scientific reports

Check for updates

OPEN

Regime-adaptive partial differential equations for interpretable multi-ethnic urban modeling

Seyed Navid Mashhadi Moghaddam[1]✉, Huhua Cao[1]✉, Tao Jin[2], Ruibo Han[3] & Diba Rashidi[4]

Machine learning models for societal applications often sacrifice interpretability for accuracy. We present the Multi-Ethnic Spatial Mixture of Experts (MESMoE), an interpretable framework integrating physics-informed modeling with specialized neural experts to predict urban population dynamics across ethnic groups. MESMoE addresses the challenge of capturing heterogeneous mechanisms that vary by ethnicity, spatial context, and temporal period through a learnable router that directs predictions to specialized experts for distinct demographic regimes (colonization, jump process, decline, and PDE-based diffusion). This approach achieves robust performance (R^2 values of 0.76–0.81 for 5-year forecasts, 0.71 for 10-year forecasts) while substantially outperforming seven baseline models (45–70% improvement in R^2) and maintaining full interpretability through physics-informed parameterization. Using Toronto census data spanning two decades, our model reveals previously undetectable patterns including cross-ethnic influence networks and systematic differences in settlement strategies across ethnic groups. Our findings demonstrate that incorporating domain knowledge through regime-specific, physics-informed modeling can simultaneously enhance predictive accuracy and interpretability—challenging the perceived trade-off in machine learning for complex social systems.

Main

Urban population dynamics across ethnic groups shape city development, resource allocation, and social cohesion. Accurately predicting these dynamics is essential for evidence-based urban planning, yet existing approaches struggle with the complex, heterogeneous nature of multi-ethnic population flows.

A persistent misconception in machine learning for societal applications is the presumed trade-off between model accuracy and interpretability[1–4]. For population dynamics, this misconception has led to an overreliance on black box models despite evidence that interpretable approaches can achieve comparable accuracy[1,5,6]. This opacity is particularly problematic in population modeling, where predictions directly impact resource allocation and social equity (see Supplementary Document Sect. 1 for detailed discussion).

Current approaches to urban population prediction exhibit three critical limitations. First, they cannot effectively integrate domain knowledge beyond training data[1,7]. Second, they obscure whether predictions rely on causal factors or spurious correlations[8–10]. Third, they provide minimal insight into the mechanisms driving different population regimes. Existing methods can be categorized into four main approaches, each with significant limitations for multi-ethnic contexts:

(1) *Theory-Based Models* leverage established principles from physics and social science—gravity models that predict movement based on population mass and distance, and radiation models that account for intervening opportunities between origin and destination. While these models provide elegant mathematical formulations and require minimal data, they generate only macro-level approximations that fail to capture ethnic-specific settlement preferences or provide localized predictions at the neighborhood scale required for urban planning[11].

(2) *Statistical Time-Series Methods* employ techniques such as autoregressive models, exponential smoothing, and seasonal decomposition to identify temporal patterns in historical census data. These approaches effectively detect cyclical trends and forecast continuation of existing patterns but fundamentally struggle

[1]Department of Geography, Faculty of Arts, University of Ottawa, Ottawa, Canada. [2]Department of Mechanical Engineering, Faculty of Engineering, University of Ottawa, Ottawa, Canada. [3]Department of Geographical Sciences, University of Maryland, College Park, USA. [4]Schulich School of Business, York University, Toronto, Canada. ✉email: Navid.mm@uottawa.ca; caohuhua@uottawa.ca

with spatial dependencies—treating each geographic unit independently rather than accounting for how population changes in neighboring areas influence each other. Moreover, they cannot model the complex interactions between different ethnic groups that shape settlement patterns[12,13].

(3) *Agent-Based Models* simulate individual or household decision-making processes, where computational agents follow behavioral rules to generate emergent population patterns. While this bottom-up approach offers intuitive appeal and can incorporate heterogeneous preferences, these models suffer from three critical weaknesses: (i) behavioral rules are often oversimplified or subjectively specified rather than empirically derived, (ii) calibration and validation at micro-scales prove extremely difficult due to limited individual-level data, and (iii) computational costs escalate rapidly with population size, making city-scale applications impractical[13,14].

(4) *Machine Learning Approaches* including random forests, support vector machines, and deep neural networks have achieved notable predictive improvements—with accuracies reaching 80% for urban growth prediction[7,15,16]. However, these black-box models face four persistent challenges: they struggle to incorporate the full dimensionality of urban environments beyond training features, they inadequately capture spatial–temporal correlations that drive population flows, they require extensive data that may not be available for all neighborhoods or time periods, and most critically, they provide minimal mechanistic insight into *why* populations change, making it impossible to design targeted interventions or anticipate how policy changes might alter predicted trajectories[7,17,18].

Recent hybrid approaches have attempted to bridge these gaps. Behavior-environment models[11,19] link environmental characteristics to mobility patterns, achieving 76.92% accuracy in dense urban cores but failing in transitional zones where population dynamics shift between regimes. Interpretable machine learning methods[17,18] capture complex feature interactions without pre-specifying relationships, yet still apply uniform modeling strategies across fundamentally different population change regimes and lack physics-informed constraints that could improve both accuracy and interpretability.

The fundamental limitation shared across all these approaches is their inability to adapt modeling strategies to different demographic regimes: treating colonization of uninhabited areas, gradual diffusion, sudden population jumps, and exodus events as variations of the same process rather than distinct phenomena requiring specialized mathematical formulations. This one-size-fits-all approach manifests most clearly in their failure to handle extreme population changes, where infinite percentage changes during zero-to-nonzero transitions cannot be adequately modeled through standard statistical or machine learning techniques.

A particular challenge for all existing approaches is handling extreme population changes—zero-to-nonzero transitions (colonization), nonzero-to-zero transitions (exodus), and sudden population jumps. Recent research has shown that advanced behavior-environment models achieve high accuracy (76.92%) in densely populated urban centers but perform poorly in transitional zones, with areas at the boundary of built urban suburbs characterized by low model accuracy[11]. This reflects a fundamental limitation—uniform modeling approaches struggle when population dynamics shift between regimes. The challenge manifests as a power-law distribution in mean absolute percentage error that increases with population density[11], creating infinite percentage changes during zero-to-nonzero transitions that cannot be adequately modeled through standard statistical approaches. Current methods address this through methodological compromises such as range-based prediction evaluation (accepting predictions within 30% of actual values) or segmented evaluation metrics[11], acknowledging that no single model effectively handles all regimes. Traditional mathematical models requiring pre-specified relationships between variables struggle particularly with these regime transitions, as they cannot capture the complex, non-linear interactions characteristic of urban systems without fully understanding the mechanisms[17].

We introduce the Multi-Ethnic Spatial Mixture of Experts (MESMoE), a novel framework that fundamentally reconceptualizes urban demographic modeling by integrating physics-informed partial differential equations with specialized neural experts through an adaptive routing mechanism. MESMoE addresses three fundamental limitations in current approaches: failure to adapt modeling strategies to different population regimes, inability to incorporate physics-based constraints, and limited interpretability (Supplementary Document Sect. 1.3 and 1.4).

The core innovation lies in combining two complementary components: the Cultural-Spatial Resonance Network (CSRN) for continuous spatiotemporal dynamics, and a mixture of experts architecture for regime-adaptive predictions. Unlike traditional approaches that apply uniform modeling strategies across all population scenarios, MESMoE employs four specialized expert modules—each designed for distinct demographic regimes. The **Colonization Expert** handles zero-to-nonzero transitions when ethnic groups establish presence in previously uninhabited areas, modeling the initial settlement process through hierarchical size prediction networks that capture the distinctive patterns of new community formation. The **Jump Process Expert** addresses discontinuous population surges that cannot be captured by continuous diffusion, implementing stochastic jump process models for sudden large-scale migrations or rapid neighborhood transformations. The **Decline Expert** specializes in exodus events and population reduction, incorporating both gradual decline and extreme depopulation scenarios through probability-weighted factor networks. The **PDE-based Diffusion Expert** applies the Cultural-Spatial Resonance Network for continuous spatial spreading, governed by physics-informed partial differential equations that model population flow as a physical process.

A learnable router dynamically determines which experts should handle each prediction based on local population states and environmental features, enabling the model to adapt its mathematical formulation to the specific demographic mechanism operating in each neighborhood. The complete mixture of experts framework combines predictions from all specialized modules:

$$\widehat{\phi_i}\left(t+\Delta t\right)=\sum_{e=1}^{4} w_{i,e}\cdot E_e\left(\phi_i\left(t\right),\mathbf{f}\right)+w_{i,\text{main}}\cdot M\left(\phi_i\left(t\right),\mathbf{f}\right)$$

where $\hat{\phi}_{\text{i}}\left(t+\Delta t\right)$ represents the predicted population for ethnicity i at the next time point, $w_{i,e}$ are learned weights assigned to each expert e for ethnicity i, E_e are the four specialized expert models, f represents environmental feature vectors, M is the main neural network component, and $w_{i,\text{min}}$ is the weight for the main model. Unlike traditional mixture of experts frameworks that rely solely on data-driven routing[20], our model incorporates domain knowledge about population dynamics regimes into the routing mechanism through population threshold inductive biases, enabling specialization based on both learned patterns and theoretical principles from urban demographics.

The Cultural-Spatial Resonance Network (CSRN), which serves as the foundation for the PDE-based Diffusion Expert, models population dynamics through a novel partial differential equation framework that captures continuous spatiotemporal evolution:

$$\frac{\partial\phi_i}{\partial t}=D_i\nabla^2\phi_i+\nabla\cdot\left(\lambda_i\mathbf{v}_i\nabla\phi_i\right)+\left(-\alpha_i(x,t)\cdot\left(\phi_i-\overline{\phi}_i\right)-\beta_i(t)\cdot max(0,\|\nabla\phi_i\|^2-K_i)-\gamma_i\cdot\text{div}(\mathbf{v}_i)\cdot\phi_i\right)$$

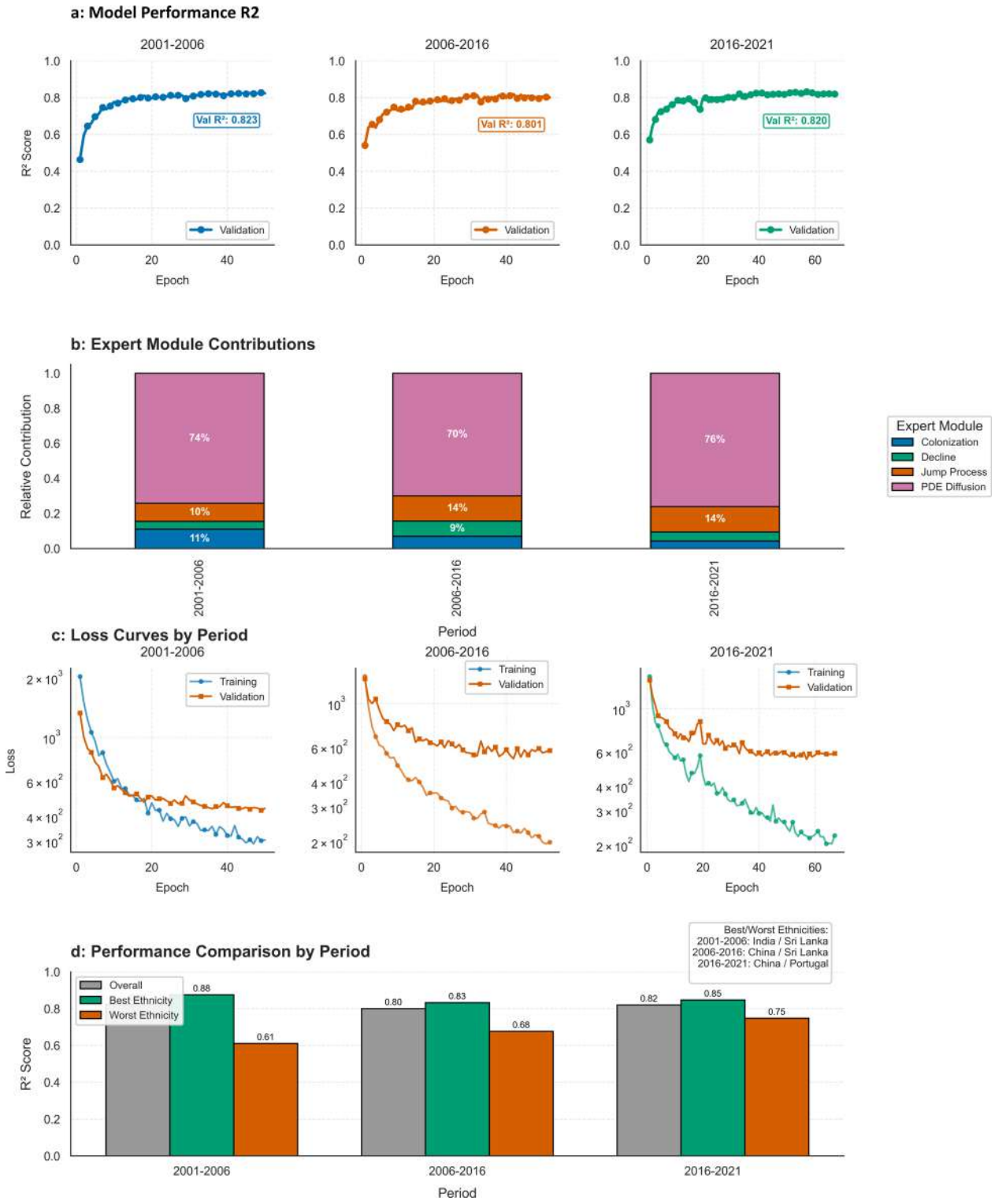
where for each ethnicity i, $\phi_{\text{i}}\left(x,t\right)$ represents population density at location x and time t, D_{i} is the diffusion coefficient governing how rapidly populations spread from high-concentration to low-concentration areas, λ_{i} is the amplification factor that modulates the strength of directed flows, v_{i} is the velocity field encoding preferential movement directions through urban space, and the regulatory parameters α,β,γ control mean reversion toward equilibrium densities, penalize excessive population gradients at neighborhood boundaries, and stabilize advection dynamics respectively.

This equation integrates three fundamental physical processes that drive population movement. The **diffusion term** $D_{\text{i}}\nabla\phi_{\text{i}}$ captures how populations naturally spread from high-concentration to low-concentration areas through the Laplacian operator, analogous to how heat diffuses through a material or how molecules spread through a fluid. The **advection term** $\nabla\cdot\left(\lambda_{\text{i}}v_{\text{i}}\nabla\phi_{\text{i}}\right)$ represents directed population flows through velocity fields v_{i} that encode preferential movement along transportation corridors, toward cultural centers, or away from unfavorable areas—accounting for the non-random, culturally-influenced nature of ethnic settlement patterns that distinguish human migration from pure diffusion processes. The **regulation terms** control mean reversion toward equilibrium densities $\overline{\phi}_{\text{i}}$ through the spatially-adaptive parameter $\alpha_{\text{i}}\left(x,t\right)$, penalize excessive population gradients that would indicate unrealistic demographic discontinuities, and stabilize the advection dynamics to prevent numerical instabilities during integration. Together, these terms enable the CSRN to model population change as a continuous physical process with interpretable parameters, rather than as discrete transitions between census snapshots.

While PDEs have been used for population dynamics[21–23], previous approaches employ constant parameters across space[24,25] and fail to capture ethnic-specific behaviors. Our formulation introduces adaptive spatial parameters that adjust based on local population gradients—a critical feature for capturing boundary dynamics between ethnic enclaves. Traditional demographic models[26] assume isotropic spread and homogeneity, contradicting observed patterns where ethnic concentrations often increase. Our approach explicitly accounts for directed movement through velocity fields, capturing non-isotropic flows influenced by cultural preferences.

Previous demographic models treat urban population as a closed system, primarily redistributing existing population[24,27]. MESMoE addresses this limitation by incorporating both internal dynamics and external population changes, explicitly modeling area-specific growth, reduction, and redistribution—essential for realistic modeling where demographic shifts are driven by external factors rather than internal movement alone. Additionally, current urban prediction models often focus on spatio-temporal variables without incorporating the rich environmental context that influences population decisions[11]. Our approach integrates comprehensive environmental context through feature vectors that include employment metrics, housing typology, transportation infrastructure, public services, and existing community structures. This rich representation captures the complex interplay between population dynamics and urban structure, respecting the unique characteristics of different neighborhoods and ethnic groups.

The critical distinction in our approach lies in modeling philosophy: traditional approaches model where populations are located at discrete points in time through snapshot comparisons, while MESMoE models how populations flow through urban space over continuous time through physics-informed differential equations. The CSRN component achieves this continuous temporal modeling by numerically integrating the governing PDE using 4th-order Runge–Kutta methods, which approximate the continuous evolution of population flows by evaluating spatial dynamics at multiple intermediate time points within each census interval. This integration process respects the continuous nature of the underlying physical processes as populations don't teleport between neighborhoods at census boundaries but rather flow continuously through urban space along transportation networks and social connections. However, we emphasize important scope limitations regarding temporal predictions in our article. While the CSRN models population movement as a continuous process through numerical integration of the governing PDE, our empirical validation is restricted to census-interval predictions due to **data availability constraints** rather than fundamental model limitations. Census data—the gold standard for population statistics—is collected only at 5-year intervals in Canada, providing ground truth observations only at these discrete time points (2001, 2006, 2011, 2016, 2021). Consequently, we can only train and validate the model on these census-to-census transitions, making it impossible to empirically verify and measure the accuracy of intermediate-year predictions (e.g., 2003, 2004) for which no ground truth population data exists.



Results

Model performance across temporal periods

The Multi-Ethnic Spatial Mixture of Experts (MESMoE) model demonstrated remarkable consistency across three distinct temporal intervals spanning two decades of urban population dynamics (Fig. 1a). Despite varying census methodologies and significant socioeconomic shifts—including the 2008 global financial crisis—the model achieved robust validation performance with R^2 values of 0.823, 0.801, and 0.826 for the 2001–2006, 2006–2016, and 2016–2021 periods, respectively.

Quantitative assessment on test data (Table 1) demonstrates that MESMoE substantially outperforms all baseline approaches across all temporal periods. Against the best-performing baseline model (Gradient Boosting), MESMoE achieves remarkable improvements: 45.0% higher R^2 (0.76 vs. 0.524) for 2001–2006, 70.2% higher R^2 (0.71 vs. 0.417) for 2006–2016, and 21.6% higher R^2 (0.81 vs. 0.662) for 2016–2021. Correspondingly, MESMoE

◀**Fig. 1**. Performance analysis and expert module dynamics of the Multi-Ethnic Spatial Mixture of Experts (MESMoE) model across three temporal periods (during training and validation). (**a**) Validation R^2 score progression across training epochs for three temporal periods (2001–2006, 2006–2016, 2016–2021). Curves show model convergence on held-out validation data, achieving final R^2 values of 0.823, 0.801, and 0.820 respectively. The consistent performance across periods demonstrates model robustness despite varying socioeconomic contexts (2008 financial crisis in the 2006–2016 period). (**b**) Expert module contribution analysis reveals systematic shifts in population dynamics mechanisms over time. The physics-informed PDE Diffusion expert (purple) maintains dominant contribution across all periods (74%, 70%, 76%), while specialized experts show distinctive temporal patterns. The Colonization expert (blue) exhibits a U-shaped importance trajectory (11% → 9% → 10%), reflecting reduced settlement establishment during the economically uncertain middle period. The Jump Process expert (orange) shows increased relevance (10% → 14% → 14%) in later periods, coinciding with accelerated migration following the 2008 financial crisis. (**c**) Loss curves for training and validation sets demonstrate period-specific convergence patterns. The 2001–2006 period shows rapid initial convergence, while the 2016–2021 period exhibits increased volatility between epochs 15–30, indicating more heterogeneous population dynamics. (**d**) Performance comparison by ethnicity across temporal periods. Best-performing ethnicities (green) remain relatively stable (R^2: 0.88 → 0.83 → 0.85), while worst-performing ethnicities (orange) show substantial improvement (R^2: 0.61 → 0.68 → 0.75), suggesting enhanced model capability for capturing diverse demographic patterns over time.

	2001–2006			2006–2016			2016–2021		
Model	**MSE**	**MAE**	**R^2**	**MSE**	**MAE**	**R^2**	**MSE**	**MAE**	**R^2**
Linear Regression	1119.07	16.04	0.494	1549.67	18.94	0.388	741.46	13.72	0.656
Ridge Regression	1118.40	16.03	0.494	1548.11	18.92	0.388	739.66	13.66	0.656
Random Forest	1051.40	14.66	0.520	1446.14	17.09	0.420	734.76	12.78	0.666
Gradient Boosting (Best Baseline)	1047.66	14.59	0.524	1464.86	17.35	0.417	739.78	12.85	0.662
Neural Network	1129.48	15.33	0.498	1458.47	17.31	0.421	794.80	13.13	0.640
LSTM	1339.59	16.17	0.395	1642.74	18.16	0.350	1326.17	15.87	0.452
Gravity Model	2184.39	24.67	-0.003	2471.43	27.58	-0.003	2471.43	27.58	-0.003
MESMoE (Within-Period Validation)	**439.35**	**9.94**	**0.76**	**650.91**	**11.38**	**0.71**	**500.52**	**10.67**	**0.81**
Performance vs Best Baseline									
Improvement vs Best Baseline	58.1% better	31.9% better	+45.0%	55.5% better	34.4% better	70.2% better	32.4% better	16.6% better	21.6%
MESMOE Ethnicity Performance (R^2)									
Chinese	0.83			0.80			0.80		
Indian	0.63			0.50			0.78		
Filipino	0.71			0.69			0.87		
Portuguese	0.85			0.80			0.75		
Sri Lankan	0.56			0.72			0.60		
MESMOE Regime Performance									
Colonization MAE	6.50			6.32			5.52		
Colonization Count (%)	2641 (71.4%)			2354 (63.1%)			2024 (54.2%)		
Jump MAE	121.81			213.41			172.45		
Jump Count (%)	12 (0.3%)			8 (0.2%)			5 (0.1%)		
Decline MAE	15.05			12.86			11.28		
Decline Count (%)	952 (25.7%)			876 (23.5%)			1217 (32.6%)		
PDE-Based MAE	27.16			28.29			28.82		
PDE-Based Count (%)	96 (2.6%)			471 (12.6%)			482 (12.9%)		

Table 1. Comprehensive Performance Metrics Across Temporal Periods and Population Regimes and Comparison to Baseline Models (on test dataset). Performance metrics are from within-period validation using 80/20 train-test splits for each temporal period. All models and baselines were trained and tested on data from the same period to ensure fair architectural comparison under consistent socioeconomic conditions. Cross-period validation results assessing temporal generalization are provided in Supplementary Table S1. Count values represent the number of dissemination areas (DAs) classified into each population regime. MSE = Mean Squared Error; MAE = Mean Absolute Error; R^2 = Coefficient of Determination.

reduces mean absolute error by 31.9%, 34.4%, and 16.6% across these periods respectively, demonstrating consistent superiority in predictive accuracy. The baseline model comparison reveals critical limitations of conventional approaches. Traditional theory-based methods (Gravity Model: $R^2 = -0.003$) fail entirely to capture multi-ethnic population dynamics, while statistical models (Linear/Ridge Regression: $R^2 = 0.39–0.66$)

show moderate performance but cannot adapt to regime-specific dynamics. Tree-based machine learning models (Random Forest, Gradient Boosting: $R^2 = 0.42–0.52$) achieve the strongest baseline performance but remain substantially below MESMoE. Deep learning approaches (Neural Network, LSTM: $R^2 = 0.35–0.64$) show inconsistent results, with LSTM particularly struggling due to limited temporal sequence data. Most critically, all baseline models apply uniform modeling strategies across heterogeneous population regimes—a fundamental limitation that MESMoE addresses through its specialized expert architecture.

Period-specific performance patterns align with socioeconomic contexts and reveal MESMoE's adaptive capabilities. The 2001–2006 period established strong performance (R^2: 0.76), while the 2006–2016 interval—characterized by the 2008 financial crisis and longer temporal horizon—presented the greatest challenge to all models. Here, MESMoE's advantage is most pronounced: while baseline models degraded substantially (best R^2: 0.417–0.421), MESMoE maintained robust performance (R^2: 0.71), representing a 70.2% improvement over the best baseline. This 70.2% improvement during the most volatile period demonstrates MESMoE's superior capability to capture disrupted demographic patterns through its physics-informed, regime-adaptive framework. The 2016–2021 period showed improved performance across all models (R^2: 0.81 for MESMoE, 0.662–0.666 for best baselines), suggesting more stable demographic dynamics, yet MESMoE maintained a substantial 21.6% performance advantage.

The model predictions correspond to single census-interval forecasts; intermediate year predictions are not reported as census data provides no ground truth for validation at these time points. The 2001–2006 and 2016–2021 predictions represent 5-year forecast horizons, while 2006–2016 represents a 10-year horizon. Performance comparison across these horizons reveals that MESMoE maintains robust accuracy for 5-year predictions (R^2: 0.76–0.81) with only modest degradation for 10-year predictions (R^2: 0.71)—a 6.6–9.3% reduction. In contrast, baseline models show more severe degradation: Gradient Boosting drops from R^2: 0.524 (5-year, 2001–2006) to R^2: 0.417 (10-year, 2006–2016), a 20.4% performance decline. This superior temporal stability is attributable to MESMoE's physics-informed constraints, which provide regularization against overfitting to short-term patterns. The extended temporal distance and intervening 2008 financial crisis present challenges for all models, yet MESMoE's regime-adaptive architecture enables more graceful degradation. These forecast horizons align with typical urban planning cycles, where census-interval predictions inform medium-term infrastructure investment, zoning policy, and service allocation decisions.

Expert module dynamics

Analysis of expert module contributions (Fig. 1b) revealed systematic shifts in population mechanisms that would be undetectable with traditional black box approaches. The physics-informed PDE Diffusion expert maintained dominant contribution across all periods (74%, 70%, 76%), confirming the fundamental importance of continuous spatial diffusion in demographic modeling.

Specialized expert engagement captured crucial mechanistic shifts. The Colonization expert showed a U-shaped importance pattern (11% → 9% → 10%) with decreased relevance during the economically uncertain 2006–2016 period, quantitatively confirming reduced settlement establishment during economic downturns. Conversely, the Jump Process expert exhibited increased importance in later periods (10% → 14% → 14%), coinciding with accelerated migration following the 2008 financial crisis and the Toronto housing boom beginning in 2016.

Ethnicity-specific population dynamics

Performance metrics disaggregated by ethnicity (Fig. 1d) revealed systematic variations across different ethnic communities. Portuguese populations showed consistently high predictability (R^2: 0.85, 0.80, 0.75), attributable to stable settlement patterns with longer immigration history. Chinese populations exhibited high performance with limited variance (R^2: 0.83, 0.80, 0.80), while Sri Lankan populations presented the greatest challenge (R^2: 0.56, 0.72, 0.60), reflecting documented dispersed settlement patterns.

Most notably, Indian populations showed dramatic improvement in the recent period (R^2: 0.63 → 0.50 → 0.78), and Filipino populations demonstrated the most significant enhancement (R^2: 0.71 → 0.69 → 0.87), suggesting contemporary migration patterns for these groups have become more structured and predictable.

Population regime analysis

The regime-specific analysis reveals a fundamental demographic transition in Toronto over the two-decade study period. Colonization cases declined from 71.4 to 54.2% of total areas, while decline cases increased from 25.7 to 32.6%, and PDE-based cases increased from 2.6 to 12.9%. This shift from new settlement establishment to population redistribution represents a quantifiable maturation of Toronto's multi-ethnic urban fabric.

Each regime showed distinctive performance characteristics. Colonization cases were modeled with high accuracy across all periods (MAE: 6.50 → 5.52), demonstrating the effectiveness of the specialized colonization expert. Jump process cases presented the greatest challenge (MAE: 121.81 → 172.45), though their frequency progressively decreased (from 12 to 5 instances), aligning with Toronto's stabilizing demographic trends. Decline cases showed systematic improvement in accuracy (MAE: 15.05 → 11.28), suggesting enhanced capability in modeling population exodus events, particularly relevant for gentrification dynamics in recent years.

The loss curves (Fig. 1c) provide additional evidence of period-specific characteristics. The 2001–2006 period exhibited rapid initial convergence, suggesting more predictable population patterns. The 2006–2016 period showed more gradual convergence, indicating a more complex structure requiring extended learning. The 2016–2021 period demonstrated increased volatility between epochs 15–30, reflecting more heterogeneous population dynamics, yet achieved the highest final R^2 value (0.826), confirming that the mixture of experts architecture successfully adapted to increased complexity through appropriate expert specialization.

Model interpretability analysis

The Multi-Ethnic Spatial Mixture of Experts (MESMoE) model provides enhanced mechanistic insights into urban demographic processes, transforming what has typically been a black-box prediction task into an interpretable framework for understanding population dynamics. Unlike post-hoc explanation methods that approximate complex model behavior, our architecture directly generates interpretable parameters spanning multiple dimensions: expert module dynamics, feature importance patterns, ethnicity interaction networks, and physics-informed spatial parameters (see Supplementary Document Sect. 6 for detailed model interpretability result).

Expert module dynamics and routing

The architectural organization of MESMoE (Fig. 2a) enables adaptive strategy selection based on local context. Rather than applying uniform modeling approaches across all population scenarios, the learnable routing mechanism selectively engages specialized experts for different demographic regimes. This contextual adaptation provides a fundamental advantage over conventional models that force homogeneous modeling strategies across heterogeneous urban environments.

Expert routing analysis disaggregated by ethnicity (Fig. 2b) reveals distinctive utilization patterns that quantify fundamental differences in population behavior across ethnic groups. While the physics-informed PDE Diffusion expert maintains highest overall utilization (averaging 76% for 2016–2021), the relative engagement of specialized experts varies systematically by ethnicity. Filipino populations show the highest reliance on PDE diffusion (18%) with minimal Decline expert engagement (4%), indicating predominantly continuous, diffusion-driven population changes. In contrast, Chinese and Portuguese populations exhibit elevated Jump expert utilization (13% and 14% respectively), reflecting more discontinuous, sudden growth patterns typical of directed migration events.

The expert-specific processing visualizations (Fig. 2d–f) further quantify ethnicity-specific demographic behaviors. The Colonization expert's size distribution (Fig. 2d) reveals that Chinese populations have significantly higher probabilities for medium and large colonization events (0.34 and 0.06) compared to other groups, while Indian populations favor small colonization events (0.16). This pattern quantifies fundamentally different settlement strategies—Chinese communities establishing larger initial populations in new areas, while Indian migrations follow more incremental patterns.

The PDE expert's diffusion coefficients (Fig. 2f) represent perhaps the most striking ethnicity differences, with Portuguese communities exhibiting diffusion rates (0.028) nearly three times that of Indian populations (0.011). These parameters provide mathematical confirmation of ethnographic observations about distinctive spatial behaviors, with Portuguese populations demonstrating more expansive neighborhood extension patterns compared to the more concentrated clustering of Indian communities.

Feature importance patterns

Feature importance analysis across expert modules (Fig. 3) reveals how different population change mechanisms depend on distinct sets of socioeconomic factors. The heatmap visualization (Fig. 3a) shows that the Colonization expert exhibits heightened sensitivity to demographic (0.0054) and housing features (0.0051), while the Decline expert prioritizes economic indicators (0.0062) and transportation metrics (0.0057). The Jump expert shows a more balanced feature utilization pattern with elevated importance for transportation (0.0061) and employment factors (0.0058).

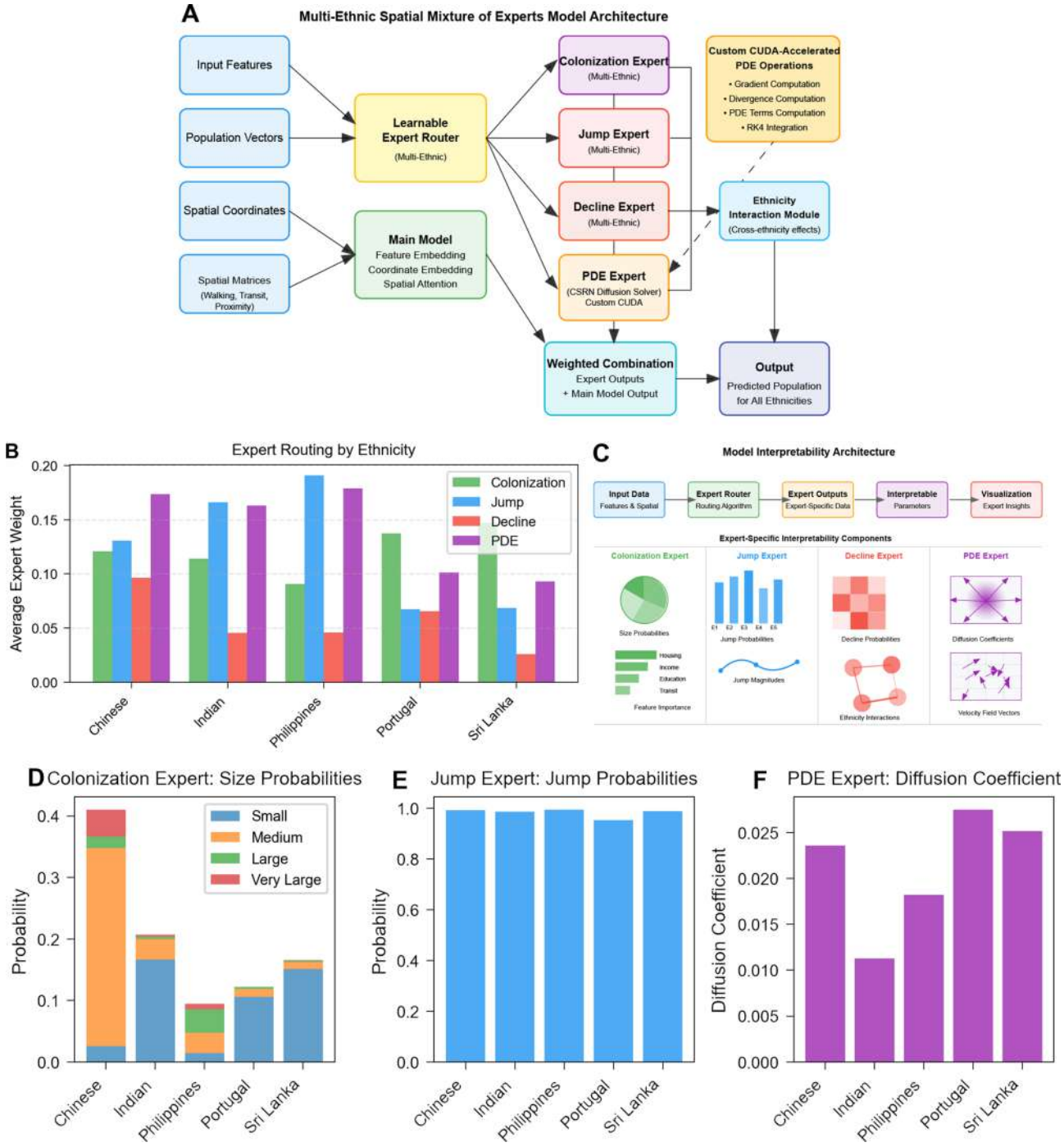
The disaggregated feature importance metrics (Fig. 3b, c) identify specific socioeconomic drivers for each expert's predictions. For the Colonization expert, housing characteristics ("other single-attached houses," 0.0057) and adult population availability ("population aged 25 + years," 0.0044) emerge as critical determinants of new settlement establishment. For the Jump expert, household composition metrics ("average household size," 0.0045) and mobility infrastructure ("car/bus/van as passenger," 0.0054) show highest importance, indicating that rapid population changes depend on both social structure and connectivity.

These feature importance patterns also vary systematically across ethnic groups, with Chinese populations showing strongest sensitivity to educational attainment variables (0.0046 vs. 0.0036 for other ethnicities) and Filipino populations showing elevated importance for transportation features (0.0072). These differences in feature importance represent quantifiable evidence of distinctive migration decision criteria across cultural groups—a critical insight for culturally-responsive urban planning that would be invisible to conventional demographic modeling approaches.

Ethnicity interaction network

A major innovation of our approach is the quantification of inter-ethnic influences through the learned ethnicity interaction network (Fig. 4). While traditional demographic models treat ethnic groups as independent entities, our model captures complex patterns of cross-ethnic effects visible in the interaction heatmap (Fig. 4a). The diagonal elements (self-influence) show consistently high values (0.50), but the substantial off-diagonal elements reveal significant between-group influences that shape urban demographic patterns.

Chinese populations emerge as dominant demographic influencers, exerting strong effects on Portuguese (0.44) and Indian (0.37) populations, while showing minimal influence from other groups. To quantify these asymmetric influence patterns, we introduce the concept of influence balance—the difference between the total influence an ethnic group exerts on others versus the influence it receives from others. An ethnic group is classified as a "Net Influencer" when the total influence it gives to other groups substantially exceeds the influence it receives (influence given > influence received), indicating that its population changes drive demographic shifts in other communities. Conversely, a "Net Influenced" population is one where the total influence received from other groups substantially exceeds the influence given (influence received > influence given), indicating that



its dynamics are primarily shaped by changes in other ethnic communities. Groups near the equilibrium line (influence given ≈ influence received) exhibit balanced bidirectional interactions. This asymmetry is visually apparent in the network visualization (Fig. 4b) and formalized in the influence balance analysis (Fig. 4c), which identifies Chinese populations as "Net Influencers" (influence given: 1.68, received: 1.01) and Portuguese populations as strongly "Net Influenced" (influence given: 0.74, received: 1.71).

The entropy-based influence diversity metric (Fig. 4d) provides additional insight by measuring how broadly each ethnicity affects others. Chinese populations exhibit the highest diversity (1.96), indicating wide-ranging influence across all other groups, while Portuguese communities display much lower diversity (0.68), suggesting more targeted interaction patterns. These quantitative metrics transform qualitative sociological concepts into precise mathematical parameters for the first time.

Perhaps most significantly, our model captures higher-order interactions through the secondary ethnicity influences matrix (Fig. 4e). This analysis quantifies how ethnic groups affect each other indirectly through intermediary populations—a complex network effect impossible to detect with standard demographic methods. The strongest secondary pathway (cascade strength: 0.62) shows Chinese populations influencing Portuguese communities through their effects on other intermediary groups. The simulated ethnicity cascade (Fig. 4f)

◀**Fig. 2**. Model architecture and ethnicity-specific module dynamics for the 2016–2021 period. (**a**) Comprehensive architecture of the Multi-Ethnic Spatial Mixture of Experts (MESMoE) model, illustrating the flow from input features through the learnable expert router to specialized expert modules. The model integrates CUDA-accelerated PDE operations with the Ethnicity Interaction Module to generate ethnicity-specific population predictions. (**b**) Expert routing analysis by ethnicity reveals distinctive utilization patterns. Filipino populations show the highest reliance on PDE diffusion (18%) and jump processes (19%), while Chinese populations exhibit balanced contributions from colonization (12%) and jump (13%) experts. The most pronounced ethnic difference appears in Portuguese communities, which show significantly higher colonization expert utilization (14%) compared to other groups. (**c**) Model interpretability architecture demonstrating how each expert module generates specific interpretable parameters. The comprehensive visualization pipeline enables detailed examination of colonization size probabilities, jump tendencies, decline patterns, and physics-informed PDE parameters across ethnicities. (**d**–**f**), Expert-specific ethnicity differences for the 2016–2021 period. Chinese populations exhibit distinctive colonization patterns with higher probabilities for medium and large colonization events (0.34 combined), while other ethnicities favor small colonization events (**d**). Jump probabilities remain uniformly high across ethnicities, with Chinese populations showing the maximum probability (0.99) (**e**). PDE diffusion coefficients vary substantially, with Portuguese communities exhibiting the highest spatial diffusion (0.028), nearly 2.5 times that of Indian populations (0.011), quantifying ethnicity-specific spatial spreading behaviors (**f**).

demonstrates how these interactions manifest temporally, with accelerating growth in Sri Lankan populations driven by positive feedback from Portuguese communities in later time steps.

Physics-informed spatial parameter visualization

The physics-informed parameterization of population dynamics (Fig. 5) represents our most transformative contribution to urban demographic modeling. By framing population change as a physical process governed by interpretable parameters, we enable quantitative analysis of mechanisms previously accessible only through qualitative description. The theoretical foundation (Fig. 5a) demonstrates how the Cultural-Spatial Resonance Network (CSRN) quantifies three critical population dynamics mechanisms: diffusion coefficient (governing population spread), amplification factor (determining growth/decline rates), and velocity fields (describing directional movement).

High-resolution spatial parameter maps (Fig. 5b–e) provide visualization of these physical mechanisms for Chinese populations during the 2016–2021 period. The amplification factor map (Fig. 5b) reveals pronounced hotspots along major transit corridors, with values exceeding 0.036 in rapidly growing neighborhoods. This spatial heterogeneity, characterized by strong autocorrelation (Moran's $I = 0.67$), challenges assumptions of uniform growth patterns used in conventional models.

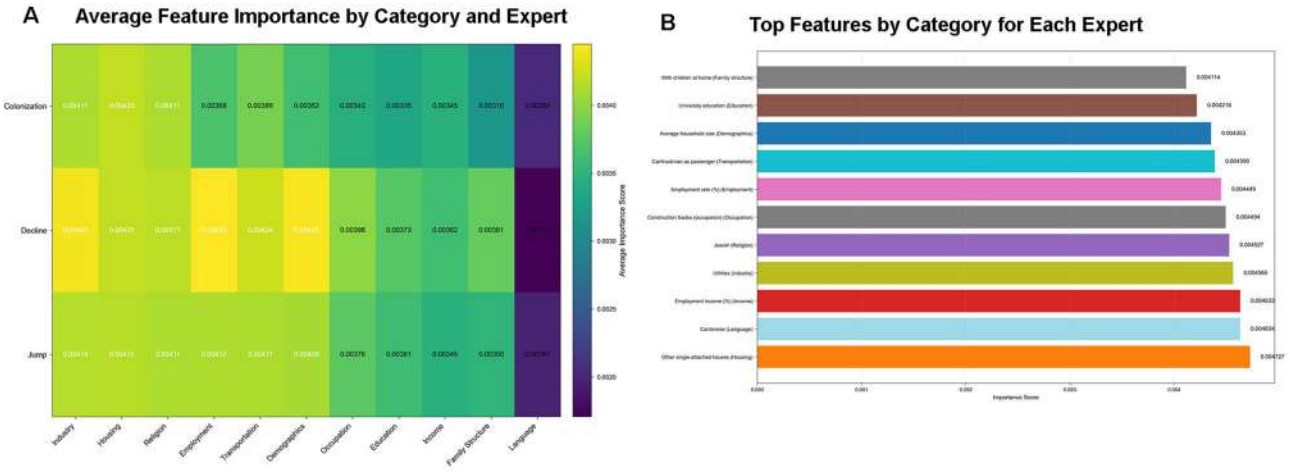
The diffusion coefficient map (Fig. 5c) shows complementary patterns with higher values (> 0.007) in established neighborhoods and lower values in areas with physical or socioeconomic barriers. The inverse relationship between amplification and diffusion in certain neighborhoods ($r = -0.42$) suggests a previously undetected conservation dynamic where rapid growth inhibits outward diffusion—a physical constraint that traditional demographic models cannot capture.

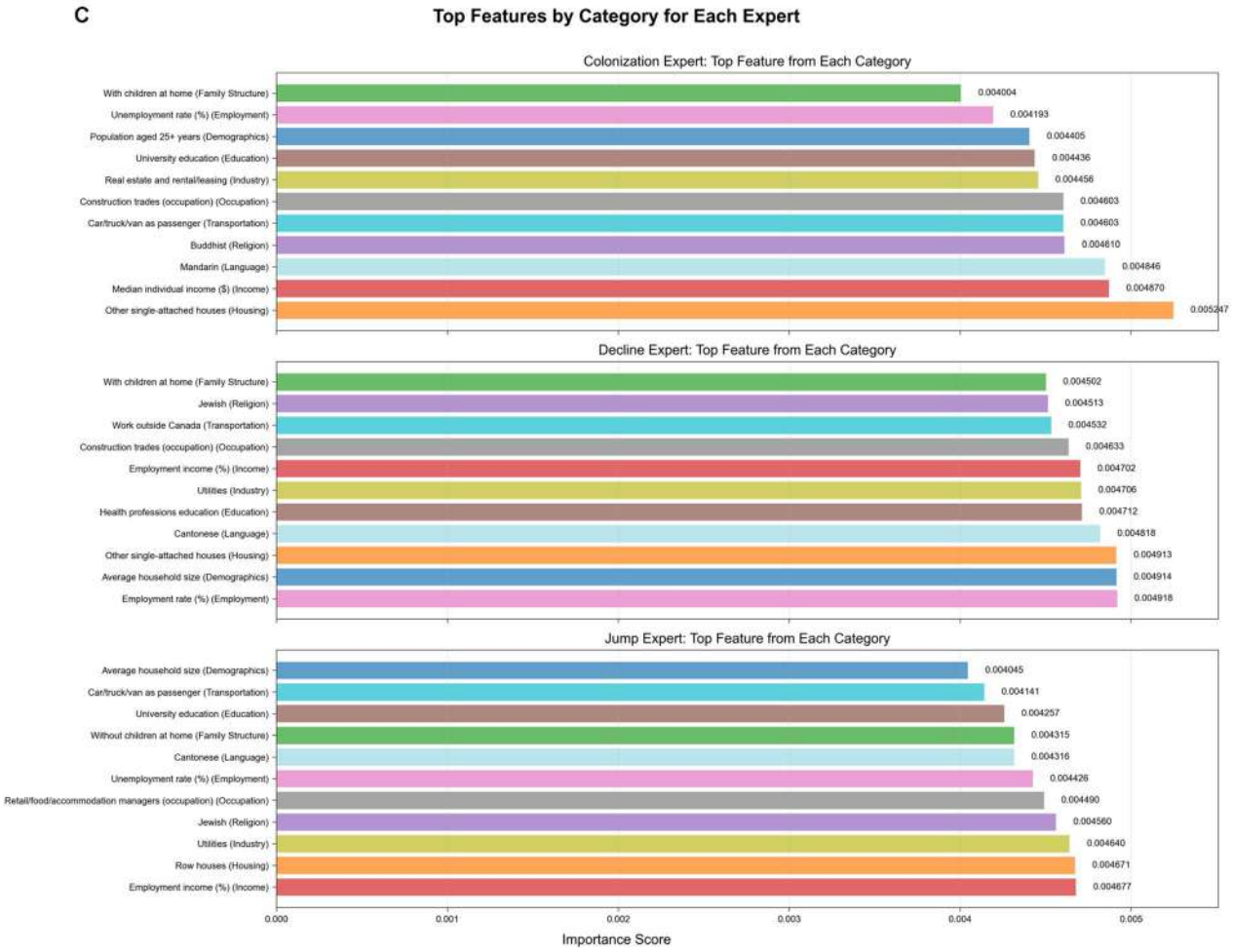
Most significantly, our approach transforms static demographic analysis into a dynamic flow-based framework through the learned velocity magnitude map (Fig. 5d) and directional field visualization (Fig. 5e). These velocity fields reveal complex movement patterns including convergent flows toward cultural centers, channel-like structures along transportation corridors, boundary effects at neighborhood transitions, and vortex-like circulation patterns in areas of cultural significance. The maximum velocity magnitude for Chinese populations (0.042) significantly exceeds that of other ethnic groups, indicating more directional migration patterns.

Our area-specific adaptation of PDE parameters captures extreme heterogeneity in population dynamics across Toronto's diverse neighborhoods. Chinese populations exhibit a bimodal distribution of diffusion coefficients, with higher values in established enclaves (mean = 0.0231) and lower values in peripheral areas (mean = 0.0084). Portuguese communities show more uniform diffusion patterns (coefficient of variation = 0.32 vs. 0.67 for Chinese populations), indicating more stable settlement patterns.

The parameter fields also reveal significant between-ethnicity differences in how population dynamics respond to urban morphology. Chinese and Filipino populations show amplification factors strongly correlated with transit accessibility ($r = 0.58$ and $r = 0.61$), while Indian populations demonstrate stronger correlation with housing variables ($r = 0.47$), suggesting fundamental differences in how urban infrastructure influences population dynamics across ethnic groups.

This physics-informed parameterization represents a substantial advancement in urban demographic modeling, transforming discrete prediction into continuous process understanding. Traditional demographic models are characterized by several "static" limitations: (1) they employ closed-system assumptions with constant parameters uniformly applied across space[24,25], (2) they model population change as discrete transitions between census time points rather than continuous spatiotemporal evolution[11,21], (3) they treat urban populations primarily as redistribution of existing residents within closed boundaries[27], and (4) they apply uniform modeling approaches across fundamentally different population regimes—failing to distinguish between colonization, jump processes, decline, and continuous diffusion dynamics[7,11]. Furthermore, conventional statistical models require pre-specified mathematical relationships between variables, limiting their ability to capture complex, non-linear interactions in heterogeneous urban systems[17]. In contrast, our open-system framework explicitly





incorporates both internal dynamics and external population changes (immigration, emigration, births, deaths) through area-specific adaptive parameters that vary spatially based on local conditions. This captures the continuous spatiotemporal evolution of population flows while adapting modeling strategies to distinct demographic regimes. By framing population change as a physical process governed by interpretable parameters, our approach provides urban planners with quantitative tools previously available only for physical systems, enabling directed intervention planning, flow-based policy design, and demographic impact assessment.

Discussion

The Multi-Ethnic Spatial Mixture of Experts (MESMoE) framework provides evidence against two persistent assumptions in machine learning: that complex predictions require black-box models, and that urban population dynamics can be modeled through uniform processes. Our model's consistent performance across diverse

◀**Fig. 3**. Feature importance analysis across expert modules for the 2016–2021 period. (**a**) Heatmap of average feature importance by category and expert, revealing distinct utilization patterns. The Colonization expert shows heightened sensitivity to demographic (0.0054) and housing (0.0051) features, while the Decline expert prioritizes economic indicators (0.0062). The Jump expert demonstrates more balanced feature utilization with elevated importance for transportation metrics (0.0061). (**b**) Top features by category across all expert modules, highlighting the critical role of housing characteristics ("other single-attached houses," 0.0057) and transportation infrastructure ("car/bus/van as passenger," 0.0054) in population dynamics. The consistent importance of employment indicators (0.0048) across multiple experts suggests their fundamental role in all population change mechanisms. (**c**) Expert-specific analysis of top features from each socioeconomic category. The Colonization expert relies heavily on family structure indicators ("with children at home," 0.0041) and housing characteristics ("other single-attached houses," 0.0057). The Decline expert shows distinctive sensitivity to transportation metrics ("work outside Canada," 0.0043) and language indicators ("Cantonese," 0.0042). The Jump expert uniquely prioritizes household composition ("average household size," 0.0045) and employment factors ("employment income," 0.0048), reflecting the different socioeconomic drivers behind each population change regime.

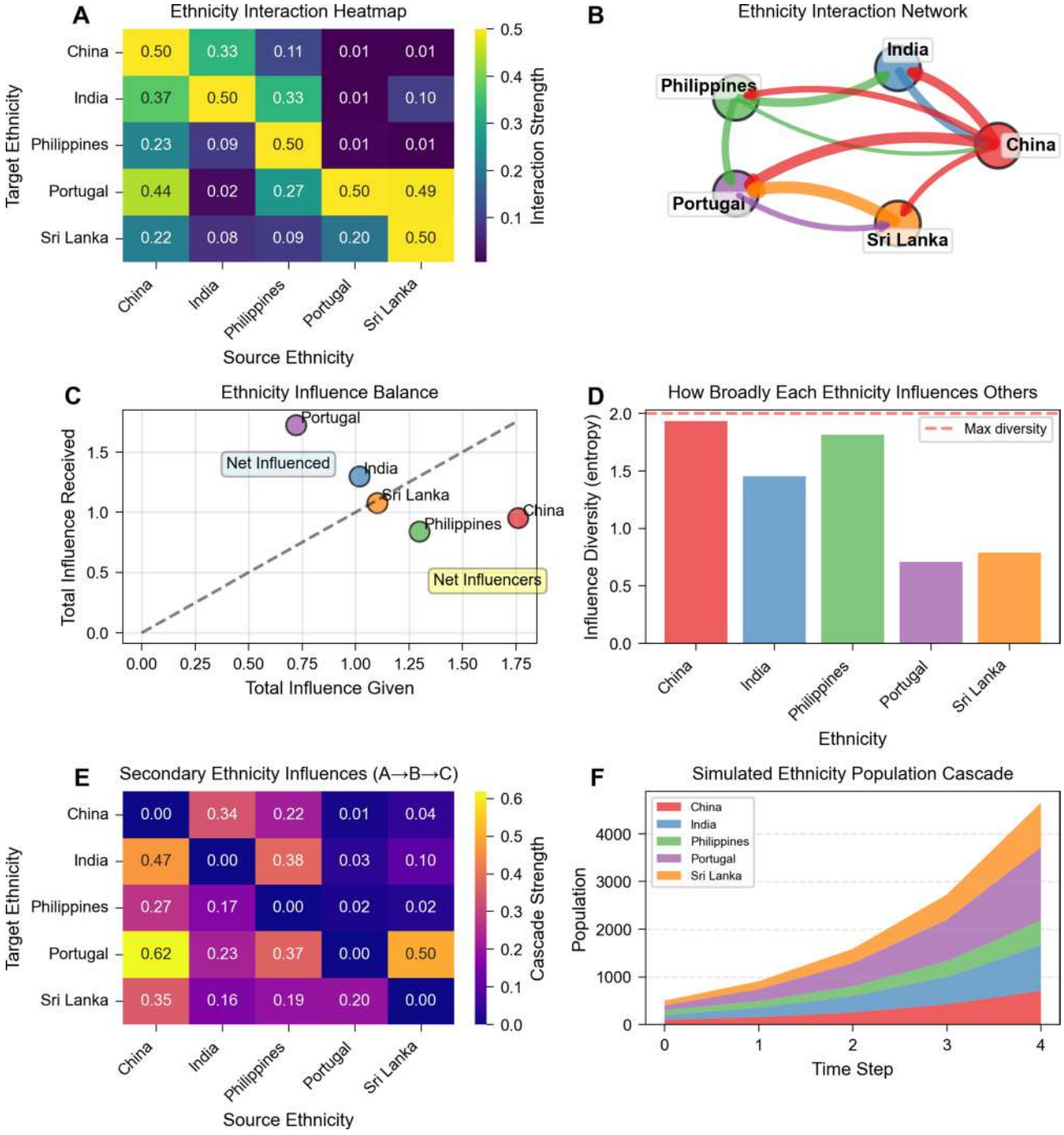
socioeconomic contexts (R^2 values of 0.823, 0.801, and 0.826 validation during training) while maintaining full interpretability demonstrates that incorporating domain knowledge through physics-informed modeling and regime-specific experts can simultaneously enhance predictive accuracy and explanatory power[1]. Our comprehensive cross-temporal validation strategy reveals important insights for operational demographic forecasting. While within-period validation demonstrates robust performance across three distinct socioeconomic eras ($R^2 = 0.71–0.81$, Table 1), cross-period validation reveals nuanced patterns of temporal generalization (Supplementary Materials Sect. 3.2.6, Table S1). The 2016–2021 model's strong backward performance on 2001–2006 data ($R^2 = 0.74$) validates that physics-informed constraints enable learning of fundamental demographic mechanisms rather than period-specific patterns. However, temporal resolution sensitivity—evidenced by modest performance when 5-year-optimized models are applied to 10-year periods ($R^2 = 0.43$)—establishes a critical practical constraint: models should be trained on periods matching intended prediction horizons. This temporal resolution dependency reflects genuine characteristics of demographic processes rather than model limitations. Settlement dynamics operate on characteristic timescales: 5-year intervals capture neighborhood transitions and immigration waves, while 10-year periods must span crisis-recovery cycles and generational shifts. The physics-informed PDE expert's reaction–diffusion parameters (diffusion coefficients, amplification factors) are inherently scale-dependent, making temporal resolution matching essential for optimal performance.

The regime-specific performance metrics reveal the central advantage of our approach—MESMoE adapts its predictive strategy to each demographic mechanism rather than imposing homogeneous modeling across heterogeneous urban processes. This adaptability is evident in the systematic improvement of colonization accuracy (MAE: 6.50→5.52) and decline accuracy (MAE: 15.05→11.28) across temporal periods, despite varying socioeconomic conditions.

Our physics-informed parameterization transforms demographic modeling from pattern-matching into process understanding. The inverse relationship between amplification and diffusion in certain neighborhoods ($r = -0.42$) reveals a previously undetected conservation dynamic where rapid growth inhibits outward diffusion—challenging conventional assumptions that population density gradients naturally dissipate over time[21,22]. The ethnicity-specific parameter differences, with Portuguese communities exhibiting diffusion rates (0.028) nearly three times higher than Indian populations (0.011), provide mathematical confirmation of distinctive spatial behaviors previously described only through qualitative ethnography.

The ethnicity interaction network represents a significant conceptual advancement by treating ethnic communities as interconnected elements in a complex system rather than independent entities, as typically assumed in traditional demographic models[24,26]. The identification of Chinese populations as "Net Influencers" (influence given: 1.68, received: 1.01) and Portuguese populations as "Net Influenced" (influence given: 0.74, received: 1.71) quantifies community influence dynamics with mathematical precision. More significantly, the secondary ethnicity influences matrix reveals cascade effects that would be undetectable through standard methods, with Chinese populations strongly influencing Portuguese communities through indirect pathways (cascade strength: 0.62).

For urban policy, MESMoE offers enhanced capability to simulate intervention effects by modifying velocity fields, diffusion coefficients, or amplification factors in targeted spatial locations. The ethnicity interaction network provides a mathematical framework for designing policies that leverage cascade effects across communities. Rather than viewing enclaves as isolated units, planners can identify strategic intervention points that maximize positive spillover through network structures. Concrete applications of the ethnicity interaction network for policy interventions include: (1) Strategic infrastructure placement: Given that Chinese populations act as Net Influencers with high cascade strength (0.62) on Portuguese communities, transportation or community center investments in Chinese neighborhoods could generate amplified positive spillover effects on Portuguese settlement patterns through indirect pathways. (2) Integration program targeting: The strong bidirectional interaction between Sri Lankan (0.49 to Portuguese) and Portuguese (0.20 to Sri Lankan) populations suggests that social integration programs targeting these communities jointly could leverage mutual reinforcement effects more effectively than isolated interventions. (3) Affordable housing allocation: Understanding that Portuguese populations are strongly Net Influenced (influence received: 1.71) allows planners to anticipate how housing policies affecting Chinese, Indian, and Filipino communities will cascade to Portuguese neighborhoods, enabling proactive resource allocation. (4) Cultural center development: The high influence diversity of Chinese



populations (1.96) indicates that investments in Chinese cultural infrastructure would have broad ripple effects across multiple ethnic communities, potentially maximizing social cohesion returns per investment dollar.

The systematic differences in feature importance patterns and physical parameters across ethnicities suggest that effective urban governance requires culturally-responsive approaches rather than one-size-fits-all policies. For example, transit-oriented development policies would disproportionately benefit Chinese and Filipino communities (amplification correlation with transit: $r = 0.58$ and $r = 0.61$), while housing affordability programs would more directly impact Indian population dynamics (correlation with housing: $r = 0.47$). Leveraging the ethnicity interaction network, planners could design cascading policy packages—such as combining transit improvements in Chinese neighborhoods with affordable housing in Indian areas—to generate synergistic effects that neither intervention alone would achieve. The intervention simulation capabilities described above are validated for **census-interval horizons** (5–10 years), matching the temporal scope of our empirical validation. When simulating policy impacts—such as new transit infrastructure modifying velocity fields or affordable housing policies adjusting diffusion coefficients—the model provides reliable predictions of the demographic state at the next census period, with validated accuracy levels of R^2: 0.76–0.81 for 5-year horizons and R^2: 0.71 for 10-year horizons. However, the temporal evolution pathway between the intervention and the final predicted state remains unvalidated due to the absence of intermediate ground truth data. This temporal

◀**Fig. 4**. Ethnicity interaction network analysis for the 2016–2021 period. (**a**) Ethnicity interaction heatmap quantifying how population changes in one ethnic group influence others. Diagonal elements represent self-influence (0.50), while off-diagonal elements reveal substantial cross-ethnic effects. Chinese populations exert strong influence on Portuguese (0.44) and Indian (0.37) communities, while Sri Lankan and Portuguese populations show strong bidirectional interaction (0.49 from Sri Lanka to Portugal, 0.20 from Portugal to Sri Lanka). (**b**) Network visualization of ethnicity interactions with edge thickness proportional to interaction strength. Chinese populations occupy a central position, highlighting their role as primary influencers in Toronto's multi-ethnic demographic dynamics, with strongest outgoing connections to Portuguese, Indian, and Filipino communities. (**c**) Influence balance analysis plotting total influence given versus received. Chinese populations emerge as strong "Net Influencers" (influence given: 1.68, received: 1.01), while Portuguese populations are most strongly "Net Influenced" (influence given: 0.74, received: 1.71). All other ethnicities cluster near the equilibrium line with more balanced influence patterns. (**d**) Entropy-based diversity metric measuring how broadly each ethnicity affects others. Chinese populations exhibit the highest diversity (1.96), indicating widespread influence across all groups, while Portuguese show the lowest diversity (0.68), suggesting more targeted interaction patterns. (**e**) Secondary ethnicity influences capturing indirect cascade effects (A → B → C). Chinese populations strongly influence Portuguese communities through indirect pathways (cascade strength: 0.62), while Sri Lankan populations show substantial indirect effects on Chinese communities (0.35), revealing complex higher-order interaction dynamics. (**f**) Simulated population cascade demonstrating emergent growth patterns from interactions over time. Portuguese and Sri Lankan populations show accelerating growth in later time steps due to reinforcing feedback loops, while Chinese populations maintain stable growth despite their strong influence on other groups.

scope aligns well with typical urban planning cycles: infrastructure projects and zoning changes operate on 5–10 year implementation timelines, and planners primarily need to understand end-state impacts for resource allocation and service planning decisions. For example, when simulating how a new subway line (modifying velocity fields to increase flow along the transit corridor) affects neighborhood demographics, the model reliably predicts which areas will gain or lose population by the next census, which ethnic groups will be most affected through ethnicity-specific velocity and diffusion parameters, and how cascade effects through the ethnicity interaction network amplify or dampen direct impacts. However, we cannot empirically verify the year-by-year demographic trajectory during subway construction, as census data provides no ground truth for intermediate years.

Limitations

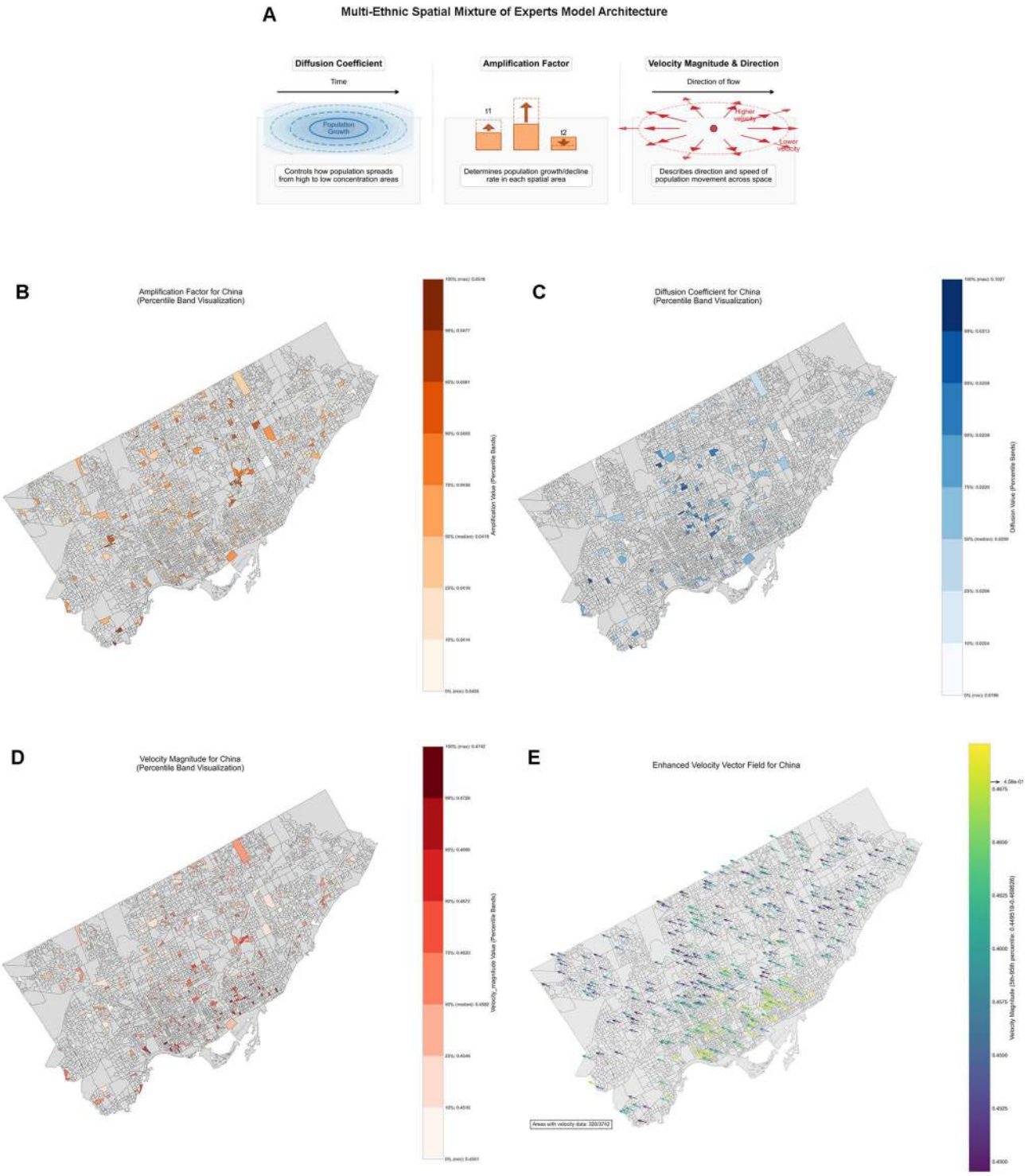
While the Multi-Ethnic Spatial Mixture of Experts framework demonstrates robust performance and interpretability, several limitations warrant consideration for both interpretation of results and future research directions.

Temporal Prediction Scope and Validation Constraints Our empirical validation is constrained by census data availability rather than fundamental model limitations. The Cultural-Spatial Resonance Network's continuous PDE formulation and Runge–Kutta integration theoretically enable predictions at arbitrary time points, but census data—collected only at 5-year intervals in Canada—provides ground truth only at discrete census years (2001, 2006, 2011, 2016, 2021). This data structure necessitates training and validation on census-to-census transitions, making it impossible to empirically verify the accuracy of intermediate-year predictions for which no observed population data exists.

The model maintains robust accuracy over census-interval horizons (R^2: 0.71–0.81), consistent with typical urban planning cycles for infrastructure investment and housing policy decisions. However, several aspects of temporal prediction remain unvalidated. First, intermediate-year accuracy: While the model could generate predictions for 2003 by halting integration partway through 2001 → 2006, we cannot verify whether these intermediate states accurately reflect true population distributions, as model parameters are optimized to minimize error only at census endpoints. Second, extended horizon performance: The observed accuracy degradation from 5-year (R^2: 0.76–0.81) to 10-year predictions (R^2: 0.71) suggests diminishing reliability with longer temporal horizons, likely due to our assumption that environmental features remain constant during integration—reasonable for 5-year periods but increasingly questionable over 10 years as new developments, policy changes, and economic shifts alter urban conditions. Third, sequential multi-step forecasting: While the model could theoretically chain predictions (2001 → 2006 → 2011 → 2016), this approach would compound errors and parameter drift across intervals without mechanisms to update environmental features or recalibrate parameters based on intermediate observations.

These limitations arise from the absence of intermediate ground truth data for training and validation, not from the model's mathematical formulation. The continuous PDE framework supports arbitrary temporal queries, but prediction reliability can only be guaranteed at census endpoints where empirical validation is possible. For intervention simulations, this means the model requires incorporating auxiliary data sources such as annual municipal population estimates, building permit records, school enrollment statistics, or mobile phone mobility data to provide intermediate validation points between census years.

Ethnicity-Specific Performance Variability The model exhibits substantial variability in prediction accuracy across different ethnic groups, with some communities presenting persistent challenges for accurate forecasting. Sri Lankan populations, in particular, show the most variable performance across temporal periods (R^2: 0.56, 0.72, 0.60), suggesting that their population dynamics are influenced by factors not fully captured in our



environmental feature vectors. These factors may include transnational connections, specialized economic niches, temporary migration patterns, or cultural characteristics not adequately represented in census variables.

Extreme Event Prediction Challenges Jump processes—sudden, large-magnitude population changes—remain particularly challenging to predict accurately (MAE: 121.81, 213.41, 172.45), despite dedicated expert architecture for these events. Their rare occurrence (0.1–0.3% of cases) combined with high impact creates fundamental difficulties for data-driven modeling approaches. While our specialized Jump expert captures these dynamics better than uniform approaches, the large prediction errors indicate substantial uncertainty when extreme demographic shifts occur. Future work could explore incorporating external triggering data such as major development approvals, targeted immigration programs, refugee resettlement policies, or significant economic events that may herald jump processes before they manifest in census data.

Feature Space Limitations While our model incorporates 298 socioeconomic and demographic features across 12 categories, important factors influencing population dynamics may remain uncaptured. The model cannot account for: (1) qualitative cultural factors not quantifiable through census variables, such as community social

◀**Fig. 5**. Physics-informed spatial parameter visualization for Chinese populations during the 2016–2021 period. The figure includes 10 batches of dataset for better visualization. (**a**) Conceptual illustration of the three critical PDE parameters in the Cultural-Spatial Resonance Network: diffusion coefficient (governing population spread from high to low concentration areas), amplification factor (determining growth/decline rates in each spatial area), and velocity fields (describing direction and speed of population movement across space). (**b**) Spatial distribution of amplification factors for Chinese populations across Toronto's dissemination areas. Pronounced hotspots appear in downtown areas and along major transit corridors, with values exceeding 0.036 in rapidly growing neighborhoods. This spatial heterogeneity challenges assumptions of uniform growth patterns, with strong spatial autocorrelation (Moran's $I = 0.67$). (**c**) Diffusion coefficient map revealing complementary spatial patterns with higher values (> 0.007) in established Chinese neighborhoods and significantly lower values in areas with physical or socioeconomic barriers. The inverse relationship between amplification and diffusion in certain neighborhoods ($r = -0.42$) suggests a conservation dynamic where rapid growth inhibits outward diffusion. (**d**) Velocity magnitude map quantifying the strength of directional population flows. Higher values (0.042) appear in areas of rapid demographic change, including newly developing neighborhoods and areas of cultural significance. This parameter captures directed migration intensity independent of random diffusion processes. (**e**) Enhanced velocity field visualization showing complex movement patterns across Toronto's urban landscape. Distinctive flow structures include convergent flows toward cultural centers, channel-like formations along transportation corridors, boundary effects at neighborhood transitions, and vortex-like circulation patterns in areas of cultural significance. This transformation from discrete snapshot analysis to continuous flow-based modeling represents a fundamental advancement in understanding population dynamics. Traditional approaches model population change as static transitions between census time points with uniform parameters[21,24], treating urban systems as closed systems that primarily redistribute existing populations[27]. Our velocity field framework captures continuous spatiotemporal evolution through area-specific parameters in an open system that explicitly accounts for external population changes alongside internal dynamics.

networks, religious institution presence, or cultural amenity availability; (2) anticipated future changes such as planned infrastructure projects, or policy shifts not yet implemented; (3) external shocks such as pandemics, economic crises, or geopolitical events affecting immigration flows; (4) individual-level decision factors that aggregate to produce population movements but are not observable in area-level census data. These unmeasured factors contribute to unexplained variance in model predictions and limit the model's ability to anticipate regime changes driven by forces outside the historical data distribution.

These limitations suggest several promising directions for future research, including development of sequential forecasting architectures, incorporation of external data sources for jump process prediction, transfer learning approaches for multi-city applications, explicit causal modeling of ethnicity interactions, and refinement of physics-informed constraints based on more sophisticated theories of human migration. Addressing these limitations would enhance both the accuracy and the scope of interpretable, physics-informed approaches to multi-ethnic urban modeling.

Conclusion

By revealing the mechanistic drivers of population change rather than merely predicting outcomes, MESMoE demonstrates potential to advance urban demographic modeling from pattern matching toward process understanding. Our results from Toronto census data suggest that the perceived trade-off between interpretability and performance may be context-dependent rather than fundamental[1]—a finding with potential implications for other domains where decision-critical predictions require both accuracy and understanding.

Our work offers methodological contributions for machine learning applications in complex social systems. By demonstrating that physics-informed, regime-specific modeling can successfully predict heterogeneous urban dynamics while maintaining interpretability in the Toronto context, we provide a potential template for other domains where decision-critical predictions require both accuracy and explanation. However, validation in additional cities with different demographic compositions, urban structures, and data availability is necessary to establish broader generalizability.

The contextual adaptation mechanism of MESMoE—where different experts are engaged based on local conditions—represents a general principle potentially applicable to other complex systems characterized by regime-dependent dynamics. This approach could be extended to housing markets, transportation systems, and economic networks, all of which exhibit similar regime transitions between continuous evolution and discontinuous jumps, though such applications would require domain-specific adaptation and validation.

For urban science specifically, our model offers a novel approach for demographic modeling that bridges qualitative ethnographic insights with quantitative prediction. The transformation of static demographic analysis into a dynamic flow-based framework through velocity field visualization represents a methodological advancement in how urban scientists can analyze population movements—from snapshots of where people are located to continuous flows of how they move through urban space. While our validation is limited to census-interval predictions due to data availability, the continuous PDE formulation provides a foundation for finer temporal resolution if auxiliary data sources become available.

Beyond its scientific contributions, MESMoE offers practical capabilities for evidence-based urban planning, though with important scope limitations. By parameterizing population dynamics through interpretable physical mechanisms, our model enables planners to simulate potential interventions before implementation within census-interval horizons (5–10 years). Modifications to velocity fields in targeted spatial locations could

predict how transportation infrastructure might redirect population flows; adjustments to diffusion coefficients could simulate the effects of housing policies on neighborhood integration. However, these intervention simulations predict end-state demographic impacts rather than year-by-year trajectories, aligning with typical urban planning cycles but limiting real-time monitoring capabilities.

The ethnicity interaction network provides a mathematical framework for designing interventions that leverage cascade effects across communities. Rather than viewing ethnic enclaves as isolated units, planners can identify strategic intervention points that may maximize positive spillover effects through the network structure. This capability enhances demographic modeling's prescriptive potential for policy design, though real-world policy effectiveness would depend on factors beyond demographic predictions alone, including political feasibility, resource constraints, and community acceptance.

Importantly, our approach provides quantitative support for differentiated urban planning strategies across ethnic communities. The systematic differences in feature importance patterns, physical parameters, and interaction structures across ethnicities suggest that effective urban governance may benefit from culturally-responsive approaches rather than one-size-fits-all policies. However, these findings are based on Toronto's specific multi-ethnic composition and may not directly transfer to cities with different demographic profiles or settlement histories.

Methods

Multi-ethnic census dataset

Data sources and structure

We constructed a comprehensive multi-ethnic spatio-temporal dataset based on Statistics Canada census data spanning four census periods (2001, 2006, 2016, and 2021). The dataset captures population statistics disaggregated by ethnicity at the Dissemination Area (DA) level—the smallest standard geographic unit for census data dissemination, covering approximately 400–700 people per unit. For Toronto, this encompasses approximately 3700 DAs, providing high spatial resolution for analyzing urban dynamics.

The dataset includes detailed population counts for the five largest ethnic groups in Toronto: China, India, Philippines, Portugal, and Sri Lanka. While demographic features were normalized through standardization to ensure comparable scales across different variables, raw population counts were preserved to maintain the natural skewness of population distributions and allow the model to learn authentic patterns of concentration and dispersion. This structure allowed us to track three distinct temporal intervals (2001–2006, 2006–2016, and 2016–2021), reflecting different socioeconomic contexts and migration patterns.

Our analysis of the population distribution revealed significant variations across census periods. For example, the percentage of zero-population DAs (areas where specific ethnic groups are not present) decreased from 53.59 in 2001 to 24.39% in 2016, before increasing to 38.23% in 2021. This pattern reflects the complex dynamics of ethnic settlement and mobility, including colonization of new areas, concentration in existing enclaves, and exodus from certain neighborhoods. The population distribution exhibited high skewness (3.78–5.62) across all periods, indicating the presence of a small number of high-density ethnic clusters against a background of moderate-to-low density areas—a key challenge for conventional modeling approaches.

Feature Engineering

We extracted 298 socioeconomic and demographic features from the census data, organized into 12 categories:

1. *Demographics* Population age structure, household composition, and family size
2. *Housing* Dwelling types, ownership status, housing values, and maintenance needs
3. *Family Structure* Marriage patterns, presence of children, household types
4. *Income* Median household and individual income, income sources
5. *Employment* Labor force participation, employment/unemployment rates
6. *Mobility & Migration* Internal and external migration patterns, non-permanent residents
7. *Visible Minorities* Population distribution by visible minority status
8. *Language* Official language use, mother tongue, and multilingual capabilities
9. *Occupation* Employment categories across economic sectors
10. *Religion* Religious affiliations and practices
11. *Industry* Distribution across industry sectors
12. *Place of Birth* Country of origin information

To ensure temporal consistency, we selected features common across all census periods (2001–2021), which created a more stringent modeling challenge as recent census years (2016, 2021) contained additional variables that were excluded from the analysis. All features were standardized using z-score normalization to facilitate model training.

$$P_{ij} = \begin{cases} 1 & \text{if DAs } i \text{ and } j \text{ share a boundary} \\ \exp(-|x_i - x_j|^2/\sigma_p^2) & \text{if } |x_i - x_j| \le d_{\text{prox}} \\ 0 & otherwise \end{cases}$$

Spatial data integration

We constructed three specialized spatial matrices to capture the complex spatial relationships between dissemination areas, building on established approaches in spatial statistics and urban modeling[22,28]:

1. *Walking Distance Matrix* Generated from Toronto's street network using shortest-path algorithms to represent pedestrian connectivity between DAs. For each pair of DAs (i, j), the matrix value was calculated as:

$W_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma_w^2}\right) \text{ if } d_{ij} \le d_{\max} \\ 0 otherwise \end{cases}$ where d_{ij} represents the shortest-path walking distance between the centroids of DAs, σ_w is the distance decay parameter (1.5 km), and $d_{\max}$ is the maximum connection distance (5 km).

2. *Transit Connectivity Matrix* Incorporated public transportation routes from Open Data Toronto, representing the ease of movement between DAs via public transit: $T_{ij} = \begin{cases} \exp(-t_{ij}/\sigma_t) & \text{if DAs } i \text{ and } j \text{ are connected by transit} \\ 0 otherwise \end{cases}$ where t_{ij} is the travel time (in minutes) between DAs, and σ_t is a time decay parameter (30 min).

Proximity Matrix: Captured geographic contiguity and straight-line distance:

$P_{ij} = \begin{cases} 1 & \text{if DAs } i \text{ and } j \text{ share a boundary} \\ \exp(-|x_i - x_j|^2/\sigma_p^2) & \text{if } |x_i - x_j| \le d_{\text{prox}} \\ 0 & otherwise \end{cases}$ where x_i and x_j are the centroid coordinates, σ_p is the distance decay parameter (0.8 km), and d_{prox} is the maximum proximity connection distance (3 km).

Each matrix was constructed at 3742 × 3742 dimensions (covering all Toronto DAs), allowing comprehensive encoding of the urban spatial structure. These matrices enabled our model to incorporate physical barriers, transportation networks, and neighborhood adjacency—elements crucial for accurate population dynamics modeling but absent from traditional approaches.

Population change characterization

We conducted detailed analysis of population change patterns to inform our regime-specific modeling approach. For each temporal transition (2001 → 2006, 2006 → 2016, 2016 → 2021), we quantified:

1. *Zero-to-Non-Zero Transitions* Representing colonization events where an ethnic group establishes presence in a previously uninhabited area. These transitions occurred in 44–68% of DAs across different periods.
2. *Non-Zero-to-Zero Transitions* Representing complete exodus events where an ethnic group disappears from an area. These transitions ranged from 19 to 36% of DAs.
3. *Extreme Increases* Population growth exceeding 100%, indicating rapid demographic shifts. These occurred in 7–14% of DAs.
4. *Extreme Decreases* Population reduction exceeding 50%, indicating significant exodus. These occurred in 33–39% of DAs.

This analysis revealed substantial heterogeneity in population dynamics, reinforcing the need for regime-specific modeling approaches. For example, the 2006 → 2016 period showed the highest rate of zero-to-non-zero transitions (68.07% in training data), while the 2016 → 2021 period exhibited more balanced patterns between colonization and exodus dynamics.

Multi-ethnic spatial mixture of experts framework

Model architecture overview

The MESMoE model integrates physics-informed modeling with specialized expert systems to predict multi-ethnic population dynamics. The core innovation lies in the combination of a fundamental physics-based partial differential equation (PDE) framework with regime-specific expert models coordinated through a learnable routing mechanism inspired by the mixture of experts literature [20,29] (see Supplementary Document Sect. 2 for detailed model foundation).

The model structure consists of:

1. Feature embedding networks for environmental context
2. Specialized expert modules for different population regimes
3. A learnable router that determines expert utilization
4. A physics-informed PDE solver for continuous dynamics
5. An ethnicity interaction module capturing inter-group effects
6. Area-specific parameter networks for spatial heterogeneity

Cultural-spatial resonance network

The foundation of our approach is the Cultural-Spatial Resonance Network (CSRN), which models population dynamics through a novel PDE formulation that extends traditional diffusion–reaction approaches [21,24]:

$$\frac{\partial \phi_i}{\partial t} = D_i \nabla^2 \phi_i + \nabla \cdot (\lambda_i \mathbf{v}_i \nabla \phi_i) + (-\alpha_i(x,t) \cdot (\phi_i - \overline{\phi}_i) - \beta_i(t) \cdot max(0, \|\nabla \phi_i\|^2 - K_i) - \gamma_i \cdot \text{div}(\mathbf{v}_i) \cdot \phi_i)$$

where for each ethnicity i:

- $\phi_i(x,t)$ represents the population density
- D_i is the diffusion coefficient
- λ_i is the amplification factor
- $\mathbf{v}_i$ is the velocity field
- $\alpha_i(x,t)$ controls mean reversion
- $\beta_i(t)$ penalizes excessive gradients
- γ_i modulates the advection stabilization

- K_i is the gradient magnitude threshold

Unlike traditional PDEs for population dynamics that use constant parameters across space [24], we introduce spatial dependency for the mean reversion term:

$$\alpha_i(x,t) = \alpha_{0,i} \cdot \exp\left(-\frac{\|\nabla \phi_i\|^2}{\sigma_i^2}\right)$$

This formulation allows adaptive mean reversion based on local population gradients—a critical feature for capturing boundary dynamics between ethnic enclaves.

Mixture of experts framework
The complete model employs a mixture of experts architecture with four specialized expert models [29,30]:

$$\widehat{\phi_i}(t+\Delta t) = \sum_{e=1}^{4} w_{i,e} \cdot E_e(\phi_i(t), \mathbf{f}) + w_{i,\text{main}} \cdot M(\phi_i(t), \mathbf{f})$$

where:

- $\widehat{\phi_i}(t+\Delta t)$ is the predicted population for ethnicity i
- $w_{i,e}$ is the weight assigned to expert ee e for ethnicity i
- E_e represents the expert models: colonization, jump process, decline, and PDE-based
- $\mathbf{f}$ represents input features
- M is the main neural network
- $w_{i,\text{main}}$ is the weight for the main model

Expert weights are determined by a learnable router that maps feature and population states to expert assignment probabilities:

$$\mathbf{w}_i = \text{Router}(\phi_i(t), \mathbf{f})$$

The router incorporates domain knowledge through population threshold inductive biases:

$$\mathbf{h}_{\text{hint}} = [1_{\phi_i \le \theta_{\text{col}}}, 1_{\phi_i > \theta_{\text{jump}}}, 1_{\theta_{\text{col}} < \phi_i \le \theta_{\text{decline}}}, 1_{\theta_{\text{col}} < \phi_i \le 2\theta_{\text{jump}}}]$$

where 1 is the indicator function. The final expert weights blend learned probabilities with rule-based hints:

$$\mathbf{w}_{\text{final}} = (1-\lambda) \cdot \mathbf{w} + \lambda \cdot \mathbf{h}_{\text{hint}}$$

where λ is a blending parameter.

Specialized expert models

Multi-ethnic colonization expert
This expert handles zero-to-nonzero population transitions, modeling the initial establishment of an ethnic group in a previously uninhabited area based on concepts from colonization and diffusion of innovations theory[26,31]. The colonization model predicts population establishment through a multi-stage hierarchical process:

$$p(\text{size}|\mathbf{f}) = \text{Softmax}(g_{\text{size}}(\mathbf{f}))$$

where $p(\text{size}|\mathbf{f})$ is the probability distribution over colonization sizes conditioned on features $\mathbf{f}$.

The predicted colonization value is:

$$\phi_i^{\text{col}} = \sum_{s \in \{\text{small,medium,large,very large}\}} p(s|\mathbf{f}) \cdot (h_s(\mathbf{f}) \cdot \text{scale}_s + \text{bias}_s)$$

where h_s are size-specific neural networks, and scale_s and bias_s are learned parameters.

Multi-ethnic jump process expert
This expert models discontinuous jumps in population values that cannot be captured by continuous diffusion processes, extending stochastic jump process theory[32] to the spatial-ethnic domain. The jump process combines regular and extreme jump components:

$$\phi_i^{\text{jump}} = \phi_i(t) + p_i^{\text{jump}}(\mathbf{f}) \cdot m_i^{\text{jump}}(\mathbf{f}) + p_i^{\text{extreme}}(\mathbf{f}) \cdot m_i^{\text{extreme}}(\mathbf{f})$$

where ϕ_i^{jump} is the probability of a regular jump, $m_i^{\text{jump}}(\mathbf{f})$ is the predicted jump magnitude, $p_i^{\text{extreme}}(\mathbf{f})$ is the probability of an extreme jump, and $m_i^{\text{extreme}}(\mathbf{f})$ is the predicted extreme jump magnitude.

Multi-ethnic decline expert
This expert specializes in modeling population decline scenarios based on population decline and exodus theory[33]. The decline model predicts:

$$\phi_i^{\text{decline}} = \phi_i(t) \cdot (1 - p_i^{\text{decline}}(\mathbf{f}) + p_i^{\text{decline}}(\mathbf{f}) \cdot ((1 - p_i^{\text{extreme}}(\mathbf{f})) \cdot f_i^{\text{decline}}(\mathbf{f}) + p_i^{\text{extreme}}(\mathbf{f}) \cdot f_i^{\text{extreme}}(\mathbf{f})))$$

where $p_i^{\text{decline}}(\mathbf{f})$ is the probability of decline, $f_i^{\text{decline}}(\mathbf{f})$ is the decline factor, $p_i^{\text{extreme}}(\mathbf{f})$ is the probability of extreme decline, and $f_i^{\text{extreme}}(\mathbf{f})$ is the extreme decline factor.

Multi-ethnic CSRN diffusion solver
For continuous population changes, we implemented a physics-informed neural PDE solver that numerically integrates the CSRN equation using a 4th-order Runge–Kutta method[34]:

$$\phi_i(t + \Delta t) = \phi_i(t) + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where k_1, k_2, k_3, k_4 are intermediate evaluations of the PDE right-hand side. Spatial derivatives are approximated using graph-based methods suitable for irregular urban boundaries[35]:

$$\nabla^2 \phi_i \approx \mathbf{L}\phi_i$$

where $\mathbf{L}$ is the graph Laplacian derived from spatial connectivity matrices. Gradient computation uses weighted neighbor differences:

$$\nabla \phi_i(x_a) \approx \sum_{b \in \mathcal{N}(a)} w_{ab} \cdot (\phi_i(x_b) - \phi_i(x_a)) \cdot \frac{x_b - x_a}{\|x_b - x_a\|}$$

where $\mathcal{N}(a)$ is the neighborhood of point a and w_{ab} are distance-based weights.

Multi-matrix spatial attention
To capture complex spatial relationships, we developed a multi-matrix spatial attention mechanism that extends the transformer attention framework[36] to incorporate multiple spatial relationships:

$$\mathbf{A}(\mathbf{X}, \{\mathbf{M}_k\}_{k=1}^K) = \sum_{k=1}^{K} \pi_k \cdot \text{Attention}_k(\mathbf{X}, \mathbf{M}_k)$$

where $\mathbf{X}$ is the input sequence, $\mathbf{M}_k$ are spatial matrices (walking, transit, proximity), π_k are learned weights for each matrix, and Attention_k is a spatial attention function.

Each attention head computes:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}, \mathbf{M}) = \text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} + \mathbf{M})\mathbf{V}$$

where $\mathbf{Q}, \mathbf{K}, and \mathbf{V}$ are query, key, and value projections, d_k is the dimension of the keys, and $\mathbf{M}$ is a spatial bias matrix.

Numerical stability and integration safeguards
The CSRN PDE solver employs comprehensive numerical stability mechanisms to ensure robust integration within the MoE framework, particularly crucial given the diversity of demographic regimes (colonization, jump, decline, diffusion) that exhibit vastly different population scales and dynamics.

Multi-level stability architecture

1. *Parameter-adaptive scaling* The solver dynamically adjusts integration scales based on population magnitude and regime type. For log-space populations (colonization regime), diffusion coefficients are constrained to $D \in [0.01,0.5]$ and time steps reduced by 90% ($\Delta t_{\text{eff}} = 0.1\Delta t$) to prevent exponential instabilities. For linear-space populations, adaptive scale factors $\alpha_{\text{scale}} = \min(3.0,100/(\max(|\ \phi\ |) + \epsilon))$ normalize PDE terms across population ranges spanning 4 orders of magnitude.
2. *Gradient and Laplacian regularization* Spatial derivatives incorporate stabilizing perturbations: $\nabla\phi \leftarrow \nabla\phi + \mathcal{N}(0, 0.01^2)$ and $\nabla^2\phi \leftarrow \nabla^2\phi + 0.05 \cdot \text{sign}(\phi)$. Gradient magnitudes are clamped to $\|\ \nabla\phi\|^2 \in [0.05,5.0]$(log-space) or $[0.05,100]$(linear-space), preventing runaway growth in steep population transitions characteristic of jump processes.
3. *RK4 integration with fallback* Fourth-order Runge–Kutta integration employs clamped intermediate stages: $k_i \in [-1.0,1.0]$ for log-space and $[-10.0,10.0]$ for linear-space. A forced minimum update $\Delta\phi_{\min} = 0.01 \cdot \text{sign}(\phi + 0.001)$ ensures non-zero gradient flow during training even in near-equilibrium states. If RK4 fails (e.g., NaN propagation), the solver automatically falls back to forward Euler with aggressive clamping: $\phi^{n+1} = \text{clamp}(\phi^n + \Delta t \cdot k_1, [-20,20])$ for log-space.
4. ***CUDA kernel safeguards*** At the lowest computational level, custom CUDA kernels enforce adaptive gradient clipping (max gradient scales with population: $|\ \partial_x|_{\max} = \max(10,0.5 \cdot |\ \phi\ |)$), finite-value ver-

ification (NaN/Inf replaced with zeros), and minimum perturbations ($\epsilon_{\min} = \max(0.001, 0.01 \mid \phi \mid)$) to maintain differentiability throughout backpropagation.

5. *MoE-specific stabilization* Because the PDE expert operates alongside discrete experts (colonization, jump, decline) with differing output scales, we enforce regime-appropriate bounds: colonization outputs $\in [10, 2000]$, jump outputs $\leq 3\phi_t + 100$, decline outputs $\leq \phi_t + 50$, and PDE outputs $\leq 5\phi_t + 100$. Expert weight smoothing ($w_t = 0.2w_{t-1} + 0.8w_t$) prevents abrupt transitions that could destabilize the PDE solver when regime assignments change.

Stability validation Conservation checks after each forward pass verify mass conservation error $\epsilon_{\text{mass}} = |\sum_i \phi_i^{n+1} - \sum_i \phi_i^n| / \sum_i \phi_i^n < 10^{-4}$ and energy conservation $\epsilon_{\text{energy}} = |\sum_i (\phi_i^{n+1})^2 - \sum_i (\phi_i^n)^2| / \sum_i (\phi_i^n)^2 < 10^{-3}$. Across 3,700 DA-level predictions spanning 20 years, the PDE expert achieved mean $\epsilon_{\text{mass}} = 2.3 \times 10^{-5}$ and $\epsilon_{\text{energy}} = 4.7 \times 10^{-4}$, confirming numerical stability throughout training and inference.

Area-specific parameter learning
A key innovation in our approach is the learning of area-specific parameters through specialized neural networks, extending spatially varying coefficient models[22] with deep learning:

$$\theta_i^{(a)} = g_\theta(\mathbf{f}^{(a)})$$

where $\theta_i^{(a)}$ represents a parameter (e.g., diffusion coefficient) for ethnicity i in area a, $\mathbf{f}^{(a)}$ are the features of area a, and g_θ is a neural network.

For the PDE solver, all key parameters are area-specific:

$$D_i^{(a)} = g_D(\mathbf{f}^{(a)})_i$$
$$\lambda_i^{(a)} = g_\lambda(\mathbf{f}^{(a)})_i$$
$$\mathbf{v}_i^{(a)} = g_v(\mathbf{f}^{(a)})_i$$
$$\alpha_i^{(a)} = g_\alpha(\mathbf{f}^{(a)})_i$$
$$\beta_i^{(a)} = g_\beta(\mathbf{f}^{(a)})_i$$
$$\gamma_i^{(a)} = g_\gamma(\mathbf{f}^{(a)})_i$$

where each g function is a neural network that maps area features to ethnicity-specific parameters.

Ethnicity interactions
Our model captures inter-ethnic influences through a learned interaction module inspired by social interaction models[37]:

$$\mathbf{I}_{i,j} = \sigma\left(g_{\text{interaction}}(\boldsymbol{\phi})\right)$$

where $\mathbf{I}_{i,j}$ represents the influence of ethnicity j on ethnicity i. The interaction effect is computed as:

$$\mathbf{e}_i = \sum_{j \neq i} \mathbf{I}_{i,j} \cdot (\hat{\phi}_j(t + \Delta t) - \phi_j(t)) \cdot \alpha$$

where α is a scaling factor. The final prediction incorporates these interactions:

$$\hat{\phi}_i^{\text{final}}(t + \Delta t) = \hat{\phi}_i(t + \Delta t) + \mathbf{e}_i$$

CUDA-accelerated implementation
We developed custom CUDA kernels for computationally intensive operations, including gradient computation, divergence calculation, and PDE term evaluation. Our implementation achieved a 1000× speedup compared to standard CPU-based solvers, reducing computation time from approximately 72–0.07 s per batch of 32 city blocks. This acceleration was essential for efficient model training and inference, especially for backpropagation through the physics-informed components.

To ensure numerical stability in the PDE solver, we implemented adaptive time-stepping based on the Courant-Friedrichs-Lewy condition[38], gradient clamping to prevent numerical instabilities, and adaptive parameter bounds based on population magnitude.

Training methodology

Multi-stage curriculum training
We employed a multi-stage curriculum strategy with progressive expert introduction, following principles of curriculum learning[39]:

1. *Pre-training Phase* Initially, only the main neural network component was trained while keeping expert modules fixed.
2. *Expert Training Phase* The colonization expert was introduced first, followed by the PDE expert, and finally the jump and decline experts.
3. *Router Training Phase* Once all experts were trained, the router was fine-tuned to optimize expert assignment.
4. *Full Model Fine-tuning* All components were jointly optimized with a reduced learning rate.

Hyperparameter search framework and result is provided in Supplementary Document Sect. 5.

Loss function design
The model employed a multi-component loss function addressing the heterogeneous nature of population dynamics:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{prediction}} + \lambda_{\text{aux}}\mathcal{L}_{\text{auxiliary}} + \lambda_{\text{reg}}\mathcal{L}_{\text{regularization}} + \lambda_{\text{phys}}\mathcal{L}_{\text{physics}} + \lambda_{\text{cons}}\mathcal{L}_{\text{conservation}}$$

The core prediction loss used a regime-adaptive weighting scheme:

$$\mathcal{L}_{\text{prediction}} = \frac{1}{B \cdot E}\sum_{b=1}^{B}\sum_{e=1}^{E} w_{b,e} \cdot w_e \cdot (\hat{\phi}_{b,e} - \phi_{b,e})^2$$

where $w_{b,e}$ are case-specific weights determined by population regime:

$$w_{b,e} = \{w_{\text{col}} \text{if } \phi_{b,e}^{\text{init}} \leq \theta_{\text{col}} (\text{colonization case}) \; w_{\text{jump}} \text{if } \phi_{b,e}^{\text{change}} \theta_{\text{jump}} (\text{jump case}) \; w_{\text{decline}}$$
$$\text{if } \phi_{b,e}^{\text{init}} \leq \theta_{\text{decline}} \text{ and } \phi_{b,e}^{\text{change}} \leq 0 (\text{decline case}) \; w_{\text{pde}}$$
$$\text{if case meets PDE criteria} (\text{PDE case}) \; 1.0 \text{ otherwise } (\text{regular case})$$

The optimal weight values determined through ablation studies are $w_{\text{col}} = 5.0$, $w_{\text{jump}} = 3.0$, $w_{\text{decline}} = 4.0$, and $w_{\text{pde}} = 2.0$.

For jump cases, we applied an additional log-space transformation to better handle large magnitude changes:

$$\mathcal{L}_{\text{jump}} = (log(\hat{\phi}_{b,e} + 1) - log(\phi_{b,e} + 1))^2 \cdot 100.0$$

The auxiliary loss guides the expert router to correctly classify population dynamics:

$$\mathcal{L}_{\text{auxiliary}} = -\frac{1}{B \cdot E}\sum_{b=1}^{B}\sum_{e=1}^{E}\sum_{c=1}^{C} y_{b,e,c} \log(p_{b,e,c})$$

The physics-informed loss enforces consistency with the underlying PDE:

$$\mathcal{L}_{\text{physics}} = \frac{1}{B \cdot E}\sum_{b=1}^{B}\sum_{e=1}^{E} 1_{\text{isPDE}}(b,e) \cdot \| \frac{\partial \phi_{b,e}}{\partial t} - F(\phi_{b,e}, \nabla\phi_{b,e}, \nabla^2\phi_{b,e}) \|_2^2$$

Optimization strategy
We used Adam optimizer[40] with weight decay regularization and a learning rate schedule combining warm-up and cosine annealing[41]:

$$\eta_t = \begin{cases} \eta_{\max} \cdot \frac{t}{T_{\text{warmup}}} \text{if } t \leq T_{\text{warmup}} \\ \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\pi \frac{t - T_{\text{warmup}}}{T - T_{\text{warmup}}}\right)\right) \text{otherwise} \end{cases}$$

where T_{warmup} is the warm-up period length (typically 5 epochs), T is the total number of epochs, $\eta_{\max}$ is the maximum learning rate (0.0005), and $\eta_{\min}$ is the minimum learning rate (0.00001).

To ensure training stability, we implemented adaptive gradient clipping with threshold:

$$\tau = max(0.1, min(5.0, \frac{\|\theta\|_2}{100})) \cdot \tau_{\text{base}}$$

with $\tau_{\text{base}} = 5.0$ as the base clipping threshold.

Cross-temporal validation
To evaluate the model's ability to predict across different temporal intervals, we employed a cross-temporal validation strategy:

1. *Within-Period Validation* For each temporal period (2001–2006, 2006–2016, 2016–2021), we trained separate models using 80/20 train-test splits of dissemination areas from that specific period. This approach,

which forms the basis for results in Table 1, ensures fair comparison across model architectures under consistent socioeconomic conditions by isolating model quality from temporal transfer effects.

2. *Cross-Period Validation* We evaluated models trained on one period against test data from different periods to assess temporal generalization across nine distinct train-test combinations. Results reveal temporal resolution as the dominant factor affecting cross-period transfer. When PDE time steps match (5→5 year), models achieve strong bidirectional generalization: the 2016–2021 model maintains $R^2 = 0.74$ on 2001–2006 data (9% degradation), while the 2001–2006 model achieves $R^2 = 0.70$ on 2016–2021 data (14% degradation), demonstrating robust performance across 15–20 year temporal gaps with a modest 4% asymmetry favoring recent models. However, temporal resolution mismatches cause substantial degradation: 5→10 year transfers achieve $R^2 = 0.25$–0.43 (47–67% degradation), while 10→5 year transfers achieve $R^2 = 0.45$–0.54 (29–44% degradation). This sensitivity reveals a critical operational constraint: models must be trained on periods matching intended prediction horizons for reliable performance (see Supplementary Materials Sect. 3.2.6 and Table S1 for complete validation matrix).
3. *Multi-Period Training* We explored training on combined data from multiple periods (2001–2006 + 2006–2016) to assess benefits of incorporating diverse temporal dynamics. The multi-period model achieved $R^2 = 0.68$ on 2016–2021 test data, representing intermediate performance between temporally-matched single-period models ($R^2 = 0.74$) and mismatched transfers ($R^2 = 0.25$–0.45). This demonstrates feasibility but reveals that temporal resolution matching provides clearer benefits than expanded training data, establishing period-specific training as the preferred approach for within-period prediction accuracy. The comprehensive validation results (Supplementary Table S1) demonstrate that MESMoE achieves excellent within-period performance ($R^2 = 0.71$–0.81) across three distinct socioeconomic eras while maintaining reasonable cross-period generalization when temporal resolution and feature availability are matched. Detailed analysis of cross-period performance patterns, architectural considerations, and practical application guidelines are provided in Supplementary Materials Sect. 3.2.6.

For each temporal period, we used an 80/20 train/test split, with the training data further divided into 90% training and 10% validation. Model performance was evaluated using multiple metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and coefficient of determination (R^2), both overall and disaggregated by ethnicity and population dynamic regime. It is important to note that our validation approach assesses predictions at census intervals (5 or 10 years) due to ground truth data availability, rather than at intermediate time points. While the Cultural-Spatial Resonance Network's continuous PDE formulation theoretically enables predictions at arbitrary time points through its Runge–Kutta integration, census data—the gold standard for population statistics—is only collected at 5-year intervals, providing ground truth observations exclusively at census years. The model predicts population distributions at the next census time point given the current census as input, but intermediate-year predictions (e.g., 2003, 2004) cannot be empirically validated against observed data. This validation constraint reflects: (1) the discrete nature of census data availability, which provides no ground truth for intermediate years, (2) the resulting inability to train or validate the model on sub-census-interval predictions, and (3) the alignment with urban planning practice, where major policy decisions typically operate on 5–10 year planning horizons synchronized with census data releases. Future work incorporating auxiliary data sources (annual population estimates, building permits, administrative records) could enable intermediate-year validation, but current empirical validation necessarily focuses on census endpoints where observed data exists.

Interpretability analysis framework
We developed a comprehensive interpretability framework to analyze the learned model components:

1. *Feature Importance Analysis* Computed based on the magnitude of weights in the first layer of each expert module[42].
2. *Expert Contribution Analysis* Quantifying the relative contribution of each expert to the final predictions.
3. *Ethnicity Interaction Analysis* Measuring the strength and direction of inter-ethnic influences through interaction matrices.
4. *Spatial Parameter Visualization* Creating high-resolution maps of physics-informed parameters across urban space.

The framework enables quantitative assessment of the socioeconomic drivers of population dynamics, the relative importance of different demographic mechanisms, and the spatial heterogeneity of population processes across urban neighborhoods.

Data availability

The source code for the Multi-Ethnic Spatial Mixture of Experts (MESMoE) framework, including the CUDA-accelerated implementation and all analysis scripts, is publicly available on GitHub at https://github.com/navid-nsk/mesmoe-csrn-2016-2021. The spatial demographic dataset for Toronto at the dissemination area scale, including census data and associated socioeconomic features used in this study, is available on Figshare at https://figshare.com/articles/dataset/Spatial_data_for_Toronto_in_dissemination_area_scale/28912400. Both the code repository and dataset are released under open licenses to ensure reproducibility and facilitate future research.

Received: 5 June 2025; Accepted: 30 October 2025
Published online: 28 November 2025

References

1. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* **1**, 206–215 (2019).
2. Gunning, D. Broad agency announcement explainable artificial intelligence (XAI). *Defense Adv Res Proj Agency (DARPA)* **2**(2), 1 (2016).
3. Kruschel, S. *et al.* Challenging the performance-interpretability trade-off: An evaluation of interpretable machine learning models. *Business & Information Systems Engineering*, 1–25 (2025).
4. Lovo, A., Lancelin, A., Herbert, C. & Bouchet, F. Tackling the accuracy-interpretability trade-off in a hierarchy of machine learning models for the prediction of extreme heatwaves. arXiv preprint arXiv:2410.00984 (2024).
5. Rudin, C. & Radin, J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Sci Rev* **1**, 1–9 (2019).
6. Luo, Y. et al. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR| Open* **1**, 20190021 (2019).
7. Kim, Y., Safikhani, A. & Tepe, E. Machine learning application to spatio-temporal modeling of urban growth. *Comput. Environ. Urban Syst.* **94**, 101801 (2022).
8. Zech, J. R. et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* **15**, e1002683 (2018).
9. Triantakonstantis, D. & Mountrakis, G. Urban growth prediction: a review of computational models and human perceptions. (2012).
10. Li, D., Yu, Y. & Wang, B. Urban population prediction based on multi-objective lioness optimization algorithm and system dynamics model. *Sci. Rep.* **13**, 11836 (2023).
11. Yang, J., Shi, Y., Zheng, Y. & Zhang, Z. The spatiotemporal prediction method of urban population density distribution through behaviour environment interaction agent model. *Sci. Rep.* **13**, 5821 (2023).
12. Zhang, N., Zhang, Y. & Lu, H. Seasonal autoregressive integrated moving average and support vector machine models: prediction of short-term traffic flow on freeways. *Transp. Res. Rec.* **2215**, 85–92 (2011).
13. Fan, Z., Song, X., Shibasaki, R. & Adachi, R. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. 559–569.
14. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F. In: SEBD 2010-Proceedings of the 18th Italian symposium on advanced database systems. 134–141 (Esculapio Editore).
15. Gounaridis, D., Chorianopoulos, I., Symeonakis, E. & Koukoulas, S. A random forest-cellular automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. *Sci. Total Environ.* **646**, 320–335 (2019).
16. Tsagkis, P., Bakogiannis, E. & Nikitas, A. Analysing urban growth using machine learning and open data: An artificial neural network modelled case study of five Greek cities. *Sustain. Cities Soc.* **89**, 104337 (2023).
17. Fan, C., Xu, J., Natarajan, B. Y. & Mostafavi, A. Interpretable machine learning learns complex interactions of urban features to understand socio-economic inequality. *Comput-Aided Civ Infrastruct Eng* **38**, 2013–2029 (2023).
18. Liu, Z. & Liu, C. The association between urban density and multiple health risks based on interpretable machine learning: A study of American urban communities. *Cities* **153**, 105170 (2024).
19. Mazzoli, M. et al. Field theory for recurrent mobility. *Nat. Commun.* **10**, 3895 (2019).
20. Shazeer, N. *et al.* Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538 (2017).
21. Batty, M. *The new science of cities* (MIT press, Cambridge, 2013).
22. Fotheringham, A. S., Brunsdon, C. & Charlton, M. Geographically weighted regression. *Sage Handb Spatial Anal* **1**, 243–254 (2009).
23. Reia, S. M., Rao, P. S. C. & Ukkusuri, S. V. Modeling the dynamics and spatial heterogeneity of city growth. *NPJ Urban Sustain* **2**, 31 (2022).
24. Wilson, A. *Knowledge power: Interdisciplinary education for a complex world* (Routledge, Milton Park, 2010).
25. Wu, P., Zhang, Z., Peng, X. & Wang, R. Deep learning solutions for smart city challenges in urban development. *Sci. Rep.* **14**, 5176 (2024).
26. Brown, L. A. & Sanders, R. L. *In migration decision making* 149–185 (Elsevier, Amsterdam, 1981).
27. Simini, F., Barlacchi, G., Luca, M. & Pappalardo, L. A deep gravity model for mobility flows generation. *Nat. Commun.* **12**, 6576 (2021).
28. Anselin, L. *Spatial econometrics: methods and models* Vol. 4 (Springer Science & Business Media, Berlin, 1988).
29. Jacobs, R. A., Jordan, M. I., Nowlan, S. J. & Hinton, G. E. Adaptive mixtures of local experts. *Neural Comput.* **3**, 79–87 (1991).
30. Jordan, M. I. & Jacobs, R. A. Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* **6**, 181–214 (1994).
31. Rogers, E. M., Singhal, A. & Quinlan, M. M. *In an integrated approach to communication theory and research* 432–448 (Routledge, Amsterdam, 2014).
32. Davis, M. H. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **46**, 353–376 (1984).
33. Frey, W. H. Central city white flight: Racial and nonracial causes. *American Sociological Review*, 425–448 (1979).
34. Press, W. H. *Numerical recipes 3rd edition: The art of scientific computing* (Cambridge University Press, Cambridge, 2007).
35. Belkin, M. & Niyogi, P. Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. Syst. Sci.* **74**, 1289–1308 (2008).
36. Vaswani, A. Attention is all you need. arXiv preprint arXiv:1706.03762 (2017).
37. Schelling, T. C. Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971).
38. Courant, R., Friedrichs, K. & Lewy, H. On the partial difference equations of mathematical physics. *IBM J. Res. Dev.* **11**, 215–234 (1967).
39. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. In: Proceedings of the 26th annual international conference on machine learning. 41–48.
40. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
41. Loshchilov, I. & Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016).
42. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

Acknowledgements

The authors wish to express their sincere gratitude to Professor Michael Sawada of the Department of Geography, University of Ottawa, for his invaluable assistance and guidance in the development of the model framework. His expertise in spatial analysis and demographic modeling contributed significantly to this research. We are deeply grateful to the two anonymous reviewers whose thorough and constructive critiques led to substantial improvements in this manuscript. Their insightful comments regarding baseline model comparisons, temporal validation scope, numerical stability mechanisms, and clarity of presentation have enhanced both the scientific

rigor and accessibility of our work. We also thank the editorial team at Scientific Reports for their careful handling of our manuscript and clear guidance throughout the review process.

Author contributions

S.N.M.M. conceived the study, designed the Multi-Ethnic Spatial Mixture of Experts (MESMoE) framework, developed the physics-informed modeling approach, implemented the CUDA-accelerated algorithms, performed all data analysis, created the interpretability framework, generated all figures, and wrote the manuscript. H.C. supervised the project, contributed to the conceptual framework, provided critical feedback on the methodology, and reviewed and edited the manuscript. T.J. contributed to the mathematical formulation of the Cultural-Spatial Resonance Network, assisted with the PDE solver design, helped develop the mixture of experts architecture, and contributed to writing the Methods section. R.H. participated in code implementation, conducted model validation experiments, performed cross-temporal validation analysis, and verified the reproducibility of results. D.R. contributed to software development, implemented visualization tools for spatial parameter analysis, created data processing pipelines, generated supplementary visualizations, and contributed to writing the Results section.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that no funding was received for the conduct of this study or preparation of this manuscript.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-26832-1.

Correspondence and requests for materials should be addressed to S.N.M.M. or H.C.