Research Paper
Computational Methods for Differential Equations
http://cmde.tabrizu.ac.ir
Vol. *, No. *, *, pp. 1-22
DOI:10.22034/cmde.2026.74842.3889

Arctangent loss for parameter estimation and model selection in nonlinear dynamical systems

Saimir Tola[1], **Klodian Dhoska**[2,*], **Aleksander Vokhmintcev**[3], **Mostafa Abotaleb**[4,*], **Veena Beleyur**[5], **and Nigar Nagiyeva**[6]

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

[2]Department of Mechanics, Polytechnic University of Tirana, Albania.

[3]Chelyabinsk State University, Chelyabinsk, Russia.

[4]Engineering School of Digital Technologies, Yugra State University, Khanty-Mansiysk, Russia.

[5]Symbiosis Institute of Technology, Pune campus, Symbiosis International (Deemed University), Pune, India.

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

Abstract

A local minimum of arctangent general least deviations (GLDM) loss need not determine which of two dynamical models fits better. For residuals affine in the coefficients, a chord drawn across each residual interval yields a weighted absolute-deviation relaxation, whose dual gives a lower bound on the global minimum over a prescribed box. Together with a feasible upper value, this bound can settle model comparisons before all competing minima have been computed to the same tolerance. In 1,023 measured tank transitions the score intervals separate two candidate models at five penalties. For 26 annual May NDVI compositions in Khanty-Mansiysk, adding a penalty on changes between adjacent coefficients reduces degree-five GLDM holdout error from 12.287 to 5.495 percentage points, while the historical mean gives the smallest rolling error among nine procedures. The regional results provide a benchmark for vegetation monitoring and show which changes in the reported shares require spatial follow-up.

Keywords. Nonlinear differential equations, Robust identification, Global optimization, Model selection, Majorization–minimization, NDVI, Khanty-Mansiysk.

2010 Mathematics Subject Classification. 65L09, 65G20, 90C26, 62M10.

Uncorrected Proof

1. Introduction

Two numerical fits can differ because their models describe the observations differently, or because one solver stopped farther from the minimum of its criterion. This distinction matters for GLDM: the bounded arctangent loss reduces the contribution of a large residual, but its nonconvexity prevents a local stationarity test from settling a global comparison. We retain the established GLDM loss and its reciprocal representation [1, 16, 17] and study how to enclose the value of a fit. Related approaches address robust dynamical estimation, sparse model selection and weak-form identification [4, 7, 12, 13].

The central construction places a chord below each loss on its attainable residual interval and combines those chords through a common coefficient vector. A feasible dual witness gives an objective lower bound, and a curvature estimate distinguishes the error due to the chords from the error due to the numerical witness. We give the assumptions under which optimization accuracy can be translated into statistical selection or a bound on a trajectory, then test the finite-box calculation on synthetic data and measured tank transitions.

For the Yugra project on overmoistening, waterlogging and forest–wetland degradation, the separate empirical question is whether a calibrated trajectory improves on the historical May composition in Khanty-Mansiysk. Section 10

Received: 20 September 2026; Accepted: 16 November 2026.
∗ Corresponding authors. Emails: kdhoska@fim.edu.al, abotalebmostafa@bk.ru.

addresses it with the reported 2000–2025 record, comparing forecast errors at chronological origins and retaining the maps against which changes in the regional percentages can be examined.

2. Identification problem and robust loss

2.1. **Differential equations, weak observations, and recurrences.** Let m index a finite model family $\mathcal{M}$. A vector-field candidate is

$$\dot{x}(t) = f_m(t, x(t); \theta) = f_0(t, x(t)) + A_m(t, x(t))\theta, \qquad \theta \in B_m = [\ell_m, u_m] \subset \mathbb{R}^{q_m}. \tag{2.1}$$

The coefficient box is compact and fixed before fitting. Nonlinearity in the state is unrestricted subject to the regularity stated for each result. Affinity in θ is essential for the arrangement and convex-surrogate results. Stacking observed derivatives and subtracting f_0 gives a scalar regression $y_i = X_{i,m}\theta + e_i$. Component-specific scales can be absorbed into rows of X and y.

State observations lead to another affine regression. For a test function $\varphi_i \in C^1$ supported on $[a_i, b_i]$ with zero endpoint values, integration by parts gives

$$y_i = -\int_{a_i}^{b_i} \varphi_i'(t)x(t)\,\mathrm{d}t - \int_{a_i}^{b_i} \varphi_i(t)f_0(t, x(t))\,\mathrm{d}t, \tag{2.2}$$

$$X_{i,m} = \int_{a_i}^{b_i} \varphi_i(t)A_m(t, x(t))\,\mathrm{d}t, \qquad y_i = X_{i,m}\theta. \tag{2.3}$$

Each vector equation is stacked componentwise. In practice, measured states and quadrature replace the integrals. The resulting problem has errors in variables; robust residual fitting does not by itself remove that bias.

For a scalar recurrence with observed exogenous input u_t, write

$$Y_t = u_t + \theta^{\mathsf{T}}\Phi_{p,d}(Z_t) + e_t, \qquad Z_t = (Y_{t-1}, \ldots, Y_{t-p})^{\mathsf{T}}. \tag{2.4}$$

The full polynomial dictionary without an intercept has $q(p, d) = \binom{p+d}{d} - 1$ coefficients. In particular, $q(p, 2) = p(p + 3)/2$. All candidates use the same response times, the same scale, and weights computed once from the largest dictionary. Otherwise raw empirical sums need not be comparable.

2.2. **Frozen scale and covariate weights.** For one candidate, write $q = q_m$, omit m and define

$$F(\theta) = \sum_{i=1}^{N} v_i\rho_a(r_i(\theta)), \qquad r_i(\theta) = y_i - X_i\theta, \qquad \rho_a(r) = \arctan(|r|/a), \quad a > 0, \quad 0 \le v_i \le 1. \tag{2.5}$$

The weights and a are fixed throughout optimization and certification. A practical scale is $a = cs$, where s is the median absolute deviation of pilot LAD residuals multiplied by 1.4826. A numerical floor is recorded. Optimizing an unpenalized scale jointly with θ is inadmissible: $F(\theta, a) \to 0$ as $a \to \infty$.

For a positive diagonal feature scale matrix D and full-dictionary row Φ_i, use

$$v_i = \min\{1, \kappa/\left\|D^{-1}\Phi_i\right\|_2\}, \qquad v_i = 1 \text{ if } \Phi_i = 0. \tag{2.6}$$

Then $v_i \left\|D^{-1}\Phi_i\right\|_2 \le \kappa$. This limits a row's coefficient-score contribution, including that of polynomial leverage points. It does not establish a finite-sample breakdown point. Estimation of D adds uncertainty; fixed-weight deterministic results hold conditional on its realized value.

Lemma 2.1 (Loss and scale perturbations)**.** *For $a, b > 0$ and $r, s \in \mathbb{R}$,*

$$0 \le \rho_a(r) < \pi/2, \qquad |\rho_a(r) - \rho_a(s)| \le |r - s|/a, \tag{2.7}$$

$$\sup_r |\rho_a(r) - \rho_b(r)| \le \tfrac{1}{2}|\log(a/b)|. \tag{2.8}$$

Where $r \ne 0$, $\psi_a(r) = \rho_a'(r) = a\,\mathrm{sign}(r)/(a^2 + r^2)$. In coordinates $\gamma = D\theta$, the norm of an individual weighted score is at most κ/a when X_i is a subrow of Φ_i.

Uncorrected Proof

Proof. The derivative of $t \mapsto \arctan(t/a)$ on $t \geq 0$ is at most $1/a$; compose with the absolute value. Differentiation with respect to $\log a$ gives $-a|r|/(a^2+r^2)$, whose magnitude is at most $1/2$. Integrate this derivative between $\log a$ and $\log b$. Finally, the score in γ coordinates is $-v_i D^{-1} X_i^{\mathsf{T}} \psi_a(r_i)$, and apply (2.6). □

On a bounded residual range, $a\rho_a(r) \to |r|$ as $a \to \infty$, so the limiting loss after rescaling is absolute rather than squared error.

2.3. **Reciprocal GLDM representation and inverse-power loss.** Ordinary inverse GLDM is a complementary representation of the same objective, not a second estimator. For $u \geq 0$ define

$$G_a(u) = \frac{2}{\pi}\arctan(a/u), \qquad \phi_{a,q_\rho}(u) = \frac{\pi}{2q_\rho}\{1 - G_a(u)^{q_\rho}\}, \quad q_\rho \geq 1, \qquad G_a(0) = 1. \tag{2.9}$$

The exponent q_ρ is distinct from the coefficient dimension q_m. For $q_\rho = 1$, $\phi_{a,1}(u) = \arctan(u/a)$ and minimizing the loss is exactly equivalent to maximizing $\sum_i v_i G_a(|r_i|)$. Exponents greater than one change the estimator. The factor $1/q_\rho$ keeps the right derivative at zero equal to $1/a$; hence comparisons at a common penalty do not merely change that local slope.

Proposition 2.2 (Geometry of the inverse-power family)**.** *For $a > 0$ and $q_\rho \geq 1$, ϕ_{a,q_ρ} is increasing, concave and $1/a$-Lipschitz on $[0,\infty)$. Its upper limit is $\pi/(2q_\rho)$, and*

$$\phi'_{a,q_\rho}(u) = \frac{aG_a(u)^{q_\rho-1}}{a^2+u^2}, \tag{2.10}$$

$$\phi''_{a,q_\rho}(u) = -\frac{2auG_a(u)^{q_\rho-1}}{(a^2+u^2)^2} - \frac{2(q_\rho-1)a^2G_a(u)^{q_\rho-2}}{\pi(a^2+u^2)^2}. \tag{2.11}$$

For $q_\rho = 1$ the second term is identically zero. The score of $r \mapsto \phi_{a,q_\rho}(|r|)$ decays as $O(|r|^{-q_\rho-1})$. Its tangent yields a weighted absolute-deviation upper majorizer. For affine residuals without an added convex penalty, residual-cell concavity and the shifted-chord dual construction remain valid with ϕ replaced by ϕ_{a,q_ρ}.

Proof. Differentiate G_a to obtain $G'_a(u) = -2a/[\pi(a^2+u^2)]$ and then (2.10)–(2.11). The first derivative lies in $(0, 1/a]$ and the second is nonpositive. The limit follows from $G_a(u) \to 0$. Since $G_a(u) \sim 2a/(\pi u)$, substitution in (2.10) gives the tail order. Concavity supplies tangent upper bounds and chord lower bounds. On each residual-sign cell, $|r_i|$ is affine, so composition with the increasing concave function preserves concavity. The proofs of Theorems 4.1 and 4.2 then apply with the new chord slopes. Adding a convex coefficient penalty need not preserve cell concavity, so the vertex assertion is not claimed for the penalized fits below. □

At $u = a$, increasing the exponent multiplies the ordinary score by 2^{1-q_ρ}, so the change affects moderate residuals as well as extreme observations, even though the cusp slope remains $1/a$. Since the loss range decreases to $\pi/(2q_\rho)$, a fixed numerical penalty becomes stronger relative to that range. For $q_\rho = 2$, the formula $A - A^2/\pi$, with $A = \arctan(u/a)$, avoids cancellation near zero; smoothing $|r|$ to $\sqrt{r^2+\epsilon^2}$ changes a weighted objective by at most $\epsilon \sum_i v_i/a$.

2.4. **Population curvature despite the cusp.**

Proposition 2.3 (Density smoothing)**.** *Let an error have bounded continuous density g, and set $\ell(b) = \mathbb{E}\rho_a(e-b)$. Then ℓ is twice continuously differentiable and*

$$\ell''(b) = \frac{2g(b)}{a} - 2a\int_{\mathbb{R}} \frac{|u|g(b+u)}{(a^2+u^2)^2}\,\mathrm{d}u. \tag{2.12}$$

If g is symmetric about zero, $\ell'(0) = 0$. If also $g(u) \leq g(0)$ and the inequality is strict on a set of positive measure, then $\ell''(0) > 0$.

Proof. The score has jump $2/a$ at zero and ordinary derivative $-2a|u|/(a^2+u^2)^2$ elsewhere. Its distributional derivative is the sum of that integrable kernel and $(2/a)\delta_0$. Convolving with the bounded continuous density gives (2.12). Continuity follows by dominated convergence, first on compact intervals and then using integrable tails. Symmetry makes the first derivative vanish. Since $\int_{\mathbb{R}} 2a|u|/(a^2+u^2)^2\,\mathrm{d}u = 2/a$, replacing $g(u)$ by $g(0)$ bounds the integral term by $2g(0)/a$, strictly under the stated condition. □

Uncorrected Proof

For conditionally symmetric errors, positive conditional curvature and a nonsingular weighted feature Gram matrix give positive population curvature. This gives local population identifiability and quadratic risk growth. Proposition 6.3 states conditions that carry this curvature into the empirical growth requirement. It does not automatically supply the deterministic empirical constant μ required in Theorem 7.3.

3. Local fitting and the nonsmooth limit

3.1. **Proximal weighted absolute deviations.** Let $\phi(t) = \arctan(t/a)$ for $t \geq 0$. Its derivatives are

$$\phi'(t) = \frac{a}{a^2 + t^2}, \qquad \phi''(t) = -\frac{2at}{(a^2 + t^2)^2}, \qquad \sup_{t \geq 0} |\phi''(t)| = \frac{9}{8\sqrt{3}\, a^2} = C_a. \tag{3.1}$$

The tangent to this concave function is an upper bound. Given θ^k, define

$$w_i^k = \frac{v_i a}{a^2 + r_i(\theta^k)^2}, \tag{3.2}$$

$$\theta^{k+1} = \arg\min_{\theta \in B} \left\{ \sum_i w_i^k |r_i(\theta)| + \frac{\mu_k}{2} \left\| \theta - \theta^k \right\|_2^2 \right\}, \quad 0 < \underline{\mu} \leq \mu_k \leq \overline{\mu}. \tag{3.3}$$

The subproblem is strongly convex and has a unique solution. Introduce auxiliary $t_i \geq \pm r_i(\theta)$ to obtain a box-constrained convex quadratic program. The proximal term stabilizes computation and is absent from the final statistical objective.

Proposition 3.1 (Descent and relative stationarity)**.** *Exact solutions of (3.3) satisfy, with* $d_k = \left\| \theta^{k+1} - \theta^k \right\|_2$,

$$F(\theta^k) - F(\theta^{k+1}) \geq \tfrac{1}{2} \underline{\mu} d_k^2, \tag{3.4}$$

$$\operatorname{dist}(0, \partial(F + \iota_B)(\theta^{k+1})) \leq C d_k, \quad C = \overline{\mu} + C_a \sum_i v_i \left\| X_i \right\|_2^2. \tag{3.5}$$

Here ∂ is the limiting subdifferential and ι_B is zero on B and infinity elsewhere. For $\theta \in B$, the convex normal cone is $N_B(\theta) = \{n : \langle n, z - \theta \rangle \leq 0$ for all $z \in B\}$. Unsubscripted vector norms are Euclidean.

Proof. Concavity gives $\phi(t) \leq \phi(t_k) + \phi'(t_k)(t - t_k)$. Summing and comparing the subproblem at θ^{k+1} and θ^k proves (3.4). At the new point, optimality supplies $s_i \in \partial |r_i(\theta^{k+1})|$ and $n \in N_B(\theta^{k+1})$ such that $\sum_i w_i^k s_i + \mu_k(\theta^{k+1} - \theta^k) + n = 0$. Replacing w_i^k by w_i^{k+1} produces a subgradient of the target extended objective. The difference has norm at most $(\overline{\mu} + C_a \sum_i v_i \left\| X_i \right\|^2) d_k$. The subgradient sum is exact here: the loss is a smooth function plus a positive multiple of an absolute value locally at each zero residual; equivalently a sufficiently large quadratic makes it convex. Convex normal-cone calculus then applies on the box [19]. □

Theorem 3.2 (Convergence of the complete PMM sequence)**.** *On a nonempty compact box, the exact iteration (3.3) has finite length, converges to one stationary point θ^∞, and satisfies $\sum_k d_k < \infty$. If its local KL desingularizer is $\varphi(s) = As^{1-\vartheta}$, $A > 0$, $0 \leq \vartheta < 1$, then the objective gap $e_k = F(\theta^k) - F(\theta^\infty)$ has finite termination when $\vartheta = 0$, geometric decay when $0 < \vartheta \leq 1/2$, and*

$$e_k = O(k^{-1/(2\vartheta - 1)}), \qquad \left\| \theta^k - \theta^\infty \right\| = O(k^{-(1-\vartheta)/(2\vartheta - 1)}) \quad \text{when } \vartheta > 1/2. \tag{3.6}$$

Proof. The restricted analytic arctangent, absolute affine residuals, and box indicator define a subanalytic extended objective on a neighborhood of B, so the KL property applies [3]. Descent gives convergence of objective values and $d_k \to 0$. Compactness and (3.5) imply that cluster points are stationary and have the same objective value. Uniformizing the KL inequality over that compact cluster set gives, for all large k, $1 \leq \varphi'(e_k) C d_{k-1}$. With $a_0 = \underline{\mu}/2$, concavity of φ yields

$$d_k^2 \leq (C/a_0) d_{k-1} [\varphi(e_k) - \varphi(e_{k+1})]. \tag{3.7}$$

The inequality $\sqrt{uv} \leq (u + v)/2$ and summation show that the tail length is finite; the sequence is Cauchy. For the power desingularizer, (3.4)–(3.5) yield $e_{k-1} - e_k \geq c_0 e_k^{2\vartheta}$. Comparison of this scalar recurrence with a geometric

Uncorrected Proof

or power sequence gives the stated objective rates; summing (3.7) gives the iterate rate. This is the standard KL argument specialized to the present constants [2]. □

For fixed proximal curvature, a zero surrogate step is equivalent to stationarity because the surrogate and target optimality conditions then agree. An inexact update needs a descent check and a bound on its subgradient residual.

3.2. **Smooth computational approximation.** For dimensionless $\delta > 0$, set

$$F_\delta(\theta) = \sum_i v_i \arctan \sqrt{(r_i(\theta)/a)^2 + \delta^2}. \tag{3.8}$$

The function $q \mapsto \arctan \sqrt{q + \delta^2}$ is increasing and concave for $q \geq 0$. Its tangent gives a weighted least-squares surrogate with weights proportional to

$$\omega_i^k = \frac{v_i}{t_i^k[1 + (t_i^k)^2]}, \qquad t_i^k = \sqrt{(r_i(\theta^k)/a)^2 + \delta^2}. \tag{3.9}$$

A positive proximal quadratic and box constraints preserve strong convexity. This smooth update has a different inner problem from (3.3), so the exact nonsmooth convergence theorem does not by itself certify a finite-precision IRLS run.

Proposition 3.3 (Uniform smoothing and stationary limits)**.** *Writing $V = \sum_i v_i$, one has $0 \leq F_\delta - F \leq \delta V$ uniformly on B. If $\widehat{\theta}_\delta$ is an η-global minimizer of F_δ, then*

$$F(\widehat{\theta}_\delta) - \min_B F \leq \eta + \delta V. \tag{3.10}$$

If $\delta_j \downarrow 0$, $\theta_j \to \bar{\theta}$, and $\operatorname{dist}(0, \nabla F_{\delta_j}(\theta_j) + N_B(\theta_j)) \to 0$, then $0 \in \partial(F + \iota_B)(\bar{\theta})$.

Proof. Use $0 \leq \sqrt{z^2 + \delta^2} - |z| \leq \delta$ and the unit Lipschitz constant of arctangent. Comparing the two minima proves (3.10). Each residual derivative is bounded by v_i/a. On a convergent subsequence it tends to the ordinary derivative if the limiting residual is nonzero, and to a value in $[-v_i/a, v_i/a]$ otherwise. The normal vectors are bounded by stationarity and the bounded gradient; their limits belong to the closed graph of the box normal cone. The resulting sum is an exact subgradient as in Proposition 3.1. □

The scale, weights and coefficient box define the criterion and are frozen before fitting, whereas proximal curvature controls the update and smoothing introduces a recorded approximation allowance. An objective allowance ε_s suggests $\delta \leq \varepsilon_s/V$, although smaller smoothing widths can worsen conditioning near zero residuals. In implementation, multiple feasible starts are compared using the original loss, with surrogate feasibility, descent and subproblem residuals recorded at every accepted update.

3.3. **Nonlinear observation maps and stationarity checks.** Even when the vector field is affine in its rates, its integrated observation map is generally nonlinear in those rates. For such maps write

$$H(\theta) = \sum_i v_i \phi_{a,q_\rho}(|y_i - g_i(\theta)|) + P(\theta), \qquad \theta \in B, \tag{3.11}$$

where B is a nonempty compact convex box, P is convex and continuously differentiable on a neighborhood of B, and g_i is continuously differentiable with L_i-Lipschitz gradient there. Put $K = \sum_i v_i L_i/a$ and $r_i^k = y_i - g_i(\theta^k)$.

Theorem 3.4 (Nonlinear inverse-power descent)**.** *Under the preceding assumptions, choose $\mu_k \in [K + \eta, \overline{\mu}]$ with $\eta > 0$ and finite $\overline{\mu}$. Let θ^{k+1} be the exact minimizer on B of*

$$Q_k(\theta) = \sum_i v_i \phi'_{a,q_\rho}(|r_i^k|)|r_i^k - \nabla g_i(\theta^k)^\mathsf{T}(\theta - \theta^k)| + P(\theta) + \frac{\mu_k}{2}\left\|\theta - \theta^k\right\|^2. \tag{3.12}$$

Then $H(\theta^k) - H(\theta^{k+1}) \geq \eta \left\|\theta^{k+1} - \theta^k\right\|^2/2$, the squared step lengths are summable, and every cluster point is Clarke stationary on B. No assertion of global optimality or statistical consistency follows. Backtracking that starts at a positive lower curvature and increases geometrically until the corresponding majorization inequality holds has a finite acceptance threshold under these assumptions.

Uncorrected Proof

Proof. Taylor's remainder gives $|g_i(\theta^k+d) - g_i(\theta^k) - \nabla g_i(\theta^k)^\mathsf{T} d| \le L_i \|d\|^2/2$. The absolute value is 1-Lipschitz. Use the tangent inequality in Proposition 2.2 and $\phi' \le 1/a$ to obtain

$$H(\theta^k+d) \le H(\theta^k) + Q_k(\theta^k+d) - Q_k(\theta^k) - \tfrac{1}{2}(\mu_k - K)\|d\|^2. \tag{3.13}$$

Exact minimization implies $Q_k(\theta^{k+1}) \le Q_k(\theta^k)$. Summation proves the descent and square-summability claims because H is bounded below on B. The convex subproblem optimality condition, with $d^k = \theta^{k+1} - \theta^k$, contains signs $s_i^k \in \operatorname{Sign}(r_i^k - \nabla g_i(\theta^k)^\mathsf{T} d^k)$, with $\operatorname{Sign}(0) = [-1,1]$, and a box normal. Along a convergent subsequence, $d^k \to 0$; extract convergent signs in $[-1,1]$. The remaining terms are bounded, so the corresponding normals are bounded and have a convergent subsequence. Continuity and closedness of the sign and convex-normal graphs yield

$$0 \in -\sum_i v_i \phi'_{a,q_\rho}(|r_i(\bar\theta)|)\operatorname{Sign}(r_i(\bar\theta))\nabla g_i(\bar\theta) + \nabla P(\bar\theta) + N_B(\bar\theta). \tag{3.14}$$

Near a zero residual the loss is a positive multiple of $|r_i|$ plus a continuously differentiable remainder with zero derivative at the crossing. Thus this condition implies Clarke stationarity. The Taylor bound ensures acceptance whenever the curvature exceeds K by the required margin, proving finite backtracking. □

The theorem concerns exact convex subproblems. An implementation can monitor the subproblem's primal–dual gap and the first-order inclusion (3.14) at the returned point. Neither diagnostic bounds the objective at other points in the coefficient box, which is the purpose of the next construction.

4. Global enclosures on a finite coefficient box

4.1. **Geometry and a strengthened chord bound.** For a residual-sign vector σ, define $C_\sigma = B \cap \{\theta : \sigma_i r_i(\theta) \ge 0\}$. Within this polytope the loss is a sum of concave functions of nonnegative affine arguments.

Theorem 4.1 (Weighted finite-arrangement principle)**.** *There is a global minimizer of F at a vertex of the arrangement of the residual hyperplanes $X_i\theta = y_i$ and box facets. Enumerating intersections of q independent active hyperplanes therefore suffices in exact arithmetic; there are at most $\binom{N+2q}{q}$ candidate intersections before filtering.*

Proof. A global minimizer exists by continuity and compactness and lies in some C_σ. The function is concave on that compact polytope because $v_i \ge 0$. Every point is a convex combination of its vertices; concavity implies that at least one vertex has value no larger. Each vertex is determined by q linearly independent active residual or box hyperplanes. Degenerate cells cause duplicate intersections, not additional candidates. □

This finite principle supplies an independent small-dimensional diagnostic. Floating-point enumeration without rank and feasibility enclosures is not a verified certificate.

For a subbox $B' = [l,u]$, let $z = (l+u)/2$ and define

$$h_i = \tfrac{1}{2}\sum_j |X_{ij}|(u_j - l_j), \qquad [\underline{r}_i, \overline{r}_i] = [r_i(z) - h_i, r_i(z) + h_i], \tag{4.1}$$

$$m_i = \operatorname{dist}(0, [\underline{r}_i, \overline{r}_i]), \qquad M_i = \max\{|\underline{r}_i|, |\overline{r}_i|\}. \tag{4.2}$$

Set $L_I(B') = \sum_i v_i \phi(m_i)$. Unlike an origin chord, the chord on $[m_i, M_i]$ uses both residual endpoints:

$$\alpha_i = v_i \frac{\phi(M_i) - \phi(m_i)}{M_i - m_i}, \qquad b_i = v_i\phi(m_i) - \alpha_i m_i. \tag{4.3}$$

If $M_i = m_i$, take $\alpha_i = 0$, $b_i = v_i\phi(m_i)$. Concavity gives $v_i\phi(t) \ge b_i + \alpha_i t$ for $m_i \le t \le M_i$. The shifted chord is at least as high on this interval as the chord from zero to M_i.

Theorem 4.2 (Coupled dual lower enclosure)**.** *For every ζ with $|\zeta_i| \le \alpha_i$,*

$$L_D(B';\zeta) = \sum_i b_i + y^\mathsf{T}\zeta - \sum_j \max\{l_j(X^\mathsf{T}\zeta)_j, u_j(X^\mathsf{T}\zeta)_j\}, \tag{4.4}$$

$$L(B') = \max\{0, L_I(B'), L_D(B';\zeta)\} \le \min_{\theta\in B'} F(\theta). \tag{4.5}$$

Moreover, maximizing (4.4) over its box constraints equals $\sum_i b_i + \min_{\theta\in B'}\sum_i \alpha_i |y_i - X_i\theta|$.

Uncorrected Proof

Proof. For every feasible θ, concavity gives $F(\theta) \geq \sum_i (b_i + \alpha_i |r_i|)$. Since $\zeta_i r_i \leq \alpha_i |r_i|$, the right-hand side is at least $\sum_i b_i + y^{\mathsf{T}}\zeta - \theta^{\mathsf{T}} X^{\mathsf{T}}\zeta$. Minimizing this affine expression over $[l, u]$ yields (4.4): in coordinate j the contribution is $-\max\{l_j d_j, u_j d_j\}$ for $d = X^{\mathsf{T}}\zeta$. The other two lower values in (4.5) follow from nonnegativity and $|r_i| \geq m_i$. Finally, writing $\alpha_i |r_i| = \max_{|\zeta_i| \leq \alpha_i} \zeta_i r_i$ identifies the weighted-LAD problem and its linear-program dual, so strong duality gives the claimed equality [5]. □

The LP supplies a proposal, while feasibility of the dual witness supplies the guarantee. This specialization of convex underestimation [14, 15] also applies to the concave positive-side losses of Proposition 2.2.

Proposition 4.3 (Chord error and witness accuracy)**.** *For $\phi(t) = \arctan(t/a)$ and the intervals in (4.2), let*

$$S(\theta) = \sum_i \{b_i + \alpha_i |r_i(\theta)|\},$$

$$C_i = \max_{m_i \leq t \leq M_i} \frac{2at}{(a^2 + t^2)^2}, \qquad D_{\mathrm{ch}}(B') = \frac{1}{8} \sum_i v_i C_i (M_i - m_i)^2. \tag{4.6}$$

The maximum defining C_i is attained at the projection of $a/\sqrt{3}$ onto $[m_i, M_i]$, and $C_i \leq 9/(8\sqrt{3}\, a^2)$. Put $S^\star = \min_{B'} S$. Then

$$0 \leq \min_{B'} F - S^\star \leq D_{\mathrm{ch}}(B'). \tag{4.7}$$

For any feasible primal point $\widehat{\theta} \in B'$ and dual witness ζ, define $\Gamma = S(\widehat{\theta}) - L_D(B'; \zeta)$. Then

$$0 \leq \min_{B'} F - L_D(B'; \zeta) \leq \Gamma + D_{\mathrm{ch}}(B'). \tag{4.8}$$

These statements use exact chord coefficients; outward upper bounds on Γ and D_{ch} give their verified numerical versions. If an implemented slope $\widetilde{\alpha}_i \leq \alpha_i$ is anchored at m_i, replace S and L_D consistently by their rounded-slope versions and add

$$D_{\mathrm{round}} = \sum_i (\alpha_i - \widetilde{\alpha}_i)(M_i - m_i) \tag{4.9}$$

to the right-hand side of (4.8). An outward upper enclosure of D_{round} accounts for this additional arithmetic slack.

Proof. Linear interpolation error on $[m_i, M_i]$ is nonnegative by concavity and at most $v_i C_i (t - m_i)(M_i - t)/2 \leq v_i C_i (M_i - m_i)^2/8$. For a degenerate interval it is zero. Summing gives $S \leq F \leq S + D_{\mathrm{ch}}$ on B'. Evaluate the upper inequality at a minimizer of the continuous function S on the compact box to prove (4.7). Weak duality gives $L_D \leq S^\star$, while evaluating at $\widehat{\theta}$ gives $\min F \leq F(\widehat{\theta}) \leq S(\widehat{\theta}) + D_{\mathrm{ch}} = L_D + \Gamma + D_{\mathrm{ch}}$. Differentiating $2at/(a^2 + t^2)^2$ gives the stated curvature maximum. For the last assertion, the difference between an exact and an anchored rounded chord is $(\alpha_i - \widetilde{\alpha}_i)(t - m_i)$, between zero and $(\alpha_i - \widetilde{\alpha}_i)(M_i - m_i)$. Sum this difference and repeat the preceding argument. □

The decomposition identifies where further work can improve a weak bound: subdivision reduces residual widths and D_{ch}, whereas a better LP witness reduces Γ. When an interval crosses zero, $m_i = 0$ and the shifted and origin chords coincide, so additional LP accuracy cannot remove the interval-width contribution.

Proposition 4.4 (Consistency radius)**.** *For the midpoint z of B', define*

$$\Delta(B') = \frac{1}{2a} \sum_j (u_j - l_j) \sum_i v_i |X_{ij}|. \tag{4.10}$$

Then $0 \leq F(z) - L_I(B') \leq \Delta(B')$.

Proof. One has $m_i = \max\{|r_i(z)| - h_i, 0\}$. The Lipschitz bound in Lemma 2.1 gives $0 \leq v_i \phi(|r_i(z)|) - v_i \phi(m_i) \leq v_i h_i / a$. Sum and interchange the two finite sums. □

Uncorrected Proof

Algorithm 1 Verified minimization on a prescribed coefficient box.

Inputs: fixed X, y, v, a, B, target gap ε and expansion budget.

(1) Obtain a feasible local incumbent, evaluate its outward upper value U, and place the root box with its lower bound (4.5) in a priority queue.
(2) Set $L = \min\{U, \min_{B' \text{ in queue}} L(B')\}$; return the enclosure and its status if $U - L \le \varepsilon$ or the work budget is exhausted.
(3) Remove a least-bound box, discarding it only when its lower bound is at least U, and choose an influential coordinate by weighted width or the safeguarded rule (4.11).
(4) Bisect at the coordinate midpoint, returning an unresolved rounding status if no influential coordinate can be bisected at the working precision.
(5) Bound both children, test their midpoints, update U from feasible points and retain every child with lower bound below U, then repeat from step 2.

An empty queue completes the enclosure, while budget and rounding stops retain the unresolved gap.

4.2. **Verified arithmetic and subdivision.** The numerical inputs are the stored binary values, which are converted to Arb balls before residual, arctangent, chord and support-function evaluation at 100-bit precision [11, 20]. Dual proposals are clipped to downward-enclosed slopes and the anchored intercept is recomputed consistently, while exported endpoints are expanded outward to binary64. A numerical LP optimum is never substituted for a verified lower value.

Computing residual intervals and a sign-based dual witness costs $O(Nq_m)$ arithmetic operations per box. An optimized witness additionally solves an LP with $N + q_m$ primal variables and $2N$ residual inequalities. A queue of K boxes needs $O(Kq_m)$ storage, apart from shared data and solver workspaces. Refining each coordinate to width η_j may require a grid with $\prod_j \lceil (u_j - l_j)/\eta_j \rceil$ cells; therefore finite termination does not imply scalable runtime.

The computational study in Section 9 reports unresolved outcomes at the work limit, since finite termination in exact arithmetic does not provide a practical dimension threshold. Comparisons over a larger candidate family also require sufficient lower scores for every excluded model.

Theorem 4.5 (Global invariant and termination)**.** *Algorithm 1 maintains $L \le F^\star \le F(\widehat{\theta}) \le U$. Thus $U - L \le \varepsilon$ certifies an ε-global solution on the stated box. In exact arithmetic, with no node cap and positive ε, weighted-width bisection and midpoint testing terminate after finitely many expansions.*

Proof. Initially the root covers the feasible set. Splitting preserves that cover, and discarding a box with lower bound at least U cannot discard an improving point. The lower bound is the minimum of active bounds and U, which establishes the invariant. For termination, suppose infinitely many boxes are expanded. The finitely branching tree has an infinite nested path. Along it the largest influential weighted width tends to zero, since each maximal width is halved and there are finitely many coordinates. Hence $\Delta(B') \to 0$. The midpoint was tested when each node was created, so, when that node is later selected as a least-bound box, $U - L \le F(z) - L(B') \le \Delta(B')$. Eventually this is below ε, a contradiction. Zero-influence coordinates need not be split. □

At finite precision, a rounding floor or node budget may leave the global objective gap unresolved. Local refinement improves upper values but does not determine lower-bound validity.

4.3. **Safeguarded geometry-based branching.** Let $d_j = (u_j - l_j) \sum_i v_i |X_{ij}|$. Weighted-width branching chooses a largest d_j. To target residual uncertainty more directly, restrict candidate coordinates to

$$\mathcal{J}(B') = \{j : d_j > 0,\ d_j \ge \eta \max_k d_k\}, \qquad \eta = 1/4. \tag{4.11}$$

Denote the two children from bisecting coordinate j by $B'_{j,-}, B'_{j,+}$, and retain the residual half-width h_i in (4.1). Two alternatives are

$$j_R \in \arg\min_{j \in \mathcal{J}(B')} \max_i v_i \left(h_i - \frac{1}{4} |X_{ij}| (u_j - l_j) \right),$$
$$j_D \in \arg\min_{j \in \mathcal{J}(B')} \max_{s \in \{-,+\}} D_{\mathrm{ch}}(B'_{j,s}). \tag{4.12}$$

Uncorrected Proof

The first minimizes the largest weighted child residual half-width. The second minimizes the worst-child chord allowance; ties use the smaller worst-child zero-crossing fraction, then the larger d_j. The width rule also breaks ties by larger d_j. Child half-widths do not depend on which half is retained. Chord allowances do, because the interval centers and zero crossings change. These scores are floating-point branching proposals, not lower bounds. Every retained child still receives the same outward evaluation in Theorem 4.2. The direct chord look-ahead implementation costs $O(Nq_m^2)$ operations per split and makes no extra LP calls.

Corollary 4.6 (Termination under safeguarded branching)**.** *In exact arithmetic, replace weighted-width branching in Algorithm 1 by any rule satisfying (4.11) with fixed $0 < \eta \le 1$. Under the assumptions of Theorem 4.5, midpoint testing and no node cap, the invariant and finite termination for every positive target gap remain valid.*

Proof. The enclosure invariant depends on valid bounds and the box cover, not the branching score. Along any infinite nested path, each selected d_j is halved. For any $b > 0$, a coordinate can be selected only finitely often while its $d_j \ge \eta b$. If the maximum weighted width remained at least b, the safeguard would require infinitely many such selections among finitely many coordinates, a contradiction. Thus all influential widths tend to zero, and so does $\Delta(B')$ in (4.10). The midpoint argument in Theorem 4.5 proves finite termination. This exact-arithmetic result does not remove a finite-precision or work-budget stopping limit. □

The coefficient box must be declared in the coordinates used by the search. Physical restrictions or a training-only pilot may define its endpoints, but an interior solution does not establish unrestricted optimality; an enlargement defines a new problem and needs its own enclosure.

4.4. **Residual and quadrature uncertainty.**

Proposition 4.7 (Transport of numerical enclosures)**.** *Suppose computed residuals $\widetilde{r}_i$ obey $\sup_{\theta\in B} |r_i(\theta) - \widetilde{r}_i(\theta)| \le e_i$. Put $\omega = \sum_i v_i e_i/a$. If $[\widetilde{L}_\delta, \widetilde{U}_\delta]$ encloses the minimum of the computed smoothed objective and its feasible upper value, then*

$$\max\{0, \widetilde{L}_\delta - \omega - \delta V\} \le \min_B F \le F(\widehat{\theta}) \le \widetilde{U}_\delta + \omega. \tag{4.13}$$

For affine data errors, one may take $e_i = |y_i - \widetilde{y}_i| + \sum_j \max\{|\ell_j|, |u_j|\}|X_{ij} - \widetilde{X}_{ij}|$.

Proof. Lemma 2.1 bounds the nonsmoothed objective perturbation by ω uniformly. Proposition 3.3 bounds smoothing above by δV and below by zero. Combine the two inequalities, then take minima or evaluate the feasible point. The last assertion is the triangle inequality and the coefficient box bound. □

In a weak formulation, e_i can include a validated quadrature enclosure for a specified interpolated trajectory. It is not automatically a bound on unobserved measurement error. Ordinary quadrature without a validated remainder does not supply such a certificate.

5. Certified comparisons of finite-box criteria

5.1. **Finite, computable comparison.**

For a nonnegative penalty multiplier λ_N, define the exact criterion

$$C_m = 2F_m^\star + \lambda_N q_m. \tag{5.1}$$

The factor two is a convention; a bounded arctangent loss is not a normalized negative log-likelihood. If $F_m^\star \in [L_m, U_m]$, define $\underline{C}_m = 2L_m + \lambda_N q_m$ and $\overline{C}_m = 2U_m + \lambda_N q_m$.

Theorem 5.1 (Certified selection and regret)**.** *If $\overline{C}_{\widehat{m}} < \min_{m\ne\widehat{m}} \underline{C}_m$, then $\widehat{m}$ uniquely minimizes (5.1). Without separation, choosing the smallest upper score gives the deterministic regret bound*

$$0 \le C_{\widehat{m}} - \min_m C_m \le \overline{C}_{\widehat{m}} - \min_m \underline{C}_m. \tag{5.2}$$

If all score intervals shrink to their exact values and the smallest exact score has a positive gap from every competitor, the separation test eventually succeeds.

Uncorrected Proof

Proof. Every true score lies in its interval. Strict interval separation implies strict true-score ordering. For regret, use $C_{\widehat{m}} \le \overline{C}_{\widehat{m}}$ and $\min_m C_m \ge \min_m \underline{C}_m$. With a finite family and true margin $\gamma > 0$, interval errors smaller than $\gamma/3$ on both sides establish separation. □

After computing score intervals, refine overlapping competitive candidates and stop only when separation holds or the work budget is exhausted, returning the regret bound when the ordering remains unresolved. Theorem 5.1 concerns finite-data criteria, while the additional statistical conditions in Section 6 are needed to discuss generating-model recovery.

5.2. **Criterion units and a usable tolerance.** For a sum of losses, $\lambda_N = \tau \log N$ charges additional coefficients without interpreting arctangent loss as a likelihood. Rescaling the objective by $d > 0$ preserves score order and stopping decisions only if the penalty and tolerance are rescaled by d as well. The numerical target 0.02 in Section 9 is an optimization allowance; converting it to a coefficient or trajectory error requires additional growth and identification assumptions, as in Theorem 7.3.

6. Conditional statistical reference results

The following results give sufficient conditions for statistical selection consistency, whereas the deterministic enclosures above require no stochastic assumptions. The contracting recurrence supplies a qualitative example of those conditions, without calibrating the NDVI sample size, its penalty or its forecasting uncertainty.

6.1. **Consistency under dependence: explicit additional assumptions.** Consider stationary observations and a fixed finite candidate family, and $R_m(\theta) = \mathbb{E}[v_t \rho_{a_0}(r_{t,m}(\theta))]$. Write $R_m^* = \min_{\theta \in B_m} R_m(\theta)$ and let m_0 be the unique smallest risk-correct model. Every other candidate either properly contains m_0 with $q_m > q_{m_0}$, or has $R_m^* - R_{m_0}^* > 0$. Such separation is an assumption about the data-generating mechanism, not a consequence of bounded loss.

Lemma 6.1 (Uniform law with estimated common scale)**.** *Assume stationary ergodicity, compact boxes,* $\mathbb{E}[v_t \|X_{t,m}\|] < \infty$, $a_N \to a_0 > 0$, *and* $N^{-1}\sum_t |v_{t,N} - v_t| \to 0$ *in probability. Then*

$$\max_{m \in \mathcal{M}} \sup_{\theta \in B_m} \left| N^{-1} F_{N,m}(\theta) - R_m(\theta) \right| \longrightarrow 0 \quad \text{in probability.} \tag{6.1}$$

Proof. A finite parameter net and the ergodic theorem give convergence at all net points. The loss oscillation within radius η is bounded by $(\eta/a_0) N^{-1} \sum_t v_t \|X_{t,m}\|$, whose limit is finite. Let the mesh tend to zero. Estimated weights contribute at most $(\pi/2) N^{-1} \sum_t |v_{t,N} - v_t|$. The scale replacement is bounded uniformly, without residual moments, by $\frac{1}{2} |\log(a_N/a_0)|$ using (2.8). Finally take the maximum over the finite family. □

Mixing is one route to ergodicity and score limit theorems [6, 18]. For an overfitted model, a uniform law alone is inadequate. We require the following local empirical lower-growth condition in addition to unique population identification. For some neighborhood of its embedded true parameter θ_m^0, with probability tending to one,

$$F_{N,m}(\theta) - F_{N,m}(\theta_m^0) \ge Nb \left\|\theta - \theta_m^0\right\|^2 - \sqrt{N} A_{N,m} \left\|\theta - \theta_m^0\right\|, \quad b > 0, \quad A_{N,m} = O_{\mathbb{P}}(1). \tag{6.2}$$

The exact global minimizers must enter that neighborhood, as follows from (6.1) when the population minimizer is unique. Local expansions valid only for bounded $\sqrt{N}(\theta - \theta^0)$ do not, without localization, establish (6.2).

Theorem 6.2 (Consistency with a recorded optimization budget)**.** *Assume Lemma 6.1, positive underfit risk gaps, unique embedded population minima, and (6.2) for models containing* m_0. *If*

$$\lambda_N \to \infty, \qquad \lambda_N / N \to 0, \qquad \max_m \{F_{N,m}(\widehat{\theta}_m) - F_{N,m}^\star\} = o_{\mathbb{P}}(\lambda_N), \tag{6.3}$$

then minimizing $2F_{N,m}(\widehat{\theta}_m) + \lambda_N q_m$ *selects* m_0 *with probability tending to one.*

Proof. For an underfit, the uniform law gives $F_{N,m}^\star - F_{N,m_0}^\star = N\Delta_m + o_{\mathbb{P}}(N)$ with $\Delta_m > 0$; penalties and optimization errors are smaller than order N. For an overfit, write $t = \sqrt{N}\|\theta - \theta_m^0\|$. The growth lower bound is

$$bt^2 - A_{N,m} t = b\left(t - \frac{A_{N,m}}{2b}\right)^2 - \frac{A_{N,m}^2}{4b} \ge -\frac{A_{N,m}^2}{4b}. \tag{6.4}$$

Uncorrected Proof

Thus the improvement below the common embedded true value is at most $A_{N,m}^2/(4b) = O_{\mathbb{P}}(1)$. Both exact minima are localized by population identification. Their difference is therefore $O_{\mathbb{P}}(1)$, whereas the added dimension incurs $\lambda_N(q_m - q_{m_0}) \to \infty$. The recorded optimization errors are negligible relative to that penalty. A finite union completes the proof. □

A recorded enclosure supplies the optimization requirement whenever its width is $o_{\mathbb{P}}(\lambda_N)$, since

$$0 \le F_{N,m}(\widehat{\theta}_m) - F_{N,m}^{\star} \le \Delta_{N,m} := U_{N,m} - L_{N,m}. \tag{6.5}$$

A finite-sample enclosure or a local stationarity residual alone does not establish this asymptotic relation.

Proposition 6.3 (A sufficient condition for empirical lower growth)**.** *Suppose (X_i, e_i) are iid, the fixed-dimensional full features are bounded, and the conditional error densities g_x are symmetric, continuous, and uniformly bounded. Let $v_\eta(x)$ be the rule (2.6), with nuisance parameter $\eta = (a, D, \kappa)$ in a compact set having a, D_{jj}, κ bounded away from zero. Assume that at an interior θ^0 and limiting nuisance η_0,*

$$J_{\eta_0} = \mathbb{E}\{v_{\eta_0}(X) X X^{\mathsf{T}} \ell''_{a_0,X}(0)\} \succ 0, \tag{6.6}$$

where the conditional curvature is given by (2.12). Then for any consistent nuisance estimates $\widehat{\eta}_N \to \eta_0$, condition (6.2) holds on a sufficiently small fixed neighborhood. For dependent observations, the same conclusion holds if the corresponding score class obeys the uniform root-N empirical-process bound used in the proof.

Proof. Conditional symmetry makes the population score zero at θ^0 for every admissible scale and covariate-only weight. Continuity and (6.6) give a common neighborhood where the population Hessian is at least $2bI$, hence $R_\eta(\theta^0 + h) - R_\eta(\theta^0) \ge b\|h\|^2$. Define the bounded score class

$$\xi_{\theta,\eta}(X, e) = -v_\eta(X) X \, \psi_a\{e - X^{\mathsf{T}}(\theta - \theta^0)\}. \tag{6.7}$$

Set $\psi_a(0) = 0$ for this empirical-process calculation. Here P_N and P denote empirical averaging and population expectation. Each component is a finite-dimensional, bounded VC-subgraph class: after splitting the sign and cutoff cases, its subgraphs are described by finitely many polynomial inequalities of bounded degree. The usual iid maximal inequality therefore gives $A_N = \sup_{\theta,\eta} \left\| \sqrt{N}(P_N - P)\xi_{\theta,\eta} \right\| = O_{\mathbb{P}}(1)$ [24]. Along a line segment, the empirical loss is absolutely continuous; its finitely many residual crossings do not affect integration of the almost-everywhere derivative. Integrating the centered score bounds the centered loss increment by $A_N \|h\| / \sqrt{N}$. Multiplying by N proves (6.2), uniformly in the nuisance neighborhood, which contains $\widehat{\eta}_N$ with probability tending to one. The same integration proves the dependent-data assertion whenever its stated maximal bound holds. □

The distinction is between stochastic risk growth and deterministic empirical curvature. Between residual hyperplanes the arctangent objective is concave; positive population curvature comes from averaging cusp crossings. The fluctuation term in (6.2) is essential. The following result verifies it qualitatively for the clean uniform-innovation G2 generator.

Proposition 6.4 (Qualitative growth for the clean contracting recurrence)**.** *Consider the contracting recurrence*

$$Y_t = f(Y_{t-1}) + u_t + e_t, \quad f(x) = 0.35x + 0.25x^2, \quad u_t \sim U[-0.3, 0.3], \quad e_t \sim U[-\sigma_e, \sigma_e], \quad \sigma_e = 0.04. \tag{6.8}$$

The two innovation sequences are mutually independent and iid. Start in $[-1, 1]$, use responses $Y_t - u_t$, and let $X_{t,m}$ contain the full monomials of degrees one and two in $p \le 3$ past states. The coefficient box is $[-3, 3]^{q_m}$ and contains the embedded true coefficient vector in its interior. Use either unit weights or (2.6) from the common nine-feature dictionary, with fixed positive a, D, κ, or consistent estimates of these quantities. Then the clean process has a unique geometrically absolutely regular stationary law, and the empirical growth condition (6.2) holds for every candidate containing the true model $(p, d) = (1, 2)$.

More explicitly, put $\phi_p(z) = \Phi_{p,2}(z)$, $M_p = \int_{[-0.1,0.1]^p} \phi_p(z)\phi_p(z)^{\mathsf{T}} \, \mathrm{d}z$, $c_p = (5/3)^p 15^{-6}$, and $v_* = \min\{1, \kappa \min_j D_{jj}/3\}$ (or $v_* = 1$ for unit weights). At the limiting nuisance values,

$$J = \frac{a}{\sigma_e(a^2 + \sigma_e^2)} \mathbb{E}[v X X^{\mathsf{T}}] \succeq j_* I, \qquad j_* = \frac{a \, v_* c_p \lambda_{\min}(M_p)}{\sigma_e(a^2 + \sigma_e^2)} > 0. \tag{6.9}$$

Uncorrected Proof

One may take $b = j_*/4$ on a sufficiently small fixed neighborhood. This is a positive population lower bound, not a computed curvature bound for a realized empirical objective.

Proof. Invariant set and mixing. On $[-1,1]$, $|f(x)| \le 0.6|x|$, $|f'(x)| \le 0.85$ and $|u_t + e_t| \le 0.34$. The interval is invariant. Synchronous copies contract by 0.85 per step, giving a unique stationary solution. The innovation density k is the convolution of two uniform densities. It is $5/3$ on $[-0.26, 0.26]$, has support $[-0.34, 0.34]$, and has finite total variation. Therefore

$$\|P(x,\cdot) - P(x',\cdot)\|_{\mathrm{TV}} \le C|f(x) - f(x')|, \qquad \sup_{x,x' \in [-1,1]} \|P^n(x,\cdot) - P^n(x',\cdot)\|_{\mathrm{TV}} \le 2C(0.85)^n. \tag{6.10}$$

The second inequality couples the first $n-1$ innovations synchronously and uses the density-translation bound for the last transition. This uniform bound implies geometric absolute regularity of the stationary chain. Appending finitely many lagged states and the current independent innovation pair preserves geometric absolute regularity, with a finite lag shift. A sufficiently long block has a uniform total-variation contraction strictly below one. Blockwise maximal coupling then couples any deterministic initial state to stationarity in a time with a geometric tail. After coupling, use common innovations; the finite-lag score records coincide after at most three additional steps. Consequently bounded empirical averages and root-N score bounds have the same limits for the initialized experiment. A finite burn-in is not asserted to create exact stationarity.

A positive Gram matrix. Six consecutive innovations with $|u_t + e_t| \le 0.02$ place every initial state in $[-0.1, 0.1]$, since $0.05 + 0.95(0.6)^6 < 0.1$. Each such event has probability $1/15$. The stationary probability of that interval is at least 15^{-6}. Transitions between two points of the interval have density $5/3$, because $|z - f(x)| \le 0.1375 < 0.26$. The stationary p-lag density is therefore at least c_p on $[-0.1, 0.1]^p$. A nonzero polynomial cannot vanish throughout an open cube, so $M_p \succ 0$. Since the full feature norm is at most three, $v \ge v_*$ and $\mathbb{E}[vXX^\mathsf{T}] \succeq v_* c_p M_p$. Every entry of M_p is an elementary monomial integral; odd powers integrate to zero and $\int_{-d}^{d} z^{2k}\,\mathrm{d}z = 2d^{2k+1}/(2k+1)$. For $d = 0.1$, the resulting smallest eigenvalues are 4×10^{-6}, $16/(45 \times 10^6)$ and $32/(45 \times 10^7)$ for $p = 1, 2, 3$. They follow by splitting the Gram matrix into linear, square and cross-product blocks; the square block has diagonal $2^p d^{p+4}/5$ and off-diagonal $2^p d^{p+4}/9$.

Population curvature. For $\ell_a(t) = \mathbb{E}[\rho_a(e-t)]$ with $e \sim U[-\sigma_e, \sigma_e]$, symmetry and integration over the finite interval give

$$\ell_a'(t) = \frac{\rho_a(t+\sigma_e) - \rho_a(t-\sigma_e)}{2\sigma_e}, \qquad \ell_a''(0) = \frac{a}{\sigma_e(a^2 + \sigma_e^2)} > 0. \tag{6.11}$$

The derivative is positive for $t > 0$, because $|t + \sigma_e| > |t - \sigma_e|$. Thus the true conditional location uniquely minimizes risk. The second derivative is continuous near zero; the density's endpoint jumps do not affect this local calculation. Combining (6.11) with the Gram bound proves (6.9). Continuity in the positive nuisance parameters gives a common neighborhood with Hessian at least $(j_*/2)I$. Integration twice gives a population increment at least $(j_*/4)\|h\|^2$. Polynomial independence and compact boxes also give positive risk gaps for candidates omitting the quadratic term.

The stochastic increment. Consider each component of the bounded score class (6.7), with nuisance parameters in a compact positive neighborhood. A parameter-grid cell of diameter ϵ admits a pointwise bracket of width bounded by

$$C\epsilon + C\,\mathbf{1}\{|e - X^\mathsf{T}(\theta - \theta^0)| \le C\epsilon\}. \tag{6.12}$$

Away from the indicated sign change, the score and weight are uniformly Lipschitz. Conditional on the lags, e has density at most $1/(2\sigma_e)$, so the event has probability $O(\epsilon)$. There are polynomially many grid cells. For geometric absolute regularity, the squared $L_{2,\beta}$ norm of this bracket is $O(\epsilon \log(1/\epsilon))$: the inverse mixing-rate function is bounded by $C\{1 + \log(1/u)\}$, and integration over an event of probability $O(\epsilon)$ gives this bound. Consequently the square-root bracketing-entropy integral is finite. The absolutely regular empirical-process theorem of Doukhan, Massart and Rio [8] gives

$$\sup_{\theta,\eta} \|\sqrt{N}(P_N - P)\xi_{\theta,\eta}\| = O_{\mathbb{P}}(1). \tag{6.13}$$

Uncorrected Proof

Integrating this centered score along θ^0+th, as in the preceding proof, gives the linear stochastic term in (6.2). Coupling changes this uniform score bound by $O_{\mathbb{P}}(N^{-1/2})$ for the zero initialization. This proves the asserted empirical growth.

Estimated nuisance quantities. For clean G2, the common full-dictionary LAD pilot has a unique population minimum: the conditional absolute-loss risk is strictly minimized at zero displacement, and the lag density makes distinct coefficient polynomials distinguishable. Compactness, bounded features and ergodicity give a uniform law and consistency of this convex pilot. Its residual median and MAD converge to 0 and $\sigma_e/2$. The stationary support is an interval: the map sending a compact set to $f(K)+[-0.34, 0.34]$ contracts in Hausdorff distance and preserves intervals. The positive innovation density gives a positive stationary density in the support interior. Successively adjoining admissible innovation intervals gives a connected finite-lag support with a density on its interior. A nonconstant polynomial has null level sets; its absolute value has an interval image. These facts give continuous absolute-monomial distributions with unique positive medians. Ergodic quantiles therefore give convergence of the feature scales, including their fixed positive floors. The empirical-process bound is uniform in a neighborhood of these limits, so substituting the estimated scales needs consistency, not a root-N nuisance rate. This proves the stated pilot-scale extension. No analogous claim is made for the contaminated variants or for the optimization errors of the local algorithms. □

For three lags, $\sigma_e = 0.04$ and $a_0 = 1.4826\sigma_e/2 = 0.029652$, (6.9) gives $j_*/v_* \approx 8.642 \times 10^{-12}$. This establishes positive growth for the clean recurrence, without supplying a practical finite-sample tolerance for the regional series.

A long-run score covariance would enter a distributional approximation under an additional mixing central limit theorem and a localized score expansion [18, 23]. We do not use such an approximation for the May forecasts.

7. Conditional trajectory-error consequences

7.1. **Exact sensitivity and discrete dynamics.** For a differentiable field, the parameter sensitivity $S(t) = \partial x(t;\theta)/\partial\theta^{\mathsf{T}}$ solves

$$\dot{S} = f_x(t, x; \theta)S + A_m(t, x), \qquad S(0) = \partial x_0/\partial\theta^{\mathsf{T}}. \tag{7.1}$$

For a recurrence state map $z_{k+1} = G_\theta(z_k)$, its counterpart is

$$S_{k+1} = J_k S_k + G_{\theta,k}, \qquad S_h = P_{h:0}S_0 + \sum_{j=0}^{h-1} P_{h:j+1}G_{\theta,j}, \tag{7.2}$$

where $J_k = \partial_z G_\theta(z_k)$, $G_{\theta,k} = \partial_\theta G_\theta(z_k)$, $P_{h:j} = J_{h-1}\cdots J_j$ and $P_{h:h} = I$. These equations follow by the chain rule and induction. They compute the derivative of a forecast, not a global uncertainty region.

Proposition 7.1 (Linearization remainder). *For a fixed finite horizon, suppose the flow or recurrence is twice continuously differentiable jointly in state and parameter on a compact neighborhood of the relevant trajectories. Then*

$$x(T;\theta + d) - x(T;\theta) = S(T)d + O(\|d\|^2), \tag{7.3}$$

uniformly for sufficiently small d. The same statement holds for a fixed number of recurrence steps.

Proof. Differentiating the ODE yields (7.1); differentiating a finite composition gives (7.2). The second derivative is bounded on the compact neighborhood. Taylor's theorem with integral remainder gives (7.3). □

For lag states, the Euclidean one-step condition $\|J_k\|_2 < 1$ is often impossible because the shift block contains unit entries. It must not be asserted as a generic property of autoregressions. Products of Jacobians or an adapted norm are the appropriate objects.

Proposition 7.2 (A useful adapted norm). *If a fixed Jacobian J_* has spectral radius below one, there is $P \succ 0$ solving $J_*^{\mathsf{T}} P J_* - P = -I$, and*

$$\|J_* z\|_P^2 \le \left(1 - \lambda_{\max}(P)^{-1}\right)\|z\|_P^2, \qquad \|z\|_P = (z^{\mathsf{T}} P z)^{1/2}. \tag{7.4}$$

A strict contraction persists for nearby Jacobians. In Euclidean coordinates a Jordan block may produce transient polynomial growth even when all eigenvalues lie inside the unit disk.

Uncorrected Proof

Proof. The convergent series $P = \sum_{k=0}^{\infty} (J_*^k)^{\mathsf{T}} J_*^k$ solves the equation and satisfies $P \succeq I$. Hence $\|J_* z\|_P^2 = \|z\|_P^2 - \|z\|_2^2$, giving (7.4). Strict inequalities persist by continuity. A Jordan block $(\lambda I + N)^h = \sum_{j=0}^{q-1} \binom{h}{j} \lambda^{h-j} N^j$ explains the polynomial factor. Numerically, Schur factors or a Lyapunov solve are preferable to computed Jordan chains [9]. □

7.2. **An integrated deterministic error budget.**

Theorem 7.3 (Identification, optimization, and propagation)**.** *Fix a selected model and let θ^* be a global minimizer of its empirical loss. Suppose the returned point belongs to a region on which*

$$F(\theta) - F(\theta^*) \geq \frac{\mu}{2} \|\theta - \theta^*\|^2, \qquad \mu > 0. \tag{7.5}$$

Let $F(\widehat{\theta}) - F(\theta^*) \leq \varepsilon_{\mathrm{opt}}$. For a reference parameter θ^0 assume an available bound $\|\theta^* - \theta^0\| \leq r_{\mathrm{id}}$. Suppose on a common convex trajectory domain the reference field has one-sided Lipschitz constant ν and $\|f(t,x;\theta) - f(t,x;\vartheta)\| \leq M\|\theta - \vartheta\|$. Let $E_0 = \|\widetilde{x}(0) - x(0;\theta^0)\|$. If a continuous numerical reconstruction $\widetilde{x}$ has defect $\left\|\dot{\widetilde{x}} - f(t,\widetilde{x};\widehat{\theta})\right\| \leq d(t)$, then

$$\left\|\widetilde{x}(T) - x(T;\theta^0)\right\| \leq e^{\nu T} E_0 + \int_0^T e^{\nu(T-t)} \left[M\left(r_{\mathrm{id}} + \sqrt{2\varepsilon_{\mathrm{opt}}/\mu} \right) + d(t) \right] \mathrm{d}t. \tag{7.6}$$

For $d(t) \leq \bar{d}$, the integral is bounded by $K_\nu(T)[M(r_{\mathrm{id}} + \sqrt{2\varepsilon_{\mathrm{opt}}/\mu}) + \bar{d}]$, where $K_\nu(T) = (e^{\nu T} - 1)/\nu$ and $K_0(T) = T$.

Proof. Quadratic growth gives $\left\|\widehat{\theta} - \theta^*\right\| \leq \sqrt{2\varepsilon_{\mathrm{opt}}/\mu}$. The triangle inequality bounds the distance to θ^0. Subtract the reference ODE from the reconstruction, add and subtract $f(t,\widetilde{x};\theta^0)$, and use one-sided Lipschitz continuity for the common-parameter term. Writing $E(t) = \|\widetilde{x}(t) - x(t;\theta^0)\|$, the upper Dini derivative satisfies $D^+E \leq \nu E + M(r_{\mathrm{id}} + \sqrt{2\varepsilon_{\mathrm{opt}}/\mu}) + d(t)$. Multiplication by $e^{-\nu t}$ and integration proves (7.6). □

Applying this bound to measured observations requires an independently justified growth constant μ and identification radius r_{id}, neither of which follows from a small objective gap. Alternatively, surviving boxes can bound the distance from an incumbent to a global minimizer without curvature, although identification relative to the environmental reference still needs observation or parameter information. Across a cusp, a smooth interval-Hessian calculation cannot be used without accounting for smoothing.

Corollary 7.4 (Balancing computational budgets)**.** *If a trajectory tolerance ε_T is allocated only to optimization error and $MK_\nu(T) > 0$, it suffices to require*

$$\varepsilon_{\mathrm{opt}} \leq \frac{\mu}{2} \left(\frac{\varepsilon_T}{MK_\nu(T)} \right)^2. \tag{7.7}$$

With smoothed and perturbed residuals, the certified optimization allowance must include δV and both sides of the objective perturbation from (4.13).

Proof. Isolate the optimization term in (7.6) and solve the resulting scalar inequality. The extra allowances follow by comparing perturbed and exact minima and feasible values. □

Proposition 7.5 (A directly checkable Euler envelope)**.** *Suppose the reference flow contracts by $e^{\nu h}$ in the chosen norm, and its exact one-step Euler defect is at most $Dh^2/2$. If $\sup_{x \in \mathcal{D}, t} \left\|f(t,x;\widehat{\theta}) - f(t,x;\theta^0)\right\| \leq b$, and numerical nodes and the intervening reference flows stay in $\mathcal{D}$, then*

$$E_{k+1} \leq e^{\nu h} E_k + hb + \tfrac{1}{2} Dh^2. \tag{7.8}$$

For $E_0 = 0$ and $\nu \neq 0$, this gives $E_k \leq (hb + Dh^2/2)(1 - e^{\nu k h})/(1 - e^{\nu h})$.

Proof. Compare the fitted Euler step at its current node with a reference Euler step at the same node, then the exact reference flow from that node, then the exact reference flow from the true current state. The three differences are bounded by hb, $Dh^2/2$, and $e^{\nu h} E_k$. Iterate the scalar recursion. □

Uncorrected Proof

8. Bounded compositional trajectories

For five percentage components, let T_H be a requested final year fixed before its observations are seen and put $\chi(t) = (t - 2000)/(T_H - 2000)$. We use a bounded Bernstein representation

$$B_{k,d}(\chi) = \binom{d}{k}\chi^k(1-\chi)^{d-k}, \qquad z_{j,d}(t) = \sum_{k=0}^{d} \beta_{jk} B_{k,d}(\chi(t)), \quad 0 \le \beta_{jk} \le 100. \tag{8.1}$$

The nonnegative basis functions sum to one, which connects coefficient bounds to trajectory bounds throughout the requested time interval even when prediction extends beyond the training years. The representation is an empirical trajectory model with analytic derivatives, rather than a mechanistic account of vegetation dynamics.

Proposition 8.1 (Trajectory and closure bounds)**.** *For coefficients satisfying* (8.1)*,* $0 \le z_{j,d}(t) \le 100$ *on* $[2000, T_H]$*. For* $1 \le \ell \le d$*,*

$$\frac{\mathrm{d}^\ell z_{j,d}(t)}{\mathrm{d}t^\ell} = \frac{d!}{(d-\ell)!(T_H - 2000)^\ell} \sum_{k=0}^{d-\ell} \Delta^\ell \beta_{jk} B_{k,d-\ell}(\chi(t)). \tag{8.2}$$

If $u, v \in [0,\infty)^5$ *have sums at least* $s_* > 0$*, the closure map* $\mathcal{C}(u) = 100u/\sum_j u_j$ *obeys*

$$\|\mathcal{C}(u) - \mathcal{C}(v)\|_1 \le (200/s_*)\|u - v\|_1. \tag{8.3}$$

Proof. Bernstein basis functions are nonnegative and sum to one, proving the coefficient-hull bound. Differentiate once using $B'_{k,d} = d(B_{k-1,d-1} - B_{k,d-1})$, with out-of-range indices interpreted as zero, and collect adjacent coefficients. Iteration and the chain rule give (8.2). For $s = \sum u_j$ and $t = \sum v_j$, add and subtract $100v/s$ and use $|s-t| \le \|u-v\|_1$. The resulting upper bound is $200\|u - v\|_1/s$, hence (8.3). □

Writing $S(t) = \sum_j z_{j,d}(t) > 0$ and $p_j(t) = 100 z_{j,d}(t)/S(t)$ gives the differential representation

$$\dot p_j(t) = \frac{100}{S(t)^2}\left\{\dot z_{j,d}(t)S(t) - z_{j,d}(t)\sum_k \dot z_{k,d}(t)\right\}, \qquad \sum_j \dot p_j(t) = 0. \tag{8.4}$$

The quotient rule and Proposition 8.1 establish this identity, so the fitted composition preserves its total while its derivatives describe redistribution among the five reported intervals. This differential description does not introduce an independently estimated ecological mechanism, and it has no numerical ODE-integration error because the derivatives are available analytically. Joint constraints $\sum_j \beta_{jk} = 100$ would instead make closure automatic for every time in the interval.

9. Computational assessment

9.1. **Global fitting, dimensions and lower-bound cost.** The synthetic audit uses 30 seeds, 2026190000–2026190029, with 28 states drawn from $[-1.4, 1.4]^2$ and responses $-0.65u + 0.35\tanh v + \eta$, where $\eta \sim N(0, 0.035^2)$ and three responses receive signed shifts of magnitude 0.9. The ordered dictionary is $(-u, \tanh v, -u^3, uv, v^2)$, with successive coefficient bounds $[0.1, 1.5]$, $[-0.8, 0.8]$ and $[-0.4, 0.4]$ for each remaining coordinate. These declared benchmark boxes contain the generating coefficients and are not confidence intervals. All runs use $a = 0.05$, $q_\rho = 1$, frozen full-dictionary weights, target gap 0.02 and an 1,800-expansion limit, with all capped outcomes retained.

Computation used Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0 and python-flint 0.9.0 on an AMD EPYC 9V74 host with one BLAS/OpenMP thread per worker; timing comparisons use sequential searches and exclude data generation and worker startup. The protocol was fixed after exploratory work, so it is a reproducible computational comparison rather than an independently preregistered study. Table 1 summarizes each dimension over its 30 independent generator replications.

The observed limit arises during subdivision rather than solely at the root: median initial gaps remain near ten from two through five coefficients, while the final active frontier increases from 27 boxes at three coefficients to 1,062 at five and the corresponding chord allowance grows from 0.0947 to 22.2. In two instrumented searches, the relaxation

Uncorrected Proof

TABLE 1. Fresh-seed dimensional scaling, 30 searches per dimension. Time, expanded boxes and attained gap are medians over all runs, including failures at the 1800-expansion cap. Success means an outward global objective gap at most 0.02. The two-sided 95% Clopper–Pearson intervals describe generator-specific success probabilities; capped times are not times to solution.

q_m	Time (s)	Expanded	Gap	Success / 30	Two-sided 95% (%)
1	0.089	10	0.0146	30/30	88.4–100.0
2	0.255	40	0.0160	30/30	88.4–100.0
3	1.752	298	0.0192	30/30	88.4–100.0
4	8.405	1409	0.0199	22/30	54.1–87.7
5	10.908	1800	1.1083	0/30	0.0–11.6

gap Γ never exceeds 2.42×10^{-13} although the final five-coefficient chord allowance is 14.406, indicating that residual-interval geometry can dominate LP-witness accuracy. The safeguard rules in Section 4.3 close no additional four- or five-coefficient case in the paired ten-seed subset.

The bound ablation uses the same 30 datasets in two coefficients, freezes weights from the three-term dictionary, holds a common feasible arrangement incumbent fixed and permits 3,000 expansions at target 0.02. Table 2 includes capped work in its summaries, with paired reduction intervals formed from the 10th and 21st ordered reductions, giving at least 95 percent coverage for the generator median. The shifted-chord LP requires fewer boxes, but the cheaper sign witness has lower measured runtime because solving the LP adds work to every evaluated box.

TABLE 2. Fresh-seed paired lower-bound ablation, 30 datasets. Entries are medians except success counts. Reduction is paired against interval-rule expanded boxes. Brackets give distribution-free median intervals with coverage at least 95 percent under independent generator replications. Work at an unresolved cap remains in the summary.

Bound	Time (s)	Nodes	LP time (s)	Success / 30	Reduction [interval] (%)
Interval	0.625	1547.0	0.000	29/30	0.0 [0.0–0.0]
Origin chord + LP	0.912	171.5	0.805	30/30	87.8 [86.6–89.4]
Shifted chord + sign	0.039	72.5	0.000	30/30	95.1 [94.5–96.3]
Shifted chord + LP	0.221	38.5	0.194	30/30	97.4 [96.9–98.0]

Across 40 paired searches, changing Arb precision from 100 to 200 bits alters exported bounds or gaps by at most 3.36×10^{-9} and leaves stopping counts unchanged. A 1.5-fold enlargement of synthetic box half-widths also preserves the success counts in the paired subset, although unresolved gaps increase; no returned endpoint in the 450 original or 150 enlarged searches contacts a facet at normalized distance 10^{-8}. These checks concern the tested boxes, since interior incumbents cannot establish unrestricted optimality. At equal mean-objective allowance 0.02/28, four-coefficient success decreases from 10/10 at $N = 14$ to 2/10 at $N = 56$, so the dimension and sample-size limitations are not removed by tolerance normalization.

9.2. **Measured-data calibration and model comparison.** The public Cascaded Tanks record contains 1,024 input and output measurements at four-second intervals, measured in volts [21, 22]. We use all 1,023 estimation transitions and compare the empirical one-step predictors

$$\widehat{y}_k = \theta_1 y_{k-1} + \theta_2 u_{k-1}, \qquad \widehat{y}_k = \theta_1 y_{k-1} + \theta_2 u_{k-1} + \theta_3, \tag{9.1}$$

with unit weights, $a = 0.1$ V and boxes $B_2 = [0, 1.1] \times [0, 0.5]$, $B_3 = B_2 \times [-1, 1]$. Observed lagged outputs are fixed regressors, so these are calibration predictors rather than hydraulic parameter models. The boxes and criterion are fixed without using the separate benchmark test responses, and Table 3 reports the resulting objective enclosures and test errors.

Uncorrected Proof

TABLE 3. Outward global enclosures for the measured tank estimation criterion, $N = 1,023$, at absolute target 0.02. Decimal endpoints are rounded outward. Test RMSE uses all 1,023 separate test transitions and observed lags; it is not a free-running forecast or a certified prediction-error bound.

q_m	Lower	Upper	Gap upper	Expanded	Test RMSE (V)
2	384.198390	384.217151	0.018761	92	0.09477
3	356.591002	356.610673	0.019671	1431	0.08305

Both returned endpoints lie inside their boxes, with the displayed coefficient vectors

$$\widehat{\theta}_2 = (0.992765, 0.013136), \qquad \widehat{\theta}_3 = (0.995138, 0.037014, -0.086677).$$

Applying Theorem 5.1 to $C_m = 2F_m^\star + \tau q_m \log(1023)$ certifies the three-coefficient criterion at every $\tau \in \{0.25, 0.5, 0.75, 1, 1.5\}$; at $\tau = 1$, the outward score intervals are $[782.2577, 782.2953]$ and $[733.9734, 734.0129]$, while the smallest grid separation exceeds 44.7796. This establishes ordering within the stated two-model family, independently of the separate test-RMSE comparison.

The ordering remains certified when a changes to 0.05 or 0.2 V and when the box limits are enlarged to $B_m^+ = 1.5B_m$. Both enlarged searches close at gap 0.02, but the three-coefficient fit at $a = 0.05$ remains unresolved at gap 6.523707 after 1,800 expansions, even though its worst-grid score satisfies $C_3 \le 1179.9479 < 1232.0946 \le C_2$. The distinction matters in practice because a model comparison may be resolved without bringing every candidate minimum to the same tolerance.

10. MAY NDVI COMPOSITIONS FOR KHANTY-MANSIYSK

10.1. **Regional record and monitoring question.** The regional record contains 26 annual May compositions for Khanty-Mansiysk, 2000–2025, supplied as MODIS-derived NDVI counts. Holding the month fixed gives a common seasonal reference for comparing five vegetation-index shares across years.

The reported intervals are $[-0.2, 0]$, $[0, 0.2]$, $[0.2, 0.5]$, $[0.5, 0.8]$ and $[0.8, 1]$. NDVI measures red/near-infrared reflectance contrast [25], but these numerical intervals do not by themselves identify land cover or hydrological condition. From positive class counts N_{jt}, the analyzed percentages are

$$y_{jt} = 100N_{jt}/\sum_{k=1}^{5} N_{kt}, \qquad \sum_j y_{jt} = 100. \tag{10.1}$$

The reconstructed percentages agree with the supplied values within 7.2×10^{-15} percentage points, while total counts range from 20,871,575 to 20,983,263 and Class 5 averages 0.0356 percent. Table 4 retains the full series, and Figure 1 shows the supplied maps; their nine display intervals differ from the five modeled totals and therefore provide spatial context rather than a reconstruction of those totals.

Class 4 ranges from 0.8503% in 2004 to 72.9998% in 2011. The missing source rasters, quality masks and compositing metadata prevent these changes in reported percentages from being attributed to a particular environmental process.

10.2. **Chronological fitting and evaluation.** Training through 2020 is followed by one fixed holdout covering 2021–2025, and model settings are selected using the five preceding one-year origins 2015–2019. A separate rolling comparison uses ten origins, 2015–2024, with settings reselected from each origin's five preceding forecasts, so every scale, fit and validation choice uses only observations available before its target. The histories are nested and their targets overlap the fixed holdout, which makes these descriptive evaluations dependent rather than independent temporal replications.

For the Bernstein trajectories in Section 8, a degree-one box-constrained LAD pilot supplies training residuals and the common class scale

$$s_j = \max\{0.01, 1.4826022\,\mathrm{median}_t\,|e_{jt}^{\mathrm{pilot}} - \mathrm{median}_u\, e_{ju}^{\mathrm{pilot}}|\}. \tag{10.2}$$

Uncorrected Proof

TABLE 4. May NDVI-class compositions from the supplied pixel counts, in percent. C1–C5 denote the five intervals in the text. Years 2021–2025 are the final holdout.

Year	C1	C2	C3	C4	C5	Year	C1	C2	C3	C4	C5
2000	15.7673	24.6482	33.3171	26.2311	0.0363	2013	19.0769	22.8670	23.0765	34.9321	0.0476
2001	15.3943	23.2638	21.0048	40.3201	0.0171	2014	19.8310	31.0047	23.5856	25.5692	0.0095
2002	35.3782	24.9358	20.9460	18.7286	0.0114	2015	21.0531	24.1373	22.6996	32.0985	0.0115
2003	27.2784	43.5613	19.3998	9.7381	0.0224	2016	9.6650	13.4914	22.0657	54.7647	0.0132
2004	34.9255	51.7636	12.4370	0.8503	0.0236	2017	21.9735	30.0449	22.2623	25.6699	0.0495
2005	17.8299	31.3182	22.1334	28.7124	0.0061	2018	28.0303	40.0948	19.4945	12.3675	0.0129
2006	30.7549	45.5882	18.1576	5.4868	0.0125	2019	24.8960	29.7272	20.6767	24.6857	0.0144
2007	10.7314	22.2436	28.3815	38.5672	0.0763	2020	5.5034	14.5408	21.4558	58.4557	0.0442
2008	31.9932	38.3231	17.8497	11.8014	0.0325	2021	11.0745	13.2409	18.6065	56.9907	0.0875
2009	20.3692	36.3415	20.5559	22.7168	0.0166	2022	24.8189	32.4278	19.2445	23.4306	0.0781
2010	12.0720	18.7545	23.2505	45.8962	0.0268	2023	26.3217	23.0982	17.9585	32.4687	0.1530
2011	2.8842	4.5395	19.5631	72.9998	0.0135	2024	28.9112	40.3680	21.1167	9.5560	0.0481
2012	6.1817	21.3893	31.2669	41.1313	0.0308	2025	24.1684	26.8776	21.8938	27.0311	0.0291

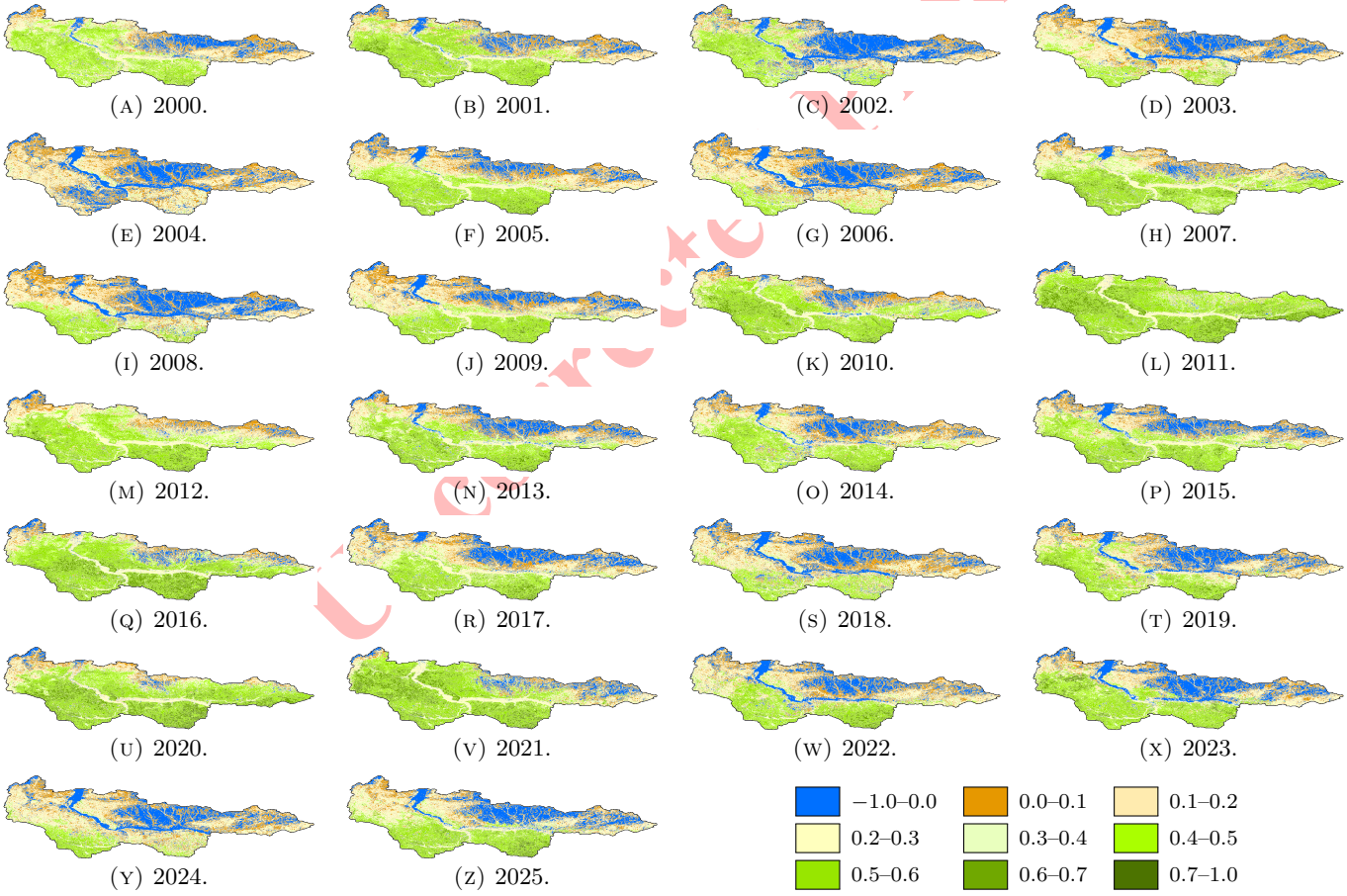


FIGURE 1. May NDVI maps of Khanty-Mansiysk, 2000–2025. The nine colors are the intervals in the supplied map key; they differ from the five classes in Table 4. Each panel shows an observed year.

In normalized coordinates $b_{jk} = \beta_{jk}/s_j$, the criterion is

$$F_{j,d}^{q_\rho,\lambda}(b_j) = \frac{1}{n_T}\sum_{t\in\mathcal{T}} \phi_{c,q_\rho}\left(\left|\frac{y_{jt}}{s_j} - \sum_{k=0}^{d} b_{jk}B_{k,d}(\chi(t))\right|\right) + \frac{\lambda d}{2}\sum_{k=0}^{d-1}(b_{j,k+1}-b_{jk})^2, \quad 0 \le b_{jk} \le 100/s_j. \tag{10.3}$$

Here $c = 1.5$, $q_\rho \in \{1,2\}$, $d \in \{1,\dots,5\}$ and $\lambda \in \{0, 0.1, 1, 10\}$, with the same pilot scale across degrees, losses and penalties at a given split and requested horizon. The coefficient-difference penalty limits trajectory variation, and validation minimizes composition MAE with exact ties resolved toward smaller degree and then larger penalty. Each fit uses four deterministic starts and smoothing widths $10^{-2}, 10^{-4}, 10^{-6}$, while the original loss ranks the final endpoints. Among 8,965 class fits and 34,260 starts, 1,682 start sets have objective spreads above 10^{-6} and the largest final smoothed projected-gradient residual is 9.48×10^{-5}, so these are locally optimized penalized fits rather than applications of the unpenalized global arrangement certificate.

OLS and LAD use the same coefficient bounds and degree candidates, while Huber and Cauchy also use the same penalty grid with normalized cutoffs 1.345 and 2.385. The historical mean, persistence and level-only exponential smoothing supply low-complexity baselines; for the latter, $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$ with $\ell_{2000} = y_{2000}$ and forecast ℓ_T, using $\alpha \in \{0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$ selected by the same historical validation [10]. Separate class forecasts are closed by $\widehat{y}_t = \mathcal{C}(z(t))$, and equal-weight percentage-point accuracy is measured by

$$\mathrm{MAE} = \frac{1}{5n_E}\sum_{t\in\mathcal{E}}\sum_{j=1}^{5}|\widehat{y}_{jt} - y_{jt}|, \quad \mathrm{RMSE} = \left[\frac{1}{5n_E}\sum_{t,j}(\widehat{y}_{jt}-y_{jt})^2\right]^{1/2}, \quad \mathrm{TV} = \frac{1}{2n_E}\sum_{t,j}|\widehat{y}_{jt}-y_{jt}|. \tag{10.4}$$

These metrics preserve the interpretation of a one-percentage-point error across components, while Table 5 exposes the rare component that an aggregate score could conceal. No confidence interval or Monte Carlo standard error is inferred from a single regional series.

10.3. **Forecasting results and regularization.** All six fitted procedures select degree one for the fixed five-year holdout. Table 5 reports the complete selection and evaluation results, where inverse-power GLDM gives holdout MAE 5.446 pp compared with 5.628 for ordinary GLDM and 5.740 for the historical mean; however, the historical mean gives the lowest rolling MAE at 6.354 pp compared with 6.423 and 6.435 for the two GLDM procedures. The contrast therefore depends on the evaluation period and supports continued comparison with the historical baseline rather than a claim of established predictive superiority.

TABLE 5. May forecasting errors in percentage points. Validation uses five one-year origins through 2019; the fixed holdout uses 2021–2025, and rolling evaluation uses ten one-year origins through 2024. The settings shown are those chosen for the fixed holdout. C1–C5 give its classwise errors; their mean is the holdout MAE.

		MAE across classes			Holdout MAE by class				
Method	Settings	Val.	Holdout	Rolling	C1	C2	C3	C4	C5
OLS	$d=1,\lambda=0$	8.418	8.332	9.527	10.625	9.487	2.049	19.449	0.05199
LAD	$d=1,\lambda=0$	8.267	5.985	7.232	5.876	8.327	2.151	13.505	0.06711
Huber	$d=1,\lambda=10$	6.843	5.656	6.451	6.615	7.625	2.206	11.778	0.05661
Cauchy	$d=1,\lambda=10$	6.878	5.656	6.435	6.628	7.601	2.181	11.811	0.05685
GLDM	$d=1,\lambda=10$	7.186	5.628	6.423	5.634	7.294	3.658	11.491	0.06445
Inverse-power GLDM	$d=1,\lambda=0.1$	6.994	5.446	6.435	4.662	7.543	3.597	11.360	0.06522
Historical mean	No tuning	–	5.740	6.354	6.869	7.559	2.311	11.905	0.05395
Persistence	No tuning	–	12.241	8.281	17.556	13.182	1.867	28.560	0.04096
Exponential smoothing	$\alpha=0.1$	7.651	6.248	6.985	7.859	7.424	3.449	12.457	0.05223

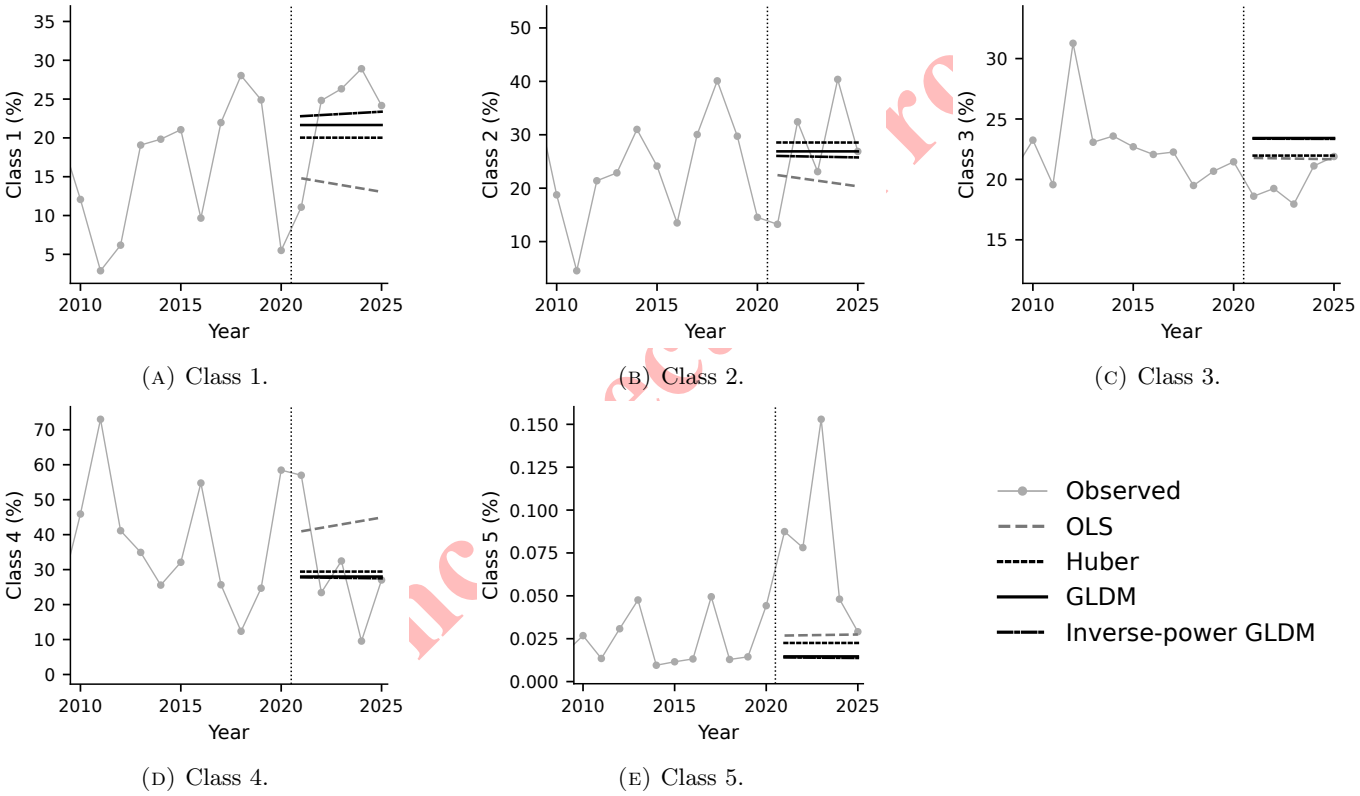
Regularization accounts for most of the degree-five reduction in holdout error, from 12.287 pp without the penalty to 5.495 pp at $\lambda = 0.1$, followed by a smaller change to 5.448 pp when the exponent changes at that same penalty. Table 6

Uncorrected Proof

retains every degree so that this improvement cannot be attributed to the loss exponent alone, and Figure 2 shows the resulting trajectories against all held-out observations. These arithmetic contrasts describe the fitted procedures, since fixing a numerical penalty does not equalize its strength relative to the different loss ranges.

TABLE 6. May holdout MAE at each degree. The common penalty separates regularization and loss-shape contrasts.

Degree	$q_\rho = 1, \lambda = 0$	$q_\rho = 1, \lambda = 0.1$	$q_\rho = 2, \lambda = 0.1$	$q_\rho = 2, \lambda = 0$
1	5.413	5.505	5.446	5.400
2	7.539	5.502	5.448	7.306
3	10.889	5.500	5.449	10.521
4	6.476	5.498	5.448	10.384
5	12.287	5.495	5.448	12.309



(A) Class 1. (B) Class 2. (C) Class 3.

(D) Class 4. (E) Class 5.

FIGURE 2. May forecasts from training through 2020. Gray points are observations; the dotted line marks the holdout.

Class 4 produces the largest absolute error for every procedure, while persistence has the smallest Class 5 error despite its poor aggregate score. The inverse-power procedure reduces the selected GLDM aggregate error but increases errors in Classes 2 and 5, which is relevant to monitoring because a regional average can obscure a component whose behavior matters locally. Each class contributes only five holdout values, so these differences identify aspects requiring further observation rather than a reliable classwise ranking.

Deleting one training year while fixing the selected degree and penalty changes the 25 holdout predictions by an average 0.224 pp for ordinary GLDM and 0.150 pp for inverse-power GLDM, with largest individual changes 0.980 and 0.899 pp. A separate coherence check uses 44 joint fits with $\sum_j \beta_{jk} = 100$, degrees one and three, both exponents and 11 evaluation splits at $\lambda = 0.1$; its maximum coefficient-sum discrepancy is 2.85×10^{-14} pp. At degree one and exponent two, joint fitting changes holdout MAE from 5.446 to 5.362 pp but rolling MAE from 6.149 to 6.796 pp in that fixed-configuration comparison, showing that exact compositional coherence and lower forecast error are separate requirements.

11. Conclusion

For an affine-residual differential model, the shifted chord and a feasible dual witness give a lower value against which to check a numerical fit. The measured tank example illustrates the practical use of this calculation: a wide interval for one model can still settle the comparison if its upper score is below its competitor's lower score. The synthetic searches show its computational limit, since the requested gap remained open in every five-coefficient case at the stated budget.

For Khanty-Mansiysk, regularizing the degree-five May trajectory reduced holdout MAE from 12.287 to 5.495 percentage points, whereas changing the loss exponent at the same penalty reduced it by only another 0.047 points. The historical mean gave the lowest rolling MAE, so it provides the appropriate benchmark for any more elaborate regional forecast. In the Yugra monitoring project, a large departure from that reference can direct attention to a particular May map, especially for Class 4, which had the largest forecast errors. Interpreting such a departure as waterlogging or forest–wetland degradation requires spatially matched water-sensitive observations or field evidence.

Data availability

The May dataset contains MODIS-derived pixel counts and percentages for five NDVI classes in Khanty-Mansiysk. The reported processing steps included NDVI calculation, image correction and combination, mapping for January, May, July and October, and classification by NDVI thresholds. This study uses only the May observations, which comprise 26 annual compositions from 2000 through 2025.

Acknowledgment

This research was supported by the Foundation for Scientific and Technological Development of Yugra under project No. 2026-252-01.

References

[1] M. Abotaleb, *Solving the optimizing parameters problem for nonlinear datasets using the high-order general least deviations method (GLDM) algorithm*, Comput. Methods Differ. Equ., *13* (2025), 940–967.

[2] 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* (2013), 91–129.

[3] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., *17* (2007), 1205–1223.

[4] 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* (2023), Article 110.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[6] R. C. Bradley, *Basic properties of strong mixing conditions: a survey and some open questions*, Probab. Surv., *2* (2005), 107–144.

[7] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Natl. Acad. Sci. USA, *113* (2016), 3932–3937.

[8] P. Doukhan, P. Massart, and E. Rio, *Invariance principles for absolutely regular empirical processes*, Ann. Inst. H. Poincaré Probab. Statist., *31* (2) (1995), 393–427. https://www.numdam.org/item/AIHPB_1995__31_2_393_0/.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, 2013.

Uncorrected Proof

[10] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed., OTexts, Melbourne, 2021. https://otexts.com/fpp3/. Accessed 4 October 2026.
[11] F. Johansson, *Arb: efficient arbitrary-precision midpoint-radius interval arithmetic*, IEEE Trans. Comput., *66* (2017), 1281–1292.
[12] C. Maier, C. Loos, and J. Hasenauer, *Robust parameter estimation for dynamical systems from outlier-corrupted data*, Bioinformatics, *33* (2017), 718–725.
[13] N. M. Mangan, J. N. Kutz, S. L. Brunton, and J. L. Proctor, *Model selection for dynamical systems via sparse regression and information criteria*, Proc. R. Soc. A, *473* (2017), Article 20170009.
[14] G. P. McCormick, *Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems*, Math. Program., *10* (1976), 147–175.
[15] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
[16] A. V. Panyukov and Y. S. 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* (2018), 35–43.
[17] A. V. Panyukov and Y. S. Mezaal, *Improving of the identification algorithm for a quasilinear recurrence equation*, in *Advances in Optimization and Applications*, Springer, Cham, 2020, 15–26.
[18] E. Rio, *Asymptotic Theory of Weakly Dependent Random Processes*, Springer, Berlin, 2017.
[19] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[20] S. M. Rump, *Verification methods: rigorous results using floating-point arithmetic*, Acta Numer., *19* (2010), 287–449.
[21] M. Schoukens, P. Mattsson, T. Wigren, and J. P. Noël, *Cascaded tanks benchmark combining soft and hard nonlinearities*, 4TU.ResearchData (2020).
[22] M. Schoukens and J. P. Noël, *Three benchmarks addressing open challenges in nonlinear system identification*, IFAC-PapersOnLine, *50*(1) (2017), 448–453.
[23] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
[24] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York, 1996.
[25] J. Weier and D. Herring, *Measuring vegetation (NDVI & EVI)*, NASA Earth Observatory (2000), updated November 14, 2025. https://science.nasa.gov/earth/earth-observatory/measuring-vegetation-ndvi-evi/ (accessed October 4, 2026).

Uncorrected Proof