mathematics

MDPI

Article

Positive Latent Suppression in Logistic Neural Differential Equations: Trajectory Reconstruction, Identifiability and Population Hindcasts

Chenhao Xu [1], Tiecheng Xia [1,2], Haijing Ma [1] and Junjie Fan [1,*]

[1] School of Science, Inner Mongolia University of Technology, Hohhot 010051, China
[2] Department of Mathematics, Shanghai University, Shanghai 200444, China
* Correspondence: fanjunjie@imut.edu.cn

Abstract

A close fit to a scalar time series does not establish either accurate extrapolation or unique mechanistic parameter recovery. We examine a logistic gray-box differential equation coupled to a positive latent suppression state whose relative rate is represented by a bounded neural network. Positivity at every finite time, a population upper bound and a local turning-point criterion are established. For freely chosen bounded closures on a finite observation interval, compatible mechanistic parameters are characterized by a convex feasible set and their nonuniqueness is established. With a positive closure bound, unrestricted positive growth parameters and a free latent initial value, every positive twice continuously differentiable trajectory on that interval is compatible with such a closure. Strictly feasible parameter pairs have two-dimensional feasible neighborhoods. These results describe identifiability boundaries in the freely chosen closure class; a complete structural-identifiability analysis of the fixed finite-width neural architecture remains a separate problem, as does selection of a unique future continuation. Numerical evaluation includes four synthetic mechanisms and current-vintage population estimates for China, India and Japan. Across six historical evaluation windows per country, the gray-box model attains mean training and test absolute percentage errors of 1.126% and 1.338%, respectively. Its mean test error in these windows is below five of the six comparison implementations but above the 0.684% of a residual universal differential equation. Equal additional training of these two models reduces their test errors to 0.714% and 0.409%. The results support conditional usefulness of the latent-suppression structure while demonstrating limitations in optimization, parameter interpretation and temporal generalization.

Keywords: neural ordinary differential equations; logistic dynamics; latent states; identifiability; time-series extrapolation

MSC: 34A55; 68T07; 92D25

Check for updates

Academic Editor: Sophia Jang

Received: 29 September 2026
Revised: 20 October 2026
Accepted: 22 October 2026
Published: 26 October 2026

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

1. Introduction

Sparse scalar observations can constrain a trajectory without uniquely determining the dynamics that generated it. We distinguish three tasks: reconstructing observed values, extrapolating to later observations and identifying mechanistic parameters. Each requires separate evidence; a small fitting error alone addresses only the first.

Neural ordinary differential equations (NODEs) parameterize a continuous-time vector field with a neural network [1]. Augmented NODEs enlarge the state space to address

restrictions on representable flows [2], while latent ODE models use continuous hidden dynamics for irregularly sampled observations [3]. These developments motivate the use of unobserved states, but do not by themselves give those states a unique mechanistic interpretation.

Combining neural components with partial first-principles models predates NODEs [4]. Universal differential equations (UDEs) express this combination through differential equations with learnable components [5], within the broader scope of physics-informed machine learning [6]. Related approaches include residual-based physics-informed neural networks [7] and sparse identification of dynamical systems [8]. The present implementation instead fits an explicitly integrated ODE: it neither penalizes a differential-equation residual at collocation points nor selects terms from a sparse library. These approaches provide context, not additional experimental baselines.

In related work on green-behavior dynamics, Zhu et al. compared CNN-based parameter estimation with optimal-control approaches and reported training-error histories and optimization trajectories [9]. This provides an applied example of reporting estimation performance together with optimization behavior.

The present model follows an existing logistic–dynamic-suppression structure. In ecological state-space UDEs, Buckner et al. used a logistic biomass equation and a learned exploitation-rate equation in a fisheries example [10]. We study a positive latent-suppression version with a bounded continuous-time closure, rather than claiming the underlying coupling as a new model family. Unlike that ecological state-space framework, the present implementation is a deterministic trajectory fit and does not separately model process and measurement noise.

Our analysis links the state-domain restrictions to the set of mechanistic parameters compatible with a scalar trajectory. Structural identifiability concerns uniqueness under ideal input–output information [11,12]; practical identifiability additionally depends on finite, noisy observations [13,14]. A low optimization loss is not a test of either property. We evaluate the structure against six baselines and two ablations using controlled synthetic data and population estimates from three countries. Population is an application-level test of aggregate trajectories, not evidence that the latent state measures a particular socioeconomic mechanism. The study addresses where the structure is useful and where it is insufficient, including unfavorable comparisons.

2. Model and Analytical Properties

2.1. Logistic Dynamics with Positive Latent Suppression

Logistic-type growth curves provide a compact description of density-dependent growth, with known restrictions on the trajectories they can represent [15]. Here we retain the logistic baseline and introduce a positive dynamic loss relative to it. For an observed scale $P(t) > 0$ and latent suppression rate $U(t) > 0$, consider

$$\dot{P} = rP\left(1 - \frac{P}{K}\right) - UP, \qquad \dot{U} = Uf_\theta(P, U, t), \tag{1}$$

with $r, K, P_0, U_0 > 0$. Here, P and K have units of population, while r, U and f_θ have units of inverse time. There is no separate positive-intervention term. The logistic term supplies baseline growth, and U is an effective loss relative to that baseline. Consequently, a fitted K is a model scale and is not automatically an independently established demographic carrying capacity.

Set $P = Sx$, $t = T\tau$ and $u = TU$. With $\rho = Tr$, $\kappa = K/S$ and $F_\theta = Tf_\theta$, the dimensionless equations are

$$x' = \rho x(1 - x/\kappa) - ux, \qquad u' = uF_\theta(x, u, \tau), \tag{2}$$

Primes denote derivatives with respect to τ. Population experiments use $T = 10$ years and a training-only mean population S. The closure is

$$\begin{aligned} F_\theta &= \tfrac{1}{2}\tanh\{W_2 \tanh(W_1 X + b_1) + b_2\}, \\ X &= (\log x, \log u + 3, \tau/20)^{\mathsf{T}}, \end{aligned} \tag{3}$$

with four hidden units. The fixed input offsets are numerical choices, not observations of the latent state. There are 21 network parameters and three additional learned quantities, $\log\rho$, $\log\kappa$ and $\log u_0$. The observed initial population is fixed. We integrate $z = \log x$ and $v = \log u$:

$$z' = \rho(1 - e^z/\kappa) - e^v, \qquad v' = F_\theta(e^z, e^v, \tau). \tag{4}$$

This representation enforces positivity in exact arithmetic, but it does not eliminate floating-point overflow or discretization error.

2.2. Positive Solutions and Turning Points

Proposition 1. *Suppose that f_θ is defined for positive states and all $t \geq 0$, is continuous in time and locally Lipschitz in the state variables, locally uniformly in time, and satisfies $|f_\theta| \leq M$. Equation (1) has a unique forward-global positive solution. For every $t \geq 0$,*

$$\begin{aligned} 0 < P(t) &\leq C := \max(P_0, K), \\ U_0 e^{-Mt} &\leq U(t) \leq U_0 e^{Mt}. \end{aligned} \tag{5}$$

Proof. On the maximal existence interval,

$$U(t) = U_0 \exp\left(\int_0^t f_\theta(P(s), U(s), s)\, ds\right),$$

which gives the bounds for U. Since $\dot{P} \leq rP(1 - P/K)$, scalar comparison gives $P \leq C$. For $M > 0$, integrating

$$\frac{\dot{P}}{P} \geq r - \frac{rC}{K} - U_0 e^{Mt}$$

gives the explicit lower bound

$$P(t) \geq P_0 \exp\left[\left(r - \frac{rC}{K}\right)t - \frac{U_0}{M}(e^{Mt} - 1)\right] > 0. \tag{6}$$

For $M = 0$, the corresponding bound is $P_0 \exp[(r - rC/K - U_0)t]$. Together with the upper bounds, this keeps both states in a compact subset of the positive domain on each finite interval of existence and prevents a finite-time exit. The solution therefore extends globally. □

Positivity at every finite time is distinguished from uniform persistence, $\inf_{t\geq 0} P(t) > 0$. For example, for the admissible constant closure $f \equiv m$ with $0 < m \leq M$, one has $U(t) = U_0 e^{mt}$ and

$$P(t) \leq P_0 \exp\left[rt - \frac{U_0}{m}(e^{mt} - 1)\right] \longrightarrow 0.$$

Thus exponential growth of U and decay of P to zero remain possible under a bounded relative rate. When $M = 0$, $U_0 > r$ also gives $P(t) \to 0$. The proposition's assumptions are

satisfied by the neural closure in (3) after time scaling; the resulting positivity guarantee applies at each finite time.

Let $g = \dot{P}/P = r - rP/K - U$. At an interior stationary point t_* with $P(t_*) > 0$ and $\dot{P}(t_*) = 0$,

$$\ddot{P}(t_*) = -P(t_*)\dot{U}(t_*). \tag{7}$$

A positive $\dot{U}(t_*)$ therefore implies a strict local maximum; a negative value implies a strict local minimum. The zero case requires higher-order information. This conditional criterion does not establish existence, uniqueness or predictive accuracy of a peak.

2.3. Observational Equivalence and Compatible Parameters

Write $a = r$, $b = r/K$, so that $g = \dot{P}/P = a - bP - U$. Throughout this subsection, compatibility is defined for an ideal positive path on a finite interval $I = [0, H]$, $H > 0$. Bounded closures are freely chosen, and the latent initial value is allowed to vary. This function class is broader than the specific neural architecture in (3), whose identifiability is a separate question.

For an admissible path and $c > 0$, set

$$\widetilde{a} = a + c, \qquad \widetilde{b} = b, \qquad \widetilde{U}(t) = U(t) + c. \tag{8}$$

The same observed P is generated on I by the time-only closure

$$\widetilde{f}(t) = \frac{\dot{U}(t)}{U(t) + c}, \qquad |\widetilde{f}(t)| \le M\frac{U(t)}{U(t) + c} \le M. \tag{9}$$

This function is independent of the state arguments and is consequently defined on the whole positive state domain for $t \in I$. A bounded continuous extension in time, for example a constant extension after H, is possible. Pathwise compatibility is thereby established using a time-only closure on the full positive state domain. Representability of the transformed closure within the fixed tanh architecture is a separate requirement.

The following characterization makes the trajectory-level restriction explicit.

Proposition 2 (Compatible-parameter set)**.** *Let $P \in C^2(I)$ be positive on $I = [0, H]$, $H > 0$, let $g = \dot{P}/P$ and let $M \ge 0$. Define*

$$\begin{aligned} \mathcal{A}_M(P) = \{(a, b) \in (0, \infty)^2 : \\ a - bP(t) - g(t) > 0, \\ |b\dot{P}(t) + \dot{g}(t)| \le M[a - bP(t) - g(t)] \\ \forall t \in I\}. \end{aligned} \tag{10}$$

A pair (a, b) belongs to this set if and only if a positive C^1 state U and a continuous time-only closure $f(t)$ with $|f(t)| \le M$ generate P through (1) with its closure replaced by $f(t)$, with $r = a$ and $K = a/b$. The set $\mathcal{A}_M(P)$ is convex.

Proof. Necessity follows from $U = a - bP - g > 0$ and $\dot{U} = -b\dot{P} - \dot{g} = Uf$. Conversely, for any feasible pair define

$$U = a - bP - g, \qquad f = -\frac{b\dot{P} + \dot{g}}{a - bP - g}.$$

The denominator has a positive minimum on I, so f is continuous, bounded by M, and yields both equations. A time-only closure is also locally Lipschitz in the state variables. For each fixed time, positivity is a strict affine half-space condition, and the absolute-value

bound is equivalent to two affine inequalities. Their intersection with the positive quadrant is convex. □

Corollary 1 (Finite-interval compatibility for $M > 0$)**.** *Let $P \in C^2([0, H])$ be positive, $H > 0$ and $M > 0$. If $a, b > 0$ have no additional upper bounds and U_0 is free, then $\mathcal{A}_M(P)$ is nonempty. In fact, for every fixed $b > 0$ there is a strictly feasible pair (a, b).*

Proof. Let $g = \dot{P}/P$ and define the finite constants

$$B = \max_{t \in I}\{bP(t) + g(t)\}, \qquad D = \max_{t \in I} |b\dot{P}(t) + \dot{g}(t)|.$$

Choose $a > \max\{0, B + D/M\}$. Then, $U = a - bP - g \geq a - B > 0$ and

$$|\dot{U}| \leq D < M(a - B) \leq MU.$$

Proposition 2 supplies the continuous time-only closure $f = \dot{U}/U$ and $r = a$, $K = a/b$. The initial value is $U_0 = a - bP(0) - g(0) > 0$. □

The restriction $M > 0$ is essential to this corollary, although Proposition 2 remains valid for $M = 0$. If $M = 0$, feasibility requires $b\dot{P} + \dot{g} = 0$; for $P(t) = e^t$, one has $g = 1$ and $b\dot{P} + \dot{g} = be^t > 0$, so no $b > 0$ is feasible. The corollary shows that, with unrestricted parameters and free bounded closures, finite-interval compatibility alone imposes no additional restriction on a positive C^2 observed path. Accurate extrapolation and identification of a suppression mechanism require additional evidence.

Corollary 2 (Local two-dimensional nonuniqueness)**.** *Suppose $(a, b) \in \mathcal{A}_M(P)$ satisfies*

$$|b\dot{P}(t) + \dot{g}(t)| < M[a - bP(t) - g(t)] \quad (t \in I).$$

Then (a, b) is an interior point of $\mathcal{A}_M(P)$ in $\mathbb{R}^2$. In particular, under Corollary 1, $\mathcal{A}_M(P)$ has nonempty interior, not merely an unbounded feasible ray.

Proof. Continuity and compactness give

$$\delta_U = \min_I (a - bP - g) > 0,$$
$$\delta_F = \min_I \{M(a - bP - g) - |b\dot{P} + \dot{g}|\} > 0.$$

Put $P_{\max} = \max_I P$ and $V_{\max} = \max_I |\dot{P}|$. For perturbations satisfying $a + \Delta a > 0$, $b + \Delta b > 0$ and

$$|\Delta a| + P_{\max}|\Delta b| < \delta_U,$$
$$M|\Delta a| + (MP_{\max} + V_{\max})|\Delta b| < \delta_F,$$

the perturbed latent state stays positive. The triangle inequality gives a perturbed feasibility margin of at least $\delta_F - M|\Delta a| - (MP_{\max} + V_{\max})|\Delta b| > 0$. These strict conditions hold in a sufficiently small open neighborhood of zero, proving the claim. Corollary 1 supplies a strictly feasible pair when $M > 0$. □

Convexity here is in the (a, b) coordinates. The smooth bijection $r = a$, $K = a/b$ preserves local nonuniqueness but does not in general preserve convexity. If (a, b) is feasible, $(a + c, b)$ remains feasible for every $c > 0$. Parameter restrictions can truncate this ray or the feasible neighborhoods; they do not provide independent observations of U.

For the neural closure defined in Equation (3), with $T = 10$ years, the dimensional bound is $M = 0.05\,\text{year}^{-1}$. Proposition 2 gives a necessary trajectory-level condition for that network. Sufficiency and the two corollaries are established in the freely chosen closure class. Realization by the fixed network within its bounded parameter search requires a separate analysis. Application to noisy annual observations would additionally require derivative regularization, beyond the present ideal-observation analysis.

Continuation Beyond the Observation Interval

The closure after H and the resulting future population remain undetermined by Proposition 2. To see this for $M > 0$, write $s = f(H)$ and, for any $q \in [-M, M]$ and $\tau_0 > 0$, extend the same observed-interval closure by

$$f_q(t) = s + (q - s)\big(1 - e^{-(t-H)/\tau_0}\big), \qquad t \geq H. \tag{11}$$

Each extension is continuous, bounded by M and agrees at H. Distinct q yield distinct future latent paths because the integrals of the extensions differ for $t > H$. Their population paths cannot remain identical on a subsequent interval: for fixed a, b and $P > 0$, the population equation would then determine the same U. The nonuniqueness concerns the choice of the future closure; solution uniqueness is retained for each fixed complete vector field. When $M = 0$, this particular family collapses to the zero closure.

For completeness, two independent latent-state calibration measurements can resolve the two mechanistic coefficients in ideal continuous observations. If $q_i = g(t_i) + U(t_i)$ and $P(t_1) \neq P(t_2)$, then

$$b = \frac{q_1 - q_2}{P(t_2) - P(t_1)}, \qquad a = q_1 + bP(t_1). \tag{12}$$

For $a, b > 0$, $r = a$ and $K = a/b$ are unique, and $U = a - bP - g$ is determined along the observed trajectory. This does not identify the off-trajectory closure or unique neural weights. If only q_i have errors bounded by ε_i,

$$|\delta b| \leq \frac{\varepsilon_1 + \varepsilon_2}{|P(t_2) - P(t_1)|}.$$

The population data contain no independent suppression measurements; this conditional algebraic result is not evidence of unique parameter recovery in the empirical study.

3. Data and Numerical Protocol

3.1. Models and Training

The six comparison models are logistic [15], Gompertz [16], Richards [17], constant-suppression logistic, black-box NODE and residual UDE. Their vector fields and parameter counts appear in Table 1; these explicit definitions avoid ambiguity between growth-curve parameterizations. Two ablations remove the explicit time input from the latent closure (autonomous) or replace it by a learned constant (constant-rate). Constant suppression $u = \gamma$ differs from constant relative suppression rate $u' = \alpha u$.

The constant-suppression control has the population vector field

$$x' = \alpha x - \beta x^2, \qquad \alpha = \rho - \gamma, \quad \beta = \rho/\kappa > 0. \tag{13}$$

For $\alpha > 0$, this is logistic with $r_{\text{eff}} = \alpha$ and $K_{\text{eff}} = \alpha/\beta = \kappa(\rho - \gamma)/\rho$. For $\alpha \leq 0$, it allows nonpositive net growth, which the positive-parameter logistic baseline excludes. Its three active stored parameters therefore enter through only two vector-field coefficients. This implementation is retained as a control for reparameterization and nonpositive net growth. Its role is distinguished from that of an independent dynamic mechanism. Different

bounded search domains, initializations and finite optimization trajectories can produce different errors even when the fitted vector fields belong to the same positive-growth family. Neither classical implementation has a weight penalty.

Table 1. Implemented comparison models. Counts exclude the fixed observed population initial value. Active counts refer to vector-field dependence and exclude zero-input weights, not parameter symmetries.

Model	Log-State Vector Field	Stored	Active
Logistic	$z' = \rho(1 - e^z/\kappa)$	2	2
Gompertz	$z' = \rho(\log\kappa - z)$	2	2
Richards	$z' = \rho[1 - (e^z/\kappa)^\nu]$	3	3
Constant suppression	$z' = \rho(1 - e^z/\kappa) - \gamma$	3	3
Black-box NODE	$z' = N_\theta(z, \tau)$	21	17
Residual UDE	$z' = \rho(1 - e^z/\kappa) + N_\theta(z, \tau)$	23	19
Main model	Equation (4)	24	24
Autonomous	Equation (4), time input set to zero	24	20
Constant rate	Equation (4), $v' = \frac{1}{2}\tanh\eta$	4	4

Active counts are not identifiable dimensions. Constant suppression has three active parameters but only two vector-field coefficients, $\alpha = \rho - \gamma$ and $\beta = \rho/\kappa$. For the main model and both dynamic ablations, $z = \log x$, $v = \log u$. All neural closures have four tanh hidden units and output bound 0.5. The black-box NODE, residual UDE and autonomous ablation each have an identically zero third input.

The Richards implementation in Table 1 is unnormalized. With $q = z - \log\kappa$, $1 - e^{\nu q} = -\nu q + O(\nu^2)$; hence its vector field tends to zero as $\nu \to 0$ with ρ fixed. The alternative normalized field

$$z' = \bar{\rho}\,\frac{1 - (e^z/\kappa)^\nu}{\nu}$$

tends to $\bar{\rho}(\log\kappa - z)$. For $\nu > 0$ the fields agree when $\bar{\rho} = \rho\nu$. The implemented bounds $\log\rho, \log\nu \in [-8, 2]$ map to a coupled region with $\log\bar{\rho} = \log\rho + \log\nu$, not the same independent box for the normalized parameters. Also, the fitted range $\nu \geq e^{-8}$ excludes the literal zero limit. The implemented baseline and its results are retained. The limiting relation is interpreted together with these parameter restrictions, under which the Gompertz and normalized Richards searches have different bounded domains.

All models use float64 fourth-order Runge–Kutta (RK4) integration [18], direct differentiation through the discrete solver and Adam [19]. The distinction between differentiating a discretized solve and discretizing continuous sensitivity equations can matter numerically [20]; we use the former, without implementing ANODE checkpointing. The basic dimensionless step is 0.5 for the dense synthetic grid and 0.1 for annual population observations. Sparse synthetic training retains the same integration grid. For each of two learning rates, 0.01 and 0.003, the model is trained for 600 epochs. Every 20 epochs, validation log mean squared error selects a checkpoint; population experiments additionally select among five initializations. The training objective is

$$\mathcal{L} = \frac{1}{n_{\mathrm{tr}}}\sum_{i\in\mathcal{I}_{\mathrm{tr}}}(\log\widehat{P}_i - \log P_i)^2 + \lambda\frac{\|\theta\|_2^2}{d_\theta}, \tag{14}$$

where $\lambda = 10^{-5}$ for the neural models and zero otherwise. The regularizer covers the stored neural weights, not the mechanistic log parameters. Gradients are clipped to norm one separately for each independent fit, and all raw trainable parameters are clipped to $[-8, 2]$ after an update. Positive mechanistic parameters are exponentials of their raw values. No parameters are shared across countries or trajectories.

The optimization budget is specified by matched learning rates and epoch counts. Equal forward right-hand-side evaluation counts across the nine models are also verified from the fixed-step implementation: each learning-rate candidate uses 144,000 training evaluations and 8060 validation evaluations. With two learning rates and five initializations, the search costs 1,520,600 such evaluations per country–window. The batched training solver always integrates 60 annual intervals, masking residuals after the relevant training cutoff; the validation solver integrates 65 intervals. These counts include those masked intervals but exclude automatic-differentiation backward work and final output checks. FLOPs, wall time, state dimension and representational capacity can differ at equal forward evaluation counts; convergence also requires separate assessment. Supplementary Material, Section S10, reports the accounting, archived timings and all-model loss histories against this cost axis. Each of the black-box NODE, residual UDE and autonomous ablation has four inactive input weights associated with its zero input. These weights remain in the implementation and regularizer; Table 1 reports stored and active counts separately. Stored and active counts describe the implementation; identifiable degrees of freedom require a separate analysis. In the 18 selected population fits, the constant-suppression coefficients satisfy $\alpha \in [0.233117, 0.420401]$, $\beta \in [0.100894, 0.271237]$ and $K_{\text{eff}} \in [1.206889, 3.134799]$ in dimensionless units. Thus every selected field lies in the positive-growth logistic family; all 90 archived initialization-specific historical checkpoints also have $\alpha > 0$. Supplementary Material, Section S9, lists the selected coefficients. For example, the constant-suppression vector field depends on its three mechanistic parameters only through $\rho - \gamma$ and ρ/κ.

3.2. Synthetic Trajectories

We use four known mechanisms: dynamic suppression with constant $d \log U/dt = \alpha$, logistic growth, Gompertz growth and periodic suppression with $d \log U/dt = 0.20 \cos(t/3)$. Five trajectories are generated per mechanism using paired seeds. Growth, carrying scale, initial suppression and initial population are sampled uniformly from $[0.35, 0.55]$, $[1.2, 1.8]$, $[0.02, 0.04]$ and $[0.16, 0.24]$, respectively. The dynamic-suppression rate is uniform on $[0.09, 0.15]$; the Gompertz coefficient is 0.16. These are controlled simulations, not real demographic records.

A high-accuracy solver generates 41 points on $[0, 20]$ at spacing 0.5. The first 21 observations are used for training, the next eight for validation and the final 12 for testing. Gaussian noise with standard deviation $\sigma \in \{0, 0.02, 0.05\}$ is added to log population, except at the fixed initial observation. Sparse training retains 11 alternating training observations while leaving validation and test observations unchanged. Ground-truth mechanistic parameters are not supplied in the main comparison. Seeds are paired across mechanisms and noise levels; the cases are not all independent replications. Each synthetic case has one neural initialization per model and learning-rate candidate.

Synthetic test errors are evaluated against the noise-free trajectory using root mean squared error normalized by mean test population. Population experiments use mean absolute percentage error:

$$\begin{aligned} \text{NRMSE} &= \frac{\sqrt{n^{-1} \sum_i (\widehat{P}_i - P_i)^2}}{n^{-1} \sum_i P_i}, \\ \text{MAPE} &= \frac{100}{n} \sum_i \frac{|\widehat{P}_i - P_i|}{P_i}. \end{aligned} \tag{15}$$

Here, n is the number of observations in the evaluated segment. The synthetic tables multiply NRMSE by 100 and report sample standard deviations across five trajectories. Population MAPE measures average pointwise relative deviation from archived estimates;

synthetic NRMSE measures root mean square deviation from the noise-free generated path, normalized by its test mean, and is more sensitive to large residuals. These percentages use different normalizations and reference targets and cannot be compared directly across the two experiments. Both metrics summarize trajectory error; mechanistic recovery, turning-point timing and predictive interval coverage require separate assessments. Their interpretations are summarized in Table 2. Neither value represents the percentage of predictions that are correct. Both metrics have additionally been recomputed from the same archived predictions, without retraining (Supplementary Material, Section S12). Forecast accuracy measures have different scaling properties and failure modes [21]. The positive population levels avoid MAPE's zero-denominator problem, but this does not make MAPE sufficient for assessing turning-point timing or uncertainty.

Table 2. Interpretation of the error metrics in Equation (15). Both are descriptive trajectory-error measures.

Metric	Question Addressed	Scaling and Sensitivity	Reference in This Study
NRMSE	How large is the root mean squared trajectory error relative to the mean target level?	Squared residuals give greater weight to large absolute deviations. A single mean target level normalizes the evaluated segment.	Noise-free synthetic test trajectory; reported as $100 \times$ NRMSE in tables.
MAPE	What is the average absolute error relative to each target value?	Each absolute residual is divided by its own positive target. Equal absolute errors receive greater weight at smaller population levels.	Archived population estimates; the factor 100 is included in the definition.

3.3. Population Records and Temporal Evaluation

The empirical data are the archived World Bank World Development Indicators series SP.POP.TOTL [22]: 66 annual observations for each of China, India and Japan from 1960 through 2025, totaling 198 records without interpolation. They are midyear population estimates compiled from statistical and demographic sources, not a sequence of annual censuses. The archived response records an update date of 13 July 2026. Its exact bytes and checksum accompany the reproducibility package. Evaluation uses this single current data vintage; it is not a historical real-time exercise based on past publication vintages.

Following chronological rolling evaluation [23], the training end years are 1990, 1995, 2000, 2005, 2010 and 2015. Each is followed by five validation years and five test years. Thus the six test blocks cover 1996–2025 without overlap within a country. Validation selects hyperparameters and checkpoints but is not used for gradient updates. The trajectory is not reset to the observed population at the test origin. Test points are therefore years 6–10 after the training cutoff, or years 1–5 after the validation cutoff. Initial-state values and scaling are obtained from the training segment. Aggregate results give equal weight to the 18 country–window combinations. Training windows overlap and country trajectories are serially dependent, so these 18 combinations are not treated as independent statistical replicates.

Evaluation schemes for time series depend on the forecasting task and dependence structure [24]. Standard cross-validation can be valid for certain autoregressive settings with uncorrelated errors [25]; those conditions are not established here. Chronological splits match our intended extrapolation task, but they do not remove the limitations of a small, dependent sample.

4. Results

4.1. Historical Population Reconstruction and Extrapolation

Table 3 jointly reports fit and test performance under the matched budget. The main model reaches 1.126% mean training MAPE and 1.338% mean test MAPE. In these 18 windows, its mean test error is lower than that of five comparison implementations, but the residual UDE is more accurate at 0.684%. The main model has lower test MAPE than the residual UDE in seven of 18 individual windows. These comparisons are interpreted descriptively because the windows are dependent; the win count is therefore reported as a summary of the evaluated windows.

Table 3. Mean population error across 18 country–window combinations. All values are MAPE (%).

Model	Training	Test
Logistic	1.620	2.119
Gompertz	1.560	2.568
Richards	1.539	1.633
Constant suppression	1.976	2.084
Black-box NODE	1.623	2.416
Residual UDE	0.980	0.684
Main model	1.126	1.338
Autonomous	1.124	1.340
Constant rate	1.378	1.626

The comparison is heterogeneous across countries (Table 4). In Japan, the main model substantially improves over the conventional growth curves, whereas logistic and Richards models already obtain comparable or lower error in China. The black-box NODE and residual UDE are more accurate in India. Figure 1 presents the full heatmap for all nine models across countries and evaluation windows. Figure 2 shows the corresponding within-window paired differences, and Figure 3 displays the window-specific errors for five representative models; all nine models are retained in the tables and Supplementary Material. Further paired comparisons are provided in Supplementary Material, Section S11. Inference based on independent windows is excluded because the observations are dependent. Variation across initializations is interpreted as optimization sensitivity, separately from population forecast uncertainty.

Table 4. Mean test MAPE (%) across six windows within each country.

Model	China	India	Japan
Logistic	0.694	2.664	2.999
Gompertz	0.799	3.971	2.933
Richards	0.675	2.113	2.111
Constant suppression	0.816	3.372	2.064
Black-box NODE	3.005	1.371	2.872
Residual UDE	0.858	0.534	0.660
Main model	0.701	2.620	0.692
Autonomous	0.720	2.619	0.681
Constant rate	0.871	3.216	0.791

For each country and evaluation window, the paired difference is defined as $\Delta = \mathrm{MAPE}_{\mathrm{main}} - \mathrm{MAPE}_{\mathrm{comparator}}$ in percentage points. Negative values indicate lower main-model error. Figure 2 displays all 18 comparisons against the residual UDE, logistic and Richards implementations. The same country and window are used for both members of each pair; the dependent windows are interpreted descriptively.



Figure 1. All nine implementations across the 18 historical country–window combinations. Cell values are test MAPE (%). The dependent windows are displayed for descriptive comparison. Constant suppression is a reparameterization control.

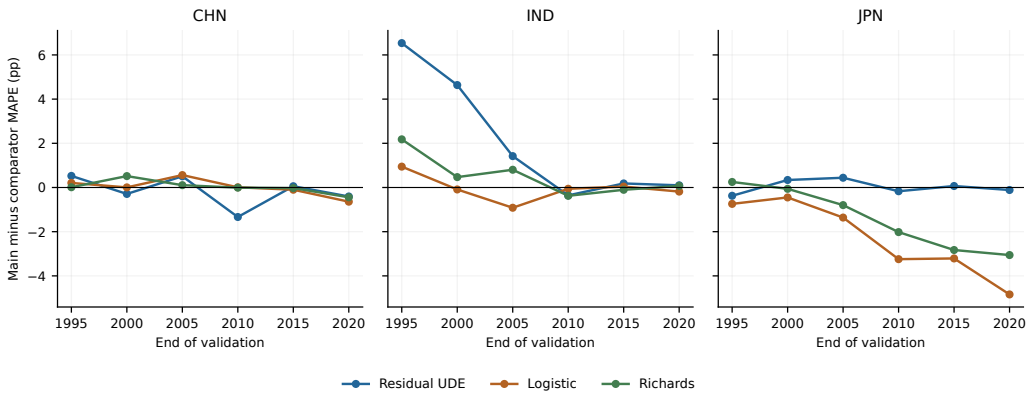


Figure 2. Within-window test-MAPE differences in percentage points, main model minus comparator. Negative values favor the main model. All six windows per country are displayed. Windows are dependent; the figure supports descriptive comparison, with no significance claim or population-level confidence interval.

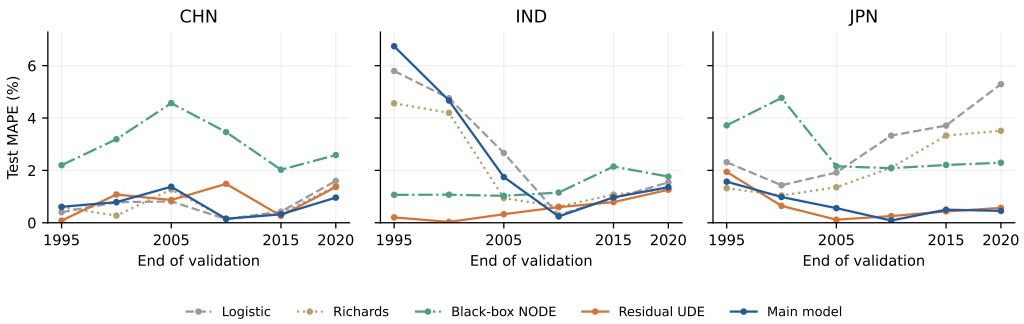


Figure 3. Historical test MAPE by country and validation cutoff. Each point summarizes the next five years. Model selection uses validation data only. Selected comparators are shown for readability; complete window scores are supplied in the Supplementary Material.

Figure 4 shows population trajectories for the latest historical training cutoff, 2015, chosen by chronology rather than test performance. These plots distinguish visual agreement at the population scale from accurate prediction of small changes. In particular, a model can have low percentage error while missing a recent reversal, so a fitted trajectory alone is not used to validate the turning-point criterion empirically.

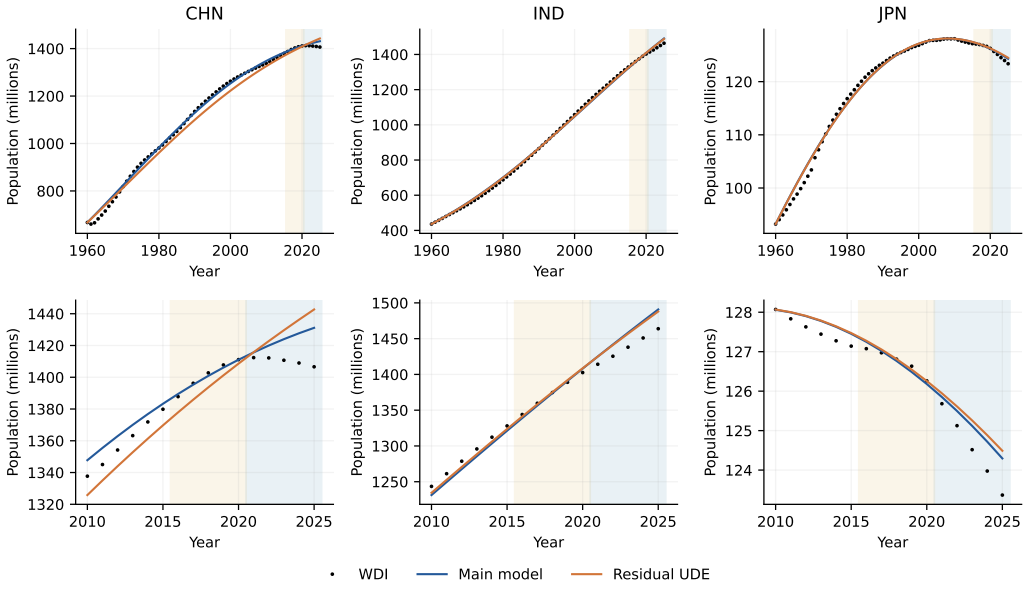


Figure 4. Validation-selected trajectories for the 2015 training cutoff. Top: 1960–2025 reconstruction. Bottom: 2010–2025 detail. Shaded regions mark validation (2016–2020) and test (2021–2025); black points are archived WDI estimates. No observation is used to reset the state at either boundary.

4.2. Controlled Synthetic Evaluation

At $\sigma = 0.02$, the main model performs well relative to several baselines for dynamic suppression and logistic generation, but it is not best for Gompertz or periodic suppression (Table 5). Its dynamic-suppression NRMSE is $20.98 \pm 22.47\%$, indicating substantial variability over only five trajectories. Consequently, the mean improvement does not establish reliable recovery for every trajectory. The five errors are 54.10%, 6.68%, 6.55%, 2.97% and 34.61%; the median is 6.68%, with quartiles 6.55% and 34.61%. The mean and standard deviation are retained alongside these summaries. All five seed-specific paired differences against the eight comparators are displayed for this setting in Figure 5. Complete noise and density settings are supplied in Supplementary Material, Section S13. The seed pairing is retained within each setting, and repeated seeds across settings are treated as related cases.

Table 5. Dense synthetic observations with log-noise standard deviation 0.02. Values are test NRMSE (%), mean ± sample standard deviation over five trajectories.

Model	Dynamic	Logistic	Gompertz	Periodic
Logistic	40.07 ± 20.04	0.80 ± 0.56	8.28 ± 1.72	1.66 ± 0.89
Gompertz	39.09 ± 27.22	5.65 ± 1.43	4.34 ± 4.20	6.57 ± 1.88
Richards	37.20 ± 22.87	1.20 ± 1.22	7.58 ± 1.85	2.72 ± 2.26
Constant suppression	41.36 ± 22.54	1.27 ± 0.66	7.36 ± 2.26	3.47 ± 4.14
Black-box NODE	32.47 ± 20.20	5.46 ± 2.54	7.32 ± 0.84	1.64 ± 1.05
Residual UDE	38.22 ± 18.55	1.66 ± 0.95	3.85 ± 3.18	3.26 ± 3.39
Main model	20.98 ± 22.47	0.56 ± 0.49	6.46 ± 2.06	2.97 ± 2.60
Autonomous	21.58 ± 21.92	0.49 ± 0.35	6.70 ± 1.36	2.64 ± 2.70
Constant rate	22.17 ± 24.22	1.74 ± 1.01	7.46 ± 2.03	2.67 ± 2.64

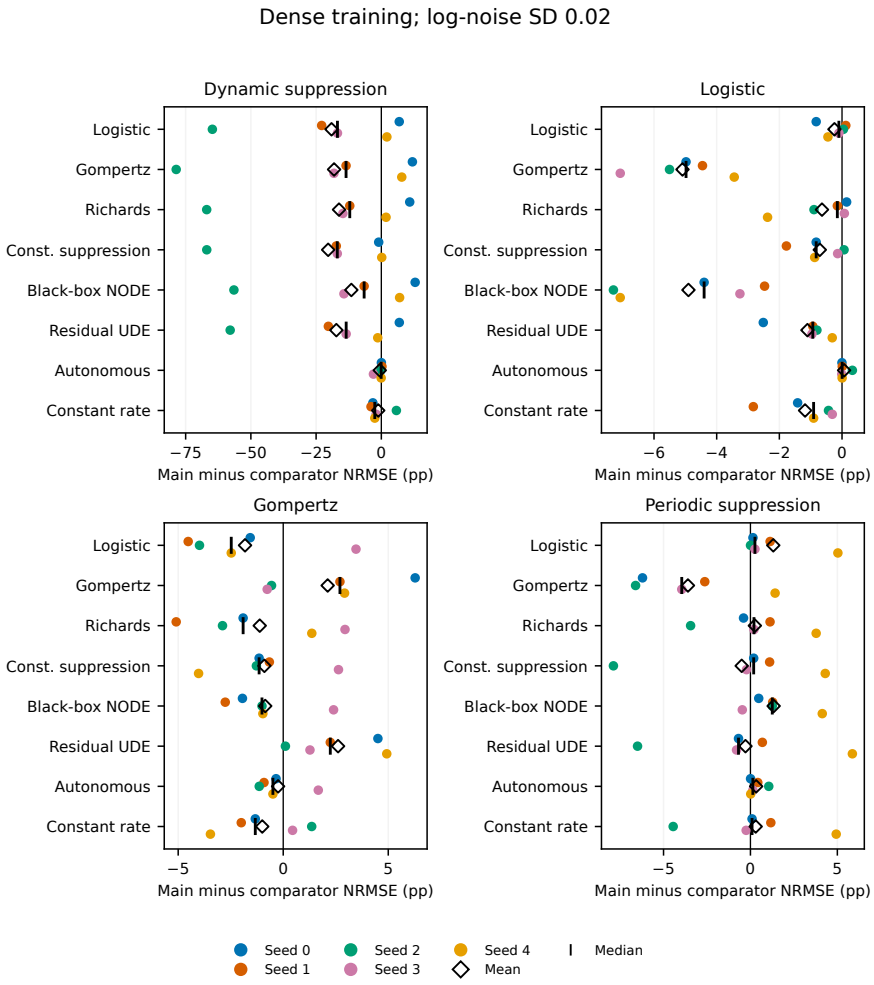


Figure 5. Dense synthetic observations at log-noise SD 0.02, matching Table 5. Each colored point is a same-seed difference in test NRMSE (percentage points), main model minus the named comparator. Hollow diamonds denote means and black ticks medians. Negative values favor the main model. All five trajectories are shown, with no confidence-interval interpretation.

With sparse training, the main-model mean errors for the four mechanisms are 20.29%, 0.70%, 6.49% and 3.69%, respectively, at the same noise level. Results for all noise levels and observation densities are retained in the Supplementary Material. This experiment addresses a specific regular subsampling pattern, not arbitrary missingness.

An additional, separate known-physics diagnostic supplies the true growth and carrying parameters for the dynamic-suppression generator. At $\sigma = 0.02$, main, autonomous and constant-rate errors are approximately 7.36%, 9.39% and 2.77%. The constant-rate

model matches the generator and benefits from this information. These results must not be pooled into the unknown-physics ranking; they show that increased neural flexibility is not always necessary.

4.3. Ablations and Optimization Sensitivity

The autonomous ablation attains 1.340% mean population test MAPE, close to the main model's 1.338%; the constant-rate ablation attains 1.626%. A lower mean error is obtained by the main model than by the constant-rate ablation in these windows. The near-equal main and autonomous errors provide insufficient evidence for a robust advantage of explicit time dependence.

A single monotone path also has limited ability to distinguish unrestricted closure dependence. If an observed continuous P is strictly monotone on I, its inverse $h : P(I) \to I$ is continuous. For any continuous time-only closure f, the function $F(p) = f(h(p))$ is continuous on $P(I)$ and satisfies $F(P(t)) = f(t)$ throughout I. An exact continuous representation is thus obtained along the observed path. Its scope is limited to that path and to continuous functions: representation by the fixed four-unit tanh network, off-path equivalence and the observed optimization errors require separate analysis. A Lipschitz inverse requires stronger assumptions than strict monotonicity, so the state regularity required by Proposition 1 must be checked separately for an autonomous extension. The inverse construction is restricted to strictly monotone paths. Equivalence of the implemented networks is therefore left unresolved.

After inspecting the original test results, we conducted a development diagnostic for the main model and strongest baseline. Both start from their own validation-selected checkpoints for all five initializations, reset the Adam state and receive 1200 additional epochs at learning rate 0.003. Validation can retain the original checkpoint. Training targets alone enter gradient updates. This diagnostic is not a new independent confirmation experiment and is not combined with unrefined models in a matched-budget leaderboard. Separating this diagnostic from the original comparison is important because repeated model selection can itself bias performance estimates [26]; validation-based checkpoint selection does not erase prior inspection of test results.

The main-model mean test error falls to 0.714%, while residual UDE error falls to 0.409% (Table 6). After refinement, the main model has country-specific errors of 0.518%, 0.924% and 0.700%, compared with 0.537%, 0.237% and 0.453% for the residual UDE. The main model is more accurate in two of 18 individual windows. Optimization budget explains part of the earlier error but does not reverse the overall comparison.

Table 6. Equal additional optimization for two models. Values are mean MAPE (%). This post-evaluation development diagnostic is separate from the matched-budget comparison.

Model	Fit Before	Fit After	Test Before	Test After
Main model	1.126	0.901	1.338	0.714
Residual UDE	0.980	0.719	0.684	0.409

4.4. Nine-Model Budget Sensitivity

A revision sensitivity analysis was conducted for all nine models at three predetermined budgets: 200, 400 and 600 updates per learning-rate and initialization candidate. The original chronological splits, two learning rates, five initializations and validation rule were retained. Checkpoints were assessed every 20 updates, including initialization. At each budget, the minimum validation log-MSE was used to select a checkpoint, learning rate and initialization independently for each country–window case. All three budgets were retained for reporting. The protocol was fixed before these runs, and test errors were com-

puted after all training had been completed. As the original test results had already been examined during development, these runs are interpreted as revision sensitivity analyses.

For $E \in \{200, 400, 600\}$, the per-candidate training and validation forward counts are $240E$ and $260(E/20 + 1)$. Across ten candidates per country–window search, the totals are 508,600, 1,014,600 and 1,520,600 forward RHS evaluations, respectively. These counts were checked by instrumenting RHS calls. Training time includes reverse automatic differentiation, clipping and optimizer updates; its cost is recorded separately from forward NFE. State dimensions and stored/active parameter counts are supplied with the timing results in Supplementary Material, Section S14. The neural implementations retain their original four-unit widths and active counts, so capacity differences remain part of the comparison.

Figure 6 shows the selected training objective, validation loss, test error and measured time at every budget. Training and validation histories for the latest historical cutoff, 2015, are displayed in Figure 7; this window is selected by chronology. All six cutoffs are included in Supplementary Material, Section S14. The complete regularized objective and the pure training residual are distinguished. Each history corresponds to the candidate selected at the 600-update budget; its checkpoint is marked, and its selected learning rate and initialization are retained in the accompanying case table. Training and test MSE histories for CNN-based estimation, together with parameter-error trajectories for GD, BB and BFGS, are reported by Zhu et al. [9]. Here, the loss histories are accompanied by instrumented forward NFE, measured wall-clock time and active parameter counts, so optimization behavior and computational cost can be assessed separately.

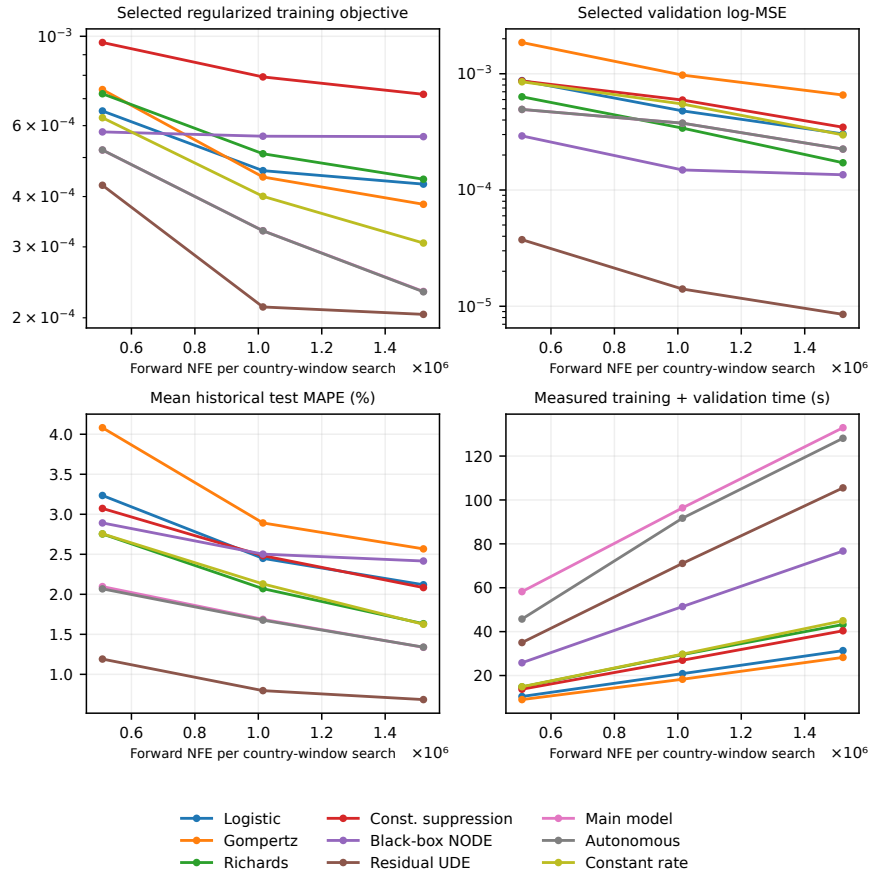


Figure 6. Nine-model revision sensitivity analysis at 200, 400 and 600 updates per candidate. The cost axis is forward NFE over ten candidates per country–window search. Accuracy and objective values are averaged over 18 dependent historical cases. Time is measured for the full 90-row batch over both learning rates, including backward work and validation. One timing run was recorded; these measurements describe this implementation and environment.

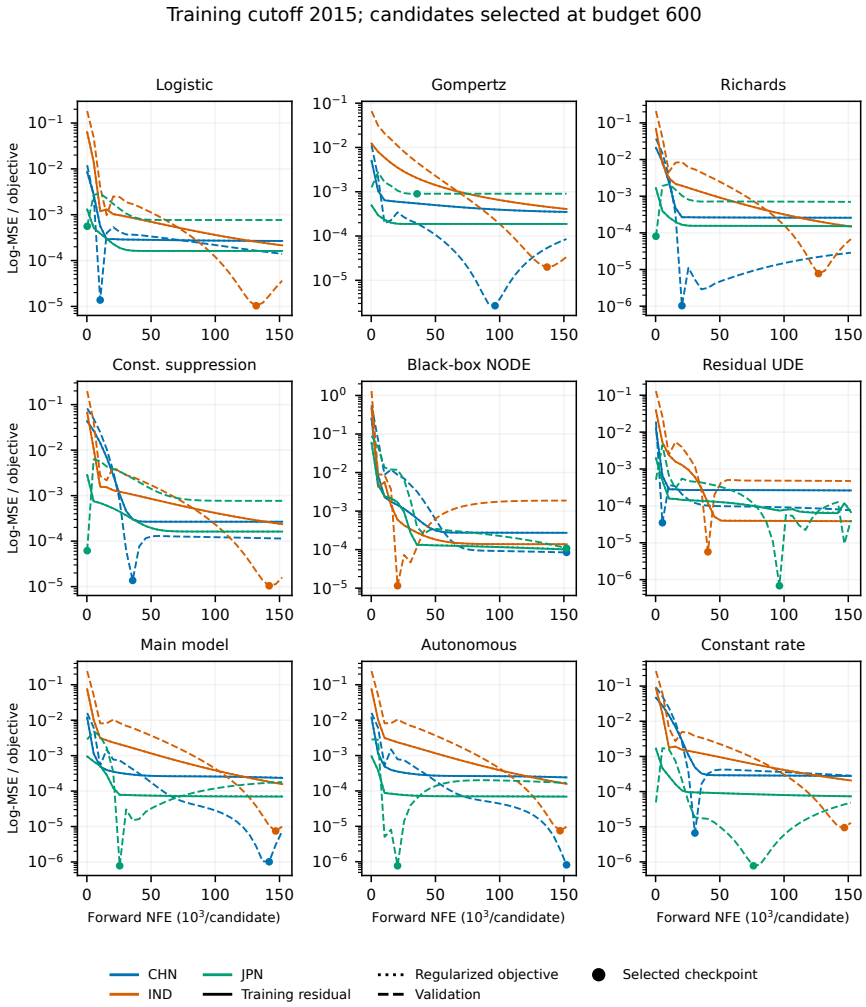


Figure 7. Nine-model training and validation histories for cutoff 2015. Colors identify countries. Solid lines show training residual log-MSE, and dashed lines show validation log-MSE. Dots mark the checkpoint selected at the largest budget. Histories are recorded every 20 updates and continue to 600. The axis counts forward NFE within one candidate.

Across the three budgets, mean test MAPE for the main model is 2.096%, 1.687% and 1.338%, compared with 1.191%, 0.796% and 0.685% for the residual UDE. The residual UDE has lower mean error at each reported budget. The curves show sensitivity to the available optimization budget; an asymptotic convergence rate is outside the scope of these measurements. Forward NFE is matched across models within each budget, while the measured time and active parameter counts describe their different evaluation costs and capacities.

The 600-update runs closely reproduce the archived population results. The main-model mean remains 1.338%; the residual-UDE mean is 0.685% in this rerun, compared with the archived 0.684%. The original tables are retained, and the rerun results are reported separately with their batch configuration and software environment.

4.5. Numerical and Reporting Checks

The original absolute gradient discrepancy of approximately 2.38×10^{-10} came from a central difference with step 10^{-6} for the terminal log population of one initialized main-model trajectory on $\mathrm{linspace}(0, 2, 5)$. The scope of that check was a single initialized trajectory. The actual regularized population training objective was subsequently checked for all nine models and all 18 historical cases, at both reconstructed initial and archived

validation-selected trained weights (324 states). Central differences use $h_j = \eta \max(1, |w_j|)$ for $\eta \in \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}\}$, without optimizer updates or parameter projection. Supplementary Material, Section S7, gives the scheme, gradient magnitudes, absolute and relative discrepanciesand the complete step sweep.

At $\eta = 10^{-5}$, 323 of 324 states satisfy both maximum absolute discrepancy $\leq 10^{-7}$ and relative vector discrepancy $\leq 10^{-5}$. The remaining state is the trained Gompertz fit for Japan with training ending in 2010: its gradient norm is only 6.17×10^{-10}. Its absolute discrepancy is 1.79×10^{-13} but its relative discrepancy is 1.45×10^{-4}. This finite-difference exception is included in the reported results. Independent 80-digit differentiation of the same discrete RK4 objective gives maximum absolute and relative vector discrepancies of 8.50×10^{-18} and 7.03×10^{-9} against automatic differentiation. The implementation is supported by these checks at the tested states. Optimization convergence and gradient accuracy at other states require separate assessment.

With weights, initial states and output times held fixed, integrations were also performed at $h, h/2, h/4, h/8$ and the following quantities were computed:

$$D_h = \max_i |P_h(t_i) - P_{h/2}(t_i)|, \qquad p_{\text{fine}} = \log_2(D_{h/2}/D_{h/4}).$$

The check covers all 162 selected historical population fits, 36 refined fits and 1080 dense/sparse synthetic fits. Population comparisons end at each case's own test endpoint; synthetic comparisons use $[0, 20]$. The resolved fine-grid orders range from 3.970 to 4.014 for the main population comparison, 3.913 to 4.027 for refinement, 3.985 to 4.103 for dense synthetic data and 4.006 to 4.099 for sparse data. Four population ratios fall below the stated roundoff screen and are not interpreted. The observed orders are consistent with fourth-order convergence on the tested trajectories; stability at arbitrary step sizes is outside the scope of this check.

The original approximately 239-person step difference is reproduced as 239.381 people when refined paths are compared through the common archived endpoint, 2025, even for earlier windows. Restricting each comparison to its own historical test endpoint gives maxima of 34.955 people for the original selected fits and 199.634 people for the refined fits. The reported maxima are therefore associated explicitly with their respective comparison intervals. The full protocol and per-case results are supplied in Supplementary Material, Section S8.

As an independent reporting check, all 162 selected population model–window errors and all 1080 dense/sparse synthetic case errors were recomputed from stored trajectories. Maximum discrepancies from the existing tables are 8.33×10^{-15} percentage points for population MAPE and 9.97×10^{-17} for dimensionless synthetic NRMSE. The recomputation uses archived estimates and saved predictions without retraining; it verifies numerical traceability, not independent data collection.

5. Discussion

The positive suppression state provides an explicit route by which learned dynamics modify logistic growth. It also restricts the correction to be a nonnegative loss relative to the fitted baseline. A residual UDE can introduce signed direct corrections and performs better overall in the population experiments. The evidence does not isolate whether this gap is caused primarily by function-class restrictions, optimization or limited observations.

The analytical results separate state admissibility from parameter uniqueness. Positive solutions and population bounds describe what trajectories the equations permit. The compatible-parameter set describes how different mechanistic interpretations can reproduce the same ideal observation. For $M > 0$ with unrestricted growth parameters and free

latent initialization, every positive C^2 path on a finite interval admits a bounded time-only closure, with a feasible set of nonempty interior. Compatibility is therefore interpreted as consistency with the observed trajectory, while mechanistic identification requires additional information. The bounded finite network used in the experiments belongs to a narrower class, and exact representation of these constructions within that class remains to be established. Different future continuations are admitted when the closure beyond the observed interval is unknown. Likewise, identification of a specific socioeconomic quantity and parameter recovery from noisy annual measurements require information and methods beyond this trajectory-level characterization. The model also omits explicit births, deaths, migration and age structure; the latent state can absorb effects of these omissions, data revisions and model misspecification.

Several limitations constrain interpretation. First, the empirical sample consists of three countries and current-vintage estimates, with overlapping training histories. Second, five synthetic trajectories per mechanism provide limited precision, and their paired construction precludes treating all cases as independent. Third, the nine-model forward-NFE comparison and the two-model refinement diagnostic document optimization sensitivity over the evaluated budgets; convergence beyond these budgets remains to be assessed. Fourth, the study was developed iteratively and is not a preregistered independent confirmatory comparison. Fifth, the closure's explicit time input is evaluated beyond its training range. Finally, the feasible-set proof permits a broader function class than the implemented network and should not be read as a finite-network identifiability guarantee.

Future population projections are not used as accuracy evidence because corresponding observations are unavailable. Previously computed historical-error bands are conditional empirical reference ranges rather than guaranteed probability intervals. The present conclusions concern observed historical evaluation blocks and do not imply a necessary future demographic trajectory.

6. Conclusions

A logistic gray-box NODE with positive latent suppression provides a mathematically admissible framework for reconstructing aggregate trajectories. Its positivity, population bound and local turning-point relation coexist with nonuniqueness of mechanistic explanations under flexible closure functions. The compatible-parameter characterization and its corollaries establish the nonuniqueness of mechanistic parameters for freely chosen bounded closures on a finite observation interval. A complete structural-identifiability analysis of the fixed finite-width neural architecture, including the uniqueness of its weights, remains a separate problem. Selection of a unique future continuation also requires additional information. The positivity guarantee applies at each finite time, while decay to zero over an infinite horizon remains possible. Under the specified numerical budget, the mean population hindcast error over the evaluated windows is lower than for five comparison implementations but higher than for a residual UDE; these differences are interpreted descriptively. Additional optimization improves both models without changing that conclusion. The model is therefore useful under some tested conditions, with explicit limits on what its fitted parameters and extrapolations can establish.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math14213898/s1, Supplementary Material: Additional theoretical and numerical analyses, complete experimental results, optimization histories, gradient validation, numerical step-refinement studies, parameter audits, and computational budget comparisons.

Author Contributions: Conceptualization, J.F.; Methodology, T.X. and J.F.; Software, C.X. and J.F.; Validation, H.M.; Formal analysis, H.M.; Investigation, T.X. and J.F.; Writing—original draft, C.X.

and J.F.; Writing—review and editing, J.F.; Supervision, J.F.; Funding acquisition, T.X. and H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 11975145), the Natural Science Foundation of Inner Mongolia Autonomous Region (Grant No. 2026MS0313), the Scientific Research Start-up Fund of Inner Mongolia University of Technology (Grant No. BS2025078), and the Undergraduate Innovation and Entrepreneurship Training Program of Inner Mongolia University of Technology (Project No. S202610128024).

Data Availability Statement: The data presented in this study are available via World Bank at https://data.worldbank.org/indicator/SP.POP.TOTL, reference number [21].

Acknowledgments: The authors acknowledge the Department of Mathematics, Shanghai University, Shanghai 200444, P.R. China, for its support.

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

References

1. Chen, R.T.Q.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D.K. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2018; Volume 31.
2. Dupont, E.; Doucet, A.; Teh, Y.W. Augmented neural ODEs. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2019; Volume 32.
3. Rubanova, Y.; Chen, R.T.Q.; Duvenaud, D.K. Latent ordinary differential equations for irregularly-sampled time series. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2019; Volume 32.
4. Psichogios, D.C.; Ungar, L.H. A hybrid neural network-first principles approach to process modeling. *AIChE J.* **1992**, *38*, 1499–1511. [CrossRef]
5. Rackauckas, C.; Ma, Y.; Martensen, J.; Warner, C.; Zubov, K.; Supekar, R.; Skinner, D.; Ramadhan, A.; Edelman, A. Universal differential equations for scientific machine learning. *arXiv* **2020**, arXiv:2001.04385.
6. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
7. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
8. Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932–3937. [CrossRef] [PubMed]
9. Zhu, L.; Ding, Y.; Shen, S. Green behavior propagation analysis based on statistical theory and intelligent algorithm in data-driven environment. *Math. Biosci.* **2025**, *379*, 109340. [CrossRef] [PubMed]
10. Buckner, J.H.; Meunier, Z.D.; Arroyo-Esquivel, J.; Fitzpatrick, N.; Greiner, A.; McManus, L.C.; Watson, J.R. Recovering complex ecological dynamics from time series using state-space universal dynamic equations. *Commun. Earth Environ.* **2026**, *7*, 112. [CrossRef]
11. Bellman, R.; Åström, K.J. On structural identifiability. *Math. Biosci.* **1970**, *7*, 329–339. [CrossRef]
12. Villaverde, A.F.; Barreiro, A.; Papachristodoulou, A. Structural identifiability of dynamic systems biology models. *PLoS Comput. Biol.* **2016**, *12*, e1005153. [CrossRef] [PubMed]
13. Raue, A.; Kreutz, C.; Maiwald, T.; Bachmann, J.; Schilling, M.; Klingmüller, U.; Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **2009**, *25*, 1923–1929. [CrossRef] [PubMed]
14. Wieland, F.-G.; Hauber, A.L.; Rosenblatt, M.; Tönsing, C.; Timmer, J. On structural and practical identifiability. *Curr. Opin. Syst. Biol.* **2021**, *25*, 60–69. [CrossRef]
15. Tsoularis, A.; Wallace, J. Analysis of logistic growth models. *Math. Biosci.* **2002**, *179*, 21–55. [CrossRef] [PubMed]
16. Tjørve, K.M.C.; Tjørve, E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS ONE* **2017**, *12*, e0178691. [CrossRef] [PubMed]
17. Richards, F.J. A flexible growth function for empirical use. *J. Exp. Bot.* **1959**, *10*, 290–301. [CrossRef]
18. Hairer, E.; Nørsett, S.P.; Wanner, G. *Solving Ordinary Differential Equations I: Nonstiff Problems*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 1993. [CrossRef]
19. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

20. Gholaminejad, A.; Keutzer, K.; Biros, G. ANODE: Unconditionally accurate memory-efficient gradients for neural ODEs. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 730–736. [CrossRef] [PubMed]
21. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
22. World Bank. World Development Indicators: Population, Total (SP.POP.TOTL). 2026. Available online: https://data.worldbank.org/indicator/SP.POP.TOTL (accessed on 28 September 2026).
23. Hyndman, R.J.; Athanasopoulos, G. Section 5.10: Time series cross-validation. In *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021.
24. Cerqueira, V.; Torgo, L.; Mozetič, I. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Mach. Learn.* **2020**, *109*, 1997–2028. [CrossRef]
25. Bergmeir, C.; Hyndman, R.J.; Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.* **2018**, *120*, 70–83. [CrossRef]
26. Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.

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