Human population dynamics as a Bayesian inverse transport problem

Chong Qi*
Department of Physics, KTH Royal Institute of Technology, SE-10691 Stockholm, Sweden

(Received 14 July 2026; revised 28 September 2026; accepted 20 October 2026; published 2 November 2026)

Many open problems across physical, biological, and engineered systems involve nonequilibrium transport processes where the governing conservation laws are known, but the underlying constitutive relations remain latent and time varying. Conventional data-driven approaches like deep neural networks capture statistical patterns but routinely violate fundamental mass conservation. Here I introduce a unified Bayesian inverse transport framework that resolves this by embedding Bayesian neural networks directly within exact partial differential equations in age-time space. By evaluating this framework on complex, real-world human cohort advection across China, Japan, and South Korea, I demonstrate that this physical constraint enables consistent uncertainty propagation and missing-data reconstruction from sparse observations. Beyond demography, this framework provides a generalizable foundation for observing and forecasting nonequilibrium boundary dynamics across various fields.

DOI: 10.1103/nrg2-2xj9

I. INTRODUCTION

Many dynamical systems in physics are naturally described by transport equations. Examples include neutron transport in nuclear reactors, radiative transfer in stellar atmospheres, charge carrier transport in semiconductors, and stress-strain relations in continuum mechanics. Once the transport mechanism is established, the scientific problem shifts from determining the equation of motion to identifying the system-specific constitutive relations, such as cross sections, opacities, or thermal conductivities. In other words, while the governing transport equation is assumed to be known, the constitutive relations must be inferred from observations.

In this context, human demographics represent a highly nonequilibrium advective flow governed by latent boundary conditions. Under this transport-theoretic view, individuals move deterministically through age space with unit velocity, while births introduce new individuals at the boundary and mortality acts as a time-varying internal sink. The system is governed by a first-order partial differential equation (PDE) for age-structured population transport [1,2]. Here age a serves as the transport coordinate, analogous to spatial position in a one-dimensional advection equation, with new births acting as a boundary influx at $a = 0$. In practice, these continuous dynamics are often approximated using a discrete matrix formulation [3] that maps age-specific survival and fertility into a linear state transition. Existing demographic models generally treat fertility and mortality as prescribed inputs. The cohort-component method projects an observed age distribution forward using deterministic, scenario-based schedules, lacking a native representation of uncertainty [4]. Reference [5] factorizes age-specific vital rates via low-rank decompositions with univariate time-series models. The United Nations Population Division applies hierarchical Bayesian models to scalar summary indicators including the total fertility rate (TFR) and life expectancy [6–8]. Machine-learning approaches often bypass mechanisms altogether and learn population trajectories directly from data. These approaches share common limitations: (i) They treat age-specific vital rates as independent parameters rather than a coupled two-dimensional age-time field, losing cross-cohort correlation structure; (ii) they do not enforce demographic consistency as a hard physical constraint, allowing mass conservation to be violated; and (iii) they rely on empirical projections that lack physical interpretability and mechanistic feedback.

In this work, instead of prescribing explicit functional forms for fertility and mortality, both are parameterized using Bayesian neural networks and, within the Bayesian inverse problem framework [9], embedded inside the exact transport PDE so that every posterior sample automatically satisfies demographic mass conservation. This formulation unifies historical reconstruction, forecasting, and uncertainty quantification in a single framework: Missing data are recovered through Bayesian inference rather than interpolation, and future projections follow from propagating posterior samples through the transport dynamics. The posterior over neural network weights is sampled using the No-U-Turn Sampler (NUTS) [10,11], a Hamiltonian Monte Carlo method previously applied by the present author to Bayesian inference of atomic masses [12]. While several recent physics-inspired models have used nonlinear dynamics to predict global

*Contact author: chongq@kth.se

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by Bibsam.

population trajectories [13–15], the focus of this work is on the detailed, age-structured density distribution $\rho(a,t)$, which is the critical state variable governing future labor capacity, dependency dynamics, and structural stability. This approach is applied to three East Asian economies, China, Japan, and South Korea (the CJK triad), which currently have the lowest fertility rates among major economies and represent extreme cases of demographic contraction and labor support stress.

II. METHODOLOGY

The foundation of demographic modeling rests on exact conservation laws. These dynamics were originally formulated as a continuous age-structured transport PDE [1], which was later generalized across cellular and ecological contexts [2]. In the statistical physics of society and evolutionary biology, demographic fluctuations and population dynamics have also been modeled using macroscopic advection-diffusion equations [16,17] as well as agent-based or Monte Carlo simulations, most notably the Penna bit-string model of aging and mutation accumulation [18]. This approach reflects the long tradition of applying physical and mathematical methods [19] to demographic dynamics. The present model treats these transport equations as exact physical constraints (constituting the "known physics" of the system) rather than treating population growth as an unconstrained regression or generic time-series forecasting problem.

A. Age-structured transport dynamics

Let $\rho(a,t)$ denote the age distribution density at age a and year t. The governing transport PDE reads

$$\partial_t \rho + \partial_a \rho = -\mu(a,t)\rho, \tag{1}$$

where $\mu(a,t)$ is the hazard rate of mortality.

In discrete annual form, with ages $a \in \{0, 1, \ldots, A_{\max}\}$ and years t, the dynamics become the following:

$$\rho(a+1, t+1) = s(a,t)\rho(a,t), \tag{2}$$

where $s(a,t) = \exp(-\int_a^{a+1} \mu(a',t)da') \in [0,1]$ is the annual survival probability. At the terminal age boundary $A_{\max}$ (here set to 100), the surviving cohorts are aggregated:

$$\begin{aligned}\rho(A_{\max}, t+1) = {} & s(A_{\max}-1, t)\rho(A_{\max}-1, t) \\ & + s(A_{\max}, t)\rho(A_{\max}, t).\end{aligned} \tag{3}$$

The birth boundary condition closes the system at age 0:

$$\rho(0, t+1) = B(t), \tag{4}$$

where the annual births $B(t)$ are defined through the fertility kernel $f(a,t)$:

$$B(t) = \int_0^{A_{\max}} f(a,t)\rho(a,t)da. \tag{5}$$

To make the Markov-chain Monte Carlo sampler computationally feasible over long time horizons, the discrete transport equations are formulated using JAX's loop construct `jax.lax.scan`. The step function maps the state ρ_t and time index t to the next state ρ_{t+1}:

$$\rho_{t+1}, (\rho_t, B_t, D_t) = \text{TransportStep}(\rho_t, t). \tag{6}$$

Compiling this sequential transport into a single XLA loop prevents graph unrolling, which dramatically accelerates gradient evaluation and reduces sampling time.

B. Bayesian inverse formulation and priors

The present approach builds on the mathematical theory of Bayesian inverse problems in PDEs [9,20] and dynamical Bayesian inference [21], treating PDE-constrained parameters as latent random fields with smooth priors. This also draws inspiration from recent advances in deep neural operators for PDEs [22]. Exact cohort transport is used to reconstruct the latent age-structured density $\rho(a,t)$ from sparse, noisy total population and birth and death statistics. Unlike inferring static spatial coefficients, latent, time-varying constitutive laws that govern boundary fluxes and sinks are inferred, using macroscopic transport principles to analyze real-world census and vital registration data.

The state trajectory is defined as $\rho = \{\rho(a,t)\}_{a,t}$. Let $\mathbf{Y}$ denote the collection of heterogeneous demographic observations. The posterior distribution over the Bayesian neural networks (BNN) weights $\mathbf{W} = \{\mathbf{W}_s, \mathbf{W}_f\}$ and the initial population profile $\rho_0(a)$ is as follows:

$$P(\mathbf{W}, \rho_0|\mathbf{Y}) \propto P(\mathbf{Y}|\mathbf{W}, \rho_0)P(\mathbf{W})P(\rho_0). \tag{7}$$

The likelihood $P(\mathbf{Y}|\mathbf{W}, \rho_0)$ is constructed from observation operators matching the specific data available for each country.

Normal priors are placed on all BNN weights and biases. To prevent overfitting and ensure smooth, physically realistic schedules, the weight priors have variance scaling inversely with the input layer dimension, $W_{ij} \sim \mathcal{N}(0, 1/n_{\text{in}})$, and standard normal priors on the biases $b_i \sim \mathcal{N}(0,1)$. These priors act as an L_2 regularization (weight decay) on the neural network parameters, restricting the BNN to learning smooth, slowly varying schedules over the age and year coordinates. The implied prior over the demographic trajectories $\rho(a,t)$ is therefore a smooth, continuous field that respects cohort advection.

To verify the framework's robustness, sensitivity checks were performed under alternative prior strengths (scaling the prior standard deviation by 0.5 and 2.0) and under various hidden layer configurations. The resulting projections and nonequilibrium metrics remained stable. This confirms that the physical transport constraints dominate the posterior inference, rendering the core findings insensitive to the specific choice of neural architecture or hyperparameters.

C. Neural parametrization of constitutive laws

This work is related to recent attempts to apply deep learning to demographic forecasting [23,24] and social dynamics [25–30]. However, many approaches either treat the system as a purely data-driven black box or rely on unconstrained agent-based simulations. In this hybrid design, the transport dynamics are strictly fixed, and the neural networks are used solely to parametrize the unknown constitutive laws. This separates the universal physical transport dynamics from system-specific biological and social behaviors. It also represents a fully nonparametric generalization of linear

factorizations [5], as it can capture nonlinear age-time interactions natively.

The fertility kernel $f(a,t)$ is assumed to be governed by a parametric distribution representing reproductive behaviors, scaled by the TFR at a given time t (TFR_t):

$$f(a,t) = 0.5\,\mathrm{TFR}_t\, g(a;\nu_t,\sigma_t), \tag{8}$$

where $g(a;\nu_t,\sigma_t)$ is a Gaussian density function representing childbearing age preferences, masked to vanish outside reproductive ages ($15 \leqslant a \leqslant 49$) and normalized. The latent parameters $\theta_f(t) = (\ln \mathrm{TFR}_t, \mathrm{logit}(\nu_t), \mathrm{logit}(\sigma_t))$ are parameterized as outputs of a multilayer perceptron (MLP) mapping normalized year t to the reproductive space:

$$\theta_f(t) = \mathrm{MLP}_f(t;\mathbf{W}_f,\mathbf{b}_f). \tag{9}$$

The neural network MLP_f comprises two hidden layers, each containing 16 units and using ReLU activation functions.

When complete annual survival grids are observed, as for Japan and South Korea, the survival probability $s(a,t)$ is directly modeled via a BNN mapping normalized age and year to survival logits:

$$s(a,t) = \sigma(\mathrm{MLP}_s(a,t;\mathbf{W}_s,\mathbf{b}_s)), \tag{10}$$

where $\sigma(\cdot)$ is the sigmoid function and MLP_s has an architecture of [2, 16, 16, 1]. Conversely, when complete annual survival rates are unobserved, as is the case for China, this direct parametrization leads to numerical instability. Instead, survival is modeled as a perturbation around a standard baseline hazard function, $\mu_0(a) = 0.005 + 8\times10^{-5}e^{0.082a}$:

$$s(a,t) = \exp(-\mu_0(a)\,\lambda(a,t)), \tag{11}$$

where the mortality multiplier $\lambda(a,t) = \exp(\mathrm{MLP}_s(a,t;\mathbf{W}_s,\mathbf{b}_s))$ is parameterized by the survival BNN.

This asymmetry is due to data availability: Complete annual survival grids simply do not exist for China over 1950–2024, and directly parametrizing $s(a,t)$ as a free BNN output destabilizes the sampler when data are this sparse. Anchoring to this hazard function provides a numerically stabilizing reparameterization rather than a restrictive structural prior: The multiplier $\lambda(a,t)$ retains the full flexibility of the BNN to depart from $\mu_0(a)$. For instance, the reconstructed annual survival probability at age 95 falls from $\approx$0.39 in 1950 to $\approx$0.32 during the 1959–1961 famine before recovering to $\approx$0.78 by 2024, a range the fixed baseline alone could not produce. This parametrization imposes a smoother effective prior on $s(a,t)$ than the direct sigmoid link used for Japan and South Korea (whose weight priors are themselves tightened relative to Sec. II B, for the numerical-stability reasons detailed in Sec. III C, though the link function remains the free sigmoid rather than an anchored one). Absolute cross-country comparisons of derived quantities (e.g., the magnitude of S_{total} or D_{KL} introduced in Sec. IV) should be interpreted with this asymmetry in mind. However, the qualitative conclusions of this work, including the timing of structural transitions and the relative ordering of the three countries' demographic pressure, are determined primarily by the observed vital data and the exact transport constraint rather than by this prior choice and are therefore robust to it.

D. Posterior inference

To sample from the highly nonlinear and high-dimensional posterior distribution $P(\mathbf{W},\rho_0|\mathbf{Y})$, NUTS is employed, an advanced variant of Hamiltonian Monte Carlo (HMC). HMC uses the gradients of the log-posterior density to generate proposal states, mapping the parameter space to a physical system where the parameters act as coordinates and auxiliary variables act as momentum. By simulating Hamiltonian dynamics using a leapfrog integrator, HMC can make distant proposals with high acceptance probabilities, bypassing the random-walk behavior of traditional Metropolis-Hastings algorithms. The sampler automates the tuning of HMC parameters by adaptively building a recursive search tree of leapfrog steps, stopping automatically when the trajectory begins to double back on itself (the "no-U-turn" condition). This prevents redundant calculations and ensures efficient exploration of the posterior.

In this work, the framework is implemented across all three empirical settings using JAX for high-performance compiler acceleration and autodifferentiation, running the NUTS sampler via NumPyro. For each setting, the posterior is sampled by running four parallel chains, each performing 1000 warmup iterations to adapt the step size and mass matrix, followed by 1000 sampling iterations, yielding a total of 4000 posterior samples for uncertainty quantification.

The architecture and implementation details of both networks are summarized as follows: MLP_f ([1,16,16,3]) and MLP_s ([2,16,16,1]) each use ReLU activations on their two hidden layers and a linear (identity) output layer; the physical link functions described above [$\exp(\cdot)$, the logistic-scaled outputs for ν_t and σ_t, and the sigmoid or baseline-hazard multiplier for $s(a,t)$] are applied outside the network to these raw linear outputs. The inputs are normalized to the unit interval with age as $a/A_{\max}$ and year as $(t-t_{\mathrm{start}})/(t_{\mathrm{end}}-t_{\mathrm{start}})$. MLP_f is assigned the weight and bias priors of Sec. II B unmodified in all three countries. For MLP_s in Japan and South Korea, where $s(a,t)$ uses the direct sigmoid link of Eq. (10), scaling down the weight and bias prior standard deviations by a factor of 0.10 (Japan) and 0.15 (South Korea) relative to Sec. II B was found necessary. China's anchored parametrization [Eq. (11)] uses the unmodified prior throughout. The sampler is initialized at the prior median (`init_to_median`) rather than at a random draw, which avoids the population collapsing to zero during the earliest warmup steps.

E. Generalization and future extensions

The Bayesian inverse transport framework is general and can be extended in several directions. First, it can be applied to other countries or regional systems by assimilating localized vital statistics and census datasets. Second, the transport equation can be augmented with a spatial advection or diffusion term to model migration-augmented transport:

$$\partial_t\rho + \partial_a\rho = -\mu(a,t)\rho + M(a,t), \tag{12}$$

where $M(a,t)$ is the net migration flux. Third, the latent constitutive laws can be coupled to other socioeconomic variables (such as GDP, housing price index, or female labor participation) by feeding them as inputs to the BNNs, enabling a fully closed-loop sociophysics modeling of demographic feedback.

III. EMPIRICAL CASE STUDIES AND MODEL DIAGNOSTICS

The choice of the CJK triad as the empirical case studies is motivated by their roles as the preeminent industrial powerhouses and "world factories" of the modern era, all currently experiencing a rapid collapse in fertility. Rather than aiming to artificially restore birth rates to the classical replacement level of 2.1, an objective that may no longer be realistic or structurally meaningful, the interest here lies in characterizing the nonequilibrium age structures before these systems stabilize. From a physical and economic perspective, the primary concern is the potential support limits (the ratio of active labor cohorts to elderly dependents). While automation, robotics, and artificial intelligence may reduce the absolute labor demand for social maintenance, and indeed shrinking workforces might even coincide with economic growth booms [31], understanding the transient age structure remains essential for managing the transition.

A. Logistic transition rates and demographic pressure

To extract the macroscopic parameters governing the collapse of the TFR boundary condition, the long-term empirical fertility transition is modeled using a generalized logistic decay,

$$B(t) = B_{\min} + \frac{B_{\max} - B_{\min}}{1 + e^{k(t-t_0)}}, \tag{13}$$

where $B_{\max}$ and $B_{\min}$ represent the pretransition (natural) and post-transition (modern) asymptotic fertility bounds, respectively, t_0 is the transition midpoint, and k is the maximum transition rate.

In this formulation, the decay rate k is a quantitative metric for the severity of demographic pressure (e.g., high urbanization pressure, high housing costs, and intense competition for resources) driving the fertility transition. A higher value of k indicates a more rapid, steplike collapse in fertility, locking the system into a demographic transition over a very short time window.

By fitting the transition rate k to historical data, the demographic transitions of China, Japan, and South Korea can be empirically classified. South Korea ($k \approx 0.35$) and China ($k \approx 0.28$) exhibit extremely rapid decay behavior, indicating intense demographic pressure that compresses the transition into a few decades. In contrast, Japan ($k \approx 0.12$) represents a more gradual relaxation. This macroscopic parametrization provides a useful tool to classify demographic crises by their transition rates and forecast the resulting structural imbalances.

B. Historical population reconstructions

1. Japan (1970–2024)

Japan's dataset from the Human Mortality Database (HMD) [32] provides complete annual population grids $Y_{\text{pop}}(a,t)$, births $Y_B(t)$, and deaths $Y_D(a,t)$. The likelihood is constructed using log-normal distributions:

$$\ln Y_{\text{pop}}(a,t) \sim \mathcal{N}\big(\ln \rho(a,t), \sigma_p^2\big), \tag{14}$$

with HalfNormal priors on observation noise parameters.

As shown in Fig. 1, the model successfully reconstructs the historical macroscopic vital events and age density profiles. Furthermore, the model infers a decline in TFR from ≈2.1 in 1970 to ≈1.15 in 2024, accompanied by a structural shift in the mean childbearing age from 26.2 to 31.9 years.

2. South Korea (2003–2024)

South Korea's HMD dataset [32] is aligned from 2003 to 2024. The training setup matches Japan's complete grid likelihood.

Figure 2 demonstrates the model's ability to accurately infer South Korea's macroscopic vital events and age density profiles directly from the inversion of the age-structure PDE. The model also infers a world-record drop in TFR from ≈1.18 in 2003 to ≈0.67 in 2024.

3. China (1950–2024): Reconstructing incomplete data

China represents a highly challenging inverse problem. No annual age-structured population grid is available. Only annual births, annual deaths, and total population size are observed, aggregated from multiple conflicting sources (specifically, manually compiled records from Chinese vital statistics and demographic columns [33–38]), along with coarse five-year census age brackets for 2000, 2010, and 2020.

The initial state $\rho(a, 1950)$ is parameterized as a latent stable pyramid:

$$\rho(a, 1950) = N_{\text{scale}} \frac{e^{-ra} S_{\text{base}}(a)}{\sum_{a'} e^{-ra'} S_{\text{base}}(a')}, \tag{15}$$

where the initial growth rate r and scale N_{scale} are sampled. Here $S_{\text{base}}(a) = \exp(-\sum_{a'=0}^{a-1} \mu_0(a'))$ is the cumulative survivorship to age a implied by the same baseline hazard $\mu_0(a) = 0.005 + 8 \times 10^{-5} e^{0.082a}$ used in Eq. (11) above. It supplies a demographically realistic, numerically stable shape for the unobserved 1950 age pyramid and introduces no additional free parameters. To resolve numerical ill conditioning caused by the population scale (10^9) relative to the BNN parameters (1.0), $N_{\text{scale}} \sim \mathcal{N}(5.52, 0.27)$ is sampled and scaled up by 10^8. To prevent vanishing gradients during chain initialization (where a default 0.5 survival rate causes the population to shrink to zero immediately, zeroing out later gradients), the baseline survival constraint is applied and the sampler is initialized using the prior medians (`init_to_median`).

The coarse census age groups $C_g(t)$ for years $t \in \{2000, 2010, 2020\}$ are matched to the model cohorts by aggregating the corresponding one-year bins (excluding age 0):

$$C_{1\text{-}4}(t) = \sum_{a=1}^{4} \rho(a,t), \quad C_{5\text{-}9}(t) = \sum_{a=5}^{9} \rho(a,t), \quad \ldots. \tag{16}$$

This allows the model to be evaluated against independent census counts, accounting for adjustments for infant and child undercounting [39,40], as well as broader evaluations of Chinese demographic forecasts [41], under a log-normal likelihood with a 5% coefficient of variation.

Despite the absence of comprehensive annual age data, the model successfully reconstructs the complete historical age-structured grid $\rho(a,t)$ over 75 years. As illustrated in Fig. 3, the model accurately captures macroscopic vital events. In

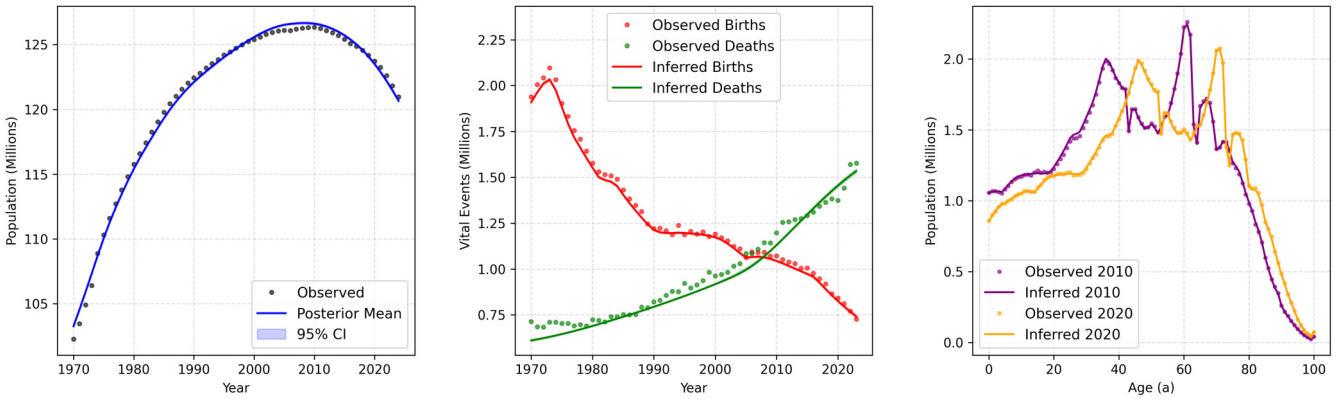


FIG. 1. Validation of the model for Japan. The left panel compares the inferred total population (blue, with shaded 95% credible interval) against observed data (black dots). The middle panel plots the macroscopic vital events (annual births in red, deaths in green, both with 95% credible intervals). Because the total births and deaths were not explicitly provided in the likelihood function (which relies solely on the cross-sectional age grid), their accurate reconstruction confirms that the transport PDE conserves mass while advecting the population cohort backward in time. The right panel shows the structural fit of the continuous age density to empirical 1-year age buckets for the years 2010 and 2020. Credible intervals are narrow and indistinguishable from the mean at this scale.

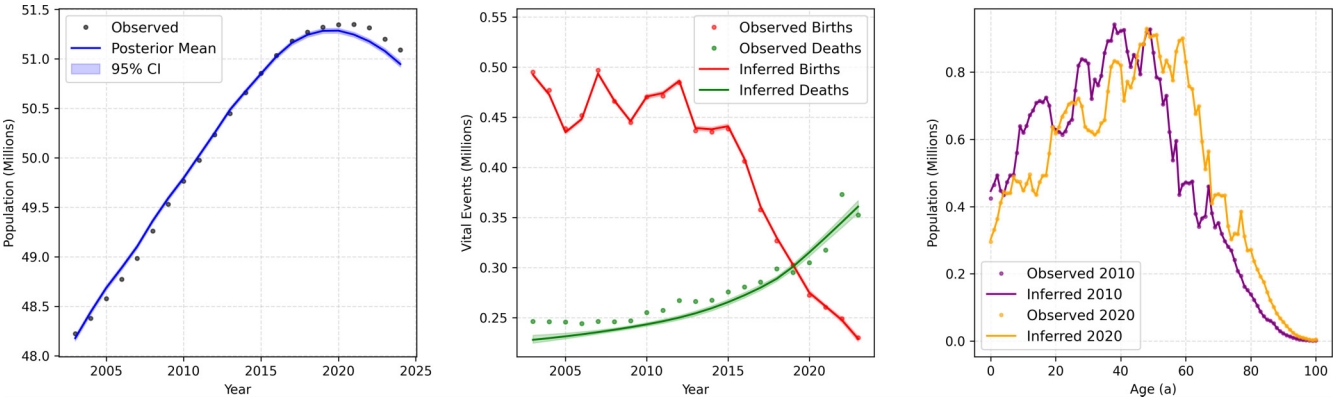


FIG. 2. Validation of the model for South Korea. The left and middle panels show the model's ability to infer macroscopic total population and vital events (births and deaths, shown with 95% credible intervals) directly from the inversion of the age-structure PDE without explicit observation of those totals. The right panel shows the tight alignment between the continuous age density and the empirical data for 2010 and 2020. Credible intervals are narrow and indistinguishable from the mean at this scale.

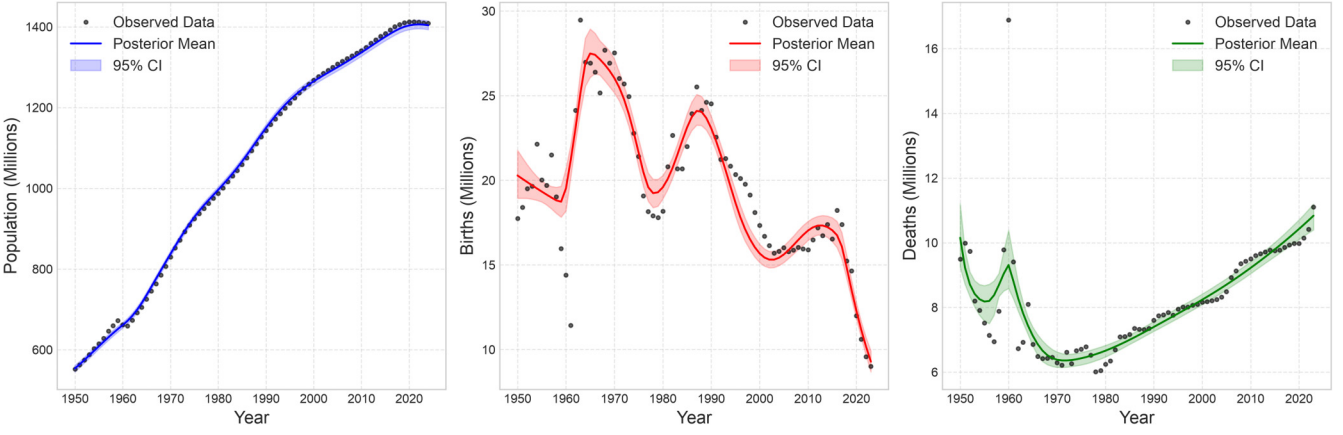


FIG. 3. Posterior predictive check for China's macroscopic vital events. The model's inferred total population, annual births, and annual deaths (solid blue, red, and green lines with 95% credible intervals) are plotted against historical observed data (black circles) derived from national statistics. The continuous inverse transport dynamics accurately capture massive historical shocks, such as the sharp demographic fluctuations around 1960.

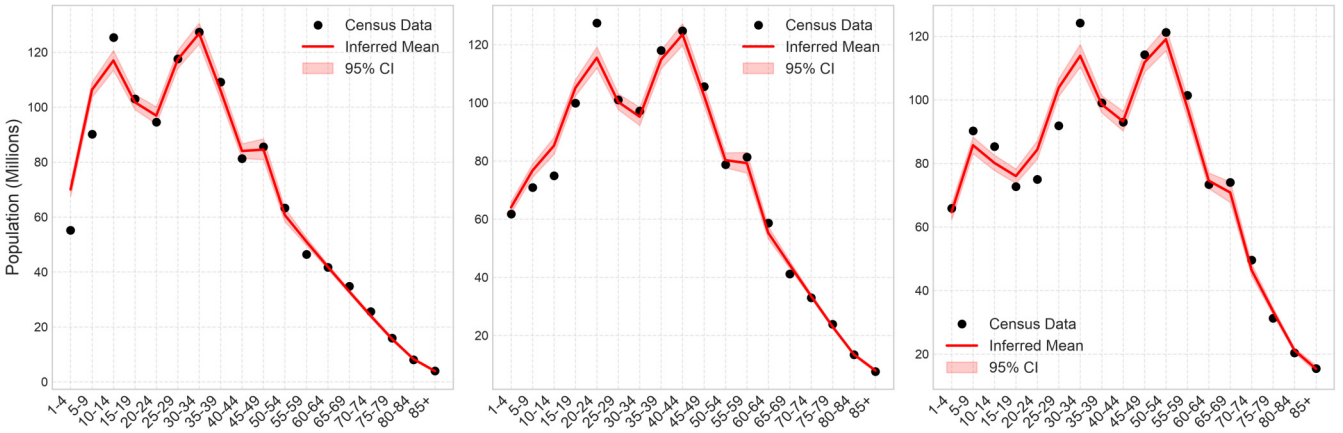


FIG. 4. Historical backtesting of China's age structure. The model's reconstructed continuous age density aggregated into standard 5-year brackets (red lines with 95% credible intervals) is overlaid directly on the raw empirical census data (black circles) for the years 2000, 2010, and 2020. The alignment confirms that the model accurately captures complex cohort advection over decades.

addition, the reconstructed age distributions match the independent census observations with high precision, successfully capturing the major historical cohort bulges and troughs (Fig. 4).

C. Quantitative evaluation and diagnostics

To establish the statistical validity of the proposed method, chain convergence is verified quantitatively using the split-$\hat{R}$ statistic and the effective sample size (ESS) computed by `numpyro.diagnostics` for all raw NUTS-sampled latent variables (BNN weights and biases, and any per-country noise or initial-state parameters) across the four sampled chains, recovering the native 4×1000 chain layout from the stored posterior draws. Table I reports, for each country, the worst-case (maximum) $\hat{R}$ and worst-case (minimum) ESS both over the raw latent weights and over the physically reported constitutive quantities $[\mathrm{TFR}_t, \nu_t, \sigma_t, s(a,t)]$.

China's chains mixed acceptably from the outset ($\hat{R} \lesssim 1.5$). Japan and South Korea initially showed poor raw mixing ($\hat{R}$ up to several hundred): Each chain explored only a narrow region of weight space, and different chains settled in different regions, most visibly for the childbearing-age width σ_t. This was traced to the direct sigmoid link used for $s(a,t)$ in these two countries [Eq. (10)], which drove 35% (Japan) and 37% (South Korea) of the age-time survival grid into the saturated tails of the sigmoid, compared to only 4% for China's anchored parametrization [Eq. (11)]. This saturation creates flat, ridgelike regions in the posterior over BNN weights that a fixed-length (1000-iteration) NUTS warmup cannot adapt to.

This was resolved by scaling down the weight and bias prior standard deviations of MLP_s by a factor of 0.10 (Japan) and 0.15 (South Korea), with NUTS run under otherwise identical settings. The physically meaningful quantities mix well for all three countries ($\hat{R} \leqslant 2.3$, comparable across countries). The raw-weight $\hat{R}$ remains higher (19–23 for Japan and South Korea, 3.2 for China) because the ordinary permutation and sign symmetries of the hidden layer make individual weights nonidentifiable even in a well-behaved posterior.

Compared to standard UN World Population Prospects and traditional time-series forecasts, the present PDE-constrained inverse framework differs in two fundamental aspects. First, traditional models project cohort trajectories as independent statistical trends, so forecast uncertainty is uncorrelated across ages. In contrast, the present framework enforces exact cohort transport, which guarantees that uncertainty in fertility at year t propagates downstream along the characteristics of the advection equation, preserving the structural cross-cohort correlations over time. Second, while standard models can project unconstrained mortality schedules that violate biological consistency (such as crossing cohort profiles), the BNN-parameterized constitutive laws are regularized by both the baseline survival constraints and the transport dynamics, yielding physically consistent age profiles even under sparse observations.

TABLE I. Convergence diagnostics: worst-case split-$\hat{R}$ and minimum ESS (of 4000 total draws) over the raw BNN weights or biases and over the physically reported constitutive quantities $[\mathrm{TFR}_t, \nu_t, \sigma_t, s(a,t)]$.

	Raw BNN weights		Physical quantities	
Country	Max $\hat{R}$	Min ESS	Max $\hat{R}$	Min ESS
China	3.2	2.3	1.5	4.0
Japan	23.3	2.0	2.3	2.5
S. Korea	19.7	2.0	2.1	2.4

IV. STRUCTURAL DYNAMICS AND COHORT FORECASTS

Having established the model's ability to reconstruct historical demographic states, the inferred constitutive laws are now applied to project future structural dynamics. In this section, a simplified mathematical intuition is first developed for how fertility crashes propagate through age structures over time. The model's predictive power is then validated against historical TFR data, medium-horizon cohort shifts are projected, and the resulting long-term demographic pyramids are compared.

A. A simplified population model for a substantially reduced TFR

To build intuition for the structural transitions projected by the continuous transport model, a simplified, discrete 80-cohort model is considered. Suppose each individual survives exactly from age 0 to 79 and produces one child at age 25 (i.e., replacement fertility of two children per couple). From time T_0 to $T_0 + D$, fertility is halved to one child per couple, after which it returns to replacement.

Because the birth sequence outside the dip is exactly 26-periodic, the total population eventually converges to a permanently lower plateau. The loss is effectively a discretized exponential, with the population decaying as

$$P(D) \approx P_0\, 2^{-D/26} \approx P_0(1 - 0.02D) \quad (\text{for } D \lesssim 26). \tag{17}$$

This indicates a demographic half-life of 26 years: A one-child generation lasting 26 years halves the total population. In turn, the deficit created during the dip echoes every 26 years. The population reaches its new steady state roughly two echo cycles ($2 \times 26 = 52$ years) after the dip ends, yielding a settling time of $T_{\text{settle}}(D) \approx D + 50$.

While the total population declines smoothly, the dependency ratio experiences a delayed, transient crisis. Defining the working-age population (W) as ages 15–64 and the elderly (E) as ages 65–79, the baseline ratio is $W/E = 50/15 \approx 3.33$. When the fertility dip begins, W/E remains stable for 15 years. The ratio then falls steadily as the undersized cohorts enter the workforce while the elderly band remains full size. The crisis point (the minimum of W/E) occurs roughly 65 years after the gap starts ($t_{\min} \approx T_0 + 65$), when the first undersized cohort turns 65 and leaves the workforce. The depth of this crisis worsens with the duration of the dip, flattening near $W/E \approx 1.27$ for $D \gtrsim 50$, because the entire 50-year working-age band becomes depleted.

Ultimately, this toy model demonstrates that while population loss is an immediate and permanent consequence of a fertility shock, the dependency ratio crisis is a delayed effect that peaks roughly two generations (≈65 years) later.

B. Validation via TFR

To benchmark the posterior predictive power of the model, the model's internally inferred TFR is compared with actual empirical records from the World Bank. Historical TFR data are not included in the training of the model; the inverse framework only observes aggregated age-structure census constraints. That the transport dynamics correctly deduce the historical fertility trajectories, including the exact timing and depth of South Korea's fertility crash and China's subreplacement shift, validates the structural consistency of the model (Fig. 5).

C. Medium-horizon cohort forecasts

A central application of the proposed framework is the prediction of age composition over the next one to two generations. Because the model evolves the full age density $\rho(a, t)$ through an exact transport equation, future cohort structure is strongly constrained by the currently observed age pyramid and by the inferred fertility and mortality laws. This makes

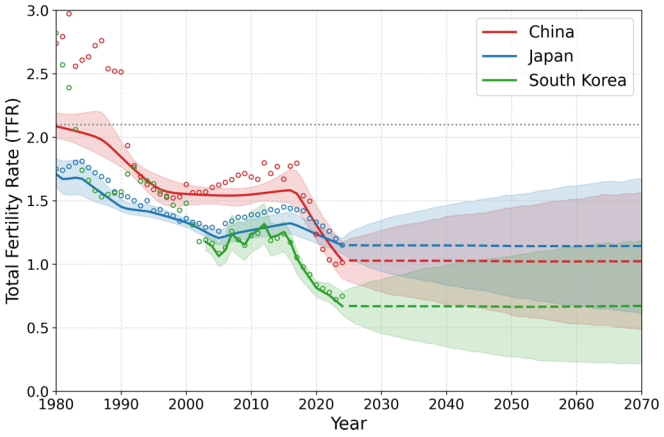


FIG. 5. Model-inferred historical (solid lines) and projected (dashed lines) TFR (with 95% credible intervals) for China, Japan, and South Korea, overlaid with external empirical data points (open circles) from the World Bank (SP.DYN.TFRT.IN) [42]. Note that TFR was not provided as a training input to the model.

forecasts of child, working-age, and elderly shares more structurally reliable than unconstrained long-range projections based only on aggregate indicators. Rather than treating population growth as a single scalar trajectory, this framework resolves the full demographic pipeline and can therefore predict the future age composition.

Using the posterior samples, the demographic trajectories are propagated to 2070 under the Status Quo random walk scenario. Figure 6 shows the comparative age structures for all three countries in 2050 (while the kinematic heatmaps extend to 2070). China and South Korea exhibit highly inverted pyramids with a hollow base, representing decades of subreplacement fertility, while Japan shows a stable, aged pillar structure.

The 25- to 50-year horizon is especially important because it lies between near-term policy response and full generational turnover. Because new births enter the labor force only after a delay of roughly 20–25 years, today's fertility regime already determines a substantial part of the future workforce: The children who will be working-age adults in 2050 are being born now or have already been born. The transport dynamics encode this structural inertia exactly, giving the model strong predictive power over the age structure even when long-run uncertainty in vital rates grows.

Table II and Fig. 7 present the medium-horizon cohort forecasts for all three countries under the Status Quo scenario. Here the elderly dependency ratio (EDR) is introduced, representing the ratio of the elderly population (65+) to the working-age population (15–64):

$$\text{EDR}(t) = \frac{\sum_{a=65}^{A_{\max}} \rho(a,t)}{\sum_{a=15}^{64} \rho(a,t)}, \tag{18}$$

and the potential support ratio (PSR), representing the number of working-age adults per elderly person:

$$\text{PSR}(t) = \frac{\sum_{a=15}^{64} \rho(a,t)}{\sum_{a=65}^{A_{\max}} \rho(a,t)}. \tag{19}$$

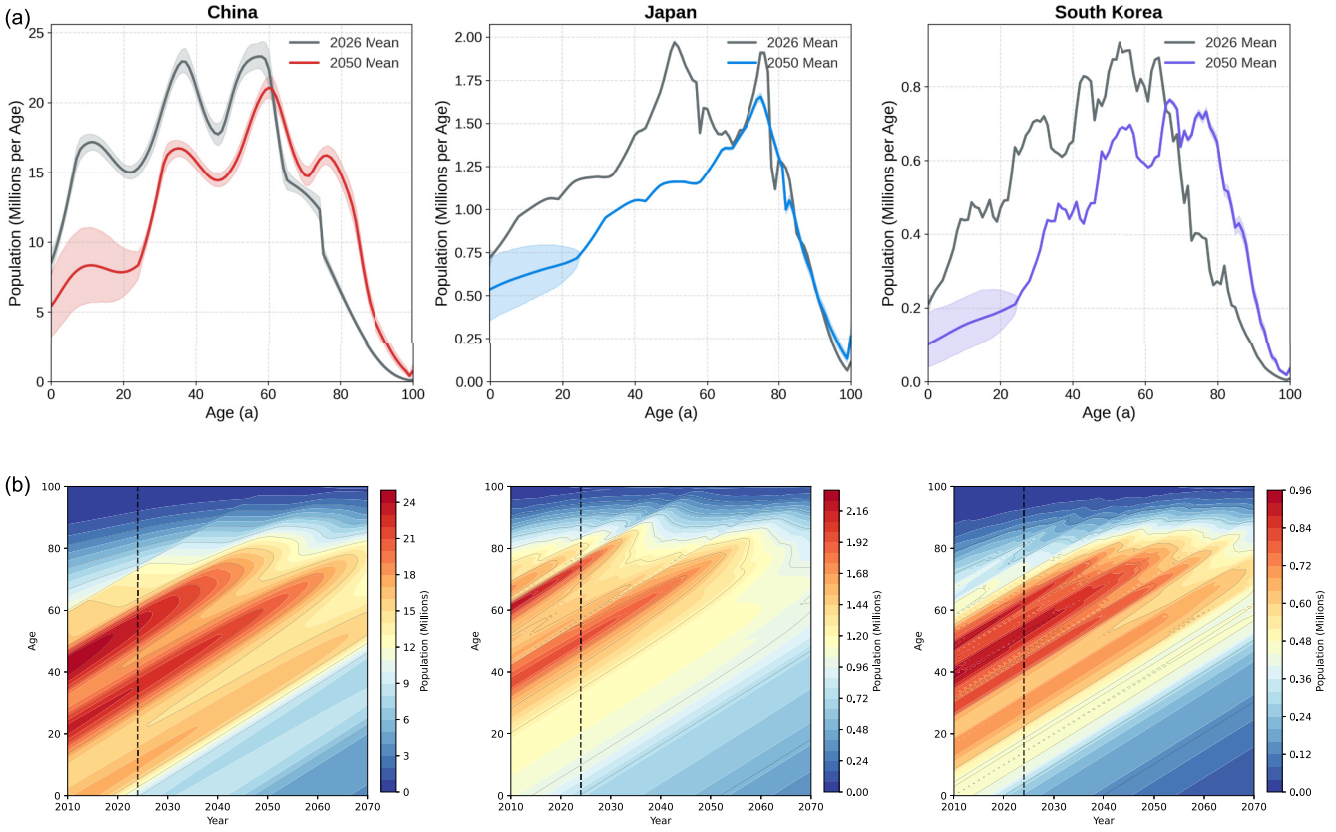


FIG. 6. (a) Comparative 2050 population pyramids (mean and 95% credible intervals) for China (red), Japan (blue), and South Korea (purple) under the Status Quo scenario. (b) Kinematic Density Wave heatmaps. The contour shading traces the evolution of the population age structure $\rho(a,t)$ over time. Bright red regions indicate large cohort masses (population bulges) propagating diagonally as they age. The dashed black line separates the structurally inferred historical dynamics from the Status Quo forecast.

TABLE II. Medium-horizon cohort forecasts (mean ± SD from posterior samples) under the Status Quo scenario. Population in millions; shares in percentage of total.

Country	Year	Pop. (M)	0-14 (%)	15-64 (%)	65+ (%)	75+ (%)	EDR	PSR
China	2026	1398 ± 6	14.7 ± 0.2	68.9 ± 0.2	16.4 ± 0.2	6.7 ± 0.2	0.24 ± 0.00	4.20 ± 0.07
	2030	1381 ± 7	12.3 ± 0.3	68.3 ± 0.3	19.5 ± 0.3	8.3 ± 0.2	0.29 ± 0.00	3.51 ± 0.06
	2040	1309 ± 17	9.3 ± 1.0	63.6 ± 0.8	27.1 ± 0.5	12.2 ± 0.3	0.43 ± 0.01	2.34 ± 0.04
	2050	1186 ± 27	9.1 ± 1.3	60.4 ± 0.9	30.5 ± 0.8	17.2 ± 0.6	0.51 ± 0.01	1.98 ± 0.05
	2060	1020 ± 38	7.0 ± 1.5	57.1 ± 0.7	35.9 ± 1.4	18.6 ± 0.9	0.63 ± 0.03	1.59 ± 0.07
	2070	855 ± 51	7.0 ± 2.1	55.5 ± 0.9	37.5 ± 2.3	22.5 ± 1.5	0.68 ± 0.05	1.49 ± 0.10
Japan	2026	119 ± 0	11.1 ± 0.0	58.8 ± 0.0	30.1 ± 0.0	17.8 ± 0.0	0.51 ± 0.00	1.96 ± 0.00
	2030	115 ± 0	10.3 ± 0.2	58.7 ± 0.1	30.9 ± 0.1	18.8 ± 0.1	0.53 ± 0.00	1.90 ± 0.01
	2040	104 ± 1	9.5 ± 0.8	55.6 ± 0.5	34.9 ± 0.3	19.2 ± 0.2	0.63 ± 0.00	1.59 ± 0.01
	2050	92 ± 2	9.5 ± 1.2	52.8 ± 0.6	37.7 ± 0.7	22.4 ± 0.5	0.72 ± 0.01	1.40 ± 0.02
	2060	80 ± 3	9.1 ± 1.4	52.6 ± 0.6	38.3 ± 1.2	24.5 ± 0.8	0.73 ± 0.02	1.38 ± 0.04
	2070	68 ± 3	8.7 ± 1.8	52.6 ± 0.7	38.7 ± 2.0	23.7 ± 1.3	0.74 ± 0.04	1.36 ± 0.08
S. Korea	2026	51 ± 0	10.1 ± 0.1	68.8 ± 0.0	21.1 ± 0.0	8.5 ± 0.0	0.31 ± 0.00	3.26 ± 0.00
	2030	50 ± 0	8.1 ± 0.2	66.9 ± 0.2	25.0 ± 0.1	10.0 ± 0.0	0.37 ± 0.00	2.68 ± 0.01
	2040	46 ± 0	5.8 ± 0.9	59.8 ± 0.6	34.4 ± 0.3	17.1 ± 0.2	0.57 ± 0.00	1.74 ± 0.01
	2050	40 ± 1	5.0 ± 1.2	54.2 ± 0.5	40.8 ± 0.8	23.5 ± 0.5	0.75 ± 0.01	1.33 ± 0.02
	2060	32 ± 1	4.0 ± 1.4	50.8 ± 0.5	45.2 ± 1.5	27.1 ± 0.9	0.89 ± 0.03	1.12 ± 0.04
	2070	25 ± 1	3.5 ± 1.5	45.7 ± 1.3	50.9 ± 2.5	30.1 ± 1.6	1.12 ± 0.09	0.90 ± 0.07

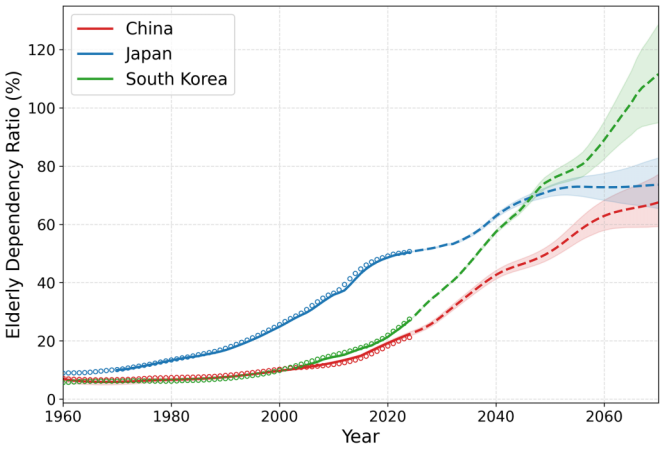


FIG. 7. EDR for China, Japan, and South Korea. Solid lines (with 95% CI shading) show the model's structurally inferred historical EDR, aligning with independent World Bank data points (circles). Dashed lines project the future growth of the dependency burden under the Status Quo scenario as the rapid fertility collapse permanently inverts the demographic pyramids.

The most important demographic changes over the next quarter-century are not simply declines in total population, but pronounced shifts in age composition.

China's working-age population (15–64) peaked around 2015 and is projected to contract by 249 ± 21 million (-25.8%) between 2024 and 2050, representing a loss roughly equal to the entire population of Indonesia. The child share (0–14) falls from 15.9% to 9.1%, while the elderly share (65+) doubles from 15.4% to 30.5%. Japan, which has been aging since the 1990s, sees its working-age share decline from 58.8% to 52.8%, a loss of 22.3 ± 0.5 million (-31.4%). South Korea undergoes the most dramatic structural transformation: Its working-age population contracts by 14.1 ± 0.3 million (-39.6%), the child share collapses from 11.0% to 5.0%, and the elderly share more than doubles from 18.9% to 40.8%.

To place these projections in a global context, typical advanced economies maintain significantly lower and more stable dependency ratios. As of 2024, the EDR is approximately 28.4% in the United States and 34.5% in the European Union. While these Western regions are also aging, their demographic transitions are buffered by steady immigration and less extreme historical fertility declines. In stark contrast, the CJK triad faces an unbuffered structural inversion. With South Korea's EDR projected to skyrocket past 80% and Japan's approaching 72% by 2070, the macroeconomic burden placed on their active labor forces will vastly exceed the aging challenges currently anticipated in the West.

The posterior samples allow the timing and uncertainty of key structural milestones to be estimated. All three countries have already undergone or are currently undergoing *cohort inversion*, the point at which the elderly (65+) outnumber children (0–14). Japan crossed this threshold in 1998, South Korea in 2018, and China is crossing it now (2024–2026, 95% CI). The PSR provides a direct measure of this labor-force burden. Japan's PSR has already fallen below 2.0 in 2023–2024. South Korea's PSR is projected to breach 2.0 by 2037

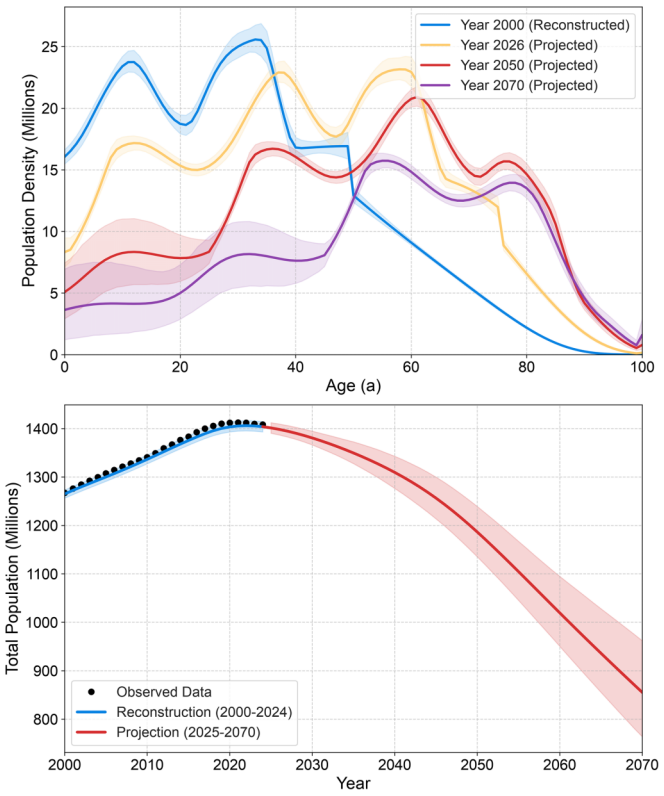


FIG. 8. Reconstructed and projected age-structured population density $\rho(a, t)$ for China. Comparison of Year 2000 (reconstructed historical mean), Year 2026 (short-term projection), and Year 2050 (long-term projection). Shaded areas represent the 95% credible intervals.

(tightly constrained, with all posterior samples agreeing). China's PSR approaches the 2.0 threshold by 2049 [2047–2050], though this milestone carries the widest uncertainty among the three countries, reflecting the greater epistemic uncertainty in China's historical vital statistics.

These forecasts show more than how many people there will be: They show how the population will be distributed across age groups that matter for schooling, labor supply, taxation, care needs, and retirement systems. Because the full age-time density field is resolved, one can extract credible intervals for any age-specific quantity: the size of the university-age cohort for education planning, the 25–54 prime working-age population for labor-market policy, or the 80+ population for long-term care projections. The inferred age-time field is thus a planning tool that extends well beyond conventional population forecasts.

1. Comparison with existing models and scenario projections for China

To contextualize the results, the framework is applied to predict China's future age density (Fig. 8) and compare these projections to existing independent models and institutional reports (summarized in Table III). These alternative models yield significantly different results, reflecting the diverse assumptions embedded in their structures.

TABLE III. Comparison of different demographic model projections for China in 2050. Population is in billions, and age shares or dependency ratios are in percent.

Model framework	Total Pop. (N_{2050})	Structural aging metric	Key system insight or stress point
UN WPP (2024 Revision)	1.21–1.26	$P_{65+} \approx 30\%$	Underestimates rapid fertility decline
YuWa Population (2023)	1.17–1.23	EDR $\approx 52.4\% - 54.3\%$	Severe labor constriction (EDR >50%)
Institute for Health Metrics and Evaluation (Vollset *et al*. 2020)	1.18	$P_{65+} > 32\%$, Median Age >50	Highly inverted population pyramid
China Population and Development Research Center (PADIS-INT 2023)	1.24	$P_{60+} \approx 38\% - 40\%$	Hyperaged segment peak (≈500 M)
This work (posterior)	1.186 ± 0.027	$P_{65+} = 30.5\% \pm 0.8\%$	Exact cohort-transport uncertainty

For instance, the UN World Population Prospects (WPP 2024 Revision) [34] projects China's 2050 population (N_{2050}) to be between 1.21 and 1.26 billion under its medium-variant projection, with the proportion of the population aged 65 and above (P_{65+}) reaching approximately 30%. A key methodological feature of the UN framework is its reliance on three discrete scenario variants (high, medium, and low fertility paths), constructed by applying fixed offsets to the TFR rather than using a continuous probabilistic representation of vital rate uncertainty. While this approach offers simple bounds, it does not propagate joint parameter uncertainty or capture structural cohort-transport dependencies.

Alternative models project much steeper declines. The YuWa Population Research Group (2023 Report) [43], assuming a sustained low-fertility trap (TFR ≈ 1.05), forecasts N_{2050} to be between 1.17 billion (low variant) and 1.23 billion (Medium variant). They project a 25% contraction of the working-age population (15–64) by 2050, resulting in an EDR of 52.4% to 54.3%, meaning roughly two active workers must support one elderly dependent. Similarly, the Institute for Health Metrics and Evaluation global model [44], which models fertility as a function of female education, contraceptive access, and urbanization rather than direct extrapolation, projects $N_{2050} \approx 1.18$ billion, with $P_{65+} > 32\%$ and a median age exceeding 50 years.

Institutional models focused on specific sectors highlight extreme system stresses. The Chinese Academy of Social Sciences pension model [45] projects that the pension system dependency ratio, which measures the ratio of retirees to active pension contributors, will rise exponentially to reach 96.3% by 2050. This represents a transition from a 2:1 support matrix in 2019 to an unsustainable 1:1 limit equilibrium, projecting a depletion of the enterprise pension fund by 2035. Additionally, the China Population and Development Research Center PADIS-INT model [35] projects a 2050 population of 1.24 billion but notes that the population aged 60 and over (P_{60+}) will peak near 500 million, or 38% to 40% of the total population, driven by the rapid aging of the 1980s baby-boom generation.

Compared to these models, the Bayesian inverse transport framework yields a posterior forecast of 1.186 ± 0.027 billion for China's 2050 population, with $P_{65+} = 30.5\% \pm 0.8\%$. The present model does not rely on scenario variants; instead, it directly infers latent fertility and survival laws from historical data and propagates continuous, calibrated probability distributions through the exact age-time transport dynamics, capturing both parametric and cohort-transport uncertainties.

D. Structural age-structure entropy and macroscopic Lotka divergence

To establish a rigorous, model-independent physical quantification of the structural demographic transition, three information-theoretic and structural metrics are introduced, derived directly from the inferred age density field $\rho(a,t)$.

1. Total demographic entropy and phase volume

While standard Shannon entropy $S = -\sum p(a) \ln p(a)$ quantifies the dispersion of a distribution, it is structurally symmetric: A sharply peaked young population and a sharply peaked elderly population yield identical values, ignoring the biological arrow of time. To physically distinguish these states, the concept of the *demographic phase volume* or internal microstate potential, $\Omega(a) = A_{\max} - a$, is introduced, representing the remaining life potential (or future possible trajectories) of an individual at age a, normalized to a reference age distribution $Q(a) = \Omega(a)/\sum_{a'=0}^{A_{\max}} \Omega(a')$.

Using the normalized age density $p(a,t) = \rho(a,t)/N(t)$, the per-capita total demographic entropy is defined as the sum of the Shannon entropy and the expected structural potential:

$$S_{\text{total}}(t) = -\sum_{a=0}^{A_{\max}} p(a,t) \ln p(a,t) + \sum_{a=0}^{A_{\max}} p(a,t) \ln Q(a). \tag{20}$$

Under this definition, a young, growing population (a broad pyramid) possesses maximum demographic entropy: The population mass is concentrated where internal phase volume $\Omega(a)$ is largest. As fertility collapses and the population ages, the mass shifts into older cohorts where $\Omega(a)$ approaches zero. This transition acts as a structural contraction process, systematically stripping the society of its demographic potential, driving it toward a low-entropy state characterized by an inverted age pyramid.

As shown in Fig. 9 and Table IV, Japan began this structural contraction in the 1980s. China's entropy did not plummet during the strict implementation of the One-Child policy, but in the 2010s, and South Korea exhibits an unprecedented, almost vertical drop starting in the 2000s, reflecting its world-record fertility crash. By 2070, all three systems converge toward the minimal entropy state, structurally locked

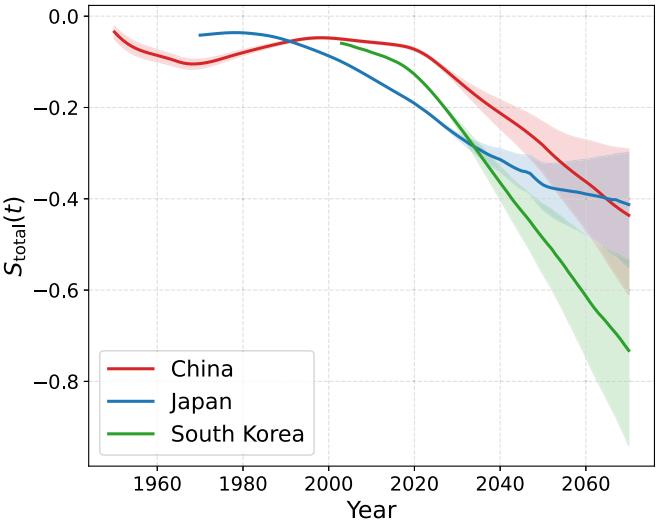


FIG. 9. Total demographic entropy $S_{\text{total}}(t)$ (with 95% credible intervals) from 1950 to 2070. The rapid fertility collapse in East Asia drives a steep structural contraction phase, rapidly stripping the populations of their internal potential as they transition into hyper-aged demographic states.

into rigid, top-heavy distributions with little remaining demographic potential.

Crucially, this entropy contraction acts as a leading indicator of macroscopic demographic decline, preceding both working-age and total population contractions. In Japan, the entropy contraction began in the 1980s, well before the working-age population (15–64) peaked in 1995 (at ≈86 million) and the total population peaked in 2008 (at ≈126.7 million), which then began its uninterrupted decline in 2010-2011. Similarly, China's entropy contraction accelerated in the 2010s, foreshadowing the peak of its working-age population in 2015 (at just under 1 billion) and its first total population decline in 2022 (peaking at ≈1.426 billion in 2021). For South Korea, the vertical entropy drop of the 2000s anticipated the working-age peak in 2017 (at ≈37.4 million) and the total population decline in 2021 (peaking at ≈51.3 million in 2020). The fact that structural entropy peaks and declines decades before physical population peaks highlights its utility as a critical precursor for identifying systemic demographic turning points.

TABLE IV. Nonequilibrium physics and aging metrics (mean ± SD) across case studies for years 2000–2070.

Country	Year	Total entropy S_{total}	D_{KL}	Aging index
China	2000	−0.05(0)	0.20(2)	0.27(0)
	2026	−0.11(0)	0.26(5)	1.11(2)
	2030	−0.14(1)	0.22(6)	1.59(5)
	2040	−0.21(2)	0.14(7)	2.95(38)
	2050	−0.27(3)	0.09(8)	3.43(62)
	2060	−0.34(5)	0.07(7)	5.42(149)
	2070	−0.39(8)	0.05(7)	6.08(264)
Japan	2000	−0.09(0)	0.11(1)	1.15(0)
	2026	−0.23(0)	0.04(1)	2.71(1)
	2030	−0.26(0)	0.03(1)	2.99(6)
	2040	−0.31(1)	0.02(2)	3.71(34)
	2050	−0.37(3)	0.02(2)	4.04(60)
	2060	−0.39(4)	0.02(2)	4.37(89)
	2070	−0.41(6)	0.02(3)	4.72(136)
S. Korea	2026	−0.19(0)	0.44(11)	2.09(1)
	2030	−0.24(0)	0.38(13)	3.07(9)
	2040	−0.36(2)	0.25(16)	6.10(102)
	2050	−0.49(4)	0.17(16)	8.78(257)
	2060	−0.61(7)	0.12(14)	13.23(611)
	2070	−0.73(10)	0.10(12)	18.76(1190)

2. Distance from Lotka equilibrium

In classical mathematical demography [4], if the fertility $f(a,t)$ and mortality schedules were frozen at their instantaneous values at year t, then the population would asymptotically converge to the Lotka stable age distribution:

$$p_t^*(a) = C_t e^{-r_t a} l(a,t), \tag{21}$$

where $l(a,t) = \prod_{i=0}^{a-1} s(i,t)$ is the survival probability to age a, r_t is the intrinsic rate of natural increase solving the Euler-Lotka equation $\sum_a e^{-r_t a} f(a,t) l(a,t) = 1$, and C_t is the normalization constant. The nonequilibrium deviation of the actual age structure from this instantaneous stable attractor is quantified using the Kullback-Leibler (KL) divergence:

$$D_{\text{KL}}(p(\cdot,t) \parallel p_t^*(a)) = \sum_{a=0}^{A_{\max}} p(a,t) \ln\left(\frac{p(a,t)}{p_t^*(a)}\right). \tag{22}$$

This formulation mirrors concepts in nonequilibrium statistical mechanics and information theory, where the KL divergence measures the distance of a transient state from its steady-state Gibbs distribution [46], so that D_{KL} can be treated as a physical measure of demographic driving force.

For physicists, D_{KL} quantifies this driving force, indicating how far a rapid change (such as South Korea's birth rate crash) has pushed the system from its steady-state attractor. For policymakers, a high D_{KL} represents a severe structural mismatch: The current age structure is out of sync with the underlying vital rates. This mismatch implies that even if fertility rates were immediately restored to replacement level, the population would continue to contract for decades due to population momentum (depleted childbearing cohorts). This suggests that immediate policy interventions will have long lags before yielding visible stabilization.

Presently, South Korea shows a massive nonequilibrium deviation ($D_{\text{KL}} \approx 0.474$ in 2024), driven by the rapid collapse of its birth rate. China also exhibits a high divergence ($D_{\text{KL}} \approx 0.290$ in 2024). In contrast, Japan's slow transition keeps it closer to equilibrium ($D_{\text{KL}} \approx 0.040$ in 2024). By 2050, all three countries show a relaxation in D_{KL} (e.g., dropping to 0.17 in South Korea), indicating that the bulk age distributions are catching up to their low-fertility attractors.

3. Labor scarcity and social support dynamics

To link these physical state variables to macrosocial constraints, the Aging Index is reported in Table IV:

$$\mathrm{AgeIdx}(t) = \frac{\sum_{a=65}^{A_{\max}} \rho(a,t)}{\sum_{a=0}^{14} \rho(a,t)}. \tag{23}$$

Under the Status Quo forecast, South Korea shows a transition into a hyper-aged, highly ordered state with declining entropy and extremely low potential support ratios. Its Aging Index is projected to reach 878.2% by 2050 (representing over 8.7 elderly citizens for every child), with its PSR dropping to 1.33. This extreme contraction of the active labor pool suggests that labor scarcity, rather than fiscal resources, becomes the primary physical bottleneck of the social system.

V. DISCUSSION AND CONCLUSIONS

In this work, cohort transport dynamics are shown to be castable as a class of nonequilibrium boundary flux problems where the governing conservation law is exact, but the underlying constitutive relations (fertility and mortality) remain latent and time varying. By parametrizing these constitutive laws via Bayesian neural networks embedded directly within the transport PDE, the framework guarantees mass conservation while capturing complex, nonlinear age-time interactions. The resulting Bayesian inverse formulation provides a mathematically rigorous way to reconstruct missing state variables from sparse, noisy observations and propagate full posterior predictive uncertainty.

An alternative design could also be considered, in which the transport PDE is enforced only as a soft residual penalty, in the spirit of physics-informed neural networks (PINNs). The network weights would then be treated as point estimates obtained by optimization, rather than as posterior samples. Such an approach would substantially reduce the computational cost relative to high-dimensional NUTS sampling. However, the transport constraint would then only be approximately satisfied, up to the residual penalty weight, rather than exactly enforced at every posterior sample as in Eqs. (2)–(4). A point estimate of the weights $\mathbf{W}$ also provides no direct route to calibrated predictive uncertainty. This would instead require a separate *post-hoc* procedure, such as deep ensembles or a Laplace approximation, whose calibration is not guaranteed to match a full Bayesian treatment. The fully Bayesian formulation is therefore retained here, since it ensures exact mass conservation and calibrated uncertainty propagation. The PINN-regularized point-estimate approach nonetheless remains a promising direction for scaling this framework to higher-dimensional or higher-resolution transport problems, where full posterior sampling becomes prohibitive.

Applying this framework to the CJK triad, these systems are shown to undergo a marked structural contraction. By introducing a definition of total demographic entropy, how age structures collapse into highly ordered, low-entropy states under environmental and socioeconomic constraints is characterized. The model's validation against independent World Bank data confirms its ability to infer latent vital rates from age-structured transport constraints alone. Specifically, Japan's gradual transition since the 1980s leads to a stable but heavily aged pillar structure. In contrast, South Korea's sharp entropy drop of the 2000s locks the system into a rigid structural state with a projected PSR of 1.33 by 2050. China's transition, accelerating in the 2010s, demonstrates a similar structural contraction, with the working-age cohort peaking in 2015 and total population entering a historical decline in 2022.

Looking forward, the mathematical structure developed here is universal, extending beyond human demographics. Any nonequilibrium system characterized by advective transport, time-varying boundary generation, and internal advective sinks under mass conservation can be modeled using this unified Bayesian inverse framework. Potential applications include tracking the age- or size-structured density of cell divisions and apoptotic sinks in cellular flows and active matter, modeling polymer chain-length degradation and fragmentation under dynamic thermal stresses, and forecasting the structured transport of viral loads within hosts under latent mutation rates and selective pressures. By unifying the flexibility of Bayesian neural networks with the rigorous constraints of physical conservation laws, this framework offers a generalizable paradigm for observing, modeling, and forecasting bounded transport networks across the physical and biological sciences.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

[1] A. G. M'Kendrick, Applications of mathematics to medical problems, Proc. Edinb. Math. Soc. **44**, 98 (1925).

[2] H. von Foerster, Some remarks on changing populations, in *The Kinetics of Cellular Proliferation*, edited by F. Stohlman Jr. (Grune & Stratton, New York, 1959), pp. 382–407.

[3] P. H. Leslie, On the use of matrices in certain population mathematics, Biometrika **33**, 183 (1945).

[4] S. H. Preston, P. Heuveline, and M. Guillot, *Demography: Measuring and Modeling Population Processes* (Blackwell Publishers, Oxford, 2001).

[5] R. D. Lee and L. R. Carter, Modeling and forecasting U. S. mortality, J. Am. Stat. Assoc. **87**, 659 (1992).

[6] L. Alkema, A. E. Raftery, P. Gerland, S. J. Clark, F. Pelletier, T. Buettner, and G. K. Heilig, Probabilistic projections of the total fertility rate for all countries, Demography **48**, 815 (2011).

[7] A. E. Raftery, N. Li, H. Ševčíková, P. Gerland, and G. K. Heilig, Bayesian probabilistic population projections for all countries, Proc. Natl. Acad. Sci. USA **109**, 13915 (2012).

[8] A. E. Raftery, L. Alkema, and P. Gerland, Bayesian population projections for the United Nations, Stat. Sci. **29**, 58 (2014).

[9] A. M. Stuart, Inverse problems: A Bayesian perspective, Acta Numer. **19**, 451 (2010).

[10] R. M. Neal, MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman,

G. L. Jones, and X.-L. Meng (CRC Press, Boca Raton, FL, 2011), pp. 113–162.
[11] M. D. Hoffman and A. Gelman, The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, J. Mach. Learn. Res. **15**, 1593 (2014).
[12] M. Storbacka and C. Qi, A flexible Bayesian framework for atomic masses by locally inferring configuration mixing, Commun. Phys. **9**, 143 (2026).
[13] A. A. Sojecka and A. Drozd-Rzoska, Global population: From Super-Malthus behavior to doomsday criticality, Sci. Rep. **14**, 9853 (2024).
[14] V. M. Yakovenko, The end of hyperbolic growth in human population and CO_2 emissions, Physica A **661**, 130412 (2025).
[15] A. Zaccone and K. Trachenko, Global population crisis scenarios predicted by a general nonlinear dynamical model, Chaos Solitons Fract. **209**, 118542 (2026).
[16] E. Hernández-García and C. López, Clustering, advection, and patterns in a model of population dynamics with neighborhood-dependent rates, Phys. Rev. E **70**, 016216 (2004).
[17] P. Olla, Demographic fluctuations in a population of anomalously diffusing individuals, Phys. Rev. E **85**, 021125 (2012).
[18] T. J. P. Penna, A bit-string model for biological aging, J. Stat. Phys. **78**, 1629 (1995).
[19] C. Qi, *Mathematical Methods for Physics and Engineering: Practical Applications* (CRC Press, Boca Raton, FL, 2026).
[20] M. Dashti and A. M. Stuart, The Bayesian approach to inverse problems, in *Handbook of Uncertainty Quantification*, edited by R. Ghanem, D. Higdon, and H. Owhadi (Springer, Cham, 2017), pp. 311–428.
[21] A. Duggento, T. Stankovski, P. V. E. McClintock, and A. Stefanovska, Dynamical Bayesian inference of time-evolving interactions: From a pair of coupled oscillators to networks of oscillators, Phys. Rev. E **86**, 061126 (2012).
[22] L. Lu, R. Pestourie, S. G. Johnson, and G. Romano, Multifidelity deep neural operators for efficient learning of partial differential equationswith application to fast inverse design of nanoscale heat transport, Phys. Rev. Res. **4**, 023210 (2022).
[23] D. Ciganda, F. Morini, F. Piriz, H.-A. Schubert, U. Basellini, and M. Myrskylä, Anticipating continued global fertility decline via neural forecasting, arXiv:2605.23858.
[24] F. Perla, R. Richman, S. Scognamiglio, and M. V. Wüthrich, Time-series forecasting of mortality rates using deep learning, Scand. Actuar. J. **2021**, 572 (2021).
[25] S. Galam, Sociophysics: A review of Galam models, Int. J. Mod. Phys. C **19**, 409 (2008).
[26] D. Stauffer, Social applications of two-dimensional Ising models, Am. J. Phys. **76**, 470 (2008).
[27] J. C. Micó, D. Soler, and A. Caselles, Age-structured human population dynamics, J. Math. Sociol. **30**, 1 (2006).
[28] K. Itao, Two universal pathways in demographic transition, Evol. Hum. Sci. **8**, e28 (2026).
[29] G. W. A. Constable and A. J. McKane, Exploiting fast-variables to understand population dynamics and evolution, J. Stat. Phys. **172**, 3 (2018).
[30] M. Jusup, P. Holme, K. Kanazawa, M. Takayasu, I. Romić, Z. Wang, S. Geček, T. Lipić, B. Podobnik, L. Wang, W. Luo, T. Klanjšček, J. Fan, S. Boccaletti, and M. Perc, Social physics, Phys. Rep. **948**, 1 (2022).
[31] D. Acemoglu, D. Autor, K. Beirne, and A. Scott, Baby busts and growth booms: Demographic change and the macroeconomy, NBER Working Paper No. 35401 (2026), https://www.nber.org/system/files/working_papers/w35401/w35401.pdf.
[32] Human Mortality Database, Max Planck Institute for Demographic Research (Germany) and University of California, Berkeley (USA), available at: https://www.mortality.org (last access date: 2026).
[33] National Bureau of Statistics of China (NBS), *China Statistical Yearbook* (China Statistics Press, Beijing, 1980–2024).
[34] United Nations Department of Economic and Social Affairs, Population Division, *World Population Prospects 2024: Methodology* (United Nations, New York, 2024).
[35] China Population and Development Research Center (CPDRC), Population forecasting via PADIS-INT modeling platform, *China Population Statistics Compilation (Beijing, 1990-2024)* (CPDRC Government Briefing, Beijing, 2023).
[36] National Bureau of Statistics of China (NBS), *Tabulation on the 2020 Population Census of the People's Republic of China*, (China Statistics Press, Beijing, 2022).
[37] Chinese Academy of Social Sciences, *Green Book of Population and Labor* (Social Sciences Academic Press, Beijing, 2023).
[38] Ministry of Civil Affairs of China, *Statistical Bulletin on Civil Affairs Development* (Beijing, 2010–2024).
[39] Y. Cai, China's new demographic reality: Learning from the 2010 census, Popul. Dev. Rev. **39**, 371 (2013).
[40] W. Feng, Y. Cai, and B. Gu, Population, policy, and politics: How will history judge China's one-child policy? Popul. Dev. Rev. **38**, 115 (2013).
[41] W. Zhou, A physics model for analysis on demographic structures, Mod. Appl. Sci. **11**, 60 (2017).
[42] https://data.worldbank.org/indicator/SP.DYN.TFRT.IN
[43] J. Liang, Z. Ren, *et al.* (YuWa Population Research), *China Population Prediction Report 2023 Edition,* YuWa Population Research Institute.
[44] S. E. Vollset, E. Goren, C.-W. Yuan, J. Cao, A. E. Smith, T. Hsiao, C. Bisignano, G. S. Azhar, E. Castro, J. Chalek, A. J. Dolgert, T. Frank, K. Fukutaki, S. I. Hay, R. Lozano, A. H. Mokdad, V. Nandakumar, M. Pierce, M. Pletcher, T. Robalik, *et al.*, Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease Study, Lancet **396**, 1285 (2020).
[45] *China Actuarial Report on Pensions 2019-2050*, edited by B. Zheng (China Labor and Social Security Publishing House, Beijing, 2019).
[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2006).