Bridging Tradition and Modernity: The Application and Reflection of Information Technology in Census Practices

Jiahui Chen[a]

College of Art & Sciences, Department of Communication, University of Washington, Seattle, U.S.A.

Keywords: Population Forecasting, Demographic Models, Small-Area Projections, Machine Learning, Probabilistic Methods.

Abstract: Countries and governments need to regularly understand population changes and predict population trends. This is very important for planning housing, employment, health care and other public services. Population forecasting helps to reasonably and accurately allocate resources to maximize utilization. Traditional statistical models such as cohort distribution cannot explain population trends caused by global events such as politics, economy, and immigration. LSTM networks provide a computational method for sustainable observation to adapt to complex population patterns. Artificial intelligence and scientific statistics have made predictions easier, but they still require a lot of data support. Population forecasting and census are more complex issues. This paper studies the limitations and reliability of different measurement methods, and considers other external factors such as climate, policy, and economic factors that affect population size, such as real challenges. By studying and summarizing these factors, this study will emphasize the differences in forecasting methods and consider how to improve the accuracy and adaptability of population forecasts and censuses in the future.

1 INTRODUCTION

Being able to predict population trends is crucial for governments, businesses, and planners. Without accurate forecasts, cities might not build enough schools, hospitals, or housing to keep up with demand. Businesses also rely on population predictions to decide where to expand or what products to offer. Governments use these forecasts to plan long-term policies on infrastructure, healthcare, and economic development.

For decades, most population predictions were done using demographic models, like the cohort-component method, which estimates future populations based on birth rates, death rates, and migration patterns(Aryal, 2020). These models work well for large areas with stable trends but don't always perform as well for smaller regions or places experiencing sudden changes, like economic booms or natural disasters. That's why researchers have been developing more advanced forecasting methods that consider uncertainty, spatial variation, and machine learning techniques(Yu et al., 2023).

Population changes vary widely depending on location. In some places, like South Korea, aging populations are a major concern, and forecasting helps governments prepare for rising healthcare costs and pension demands (Kim & Kim, 2020).

Meanwhile, in fast-growing urban areas, projections need to factor in high migration rates and shifts in birth patterns (Sang et al., 2024). Because of these differences, improving forecasting methods is necessary to ensure accurate predictions for different regions and circumstances.

There exist a number of methodologies employed in population projection with their respective merits and demerits. Aryal utilized the cohort-component method, one of the most popular demographic approaches, to project future populations by an analysis of historical data of birth rates, death rates, and migration(Aryal, 2020). This approach has shown itself to be trustworthy for forecasting long- term trends, especially in settled areas, but is frequently not able to manage unexpected demographic disruptions or sudden shifts, particularly those instigated by political or environmental crises.

[a] https://orcid.org/0009-0008-8576-5614

Chen, J.
Bridging Tradition and Modernity: The Application and Reflection of Information Technology in Census Practices.
DOI: 10.5220/0013688000004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 285-291
ISBN: 978-989-758-765-8

To help alleviate the limitations of single-point projections, Vollset et al. put forward a probabilistic projection framework that could produce a range of possible population scenarios. The study, which included 195 countries and territories, demonstrated that the inclusion of uncertainty regarding fertility and migration substantially improves accuracy at both the national and regional levels (Vollset, 2020).

Pointing to the difficulties with making projections at more localized levels, Kim and Kim and Sang et al. considered small-area and spatial projection models (Kim & Kim, 2020; Sang et al., 2024). These models synthesize localized trends and geographic information to make improved predictions for individual municipalities or counties, especially where population trends are very different from the national average.

In the last several years, scholars have increasingly applied artificial intelligence to improve population projection. Grossman et al. applied long short-term memory (LSTM) neural networks to small-area forecasting and concluded that AI-based models have the potential to outperform conventional methods in uncovering intricate patterns in big data (Grossman et al., 2023). Nevertheless, the success of such approaches hinges greatly on data quality and availability, and the interpretability of such models continues to challenge demographers and policymakers.

This paper will explore different population forecasting techniques and compare their effectiveness. The main areas of focus include:

1. How traditional demographic models compare to newer probabilistic and AI-driven techniques.
2. The challenges in forecasting smaller populations and improving prediction accuracy at the local level.
3. The potential for AI to enhance population forecasting and whether it can work alongside traditional models.

By analyzing these methods, this paper aims to provide insight into the best approaches for different forecasting needs and how they can be refined. Since reliable population predictions are necessary for effective decision-making, improving forecasting techniques will remain an important area of research.

2 RESEARCH METHODS AND FORECASTING MODELS

Population forecasting has evolved significantly over the past decades, expanding from deterministic demographic models to probabilistic approaches and artificial intelligence (AI)-driven algorithms. This section systematically reviews three main methodological categories: traditional cohort-based projections, probabilistic and spatial forecasting techniques, and machine learning-based models. Each has unique strengths and weaknesses depending on the forecasting scale, data availability, and policy applications.

2.1 Cohort-Based and Structured Demographic Models

Traditional demographic forecasting often begins with the cohort-component model, which segments the population by age and sex and applies projected fertility, mortality, and migration rates to simulate future changes. Aryal outlines the foundational role this method plays in national statistics offices, noting its relative simplicity and strong performance in contexts with stable demographic trends (Arya, 2020). It is widely adopted due to its transparency and compatibility with census data.

However, its limitations have become increasingly apparent, particularly in cases where rapid change or local heterogeneity undermines the assumption of linearity. Ellner et al. expand on these structural models by comparing integral projection models (IPMs) and matrix population models (MPMs), arguing that while both frameworks offer mathematical precision, they assume stable environments and are thus ill-equipped to model abrupt demographic shifts caused by unexpected events, such as the COVID-19 pandemic or sudden policy shifts (Ellner et al., 2022).

Kim and Kim highlighted these limitations in their sub-national study of South Korea's aging population. Their research revealed that while national-level projections often suggest moderate trends, local regions experience much sharper demographic transitions, requiring forecasting methods that account for age-specific dynamics and spatial differentiation (Kim & Kim, 2020).

Overall, while structured demographic models remain a cornerstone in official forecasting, their reliance on historic trend stability, and their limited adaptability to granular regional differences, limits their utility in increasingly volatile demographic landscapes.

2.2 Probabilistic Forecasting and Spatial Models

In contrast to deterministic models, probabilistic forecasting techniques explicitly account for uncertainty in demographic variables. Rather than offering a single-point estimate, these models produce distributions of outcomes based on simulations or Bayesian inference frameworks. Vollset et al. introduced a comprehensive probabilistic approach for 195 countries in the Global Burden of Disease study, modeling population scenarios from 2017 to 2100 (Vollset et al., 2020). Their framework incorporates stochastic variation in fertility, mortality, and migration assumptions, yielding more informative forecasts, especially for long-term planning.

Yu et al. extend probabilistic modeling to the county level in the U. S., emphasizing that small-area forecasts are highly sensitive to local migration patterns and fertility trends (Yu et al., 2023). By using Bayesian hierarchical models, they are able to "borrow strength" from neighboring areas to improve the robustness of projections in data-sparse regions. Their work demonstrates how probabilistic models are adaptable to different scales, providing a more flexible alternative to rigid deterministic structures.

Complementary to probabilistic models are spatially explicit projections, which integrate geographical variation into forecasting. Sang et al. developed a county-level population projection framework in China that accounts for both spatial and temporal dynamics using fine-grained geographic data (Sang et al., 2024). Their findings illustrate that spatially-aware models can capture urbanization trends, population clustering, and regional disparities more effectively than conventional national-level approaches. Chen et al. similarly adopted gridded projections under shared socioeconomic pathways (SSPs) for China, creating high-resolution future population scenarios aligned with global climate frameworks (Chen et al., 2020).

Despite their strengths, probabilistic and spatial models face challenges in data requirements, computational complexity, and interpretability. Many countries lack consistent subnational demographic data, making localized modeling difficult. Moreover, policymakers may struggle to translate probabilistic ranges into actionable decisions without additional interpretive tools.

2.3 Machine Learning and AI-Enhanced Forecasting

Recent years have witnessed a growing interest in machine learning (ML) and artificial intelligence (AI) approaches in population forecasting, especially where data complexity and volume exceed the capabilities of traditional models. One prominent method is the Long Short-Term Memory (LSTM) neural network, which excels at modeling temporal dependencies in time series data. Grossman et al. applied LSTM to small-area population forecasting in Australia and demonstrated that these models outperform conventional techniques in terms of predictive accuracy, particularly for short-term forecasts (Grossman et al., 2023).

LSTM models are capable of learning non-linear relationships and incorporating lagged effects, making them suitable for detecting sudden demographic shifts or non-monotonic migration flows. Their adaptability is especially useful in small-area forecasts, where local trends may diverge substantially from national averages. However, Grossman and colleagues caution that LSTMs require large, high-quality training datasets and may be difficult to interpret—posing barriers for their integration into official statistical systems.

Beyond LSTMs, other AI models such as decision trees, random forests, and support vector machines have also been applied in demographic contexts. For instance, Papastefanopoulos et al. evaluated multiple time series models for forecasting COVID-19 case proportions and demonstrated that machine learning techniques, including LSTMs and autoregressive models, outperformed classical statistical models during rapidly evolving public health scenarios (Papastefanopoulos et al., 2020). Their work illustrates the potential transferability of these models to population forecasting during periods of demographic disruption.

Despite their promise, AI models are not without caveats. O'Sullivan warned that overreliance on opaque models could exacerbate demographic misinterpretation, particularly when policymakers treat forecasts as deterministic outcomes. Additionally, AI techniques often require extensive computational resources and suffer from "black box" issues, making them difficult to audit or validate through traditional demographic lenses (O'Sullivan, 2023).

To mitigate these concerns, researchers have proposed hybrid models that integrate traditional demographic techniques with machine learning enhancements. For example, baseline population projections can be generated using cohort-component models and adjusted dynamically using LSTM networks fed with real-time migration, birth, and administrative data. Such integration improves both the accuracy and responsiveness of forecasts.

3 LITERATURE ANALYSIS: APPLICATIONS AND FORECASTING RESULTS

3.1 Datasets Used

The effectiveness of any population forecasting model depends significantly on the quality, resolution, and granularity of the datasets used. Across the reviewed literature, we observe a diversity of data sources ranging from national census and administrative registers to high-resolution spatial grids and real-time datasets.

Traditional demographic projections, as represented by Aryal, mainly rely on census data, vital registration systems (including birth and death registrations), and migration statistics aggregated at the national or subnational levels. These datasets are typically collected every ten years and constitute the basis for organized forecasting models, such as cohort-component projections. While these sources are marked by standardization and consistency, they often suffer from a lack of timeliness and are not sufficient to capture rapid changes in population dynamics, especially in times of crisis.

Probabilistic and spatial modeling approaches, which are represented by the research works of Vollset et al. and Chen et al. utilize more extensive demographic and socioeconomic information. For example, the Global Burden of Disease study utilized the World Population Prospects, United Nations projections, and national statistical databases to make long-term demographic projections for 195 countries. Chen et al. extended this methodology in the Chinese setting by integrating gridded population data with Shared Socioeconomic Pathways (SSPs) to facilitate high-resolution projections under alternative climate and development futures. Such models call for harmonization of heterogeneous data sources, subject to intricate preprocessing and spatial coordination procedures (Vollset et al., 2020; Chen et al., 2020).

Spatial studies, such as Sang et al., relied on county-level population data in China combined with satellite-based urbanization metrics and geocoded administrative boundaries (Sang et al., 2024). Such fine-grained datasets are essential for modeling urban growth and spatial population distribution, but they are limited by availability and standardization, especially across developing countries.

AI and machine learning methods rely even more heavily on comprehensive and real-time data. Grossman et al. used small-area data from Australia's national statistical agency, which included historical population counts and migration flows over multiple decades (Grossman et al., 2023). The LSTM model they developed required time series input data structured into temporal windows, making continuous, high-frequency datasets essential for model training. Similarly, Papastefanopoulos et al. evaluated forecasting models using COVID-19 case data expressed as percentages of active cases per population. The time-sensitive nature of such data reflects the strengths of AI in responding to fast-moving demographic phenomena.

Overall, data quality and availability remain critical barriers to forecasting performance. While traditional models tolerate coarse, static data, probabilistic and AI-based approaches demand detailed, structured, and consistent inputs that are not universally accessible.

3.2 Comparative Analysis of Forecasting Results

The reviewed studies highlight distinct performance characteristics across different forecasting techniques, each with context-dependent strengths.

In deterministic models such as the cohort-component approach, Aryal found that forecasts were generally accurate in countries with stable population structures and low migration volatility (Aryal, 2020). However, these models underperformed in regions experiencing sudden demographic shifts. For example, Kim and Kim showed that even when applied sub-nationally, deterministic projections underestimated the speed of population aging in South Korea's rural areas (Kim & Kim, 2020). The model's simplicity, while advantageous for interpretability, leads to rigidity in uncertain or rapidly evolving conditions.

Probabilistic models introduce robustness by providing forecast intervals rather than single estimates. Vollset et al. demonstrated that this approach reduced errors in long-term global projections. Their simulations, incorporating uncertainty in fertility and migration, produced more realistic estimates, especially for developing countries with unstable demographic indicators (Vollset et al., 2020). Yu et al. validated this strength at the county level, where their Bayesian model improved forecast precision through spatial smoothing and uncertainty quantification (Yu et al., 2023).

Spatial models also offer improvements in forecast precision at fine geographical scales. Sang et al. showed that combining spatial autocorrelation structures with temporal trend modeling allowed their model to better capture migration-driven population clustering in China (Sang et al., 2024). Chen et al. 's gridded projections further illustrate how the spatial layout of population growth varies significantly under different socioeconomic scenarios. However, these models often require substantial computational resources and geographic data integration expertise (Chen et al., 2020).

Machine learning techniques have also demonstrated improved short-term forecasting in small-area settings. Grossman et al. likened predictions using LSTMs to conventional statistical models and observed that the AI approach lowered mean absolute percentage error (MAPE) by up to 30% for certain regions. This is attributed to the model's ability to learn nonlinear patterns and lagged effects over time (Grossman et al., 2023). But the trade-off is one of decreased transparency—AI predictions are frequently opaque about causal reasons, with ominous implications for use in policy-sensitive areas.

Papastefanopoulos et al. also reported AI models outperforming classical time series approaches like ARIMA in capturing the effects of the pandemic on population movement. They, however, insisted on the importance of model validation and warned against blind trust in accuracy metrics (Papastefanopoulos et al., 2020).

O'Sullivan presents a critical perspective, cautioning that excessively positive population forecasts—regardless of whether they are founded on conventional or AI-driven models—might overlook significant long-term sustainability issues. His argument highlights the need to connect model outputs with inclusive policy development and scenario evaluation (O'Sullivan, 2023).

In summary, forecasting accuracy and adaptability vary across models:

Deterministic models provide transparency and simplicity but perform poorly under uncertainty.

Probabilistic and spatial models improve reliability and geographic resolution but require complex data and calibration.

AI models deliver higher predictive accuracy, especially short-term, but face interpretability and data availability issues.

These trade-offs suggest that no single method is universally superior. Instead, hybrid strategies combining structured demographic reasoning with adaptive AI tools may provide the most effective path forward.

4 RECOMMENDATIONS AND FUTURE DIRECTIONS

The comparative analysis of existing forecasting methods reveals a dynamic and rapidly evolving field, but it also underscores persistent limitations that hinder the development of robust, responsive, and policy-relevant population projections. To move the field forward, several key areas warrant attention.

4.1 Addressing Data Limitations and Standardization

A consistent issue spanning traditional, probabilistic, and AI - based population forecasting methods is their reliance on high - quality, uniform data. Aryal points out that even the well - established cohort - component models encounter difficulties when dealing with incomplete vital statistics or inconsistent census intervals (Aryal， 2020). This problem is far more acute in machine learning models. For instance, as Grossman et al. show, in such models, real - time and temporally consistent input data are essential for model training and validation (Grossman et al., 2023).

To foster progress in these methodologies, national statistical agencies and international bodies should make data modernization a priority. This involves digitizing records, enhancing population estimates between censuses, and implementing open data standards. The effectiveness of probabilistic models, such as those developed by Vollset et al. and Yu et al., hinges largely on having detailed and consistent input variables across various regions and time periods (Vollset et al., 2020; Yu et al., 2023).

Moreover, global investments in high - resolution spatial data, similar to those used in the studies by Chen et al. and Sang et al., can enhance small - area projections and allow for integration with environmental, economic, and urban planning models. The future of population forecasting lies in the real - time integration of data from multiple sources (Chen et al., 2020; Sang et al., 2024).

4.2 Enhancing Model Interpretability and Transparency

AI and machine learning models have shown great potential in improving short-term forecasting accuracy, especially in small-area contexts (Grossman et al., 2023; Papastefanopoulos et al., 2020). However, their "black-box" nature often makes them unsuitable for high-stakes policy environments, where decision-makers require clear, interpretable evidence for interventions.

Developing explainable AI (XAI) frameworks tailored to demographic forecasting is therefore essential. These could involve hybrid approaches, where traditional demographic structures are used to anchor AI models, offering both predictive power and theoretical transparency. For example, using LSTM-based adjustments on top of deterministic cohort estimates may allow practitioners to preserve interpretability while enhancing responsiveness to new data inputs.

Moreover, as O'Sullivan cautions, overconfidence in model precision—whether from classical or AI-based methods—can distort policy planning and delay recognition of demographic risks. Researchers should routinely publish uncertainty estimates, model assumptions, and validation metrics alongside forecasts (O'Sullivan, 2023).

4.3 Bridging Scale and Scenario Gaps

Another persistent challenge is the mismatch between global or national forecasting and the needs of local policymakers. As demonstrate in their study of South Korea, national trends often conceal sharp subnational divergence, particularly in aging, urbanization, and fertility. To address this, models must become more spatially adaptive, integrating geographic heterogeneity and local socioeconomic indicators.

Scenario-based forecasting also deserves more attention. While Chen et al. adopted Shared Socioeconomic Pathways (SSPs) in population modeling for China, few national statistical agencies implement scenario planning into their population projections. Probabilistic methods and spatial models are particularly well-suited to scenario modeling, offering a pathway to incorporate uncertainty in fertility preferences, climate impact, and migration policy shifts (Chen et al., 2020).

Developing standardized scenario frameworks, similar to those in climate modeling, could significantly enhance the robustness and relevance of demographic forecasts.

4.4 Strengthening Interdisciplinary Integration

Demographic forecasting must increasingly operate at the intersection of public health, climate science, urban planning, and artificial intelligence. Vollset et al. provide an excellent model by integrating population projections into health burden forecasting (Vollset et al., 2020), while Sang et al. show how urban expansion models can inform localized demographic dynamics (Sang et al., 2024).

Future forecasting frameworks should support plug-and-play interoperability with other models—such as those used in epidemiology, infrastructure planning, and environmental simulation. This requires not only methodological compatibility but also institutional collaboration and shared data platforms.

4.5 Supporting Global Equity in Forecasting Capacity

Many of the advanced models and datasets examined in this paper originate from high - income countries boasting well - established statistical systems. However, as global demographic trends increasingly center on low - and middle - income regions, it is crucial to enhance forecasting capabilities in areas with limited data.

Efforts to create open - source modeling tools, implement capacity - building programs for national statistics offices, and establish collaborative international datasets (such as those in the Global Burden of Disease study) are essential. These initiatives ensure that all countries, regardless of their income level or data availability, can generate population forecasts that are both reliable and relevant to policy - making.

5 CONCLUSIONS

Population forecasting is vital for planning infrastructure, healthcare, labor markets, and social services. In a world of demographic uncertainty and rapid global change, accurate, adaptable, and clear forecasting tools are urgently needed. This paper compares three main population forecasting methods: traditional demographic models, probabilistic and spatial models, and machine - learning - based ones. Each method has its own strengths, depending on application scale, data availability, and population trend volatility.

Traditional cohort - component models are fundamental as they are interpretable and work well with census data. However, they often can't capture sudden demographic shifts, like those from unexpected migrations. Also, they struggle with local - level variations where regional differences in economic or cultural factors affect population trends.

Probabilistic approaches, such as those by Vollset et al. and Yu et al., address these issues. By factoring in uncertainty, they provide a range of possible future population scenarios, thus enhancing forecast reliability, especially at national and regional levels. Spatial forecasting and gridded projections enable more detailed demographic planning, crucial for rapidly urbanizing, diverse countries. Machine - learning techniques like LSTM networks are good at catching non - linear trends for small - area short - term forecasts. But, as Papastefanopoulos et al. and O'Sullivan note, they have challenges in data needs and interpretability.

Integrating these methods holds promise. Hybrid models combining demographic theory and AI can offer machine - learning accuracy and traditional - model transparency. Improving data infrastructure, investing in open - access tools, and developing explainable AI will boost forecasting systems. In essence, refining population forecasting is a societal must for creating resilient, inclusive 21st - century policies as demographic shifts shape our future.

REFERENCES

Aryal, G. R. 2020. Methods of Population Estimation and Projection. Journal of Population and Development, 1(1), 54–61.

Chen, Y., Guo, F., Wang, J., & et al. 2020. Provincial and gridded population projection for China under shared socioeconomic pathways from 2010 to 2100. Scientific Data, 7, 83.

Ellner, S. P., Adler, P. B., Childs, D. Z., Hooker, G., Miller, T. E. X., & Rees, M. 2022. A critical comparison of integral projection and matrix projection models for demographic analysis: Comment. Ecology, 103(10), 1–7.

Grossman, I., Wilson, T., & Temple, J. 2023. Forecasting small area populations with long short-term memory networks. Socio-Economic Planning Sciences, 88, 101658.

Kim, K. W., & Kim, O. S. 2020. Super aging in South Korea unstoppable but mitigatable: A sub-national scale population projection for best policy planning. Spatial Demography, 8(2), 155–173.

O'Sullivan, J. N. 2023. Demographic Delusions: World Population Growth Is Exceeding Most Projections and Jeopardising Scenarios for Sustainable Futures. World, 4(3), 545-568.

Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. 2020. COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. Applied Sciences, 10(11), 3880.

Sang, M., Jiang, J., Huang, X., & et al. 2024. Spatial and temporal changes in population distribution and population projection at county level in China. Humanities and Social Sciences Communications, 11, 288.

Vollset, S. E., Goren, E., Yuan et al., 2020. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease Study. The Lancet, 396(10258), 1285–1306.

Yu, C. C., Ševčíková, H., Raftery, A. E., & Curran, S. R. 2023. Probabilistic county-level population projections. Demography, 60(3), 915–937.