}$; enlargement uses $\tau + \hat{e}_h$. For

replication b , compute every term of the transfer inequality and its signed slack:

$$\begin{aligned}\widehat{B}_{q,h}^{(b)} &= \sum_i a_i \left(\widehat{E}_{i,h}/c + \widehat{Q}_i L_q \widehat{e}_{i,h} \right), \\ S_{q,h}^{(b)} &= \widehat{R}_{\tau,h}^{q,0}(\widehat{\theta}_{q,h}^{(b)}) + \widehat{B}_{q,h}^{(b)} - \widehat{R}_{\tau+\widehat{e}_h,0}^{q,0}(\widehat{\theta}_{q,h}^{(b)}), \\ V_{q,h}^{(b)} &= \max\{0, -S_{q,h}^{(b)}\}.\end{aligned}\tag{9.5}$$

Nonnegative slack means that the numerically evaluated enlarged-tube inequality is satisfied. A positive $V_{q,h}^{(b)}$ records a numerical violation. The hats are essential: these discrepancies and LP evaluations are not outward-rounded enclosures of exact quantities.

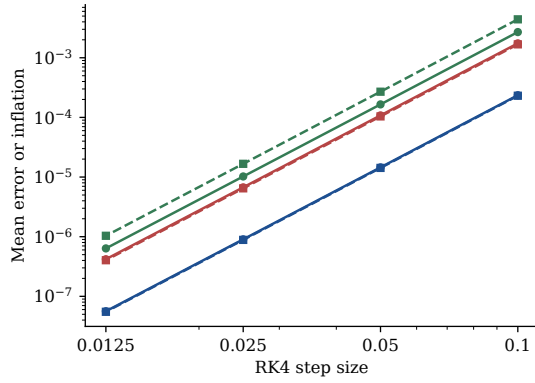
TABLE 15. Full numerical certificate-transfer verification over twenty fits per (q, h) , with $\lambda = 0$. Panel (a) reports means of all LP and inflation terms and the minimum fit-level slack $S_{\min} = \min_b S_{q,h}^{(b)}$. The middle continuous residual uses the unchanged tube; the next uses the enlarged tube. Panel (b) retains mean map error, Jacobian error and forecast RMSE. All 160 values of $V_{q,h}^{(b)}$ are zero. Values near 10^{-16} reflect LP roundoff, not validated precision. Unrounded records include Monte Carlo standard errors.

(a) Transfer terms and worst-fit slack						
q	h	$\widehat{R}_{\tau,h}^{q,0}$	$\widehat{B}_{q,h}$	$\widehat{R}_{\tau,0}^{q,0}$	$\widehat{R}_{\tau+\widehat{e}_h,0}^{q,0}$	$S_{\min}$
1	0.1	5.93e-16	0.0027	0.857	7.02e-16	0.00124
1	0.05	6.2e-16	0.000165	5.65e-16	5.65e-16	7.62e-05
1	0.025	4.91e-16	1.02e-05	4.83e-16	4.83e-16	4.72e-06
1	0.0125	4.76e-16	6.36e-07	6.29e-16	6.29e-16	2.94e-07
2	0.1	4.14e-16	0.00441	0.651	4.94e-16	0.00206
2	0.05	3.8e-16	0.00027	3.16e-16	3.16e-16	0.000126
2	0.025	4.25e-16	1.67e-05	4.63e-16	4.63e-16	7.79e-06
2	0.0125	3.48e-16	1.04e-06	4.88e-16	4.88e-16	4.85e-07

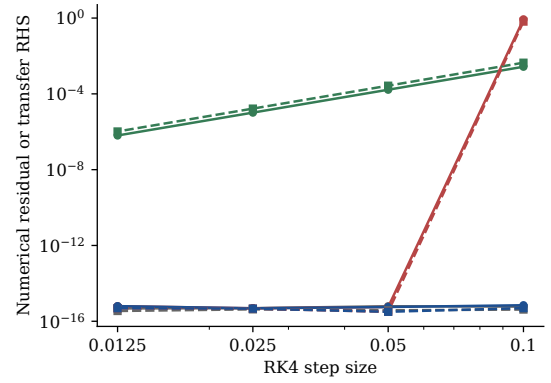
(b) Numerical reference errors and forecasting				
q	h	$\widehat{e}_h$	$\max_i \widehat{E}_{i,h}$	RMSE
1	0.1	0.000235	0.00174	0.0109
1	0.05	1.45e-05	0.000108	0.0109
1	0.025	9.03e-07	6.72e-06	0.0109
1	0.0125	5.63e-08	4.19e-07	0.0109
2	0.1	0.000231	0.00168	0.0141
2	0.05	1.42e-05	0.000104	0.0141
2	0.025	8.84e-07	6.47e-06	0.0141
2	0.0125	5.51e-08	4.03e-07	0.0141

The audit independently reevaluates both observation maps, both Jacobians, all three LP residuals and the inflation at all 160 recorded fits, without refitting. The largest absolute difference from the stored diagnostics is 1.12×10^{-14} . Across the eighty $q = 1$ fits, the minimum slack is 2.9375218×10^{-7} ; across the eighty $q = 2$ fits it is 4.8459678×10^{-7} . Across all fits, the largest signed excess (left side minus right side) is $-2.9375218 \times 10^{-7}$, and the maximum positive violation is zero. Table 15 therefore exposes the complete inequality and its worst-fit margin, rather than relying on an inequality between means.





(A) Map and sensitivity errors.



(B) Stationarity-certificate transfer.

FIGURE 12. Paired GLDM and inverse-power mesh results. Solid curves with circles use $q = 1$; dashed curves with squares use $q = 2$. Left: measured state error (blue), Jacobian error (red) and computed inflation (green). Right, on logarithmic axes: the discrete LP residual (gray), transfer right side $\hat{R}_{\tau,h}^{q,0} + \hat{B}_{q,h}$ (green), unchanged continuous residual (red) and enlarged continuous residual (blue). Curves show means; Table 15 separately reports minimum fit-level slack. Near-machine-zero LP values are numerical diagnostics.

All eighty $q = 2$ fits meet step/gap targets (maxima 1.85×10^{-6} and 9.45×10^{-15}). At $h = 0.1$, omitting enlargement violates the numerical transfer inequality in all twenty fits for each exponent. For $q = 2$, mean unchanged-tube residual, inflation and discrete residual are 0.651, 0.00441 and 4.14×10^{-16} ; enlargement reduces the continuous residual to 4.94×10^{-16} . This verifies predicted pointwise behavior numerically, not rigorously or uniformly over the parameter box.

Across eighty fits and forty forecast times, maximum error/envelope ratios are 0.0877 ($q = 1$) and 0.0878 ($q = 2$). The envelope uses the known generating parameter and is not a real-data uncertainty interval. Mean RMSE remains about 0.0109 and 0.0141, respectively, on every mesh. Transfer verification therefore implies no predictive advantage; forecast accuracy alone also cannot verify the state, sensitivity and tube terms.

The paired experiment records all 240 inverse-power starts, 80 selected fits and 3200 forecast points. It checks that the regenerated observations equal the recorded data and verifies two repeated $q = 1$ fits against their saved parameters. The run uses the same software versions and one-thread execution described in Section 9; its environment, timings, standard errors and full stopping records accompany the results.

9.8. A rigorously enclosed scalar transfer test. The scalar benchmark provides validated computation: rigorous pointwise verification of the transfer inequality using exact-rational enclosures. The chosen closed-form ODE permits every arithmetic enclosure to be checked independently. It does not validate a general nonlinear or adaptive integrator.

Consider the scalar initial-value problem

$$\dot{x} = -\theta x, \quad x(0) = 1, \quad \Theta = [1/10, 3/2], \quad t_i = i/2 \quad (i = 1, \dots, 10). \quad (9.6)$$

The data, scale and evaluation points are rational. The fixed settings are $s = 1/20$, $c = 3/2$, $a_i = 1/10$, $\tau = 10^{-4}$, $q = 1, 2$ and $\lambda = 0$. The measured values are rounded six-decimal values of $e^{-0.7t_i}$ plus the fixed perturbations listed in Section 9.8.1. A floating-point grid search and local minimization only propose a parameter; rounding it to twelve decimal places defines the exact point subsequently verified. Transfer is evaluated at each resulting rational point.



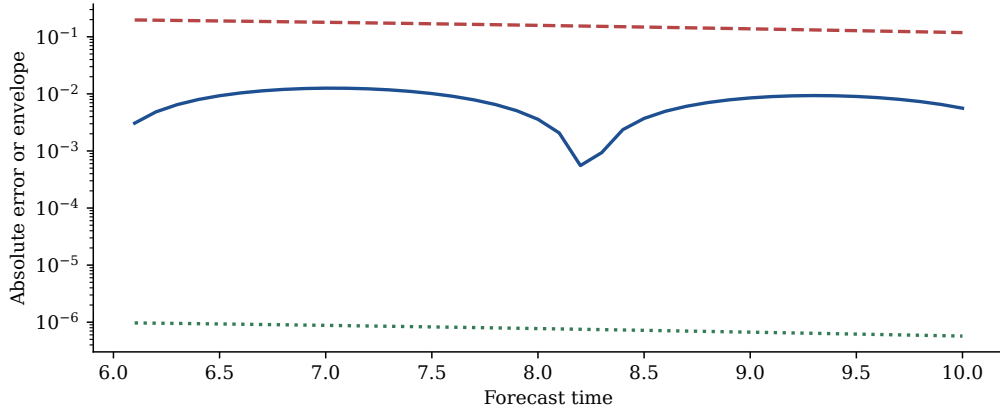


FIGURE 13. First predetermined paired-mesh dataset for $q = 1$ at $h = 0.025$. Blue: actual forecast error. Red dashed: combined envelope. Green dotted: its RK4 contribution.

Proposition 9.1 (Exact scalar observation and stationarity enclosures). *Let θ, c, s, h be positive rationals with θ in the interior of Θ , $n_i = t_i/h \in \mathbb{N}$, rational responses, positive rational weights, and nonnegative rational tube widths. For $a = \theta t_i$ and an integer $M \geq 0$ satisfying $a < M + 2$, define*

$$S_M(a) = \sum_{k=0}^M \frac{a^k}{k!}, \quad T_M(a) = \frac{a^{M+1}}{(M+1)!} \frac{1}{1 - a/(M+2)}. \quad (9.7)$$

Then $[1/(S_M + T_M), 1/S_M]$ encloses the exact state $x(t_i; \theta)$. The RK4 state and its derivative are the rational quantities

$$p(z) = 1 + z + z^2/2 + z^3/6 + z^4/24, \quad z = -\theta h, \\ x_{i,h} = p(z)^{n_i}, \quad \partial_\theta x_{i,h} = -n_i h p(z)^{n_i-1} p'(z). \quad (9.8)$$

Scaling by s gives enclosures of $g_{i,0}$, $J_{i,0} = -t_i g_{i,0}$, and their differences from $g_{i,h}$, $J_{i,h}$. For either map, at $q = 1, 2$, if all interval tests resolve membership in the tube and the signs outside it, the scalar stationarity LP has the exact value

$$R_\tau^{q,0} = \max\{0, |C| - D\}, \quad C = - \sum_{|r_i| > \tau} a_i J_i \operatorname{sgn}(r_i) w_q(|r_i|), \quad D = \sum_{|r_i| \leq \tau} a_i |J_i|/c. \quad (9.9)$$

Let $[R_h, \bar{R}_h]$ and $[\underline{R}_0, \bar{R}_0]$ enclose the discrete and enlarged-tube continuous residuals. If $B_{q,h}^+$ is computed with validated map/Jacobian bounds, upper bounds for $\|J_{i,0}\|$ and an upper bound for L_q , then

$$S^- := \underline{R}_h + B_{q,h}^+ - \bar{R}_0 \geq 0 \implies R_{\tau+e_h^{\max},0}^{q,0} \leq R_{\tau,h}^{q,0} + B_{q,h}^+. \quad (9.10)$$

Here the scalar radius increment is $e_h^{\max} = \max_i e_{i,h}$, applied to every component; it is at least the componentwise increment required by Theorem 8.1. All assertions are pointwise at the given rational parameter.

Proof. The exponential tail after degree M is positive. Its successive ratios are at most $a/(M+2) < 1$, so a geometric series bounds it by T_M . Inversion of the two positive endpoints proves Eq. (9.7). Applying classical RK4 to a scalar linear equation gives the stability polynomial p , and differentiation gives Eq. (9.8). At an interior point the normal cone is zero. Each tube residual contributes the interval $[-a_i |J_i|/c, a_i |J_i|/c]$, and their sum is $[-D, D]$. The distance of $-C$ to that interval is Eq. (9.9). Inclusion-preserving interval arithmetic then encloses the two distances. Finally $R_{\tau,h}^{q,0} \geq \underline{R}_h$ and $R_{\tau+e_h^{\max},0}^{q,0} \leq \bar{R}_0$ prove the implication in Eq. (9.10). Theorem 8.1 independently gives the same transfer statement from the validated map/Jacobian bounds. $\square$



The implementation uses $M = 100$, 60-term alternating arctangent series after rational argument reduction to magnitude at most $1/2$, and the identity $\pi = 16 \arctan(1/5) - 4 \arctan(1/239)$. Integer numerator/denominator arithmetic and outward rounding to a grid of spacing 10^{-32} are used throughout the validation. An explicit rational lower bound for $\sqrt{3}$ and the lower enclosure endpoint for π give an upper bound for L_q . The construction encloses all transcendental terms; agreement between two floating-point solvers is not used as an error bound. Algorithm 6 states the verification steps, and Section 9.8.1 records the inputs.

Algorithm 6 Rational scalar certificate-transfer verification.

Require: Rational $\theta, h, t_i, y_i, s, c, \tau, a_i$; $q \in \{1, 2\}$; $\theta \in \text{int } \Theta$.

- 1: Evaluate rational RK4 states and derivatives using Eq. (9.8).
- 2: Enclose $e^{-\theta t_i}$ using Eq. (9.7); scale states and derivatives consistently.
- 3: Bound $e_{i,h}$, $E_{i,h}$ and $|J_{i,0}|$ by interval subtraction and absolute values.
- 4: Enclose π , the arctangent terms and L_q ; round every interval operation outward.
- 5: Resolve all residual signs and tube memberships; if unresolved, refine the enclosures or return “undetermined.”
- 6: Enclose the discrete and enlarged continuous LP distances using Eq. (9.9).
- 7: Form $B_{q,h}^+$ and S^- in Eq. (9.10).

Ensure: A verified transfer inequality if $S^- \geq 0$; otherwise an undetermined result.

TABLE 16. Rigorously enclosed scalar transfer tests. “Fit” denotes the rational rounded fit; $+10^{-3}$ denotes that point plus an exact rational perturbation. The table displays a lower enclosure endpoint for the discrete residual, an upper endpoint for the enlarged continuous residual, the validated inflation and the lower slack. Displayed decimals are rounded for readability; the sign tests use the exact rational endpoints.

q	h	Point	$\underline{R}_h$	$\overline{R}_0$	$B_{q,h}^+$	S^-
1	0.500	fit	0.000000	0.000000	4.683e-03	4.683e-03
1	0.500	$+10^{-3}$	0.575836	0.575765	4.700e-03	4.771e-03
1	0.250	fit	0.000000	0.000000	2.471e-04	2.471e-04
1	0.250	$+10^{-3}$	0.576291	0.576287	2.479e-04	2.524e-04
1	0.125	fit	0.000000	0.000000	1.419e-05	1.419e-05
1	0.125	$+10^{-3}$	0.576315	0.576315	1.424e-05	1.452e-05
2	0.500	fit	0.000000	0.000000	6.513e-03	6.513e-03
2	0.500	$+10^{-3}$	0.627082	0.625788	6.536e-03	7.831e-03
2	0.250	fit	0.000000	0.000000	3.460e-04	3.460e-04
2	0.250	$+10^{-3}$	0.627570	0.627499	3.472e-04	4.179e-04
2	0.125	fit	0.000000	0.000000	1.995e-05	1.995e-05
2	0.125	$+10^{-3}$	0.627595	0.627591	2.001e-05	2.414e-05

All twelve checks have strictly positive certified slack; the minimum exceeds 1.4193×10^{-5} . At the six fitted points both tube residuals enclose exactly zero. The perturbed points have nonzero residuals, so the transfer check is not limited to zero-versus-zero comparisons. At $h = 1/2$, retaining the original continuous tube instead would give residuals greater than 0.5808 for $q = 1$ and 0.6423 for $q = 2$ at the fitted points, far exceeding the respective inflations. These are rigorous pointwise examples of the need for enlargement. They certify neither the optimizer’s global optimality nor every parameter in the box. The construction exploits the closed-form scalar solution and does not supply validated integration for STAT5 or the general nonlinear systems.



9.8.1. *Exact inputs and enclosure arithmetic for the scalar benchmark.* The ten observations in Eq. (9.6) are the exact decimal numbers

$$(y_i)_{i=1}^{10} = (0.712688, 0.484585, 0.354938, 0.546597, 0.166774, \\ 0.133456, -0.163706, 0.064810, 0.039852, 0.180197). \quad (9.11)$$

They equal six-decimal rounded exponential states plus the rational perturbations

$$(8, -12, 5, 300, -7, 11, -250, 4, -3, 150)/1000. \quad (9.12)$$

The rounded fit points, in decreasing order of step size $h = 1/2, 1/4, 1/8$, are

$$q = 1 : \quad 0.690650694331, \quad 0.690547336069, \quad 0.690541775274, \\ q = 2 : \quad 0.690650691741, \quad 0.690547345793, \quad 0.690541774158. \quad (9.13)$$

Each decimal in Eqs. (9.11)–(9.13) denotes an exact rational number, and each perturbed check adds exactly $1/1000$. These data specify the verification independently of the floating-point fitting procedure.

For $0 \leq x \leq 1/2$, the 60-term alternating series for $\arctan x$ and that partial sum plus the next term bracket the true value. Larger nonnegative arguments use $\arctan x = \pi/2 - \arctan(1/x)$ when $x > 1$ and

$$\arctan x = \arctan(1/2) + \arctan\left(\frac{x - 1/2}{1 + x/2}\right) \quad (1/2 < x \leq 1). \quad (9.14)$$

Monotonicity encloses interval arguments. The rational number $1.7320508075688772935274463415058$ has square less than 3 and is used as a lower bound for $\sqrt{3}$. Every lower endpoint is rounded down and every upper endpoint up to the grid $10^{-32}\mathbb{Z}$ using integer arithmetic; no binary floating-point operation is used in these enclosure decisions. The maximum state-map and Jacobian bounds at fitted points decrease, respectively, from 1.164×10^{-3} and 7.296×10^{-3} at $h = 1/2$ to 3.656×10^{-6} and 2.184×10^{-5} at $h = 1/8$, for both exponents. Table 16 reports the full stationarity inequality, rather than only these map bounds.

9.9. Energy stability of the fitted oscillator. This numerical verification asks whether stable forward propagation is sufficient for nonsmooth stationarity transfer. Energy and spectral factors test propagation stability; the separate tube calculation tests stationarity.

Proposition A.1 separates dissipation of the mechanical system from stability of its numerical propagation. Both factors in Eq. (A.6) are evaluated for all 160 fitted maps in the paired $q = 1, 2$ mesh experiment, with the data, parameters and selected starting points unchanged. All 160 maps satisfy $\gamma_\theta(h) < 1$ and $\varrho_\theta(h) < 1$. Their largest energy and spectral factors are 0.9999997005 and 0.9971053490, respectively. These are pointwise floating-point checks at the recorded fits, not a uniform proof over the whole parameter box.

The reference calculation extends free decay to time 20. At $h = 0.1$, the spectral radius is 0.975115 and the energy ratio is 3.5542×10^{-5} at the final time. For the deliberately excessive step $h = 2.5$, the spectral radius becomes 3.099155, and the final energy ratio is 8.3532×10^7 . The exact unforced system cannot produce such energy growth. It is an integration artifact, not evidence of negative damping or external excitation. This example also explains why an estimator with a bounded residual loss cannot repair an unstable forward solver. Conversely, every fitted map is energy-contracting on the calibration meshes, yet the coarse-mesh stationarity-transfer failures in Section 9.7 still occur without tube enlargement. Stable propagation alone does not certify nonsmooth stationarity.

9.10. Tuning and numerical tolerances. This study separates empirical sensitivity to the loss scale from numerical sensitivity to the inner-solver tolerance. The scale changes the criterion; the requested QP tolerance controls only an approximate subproblem solution.

Twenty separate logistic datasets with serial errors and outliers cross $c = 1, 1.5, 2.5$ with requested QP tolerances $10^{-6}, 10^{-8}, 10^{-10}$, from a common center start. These 180 runs do not change the primary tuning. The independently evaluated gap target remains 2×10^{-7} ; a requested QP tolerance is not a substitute for that check. Records retain runs that fail the finite stopping conditions.



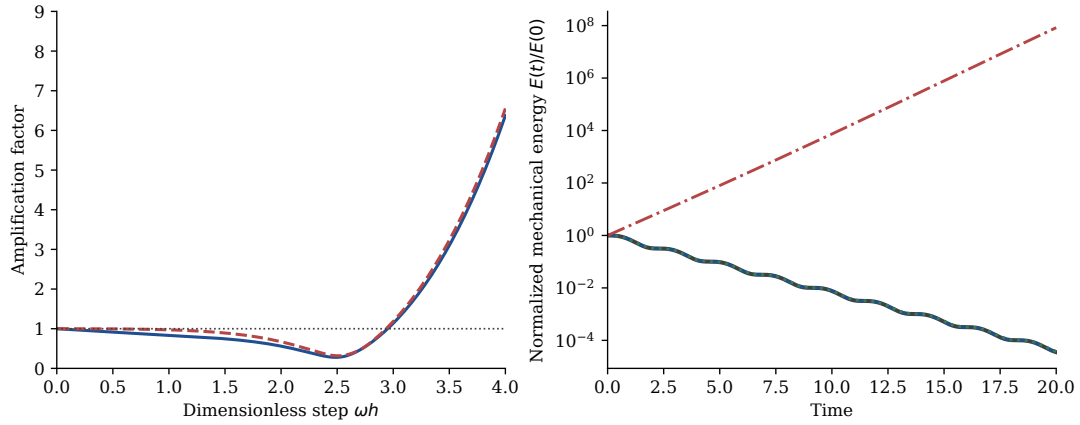


FIGURE 14. Mechanical and numerical stability at $\kappa_d = 0.18$, $\omega = 1.4$. Left: RK4 spectral factor ϱ_θ (blue solid) and energy-norm factor $\sqrt{\gamma_\theta}$ (red dashed); the dotted threshold is one. Right: exact normalized energy (black) and RK4 energy at $h = 0.1$ (blue dashed), $h = 0.5$ (green dotted) and $h = 2.5$ (red dash-dotted). The two fine-mesh curves nearly overlap the reference. The intentionally excessive step is not used for parameter estimation.

TABLE 17. Separate logistic tuning at requested QP tolerance 10^{-8} . Error is mean (MCSE); the final column counts runs satisfying both stopping conditions out of twenty.

c	Relative parameter error	Mean steps	Stopped
1.0	0.0390	7.10	20
	(0.0062)		
1.5	0.0317	6.15	20
	(0.0055)		
2.5	0.0301	5.25	20
	(0.0051)		

Standardize first, choose a short candidate grid using earlier held-out time blocks, and freeze c and the scale within each fit. A large c makes GLDM locally LAD-like, not least-squares-like. For a stationary AR(1) error with variance σ^2 and coefficient ϱ ,

$$\text{Var}(\Delta^2 e) = \sigma^2(6 - 8\varrho + 2\varrho^2). \quad (9.15)$$

At $\varrho = 0.65$, the division by $\sqrt{6}$ in the pilot gives about 0.524σ for the noise component. The trajectory's deterministic second difference also enters. This is a common normalization rule, not a universally consistent marginal-noise estimator. The original cusp is retained by PMM. The separately identified continuation studies below use $\delta = 10^{-6}$ and report their smoothing and numerical-stationarity limitations explicitly.

9.11. Experimental STAT5 calibration and chronological prediction. The STAT5 application is an optimization-sensitivity demonstration on experimental observations. Its central question is whether broader local search and a lower training objective improve chronological prediction. The original and expanded protocols are reported together in Table 23; prediction errors are empirical observations, while solver comparisons are numerical verification.

Experimental STAT5 measurements [8] provide a nonlinear ODE application distinct from the synthetic systems and the January polynomial forecasts. The public PTab benchmark representation [38, 44] contains 48 experimental measurements of three observables at 16 times between 0 and 240 minutes. The system describes STAT5A/STAT5B



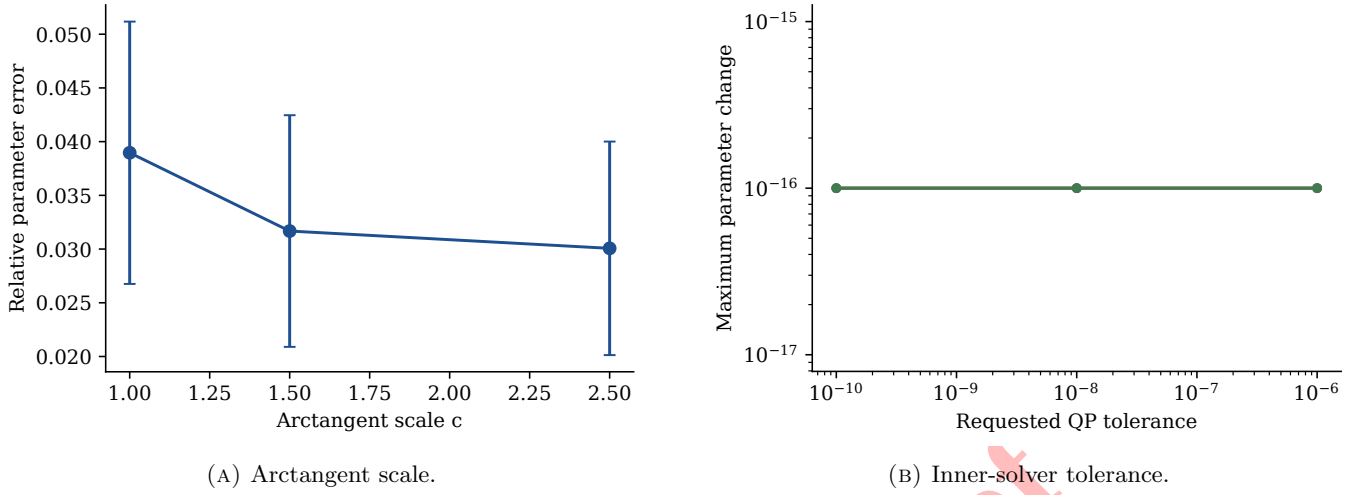


FIGURE 15. Left: scale sensitivity with 1.96 MCSE bars. Right: largest parameter difference from the 10^{-10} QP-tolerance fit across paired datasets, for $c = 1, 1.5, 2.5$ in blue, orange and green. Plotting zeros are floored at 10^{-16} .

phosphorylation, dimerization and nuclear transport in erythropoietin-stimulated BaF3 cells. The experimental measurement table is the calibration input. The separate simulated benchmark table is used only to check the implementation. No synthetic outliers are added to the experimental data.

All concentrations are normalized by the common initial total 207.6. The cytoplasmic states are a, b, d_{AB}, d_{AA} and d_{BB} ; the nuclear states are n_{AA}, n_{AB} and n_{BB} . Let $\eta = V_n/V_c = 0.45/1.4$, let $k_d, k_{xh}, k_{xt}, k_{ih}, k_{it}$ and k_p be positive degradation, export, import and phosphorylation rates, and set

$$\mathcal{E}(t) = 1.25 \times 10^{-7} e^{-k_d t}, \quad v_{AA} = 207.6 k_p \mathcal{E}(t) a^2, \quad v_{AB} = 207.6 k_p \mathcal{E}(t) ab, \quad v_{BB} = 207.6 k_p \mathcal{E}(t) b^2. \quad (9.16)$$

In these coordinates the supplied mass-action model is

$$\begin{aligned} \dot{a} &= -2v_{AA} - v_{AB} + 2\eta k_{xh} n_{AA} + \eta k_{xt} n_{AB}, \\ \dot{b} &= -v_{AB} - 2v_{BB} + \eta k_{xt} n_{AB} + 2\eta k_{xh} n_{BB}, \\ \dot{d}_{AB} &= v_{AB} - k_{it} d_{AB}, & \dot{d}_{AA} &= v_{AA} - k_{ih} d_{AA}, \\ \dot{d}_{BB} &= v_{BB} - k_{ih} d_{BB}, & \dot{n}_{AA} &= \eta^{-1} k_{ih} d_{AA} - k_{xh} n_{AA}, \\ \dot{n}_{AB} &= \eta^{-1} k_{it} d_{AB} - k_{xt} n_{AB}, & \dot{n}_{BB} &= \eta^{-1} k_{ih} d_{BB} - k_{xh} n_{BB}. \end{aligned} \quad (9.17)$$

Initial values are $a(0) = r_0 = 0.693$, $b(0) = 1 - r_0$ and zero for all dimers. With the benchmark specificity constant $s_0 = 0.107$, define $A = s_0(a + 2d_{AA}) + d_{AB}$ and $B = (1 - s_0)(b + 2d_{BB}) + d_{AB}$. The three measured percentages are

$$h_A = 100 \frac{d_{AB} + 2s_0 d_{AA}}{A}, \quad h_B = 100 \frac{d_{AB} + 2(1 - s_0) d_{BB}}{B}, \quad h_R = 100 \frac{A}{A + B}. \quad (9.18)$$

This is an eight-state nonlinear system with nonlinear observation maps. The estimated coordinates are the base-ten logarithms of the six rates, each restricted to $[-5, 5]$ in the main comparison, as in the benchmark rate bounds. A separate forecast sensitivity check widens all six intervals to $[-6, 6]$. The ratio, specificity, compartment volumes and initial total remain fixed. The benchmark also contains three noise parameters; here they are replaced by fixed training-based residual scales, so this is a stated robust reanalysis rather than an exact replication of its Gaussian likelihood fit.



Proposition 9.2 (Positivity and conservation of the STAT5 system). *Fix a finite time interval, positive finite rate bounds, $0 < r_0 < 1$, $0 < s_0 < 1$ and $\eta > 0$. For the stated initial condition, Eq. (9.17) has a unique nonnegative solution throughout the interval and preserves*

$$\begin{aligned} a + d_{AB} + 2d_{AA} + \eta(n_{AB} + 2n_{AA}) &= r_0, \\ b + d_{AB} + 2d_{BB} + \eta(n_{AB} + 2n_{BB}) &= 1 - r_0. \end{aligned} \quad (9.19)$$

The denominators in Eq. (9.18) are positive. On a compact rate box, the finite-time observation map is twice continuously differentiable with respect to the log-rate parameters and has a Lipschitz Jacobian.

Proof. At each zero concentration, all negative reaction terms containing that concentration vanish and all remaining terms are nonnegative. Local uniqueness for the smooth vector field therefore makes the nonnegative orthant invariant. Differentiating the two left-hand sides in Eq. (9.19) cancels each reaction flux and proves the identities. They bound all eight concentrations and preclude finite-time escape. On the resulting bounded set, $\dot{a} \geq -Ca$ and $\dot{b} \geq -Cb$ for a finite constant C ; hence their positive initial values remain positive at finite times. It follows that $A \geq s_0 a > 0$ and $B \geq (1 - s_0)b > 0$. Smooth dependence of the ODE solution on positive rates, followed by the logarithmic parameterization and the smooth observation functions, proves the final assertion. Compactness of the rate box and time interval makes the denominator minima positive. The smooth vector field and observation functions give continuous second parameter derivatives on a neighborhood of that box. Their boundedness on the compact box yields the Lipschitz Jacobian. These are well-posedness properties, not a parameter-identifiability proof. $\square$

Two protocols are fixed before fitting. The descriptive fit uses all 48 measurements. The forecast protocol uses the 36 measurements at times up to 100 minutes and withholds the 12 measurements at 120, 160, 200 and 240 minutes. A six-start least-squares pilot is fitted separately on each training set. For observable j , its residuals e_{ij} set the common scale

$$s_j = \max\{1, 1.4826022185 \text{ median}_i |e_{ij} - \text{median}_\ell e_{\ell j}|\} \text{ percentage points.} \quad (9.20)$$

All five methods then use these fixed scales, uniform observation weights, the same rate box, no penalty and seven starts: six Latin-hypercube points in the log-rate box with seed 2026092731, plus the training-only pilot. GLDM and inverse-power GLDM use $c = 1.5$ and $q = 1, 2$, respectively; Huber and Cauchy use the constants specified above. Neither the published nominal fitted rates nor the held-out responses set starts, scales or tuning parameters. The published nominal rates are used only for an implementation check against the simulated reference file.

LSODA integrates the states and forward sensitivities at relative tolerance 2×10^{-9} and absolute tolerance 2×10^{-11} . The nonlinear observation Jacobian follows by the chain rule. Each start uses L-BFGS-B with at most 600 iterations, function tolerance 10^{-12} and gradient tolerance 10^{-7} . For the arctangent criteria, continuation at $\delta = 10^{-2}, 10^{-4}, 10^{-6}$ initializes the two candidates with the smallest training objectives. Under this original seven-start protocol, each of the two candidates receives up to 240 unsmoothed proximal weighted-LAD steps; the lower unsmoothed training objective determines the final fit. This allowance is distinct from the expanded refinement budget in Section 9.12. The descriptive $q = 2$ fit receives one further unsmoothed MM step with QP absolute and relative tolerances 10^{-11} , and is accepted only after its independently evaluated tube residual at $\tau = 10^{-4}$ falls below 10^{-4} . All starts and stopping records are retained. These are finite local-search protocols.

Nominal simulated observables match the benchmark within 2.02×10^{-6} percentage points. Independent Radau central differences give maximum normalized sensitivity discrepancy 6.61×10^{-6} . Final maps are checked against Radau at relative/absolute tolerances 2×10^{-11} and 2×10^{-13} . These measured discrepancies are not validated bounds. Errors for this single biological time course are descriptive.

Table 18 reports the original protocol; Table 23 adds expanded search. Ordinary GLDM has forecast MAE 14.195 percentage points under both protocols. Inverse-power GLDM changes from 6.668 to 13.214 when a lower training objective is selected. Thus the original difference is search-dependent. Cauchy's original-protocol forecast MAE is 6.175. The descriptive fits use all observations and answer a different question from chronological prediction.

Figures 16 and 17 show that the GLDM fit retains excessive late phosphorylation. Held-out pSTAT5A MAE is 27.585, 13.925 and 13.446 percentage points for GLDM, inverse-power GLDM and Cauchy; all miss part of the late decline. Cytoplasmic-fraction errors are smaller: 1.825, 1.361 and 1.074. Aggregate improvement therefore chiefly



TABLE 18. STAT5 errors under the original seven-start protocol, in percentage points. Descriptive fits use all 48 observations. Forecast fits use 36 training observations and 12 later held-out observations. No Monte Carlo uncertainty is attached to this single experimental time course. Table 23(b) compares training-objective-selected forecasts under the original and expanded protocols. Ordinary GLDM has held-out MAE 14.195 under both searches. Inverse-power GLDM has MAE 6.668 under the original search and 13.214 under expanded selection (Table 23).

Method	Descriptive fit		Chronological forecast		
	MAE	RMSE	Train MAE	Test MAE	Test RMSE
Least squares	3.390	4.742	3.560	6.226	8.702
Huber	3.372	4.749	3.521	6.269	8.825
Cauchy	3.369	4.746	3.503	6.175	8.634
GLDM	4.311	8.748	4.874	14.195	19.440
Inverse-power GLDM	3.367	5.117	3.507	6.668	8.954

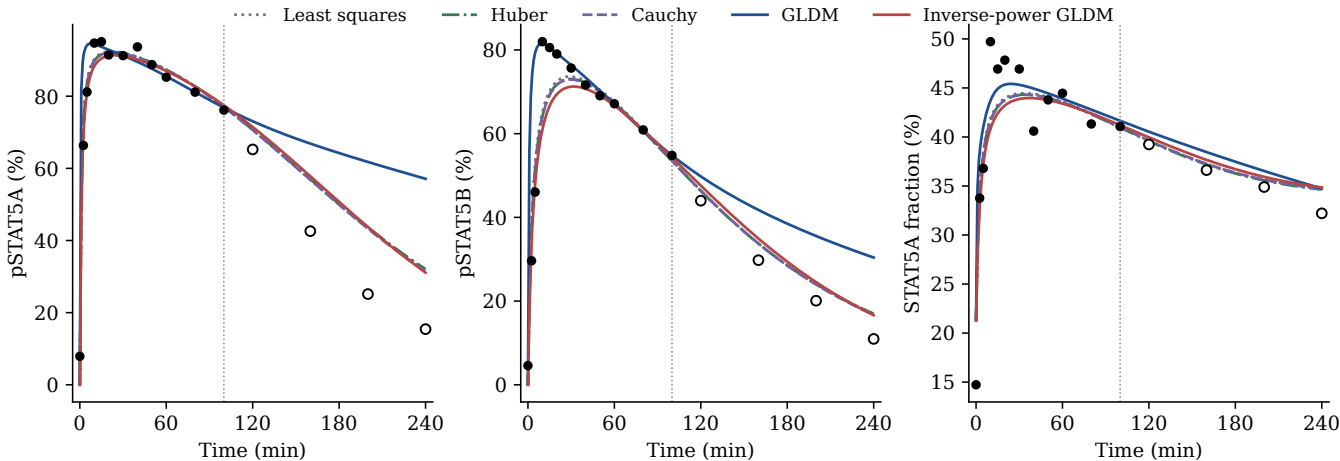


FIGURE 16. Chronological prediction of experimental STAT5 observables under the original seven-start protocol. From left to right: phosphorylated STAT5A, phosphorylated STAT5B and the cytoplasmic STAT5A fraction. Filled circles are training measurements through 100 minutes; open circles are the later held-out measurements. The vertical dotted line marks the training cutoff. Each curve uses the main $[-5, 5]$ log-rate box and is obtained by propagating the same fitted ODE without re-fitting after that cutoff. Expanded-search results are reported in Table 23 and Figure 20. Ordinary GLDM has held-out MAE 14.195 under both searches. Inverse-power GLDM has MAE 6.668 under the original search and 13.214 under expanded selection (Table 23).

concerns phosphorylation. These fits cannot distinguish noise from model inadequacy or identify the biological rate responsible.

Several rates reach the common search-box boundaries (Table 19); GLDM selects a different rate combination. The box and profile diagnostics below examine these weakly determined directions. Descriptive and forecast fits use the same time course.

All four arctangent fits meet both step/gap and independent tube targets (Table 20; $\tau = 10^{-4}$). The additional descriptive $q = 2$ MM step lowers the objective and tube residual to 9.53×10^{-7} , changing fitted observations by at most 3.84×10^{-7} percentage points; three-decimal MAE/RMSE are unchanged. Forecast residuals are 2.23×10^{-6}



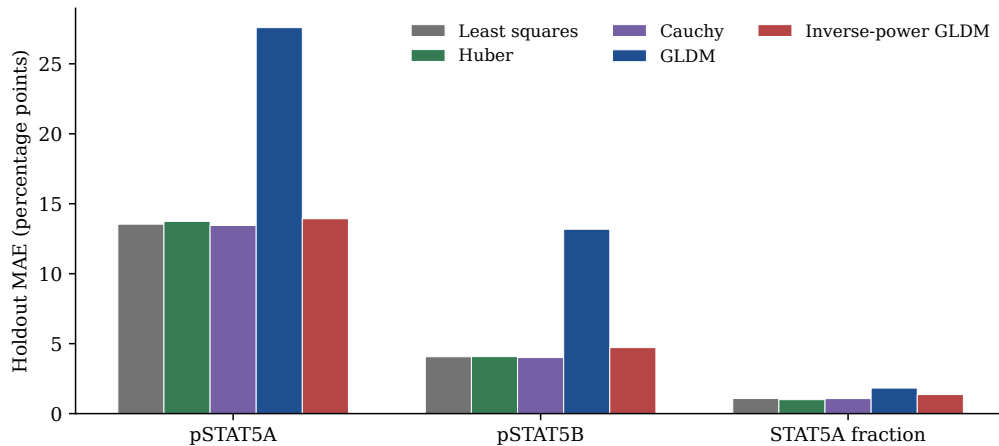


FIGURE 17. Held-out STAT5 mean absolute errors by observable under the original seven-start protocol. Each bar averages the four later measurement times for one observable. The observations form one experimental time course, so no Monte Carlo error bars or independent-replication significance claims are attached. See Table 23(b) for the expanded-search selection. Ordinary GLDM has held-out MAE 14.195 under both searches. Inverse-power GLDM has MAE 6.668 under the original search and 13.214 under expanded selection (Table 23).

TABLE 19. STAT5 forecast log-rate estimates under the original seven-start protocol. The logarithms refer to the numeric rate units of the supplied SBML model. An asterisk marks a value at a search-box boundary. These are local fitted values, not independently identified physical constants.

Rate	Least squares	Huber	Cauchy	GLDM	Inverse power
k_d	-1.827	-1.835	-1.821	-1.198	-1.764
k_{xt}	-5.000*	-5.000*	-5.000*	4.709	-5.000*
k_{xh}	-2.219	-2.227	-2.210	-1.545	-2.166
k_{it}	-1.707	-1.710	-1.718	-2.401	-1.787
k_{ih}	5.000*	5.000*	5.000*	-0.700	5.000*
k_p	4.157	4.135	4.132	4.612	4.065

TABLE 20. STAT5 numerical stopping diagnostics for the arctangent fits under the original seven-start protocol. The tube uses standardized responses, $\tau = 10^{-4}$ and log-rate coordinates. The last column is the maximum LSODA–Radau observation discrepancy in percentage points. These floating-point values are not rigorous error enclosures.

Protocol	q	Final step	Model gap	Tube residual	Solver difference
Descriptive	1	1.43×10^{-6}	1.64×10^{-15}	9.00×10^{-8}	3.27×10^{-8}
Descriptive	2	1.13×10^{-7}	7.69×10^{-13}	9.53×10^{-7}	3.73×10^{-7}
Forecast	1	2.50×10^{-7}	2.02×10^{-10}	2.23×10^{-6}	2.61×10^{-7}
Forecast	2	1.75×10^{-6}	3.19×10^{-14}	9.04×10^{-7}	3.95×10^{-7}

and 9.04×10^{-7} . Across ten fits, maximum LSODA–Radau and conservation discrepancies are 4.21×10^{-7} percentage points and 1.45×10^{-14} normalized units; evaluated concentrations are nonnegative. These pointwise diagnostics do not supply Theorem 8.1’s validated bounds or replace Section 9.7.



Parameter-box sensitivity is assessed by repeating the chronological forecast protocol with all six log-rates in $[-6, 6]$. The 36 training observations, 12 held-out observations, training-derived scales, loss constants and absence of a penalty are unchanged. The original seven common starts are retained, together with the previous fitted point and a copy with its previously active coordinates moved to the expanded limits. These two added starts coincide when the old fit is interior. The previous feasible solution is also retained as a candidate. Selection uses only the unsmoothed training objective; the held-out data are not used to select among starts. The same continuation and MM settings apply. Table 21 compares the resulting predictions on the observed times and on a one-minute grid, including the additional observation at 2.5 minutes.

TABLE 21. STAT5 forecast sensitivity to widening all six log-rate bounds from $[-5, 5]$ to $[-6, 6]$. The two displayed log-rates are from the wider-box fit; an asterisk denotes an expanded boundary. The subscripts on MAE denote the box half-width. $D_{\max}$ is the largest absolute change in a predicted observable over the 242 evaluated times in $[0, 240]$ minutes. Errors and trajectory changes are in percentage points.

Method	$\log_{10} k_{xt}$	$\log_{10} k_{ih}$	MAE ₅	MAE ₆	Δ MAE	$D_{\max}$
Least squares	-6.000*	6.000*	6.226	6.239	+0.0128	0.0393
Huber	-6.000*	6.000*	6.269	6.285	+0.0161	0.0396
Cauchy	-6.000*	6.000*	6.175	6.191	+0.0161	0.0392
GLDM	4.709	-0.700	14.195	14.195	-0.0001	0.0002
Inverse power	-6.000*	5.004	6.668	6.673	+0.0049	0.0326

Wider-box least squares, Huber and Cauchy retain active $\log_{10} k_{xt} = -6$ and $\log_{10} k_{ih} = 6$. Inverse-power GLDM retains k_{xt} 's lower bound but has interior $\log_{10} k_{ih} = 5.004$; GLDM has no active expanded bound, and no other coordinate is active. Maximum MAE and trajectory changes are below 0.0162 and 0.0396 percentage points, small against errors of 6–14; rankings are unchanged. Both arctangent tube targets still pass. This supports prediction stability for this expansion, not rate identifiability or a rate-uncertainty assessment.

At each wider-box fit, $D \in \mathbb{R}^{36 \times 6}$ is the training Jacobian in base-ten log-rate coordinates, divided rowwise by frozen observable scales, without loss weights. Table 22 reports its extreme singular values. Their ratios, 2.62×10^6 – 9.34×10^7 , indicate weak observational sensitivity in some directions. This is local ill-conditioning, not structural nonidentifiability or a covariance estimate.

TABLE 22. Standardized training-sensitivity spectra for the wider-box STAT5 forecast fits. Derivatives are with respect to the six base-ten log-rates. The condition number is $\sigma_{\max}(D)/\sigma_{\min}(D)$, without a rank truncation.

Method	$\sigma_{\max}(D)$	$\sigma_{\min}(D)$	Condition number
Least squares	49.706	1.759e-06	2.826e+07
Huber	49.904	1.790e-06	2.788e+07
Cauchy	49.711	1.791e-06	2.776e+07
GLDM	45.271	4.845e-07	9.343e+07
Inverse-power GLDM	48.510	1.851e-05	2.621e+06

For $q = 1, 2$, local objective profiles fix either $\log_{10} k_{xt}$ or $\log_{10} k_{ih}$ at $-6, -4, -2, 0, 2, 4, 6$ and at its fitted value, then reoptimize the other five coordinates. Two initial candidates use the wider-box fit and the preceding profile point; each receives continuation radii $10^{-4}, 10^{-6}$ and at most 160 L-BFGS-B iterations per radius. The unrefined wider-box fit with the fixed coordinate replaced is also retained as a feasible candidate. The smallest unsmoothed training criterion among these candidates defines the displayed local profile. This gives 31 distinct fixed-coordinate values; it does not calculate a globally minimized profile or a likelihood-ratio confidence interval.



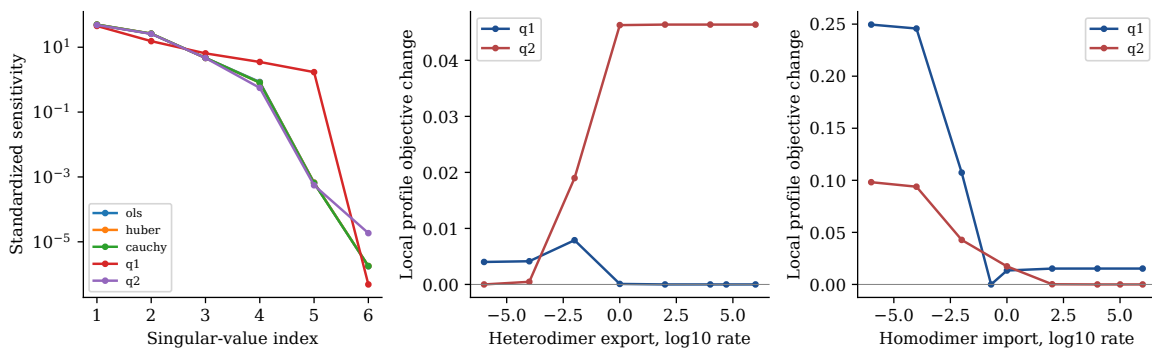


FIGURE 18. STAT5 sensitivity and local objective-profile diagnostics. Left: the six singular values for each wider-box forecast fit. Middle and right: the change in unsmoothed training objective when heterodimer export or homodimer import is fixed and the remaining rates are locally refitted. Profiles use only training observations and are not likelihood confidence curves.

The ordinary GLDM heterodimer-export profile varies by less than 3×10^{-8} above its reference criterion over log-rates 4 to 6, illustrating a nearly flat high-rate region. The $q = 2$ homodimer-import profile changes by less than 2×10^{-6} in absolute criterion value between log-rates 4 and 6. A slightly smaller criterion at one profile point reflects local-basin and numerical effects, rather than a negative uncertainty bound. These profiles reinforce the distinction between stable predictions and well-determined individual rates. The held-out observations play no role in constructing or selecting the profiles. These profiles describe the original wider-box fits; the cross-initialization experiment below separately tests alternative basins.

Software versions are as above, with pandas 2.2.3. Each fit uses one BLAS/OpenMP thread and at most four fits run concurrently. Times include integration, sensitivities and optimization; no cross-method speed claim is made. The archive retains pinned benchmark revisions/hashes, distinct experimental and simulated inputs, starts, estimates, forecasts and solver checks.

9.12. STAT5 cross-initialization and criterion-specific selection. This experiment tests search dependence while holding each target criterion fixed. Cross-initialization exposes alternative local solutions, and chronological forecasts assess the consequences of training-objective selection.

The seven-start results above compare the fits reached by the original fitting procedures. To test their basin sensitivity, each arctangent criterion is initialized both at its own recorded forecast solution and at the other criterion's solution. The 36 training measurements, frozen scales, $c = 1.5$, zero penalty and log-rate box $[-5, 5]^6$ are unchanged. Each of the four prescribed starts receives the same three continuation levels $10^{-2}, 10^{-4}, 10^{-6}$, with at most 600 L-BFGS-B iterations per level, followed by at most six MM cycles of at most 400 iterations each, with QP tolerance 10^{-11} . Thus each of the four refinement runs has an allowance of at most $6 \times 400 = 2,400$ MM iterations, separately from its continuation stages. These are permitted limits. The actual MM counts and evaluation counts are reported in Table 24; all four runs stopped in their first MM cycle. A cycle may stop at the existing step/gap rule; the independent 10^{-4} tube target is then checked. The feasible initial and continuation candidates are retained, and the smallest unsmoothed training objective among the candidates for that run is reported. Holdout values do not select a candidate. Self-initialization controls receive the same refinement budget as cross-initialization.

For ordinary GLDM, the cross-initialized fit has training objective 0.445920541 and held-out MAE 6.599. Its objective exceeds the retained value 0.431281636, so criterion-based selection keeps the forecast with MAE 14.195. For inverse-power GLDM, expanded search lowers the selected objective from 0.335620546 to 0.309651982, while held-out MAE rises from 6.668 to 13.214. Objective values are compared only within the same loss. Broader search and better training optimization therefore need not improve chronological prediction.

TABLE 23. STAT5 search-protocol comparison. Panel (a) evaluates each prescribed refinement under its target criterion. Panel (b) reports the fit selected by the smallest own unsmoothed training objective under each protocol. Compare objective values only within a fixed q . Held-out MAE is in percentage points and never selects a fit. Original-protocol errors and curves appear in Table 18 and Figures 16–17.

<i>(a) Prescribed refinements in the expanded search</i>						
Target q	Start q	Objective before	Objective after	MAE before	MAE after	Tube residual
1	1	0.431281636	0.431281636	14.195	14.195	2.83×10^{-6}
1	2	0.445921682	0.445920541	6.668	6.599	4.32×10^{-7}
2	1	0.310314941	0.309651982	14.195	13.214	6.39×10^{-9}
2	2	0.335620546	0.335620546	6.668	6.668	6.94×10^{-7}

<i>(b) Training-objective-selected forecast fits</i>			
Search protocol	Target q	Selected training objective	Held-out MAE
Original seven-start	1	0.431281636	14.195
Original seven-start	2	0.335620546	6.668
Expanded cross-initialization	1	0.431281636	14.195
Expanded cross-initialization	2	0.309651982	13.214

All four refined fits meet the independent 10^{-4} tube-residual target. This numerical stationarity evidence does not resolve the forecast differences across basins. The wider-box diagnostics in Table 21 describe the original protocol, not the expanded search. Table 24 records the actual work and Table 25 retains every saved candidate. Continuation

TABLE 24. Actual STAT5 forecast computation counts recovered from retained logs. N_{cont} counts joint objective/gradient calls in continuation, K_{MM} actual MM iterations, N_{MM} map/Jacobian queries in MM and N_{ODE} cache-miss augmented ODE solves over the fit. These counters differ and must not be added. Both MM candidates in each original search, and all four expanded runs, satisfy their recorded step/gap stopping rule.

Protocol	q	Start	N_{cont}	K_{MM}	N_{MM}	N_{ODE}
Original	1	seven	1375	24	70	1432
Original	2	seven	612	55	132	730
Expanded	1	$q = 1$	232	1	18	247
Expanded	1	$q = 2$	156	22	48	201
Expanded	2	$q = 1$	252	2	10	259
Expanded	2	$q = 2$	153	56	117	267

returns objective and gradient together, so its reported function and Jacobian call counts coincide. Cached ODE solutions can serve repeated queries. The historical continuation records retain evaluation counts and final success flags, but not per-level L-BFGS-B iteration counts or every termination message; these unrecorded quantities are not inferred from the iteration allowances. MM iteration counts, traces, gaps and step-rule outcomes are retained. The release also contains every retained candidate parameter, its criterion value and all 48 predicted measurements. Candidate objectives are recomputed without smoothing for comparison within the same q ; the original smoothed values are retained in separate fields.

STAT5 is consequently an optimization-sensitivity demonstration. The two protocols give materially different impressions of the exponent contrast, and neither is a sufficiently exhaustive global search to separate loss effects



TABLE 25. All retained arctangent STAT5 forecast candidates. Objectives are independently reevaluated without smoothing at the saved parameters, always under the row's own criterion. MAE uses the same 12 held-out measurements (percentage points). Original source numbers index the seven starts; expanded source numbers denote the original q fit. C denotes continuation, M the retained MM endpoint and I the retained initial point. These are saved endpoints, not every iterate.

Protocol	q	Source	Stage	Own objective	Held-out MAE
Original	1	2	C	0.431281649	14.195
Original	1	0	C	0.445920555	6.596
Original	1	1	C	0.590497923	8.059
Original	1	6	C	0.792255431	10.117
Original	1	5	C	1.389020637	25.920
Original	1	3	C	1.389020666	25.920
Original	1	4	C	1.389020808	25.920
Original	1	2	M	0.431281636	14.195
Original	1	0	M	0.445920541	6.599
Original	2	0	C	0.335620601	6.651
Original	2	2	C	0.342353182	20.730
Original	2	1	C	0.763475278	25.919
Original	2	6	C	0.763475477	25.918
Original	2	5	C	0.763475528	25.920
Original	2	3	C	0.763475538	25.920
Original	2	4	C	0.763475555	25.920
Original	2	0	M	0.335620546	6.668
Original	2	2	M	0.342353167	20.730
Expanded	1	1	I	0.431281636	14.195
Expanded	1	1	C	0.431281650	14.195
Expanded	1	1	M	0.431281636	14.195
Expanded	1	2	I	0.445921682	6.668
Expanded	1	2	C	0.445920555	6.596
Expanded	1	2	M	0.445920541	6.599
Expanded	2	1	I	0.310314941	14.195
Expanded	2	1	C	0.309652305	13.215
Expanded	2	1	M	0.309651982	13.214
Expanded	2	2	I	0.335620546	6.668
Expanded	2	2	C	0.335620601	6.651
Expanded	2	2	M	0.335620546	6.668

from basin effects. The original favorable $q = 2$ contrast is not evidence of predictive superiority. Its 12 held-out measurements arise from three observables at four times, not 12 independent replicated experiments.

9.13. A second public nonlinear ODE benchmark. This limited empirical illustration examines local calibration from very short observed trajectories. It does not test the transfer theorem or support epidemiological inference.

The public Bertozzi COVID-19 benchmark [7, 38] supplies an application outside biochemical signaling. Its PETab representation contains eleven daily reported counts for each of California and New York, from 20 to 30 March 2020. The analysis retains the benchmark's three-state SIR dynamics and observation mapping.

$$\begin{aligned}
 \dot{S} &= -\mathcal{R}_0 \gamma SI, & \dot{I} &= \mathcal{R}_0 \gamma SI - \gamma I, & \dot{R} &= \gamma I, \\
 (S, I, R)(0) &= (1 - I_0/N, I_0/N, 0), & H(t) &= NI(t).
 \end{aligned} \tag{9.21}$$

Here $N = 39,560,000$ for California and $19,540,000$ for New York. The reported-count table is used as supplied; retaining its NI observation convention does not independently validate the interpretation of reported cases as the



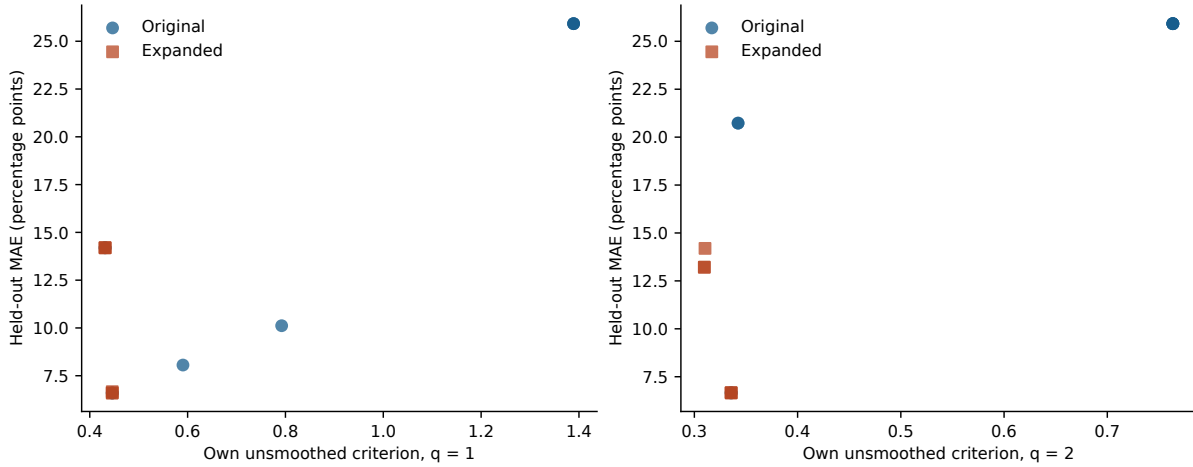


FIGURE 19. STAT5 retained candidates from the original seven-start and expanded protocols. Each panel compares only one training criterion; abscissae are not comparable across panels. Ordinary GLDM has criterion-selected MAE 14.195 under both protocols. Inverse-power GLDM has MAE 6.668 under the original protocol and 13.214 under expanded selection. Dispersion across retained candidates demonstrates search sensitivity and disagreement between training fit and chronological forecast accuracy.

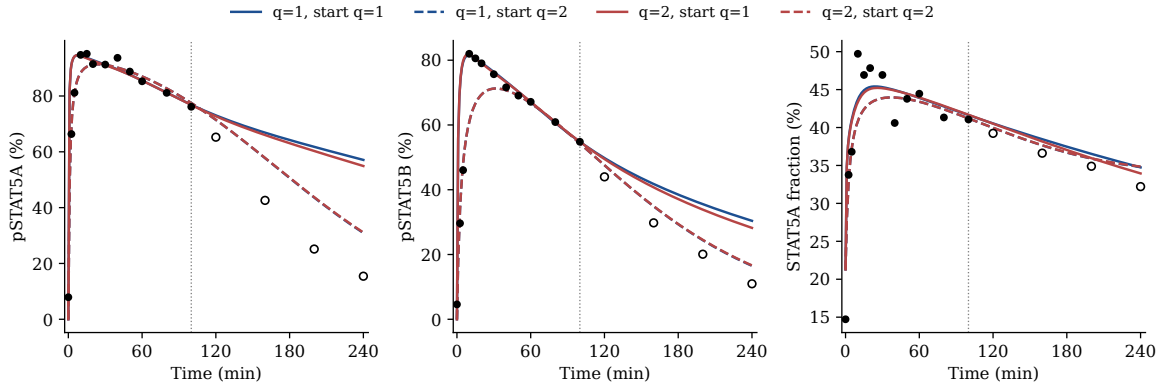


FIGURE 20. STAT5 forecasts after the four prescribed refinements. Blue denotes target $q = 1$ and red target $q = 2$; solid lines start from the original $q = 1$ fit and dashed lines from the original $q = 2$ fit. Filled observations are used for training and open observations are held out. The curves show basin dependence within each criterion, not a causal decomposition of loss effects.

infectious compartment. The variables $\mathcal{R}_0$, γ and I_0 are separately fitted in each region, with bounds $[1.5, 5]$, $[0.01, 0.2]$ and respectively $[100, 1000]$ or $[1000, 10000]$. Optimization uses log coordinates rescaled to a unit box. The positive simplex $S + I + R = 1$ is invariant because the vector field is inward pointing at its boundary and its components sum to zero. The smooth finite-time observation map therefore meets the local regularity requirements on the stated positive parameter box. These properties do not establish identifiability from eleven observations.

The first eight days are used for fitting and the last three are held out. The common nine starts are the unit-box midpoint and eight Latin-hypercube points with seed 2026100307. An OLS pilot determines the residual MAD, with a lower scale limit of 100 counts, giving fixed scales 100.000 and 272.492 for California and New York. OLS, Huber,

Cauchy, ordinary GLDM, inverse-power GLDM and Laplace correntropy then share those scales, starts, bounds and training observations. The three cusp losses use $c = 1.5$ and continuation radii $10^{-2}, 10^{-4}, 10^{-6}$; all fits use L-BFGS-B with at most 600 iterations per stage. There is no criterion penalty and no added contamination. All observation maps and forward sensitivities use DOP853 with relative/absolute tolerances 2×10^{-10} and 2×10^{-13} . Selection uses the training criterion only.

TABLE 26. Errors in reported counts for the public SIR benchmark. Each region has eight training and three held-out observations. These are descriptive errors for two short observed series, not Monte Carlo means.

Method	California		New York	
	Train MAE	Test MAE	Train MAE	Test MAE
Least squares	79.238	472.261	334.103	1963.769
Huber	74.528	399.189	316.310	1801.440
Cauchy	73.981	394.275	316.060	1801.781
GLDM	66.458	292.653	298.904	1746.043
Inverse-power GLDM	67.305	313.784	298.904	1746.043
Laplace correntropy	66.458	292.652	298.904	1746.043

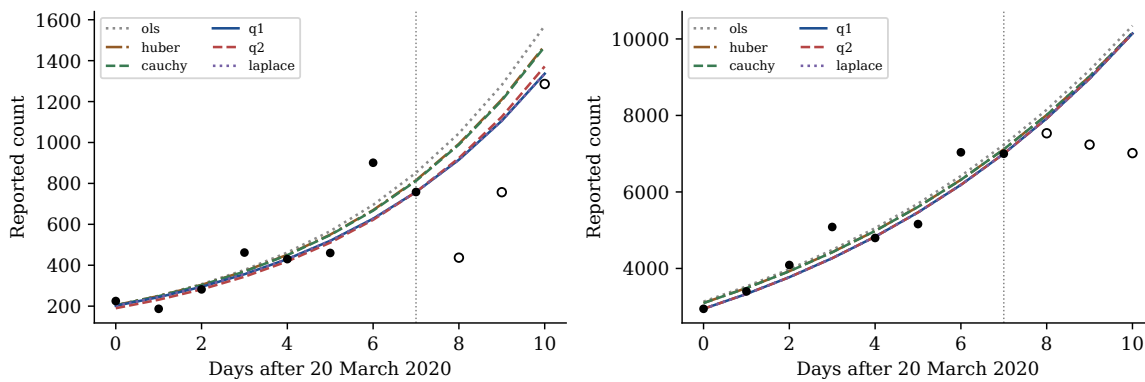


FIGURE 21. Public SIR benchmark forecasts for California (left) and New York (right). Filled circles denote training counts and open circles the three held-out counts. The vertical line marks the training cutoff. Lines join the daily numerical predictions. Several cusp-loss trajectories nearly coincide, especially in New York; Table 26 reports their errors separately.

In California, ordinary GLDM and Laplace correntropy give nearly equal held-out MAE, approximately 292.65 counts, whereas inverse-power GLDM gives 313.78. In New York the three cusp-loss forecasts agree to the reported precision, with MAE approximately 1746.04. Thus changing the scalar loss need not materially change a short-horizon prediction, and the $q = 2$ criterion is not uniformly better. During early growth, combinations of $\mathcal{R}_0$, γ and I_0 can generate similar trajectories. Several fitted values reach the specified bounds: $\gamma = 0.2$ for least squares and Huber in both regions, and $\mathcal{R}_0 = 5$ for all three cusp losses in New York. These are local benchmark fits, not independently identified epidemiological rates or operational epidemic forecasts.

Independent Radau integration at $2 \times 10^{-12}/2 \times 10^{-15}$ tolerances changes counts by at most 1.65×10^{-6} ; maximum conservation discrepancy is 2.23×10^{-16} . The largest smoothed unit-box projected-gradient residual is 2.88×10^{-4} : not every fit meets 10^{-4} , and no validated certificate is claimed. Only 22 observations and six holdouts are available. WENDy [10, 42] remains an alternative, but its comparison under partial observation requires a separate reconstruction protocol.

9.14. Application of Theorem 8.1 to Khanty-Mansiysk NDVI dynamics. This study asks whether Theorem 8.1's observation-map error, sensitivity-error inflation and tube enlargement can be evaluated directly for an ODE calibrated to the supplied January tabulations. Its evidence level is numerical verification. Forecast errors are reported separately as empirical observations.

Let d_y be the reported Class 1 percentage in year y . This class uses the reported NDVI interval -0.2 to 0.0 ; its share is neither waterlogged area nor a measurement of groundwater. The observation unit is one annual regional composition. The complementary share groups Classes 2–5 and supplies a conservation constraint, not an additional independent observation. The model below is linear in its states and nonlinear in the fitted log-rate and logit coordinates. Its purpose is to connect an ODE fit to the theorem's numerical quantities on the actual tabulations.

States, parameters and observation map. Set $t = y - 2001$ in years and let $x(t)$ denote the Class 1 share, with complement $u(t) = 1 - x(t)$. The two-state relaxation model is

$$\begin{aligned} \dot{x} &= \kappa\{\xi u - (1 - \xi)x\}, & \dot{u} &= -\dot{x}, \\ (x(0), u(0)) &= (x_{\text{init}}, 1 - x_{\text{init}}), & x_{\text{init}} &= d_{2001}/100 = 0.3101103290940961, \\ \theta &= (\eta, \beta)^\top, & \kappa &= e^\eta, \quad \xi = (1 + e^{-\beta})^{-1}. \end{aligned} \quad (9.22)$$

Here κ is a relaxation rate in year $^{-1}$ and ξ an equilibrium share. They summarize the fitted temporal curve. They are not measured rates of pixel conversion, drainage or vegetation recovery. The reduction $\dot{x} = \kappa(\xi - x)$ follows from conservation. The parameter box is

$$\Theta = [\log(0.01), \log 2] \times [\text{logit}(0.01), \text{logit}(0.99)], \quad \text{logit}(v) = \log \frac{v}{1-v}. \quad (9.23)$$

The rate interval permits relaxation times from 0.5 to 100 years, and the equilibrium bounds avoid endpoint logits. These are numerical search constraints, not biological parameter bounds. Their effect is checked below. Conditioning on the first observation as the initial state avoids adding a poorly supported third parameter.

Calibration uses the 19 subsequent years 2002–2020, at $t_i = i$ for $i = 1, \dots, 19$. The five years 2021–2025 are withheld from fitting and selection. No missing season is imputed between January observations. The fixed training-only scale is

$$\begin{aligned} s &= \max\{1, 1.482602218505602 \text{ median}_i |d_{2001+i} - \text{median}_\ell d_{2001+\ell}|\} \\ &= 6.061024396982582 \text{ percentage points.} \end{aligned} \quad (9.24)$$

Define $y_i = d_{2001+i}/s$, $g_{i,0}(\theta) = 100x(t_i; \theta)/s$ and $a_i = 1/19$. The complement is not duplicated in the objective. With $c = 1$ and $\lambda = 0$, both loss choices minimize

$$F_{q,h}(\theta) = \frac{1}{19} \sum_{i=1}^{19} \frac{\pi}{2q} \left[1 - \left\{ \frac{2}{\pi} \text{atan2}(1, |y_i - g_{i,h}(\theta)|) \right\}^q \right], \quad q \in \{1, 2\}. \quad (9.25)$$

The continuous and discrete criteria share these observations, weights, initial value, scale and constraints. This is a conditional numerical experiment with an estimated scale, not an application of the fixed-scale statistical consistency theorem.

Integration and sensitivities. The continuous reference is available in closed form,

$$x(t; \theta) = \xi + (x_{\text{init}} - \xi)e^{-\kappa t}. \quad (9.26)$$

The calibration nevertheless uses classical RK4 on the state and sensitivity equations, with $h = 1, 1/2, 1/4$ years. Integer observation times lie on each mesh. The two sensitivities $S_\eta = \partial x / \partial \eta$ and $S_\beta = \partial x / \partial \beta$ satisfy

$$\begin{aligned} \dot{S}_\eta &= -\kappa S_\eta + \kappa(\xi - x), & S_\eta(0) &= 0, \\ \dot{S}_\beta &= -\kappa S_\beta + \kappa\xi(1 - \xi), & S_\beta(0) &= 0. \end{aligned} \quad (9.27)$$



All three components use the same RK4 stages. This gives the discrete observation map and its row Jacobian

$$\begin{aligned} g_{i,h} &= 100x_h(t_i)/s, & J_{i,h} &= \frac{100}{s}(S_{\eta,h}(t_i), S_{\beta,h}(t_i)), \\ J_{i,0} &= \frac{100}{s}(\kappa t_i(\xi - x_{\text{init}})e^{-\kappa t_i}, \xi(1 - \xi)(1 - e^{-\kappa t_i})). \end{aligned} \quad (9.28)$$

Proposition 9.3 (Conservation and independently evaluable ODE maps). *For $x_{\text{init}} \in [0, 1]$ and $\theta \in \Theta$, Eq. (9.22) has a unique solution with $x, u \in [0, 1]$ and $x + u = 1$ for $t \geq 0$. Put*

$$p(z) = 1 + z + z^2/2 + z^3/6 + z^4/24, \quad n_i = t_i/h, \quad d_{i,h} = p(-\kappa h)^{n_i}. \quad (9.29)$$

At the stated meshes, RK4 preserves the share interval and has the exact-arithmetic expressions

$$\begin{aligned} x_h(t_i) &= \xi + (x_{\text{init}} - \xi)d_{i,h}, \\ S_{\eta,h}(t_i) &= (x_{\text{init}} - \xi)n_i p(-\kappa h)^{n_i-1} p'(-\kappa h)(-\kappa h), \\ S_{\beta,h}(t_i) &= \xi(1 - \xi)(1 - d_{i,h}). \end{aligned} \quad (9.30)$$

Consequently, Eqs. (9.26)–(9.30) determine the map and Jacobian discrepancies at each fixed parameter without fitting a separate reference model. For fixed observation times these discrepancies are $O(h^4)$ as $h \downarrow 0$.

Proof. The globally Lipschitz linear state equation gives uniqueness. The derivative of $x + u$ vanishes, and Eq. (9.26) is a convex combination of x_{init} and ξ . Applying an RK4 step to $x - \xi$ multiplies it by $p(-\kappa h)$. For $0 \leq v = \kappa h \leq 2$,

$$p(-v) = (1 - v/2)^2 + v^2\{(v - 2)^2 + 2\}/24 > 0, \quad 1 - p(-v) = v\{1 - v/2 + v^2(4 - v)/24\} \geq 0. \quad (9.31)$$

Thus each step is a convex combination with ξ and preserves $[0, 1]$. Differentiating the explicit RK4 recurrence is equivalent to applying its stages to Eq. (9.27), giving Eq. (9.30). The difference $p(-\kappa h) - e^{-\kappa h} = O(h^5)$, together with its parameter derivative, accumulates over $O(h^{-1})$ steps on a fixed time interval to give order four. The compact parameter box bounds the constants. $\square$

Pointwise certificate transfer. At each fitted parameter the references in Eq. (9.28) are evaluated in double precision. Set

$$\begin{aligned} \widehat{e}_{i,h} &= |g_{i,h} - g_{i,0}|, & \widehat{E}_{i,h} &= \|J_{i,h} - J_{i,0}\|_2, & \widehat{Q}_i &= \|J_{i,0}\|_2, \\ L_1 &= 9/(8\sqrt{3}) = 0.649519\dots, & L_2 &= L_1 + 2/\pi = 1.286139\dots, & \widehat{B}_{q,h} &= \frac{1}{19} \sum_{i=1}^{19} (\widehat{E}_{i,h} + \widehat{Q}_i L_q \widehat{e}_{i,h}). \end{aligned} \quad (9.32)$$

Hats on these inputs denote measured floating-point discrepancies and evaluated row norms. The closed-form expressions are independent references, but their floating-point evaluation is not an outward-rounded enclosure. They must therefore not be substituted for rigorously justified bounds in a claim of validated integration. Here they provide a numerical verification of the transfer formula at the recorded parameter.

The response tube is $\tau_i = 10^{-5}$ in standardized units. For $b = 0, h$, the LP uses the score sets of Theorem 8.1 and minimizes $\| -J_b^T \alpha + n \|_\infty$ over their product and $n \in N_\Theta(\theta)$. The continuous LP uses the componentwise radii $\tau_i + \widehat{e}_{i,h}$, not the unchanged tube. Each score outside the discrete tube is reset to its specified value, inside scores are clipped to $[-a_i, a_i]$, and the box normal is checked in the stated coordinates. Re-evaluation then supplies a feasible discrete norm $\widehat{R}_h$. The transported score preserves $\alpha_{i,h}$ inside the enlarged continuous tube and uses the continuous score outside it. We record both LP minima, the feasible witness norms, and

$$\begin{aligned} U_{q,h} &= \widehat{R}_h + \widehat{B}_{q,h}, & \widehat{\Delta}_{q,h} &= R_{\tau,h}^{q,0} + \widehat{B}_{q,h} - R_{\tau+\widehat{e}_h,0}^{q,0}, \\ v_y &= \max_i(\tau_i + \widehat{e}_{i,h}), & \epsilon_R &= 10^{-4}, & \epsilon_y &= 2 \times 10^{-5}. \end{aligned} \quad (9.33)$$

In this subsection the displayed LP residuals are floating-point LP outputs. A numerical pass requires $U_{q,h} \leq \epsilon_R$ and $v_y \leq \epsilon_y$ under Algorithm 4. Missing diagnostics or an unmet budget give an unresolved outcome. A verified outcome



would additionally require justified state, sensitivity and arithmetic bounds; no NDVI row receives that label. The exact mathematical inequality being checked remains

$$R_{\tau+e_h,0}^{q,0}(\theta) \leq R_{\tau,h}^{q,0}(\theta) + \frac{1}{19} \sum_{i=1}^{19} (E_{i,h} + Q_i L_q e_{i,h}), \quad (9.34)$$

with actual error bounds $e_{i,h}$, $E_{i,h}$, Q_i as required by Theorem 8.1.

Fitting protocol and recorded results. The nine common starts are the Cartesian product $\kappa \in \{0.05, 0.2, 0.8\}$ and $\xi \in \{0.25, 0.4, 0.6\}$. Each uses smoothing levels $10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$, up to 500 L-BFGS-B iterations at each level, followed by up to 240 unsmoothed proximal-MM iterations with inner tolerance 10^{-10} . Selection minimizes the original unsmoothed training objective within each q and mesh. Test errors do not select the parameter. These are local fits with feasible stationarity diagnostics, not global minima. The six mesh-specific fits involve 54 starting attempts. Holding each exponent's selected $h = 1/4$ parameter fixed while re-evaluating all meshes supplies a separate six-row check that isolates integration error. The fine-mesh rows occur in both records and are not independent replications.

TABLE 27. January share-relaxation ODE fits selected by each training objective. MAEs are in percentage points, using 19 training years and one five-year holdout. n_{cont} and n_{MM} are actual iterations for the selected start; continuation iterations sum four smoothing stages. Objectives are compared only within a fixed q .

q	h (years)	κ	100ξ	$F_{q,h}$	Train MAE	Test MAE	n_{cont}	n_{MM}
1	1.00	0.01000	81.825	0.499262	4.117	2.210	82	1
1	0.50	0.01000	81.825	0.499262	4.117	2.210	91	1
1	0.25	0.01000	81.825	0.499262	4.117	2.210	87	1
2	1.00	0.01000	81.825	0.378677	4.117	2.210	74	1
2	0.50	0.01000	81.825	0.378677	4.117	2.210	73	1
2	0.25	0.01000	81.825	0.378677	4.117	2.210	68	1

Both exponents yield trajectories that agree to the reported precision, with $\kappa = 0.01$ at the lower bound. The observed starting value and approximately linear fitted drift over this short record do not identify the distant equilibrium. Decreasing the lower rate bound to 0.001 and repeating the same nine starts at $h = 1/4$ moves the selected rate to 0.007319 and the equilibrium to its upper bound of 99% for both q values. The maximum 2001–2025 trajectory change is 0.110 percentage points, while five-year MAE changes from 2.210 to 2.219. The boundary shifts rather than disappears. This supports using the fit for a pointwise solver diagnostic, without treating the rate or equilibrium as identified physical quantities. Within each loss, the wider search has a lower training objective; values across different losses are not comparable.

TABLE 28. Lower-rate-bound sensitivity at $h = 1/4$ years with the same nine starts and training-only selection. The second protocol adds 18 starting attempts. Test MAE uses the same five held-out annual Class 1 values. A boundary remains active under both search boxes.

q	Lower rate	Fitted rate	Equilibrium (%)	Training objective	Test MAE	Active bound
1	0.010	0.010000	81.825	0.499262	2.210	$\kappa = 0.01$
1	0.001	0.007319	99.000	0.498773	2.219	$\xi = 0.99$
2	0.010	0.010000	81.825	0.378677	2.210	$\kappa = 0.01$
2	0.001	0.007319	99.000	0.378098	2.219	$\xi = 0.99$



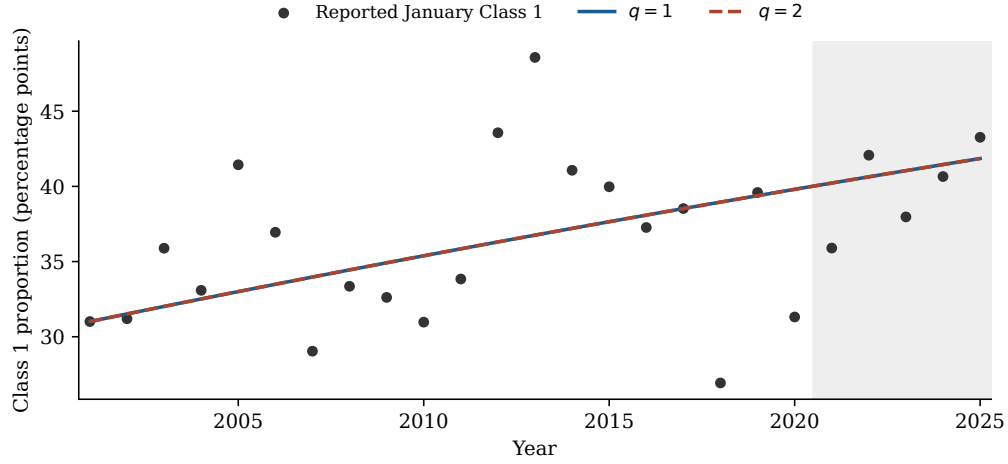


FIGURE 22. Khanty-Mansiysk January Class 1 observations and the fitted share-relaxation ODE. The initial value is fixed at 2001, calibration uses 19 annual observations for 2002–2020, and the shaded interval marks the five held-out years 2021–2025. The two loss curves overlap to plotting precision. These are regional tabulated percentages, not measured waterlogging or vegetation-state trajectories.

TABLE 29. Application of Theorem 8.1 to the January ODE at each exponent's fixed $h = 1/4$ fitted parameter. R_h and R_0 are numerical LP outputs for the discrete tube and enlarged continuous tube. $e_{\max}, E_{\max}$ are measured standardized discrepancies. The slack is $R_h + \hat{B}_{q,h} - R_0$. N denotes numerical pass under the two budgets in Eq. (9.33); no row is interval-verified. The separate mesh-refit record gives the same decisions.

q	h (years)	R_h	$e_{\max}$	$E_{\max}$	$\hat{B}_{q,h}$	R_0	Transfer slack	Decision
1	1.00	0	1.11×10^{-10}	5.34×10^{-10}	3.42×10^{-10}	0	3.42×10^{-10}	N
1	0.50	0	6.89×10^{-12}	3.32×10^{-11}	2.13×10^{-11}	0	2.13×10^{-11}	N
1	0.25	0	4.29×10^{-13}	2.07×10^{-12}	1.33×10^{-12}	0	1.33×10^{-12}	N
2	1.00	0	1.11×10^{-10}	5.34×10^{-10}	3.80×10^{-10}	0	3.80×10^{-10}	N
2	0.50	0	6.89×10^{-12}	3.32×10^{-11}	2.37×10^{-11}	0	2.37×10^{-11}	N
2	0.25	0	4.29×10^{-13}	2.07×10^{-12}	1.47×10^{-12}	0	1.47×10^{-12}	N

All six mesh-specific checks and six fixed-parameter re-evaluations have nonnegative numerical LP slack. The minimum is 1.326×10^{-12} , with maximum positive violation zero at stored precision. Recomputed discrete witness norms are below 1.3×10^{-17} , and transported continuous witnesses also satisfy the measured comparison. These values describe numerical residual evaluation, not accuracy validated to seventeen decimal places. At fixed parameters the map and sensitivity discrepancies, inflation and transferred bound decrease by factors close to sixteen under mesh halving, as expected for RK4. The one-year mesh already meets the stated numerical budgets for this slowly varying fitted model, so additional refinement is not required by those budgets.

For example, at the fixed $q = 2$ parameter and $h = 1$,

$$\begin{aligned}
 \max_i \hat{e}_{i,h} &= 1.106919 \times 10^{-10}, & \max_i \hat{E}_{i,h} &= 5.343286 \times 10^{-10}, & \max_i \hat{Q}_i &= 1.384002, \\
 \hat{R}_h &= 3.469447 \times 10^{-18}, & \hat{B}_{2,h} &= 3.802856 \times 10^{-10}, & U_{2,h} &= 3.802856 \times 10^{-10}, \\
 R_{\tau+\hat{e}_h,0}^{2,0} &= 0 \text{ (LP output)}, & v_y &= 1.000011069 \times 10^{-5}, & \hat{\Delta}_{2,h} &= 3.802856 \times 10^{-10}.
 \end{aligned} \tag{9.35}$$



The inflation decomposes as

$$\frac{1}{19} \sum_i \widehat{E}_{i,h} = 3.027723 \times 10^{-10}, \quad \frac{L_2}{19} \sum_i \widehat{Q}_i \widehat{e}_{i,h} = 7.751329 \times 10^{-11}. \quad (9.36)$$

The discrete tube contains the 2017 residual, and the feasible box normal is $n_h = (-0.00212491132728, 0)^\top$ at the active lower rate bound. The transported continuous witness has norm 2.660954×10^{-11} , so its directly recomputed slack is 3.536761×10^{-10} . Both $U_{2,h} < 10^{-4}$ and $v_y < 2 \times 10^{-5}$ hold, so the outcome is a numerical pass. The weighted state and sensitivity contributions use all 19 componentwise values, rather than substituting their maxima into the sum. Every component and transported witness is retained in the computational record.

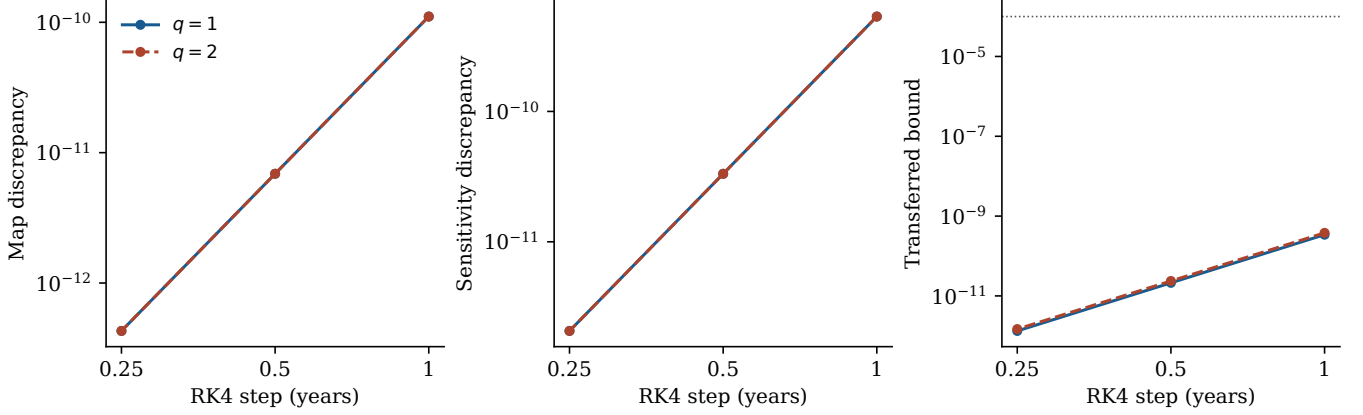


FIGURE 23. Numerical transfer verification for the January share-relaxation ODE at each exponent's fixed fine-mesh parameter. Left and middle: maximum standardized observation-map and Jacobian-row discrepancies against the analytic reference. Right: feasible transferred upper bound $U_{q,h}$; the dotted line is the stationarity budget. Each line contains three deterministic mesh evaluations of one fitted parameter, with no Monte Carlo interpretation. The error curves nearly overlap because the $q=1$ and $q=2$ fitted trajectories agree to the reported precision.

An independent stability-polynomial calculation agrees with the augmented RK4 implementation to 1.17×10^{-15} in unstandardized state/sensitivity units. Central-difference sensitivities agree to 1.28×10^{-11} , and a DOP853 evaluation with relative/absolute tolerances $10^{-12}/10^{-14}$ agrees with the analytic reference to 3.65×10^{-14} . These are floating-point implementation checks. The reported environment is the same single-thread configuration stated above. The polynomial forecasts in Section 10 remain a separate application with analytic propagation. This ODE study is the direct January demonstration of numerical stationarity transfer and neither repairs the satellite provenance nor establishes an environmental trend.

Algorithm 7 January ODE calibration and numerical transfer evaluation.

Require: January tabulations, the split, scale and box in Eqs. (9.22)–(9.24), meshes 1, 1/2, 1/4, and fixed budgets.

- 1: **for** $q \in \{1, 2\}$ and each mesh h **do**
- 2: Integrate the state and sensitivities by the same RK4 stages at each of the nine starts.
- 3: Run the finite continuation and MM solves; retain every candidate and choose the lowest unsmoothed training objective.
- 4: At that parameter evaluate the analytic reference and row sensitivities, then the measured discrepancies in Eq. (9.32).
- 5: Solve the discrete and enlarged continuous tube LPs; construct and re-evaluate feasible scores and the box normal.
- 6: Evaluate both sides of Eq. (9.34), LP slack, transported-witness slack, $U_{q,h}$ and v_y .
- 7: Apply Algorithm 4 with the evidence flag set to numerical. Record the decision and any unmet budget.
- 8: **end for**
- 9: **for** each exponent's selected fine-mesh parameter **do**
- 10: Re-evaluate all three meshes at that same parameter, preserving data, scale, weights, penalty and constraints.
- 11: **end for**
- 12: Evaluate the five held-out observations without using them for candidate selection.
- 13: Retain starts, iterations, map/Jacobian evaluations, componentwise errors, witnesses, predictions and decisions.

10. FORECASTING SUPPLIED JANUARY CLASS COMPOSITIONS

This empirical forecasting study asks how polynomial degree, regularization and loss shape affect chronological extrapolation of the supplied January compositions. It has analytic propagation and no ODE discretization error. Section 9.14 is the separate numerical application of Theorem 8.1.

The clearest methodological observation is the degree-five penalty contrast (Table 31). Adding $\lambda = 0.1$ to ordinary GLDM reduces rolling MAE from 4.013 to 2.034 percentage points. Changing from $q = 1$ to $q = 2$ at that numerical penalty gives 2.024. Most of the reported reduction therefore comes from regularization. The additional 0.010-point difference does not establish meaningful exponent superiority.

The inputs are 25 annual compositions for 2001–2025 from the January worksheet; the all-zero 2000 record is excluded. The reported preparation used MODIS imagery, NDVI calculation, correction, combination, monthly mapping and threshold classification. The processing metadata needed to reconstruct the counts are incomplete, and the nine map-display intervals cannot reproduce the five modeled classes. The analysis starts with supplied tabulations and supports no inference about preprocessing accuracy or environmental change. Appendices C–E retain all 25 maps, class-specific tables and extended sensitivity analyses. Their 900 objective spreads above 10^{-6} among 2,950 multistart class fits further restrict interpretation to local procedures. There are ten nominal one-year forecast origins, with cutoffs 2015–2024 and target years 2016–2025. Each origin contributes five linked class proportions, so the pooled rolling MAE has 50 class-year errors. Expanding training windows overlap heavily. The fixed 2020 cutoff supplies one five-year holdout with 25 class-year errors; its 2021–2025 targets overlap the last five rolling targets. These are two views of one annual record, not independent replications. An effective number of independent origins is not identified from these 25 compositions, and no effective-sample-size estimate or sampling confidence interval is assigned. Historical tuning further reuses earlier observations. Prospective performance remains unknown.

The computational release provides year, class, origin, horizon, observed proportion, prediction, signed error and procedure settings for every reported forecast, together with per-origin metrics for every principal comparator. Figure 24 shows their variation. The common numerical penalty is an explicit comparison convention. Since the loss range is $A_q = \pi/(2q)$, equal λ does not preserve the ratio of the penalty coefficient to the loss range across exponents.



TABLE 30. January validation denominators and dependence. Five class percentages are linked by closure; no row represents independent replication of a regional process.

Protocol	Cutoffs / targets	Error terms	Dependence
One-year rolling	2015–2024 / 2016–2025	$10 \times 5 = 50$	Expanding training windows; one annual series
Five-year holdout	2020 / 2021–2025	$1 \times 5 \times 5 = 25$	One block; repeats five rolling target years
Primary inner tuning	Five earlier one-year origins per fit	$5 \times 5 = 25$	Nested historical reuse; not extra test data
Horizon sensitivity	Six cutoffs, 2010–2015; horizons $H = 1, 3, 5$	$6 \times H \times 5$	Overlapping training and, for $H > 1$, forecast blocks



FIGURE 24. January performance at all ten one-year rolling origins (cutoffs 2015–2024). Top: principal drift, degree-one standard comparators and training-selected polynomial procedures; bottom: degree-five common-penalty ablation. Every point averages five class errors in percentage points. Training windows overlap and the five components are compositionally linked; lines are descriptive, not independent replicates or confidence intervals. The 2021–2025 targets also occur in the separate five-year holdout.



TABLE 31. Degree-five January rolling MAE under a common penalty, in percentage points. This ablation separates the penalty change from the exponent change. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Criterion	q	λ	MAE
Unpenalized GLDM	1	0	4.013
Penalized GLDM	1	0.1	2.034
Penalized inverse-power GLDM	2	0.1	2.024

11. NUMERICAL VERIFICATION

Verification distinguishes identities that should hold to floating-point accuracy from finite optimization targets. The numerical checks evaluate loss–gain complements, analytic derivatives, admissible score bounds, chronology, compositional sums and penalty normalization. Numerical agreement is evidence about the implementation; it is not a substitute for the proofs.

TABLE 32. Stopping and interpretation checklist. All residual quantities use the stated standardized coordinates unless marked otherwise. Iteration counts below are permitted upper limits, not numbers necessarily executed.

Quantity	Setting or check	Interpretation
Inner feasible gap	$\mathcal{G} \leq 2 \times 10^{-7}$ in the PMM experiments; penalty-aware bound in the MM cross-check	Upper bound on inner objective error
Parameter step	$\ d\ \leq 2 \times 10^{-6}$; at most 120 primary PMM or 200 polynomial cross-check steps	Finite stopping, not exact termination
STAT5 original search	Original seven-start protocol: at most 240 MM steps for each of the two selected candidates; one additional descriptive $q = 2$ polishing step	Original protocol only; Section 9.11
STAT5 expanded search	At most six cycles of 400 MM iterations, hence 2,400 per refinement run; separately, three continuation levels of at most 600 L-BFGS-B iterations each	Equal allowances for self- and cross-initialization; Section 9.12
Projected gradient	Primary continuation records $\ z - \Pi(z - \nabla J^\delta)\ _\infty$; best-fit refinement above 10^{-4}	Stationarity diagnostic for the smooth box problem
Residual tube	$\tau = 10^{-4}$, or $s_j \tau$ in physical class units	Allowed backward perturbation of responses
LP residual	Witness projected to exact score/normal bounds, then independently evaluated	Local stationarity residual for the declared tube
Smoothing	$\delta = 10^{-6}$, giving $W\delta/c = 6.67 \times 10^{-7}$ at $c = 1.5, W = 1$	Uniform objective perturbation, not parameter error
ODE discretization	Measured $\hat{e}_{i,h}, \hat{E}_{i,h}, \hat{B}_h$ for GLDM and $\hat{B}_{q,h}$ for inverse power; enlarged tube and recorded slack	Numerical verification at the tested point; rigorous application of Theorem 8.1 requires validated bounds
January ODE	$\tau = 10^{-5}$, $\epsilon_R = 10^{-4}$, $\epsilon_y = 2 \times 10^{-5}$; three RK4 meshes; nine starts; at most 500 iterations per smoothing level and 240 MM iterations	Numerical transfer checks with analytic references; all selected meshes pass these budgets; search-bound sensitivity reported
Joint constraints	Nonnegative coefficients and column sums checked independently	Compositional feasibility, not global optimality

The continuation scales are dimensionless because residuals are standardized first. Decreasing δ reduces the uniform objective error but increases the curvature bound $L_q + 1/(c\delta)$ and can worsen conditioning near interpolated



observations. A smaller smoothing level should therefore be accompanied by a stationarity check and, when relevant, a repeated-start or unsmoothed MM comparison. The cross-check in Table 55 shows that basin selection can dominate the tiny theoretical smoothing error. No general parameter stability follows without an additional local growth or identifiability condition.

Independent checks integrate polynomial derivatives, test constant-slope invariance and degree-elevation monotonicity, and verify chronological cutoffs. They compare every Bernstein derivative through degree five against a power-basis representation, test nonlinear-model and transfer inequalities (including two necessary-enlargement examples), and recompute Eq. (C.6) with $1/n\tau$ for every fit. Machine-readable checks and raw table values accompany the implementation; finite checks do not replace proofs.

11.1. Computational reproducibility. The numerical record includes fixed seeds, starting points, selected estimates, predictions, stopping diagnostics and exact enclosure endpoints. The January ODE records additionally retain all componentwise map/sensitivity discrepancies, LP scores, normal witnesses, numerical transfer slacks and mesh decisions, with the boundary-sensitivity fits reported separately. Independent checks recompute the twelve scalar transfer inequalities from rational inputs. Environment specifications and a table-to-record inventory accompany the computational implementation, allowing the reported quantities to be traced to their inputs. The scalar geometry illustrations use the deterministic observations stated in Eq. (A.27) and introduce no additional field measurements. Software access and data provenance are described in the Data and code availability statement.

A. SUPPORTING OPTIMIZATION GEOMETRY AND ENERGY CALCULATIONS

This internal appendix collects mathematical support that is not required to check the pointwise transfer theorem.

A.1. Momentum balance, energy stability and a two-state envelope. For a single translational degree of freedom, Newton's momentum balance is

$$\frac{d}{dt}(M_s \dot{u}) = F_{\text{ext}}(t) - D_s \dot{u} - K_s u, \quad M_s, D_s, K_s > 0. \quad (\text{A.1})$$

Here M_s is a constant lumped mass, D_s a viscous damping coefficient and K_s a linear stiffness. Displacement is measured from static equilibrium, so a constant gravitational force is absorbed into that equilibrium. The free-decay benchmark sets $F_{\text{ext}} = 0$. It assumes small displacement, a fixed support and linear constitutive forces; it omits external excitation, nonlinear stiffness, dry friction, state-dependent damping and spatially distributed inertia. These are modeling assumptions, not terms discarded by the estimator. Define $\omega = \sqrt{K_s/M_s}$ and $\kappa_d = D_s/(2\sqrt{M_s K_s})$, divide by M_s and set $v = \dot{u}$. The resulting underdamped model is

$$\dot{u} = v, \quad \dot{v} = -2\kappa_d \omega v - \omega^2 u, \quad (u(0), v(0)) = (1, 0), \quad 0 < \kappa_d < 1, \quad \omega > 0. \quad (\text{A.2})$$

It is a two-dimensional state system and is nonlinear in the two physical parameters through $\kappa_d \omega$ and ω^2 . Writing $\theta = (\kappa_d, \omega)$, define

$$A_\theta = \begin{pmatrix} 0 & 1 \\ -\omega^2 & -2\kappa_d \omega \end{pmatrix}, \quad a_\theta = \kappa_d \omega, \quad b_\theta = \omega \sqrt{1 - \kappa_d^2}, \quad C_\theta = 1 + \|A_\theta + a_\theta I\| / b_\theta. \quad (\text{A.3})$$

Its first state has the closed form

$$u(t; \theta) = e^{-a_\theta t} \left[\cos(b_\theta t) + \frac{\kappa_d}{\sqrt{1 - \kappa_d^2}} \sin(b_\theta t) \right]. \quad (\text{A.4})$$

Here ω is natural frequency and κ_d is damping ratio. Observations identify these two combinations, not three independent mass, damping and stiffness constants. Displacement is normalized by initial amplitude and time uses benchmark units. The state-linear oscillator is nonlinear in its calibration parameters. It models a controlled vibration experiment, not January vegetation. Steering, predictive-control, thermal and environmental applications motivate calibration [11, 22, 48, 49], but their datasets are not validated here.



Proposition A.1 (Continuous and discrete energy stability). Fix $0 < \kappa_d < 1$ and $\omega > 0$ in Eq. (A.2). Let $x = (u, v)^\top$, $G_\theta = \text{diag}(\omega^2, 1)$ and $E_\theta(x) = x^\top G_\theta x / 2$, the mechanical energy per unit mass. Then

$$\frac{d}{dt} E_\theta(x(t)) = -2\kappa_d \omega v(t)^2 \leq 0, \quad x(t) \longrightarrow 0. \quad (\text{A.5})$$

For fixed $h > 0$, set $R_h = R_4(hA_\theta)$, where $R_4(z) = 1 + z + z^2/2 + z^3/6 + z^4/24$. Define

$$\gamma_\theta(h) = \left\| G_\theta^{1/2} R_h G_\theta^{-1/2} \right\|_2^2, \quad \varrho_\theta(h) = |R_4(h(-a_\theta + ib_\theta))|. \quad (\text{A.6})$$

Then $E_\theta(R_h x) \leq \gamma_\theta(h) E_\theta(x)$ for every x . Energy is nonincreasing at every step for every initial state if and only if $\gamma_\theta(h) \leq 1$. Every discrete solution tends to zero if and only if $\varrho_\theta(h) < 1$. The energy and spectral criteria are distinct, and neither certifies parameter accuracy.

Proof. Differentiating E and using Eq. (A.2) gives $\omega^2 uv + v(-\omega^2 u - 2\kappa_d \omega v)$, proving the identity. The eigenvalues $-a_\theta \pm ib_\theta$ have negative real parts, so the continuous solution decays. Writing $z = G_\theta^{1/2} x$ makes the energy ratio a squared norm ratio for $G_\theta^{1/2} R_h G_\theta^{-1/2}$. Its supremum over $z \neq 0$ is exactly $\gamma_\theta(h)$, proving both directions of the energy assertion. The spectral mapping theorem gives the conjugate eigenvalues $R_4(h(-a_\theta \pm ib_\theta))$ of R_h . A fixed matrix power tends to zero for every initial state exactly when its spectral radius is below one, proving the last assertion. $\square$

Theorem A.2 (Combined parameter and RK4 forecast bound). Let $\hat{\theta}$ and $\theta_\star$ both be underdamped, let $h > 0$, $N \in \{1, 2, \dots\}$ and $t = Nh$, and use the common initial state $x_0 = (1, 0)^\top$. Set $R_4(z) = 1 + z + z^2/2 + z^3/6 + z^4/24$, $\hat{r}_h = |R_4(h(-\hat{a} + i\hat{b}))|$, and

$$K_h = e^h \|A_{\hat{\theta}}\| (h \|A_{\hat{\theta}}\|)^5 / 120, \quad H(t) = \begin{cases} (e^{-a_\star t} - e^{-\hat{a}t}) / (\hat{a} - a_\star), & \hat{a} \neq a_\star, \\ te^{-a_\star t}, & \hat{a} = a_\star. \end{cases} \quad (\text{A.7})$$

Then the computed RK4 state satisfies

$$\begin{aligned} \|x_h(t; \hat{\theta}) - x(t; \theta_\star)\| &\leq C_{\hat{\theta}} C_{\theta_\star} \|A_{\hat{\theta}} - A_{\theta_\star}\| H(t) \\ &\quad + C_{\hat{\theta}}^2 N K_h \max(\hat{r}_h, e^{-\hat{a}h})^{N-1}. \end{aligned} \quad (\text{A.8})$$

The same bound applies to the error in the observed first coordinate.

Proof. Let $B_\theta = A_\theta + a_\theta I$. Since $B_\theta^2 = -b_\theta^2 I$, the matrix exponential is $e^{A_\theta t} = e^{-a_\theta t} [\cos(b_\theta t) I + \sin(b_\theta t) B_\theta / b_\theta]$. Its norm is at most $C_\theta e^{-a_\theta t}$. Duhamel's identity for the difference of two matrix exponentials gives

$$e^{A_{\hat{\theta}} t} - e^{A_{\theta_\star} t} = \int_0^t e^{A_{\hat{\theta}}(t-s)} (A_{\hat{\theta}} - A_{\theta_\star}) e^{A_{\theta_\star} s} ds. \quad (\text{A.9})$$

The norm bound integrates to the first term of Eq. (A.8). The fourth-degree Taylor remainder gives $\|R_4(hA_{\hat{\theta}}) - e^{hA_{\hat{\theta}}}\| \leq K_h$. Every polynomial in $A_{\hat{\theta}}$ is a linear combination of I and $B_{\hat{\theta}}$. Taking powers therefore gives $\|R_4(hA_{\hat{\theta}})^j\| \leq C_{\hat{\theta}} \hat{r}_h^j$. The telescoping identity for the difference of the N th powers, combined with the corresponding exponential bound, yields

$$\|R_4(hA_{\hat{\theta}})^N - e^{NhA_{\hat{\theta}}}\| \leq C_{\hat{\theta}}^2 K_h \sum_{j=0}^{N-1} \hat{r}_h^{N-1-j} e^{-\hat{a}hj} \leq C_{\hat{\theta}}^2 N K_h \max(\hat{r}_h, e^{-\hat{a}h})^{N-1}. \quad (\text{A.10})$$

Multiplication by x_0 , whose norm is one, and the triangle inequality complete the proof. A coordinate error cannot exceed the full-state norm. $\square$

This deterministic envelope separates parameter mismatch from integration error. It uses the generating parameter only when evaluated as a simulation benchmark; it is not an empirical confidence interval or a usable unknown-truth guarantee for field data. When $\hat{r}_h < 1$, the discretization contribution has the expected fourth-order behavior on a fixed horizon. Its usefulness deteriorates near critical damping, where the chosen C_θ bound becomes conservative.



A.2. Reciprocal calculus and algorithmic equivalence.

A.2.1. *Calculus, stationarity, and stable evaluation.* Writing $\chi(u) = \arctan(c/u)$ for $u > 0$ and $\chi(0) = \pi/2$ gives

$$\chi'(u) = -\frac{c}{c^2 + u^2}, \quad \chi''(u) = \frac{2cu}{(c^2 + u^2)^2}, \quad \frac{d\psi_c(r)}{dr} = -\frac{c \operatorname{sgn} r}{c^2 + r^2} \quad (r \neq 0). \quad (\text{A.11})$$

Thus χ is decreasing and convex on $[0, \infty)$, whereas ψ_c has a downward cusp at zero. Its scalar Clarke set at zero is again $[-1/c, 1/c]$. In contrast to the loss, the gain is not Clarke regular at the cusp: its ordinary directional derivative there is $-|d|/c$, while its Clarke directional derivative is $|d|/c$. Optimization of $-I$ retains the regular loss calculus.

Proposition A.3 (Stationarity and certificates for the gain). *Under Definition 2.1,*

$$\partial_C I(\theta) = -\partial_C F(\theta), \quad 0 \in -\partial_C I(\theta) + N_\Theta(\theta) \iff 0 \in \partial_C F(\theta) + N_\Theta(\theta). \quad (\text{A.12})$$

The residual-tube LP, backward-data interpretation, and solver-transfer bound in Sections 7–8 therefore apply to reciprocal GLDM with unchanged numerical constants and tube radii.

Proof. The Clarke subdifferential commutes with multiplication by -1 and addition of a constant. Apply this fact to $I = \pi W/2 - F$. A constrained maximum of I is treated as a constrained minimum of $-I$, which fixes the normal-cone sign in Eq. (A.12). The score sets of $-I$ are exactly those of F . The same equality holds after enlarging scalar score intervals or replacing an observation map by its discretization. Hence the LP and the stated bounds are identical. $\square$

In particular, a small reciprocal GLDM certificate is not a global-maximization certificate. Example 2.6 becomes a stationary local minimum of the gain. For numerical evaluation, use

$$\psi_c(r) = \operatorname{atan2}(c, |r|). \quad (\text{A.13})$$

With $c > 0$, this expression directly returns the value $\pi/2$ at zero and avoids forming $c/|r|$. It also avoids the cancellation in $\pi/2 - \arctan(|r|/c)$ when the residual is very large. The January experiment evaluates the loss and the gain through separate formulas and records their complement discrepancy.

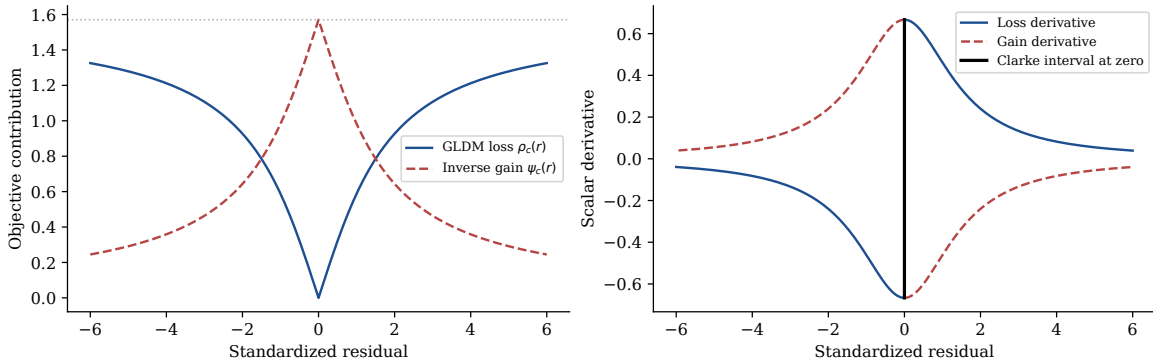


FIGURE 25. GLDM loss and reciprocal GLDM gain for $c = 1.5$. Their sum is constant and their derivatives have opposite signs. The common vertical interval at zero is a Clarke set, not an ordinary derivative.

Figure 25 shows why minimizing the loss and maximizing the gain rank all residuals in the same way. The tails have equal derivative magnitudes and therefore do not provide different outlier resistance.



A.2.2. *Minorization and convergence equivalence.* Convexity of χ implies, for all $u, v \geq 0$,

$$\chi(v) \geq \chi(u) - \frac{c}{c^2 + u^2}(v - u). \quad (\text{A.14})$$

At θ^k , let $w_i^k = a_i c / (c^2 + (r_i^k)^2)$. For affine g , a tight proximal minorizer of I is

$$M_k(z) = I(\theta^k) - \sum_i w_i^k (|r_i(z)| - |r_i^k|) - \frac{\mu_k}{2} \|z - \theta^k\|^2. \quad (\text{A.15})$$

It is strongly concave on the convex parameter set. Maximizing it is exactly the weighted-LAD minimization in Eq. (4.1). For nonlinear g , use $r_i^k - J_i^k d$ in the absolute value and accept a feasible step when

$$I(\theta^k + d) \geq I(\theta^k) - \sum_i w_i^k (|r_i^k - J_i^k d| - |r_i^k|) - \frac{\mu_k - \sigma}{2} \|d\|^2. \quad (\text{A.16})$$

Corollary A.4 (Equality of the exact iterates). *With the same initial point and proximal parameters, exact affine GLDM majorization and reciprocal GLDM minorization generate identical iterates. Their sufficient-decrease and sufficient-increase statements satisfy*

$$I(\theta^{k+1}) - I(\theta^k) = F(\theta^k) - F(\theta^{k+1}) \geq \mu_k \|\theta^{k+1} - \theta^k\|^2. \quad (\text{A.17})$$

The affine convergence and finite-termination theorems apply under their existing hypotheses. For nonlinear predictors, identical backtracking and inner-solve choices yield the same accepted sequence and the same shrinking-tube bounds. The inexact convergence result applies with the same feasible primal-dual gaps.

Proof. Dropping constants in $-M_k$ gives the strongly convex affine subproblem, whose minimizer is unique. Induction gives equality of the iterates. Eq. (3.2) converts Theorem 4.1 into Eq. (A.17). In the nonlinear case, substitution of the same identity into Eq. (A.16) gives precisely Eq. (5.2). The convex subproblem and its dual gap have not changed, so Theorems 5.2 and 6.3 apply without an additional argument about gain regularity. $\square$

Algorithm 8 Reciprocal GLDM with a loss-gain consistency check.

Require: Fixed a_i, c , observation map (g, J) , feasible start, proximal and inner tolerances.

```

1: for  $k = 0, 1, \dots$  do
2:   Evaluate  $r^k = y - g(\theta^k)$ ,  $F_k$ , and  $I_k = \sum_i a_i \operatorname{atan2}(c, |r_i^k|)$ .
3:   Check  $|F_k + I_k - \pi W/2|$  against a stated floating-point tolerance.
4:   Form  $w_i^k = a_i c / (c^2 + (r_i^k)^2)$  and solve the weighted-LAD proximal subproblem.
5:   Record the independently computed feasible primal-dual gap.
6:   For nonlinear  $g$ , increase  $\mu_k$  until Eq. (A.16) holds.
7:   Update  $\theta^{k+1}$  and record the gain increase, step, and residual-tube certificate.
8:   if the specified step, inner-gap, and certificate budgets are met then
9:     return the fit and its accuracy record.
10:  end if
11: end for
```

A.2.3. *Smooth approximation and the zero-residual limit.* Smoothing is optional and is not used for the degree-one exhaustive January fits in Section C. The higher-order extension in Section D uses the continuation rule stated there. For comparison with smooth optimization software, let $v_\delta(r) = \sqrt{r^2 + \delta^2}$, $\delta > 0$, and define

$$F_\delta(\theta) = \sum_i a_i \arctan\left(\frac{v_\delta(r_i(\theta))}{c}\right), \quad I_\delta(\theta) = \sum_i a_i \operatorname{atan2}(c, v_\delta(r_i(\theta))). \quad (\text{A.18})$$



Proposition A.5 (Uniform approximation and limiting stationarity). *For every θ ,*

$$F_\delta + I_\delta = \pi W/2, \quad 0 \leq F_\delta - F = I - I_\delta \leq W\delta/c. \quad (\text{A.19})$$

For a nonempty compact Θ and continuous g_i , suppose $F_\delta(\theta_\delta) \leq \min_\Theta F_\delta + \epsilon_\delta$. Then

$$0 \leq F(\theta_\delta) - \min_\Theta F \leq \epsilon_\delta + W\delta/c. \quad (\text{A.20})$$

If instead θ_δ are constrained stationary points of F_δ and $\delta \downarrow 0$, every accumulation point is constrained Clarke stationary for F , under Definition 2.1.

Proof. The first identity is the same scalar complement formula. Since $0 \leq v_\delta(r) - |r| \leq \delta$ and arctangent of u/c is increasing and $1/c$ -Lipschitz, summation gives the uniform bound. At a minimizer θ^* of F ,

$$F(\theta_\delta) \leq F_\delta(\theta_\delta) \leq F_\delta(\theta^*) + \epsilon_\delta \leq F(\theta^*) + W\delta/c + \epsilon_\delta. \quad (\text{A.21})$$

For stationarity, the smooth scalar loss derivative is

$$\xi_\delta(r) = \frac{cr}{v_\delta(r)[c^2 + v_\delta(r)^2]}, \quad |\xi_\delta(r)| \leq 1/c. \quad (\text{A.22})$$

At a nonzero limiting residual this derivative converges to the corresponding singleton in Eq. (2.6). At a zero limiting residual every subsequential limit belongs to $[-1/c, 1/c]$. The stationarity normals equal the bounded score–Jacobian product, so they also admit convergent subsequences. Continuity of J , the exact calculus of Proposition 2.3, and the closed graph of N_Θ give the asserted limiting inclusion. $\square$

For the scalar smooth-loss derivative ξ_δ in Eq. (A.22), differentiation also gives

$$\left| \frac{d\xi_\delta(r)}{dr} \right| \leq L_\phi + \frac{1}{c\delta}. \quad (\text{A.23})$$

Indeed, the derivative is $\phi''(v_\delta)(r/v_\delta)^2 + \phi'(v_\delta)\delta^2/v_\delta^3$, and the two terms have the stated bounds. A smaller δ controls objective approximation but can worsen conditioning near zero. To keep the uniform objective perturbation below ϵ_F , it suffices to set $\delta \leq c\epsilon_F/W$. This does not convert a finite- δ stationary point into an exact stationary point of the unsmoothed objective. The two problems must be distinguished in a numerical report.

A.2.4. Loss–gain geometry and paired optimization. For an affine model $g_i(\theta) = x_i^\top \theta$, the reciprocal identity provides a direct geometric interpretation. Write $C = \pi W/2$ and, for a common penalty coefficient $\lambda \geq 0$, define $F_\lambda = F + \lambda P$ and $I_\lambda = I - \lambda P$. Then

$$I_\lambda(\theta) = C - F_\lambda(\theta), \quad (\text{A.24})$$

$$\{\theta : I_\lambda(\theta) \geq h\} = \{\theta : F_\lambda(\theta) \leq C - h\},$$

$$I_\lambda(v) - I_\lambda(u) = F_\lambda(u) - F_\lambda(v). \quad (\text{A.25})$$

Thus reflecting the loss surface in the plane of height $C/2$ gives the gain surface. The same parameter path descends on one and ascends on the other. Every gain superlevel set is a loss sublevel set after changing its level. This explains equality of both the fitted values and within-criterion objective gaps, even when a numerical run reaches only a local solution.

Inside a residual cell where no observation is interpolated and all residual signs are fixed,

$$\nabla^2 I(\theta) = \sum_i \frac{2a_i c |r_i(\theta)|}{(c^2 + r_i(\theta)^2)^2} x_i x_i^\top \succeq 0. \quad (\text{A.26})$$

To verify the formula, differentiate each nonzero-residual gain twice and use $\nabla r_i = -x_i$. Hence the gain is convex within each cell, although it is not globally convex or concave across the downward cusps. A peak in the gain surface can occur where residual-zero lines meet. Under the full-rank and interior assumptions of Theorem 2.4, a local gain maximum must interpolate enough independent rows to span the parameter space. A smooth hill-shaped maximization picture therefore misses the geometry that motivates weighted-LAD steps.



Consider the following specified illustration with $\theta = (\beta_0, \beta_1)$, $g_i(\theta) = \beta_0 + \beta_1 x_i$, $a_i = 1/9$ and $c = 0.35$:

$$\begin{aligned} x &= (-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2), \\ y &= (-1.48, -1.04, -0.47, 0.03, 0.48, 1.04, 1.47, 4.20, 2.52). \end{aligned} \quad (\text{A.27})$$

These deterministic observations illustrate one discrepant response and are not a new statistical benchmark. With $\theta^0 = (-0.25, 0.45)$, $\mu = 0.4$, no penalty and no parameter constraints, the affine proximal update gives the trace in Table 33. The gain is evaluated independently by $\text{atan2}(c, |r|)$, rather than by subtracting the loss from C . The two surfaces in Figure 26 carry the same recorded parameter path. Their apparent ascent and descent are the same optimization operation.

TABLE 33. Ordinary loss–gain optimization for the specified nine-point illustration. Both formulations use the same subproblem, solved to the reported inner-gap accuracy. The last update has step norm 2.76×10^{-10} . This is one numerical trace, not a comparison of estimator risks.

k	β_0	β_1	$F(\theta^k)$	$I(\theta^k)$
0	-0.250000	0.450000	1.018063	0.552733
1	0.280000	0.880000	0.579488	0.991308
2	0.520000	1.000000	0.216809	1.353987
3	0.520000	1.000000	0.216809	1.353987

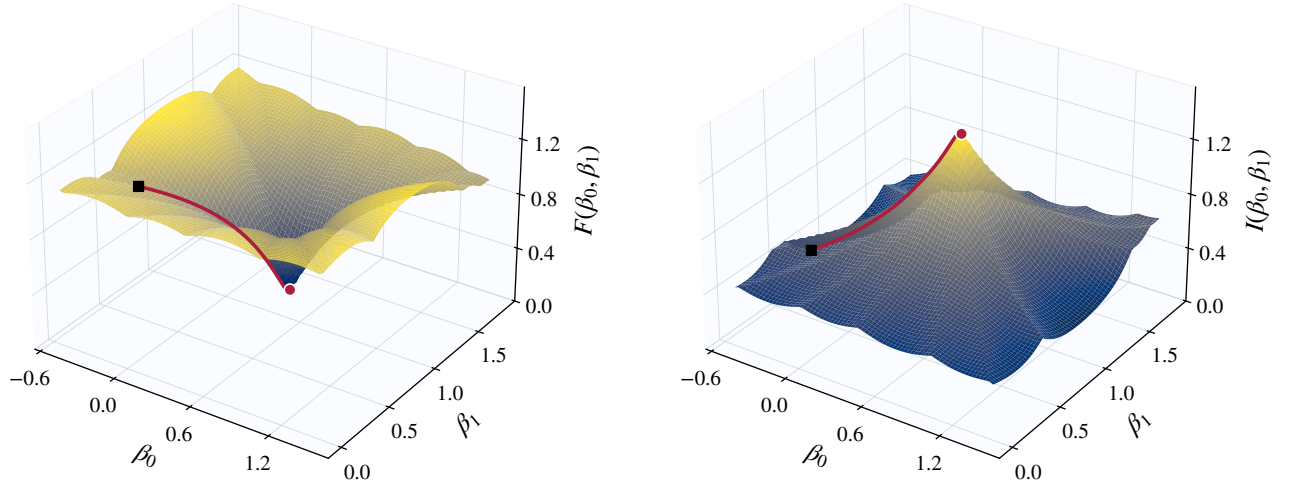


FIGURE 26. Ordinary GLDM loss F (left) and reciprocal GLDM gain I (right) for Eq. (A.27). The surfaces satisfy $F + I = \pi/2$. The red curves evaluate each objective along straight parameter segments joining the same recorded iterates; the black square marks the initial point and the red circle the last point. The illustration concerns $q = 1$ only and does not compare prediction risks.

The proximal loss majorizer and gain minorizer can also be displayed as one paired construction. At θ^k , define

$$Q_k(z) = F(\theta^k) + \sum_i w_i^k (|r_i(z)| - |r_i(\theta^k)|) + \frac{\mu_k}{2} \|z - \theta^k\|^2, \quad (\text{A.28})$$

$$M_k(z) = C - Q_k(z), \quad Q_k(z) - F(z) = I(z) - M_k(z) \geq 0. \quad (\text{A.29})$$



For affine residuals, Q_k is strongly convex and M_k strongly concave on a convex feasible set. Concavity of $u \mapsto \arctan(u/c)$ gives $Q_k \geq F$, and the complement identity gives $M_k \leq I$. Both touch their target at θ^k . Subtracting the complement identities proves equality of the two nonnegative approximation gaps in Eq. (A.29). Figure 27 shows the common touching point. This pairing also explains why an inner primal–dual gap and a nonlinear acceptance test can be reused after switching from loss to gain.

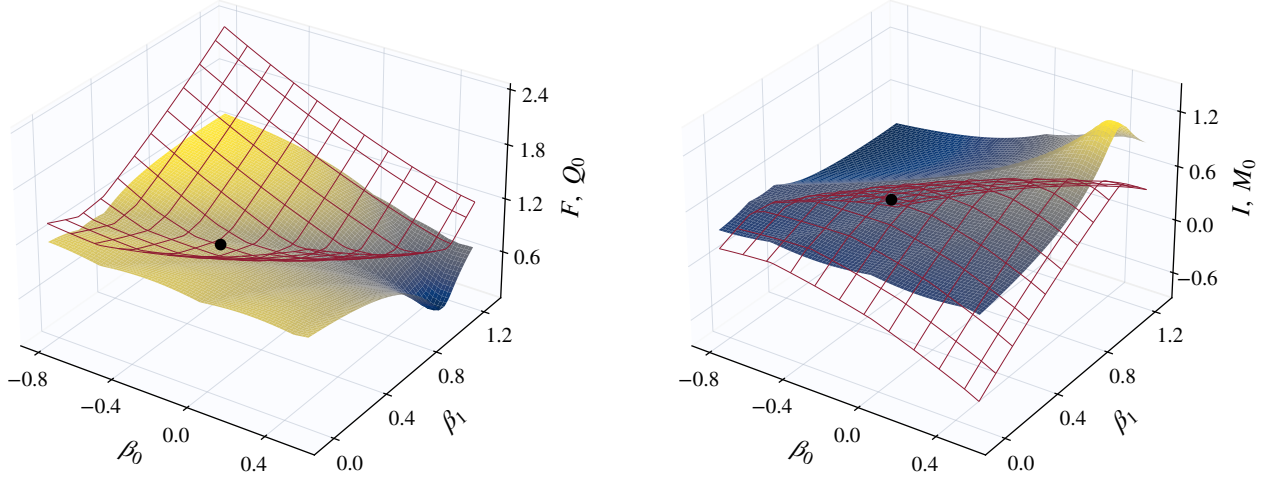


FIGURE 27. Paired proximal models for reciprocal GLDM at $\theta^0 = (-0.25, 0.45)$. The colored loss surface lies below the red wireframe majorizer (left), while the colored gain surface lies above its red wireframe minorizer (right). The black point marks their common contact in parameter coordinates. Both approximation gaps are identical by Eq. (A.29); wireframe overlap in a perspective view is not a sign change of that gap.

In this illustration the maximum independently evaluated inner gap is 1.94×10^{-10} , and the maximum sampled loss–gain complement discrepancy is 4.45×10^{-16} . These are floating-point checks of the construction. The mathematical inequality follows from concavity, rather than from a plotted grid. The trace illustrates a local numerical stopping decision and does not supply a global-optimality certificate.

A.2.5. Scale interpretation for reciprocal GLDM. At a nonzero residual, the gain and its score magnitude obey

$$\frac{\partial \psi_c(r)}{\partial c} = \frac{|r|}{c^2 + r^2} \geq 0, \quad \left| \frac{\partial \psi_c(r)}{\partial r} \right| = \frac{c}{c^2 + r^2}. \quad (\text{A.30})$$

At $r = 0$, the gain remains $\pi/2$ for every $c > 0$ and the residual derivative is replaced by a Clarke interval. Figure 28 shows the gain ridge and the positive score envelope. The envelope value $1/c$ at the ridge is the largest admissible slope magnitude, not an ordinary derivative there. Increasing c raises the gain at a fixed nonzero residual, while decreasing c concentrates sensitivity nearer zero.

The scale cannot be selected by maximizing the unadjusted training gain over $c > 0$. For any fixed parameter with finite residuals,

$$\lim_{c \rightarrow \infty} I(\theta; c) = \frac{\pi W}{2}, \quad \lim_{c \rightarrow \infty} F(\theta; c) = 0. \quad (\text{A.31})$$

Each scalar term has the displayed limit, so a finite weighted sum has it as well. If any residual is nonzero, the first derivative in Eq. (A.30) makes the gain strictly increasing with c at that parameter. Unrestricted scale enlargement

would therefore make every fixed fit approach the same gain ceiling. The empirical procedure instead fixes a training-derived scale within each fit and assesses scale choices through training-only validation. This is a property of reciprocal GLDM itself and does not require a power transformation.

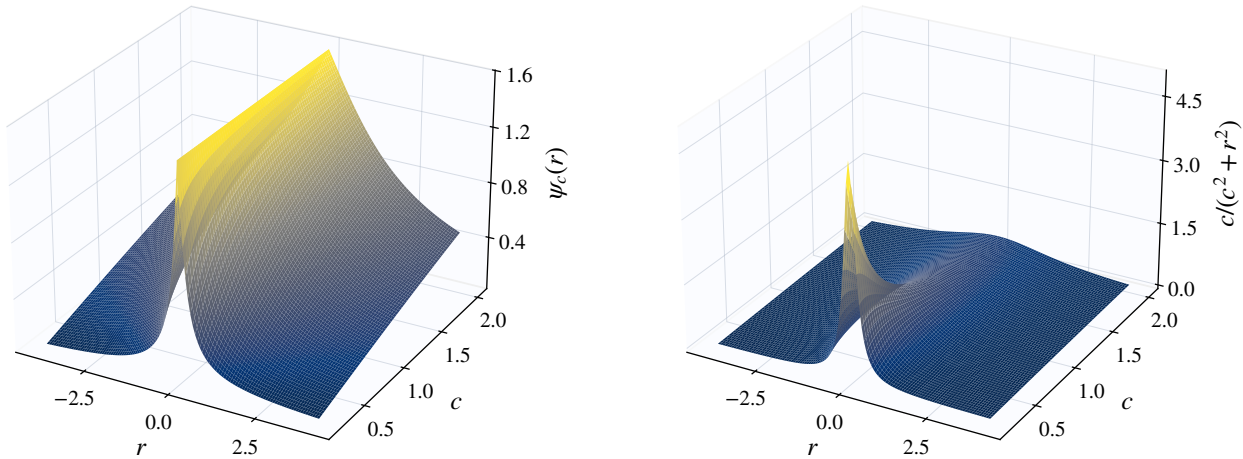


FIGURE 28. Reciprocal GLDM gain across residual and scale (left) and its nonzero-residual score-magnitude envelope (right), for $-4 \leq r \leq 4$ and $0.2 \leq c \leq 2$. The gain equals $\pi/2$ along $r = 0$. The envelope reaches $1/c$ there, but the gain has a cusp and no ordinary residual derivative at zero. The two surfaces explain scale choice without introducing an inverse-power estimator.

The relevant ODE guarantee remains Theorem 8.1. A certificate for maximizing the ordinary gain is obtained by applying that theorem to $-I = F - C$, where $C = \pi \sum_i a_i/2$ with $q = 1$. Its response tube and state/sensitivity allowances are unchanged. Equivalent objective pictures do not remove numerical integration error; they make the same transfer decision accessible from either optimization convention.

B. SUPPORTING LOSS-SHAPE CALCULATIONS

The exponent q and the scale c act differently. At $u = c$, every q uses the same residual measured in the same units, but the ratio of its weight to the ordinary-GLDM weight is 2^{1-q} . Increasing c instead changes where a residual lies relative to that reference magnitude. The polynomial degree m in the January study changes the trajectory model and is independent of both choices. Table 34 describes loss exponents one to five, not five polynomial orders.

TABLE 34. Exact scalar effects of the inverse-gain exponent at a fixed scale. All rows have cusp slope $1/c$. The relative weight compares the same residual $u = c$; the tail column omits a positive constant.

q	$\ell_{q,c}(c)$	Saturation A_q	$w_q(c)/w_1(c)$	$c^2 \ell''_{q,c}(0+)$	Tail score
1	$\pi/4$	$\pi/2$	1	0	u^{-2}
2	$3\pi/16$	$\pi/4$	$1/2$	$-2/\pi$	u^{-3}
3	$7\pi/48$	$\pi/6$	$1/4$	$-4/\pi$	u^{-4}
4	$15\pi/128$	$\pi/8$	$1/8$	$-6/\pi$	u^{-5}
5	$31\pi/320$	$\pi/10$	$1/16$	$-8/\pi$	u^{-6}

The first two propositions below make the interpretation precise. They are consequences of the scalar family and common standardization, rather than separate claims of a new general optimization principle. The last proposition explains a regularization effect that can otherwise be mistaken for a benefit of faster redescending.

Proposition B.1 (Relative score attenuation). *Fix $c > 0$, and let $q_1, q_2 \in \mathcal{Q}$ with $q_2 \geq q_1$. For every finite $u \geq 0$,*

$$\frac{w_{q_2}(u)}{w_{q_1}(u)} = G_c(u)^{q_2 - q_1}, \quad 0 \leq \ell_{q_2, c}(u) \leq \ell_{q_1, c}(u), \quad (\text{B.1})$$

$$\ell_{q, c}(c) = \frac{\pi}{2q}(1 - 2^{-q}), \quad \frac{w_q(c)}{w_1(c)} = 2^{1-q}. \quad (\text{B.2})$$

If $q_2 > q_1$ and $u > 0$, the weight ratio is strictly less than one and the loss inequality is strict. These pointwise inequalities do not order fitted parameters or forecast risks.

Proof. Cancel the common factor $c/(c^2 + u^2)$ in Eq. (3.8). Since $0 < G_c(u) \leq 1$, the weights are ordered. Both losses vanish at zero, so integration of the derivative inequality over $[0, u]$ gives the loss inequality. For $u > 0$ the weights are strictly ordered throughout $(0, u]$ when the exponents differ. Finally $G_c(c) = 1/2$ yields Eq. (B.2). The two-residual ordering example in Theorem 3.3 shows that pointwise loss ordering does not preserve sample minimizers; no forecast-risk ordering follows from it. $\square$

For the three-residual example in Eq. (3.4), changing to $q = 2$ changes the weighted-LAD coefficients to

$$\frac{1}{3c} \left(1, \frac{1}{4}, \frac{G_c(3c)}{10} \right), \quad G_c(3c) = \frac{2}{\pi} \arctan(1/3) \simeq 0.20483. \quad (\text{B.3})$$

The moderate residual c loses half of its previous weight, while the residual $3c$ retains about one fifth. The transformation therefore affects informative moderate residuals as well as extreme ones. The near-zero curvature changes even though the cusp slope stays fixed. Figure 29 shows these effects in dimensionless coordinates.

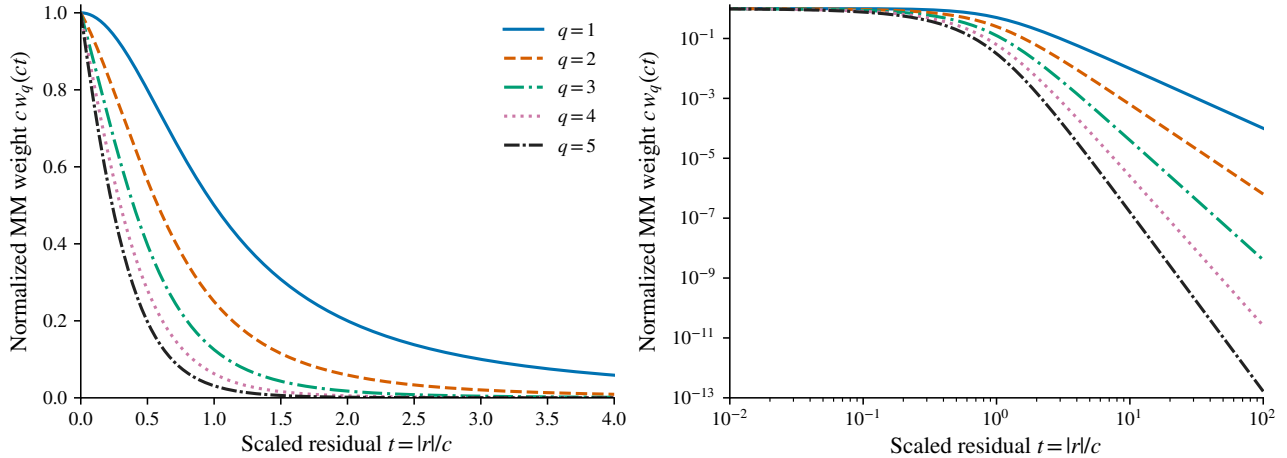


FIGURE 29. Normalized weighted-LAD coefficients $c w_q(ct)$ versus $t = |r|/c$ for $q = 1, \dots, 5$. The left panel resolves the behavior near zero, and the right panel shows algebraic tail decay on logarithmic axes. All curves start at the same cusp slope, but larger q downweights moderate as well as large residuals. These are exact scalar-function profiles, not estimation-risk comparisons.

Proposition B.2 (Invariance under a common change of observation units). *Fix $c > 0$, $q \in \mathcal{Q}$, weights $a_i > 0$, coefficient $\lambda \geq 0$ and a feasible parameter θ in a nonempty closed convex set Θ . Assume both observation maps and the convex penalty P are continuously differentiable on a neighborhood of Θ , and use positive proximal coefficients. Let $b > 0$. Multiply every response and both observation maps by b , replace c by bc , and multiply the response tube*



by b . Keep the weights, penalty, parameter coordinates and constraints unchanged. Then the loss, gain and penalized objectives, and their exact tube-stationarity residuals, are unchanged. If the original map and Jacobian bounds are $e_{i,h}, E_{i,h}, Q_i$, the transformed bounds $be_{i,h}, bE_{i,h}, bQ_i$ give the same $B_{q,h}$ in Theorem 8.1. With the same feasible start and proximal coefficients, exact affine or linearized weighted-LAD updates are unchanged.

Proof. For $u \geq 0$,

$$G_{bc}(bu) = G_c(u), \quad \ell_{q,bc}(bu) = \ell_{q,c}(u), \quad w_{q,bc}(bu) = b^{-1}w_{q,c}(u), \quad L_{q,bc} = b^{-2}L_{q,c}. \quad (\text{B.4})$$

Here the second subscript on w and L displays their scale explicitly. These identities follow by substitution. The transformed score set, including the zero interval, is b^{-1} times the original set, and the transformed Jacobian is bJ . Membership in a response tube is preserved because $|br_i| \leq b\tau_i$ exactly when $|r_i| \leq \tau_i$. Hence the score–Jacobian products, penalty gradient and normal cone entering the stationarity residual agree. The error terms satisfy

$$\frac{bE_{i,h}}{bc} + (bQ_i)(b^{-2}L_{q,c})(be_{i,h}) = \frac{E_{i,h}}{c} + Q_iL_{q,c}e_{i,h}. \quad (\text{B.5})$$

Finally each weighted absolute residual in the affine or linearized subproblem is unchanged because its weight is divided by b and its residual is multiplied by b . The common penalty and proximal terms are identical. Strict convexity of the exact subproblem gives equality of the updates. $\square$

This invariance concerns a common unit conversion with matched scales. It does not cover changing a pilot-scale estimator, adapting weights during a fit, or rescaling different observations while keeping an incompatible common c . For example, converting percentages to proportions and dividing the residual scale and tube by 100 leaves the criterion unchanged. Dividing the observations alone changes the effective loss shape and invalidates such a comparison. To preserve a smoothed criterion as well, the smoothing radius must be changed to $b\delta$.

Proposition B.3 (Fixed-penalty effect at large gain powers). *Let Θ be nonempty and compact, all residual maps and P continuous on Θ , and fix $a_i, c > 0$ and $\lambda > 0$. Let $q \rightarrow \infty$ through $\mathcal{Q}$. Suppose*

$$J_{q,\lambda}(\theta_q) \leq \min_{\theta \in \Theta} J_{q,\lambda}(\theta) + \epsilon_q, \quad \epsilon_q \geq 0. \quad (\text{B.6})$$

Writing $P_{\min} = \min_{\Theta} P$, one has

$$0 \leq P(\theta_q) - P_{\min} \leq \frac{\pi W}{2\lambda q} + \frac{\epsilon_q}{\lambda}. \quad (\text{B.7})$$

If $\epsilon_q \rightarrow 0$, every accumulation point of θ_q minimizes P on Θ .

Proof. Continuity and compactness give a minimizer θ_P of P and a minimizer of $J_{q,\lambda}$. The scalar range gives

$$\lambda P(\theta_q) \leq J_{q,\lambda}(\theta_q) \leq J_{q,\lambda}(\theta_P) + \epsilon_q \leq \lambda P_{\min} + \frac{\pi W}{2q} + \epsilon_q. \quad (\text{B.8})$$

Rearrangement proves the bound. Compactness gives accumulation points; continuity of P and the bound identify their penalty value as $P_{\min}$ when $\epsilon_q \rightarrow 0$. $\square$

This is a conditional near-global comparison, not a guarantee for the local fits in the experiments. It nevertheless explains the normalization: preserving the cusp slope does not preserve the total loss range. At the same λ , increasing q makes the penalty larger relative to the maximum data-loss contribution. Multiplying the entire criterion by $1/A_q$ gives the equivalent expression

$$\frac{J_{q,\lambda}(\theta)}{A_q} = \sum_i a_i \{1 - G_c(|r_i(\theta)|)^q\} + \frac{2q\lambda}{\pi} P(\theta). \quad (\text{B.9})$$

One can instead hold λ/A_q fixed across powers to match this saturation-based penalty ratio, but that defines a different comparison and does not ensure equal efficiency or prediction. The common- λ January ablation remains useful because it separates adding a penalty from changing the exponent under its declared convention.



Selection and reporting. Start with a scale defined from the training data, keep it fixed within each optimization, and compare candidate powers under an explicit penalty convention. Historical forecast validation can choose among them without looking at the later holdout. Record the selected scale, exponent, penalty, starting points and numerical tolerances. Report reciprocal GLDM as the same estimator when its conditions match, and report inverse-power GLDM as a distinct loss choice. Neither a smaller raw objective across different powers nor a faster-decaying scalar score establishes better prediction. The within-criterion refinements and held-out results are the relevant comparisons.

C. JANUARY DATA, MAPS AND DRIFT-MODEL FORECASTS

This section documents the supplied inputs and the complete drift-model analysis summarized in Section 10. Maps are spatial context only; the forecasting calculations start at the recorded five-class counts.

C.1. January data and preprocessing. The satellite-derived NDVI dataset covers Khanty-Mansiysk during 2000–2025. Data preparation comprised NDVI calculation from MODIS remote-sensing imagery, image correction and combination, monthly mapping for January, May, July and October, and NDVI-threshold classification. The source satellite images have been retained but are not included in the computational archive. The resulting observations are satellite-derived measurements, not simulated data. The forecasting analysis uses only the January worksheet of the supplied workbook. The accompanying cartographic series contains one January map for each usable year, 2001–2025; no January 2000 map enters the study. This worksheet records annual pixel counts, nominal areas, and percentages for five NDVI intervals, but does not contain the original rasters or individual-pixel NDVI observations. Consequently, the response variable is a January *class percentage*, not mean NDVI or an independently validated land-cover class.

For red and near-infrared reflectance, the underlying index is

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}}. \quad (\text{C.1})$$

MODIS vegetation-index products include quality information relevant to clouds and snow/ice [16]. This is particularly important when interpreting winter observations. The workbook does not specify product and collection identifiers, acquisition lists, the spatial clipping mask, or the exact monthly compositing and quality-mask rules. The existence of the source imagery does not supply these omitted processing choices. The forecasting analysis begins with the tabulated class statistics and does not repeat the image processing. Table 36 distinguishes the documented data origin from the parts of the processing record not contained in the archive. Threshold labels are retained verbatim, without inventing an endpoint-inclusion rule or assigning ecological names to the intervals. The nine map-display intervals are not a documented refinement of the five workbook classes: the map groups 0.7–1.0 together while the modeled classes split at 0.8, and the negative ranges also differ. Without pixel values, endpoint rules and a shared processing record, their class counts cannot be reconciled by simply adding the displayed colors. The maps supply context only and do not verify the modeled totals.

The sheet lists 2000–2025. In 2000 all five counts and areas are zero and all percentages are absent. This row is excluded as missing, without interpolation. This is also consistent with the published start of the Terra MOD13Q1 record in February 2000 [17], although the workbook alone does not establish that particular product as its source. The analysis therefore has 25 January compositions, covering 2001–2025. No values from May, July, or October are read or fitted. For year t and class j , preprocessing recomputes

$$N_t = \sum_{j=1}^5 N_{tj}, \quad p_{tj} = 100 \frac{N_{tj}}{N_t}, \quad \sum_{j=1}^5 p_{tj} = 100 \quad (N_t > 0). \quad (\text{C.2})$$

The largest discrepancy from the supplied percentages is 7.11×10^{-15} percentage points. The classified-pixel total ranges from 20,916,001 to 20,980,504. Percentages are conditional on those classified pixels, rather than on a verified fixed geographic area.

The supplied areas satisfy $A_{tj} = 0.0625 N_{tj}$, consistent arithmetically with 250 m square pixels. Without the original grid and mask, nominal totals of 1.307–1.311 million km² are not independently verified Khanty-Mansiysk areas. Only percentages are analyzed; no land-area or land-cover-change claim is made.



TABLE 35. January class percentages over 25 usable years, 2001–2025. The final column counts years with zero pixels in that class.

Class	NDVI interval	Mean (%)	Minimum (%)	Maximum (%)	Zeros
1	−0.2 to 0.0	36.643	26.926	48.577	0
2	0.0 to 0.2	55.928	49.415	63.365	0
3	0.2 to 0.5	7.363	1.961	15.124	0
4	0.5 to 0.8	0.065929	0.000052	0.491059	0
5	0.8 to 1.0	0.000334	0.000000	0.002179	13

The first two classes dominate (Table 35). Class 5 is zero in 13 of 25 years; valid class zeros are retained without pseudo-counts. The sampling unit is one annual composition, whose five components are dependent, not millions of independent pixels.

TABLE 36. Data lineage of the January MODIS application. Missing archive items limit processing reproducibility without changing the observational origin of the measurements.

Component	Recorded status
Origin	NDVI processing of MODIS satellite images for Khanty-Mansiysk; source images retained outside the computational archive.
Processing and months	NDVI calculation, image correction/combination, monthly mapping and threshold classification; January, May, July and October prepared, January analyzed here.
Analysis input	Annual class counts, nominal areas and percentages; 25 nonzero January rows from 2001–2025 after excluding the incomplete 2000 row.
Threshold categories	Five recorded NDVI intervals; threshold endpoint conventions are not specified in the workbook.
Cartographic exports	Twenty-five January vector map exports for 2001–2025, displayed in Figures 30–32. Their supplied legend has nine display intervals, distinct from the five forecasting classes.
Items outside this archive	Product and collection IDs, scene/acquisition inventory, red/NIR layer names and scaling, projection and resampling, clipping polygon, quality/cloud/snow masks, mosaic/composite settings and executable processing record.
Supported inference	Comparisons of forecasts of the derived January class compositions; no independent validation of preprocessing accuracy or attribution of ecological change.

Figures 30–32 contain all 2001–2025 maps and the original nine-interval legend. Those displayed bins differ from the five workbook classes, particularly below zero and at the upper boundary. The colors therefore cannot identify model responses one-to-one. Numerical results use recorded counts and Eq. (C.2), without re-digitizing map totals.

Blue, orange and pale-yellow areas provide spatial context but cannot verify the tabulated five-class totals. Regional share forecasts do not locate individual patches. Supplied orientation and extent are retained, with external year labels. These cartographic exports supply neither georeferenced distances nor areas; interannual colors do not distinguish snow, surface conditions, quality screening, compositing and vegetation effects.

C.2. A bounded drift model and a finite global search. With only 25 annual observations, a low-dimensional empirical ODE avoids an unsupported high-order model. For each class, fit

$$\dot{z}_j(t) = b_j, \quad z_j(t_0) = \alpha_j, \quad z_j(t) = \alpha_j + b_j(t - t_0), \quad t_0 = 2001. \quad (\text{C.3})$$



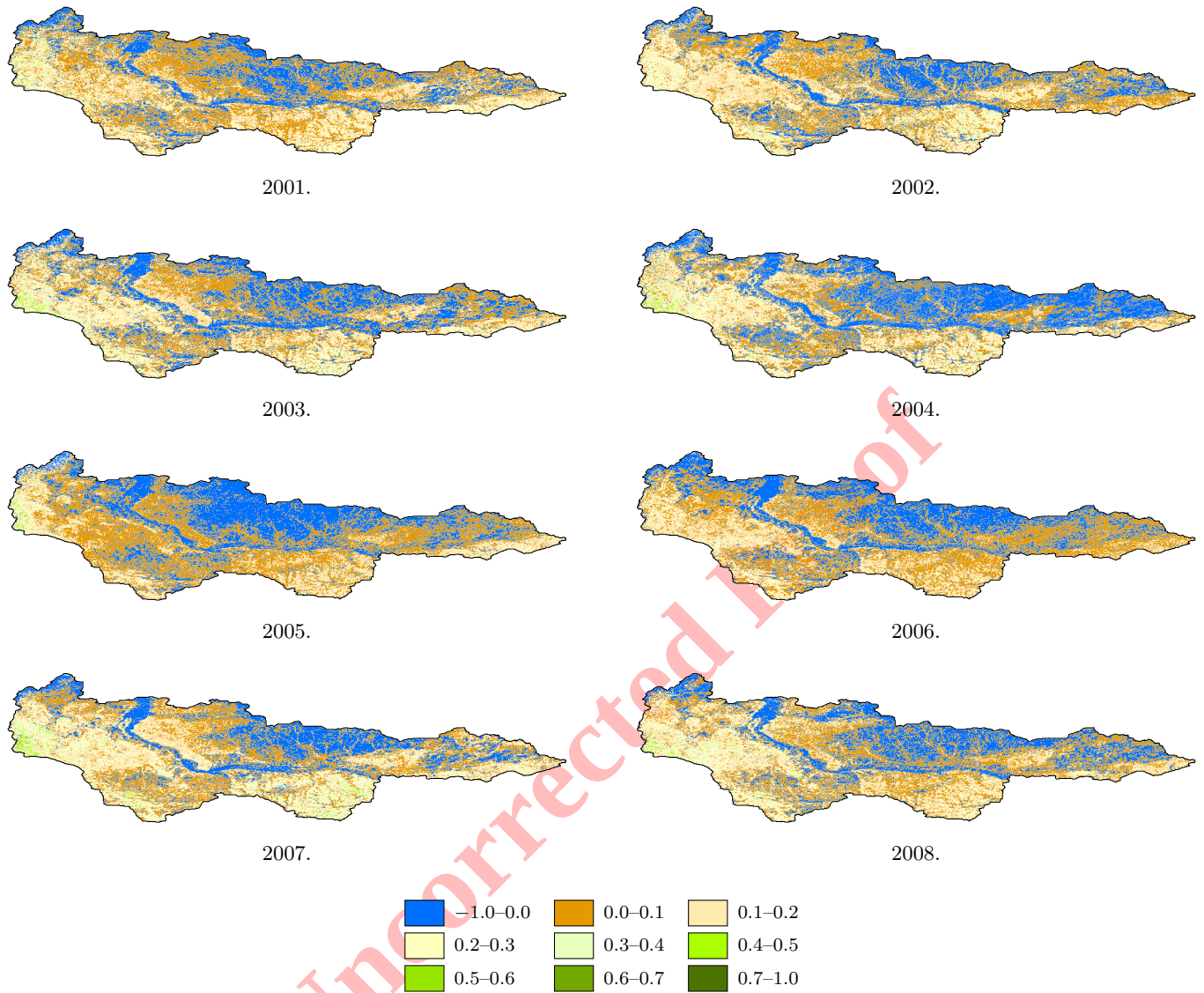


FIGURE 30. January NDVI map exports for 2001–2008 in the reported Khanty-Mansiysk study region. The common nine-interval legend differs from the five modeled classes. These exports provide spatial context only and cannot verify the numerical class totals.

For a specified evaluation horizon T , parameterize the endpoints by $\theta_j = (\alpha_j, \gamma_j) \in [0, 100]^2$, where

$$\zeta_t = \frac{t - t_0}{T - t_0}, \quad z_j(t) = (1 - \zeta_t)\alpha_j + \zeta_t\gamma_j, \quad b_j = \frac{\gamma_j - \alpha_j}{T - t_0}. \quad (\text{C.4})$$

The box guarantees a percentage between zero and 100 over the stated horizon, including the forecast years. Knowing their calendar dates does not use their response values. This is a descriptive drift model, not a mechanistic model of vegetation, snow, or land-use transitions. Its solution is analytic, so no numerical ODE discretization error enters this application.



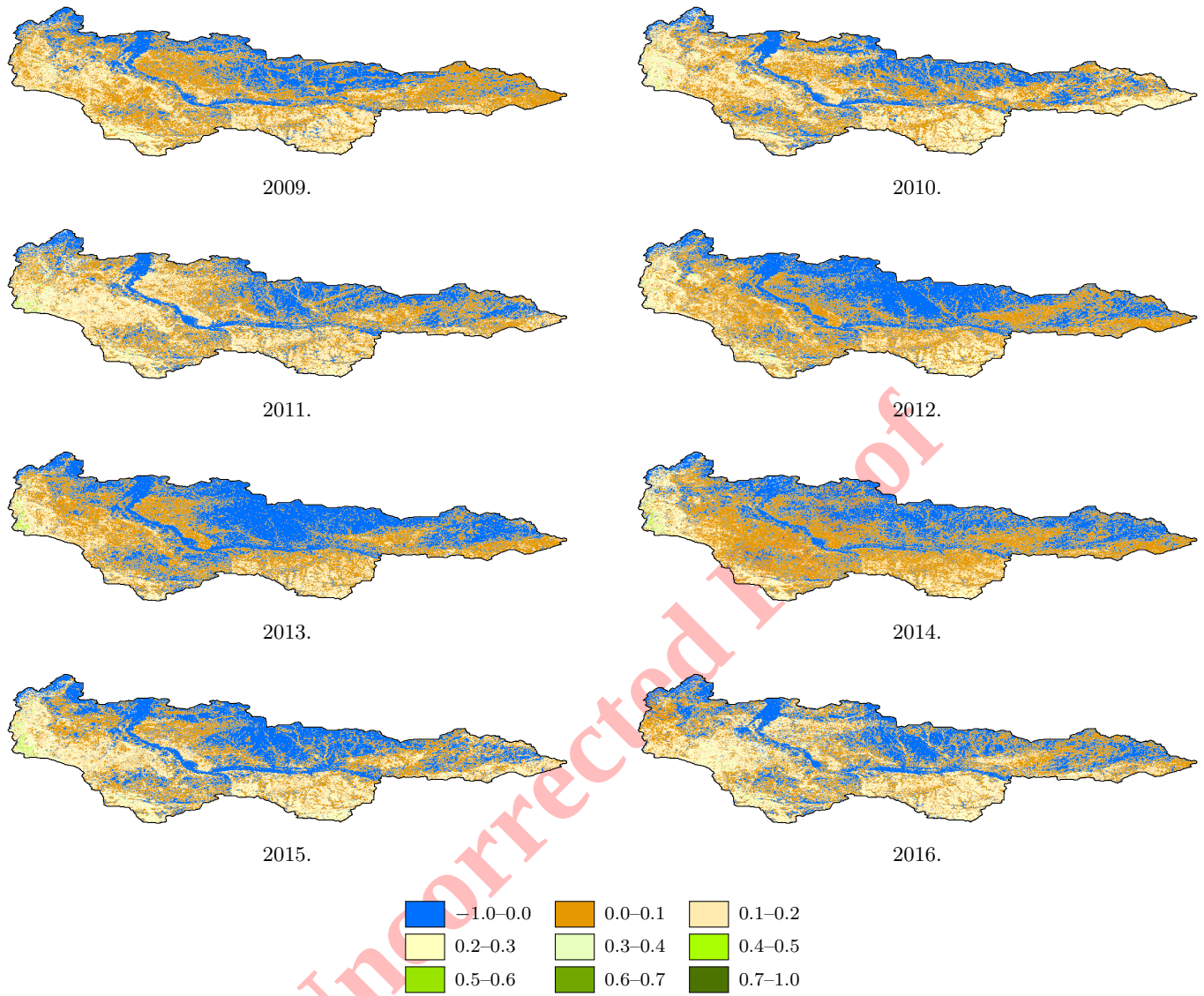


FIGURE 31. January NDVI map exports for 2009–2016 in the reported Khanty-Mansiysk study region. The supplied orientation and nine-interval color key are retained. These exports provide spatial context only and cannot verify the five modeled class totals.

For each training sample and class, first compute the box-constrained LAD fit. If e_{tj}^{LAD} are its residuals in percentage points, fix

$$s_j = \max \{0.01, 1.4826022185 \text{ median}_t |e_{tj}^{\text{LAD}} - \text{median}_u e_{uj}^{\text{LAD}}|\} . \quad (\text{C.5})$$

The floor is a stated numerical regularization in percentage points, not an estimated sensor uncertainty. It prevents the almost-always-zero class from producing a vanishing normalization.

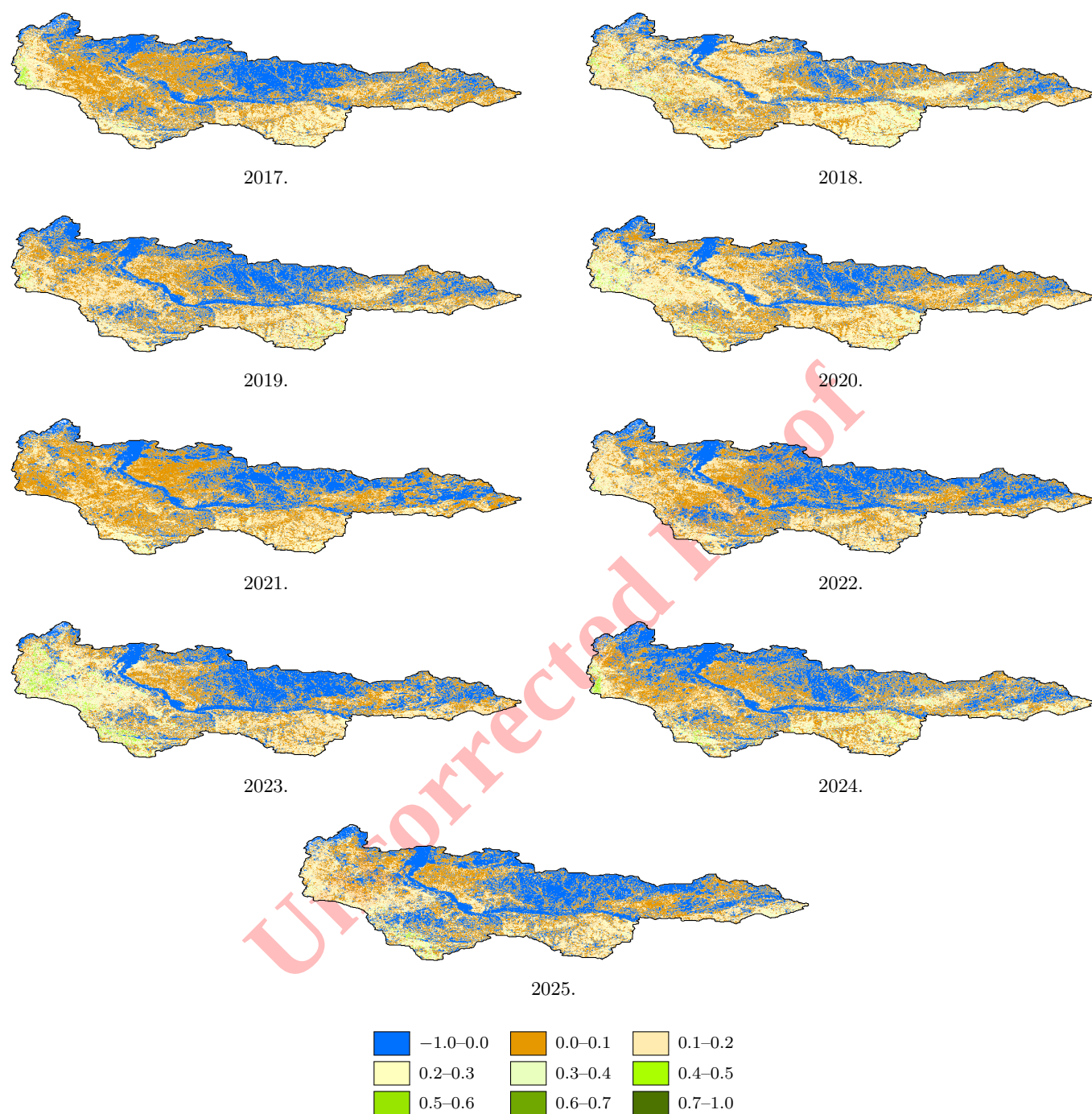


FIGURE 32. January NDVI map exports for 2017–2025 in the reported Khanty-Mansiysk study region. Years 2021–2025 form the primary chronological holdout. The panels provide spatial context, not spatial forecasts or verification of the five modeled class totals.

With $n_{\mathcal{T}} = |\mathcal{T}|$ training years, the two objectives are

$$\begin{aligned} F_j(\theta_j) &= \frac{1}{n_{\mathcal{T}}} \sum_{t \in \mathcal{T}} \arctan \left(\frac{|p_{tj} - z_j(t; \theta_j)|}{cs_j} \right), \\ I_j(\theta_j) &= \frac{1}{n_{\mathcal{T}}} \sum_{t \in \mathcal{T}} \operatorname{atan2} \left(c, \frac{|p_{tj} - z_j(t; \theta_j)|}{s_j} \right). \end{aligned} \quad (\text{C.6})$$

Here $a_t = 1/n_{\mathcal{T}}$, $W = 1$, and $F_j + I_j = \pi/2$. The two implementations use the same fixed s_j and c .

Proposition C.1 (Residual-cell enumeration on a polytope). *Let g_i be affine and let Θ be a nonempty compact polytope. Subdivide Θ by all hyperplanes $r_i(\theta) = 0$. Some minimizer of F lies at a vertex of the resulting subdivision, and some maximizer of I lies at the same vertex. For Eq. (C.4), it suffices to examine feasible intersections of pairs drawn from the $n_{\mathcal{T}}$ residual lines and the four boundary lines of $[0, 100]^2$.*

Proof. Within each closed residual-sign cell, every $|r_i|$ is an affine nonnegative function. Concavity of ϕ therefore makes F concave on that cell. Every point of a bounded polyhedral cell is a convex combination of its vertices. Concavity implies that at least one of those vertices has objective no larger than the point. There are finitely many cells, and continuity gives existence on compact Θ . Minimizing over their vertices is therefore sufficient. The loss-gain identity proves the assertion for I . In two dimensions, every vertex is an intersection of two linearly independent active residual or boundary lines; degenerate multiple intersections are included by any such pair. $\square$

There are at most $\binom{n_{\mathcal{T}}+4}{2}$ line pairs. Generating candidates costs $O(n_{\mathcal{T}}^2)$ and direct evaluation of each objective at all candidates costs $O(n_{\mathcal{T}}^2)$. This finite search is practical here because each class has only two parameters. It is not a claim of scalable global optimization for the nonlinear ODE problems in Section 9. Floating-point line intersections and objective evaluations are used, so the implementation is a numerical exhaustive search supported by the proposition, not an interval-arithmetic certificate.

Algorithm 9 January-only comparison of GLDM and reciprocal GLDM.

Require: The JAN counts, a training cutoff, forecast years, and a fixed c .

- 1: Exclude zero-total years and recompute the five class percentages.
 - 2: **for** each class $j = 1, \dots, 5$ **do**
 - 3: Construct residual lines from training years and the four endpoint-box boundary lines.
 - 4: Enumerate feasible pairwise intersections, including all box corners.
 - 5: Compute the LAD pilot and freeze the training-only scale in Eq. (C.5).
 - 6: Select a minimum of the independently evaluated F_j and a maximum of I_j .
 - 7: Record their endpoints, objective complement discrepancy, and forecast differences.
 - 8: **end for**
 - 9: Normalize the five nonnegative forecasts using Eq. (C.7).
 - 10: Compare with held-out observations and retain every year-by-class error.
-

Candidate objective ties within 64 machine epsilons are resolved by training absolute error and then by lexicographic endpoint order. The two objective branches use the same declared tie rule. An independent linear-programming LAD solution agrees with its enumerated objective to within 4.89×10^{-15} percentage points across the 225 calibration comparisons. These checks support implementation consistency without substituting for the mathematical proof.

Independent class fits need not sum to 100. For all methods, the reported composition forecast is

$$\hat{p}_j(t) = 100 \frac{\hat{z}_j(t)}{S(t)}, \quad S(t) = \sum_{\ell=1}^5 \hat{z}_\ell(t) > 0. \quad (\text{C.7})$$

Every evaluated forecast has a positive total. This explicit normalization enforces nonnegativity and closure. It is post-processing of five separate fits, not minimization of a coupled compositional GLDM objective. Its induced



differential equation is

$$\dot{\hat{p}}_j(t) = 100 \frac{b_j S(t) - \hat{z}_j(t) \sum_{\ell} b_{\ell}}{S(t)^2}, \quad \sum_j \dot{\hat{p}}_j(t) = 0. \quad (\text{C.8})$$

Both formulations yield identical raw forecasts and hence identical normalized forecasts. Raw endpoints and raw as well as normalized predictions are retained in the accompanying CSV files.

C.3. Chronological evaluation and numerical results. The primary split fits 2001–2020 and withholds all five January observations from 2021–2025. A separate expanding-window evaluation fits 2001– t and predicts $t + 1$, for $t = 2015, \dots, 2024$. These ten one-year forecasts preserve time ordering. Their last five outcomes overlap the fixed holdout, so the two evaluations are not independent replications and are not pooled. Neither future values nor other months enter a pilot, fit, or initial condition.

The primary arctangent setting is $c = 1.5$, as in the synthetic study. A prespecified sensitivity grid $c \in \{0.5, 1, 1.5, 3\}$ changes only that constant and does not replace the primary result with the best test-set setting. Comparators are box-constrained OLS, box-constrained LAD, the expanding historical mean, and persistence of the last observed composition. All predictions receive the same normalization in Eq. (C.7). If $\mathcal{H}$ contains H forecast years, the reported metrics in percentage points are

$$\begin{aligned} \text{MAE} &= \frac{1}{5H} \sum_{t \in \mathcal{H}} \sum_{j=1}^5 |\hat{p}_{tj} - p_{tj}|, \\ \text{RMSE} &= \left[\frac{1}{5H} \sum_{t \in \mathcal{H}} \sum_{j=1}^5 (\hat{p}_{tj} - p_{tj})^2 \right]^{1/2}, \quad \overline{\text{TV}} = \frac{1}{2H} \sum_{t \in \mathcal{H}} \sum_j |\hat{p}_{tj} - p_{tj}|. \end{aligned} \quad (\text{C.9})$$

The last quantity is total-variation error expressed in percentage points. If N years and five classes are scored, $\text{TV}_{\text{pp}} = (2N)^{-1} \sum_{t,j} |e_{t,j}| = (5/2)(5N)^{-1} \sum_{t,j} |e_{t,j}| = 2.5 \text{MAE}$. For proportions rather than percentage points, divide this TV value by 100. Class-specific errors accompany the aggregate scores. Percentage-relative metrics such as MAPE are omitted because some observations are exactly zero. No independent-pixel confidence intervals or Monte Carlo standard errors are assigned to this single observed time series.

TABLE 37. Raw class-drift parameters fitted to 2001–2020, with endpoint horizon 2025 and $c = 1.5$. Reciprocal GLDM gives the same values. Endpoints are percentages and slopes are percentage points per year.

Class	Scale s_j	α_j (2001)	γ_j (2025)	b_j
1	4.431791	31.011033	42.280739	0.469571
2	3.398528	58.530491	53.638491	-0.203833
3	3.302815	9.127153	5.491478	-0.151486
4	0.034685	0.002486	0.063719	0.002551
5	0.01	0	1.64×10^{-5}	6.82×10^{-7}

Table 37 gives the fitted raw drift parameters and frozen scales for the primary training period. The increasing raw class-1 trend and decreasing class-2 trend describe this calibration window only. They do not establish a physical change mechanism or a significance claim. Class 5 activates the scale floor. Its numerical parameter estimates therefore should not be interpreted as precise estimates of a rare NDVI-class component.

Figure 33 places the fitted trend and forecast against every usable January observation. The 2023 increase in class 4 is visible rather than removed as an outlier. Without the original quality layers, an unusual winter observation cannot be attributed to vegetation change or to a processing artifact. The plot shows model mismatch and temporal variation that remain after calibration.



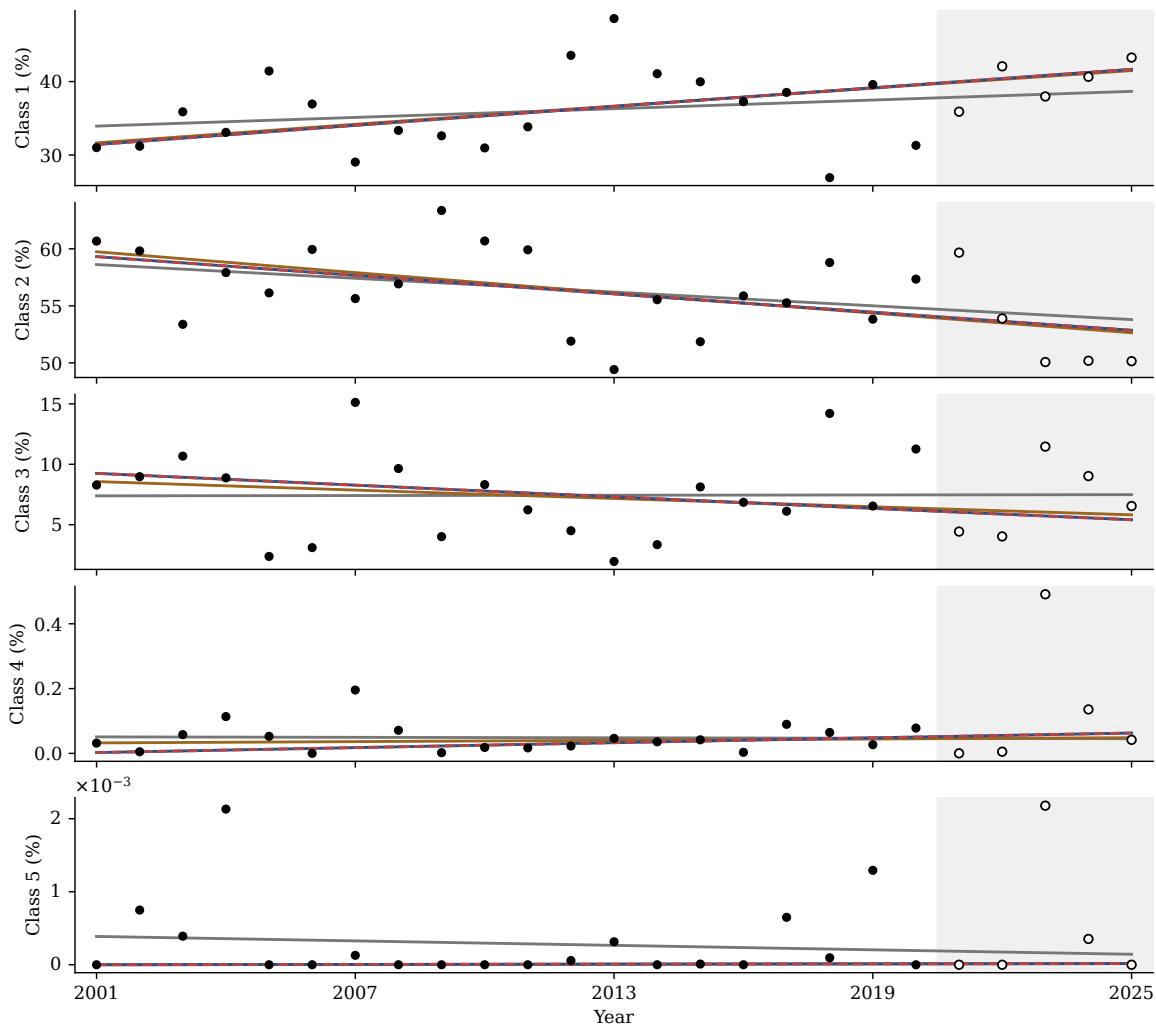


FIGURE 33. January NDVI-class percentages and normalized model trajectories. Filled points are fitting observations and open points are withheld observations. The shaded interval is 2021–2025. GLDM and reciprocal GLDM curves coincide. Gray curves show OLS, ochre curves LAD, blue curves GLDM, and red dashed curves its identical inverse form. Separate panel scales preserve visibility of rare classes.

At $c = 1.5$, GLDM and its reciprocal representation both have holdout MAE 1.608 and RMSE 2.394 percentage points. OLS has MAE 1.752 and RMSE 2.507, while LAD has the smaller MAE 1.568. For rolling forecasts, GLDM has MAE 2.036, OLS 2.131 and the historical mean 2.018. Table 38 retains all six implementations. The reciprocal row checks the loss–gain identity rather than adding an independent estimator.

Figure 34 shows that the ranking depends on the evaluation window. The lower holdout errors do not imply an improvement over every baseline, and agreement between GLDM and reciprocal GLDM is expected from Theorem 3.1 rather than evidence of two independent successes.

The full scale grid (Table 39; Figure 35) favors $c = 0.5$ descriptively. Selecting it from test errors would invalidate the holdout; the primary $c = 1.5$ remains fixed. Operational selection requires rolling validation within training years, with the chosen scale and rare-class floor reported before forecasting.

TABLE 38. January forecast errors in percentage points after composition normalization. The GLDM scale is fixed at $c = 1.5$.

Method	Holdout MAE	Holdout RMSE	Rolling MAE	Rolling RMSE
Mean	2.394	3.520	2.018	3.194
Persistence	3.703	5.332	2.574	4.111
OLS	1.752	2.507	2.131	3.525
LAD	1.568	2.332	2.046	3.363
GLDM	1.608	2.394	2.036	3.373
Reciprocal GLDM	1.608	2.394	2.036	3.373

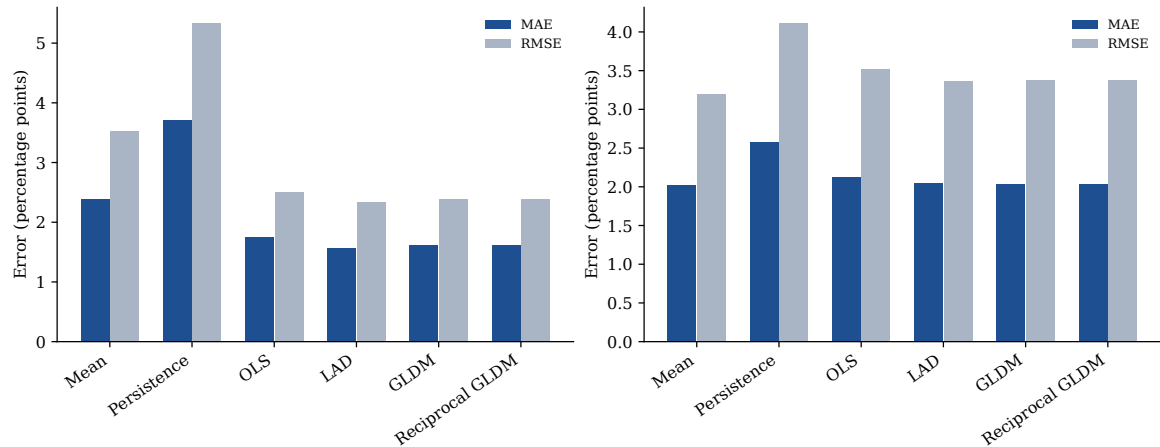


FIGURE 34. January forecast errors at $c = 1.5$. The fixed holdout contains five years and the rolling evaluation contains ten one-year forecasts. Bars are observed scores, not means over independent simulated datasets. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

TABLE 39. Scale sensitivity for GLDM and reciprocal GLDM. Both formulations give every value in this table. Errors are percentage points.

c	Holdout MAE	Holdout RMSE	Rolling MAE	Rolling RMSE
0.5	1.581	2.365	1.934	3.156
1	1.608	2.394	1.997	3.284
1.5	1.608	2.394	2.036	3.373
3	1.608	2.394	2.089	3.409

Four scales, five classes, one holdout and ten rolling origins give 220 paired calibrations; five full-period fits bring the equivalence audit to 225 pairs, all from the same 25 compositions. Every pair selects identical endpoints. The largest loss-plus-gain discrepancy is 8.89×10^{-16} . Independent forward and `atan2` formulas, finite-search proof and stored parameters support numerical equivalence.

January drift runs used an Intel Xeon Platinum 8573C (nine CPUs available), Python 3.12.14, NumPy 2.3.5 and SciPy 1.17.0, serially with one BLAS/OpenMP thread. Matplotlib 3.10.8 records regenerate vector plots. The XML



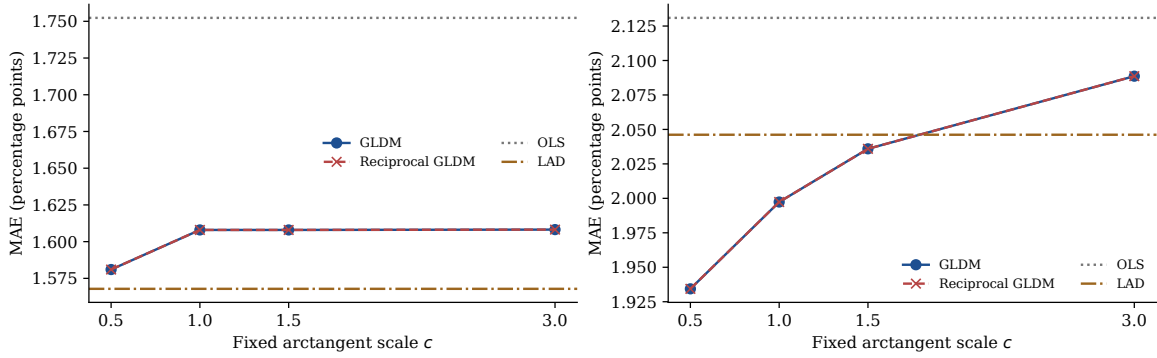


FIGURE 35. January scale sensitivity with all other settings fixed. The two arctangent implementations coincide at every tested scale. Horizontal reference lines show OLS and LAD. This is a sensitivity analysis, not test-set selection of the reported default.

reader selects only JAN; a CSV permits workbook-free reproduction. Separate synthetic and sensitivity environments are recorded in Sections 9 and E. No uncertainty-coverage, causal-vegetation or preprocessing-accuracy claim follows.

D. JANUARY POLYNOMIAL COMPARISONS AT DEGREES ONE TO FIVE

D.1. Order, constraints, and the differential-equation interpretation. This extension keeps the January data, exclusion rule, thresholds, and primary $c = 1.5$ from Section C. It compares all five polynomial degrees, without choosing a favorable degree from the held-out errors. Table 40 presents every usable annual composition. The term order below means polynomial degree, not the NDVI-class index, an autoregressive lag, or the inverse-gain exponent. The five class labels retain the intervals in Table 35; no additional land-cover attribution is introduced.

For class j , degree $m \in \{1, \dots, 5\}$, start year $t_0 = 2001$, and a specified final forecast year T , use

$$\begin{aligned} \zeta(t) &= (t - t_0)/(T - t_0), & B_{k,m}(\zeta) &= \binom{m}{k} \zeta^k (1 - \zeta)^{m-k}, \\ z_{j,m}(t; \beta_j) &= \sum_{k=0}^m \beta_{jk} B_{k,m}(\zeta(t)), & 0 \leq \beta_{jk} &\leq 100. \end{aligned} \quad (\text{D.1})$$

The forecast horizon is a calendar choice known before observing its outcomes. It enters the basis and constraints, not the responses used for calibration. The scale s_j is estimated from the training degree-one LAD pilot in Eq. (C.5), shared across all m, q, λ within the split, and frozen before fitting. Thus a higher degree cannot obtain a different loss merely by reducing its own pilot scale. The coefficient penalty is

$$P_{j,m}(\beta_j) = \frac{m}{2s_j^2} \sum_{k=0}^{m-1} (\beta_{j,k+1} - \beta_{jk})^2. \quad (\text{D.2})$$

It shrinks rapid changes in the fitted curve and also penalizes slope when $m = 1$. It does not force any individual year to be discarded.

Proposition D.1 (Bounded polynomial dynamics and variation control). *Let $m \geq 1$ be an integer, $T > t_0$, $s_j > 0$, and $0 \leq \beta_{jk} \leq 100$. For $\zeta \in [0, 1]$, Eq. (D.1) satisfies $z_{j,m} \in [0, 100]$ and*

$$\frac{d^\nu z_{j,m}}{dt^\nu} = \frac{m!}{(m-\nu)!(T-t_0)^\nu} \sum_{k=0}^{m-\nu} \Delta^\nu \beta_{jk} B_{k,m-\nu}(\zeta), \quad \nu \in \{0, \dots, m\}. \quad (\text{D.3})$$

TABLE 40. All usable January NDVI-class percentages, reconstructed from the supplied counts. The 2000 zero-total row is excluded as missing. Precision is numerical reporting precision, not a claim of measurement accuracy.

Year	Class 1	Class 2	Class 3	Class 4	Class 5
2001	31.01103	60.67896	8.27836	0.03164	0
2002	31.19727	59.82128	8.97567	0.00504	7.490×10^{-4}
2003	35.88685	53.37909	10.67582	0.05784	3.914×10^{-4}
2004	33.08789	57.91899	8.87739	0.11361	2.132×10^{-3}
2005	41.43751	56.13349	2.37635	0.05264	0
2006	36.94736	59.94740	3.10519	0.00005	0
2007	29.03986	55.64078	15.12370	0.19553	1.289×10^{-4}
2008	33.35369	56.92609	9.64894	0.07129	0
2009	32.61730	63.36504	4.01550	0.00217	0
2010	30.97109	60.69098	8.31962	0.01830	0
2011	33.84003	59.91418	6.22886	0.01693	0
2012	43.56579	51.90497	4.50639	0.02279	5.727×10^{-5}
2013	48.57739	49.41510	1.96082	0.04637	3.147×10^{-4}
2014	41.07064	55.54110	3.35261	0.03565	0
2015	39.97495	51.85468	8.12843	0.04193	9.542×10^{-6}
2016	37.26635	55.87543	6.85486	0.00337	0
2017	38.52417	55.26916	6.11634	0.08969	6.493×10^{-4}
2018	26.92623	58.79860	14.21040	0.06468	9.541×10^{-5}
2019	39.59559	53.83001	6.54634	0.02677	1.293×10^{-3}
2020	31.31533	57.34331	11.26326	0.07809	0
2021	35.89824	59.66822	4.43338	0.00015	0
2022	42.07485	53.88953	4.03045	0.00516	0
2023	37.97014	50.07212	11.46450	0.49106	2.179×10^{-3}
2024	40.65384	50.18570	9.02445	0.13567	3.533×10^{-4}
2025	43.26576	50.14674	6.54570	0.04180	0

Here $\Delta\beta_{jk} = \beta_{j,k+1} - \beta_{jk}$ and Δ^ν is its ν -fold iterate, with $\Delta^0\beta_{jk} = \beta_{jk}$. Consequently $z^{(m+1)} = 0$, equivalently a companion system of $m+1$ first-order linear ODEs. Moreover,

$$\int_{t_0}^T |z'_{j,m}(t)|^2 dt \leq \frac{m}{T-t_0} \sum_{k=0}^{m-1} (\Delta\beta_{jk})^2 = \frac{2s_j^2}{T-t_0} P_{j,m}(\beta_j). \quad (\text{D.4})$$

Proof. The Bernstein basis is nonnegative and sums to one, so z is a convex combination of its coefficients. Differentiating the basis once gives $dz/d\zeta = m \sum_k \Delta\beta_k B_{k,m-1}$, and induction proves Eq. (D.3). The next derivative vanishes. With $x = (z, z', \dots, z^{(m)})$, the companion system is $\dot{x} = (x_2, \dots, x_{m+1}, 0)$. Jensen's inequality gives $(dz/d\zeta)^2 \leq m^2 \sum_k (\Delta\beta_k)^2 B_{k,m-1}$. Each basis function has integral $1/m$ on $[0, 1]$. Integration and $dt = (T - t_0)d\zeta$ prove Eq. (D.4). $\square$

These are empirical polynomial dynamics, not a mechanistic ecosystem model. Their exact analytic propagation has no Runge–Kutta error. No forecast outside $[t_0, T]$ inherits the box bound automatically.

As before, raw positive-total class predictions are closed to a composition using Eq. (C.7). This postprocessing does not minimize a joint five-class likelihood. Its effect on errors can nevertheless be bounded.



Proposition D.2 (Stability of compositional closure). *For nonnegative vectors v, w with sums at least $M_0 > 0$, let $C(v) = 100v/(\mathbf{1}^\top v)$. Then*

$$\|C(v) - C(w)\|_1 \leq (200/M_0) \|v - w\|_1. \quad (\text{D.5})$$

Proof. Write the difference as $100(v-w)/S_v + 100w(S_w - S_v)/(S_v S_w)$. Its norm is at most $100 \|v - w\|_1 / S_v + 100 |S_w - S_v| / S_v$. Use $|S_w - S_v| \leq \|v - w\|_1$ and $S_v \geq M_0$. $\square$

D.2. Tuning and implementation without test-outcome selection. Four rows are reported for every degree. GLDM minimizes $J_{1,0}$. Reciprocal GLDM maximizes $K_{1,0}$ and has the same fit. Penalized GLDM minimizes $J_{1,\lambda}$ with a selected penalty. Inverse-power GLDM maximizes $K_{2,\lambda}$, equivalently minimizing $J_{2,\lambda}$. The third row is necessary to distinguish the effect of the penalty from the effect of the inverse power.

For an outer training cutoff t , select one common λ for all five classes from $\{0, 0.1, 1, 10\}$. For each m, q, λ , use the five inner origins $u = t - 5, \dots, t - 1$, fit only 2001– u , and predict $u + 1$. Select the smallest composition-normalized MAE across these 25 class-year errors. All inner outcomes are at or before t . Numerical ties choose the larger penalty. The same inner scores can select m , with lower degree resolving an exact tie. This optional order selection is reported separately from the complete order-by-order comparison. The choice $q = 2$ is fixed, not optimized using test outcomes.

Algorithm 10 Continuation fit for one polynomial class model.

Require: Training years and class percentages, degree m , exponent q , penalty λ , horizon T .

- 1: Compute the degree-one LAD scale from training data and set $z_k = \beta_k/s_j$.
 - 2: Build the Bernstein matrix and the bounds $0 \leq z_k \leq 100/s_j$.
 - 3: Initialize bounded OLS, bounded LAD, a constant training median, and degree-elevated degree-one LAD.
 - 4: **for** each of the four starts **do**
 - 5: **for** $\delta \in \{10^{-2}, 10^{-4}, 10^{-6}\}$ **do**
 - 6: Minimize the smooth penalized loss with analytic-gradient L-BFGS-B.
 - 7: Use the current solution as the next continuation start.
 - 8: **end for**
 - 9: **end for**
 - 10: Select the smallest final objective and evaluate its projected-gradient residual.
 - 11: If that residual exceeds 10^{-4} , attempt SLSQP refinement and retain it only if both diagnostics decrease.
 - 12: Record the coefficients, scale, solver status, gradient residual, and residual-tube check.
 - 13: **return** the selected local numerical fit.
-

L-BFGS-B has a maximum of 1,200 iterations per continuation stage, relative function tolerance 10^{-13} , projected-gradient target 10^{-7} , and at most 60 line-search steps. SLSQP refinement has a 1,000-iteration limit and function tolerance 10^{-14} . Coefficients are optimized in scale units. Different stopping diagnostics are saved separately. This implementation solves the smooth objective directly, rather than asserting that its iterates are those of the exact nonsmooth MM theorem.

D.3. All-order and class-specific results. The holdout and rolling windows are the same as in Section C. Their different forecast horizons test different questions, and their overlapping years are not independent datasets. Table 41 gives every requested GLDM and reciprocal GLDM order. Ordinary inverse results are identical by construction, with independently evaluated loss and gain values auditing the identity. The comparison does not count the identical representation as a second independently fitted estimator.

Inverse-power GLDM reduces rolling MAE at every reported degree. For $m = 1$, MAE changes from 2.035 to 1.948 percentage points, a reduction of 4.31%. For $m = 5$, it changes from 4.013 to 2.024, a reduction of 49.57%. These are descriptive differences on ten annual forecasts, not significance tests or independent Monte Carlo replications.

The five-year holdout gives a different picture. At $m = 1$, inverse-power GLDM increases MAE from 1.608 to 2.743; it also worsens degrees 2 and 3. At $m = 4$ and $m = 5$, it reduces MAE from 9.260 to 2.883 and from 9.620 to 2.747, respectively. Thus the extension controls some higher-degree extrapolation errors but does not uniformly outperform



Algorithm 11 Chronological comparison of orders one to five.**Require:** The 25 January compositions and the fixed protocol described above.

```

1: for the fixed cutoff 2020 and each rolling cutoff  $t = 2015, \dots, 2024$  do
2:   for  $m = 1, \dots, 5$  do
3:     Compute inner one-year scores for  $q \in \{1, 2\}$  and every candidate  $\lambda$ .
4:     Select  $\lambda$  separately for each  $q$  using only inner outcomes.
5:     Fit GLDM, reciprocal GLDM, penalized GLDM, and inverse-power GLDM.
6:     Close the forecast compositions and save every class-year prediction.
7:   end for
8:   Also select an order from the inner scores for each penalized family.
9: end for
10: Report all fixed and rolling MAE/RMSE values, settings, and diagnostics.

```

TABLE 41. Complete order comparison. Errors are percentage points. Reciprocal GLDM uses $q = 1$; inverse-power GLDM uses $q = 2$. Penalties use inner validation only.

m	Method	Holdout MAE	Holdout RMSE	Rolling MAE	Rolling RMSE
1	GLDM	1.608	2.394	2.035	3.373
1	Reciprocal GLDM	1.608	2.394	2.035	3.373
1	Penalized GLDM	2.573	3.820	2.033	3.368
1	Inverse-power GLDM	2.743	3.993	1.948	3.226
2	GLDM	1.687	2.392	2.414	3.828
2	Reciprocal GLDM	1.687	2.392	2.414	3.828
2	Penalized GLDM	2.598	3.885	1.964	3.058
2	Inverse-power GLDM	2.746	3.998	2.007	2.997
3	GLDM	1.853	2.741	2.798	4.295
3	Reciprocal GLDM	1.853	2.741	2.798	4.295
3	Penalized GLDM	2.612	3.924	1.954	3.054
3	Inverse-power GLDM	2.715	4.021	2.037	3.038
4	GLDM	9.260	14.328	3.975	6.457
4	Reciprocal GLDM	9.260	14.328	3.975	6.457
4	Penalized GLDM	2.621	3.937	1.957	3.055
4	Inverse-power GLDM	2.883	4.200	2.144	3.150
5	GLDM	9.620	14.507	4.013	6.671
5	Reciprocal GLDM	9.620	14.507	4.013	6.671
5	Penalized GLDM	2.851	4.156	2.182	3.295
5	Inverse-power GLDM	2.747	4.001	2.024	3.008

ordinary GLDM. Penalized GLDM itself has smaller rolling MAE than the inverse-power version at degrees 2–4. This prevents attributing every gain to the power transformation.

The degree-one values in this table use the common continuation implementation shared with higher degrees. The earlier table uses exhaustive unsmoothed vertex enumeration. Their small numerical differences arise from optimization and smoothing, not a change in data or an additional independent experiment.

Figure 36 displays the full curves instead of isolating the winning degrees. Penalization limits the high-order forecast excursions, but also biases the long-horizon low-order fits. The matched penalized-GLDM comparison shows that some of the improvement comes from shrinkage. The inverse-power loss is not the best choice at every degree or horizon.



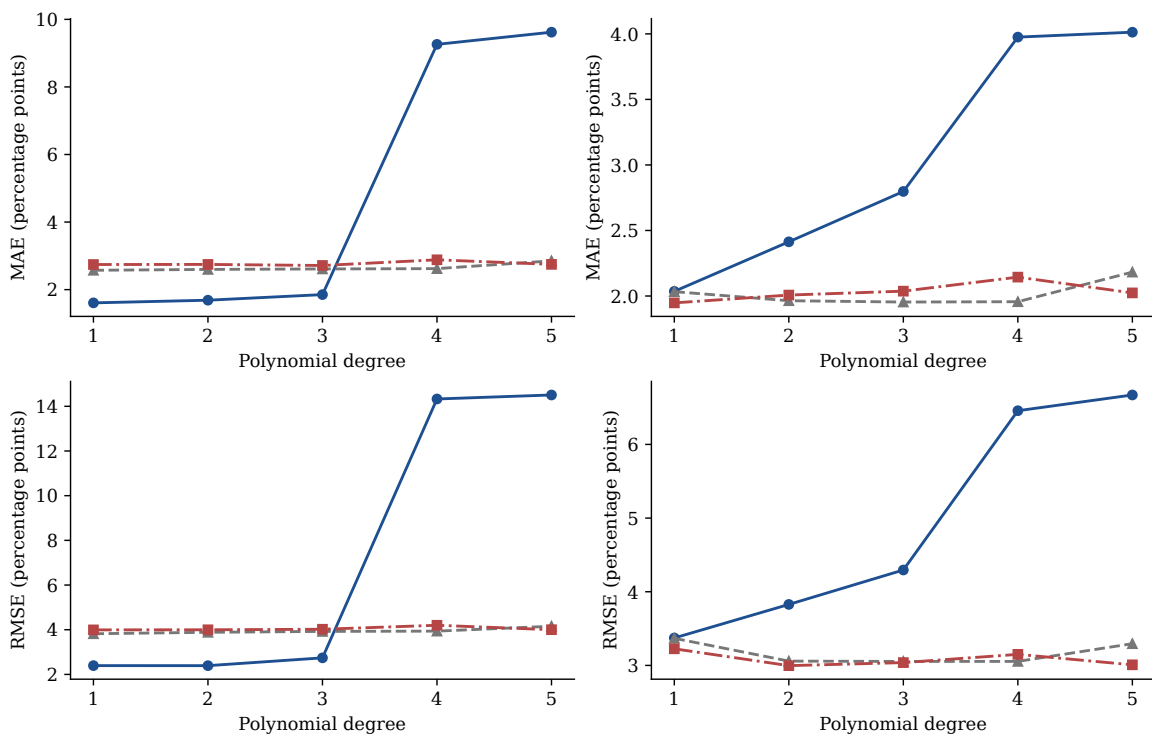


FIGURE 36. Errors over polynomial degrees one to five. Left panels use the five-year fixed holdout; right panels use the ten rolling forecasts. The upper row shows MAE and the lower row RMSE. Blue circles show GLDM and its identical ordinary inverse form, gray triangles show penalized GLDM, and red squares show inverse-power GLDM. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Tables 42 and 43 retain every class's MAE/RMSE, with scientific notation for Class 5. Reciprocal GLDM's identical class cells are not repeated, although its aggregate rows remain. Percentage-point errors weight prevalent classes more strongly and do not imply equal relative accuracy.

Figure 37 makes the class dependence visible. A lower aggregate score need not mean that every class improved. Rare-class results remain affected by the fixed scale floor and by the few nonzero observations. The complete annual observed values, unnormalized predictions, normalized predictions, and residuals are supplied as CSV files.

Figure 38 shows the source of the high-order errors. Bounded coefficients prevent predictions outside zero to 100 percent before closure, but this broad constraint does not prevent large and inaccurate extrapolation within those limits. The penalty reduces the excursions and can also flatten genuine temporal variation.

The selected-order results in Table 44 use inner validation only. They are not the minimum test scores extracted from Table 41. One-year inner validation cannot guarantee the best five-year extrapolation. A future operational design should match the validation horizon to the intended forecast horizon and assess its settings on newly observed years. Such a redesign is not selected retrospectively using this holdout.

D.4. Numerical checks and reproducibility. The experiment executes 2,950 distinct class fits, each from four starts and three smoothing stages. Repeated inner fits are cached. The maximum ordinary loss-gain complement discrepancy is 6.66×10^{-16} . Ordinary GLDM and inverse predictions agree exactly in the stored arrays. Analytic-gradient checks against central differences have maximum discrepancy 2.17×10^{-10} . The median and maximum

TABLE 42. Fixed-holdout errors for all classes and orders. Each cell gives MAE above RMSE in percentage points. G / I denotes GLDM and reciprocal GLDM, PG penalized GLDM, and PI inverse-power GLDM.

m	Method	Class 1	Class 2	Class 3	Class 4	Class 5
1	G / I	2.164	2.992	2.757	0.127	5.09×10^{-4}
		2.477	3.468	3.231	0.200	9.79×10^{-4}
1	PG	4.602	5.610	2.527	0.124	5.09×10^{-4}
		5.298	6.055	2.860	0.208	9.79×10^{-4}
1	PI	5.395	5.474	2.721	0.126	5.09×10^{-4}
		6.013	5.879	2.993	0.211	9.79×10^{-4}
2	G / I	2.605	3.267	2.429	0.132	5.10×10^{-4}
		2.862	3.617	2.702	0.188	9.76×10^{-4}
2	PG	4.639	5.738	2.489	0.124	5.10×10^{-4}
		5.341	6.223	2.860	0.208	9.75×10^{-4}
2	PI	5.408	5.471	2.724	0.126	5.10×10^{-4}
		6.032	5.873	2.998	0.211	9.75×10^{-4}
3	G / I	3.370	2.479	3.281	0.132	5.36×10^{-4}
		3.943	3.045	3.565	0.187	9.05×10^{-4}
3	PG	4.658	5.814	2.465	0.124	5.12×10^{-4}
		5.365	6.323	2.863	0.208	9.67×10^{-4}
3	PI	5.093	5.796	2.558	0.126	5.12×10^{-4}
		5.753	6.297	2.841	0.211	9.67×10^{-4}
4	G / I	22.374	11.559	12.237	0.132	5.45×10^{-4}
		25.866	13.606	13.122	0.188	8.85×10^{-4}
4	PG	4.693	5.824	2.465	0.125	5.15×10^{-4}
		5.401	6.337	2.855	0.208	9.59×10^{-4}
4	PI	5.949	5.544	2.795	0.126	5.08×10^{-4}
		6.531	5.968	3.141	0.211	9.81×10^{-4}
5	G / I	23.777	8.162	16.031	0.131	5.93×10^{-4}
		25.814	9.132	17.389	0.190	8.71×10^{-4}
5	PG	5.807	5.563	2.760	0.126	5.09×10^{-4}
		6.402	5.993	3.068	0.211	9.80×10^{-4}
5	PI	5.420	5.465	2.725	0.127	5.13×10^{-4}
		6.051	5.864	3.000	0.214	9.66×10^{-4}

selected-fit projected-gradient residuals are 7.90×10^{-7} and 9.84×10^{-5} , respectively. For the numerical backward-data LP of Proposition 3.8, with standardized tube 10^{-4} , the maximum evaluated stationarity residual is 4.76×10^{-5} . These are floating-point diagnostics, not interval-arithmetic certificates. All coefficient bounds and composition sums are checked. Bounds within 64 machine epsilons times the scale-unit upper bound are rounded to the bound before diagnostics, so a near-bound coefficient is not assigned an unjustified exact normal.

The order sweep used the Intel Xeon Platinum 8573C, Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0 and one BLAS/OpenMP thread; fitting took 58.11 seconds, excluding plots and final LPs. All coefficients, scales, penalty scores, diagnostics, predictions, class errors and checks are retained for offline January-only reproduction. Degree increases flexibility, not the 25-observation sample. No unobserved pixels, other months, vegetation labels or fabricated replications enter this study. Statistical stationarity assumptions are unverified, and exhaustive degree-one search remains distinct from local higher-degree continuation.

E. EXTENDED JANUARY SENSITIVITY ANALYSES AND COMPUTATIONAL RECORDS

The order comparison in Section D includes the poor unpenalized degree-four and degree-five results. The following experiments assess sensitivity to penalties, model structure, validation horizon and optimization. Their conclusions



TABLE 43. Rolling errors for all classes and orders. Each cell gives MAE above RMSE in percentage points. G / I denotes GLDM and reciprocal GLDM, PG penalized GLDM, and PI inverse-power GLDM.

m	Method	Class 1	Class 2	Class 3	Class 4	Class 5
1	G / I	3.982	2.715	3.395	0.085	4.69×10^{-4}
		5.372	3.195	4.217	0.158	8.34×10^{-4}
1	PG	3.951	3.049	3.083	0.084	4.69×10^{-4}
		5.226	3.800	3.865	0.155	8.34×10^{-4}
1	PI	3.752	2.737	3.162	0.086	4.60×10^{-4}
		5.005	3.514	3.821	0.156	8.33×10^{-4}
2	G / I	4.479	3.548	3.939	0.101	4.87×10^{-4}
		5.674	4.059	4.956	0.168	8.35×10^{-4}
2	PG	3.884	3.090	2.763	0.084	4.83×10^{-4}
		4.510	3.771	3.491	0.155	8.35×10^{-4}
2	PI	4.032	3.046	2.876	0.083	4.72×10^{-4}
		4.521	3.710	3.268	0.154	8.33×10^{-4}
3	G / I	5.087	4.588	4.216	0.096	5.01×10^{-4}
		6.413	5.074	5.035	0.169	8.26×10^{-4}
3	PG	3.870	3.150	2.666	0.084	4.88×10^{-4}
		4.504	3.798	3.449	0.155	8.23×10^{-4}
3	PI	4.104	2.966	3.030	0.084	4.88×10^{-4}
		4.611	3.632	3.418	0.156	8.23×10^{-4}
4	G / I	7.614	6.532	5.629	0.101	5.18×10^{-4}
		9.642	8.655	6.371	0.170	8.30×10^{-4}
4	PG	3.880	3.148	2.670	0.084	5.06×10^{-4}
		4.512	3.798	3.443	0.155	8.27×10^{-4}
4	PI	4.310	3.096	3.231	0.082	4.66×10^{-4}
		4.759	3.742	3.597	0.155	8.18×10^{-4}
5	G / I	9.114	4.799	6.050	0.101	5.81×10^{-4}
		11.431	6.689	6.861	0.170	8.61×10^{-4}
5	PG	4.433	3.373	3.023	0.083	4.63×10^{-4}
		5.095	3.868	3.652	0.154	8.17×10^{-4}
5	PI	4.070	3.042	2.921	0.084	5.02×10^{-4}
		4.551	3.682	3.309	0.155	8.25×10^{-4}

are conditional on the same 25 annual compositions; they do not add independent field replications. All tuning uses responses available before the corresponding forecast origin. Additional mathematical structure is separated from empirical performance.

E.1. Penalty scaling and a coupled compositional model. The penalty in Eq. (D.2) is $m/(2s_j^2)$ times the sum of squared differences. It is not m^2/s_j^2 times that sum. A factor m^2 appears before integrating the squared derivative; integration of a degree- $(m-1)$ Bernstein basis contributes $1/m$, producing the factor m in the penalty. Standardization by s_j^2 makes the penalty dimensionless. Equal numerical penalties across classes therefore refer to scale-adjusted variation, not equal physical percentage-point variation. They need not imply equal effective degrees of freedom.

Proposition E.1 (Degree scaling of the variation penalty). *For a fixed scale and horizon, a line represented at any degree $m \geq 1$ has the same penalty. More generally, degree elevation of a Bernstein polynomial cannot increase $m \sum_k (\Delta\beta_k)^2$.*



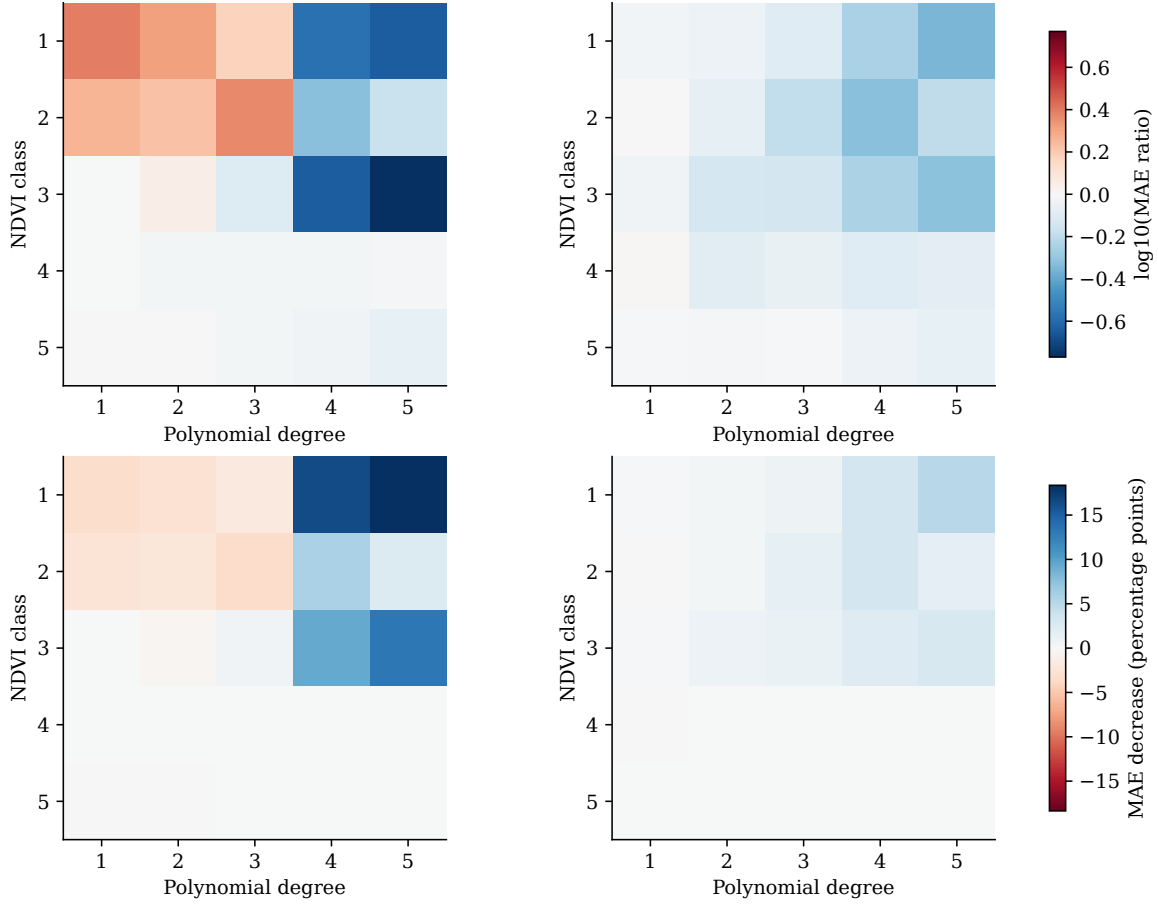


FIGURE 37. Class-specific comparisons. Columns are holdout and rolling protocols. Top: $\log_{10}$ of inverse-power to GLDM MAE ratio (negative favors inverse power). Bottom: GLDM minus inverse-power MAE in percentage points (positive favors inverse power). Relative rare-class changes should be interpreted together with absolute changes. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Proof. For a line, $\beta_k = a + (b - a)k/m$ and $m \sum_k (\Delta\beta_k)^2 = (b - a)^2$. For a degree- m polynomial elevated to $m + 1$, write

$$\gamma_k = \frac{k}{m+1} \beta_{k-1} + \left(1 - \frac{k}{m+1}\right) \beta_k, \quad \Delta\gamma_k = \frac{k\Delta\beta_{k-1} + (m-k)\Delta\beta_k}{m+1}, \quad (\text{E.1})$$

with endpoint terms omitted when their coefficients vanish. Jensen's inequality yields

$$\sum_{k=0}^m (\Delta\gamma_k)^2 \leq \frac{m^2}{(m+1)^2} \sum_{k=0}^m \left[\frac{k}{m} (\Delta\beta_{k-1})^2 + \frac{m-k}{m} (\Delta\beta_k)^2 \right] = \frac{m}{m+1} \sum_{k=0}^{m-1} (\Delta\beta_k)^2. \quad (\text{E.2})$$

This proves the assertion. The bound motivates a common candidate grid for λ but does not make different polynomial models statistically equivalent. $\square$

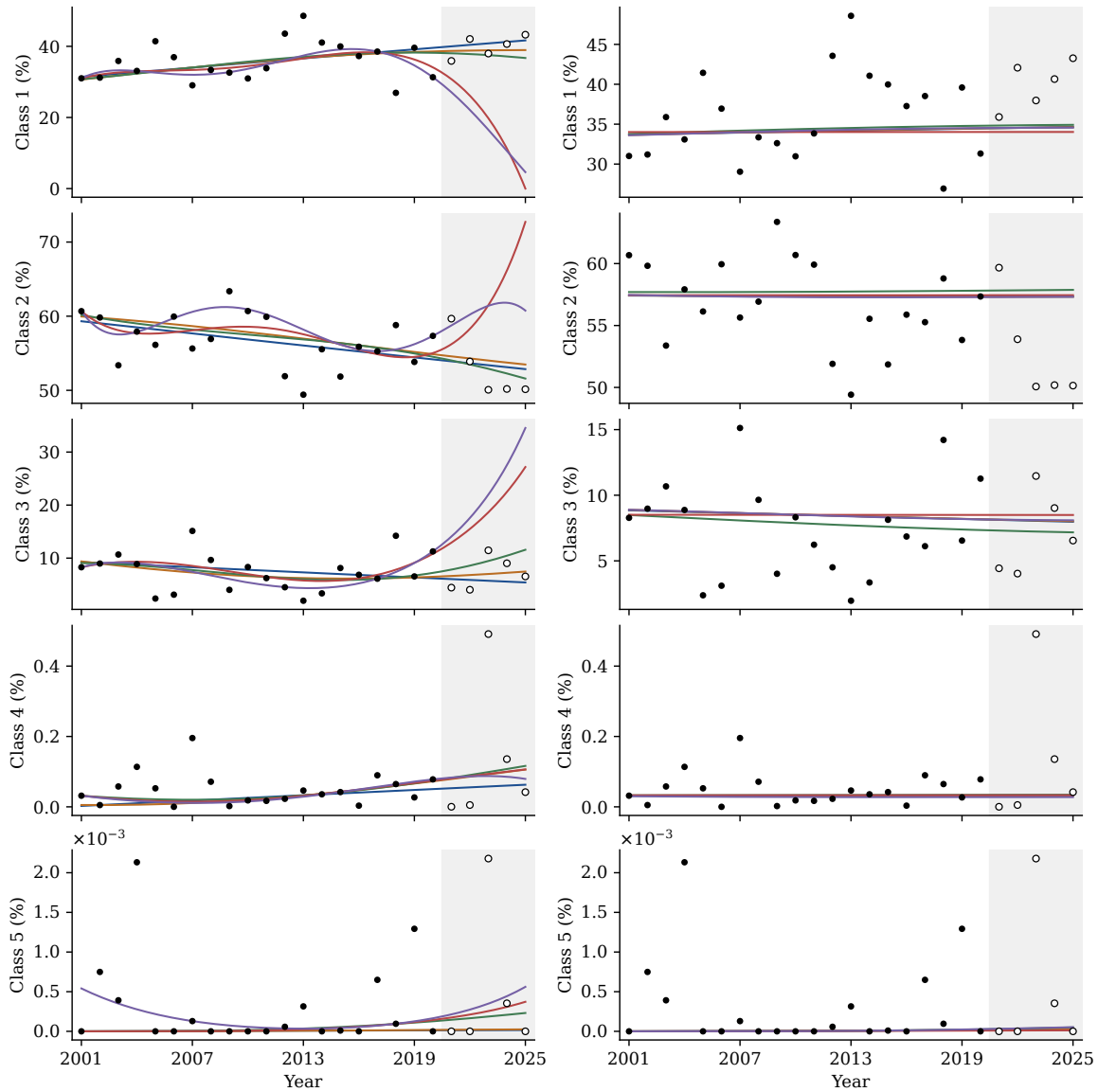


FIGURE 38. All five polynomial orders fitted through 2020. The left column shows GLDM and the identical ordinary inverse form; the right shows inverse-power GLDM with penalties selected within training. Blue, orange, green, red, and purple lines correspond to degrees 1–5. Filled circles denote training observations, open circles denote withheld observations, and shading marks 2021–2025. Each row has its own vertical scale.

For a secondary analysis, fit all five classes jointly. Let $\mathcal{B}$ be the coefficient set

$$\mathcal{B} = \{\beta : \beta_{jk} \geq 0, \sum_{j=1}^5 \beta_{jk} = 100 \text{ for every } k = 0, \dots, m\}. \quad (\text{E.3})$$

TABLE 44. Training-only order and penalty selections. The cutoff is the last training year. Selected GLDM here includes penalty selection.

Protocol and cutoff	GLDM m	GLDM λ	Inverse-power m	Inverse-power λ
Holdout 2020	4	0.1	1	0.1
Rolling 2015	1	0	1	0
Rolling 2016	1	0	1	0
Rolling 2017	1	0	1	0
Rolling 2018	1	0	1	0
Rolling 2019	5	0.1	1	0.1
Rolling 2020	4	0.1	1	0.1
Rolling 2021	3	0.1	1	0.1
Rolling 2022	1	0.1	1	0.1
Rolling 2023	1	0.1	1	0
Rolling 2024	1	0.1	1	0.1
Holdout MAE	2.621		2.743	
Rolling MAE	2.118		1.948	

Minimize on this set

$$\mathcal{L}_{q,\lambda}(\beta) = \frac{1}{5} \sum_{j=1}^5 \left[\frac{1}{n} \sum_{i=1}^n \ell_{q,c} \left(\frac{|y_{ij} - z_{j,m}(t_i; \beta_j)|}{s_j} \right) + \lambda P_{j,m}(\beta_j) \right]. \quad (\text{E.4})$$

Unlike independent fitting followed by closure, this objective constrains the calibration itself.

Proposition E.2 (Exact compositional coherence). *Fix an integer $m \geq 1$, $T > t_0$, scales $s_j > 0$, $c > 0$, $q \in \mathcal{Q}$ and $\lambda \geq 0$. The set $\mathcal{B}$ is nonempty, compact and convex. For every feasible coefficient matrix, $z_j(t) \geq 0$ and $\sum_j z_j(t) = 100$ on $[t_0, T]$. The criterion attains a minimum. With exact weighted-LAD subproblems and proximal coefficients bounded above and bounded away from zero, its unsmoothed MM algorithm has sufficient decrease and stationary accumulation points.*

Proof. Each coefficient column belongs to the closed simplex of mass 100, so the product set has the stated properties. Nonnegativity of the basis gives nonnegative predictions, and $\sum_j z_j(t) = \sum_k B_{k,m}(\zeta(t)) \sum_j \beta_{jk} = 100$. Continuity on a compact set proves attainment. Standardization only rescales the fixed affine design. The average difference penalty is convex and smooth. Theorem 3.5 therefore applies to the stacked vector and the convex set $\mathcal{B}$. $\square$

The secondary implementation uses smooth constrained continuation; the preceding theorem describes the exact MM option, not a global guarantee for that local implementation. Coherence is a structural constraint, not a guarantee of smaller predictive error.

Penalized affine MM uses the feasible bound in Proposition 6.2. Projected feasible scores and actual box activity enter the independently evaluated gap.

E.2. Separating loss power from regularization. At every degree and forecast origin, four directly paired fits use

$$(q, \lambda) \in \{(1, 0), (2, 0), (1, 0.1), (2, 0.1)\}. \quad (\text{E.5})$$

The common $\lambda^* = 0.1$ is a prespecified ablation value, not chosen by held-out performance. These comparisons use the same scales and starts. Table 45 separates the exponent effect from the penalty effect; the earlier validation-selected results remain distinct.

At degree five, adding the common penalty to GLDM changes rolling MAE from 4.013 to 2.034; changing the exponent at that same penalty changes it to 2.024. Figure 39 makes the regularization effect visible across every



TABLE 45. Same-penalty ablation, MAE in percentage points. Higher-order entries are local multi-start fits, not globally certified minima. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Protocol	m	$q = 1, \lambda = 0$	$q = 2, \lambda = 0$	$q = 1, \lambda = .1$	$q = 2, \lambda = .1$
holdout	1	1.608	1.593	2.573	2.743
holdout	2	1.687	1.607	2.598	2.746
holdout	3	1.853	1.515	2.612	2.715
holdout	4	9.260	5.036	2.621	2.748
holdout	5	9.620	9.621	2.627	2.747
rolling	1	2.035	1.998	1.963	1.985
rolling	2	2.414	2.461	1.964	2.007
rolling	3	2.798	2.814	1.954	2.037
rolling	4	3.975	3.582	1.957	2.043
rolling	5	4.013	3.936	2.034	2.024

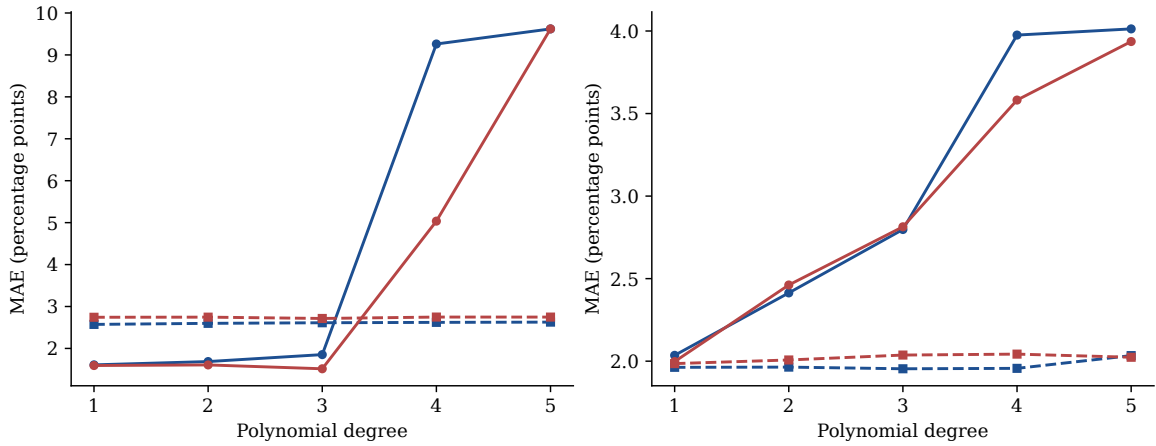


FIGURE 39. January common-penalty ablation. Left: five-year holdout; right: one-year rolling forecasts. Blue denotes ordinary GLDM ($q = 1$) and red inverse-power GLDM ($q = 2$). Solid lines with circles use $\lambda = 0$; dashed lines with squares use the common $\lambda = 0.1$. Errors are in percentage points. The pronounced high-degree reduction is mostly a penalty effect. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

degree. A smaller loss on atypical years does not by itself prevent implausible extrapolation; the difference penalty restricts variation of the forecast curve. Independently selected penalties can further change the ordering. Hence a comparison of unpenalized GLDM with penalized inverse-power GLDM is a comparison of complete procedures and cannot attribute its whole difference to q .

E.3. Historical validation horizons and forecast errors. The primary procedure uses five recent one-year validation origins. A secondary horizon experiment instead uses every origin $o = 2010, \dots, 2020 - H$ for $H \in \{1, 3, 5\}$, fitting through o and evaluating all years $o + 1, \dots, o + H$. Thus the numbers of historical validation blocks are ten, eight and six, respectively; no validation response occurs after 2020. Blocks overlap, so validation scores are descriptive averages rather than independent observations. Each horizon selects among all $m = 1, \dots, 5$, $q = 1, 2$ and $\lambda \in \{0, 0.1, 1, 10\}$.

Ties prefer lower degree, lower exponent and then larger penalty. The chosen triple is refitted through 2020 and evaluated on the common five-year holdout.

TABLE 46. Historical horizon selection and the common 2021–2025 holdout. The forecasts are local fits; the holdout does not select the displayed triple. Origin counts, error denominators and overlap follow Table 30.

Inner horizon	m	q	λ	Validation MAE	Holdout MAE
1	1	2	0.0	2.365	1.593
3	1	2	0.0	2.810	1.593
5	1	1	0.1	2.810	2.573

The one- and three-year validation rules select the same unpenalized degree-one inverse-power model in this dataset. Five-year validation selects a penalized degree-one GLDM model and produces a worse subsequent holdout score. Matching the horizon therefore changes the statistical question but does not guarantee an improvement: six historical blocks can still be unrepresentative of the later regime. The reported procedure is not changed by choosing the best horizon retrospectively.

The six overlapping five-year blocks reuse a short record and are not independent replications. Their validation ranking need not predict the later holdout: composition, influential years, local basins, model error and possible nonstationarity all contribute. One holdout cannot isolate nonstationarity. Horizon matching defines the forecasting task, but cannot establish future representativeness. Further observations should support prospective horizon-specific errors and stability assessment.

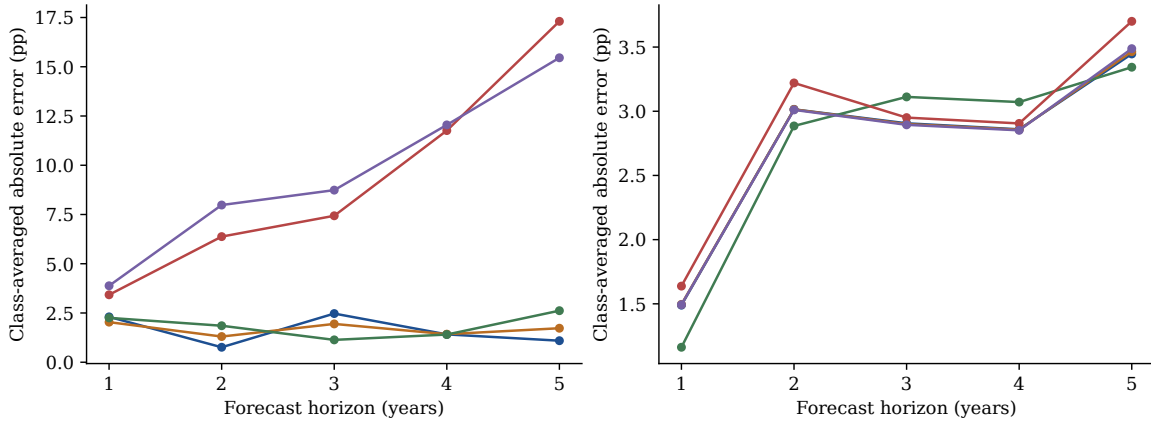


FIGURE 40. Holdout absolute error by horizon, averaged over the five classes. Left: unpenalized GLDM. Right: validation-selected inverse-power GLDM. Blue, orange, green, red and purple denote degrees one through five. These five successive errors are not independent replicates.

Figure 40 localizes the high-degree failures within the five-year horizon. In Figure 38, coefficient bounds prevent an individual raw curve from leaving $[0, 100]$ on the design interval, but permit large curvature between the observed portion and its endpoint. Bernstein coefficients near opposite box faces can still create implausible redistribution among classes. The derivative-energy bound in Eq. (D.4) explains the additional purpose of the difference penalty. It suppresses variation that the box constraint alone permits; it does not establish correctness of polynomial extrapolation.

E.4. Power, scale and standard robust comparators. Exponent sensitivity uses degree one, $q = 1, 2, 3, 4$ and $c = 1.5$. Scale sensitivity for $q = 2$ uses $c = 0.5, 1, 1.5, 3$. For each fixed candidate, the penalty is chosen from the same grid using the five pre-2021 one-year origins. Neither exponent nor scale is selected by the holdout. The $q = 2, c = 1.5$ configuration remains primary even when a sensitivity row performs better.

TABLE 47. Training-only penalty tuning within prespecified degree-one power/scale candidates. These local-fit sensitivity rows are not a post-hoc change of the primary estimator. Origin counts, error denominators and overlap follow Table 30.

q	c	λ	Validation MAE	Holdout MAE
1	1.5	0.1	1.827	2.573
2	1.5	0.1	1.828	2.743
3	1.5	0.1	1.805	2.820
4	1.5	0.1	1.776	3.229
2	0.5	0.1	1.769	2.665
2	1.0	0.1	1.837	2.800
2	3.0	0.1	1.987	2.700

Larger powers are not automatically advantageous. They preserve the slope at zero while suppressing more of the residual range, and their population curvature can become smaller. In an application with unknown contamination, a useful procedure is to compare a small, declared set of (q, c, λ) candidates on time blocks matching the intended prediction task, retain a simple near-tie solution, and report sensitivity to the final choice. A single favorable test year should not trigger retuning. For a new application, $q = 1$ provides the ordinary-GLDM reference and $q = 2$ the smallest distinct inverse-power choice.

A short scale grid around the pilot normalization, such as the declared c values here, is preferable to estimating a contamination percentage that the data cannot support. The statistical penalty λ should be chosen on earlier time blocks, separately from the algorithmic proximal coefficient and the ODE step size. The polynomial degree is limited by the number of training years and the forecast horizon. If several choices have similar validation error, prefer lower degree and report the competing fits; a larger q has no general robustness guarantee.

The smoothing parameter controls numerical approximation rather than contamination: $\delta \leq c\epsilon_F/W$ enforces a chosen objective-perturbation budget, while smaller values require tighter conditioning and stationarity checks.

Huber and Cauchy use the same frozen LAD scales and four starts at every degree, thresholds 1.345 and 2.385, and penalties separately selected on the same five historical origins. Their performance tests whether standard robust fitting plus regularization suffices. Because loss normalizations differ, the common-numerical-penalty ablation is restricted to the cusp-normalized arctangent family.

Figure 41 shows that the standard comparators are competitive. In particular, the low rolling errors of regularized Huber at intermediate degrees show that useful shrinkage is not unique to the inverse-power loss. The selected procedure should be justified by the forecast task, the contamination mechanism and the available diagnostics, rather than by its name. The numerical comparisons use the same observations, feasible sets, frozen scales, chronological splits and candidate penalty grid. Thus they compare standard loss-based estimators under a common protocol, rather than unrelated error values copied from different datasets in the literature. The weak-form and probabilistic-integrator approaches discussed in Section 1 provide methodological context and are not implemented performance baselines.

E.5. Joint compositions and delete-one-year stability. The joint model in Eq. (E.4) is evaluated at degrees one and three, with $q = 1, 2$ and the common $\lambda = 0.1$. Three feasible starts are the normalized independent coefficient matrix, a constant historical composition and their average. SLSQP uses smoothing levels $10^{-3}, 10^{-5}, 10^{-6}$, maximum 1500 iterations per level and function tolerance 10^{-11} . The best feasible objective is retained. The simplex constraints are checked independently. The result is a feasibility-checked local sensitivity fit, not a global certificate or an implementation of the exact-MM theorem.



TABLE 48. Validation-selected penalties for four robust polynomial procedures, MAE in percentage points. All are local multistart fits. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Protocol	m	GLDM	Inverse-power	Huber	Cauchy
holdout	1	2.573	2.743	2.409	2.215
holdout	2	2.598	2.746	2.413	2.243
holdout	3	2.612	2.715	2.414	2.257
holdout	4	2.621	2.883	2.415	2.265
holdout	5	2.851	2.747	2.416	2.272
rolling	1	2.033	1.948	2.151	2.183
rolling	2	1.964	2.007	1.951	1.970
rolling	3	1.954	2.037	1.953	1.970
rolling	4	1.957	2.144	1.954	1.970
rolling	5	2.182	2.024	1.955	1.971

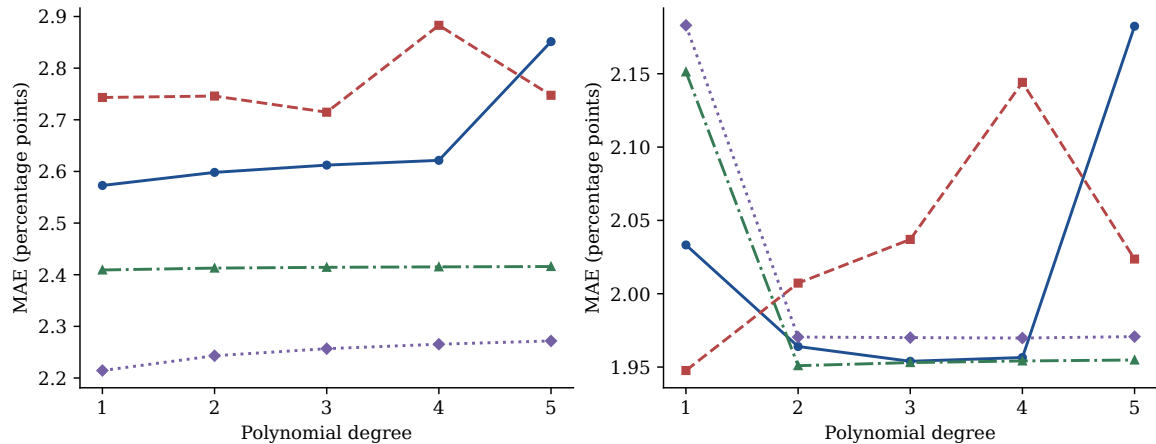


FIGURE 41. January comparisons with established robust losses under the same validation protocol. Left: holdout; right: rolling MAE. Penalized GLDM is blue with circles, inverse-power GLDM red with squares, Huber green with triangles and Cauchy purple with diamonds. Each procedure selects its penalty from earlier years. Lines connect degree-specific fits and are not confidence intervals. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

The largest coefficient-column sum discrepancy is $2.8e - 14$ percentage points across the 44 joint fits. Exact compositional structure does not improve every score. The joint restrictions can couple a difficult rare-class fit back to the dominant classes. This secondary result supports retaining the simpler independent baseline as a transparent competitor, while recognizing that post-fit closure and joint optimization solve different problems.

For training stability, each of the twenty years 2001–2020 is deleted in turn. Every degree and both exponents are refitted, with their primary penalty held fixed and the LAD scale recomputed from the retained training years. This is conditional stability of the chosen specification, not nested retuning or a sampling confidence interval. Forecasts always refer to the unchanged 2021–2025 holdout. Coefficient changes are measured in percentage points against the corresponding complete-training fit.



TABLE 49. Independent fitting plus closure versus joint compositional fitting, with matched degree, exponent and penalty. MAE is in percentage points. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Protocol	<i>m</i>	<i>q</i>	Independent	Joint
holdout	1	1	2.573	2.651
holdout	1	2	2.743	2.677
holdout	3	1	2.612	2.689
holdout	3	2	2.715	2.680
rolling	1	1	1.963	2.081
rolling	1	2	1.985	2.235
rolling	3	1	1.954	2.056
rolling	3	2	2.037	2.244

TABLE 50. Delete-one-training-year sensitivity. Maxima range over all omitted years, coefficients or holdout cells as appropriate. Origin counts, error denominators and overlap follow Table 30.

<i>m</i>	<i>q</i>	Max coefficient change	Max forecast change	Mean forecast change	MAE range
1	1	3.897	2.835	0.612	1.963–2.720
1	2	4.419	3.409	0.570	1.987–2.876
2	1	3.874	2.915	0.580	1.959–2.762
2	2	4.317	3.318	0.540	1.998–2.996
3	1	3.840	2.966	0.430	2.000–2.789
3	2	1.180	1.115	0.328	2.637–2.982
4	1	3.811	2.945	0.427	2.007–2.803
4	2	3.875	3.159	0.631	2.128–3.007
5	1	2.545	2.224	0.331	2.289–3.006
5	2	1.395	1.482	0.228	2.649–2.997

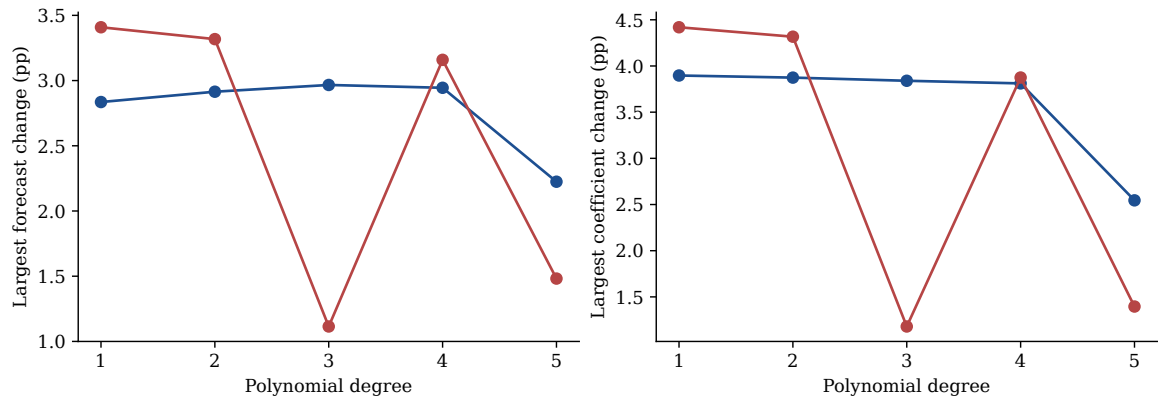


FIGURE 42. Largest change under deletion of one training year. Left: holdout cell prediction. Right: coefficient. Blue is penalized GLDM and red is inverse-power GLDM, with each primary penalty fixed.

The observed coefficient and forecast changes establish that individual training years matter; they do not support treating annual compositions as independent pixel samples. The unusual 2023 class-four observation remains in every holdout evaluation. It is not removed as an outlier merely because a model misses it.

E.6. The rare fifth class and scale-floor sensitivity. Class five is exactly zero in thirteen of the 25 annual records and has a much smaller scale than classes one and two. A floor of 0.01 percentage points can therefore dominate its residual normalization. The sensitivity experiment uses 0.005, 0.01, 0.02, 0.05 for this class only; all other class scales, degrees, primary penalties and training/holdout boundaries are fixed. Reported differences can reflect both the class-five fit and the small effect of compositional closure on other classes. This is not evidence about pixel-level detection limits, which cannot be established from the workbook alone.

TABLE 51. Degree-one scale-floor sensitivity, with the primary penalty fixed. Class-five MAE and overall MAE are in percentage points. Origin counts, error denominators and overlap follow Table 30.

q	Floor	Class 5 MAE	Overall MAE
1	0.005	0.000509	2.572830
1	0.01	0.000509	2.572830
1	0.02	0.000509	2.572830
1	0.05	0.000509	2.572830
2	0.005	0.000509	2.743200
2	0.01	0.000509	2.743200
2	0.02	0.000509	2.743200
2	0.05	0.000509	2.743200

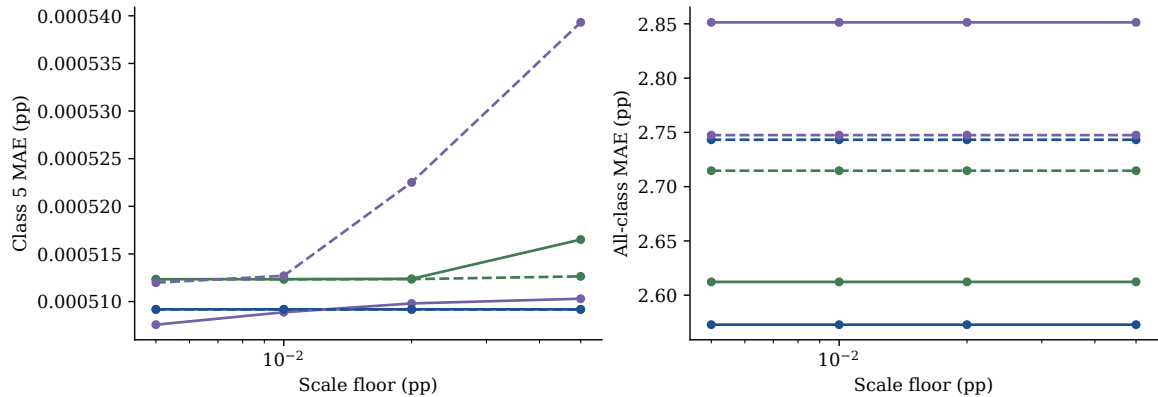


FIGURE 43. Scale-floor sensitivity for degrees one, three and five (blue, green and purple). Solid curves use $q = 1$, dashed curves $q = 2$. Left: class-five MAE; right: overall MAE.

Figure 37 shows both a logarithmic error ratio and an absolute error decrease. A large relative improvement in the rare fifth class can have negligible influence on overall percentage-point error. Reporting both views prevents that disproportion from being mistaken for a large practical gain.

E.7. Holdout compositions and forecast-error decomposition. Table 52 gives the observed and predicted compositions for representative degree-one models. All values are percentage points and rounded for display; the underlying unrounded rows sum to 100. The complete predictions at all five degrees are supplied in CSV. These tables are outputs of the prespecified fits, not additional criteria for model selection.



TABLE 52. Observed and degree-one predicted January compositions. Reciprocal GLDM predictions equal GLDM exactly and are not duplicated. Origin counts, error denominators and overlap follow Table 30.

Year	Model	Class 1	Class 2	Class 3	Class 4	Class 5
2021	Observed	35.8982	59.6682	4.4334	0.0002	0.0000
2021	GLDM	39.9996	53.9108	6.0366	0.0530	0.0000
2021	Penalized GLDM	35.2242	57.5790	7.1576	0.0391	0.0000
2021	Inverse-power	34.4981	57.3447	8.1238	0.0335	0.0000
2022	Observed	42.0749	53.8895	4.0305	0.0052	0.0000
2022	GLDM	40.4177	53.6470	5.8799	0.0554	0.0000
2022	Penalized GLDM	35.2972	57.5727	7.0907	0.0394	0.0000
2022	Inverse-power	34.5380	57.3411	8.0873	0.0336	0.0000
2023	Observed	37.9701	50.0721	11.4645	0.4911	0.0022
2023	GLDM	40.8349	53.3837	5.7234	0.0579	0.0000
2023	Penalized GLDM	35.3701	57.5664	7.0237	0.0397	0.0000
2023	Inverse-power	34.5780	57.3375	8.0509	0.0336	0.0000
2024	Observed	40.6538	50.1857	9.0244	0.1357	0.0004
2024	GLDM	41.2511	53.1211	5.5674	0.0603	0.0000
2024	Penalized GLDM	35.4432	57.5601	6.9567	0.0401	0.0000
2024	Inverse-power	34.6180	57.3339	8.0144	0.0337	0.0000
2025	Observed	43.2658	50.1467	6.5457	0.0418	0.0000
2025	GLDM	41.6664	52.8591	5.4117	0.0628	0.0000
2025	Penalized GLDM	35.5163	57.5537	6.8896	0.0404	0.0000
2025	Inverse-power	34.6580	57.3303	7.9779	0.0338	0.0000

Define the descriptive paired difference

$$\Delta_{t,j} = |e_{t,j}^{q=1,\lambda=0}| - |e_{t,j}^{q=2,\hat{\lambda}_t}|. \quad (\text{E.6})$$

Positive values favor the complete inverse-power procedure; its penalty can differ, so Eq. (E.6) is not the same-penalty effect. The average of all fifty cells is the degree-one rolling-MAE improvement. Figure 44 exposes both favorable and unfavorable class-year contributions rather than reporting only the aggregate.

The mean decrease is 0.0877 percentage points, with 23 favorable and 27 unfavorable class-year cells. They are dependent components of ten annual forecasts. No cell-wise significance test or independent-class confidence interval is attached to them.

E.8. Multistart stability and the unsmoothed MM cross-check. Among the 2,950 primary order-study class fits, 900 have a spread above 10^{-6} between the final objectives of the four starts. The largest spread is 0.270607. These spreads use the same objective within each fit and are recorded before optional best-fit polishing. A deterministic smallest-objective rule resolves starts, with ties ordered by the implementation. A winning-start frequency is not a probability of global optimality.

Twelve representative affine fits cross degrees one, three and five, exponents one and two, and classes one and three, using the fixed $\lambda = 0.1$, full pre-2021 training set and 2025 horizon. The unsmoothed penalized MM solver and continuation use identical four starts and frozen scales. MM permits 200 steps per start and checks the feasible gap bound of Proposition 6.2. Both objectives below are evaluated without smoothing at the stored parameters. The forecast difference is the maximum raw class prediction difference over 2021–2025, before closure.

Ten of the twelve pairs have objective differences below 2×10^{-8} and raw forecast differences below 0.001 percentage points. The other two inverse-power fits reach different basins, with forecast differences about 0.45 and 0.94 percentage points. One favors continuation in objective value and the other favors MM. This directly connects the theorem's



TABLE 53. Degree-one rolling absolute-error decreases by year and class, in percentage points. Negative entries favor ordinary GLDM. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

Year	Class 1	Class 2	Class 3	Class 4	Class 5	Class mean
2016	-0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000
2017	-0.0049	-0.0072	0.0009	0.0130	0.0000	0.0004
2018	0.7694	0.1311	0.6416	-0.0034	-0.0000	0.3078
2019	-0.2589	0.3688	0.6280	0.0003	-0.0000	0.1476
2020	4.3369	3.3948	0.9237	-0.0099	0.0001	1.7291
2021	3.5175	4.0948	-0.6009	-0.0019	-0.0000	1.4019
2022	-3.7367	-3.0926	-0.6434	-0.0007	0.0000	-1.4947
2023	1.0679	-4.0535	1.6140	0.0127	0.0000	-0.2718
2024	-0.0095	-0.0125	0.0011	-0.0230	-0.0000	-0.0088
2025	-3.3784	-1.0476	-0.2438	-0.0003	0.0000	-0.9340

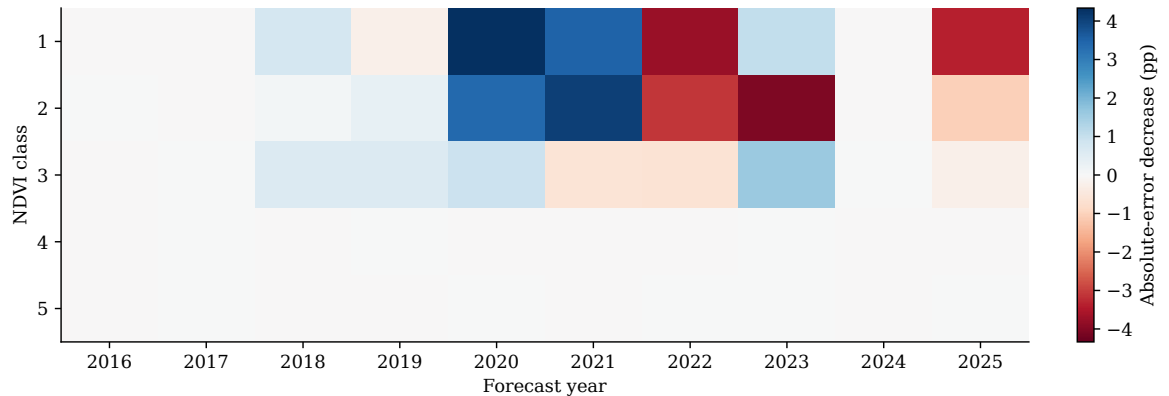


FIGURE 44. Degree-one rolling error decomposition. Blue cells favor the complete inverse-power procedure; red cells favor unpenalized GLDM. Color measures absolute percentage-point error change. Rolling summaries use ten one-year origins with overlapping training windows (50 class-year errors); the holdout, where shown, uses one five-year block (25 errors).

TABLE 54. Primary four-start diagnostics over 2,950 class fits. A non-best result exceeds the best objective by more than 10^{-6} .

Start	Selected as smallest objective	Non-best results
Bounded LS	1229	457
Bounded LAD	675	457
Constant median	598	412
Elevated LAD	448	401



TABLE 55. Unsmoothed MM versus continuation at the same criterion. Small objective differences need not imply small parameter or prediction differences; neither method is globally certified.

m	q	Class	MM objective	Continuation objective	Forecast difference
1	1	1	0.5133109	0.5133109	0.000100
1	1	3	0.4663220	0.4663220	0.000329
1	2	1	0.3961035	0.3961035	0.000002
1	2	3	0.3676399	0.3665990	0.449507
3	1	1	0.5120631	0.5120632	0.000453
3	1	3	0.4636014	0.4636014	0.000008
3	2	1	0.3954087	0.3954087	0.000449
3	2	3	0.3661457	0.3669089	0.943803
5	1	1	0.5116739	0.5116739	0.000020
5	1	3	0.4621654	0.4621654	0.000008
5	2	1	0.3951360	0.3951360	0.000523
5	2	3	0.3656975	0.3656975	0.000083

algorithm with the implemented approximation while showing why local certificates and multistart checks cannot establish unique or global solutions.

E.9. Computational timing and scaling. The records in this subsection document implementation effort. They do not test certificate validity or establish hardware-independent speed rankings. Table 56 reports the trajectory comparisons from Section 9; Table 57 gives the polynomial sensitivity timings.

TABLE 56. Measured median time (ms) and objective calls per selected trajectory fit, including all three starts. Gradient/Jacobian calls have the same count for these paired evaluators.

Method	Logistic ms	Calls	Oscillator ms	Calls	Logistic (3) ms	Calls
ols	3.80	47	21.59	43	5.75	66
huber	4.33	54	26.27	53	6.68	75
cauchy	4.10	55	28.12	56	6.41	76
tukey	5.16	69	37.72	77	8.11	95
q1	131.64	26	198.74	25	138.29	29
q2	156.78	27	190.10	25	197.12	35

The sensitivity study evaluated 7,375 additional independent-class fits, including cached validation reuse, in 120.4 seconds, as recorded for that run. This elapsed total also includes joint-model and MM sensitivity work, data handling and selection; it is not the sum of the table medians. This environment matches Section 9. The primary 2,950-fit January comparison used the separately recorded Intel Xeon environment, so timings from the two environments are not pooled as a controlled hardware comparison.

For n observations and p parameters, a dense residual/Jacobian evaluation outside the ODE solve costs $O(np)$ time and memory. L-BFGS-B uses $O(bp)$ additional memory for history size b , and each affine gradient evaluation costs $O(np)$. A sparse weighted-LAD epigraph model has $n + p$ variables and $2n$ residual inequalities plus parameter constraints, with $O(np + n)$ stored design entries; generic dense factorization can cost $O((n + p)^3)$ time and $O((n + p)^2)$ memory. This worst-case bound is not a complexity claim for OSQP's sparse implementation. Reusing symbolic factorizations and exploiting structured sensitivities can help. With K accepted MM steps, S starts and V validation fits, the work multiplies accordingly; caching repeated training splits reduces V . Forward ODE sensitivities augment a state of dimension d_x by $d_x p$ entries. Larger stiff systems may require adjoints, iterative linear algebra or parallel starts, which are not benchmarked here.



TABLE 57. Measured median cost over five classes for the 2020 cutoff, 2025 horizon and common $\lambda = 0.1$. Each fit includes all four starts, all continuation levels and any best-fit refinement. Pilot construction, validation and final diagnostics are excluded from these per-fit values.

m	q	Time (ms)	Objective calls	Gradient calls
1	1	10.69	200	200
1	2	12.40	237	237
2	1	18.01	367	367
2	2	14.16	283	283
3	1	17.24	357	357
3	2	16.50	302	302
4	1	24.65	540	540
4	2	17.12	334	334
5	1	29.46	599	599
5	2	19.36	370	370

12. DISCUSSION

Theorem 8.1 separates numerical optimization from continuous-model stationarity. Its distinguishing step is to preserve a feasible witness when discretization can move a residual across the cusp. Observation-map and sensitivity errors then give an explicit inflation, while tube enlargement preserves admissible slopes. The comparison uses one parameter throughout. This is a cusp-aware, implementable specialization of perturbation reasoning, supported by established majorization tools.

The evidence follows the levels in Table 9. Validated computation consists of twelve pointwise scalar exact-rational enclosures. Numerical verification comprises the 160 oscillator checks and the January ODE study on three RK4 meshes. The oscillator shows that omitting enlargement can fail even with stable propagation. The January ODE shows how measured error terms enter a practical budget decision. For that fitted relaxation, integration error remains below the prescribed numerical budgets, and mesh halving reduces the measured inflation by approximately sixteen. The two exponents yield trajectories that agree to the reported precision; their fitted rate and equilibrium remain sensitive to the search box.

Stationarity and prediction answer different questions. In STAT5, ordinary GLDM has held-out MAE 14.195 percentage points under both searches. Inverse-power GLDM changes from 6.668 under the original seven-start procedure to 13.214 after expanded criterion-based selection. The lower training objective gives a worse chronological forecast despite a small tube residual. Ordinary GLDM also has the lowest mean forecast RMSE in all four matched Laplace blocks. Inverse-power GLDM is therefore a diagnosable loss-shape option, not a preferred predictor on this evidence. Changing q changes curvature, saturation and tail decay together; comparisons must state the scale and penalty convention. The reciprocal GLDM representation remains an exact identity and implementation check.

The separate January polynomial study primarily demonstrates the effect of regularization. At degree five, rolling MAE falls from 4.013 to 2.034 percentage points when the penalty is added to ordinary GLDM. At the same numerical penalty, inverse-power GLDM gives 2.024. Ten origins with overlapping training windows and one five-year holdout do not establish meaningful superiority for that small difference. These forecasts describe the supplied class compositions, whereas the January ODE study evaluates certificate transfer. Incomplete satellite provenance and unmatched map-display intervals limit both to that tabulated observational basis.

The grouped scope restrictions in Table 9 govern these interpretations. In particular, the theorem does not turn local numerical stationarity into parameter accuracy or predictive reliability. Validated nonlinear state and sensitivity integration is the next computational priority. Independent applications, longer chronological records and complete satellite-processing metadata are needed for stronger empirical conclusions.



13. CONCLUSION

Theorem 8.1 transfers a feasible stationarity bound from a discretized ODE observation map to the continuous map at the same parameter. Observation-map and sensitivity errors determine the explicit inflation. Residual-tube enlargement handles cusp crossings without assuming that residual signs are preserved.

Twelve pointwise scalar inequalities have exact-rational enclosures. The 160 oscillator checks and the Khanty-Mansiysk January ODE calibration on three meshes provide numerical verification using floating-point references. These results support the transfer mechanism within their stated evidence levels; they do not validate general nonlinear integration or establish parameter accuracy and predictive superiority. The method provides a computable and explicitly scoped transfer of local stationarity evidence from a discretized ODE observation map to its continuous counterpart.

AUTHOR CONTRIBUTIONS

Tatyana Vokhmintceva, Aleksander Vokhmintcev, and Mostafa Abotaleb conceived the methodology, developed the proposed method, conducted the formal and numerical analyses, and developed the computational implementation. Alfred Daci, Klodian Dhoska, Marek Dudek, and Nigar Nagiyeva contributed to data curation, validation, visualization, interpretation of the results, and manuscript review and editing. Mostafa Abotaleb supervised the research.

DATA AND CODE AVAILABILITY

The January dataset contains MODIS-derived NDVI-class counts and percentages for Khanty-Mansiysk. The reported preparation comprised NDVI calculation, image correction and combination, mapping for January, May, July and October, and threshold classification. Only January is analyzed. The all-zero 2000 record is excluded, leaving 25 compositions for 2001–2025. The public STAT5 and SIR inputs retain their benchmark attribution [7, 8, 38, 44]. The authors' computational release has not yet received a public archival DOI; the PETab DOI identifies the external benchmark collection.

ETHICS STATEMENT

The analysis uses synthetic observations, aggregated MODIS-derived January class statistics, published cell-line measurements from the STAT5 benchmark, and public aggregate case-count series. No new human, animal or laboratory experiment was performed, and no personal data are analyzed.

FUNDING

The study was funded by the Foundation for Scientific and Technological Development of Yugra under project No. 2026-252-01.

REFERENCES

- [1] M. Abotaleb, *Proving optimal model selection and zero coefficient cases in time series forecasting with the generalized least deviation algorithm*, Adv. Math. Models Appl., 9(3) (2024), 475–524.
- [2] M. O. Abotaleb, *Solving the optimizing parameters problem for non-linear datasets using the high-order general least deviations method (GLDM) algorithm*, Comput. Methods Differ. Equ., 13(3) (2025), 940–967.
- [3] M. Abotaleb and T. Makarovskikh, *Advanced milk production modelling using high-order generalized least deviation method*, Model. Earth Syst. Environ., 10(6) (2024), 7019–7047.
- [4] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Math. Program., 137(1–2) (2013), 91–129.
- [5] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2(1) (2009), 183–202.
- [6] J. Beck, N. Bosch, M. Deistler, K. L. Kadhim, J. H. Macke, P. Hennig, and P. Berens, *Diffusion tempering improves parameter estimation with probabilistic integrators for ordinary differential equations*, Proc. Mach. Learn. Res., 235 (2024), 3305–3326. <https://proceedings.mlr.press/v235/beck24a.html>.



- [7] A. L. Bertozzi, E. Franco, G. Mohler, M. B. Short, and D. Sledge, *The challenges of modeling and forecasting the spread of COVID-19*, Proc. Natl. Acad. Sci. USA, *117*(29) (2020), 16732–16738.
- [8] M. E. Boehm, L. Adlung, M. Schilling, S. Roth, U. Klingmüller, and W. D. Lehmann, *Identification of isoform-specific dynamics in phosphorylation-dependent STAT5 dimerization by quantitative mass spectrometry and mathematical modeling*, J. Proteome Res., *13*(12) (2014), 5685–5694.
- [9] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., *18*(2) (2007), 556–572.
- [10] D. M. Bortz, D. A. Messenger, and V. Dukic, *Direct estimation of parameters in ODE models using WENDy: Weak-form estimation of nonlinear dynamics*, Bull. Math. Biol., *85*(11) (2023), 110.
- [11] J. Bundo, D. Panxhi, G. Sharko, and D. Dharmo, *Impact of operating conditions on linear ARX models for air source heat pumps*, J. Trans. Syst. Eng., *4*(1) (2026), 560–572.
- [12] Ceres Solver developers, *Modeling non-linear least squares: loss functions*, Ceres Solver documentation, https://ceres-solver.readthedocs.io/latest/nlls_modeling.html, accessed 3 October 2026.
- [13] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, *Generalized correntropy for robust adaptive filtering*, IEEE Trans. Signal Process., *64*(13) (2016), 3376–3387.
- [14] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
- [15] D. Q. F. de Menezes, D. M. Prata, A. R. Secchi, and J. C. Pinto, *A review on robust M-estimators for regression analysis*, Comput. Chem. Eng., *147* (2021), 107254.
- [16] K. Didan, A. Barreto Muñoz, R. Solano, and A. Huete, *MODIS Vegetation Index User's Guide (MOD13 Series)*, Version 3.00, Collection 6, University of Arizona, Tucson, June 2015. https://vip.arizona.edu/documents/MODIS/MODIS_VI_UsersGuide_June_2015_C6.pdf.
- [17] K. Didan, *MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061*, NASA Land Processes Distributed Active Archive Center, 2021, dataset.
- [18] D. Drusvyatskiy and C. Paquette, *Efficiency of minimizing compositions of convex functions and smooth maps*, Math. Program., *178*(1–2) (2019), 503–558.
- [19] Y. Feng, J. Fan, and J. A. K. Suykens, *A statistical learning approach to modal regression*, J. Mach. Learn. Res., *21*(2) (2020), 1–35. <https://jmlr.org/papers/v21/17-068.html>.
- [20] Y. Feng, X. Huang, L. Shi, Y. Yang, and J. A. K. Suykens, *Learning with the maximum correntropy criterion induced losses for regression*, J. Mach. Learn. Res., *16*(30) (2015), 993–1034. <https://jmlr.org/papers/v16/feng15a.html>.
- [21] F. Fröhlich, D. Weindl, Y. Schälte, D. Pathirana, L. Paszkowski, G. T. Lines, P. Stapor, and J. Hasenauer, *AMICI: High-performance sensitivity analysis for large ordinary differential equation models*, Bioinformatics, *37*(20) (2021), 3676–3677.
- [22] M. Gheibi, R. Moezzi, M. Khaleghiabbasabadi, H. Taghavian, and K. Dhoska, *Intelligent estimation of total suspended solids (TSS) in wastewater treatment plants utilizing non-linear regression analysis*, J. Trans. Syst. Eng., *1*(1) (2023), 50–55.
- [23] H. Günerhan, M. Yiğider, J. Manafian, and O. A. İlhan, *Numerical solution of fractional order logistic equations via conformable fractional differential transform method*, J. Interdiscip. Math., *24*(5) (2021), 1207–1220.
- [24] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, 2nd ed., Springer, Berlin, 1993.
- [25] F. R. Hampel, *The influence curve and its role in robust estimation*, J. Amer. Statist. Assoc., *69*(346) (1974), 383–393.
- [26] L. T. K. Hien, D. N. Phan, and N. Gillis, *An inertial block majorization minimization framework for nonsmooth nonconvex optimization*, J. Mach. Learn. Res., *24*(18) (2023), 1–41. <https://jmlr.org/papers/v24/21-0571.html>.
- [27] P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist., *35*(1) (1964), 73–101.
- [28] R. Koenker and G. Bassett, Jr., *Regression quantiles*, Econometrica, *46*(1) (1978), 33–50.
- [29] P. Lakrisenko, D. Pathirana, D. Weindl, and J. Hasenauer, *Benchmarking methods for computing local sensitivities in ordinary differential equation models at dynamic and steady states*, PLoS ONE, *19*(10) (2024), e0312148.
- [30] K. Lange, D. R. Hunter, and I. Yang, *Optimization transfer using surrogate objective functions*, J. Comput. Graph. Statist., *9*(1) (2000), 1–20.
- [31] H. Lyu and Y. Li, *Block majorization-minimization with diminishing radius for constrained nonsmooth nonconvex optimization*, SIAM J. Optim., *35*(2) (2025), 842–871.
- [32] C. Maier, C. Loos, and J. Hasenauer, *Robust parameter estimation for dynamical systems from outlier-corrupted data*, Bioinformatics, *33*(5) (2017), 718–725.
- [33] J. Manafian, *An optimal Galerkin-homotopy asymptotic method applied to the nonlinear second-order BVPs*, Proc. Inst. Math. Mech. Natl. Acad. Sci. Azerb., *47*(1) (2021), 156–182.



- [34] J. Manafian, M. Fazli, O. A. Ilhan, S. Malmir, and S. T. Ghafel, *Enhancing Kalman filter performance: Design and estimation with fading-memory Kalman filter for reduced modeling error and improved robustness*, Adv. Eng. Intell. Syst., 4(1) (2025), 1–26.
- [35] M. C. Mukkamala, J. Fadili, and P. Ochs, *Global convergence of model function based Bregman proximal minimization algorithms*, J. Global Optim., 83(4) (2022), 753–781.
- [36] A. V. Panyukov and Y. A. Mezaal, *Stable identification of linear autoregressive model with exogenous variables on the basis of the generalized least absolute deviation method*, Bull. South Ural State Univ. Ser. Math. Model. Program. Comput. Softw., 11(1) (2018), 35–43.
- [37] A. V. Panyukov and Y. A. Mezaal, *Improving of the identification algorithm for a quasilinear recurrence equation*, in: N. Olenov, Y. Evtushenko, M. Khachay, and V. Malkova (eds.), *Advances in Optimization and Applications, OPTIMA 2020*, Commun. Comput. Inf. Sci., Vol. 1340, Springer, Cham, 2020, pp. 15–26.
- [38] PETab Benchmark Collection contributors, *The PETab Benchmark Collection of parameter estimation problems*, Zenodo and GitHub, continuously updated software and model collection.
- [39] S. Pourghanbar, J. Manafian, M. Ranjbar, A. Aliyeva, and Y. S. Gasimov, *An efficient alternating direction explicit method for solving a nonlinear partial differential equation*, Math. Probl. Eng., 2020 (2020), Article ID 9647416, 12 pages.
- [40] A. Raza, M. Noor-ul-Amin, A. Ayari-Akkari, M. Nabi, and M. U. Aslam, *A redescending M-estimator approach for outlier-resilient modeling*, Sci. Rep., 14 (2024), 7131.
- [41] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
- [42] N. Rummel, D. A. Messenger, S. Becker, V. Dukic, and D. M. Bortz, *WENDy for Nonlinear-in-Parameters ODEs*, arXiv:2502.08881v3, 2025, preprint.
- [43] Y. Schälte, F. Fröhlich, P. J. Jost, J. Vanhoefer, D. Pathirana, P. Stapor, P. Lakrisenko, D. Wang, E. Raimúndez, S. Merkt, L. Schmiester, P. Städter, S. Grein, E. Dudkin, D. Doresic, D. Weindl, and J. Hasenauer, *pyPESTO: A modular and scalable tool for parameter estimation for dynamic models*, Bioinformatics, 39(11) (2023), btad711.
- [44] L. Schmiester, Y. Schälte, F. T. Bergmann, T. Camba, E. Dudkin, J. Egert, F. Fröhlich, L. Fuhrmann, A. L. Hauber, S. Kemmer, P. Lakrisenko, C. Loos, S. Merkt, W. Müller, D. Pathirana, E. Raimúndez, L. Refisch, M. Rosenblatt, P. L. Stapor, P. Städter, D. Wang, F.-G. Wieland, J. R. Banga, J. Timmer, A. F. Villaverde, S. Sahle, C. Kreutz, J. Hasenauer, and D. Weindl, *PETab—Interoperable specification of parameter estimation problems in systems biology*, PLoS Comput. Biol., 17(1) (2021), e1008646.
- [45] M. Shahriari, B. Nemati Saray, and J. Manafian, *Reconstruction of the Sturm–Liouville differential operators with discontinuity conditions and a constant delay*, Indian J. Pure Appl. Math., 51(2) (2020), 659–668.
- [46] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, *OSQP: An operator splitting solver for quadratic programs*, Math. Program. Comput., 12(4) (2020), 637–672.
- [47] S. Uzun, D. Luo, B. Açıkmeşe, and A. Y. Aravkin, *A proximal method for composite optimization with smooth and convex components*, arXiv:2512.20602v1, 2025, preprint.
- [48] T. M. Vu, R. Moezzi, J. Cyrus, and K. Dhoska, *Calculation and simulation of vehicle steering dynamics*, Int. J. Innov. Technol. Interdiscip. Sci., 5(3) (2022), 942–953.
- [49] T. M. Vu, R. Moezzi, and K. Dhoska, *Model predictive control with softened constraints for hybrid electric vehicle*, Int. J. Innov. Technol. Interdiscip. Sci., 6(2) (2023), 1130–1149.
- [50] V. J. Yohai, *High breakdown-point and high efficiency robust estimates for regression*, Ann. Statist., 15(2) (1987), 642–656.
- [51] R. Zhang, Y. Mei, J. Shi, and H. Xu, *Robustness and tractability for nonconvex M-estimators*, Statist. Sinica, 32(3) (2022), 1295–1316.
- [52] Y. Zhang, X. Zhou, Y. Zhang, S. Li, and S. Liu, *Improving time series forecasting in frequency domain using a multi resolution dual branch mixer with noise insensitive ArcTanLoss*, Sci. Rep., 15 (2025), 12557.