A METHODOLOGICAL FRAMEWORK FOR NEURAL-NETWORK-BASED LEARNING OF MULTI-FACTOR SPATIAL POPULATION DYNAMICS

ZHONGLI YAN

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

YORK UNIVERSITY
TORONTO, ONTARIO

JULY 2026

© ZHONGLI YAN, 2026

Abstract

Urban population movement is a dynamic spatial process shaped by employment, commercial activity, amenities, and transportation accessibility. Traditional spatial economic models offer valuable theoretical foundations but often rely on simplified settings, single-factor mechanisms, or equilibrium analysis. This thesis develops a neural-network-based framework for learning, comparing, and interpreting dynamic spatial population models.

The framework treats population migration as a model-selection and diagnostic problem. It evaluates whether spatial dynamics can be learned from data, identifies relevant factors, compares alternative theory-guided structures, examines how individual factors enter the model, and tests predictive accuracy and stability over time. Results show that the framework successfully learns population dynamics, identifies relevant explanatory factors, distinguishes among dynamic mechanisms, and provides interpretable diagnostics while maintaining stable predictions across updated temporal windows. Overall, the thesis offers a reusable framework for dynamic spatial population modeling that can be extended to real urban settings when suitable integrated spatiotemporal datasets become available.

Abbreviations

NEG	New Economic Geography
CP model	Core–periphery model
PDE	Partial differential equation
MLP	Multilayer perceptron
RNN	Recurrent neural network
CNN	Convolutional neural network
POI	Point of interest
GHSL	Global Human Settlement Layer
OSM	OpenStreetMap
MAE	Mean absolute error
RMSE	Root mean squared error
CV	Coefficient of variation

Table of Contents

Abstract ii
Abbreviations iii
Table of Contents iv
List of Tables vii

1 Introduction 1
1.1 Research Background 1
1.2 Research Motivation 2
1.3 Research Objectives and Contributions 2
1.4 Structure of the Thesis 3

2 Literature Review 5
2.1 NEG and Core–Periphery Foundations 5
2.2 Continuous-Space Migration and Advection–Diffusion Models 6
2.3 Multi-Factor Location Choice and Modeling Gaps 7
2.4 Summary of the Literature Gap 8

3 Neural Network Selection 9
3.1 Problem Description 9
3.2 Model Selection Principle 10
3.3 Comparison of Candidate Neural Network Architectures 11
3.3.1 Multilayer Perceptron 11
3.3.2 Temporal Models: RNN, LSTM, and GRU 11
3.3.3 Convolutional Neural Network 13
3.4 Selected Architecture in This Thesis 14

4 Model Setup 17
4.1 Spatial Domain and Population Dynamics 17
4.2 Input Factor Fields 18
4.3 True Potential Function 21

4.4 Behavior of the True Function 22
4.5 Assessment of the Reasonableness of the Simulated Data 26
4.5.1 Reasonableness of the Data-Generating Logic 26
4.5.2 Reasonableness of the Resulting Population Distribution 27
5 Experimental Pipeline and Results 28
5.1 Dataset Split and Evaluation Metrics 28
5.2 Step 1: Learnability Test 29
5.2.1 Method 1.1: Direct Prediction 29
5.2.2 Method 1.2: Derivative Learning 30
5.2.3 Results and Interpretation 32
5.3 Step 2: Factor-Relevance and Incremental Inclusion Tests 32
5.3.1 Step 2.1: Two-Factor Extension from Job Opportunity 33
5.3.2 Step 2.2: Incremental Multi-Factor Inclusion Test 39
5.3.3 Implications of Step 2 47
5.4 Step 3: Theory-Guided Model-Structure Comparison 48
5.4.1 Method 3.0: Black-Box Baseline 49
5.4.2 Method 3.1: Multi-Factor Advection–Diffusion Model 49
5.4.3 Method 3.2: Multi-Factor Wage-Difference Model 50
5.4.4 Method 3.3: Single-Factor Advection–Diffusion Model 50
5.4.5 Experimental Results 50
5.5 Step 4: Factor-Structure Test 56
5.5.1 Linear Test for w_1 59
5.5.2 Linear Test for w_2 60
5.5.3 Linear Test for w_3 61
5.5.4 Linear Test for w_4 62
5.5.5 Combined Linear Test for w_1 and w_2 63
5.6 Step 5: Rolling-Window Prediction Evaluation 65
5.6.1 Method 5.0: Baseline Multi-Factor Advection–Diffusion Model 66
5.6.2 Method 5.1: Rolling Window 1 66
5.6.3 Method 5.2: Rolling Window 2 67
5.6.4 Method 5.3: Rolling Window 3 67
5.6.5 Method 5.4: Rolling Window 4 68
5.7 Summary of Main Findings 69

6 Why Simulated Data? Real-Data Attempts and Challenges 71
6.1 Global Human Settlement Layer 71
6.2 Mobile-Phone-Based Home and Work Location Studies 72
6.3 OpenStreetMap 74
6.4 Why These Resources Are Not Yet Sufficient and Why Simulated Data Are Used 75
7 Future Extensions 77
7.1 Temporal Distance and Transportation Accessibility 77
7.2 Capacity Constraints and Upper Bounds 77
7.3 Boundary Conditions and Open Systems 78
7.4 Interaction Between Population and Spatial Factors 79
7.5 Advanced Neural Architectures and Algorithmic Improvements 79
7.6 Uncertainty and Stochastic Extensions 80
7.7 Empirical Application 81
A Algorithm Details and Limitations 82
A.1 Algorithm Details 82
A.1.1 Computing Environment and Running Time 82
A.1.2 CNN Architecture and Training 82
A.1.3 Evaluation, Numerical Constraints, and Reproducibility 83
A.2 Limitations of the Model 83
A.3 Limitations of the Dataset 84
B Diffusion: Fick's Law 87
B.1 Diffusion: Fick's Law 87
B.1.1 Fick's First Law 87
B.1.2 Fick's Second Law 87
B.2 Graphical Illustration of Fick's First Law 88
B.3 Graphical Illustration of Fick's Second Law 89
C Advection–Diffusion 91
C.1 Advection–Diffusion 91
C.1.1 Governing Equation 91
C.1.2 Graphical Illustration 92
C.1.3 Explanation 92
Bibliography 94

List of Tables

3.1 Comparison of candidate neural-network architectures 16
5.1 Step 1 results: short-horizon and long-horizon learnability test 32
5.2 Step 2.1 results: two-factor extension from job opportunity 38
5.3 Step 2.2 results: incremental multi-factor inclusion test 46
5.4 Step 3 results: comparison of model structures using MAE 55
5.5 Step 4 results: factor-structure and linearity tests 64
5.6 Step 5 rolling-window design 66
5.7 Step 5 results: baseline and rolling-window prediction performance using MAE . . 68
6.1 Examples of OSM data types and their relevance to spatial factor construction . . 74

Chapter 1

Introduction

1.1 Research Background

New Economic Geography (NEG) provides an important theoretical foundation for explaining the spatial distribution and agglomeration of economic activity. The core–periphery model of Krugman [13] shows how increasing returns to scale, transport costs, and market-size effects may generate spatial concentration of firms and workers. In this framework, agglomeration arises because firms benefit from larger markets and consumers benefit from product variety and employment access. Later work, including Fujita et al. [8], Ottaviano et al. [24], and Pflüger [25], further developed tractable NEG models and clarified the mechanisms behind spatial concentration.

A standard migration formulation in this literature describes population movement as a response to wage or utility differences. A representative reduced-form expression is

$$\frac{\partial \lambda}{\partial t}(t,x) = v\Big[w(t,x) - \bar{w}(t)\Big]\lambda(t,x), \qquad v > 0, \tag{1.1}$$

where $\lambda(t,x)$ denotes population density at location x and time t, $w(t,x)$ is local real wage or attractiveness, and $\bar{w}(t)$ is the average level. This formulation is useful for explaining region-to-region migration, but it does not explicitly describe local spatial movement, moving distance, or the multi-factor nature of residential location choice.

Another line of research moves toward continuous-space analysis. Fujita and Ogawa [7], Anas et al. [2], and Mossay [19] consider spatially distributed incentives and local movement. In an advection–diffusion interpretation, population spreads locally through diffusion and moves directionally under the gradient of an economic or spatial field. A simplified one-

dimensional expression is

$$\frac{\partial \lambda}{\partial t}(t,x) = D\frac{\partial^2 \lambda}{\partial x^2}(t,x) - \frac{\partial}{\partial x}\left(\lambda(t,x)\frac{\partial w}{\partial x}(t,x)\right), \qquad D > 0. \tag{1.2}$$

This perspective is closer to a dynamic continuous-space model, but many formulations remain one-dimensional or single-factor.

1.2 Research Motivation

This thesis is motivated by three limitations in existing modeling approaches.

First, many tractable NEG models focus on a single dominant migration factor, usually job-related attractiveness. In practice, urban population movement is influenced by multiple spatial drivers. Job opportunities, commercial activity, amenities, public services, transportation accessibility, and housing supply may all affect residential and relocation decisions [3, 12, 26, 39, 23]. Therefore, a useful model of urban migration should allow a multi-factor structure.

Second, many theoretical models use one-dimensional or highly simplified spatial settings. This is useful for analytical tractability, but real cities are two-dimensional spatial systems. Employment centers, commercial areas, amenities, transportation corridors, and residential clusters are distributed over a planar domain. A two-dimensional model is therefore more suitable for representing directional movement, spatial clustering, and local interaction in urban space.

Third, much of the theoretical literature emphasizes equilibrium. Equilibrium analysis is important because it describes the long-run spatial configuration of a system under fixed conditions. However, urban development is also a dynamic adjustment process. Population may move gradually over years or decades, while the underlying spatial factors may change before the system reaches equilibrium. Practical decision-making often requires short- and medium-horizon prediction, with new observations incorporated over time. This motivates a dynamic learning framework rather than an equilibrium-only analysis.

1.3 Research Objectives and Contributions

The objective of this thesis is to develop a methodology for studying dynamic spatial population processes, rather than to defend a single governing equation as universally correct. More specifically, the thesis constructs a general pipeline that allows researchers to specify candidate spatial factors, choose one or more candidate dynamic structures, learn the corre-

sponding dynamics from data, and compare the resulting models using both predictive and interpretive criteria.

The thesis makes four contributions.

First, it formulates population dynamics as a general model-selection and model-diagnosis problem. A candidate model may be written in the broad form

$$\partial_t \lambda(t,x) = \mathcal{M}\Big(\lambda(t,x), W(t,x)\Big), \tag{1.3}$$

where λ is population density, $W = (w_1, \ldots, w_n)$ is a collection of spatial factors, and $\mathcal{M}$ denotes a candidate dynamic operator. The operator $\mathcal{M}$ may be a black-box neural network, a wage-difference model, a single-factor advection–diffusion model, a multi-factor advection–diffusion model, or another theory-guided structure.

Second, it proposes a reusable experimental sequence: first test learnability, then test factor relevance, then compare candidate model structures, then examine factor structure, and finally evaluate prediction stability. This sequence is designed to be applicable beyond the specific simulation used in this thesis.

Third, it demonstrates the methodology using a two-dimensional multi-factor advection–diffusion system. In this demonstration, the model is not treated as the only possible migration law. Rather, it is used as a controlled benchmark because the true data-generating mechanism is known and therefore structural recovery can be evaluated.

Fourth, it emphasizes the trade-off between prediction and interpretation. A black-box model may provide flexible prediction, while a theory-guided model may provide more interpretable spatial mechanisms. The proposed framework evaluates both aspects and therefore provides a basis for choosing a model according to the research purpose.

1.4 Structure of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 reviews the literature on New Economic Geography, continuous-space migration, advection–diffusion models, and multi-factor location choice. Chapter 3 discusses neural-network selection and explains why CNN is chosen as the main architecture for learning two-dimensional spatial dynamics. Chapter 4 presents the model setup, including the spatial domain, input factor fields, true potential function, and simulated population dynamics. Chapter 5 reports the experimental pipeline and results, including learnability testing, factor-relevance tests, theory-guided model comparison, factor-structure testing, and rolling-window prediction evaluation. Chapter 6 explains why simulated data are used and discusses the limitations of existing real-world data

sources such as GHSL, mobile-phone-based studies, and OpenStreetMap. Chapter 7 outlines future extensions, including temporal distance, boundary conditions, coupled factor dynamics, advanced neural architectures, stochastic extensions, and empirical applications. Chapter 8 provides additional algorithmic details, including the CNN structure, loss function, non-negative values, and control of randomness.

Chapter 2

Literature Review

2.1 NEG and Core–Periphery Foundations

New Economic Geography (NEG) provides an important theoretical foundation for studying spatial economic agglomeration and population distribution. The core–periphery model developed by Krugman [13] shows how increasing returns to scale, transport costs, and market-size effects may generate the concentration of firms and workers in particular regions. In this framework, mobile workers are attracted to locations with better economic opportunities, while firms benefit from larger markets and lower effective trade costs. This circular causality between market access and labor concentration is one of the central mechanisms of spatial agglomeration.

Later studies further developed this theoretical tradition, including broader reviews of the frontier of NEG [6]. Fujita et al. [8] provide a systematic treatment of cities, regions, and international trade within the NEG framework. Ottaviano et al. [24] introduce a more tractable model of agglomeration and trade, while Pflüger [25] derives explicit analytical solutions in a simplified Chamberlinian agglomeration model. These studies are important because they clarify the mechanisms through which spatial concentration can emerge from economic incentives.

A representative migration mechanism in this literature can be written as

$$\frac{\partial \lambda}{\partial t}(t,x) = v\left[w(t,x) - \bar{w}(t)\right]\lambda(t,x), \tag{2.1}$$

where $\lambda(t,x)$ denotes population density at location x and time t, $w(t,x)$ represents the local wage or economic attractiveness, $\bar{w}(t)$ is the average wage or attractiveness level, and $v > 0$ is a migration sensitivity parameter. This type of formulation captures the idea that people move toward locations with above-average economic returns.

However, this formulation also illustrates a limitation of many tractable NEG models. Population movement is often explained through one dominant factor, such as wage, utility, or economic attractiveness. This is useful for theoretical analysis, but it may be restrictive for urban population modeling. In real cities, migration and residential location decisions are influenced not only by wages or employment opportunities, but also by commercial activity, amenities, transportation accessibility, housing conditions, public services, and neighborhood characteristics. Moreover, many NEG models focus on equilibrium outcomes rather than on the dynamic adjustment process by which population distribution changes over time. These limitations motivate a broader methodological framework in which different migration mechanisms can be treated as candidate structures and compared systematically.

2.2 Continuous-Space Migration and Advection–Diffusion Models

A second relevant line of literature studies population movement in continuous space. This perspective is important because urban population movement does not occur only between a small number of discrete regions. Instead, population density changes locally over a spatial domain. Fujita and Ogawa [7] and Anas et al. [2] analyze urban spatial structure by considering location decisions, commuting, and the interaction between households and firms. These studies emphasize that urban space is spatially distributed and that location decisions depend on the surrounding environment.

Mossay [19] introduces an advection–diffusion perspective into spatial economic modeling. In this type of model, population movement is interpreted as the combination of local spreading and directed movement. A simplified continuous-space formulation can be written as

$$\frac{\partial \lambda}{\partial t}(t,x) = D\Delta\lambda(t,x) - \nabla \cdot (\lambda(t,x)\nabla w(t,x)), \qquad D > 0. \tag{2.2}$$

The diffusion term $D\Delta\lambda(t,x)$ represents local spatial spreading of population density. The advection term $-\nabla \cdot (\lambda(t,x)\nabla w(t,x))$ represents directed movement induced by the gradient of an economic or spatial field $w(t,x)$. Compared with the wage-difference formulation in Equation (2.1), this model is closer to a local dynamic process because it explicitly represents spatial gradients and movement across neighboring locations.

This type of model is especially relevant for the present thesis because it provides a bridge between economic geography and dynamic spatial modeling. However, many continuous-space models remain analytically restricted. They often use one-dimensional space, a single economic factor, or a simplified functional form in order to maintain tractability. These

restrictions make theoretical analysis possible, but they also limit the model's ability to represent realistic urban systems, where multiple spatial factors jointly influence population movement over a two-dimensional domain.

Recent studies on pattern formation in continuous-space NEG further show the importance of studying dynamic processes rather than only equilibrium states [1, 20]. For example, advection–diffusion models can generate spatial patterns over time, and these patterns may depend strongly on the assumed movement mechanism and the underlying economic field. This reinforces the need to study not only the final equilibrium distribution, but also the transition behavior of population density.

2.3 Multi-Factor Location Choice and Modeling Gaps

The broader residential location choice literature clearly shows that migration and settlement decisions are multi-dimensional. Households may choose residential locations based on job accessibility, commuting cost, housing affordability, amenities, schools, public services, neighborhood quality, and life-cycle considerations. Studies such as Anas et al. [2], Bayoh et al. [3], Kim et al. [12], Schirmer et al. [26], Zondag and Pieters [39], and Orvin and Fatmi [23] all support the view that residential location choice cannot be fully explained by a single factor.

A general multi-attribute location choice formulation can be written as

$$U_i(x) = \alpha_1 w_1(x) + \alpha_2 w_2(x) + \cdots + \alpha_n w_n(x) + \varepsilon_i(x), \tag{2.3}$$

where $U_i(x)$ denotes the utility of household or individual i at location x, $w_1(x), \ldots, w_n(x)$ represent location attributes, $\alpha_1, \ldots, \alpha_n$ are preference or sensitivity parameters, and $\varepsilon_i(x)$ captures unobserved preference heterogeneity or random effects. This type of formulation reflects the idea that residential choice depends on multiple spatial attributes rather than one dominant variable.

However, multi-factor location choice models are often used to explain discrete location decisions or cross-sectional residential patterns. They do not always provide a continuous-time dynamic equation for the evolution of population density over a two-dimensional spatial domain. In other words, the location choice literature strongly supports the importance of multiple factors, but it does not directly solve the dynamic modeling problem addressed in this thesis.

This creates a methodological gap between three strands of literature. NEG models provide theoretical mechanisms of agglomeration, but often rely on simplified migration drivers

and equilibrium analysis. Continuous-space advection–diffusion models describe local movement and dynamic adjustment, but are often single-factor or low-dimensional. Multi-factor location choice studies identify many relevant explanatory variables, but are not usually formulated as dynamic population-density models.

The present thesis is motivated by this gap. Instead of assuming that one existing model is universally correct, the thesis treats different formulations as candidate model structures. Wage-difference models, single-factor advection–diffusion models, multi-factor advection–diffusion models, black-box neural models, and other future extensions can all be viewed as different ways to specify the dynamic operator governing population movement. The goal is therefore methodological: to develop a computational framework for testing whether the dynamic process is learnable, which factors are relevant, whether a proposed model structure is adequate, how factors enter the migration mechanism, and whether prediction remains stable over time.

2.4 Summary of the Literature Gap

The literature reviewed above suggests several important conclusions. First, existing spatial economic models provide strong theoretical explanations for agglomeration, but many of them focus on one dominant migration factor and long-run equilibrium outcomes. Second, continuous-space advection–diffusion models provide a useful way to describe local population movement, but many existing formulations remain restricted in spatial dimension or factor structure. Third, residential location choice studies show that real migration decisions depend on multiple factors, but these studies do not usually provide a unified dynamic model for population-density evolution.

These conclusions motivate the methodological framework developed in the following chapters. The framework does not attempt to replace existing theories with a single fixed equation. Instead, it provides a way to learn, compare, and diagnose different candidate models under a common computational pipeline. In this thesis, the two-dimensional multi-factor advection–diffusion model is used as the main controlled demonstration case, but the broader logic of the framework can be applied to other spatial population models as well.

Chapter 3

Neural Network Selection

This chapter explains why convolutional neural networks are selected as the main neural-network architecture in this thesis. The purpose of this chapter is not to claim that CNN is universally better than other neural networks. Instead, the choice is made according to the structure of the data, the nature of the prediction problem, and the methodological objective of the thesis.

In applied machine learning, an important principle is that model selection should be guided by the structure of the data and the purpose of the task [10]. A model that is powerful in one setting may not be suitable in another setting if its architecture does not match the data-generating process or the form of the input variables. In this thesis, the input variables are two-dimensional spatial fields, and the target process is a local spatial dynamic process. Therefore, the selected architecture should preserve spatial structure, capture local neighborhood relationships, and remain simple enough to support fair comparison across different model structures.

3.1 Problem Description

The learning problem studied in this thesis is a two-dimensional spatial population dynamics problem. Let $\Omega \subset \mathbb{R}^2$ denote a two-dimensional urban domain, and let

$$\lambda(t, x), \qquad x \in \Omega, \quad t \geq 0,$$

represent the population density at location x and time t. The explanatory variables are represented by spatial factor fields

$$w_1(x), w_2(x), \ldots, w_n(x),$$

where each factor is defined on the same two-dimensional grid as the population density.

The general prediction problem is to learn how the current population state and the spatial factors determine the future population distribution. In a broad form, the dynamic process can be written as

$$\frac{\partial \lambda}{\partial t} = \mathcal{M}\big(\lambda, W\big),$$

where $W = (w_1, \ldots, w_n)$ denotes the collection of spatial factor fields, and $\mathcal{M}$ represents a candidate dynamic operator. This operator may be a black-box neural network, a wage-difference model, an advection–diffusion model, or another theory-guided structure.

The problem is PDE-related because the population field evolves through local spatial interactions. Diffusion depends on local differences in population density, while advection depends on the gradient of the spatial potential. Therefore, the learning architecture should be able to process neighboring grid cells and local spatial patterns.

The experiments evaluate both short-horizon and long-horizon prediction. Short-horizon prediction tests whether the model can learn local dynamic changes over a limited forecasting period. Long-horizon prediction tests how errors accumulate when the learned model is recursively applied over many steps. Although both horizons are important, short-horizon prediction is especially important for the methodological purpose of this thesis, because the main goal is to learn the local dynamic rule. In realistic applications, prediction can also be updated when new observations become available, which motivates the later rolling-window evaluation.

3.2 Model Selection Principle

A useful machine-learning model should match the structure of the data. In business and financial applications, model selection is often guided by the type of input data, the prediction target, interpretability requirements, and the cost of overfitting [10]. The same principle applies to this thesis.

The experiments in this thesis use a 200×200 spatial grid. This choice is made for two main reasons. First, it is consistent with the type of gridded spatial information provided by large-scale population datasets such as the Global Human Settlement Layer (GHSL). For medium-sized and large cities, an urban area can be represented by a grid of this order of magnitude while still preserving meaningful spatial variation in population density and urban factors [4]. Second, the 200×200 grid is suitable from a machine-learning perspective. It is large enough to represent two-dimensional spatial patterns, local gradients, and heterogeneous factor distributions, but still small enough to make repeated model training

and rolling-window experiments computationally feasible.

Based on these requirements, this chapter compares multilayer perceptrons, temporal models, and convolutional neural networks.

3.3 Comparison of Candidate Neural Network Architectures

3.3.1 Multilayer Perceptron

A multilayer perceptron is a general-purpose neural network for nonlinear function approximation. In principle, an MLP can approximate complex relationships between inputs and outputs. A direct prediction model can be written as

$$\lambda_{t+\Delta t} = f_\theta\Big(w_1, w_2, \ldots, w_n, \lambda_t\Big),$$

where f_θ is represented by a fully connected neural network.

However, an MLP is not well suited to the spatial structure of this thesis. To use an MLP, the two-dimensional fields usually need to be flattened into vectors. For a 200×200 grid, one field already contains $40,000$ values. If several spatial factors and the current population density are included, the input dimension becomes very large. This creates a high-dimensional learning problem and increases the number of parameters required by the model.

More importantly, flattening destroys the explicit neighborhood structure of the spatial grid. In a PDE-based population model, nearby cells are not independent. Diffusion, advection, and spatial gradients are all local mechanisms that depend on neighboring locations. Once the grid is flattened, the model no longer naturally recognizes which cells are adjacent to each other.

From the perspective of applied machine learning, this means that the MLP architecture does not match the structure of the data. Although it is flexible, it does not build in the spatial inductive bias required by the problem. Therefore, MLP is not selected as the main architecture in this thesis.

3.3.2 Temporal Models: RNN, LSTM, and GRU

Temporal models are designed for data in which the order of observations matters. In many forecasting applications, the target variable is not determined only by the current input,

but also by a sequence of previous observations. Recurrent neural networks, long short-term memory networks, and gated recurrent units are common architectures for this type of sequential learning problem [17].

A recurrent neural network can be written in an abstract form as

$$h_t = R_\theta(h_{t-1}, z_t),$$

$$\widehat{\lambda}_{t+\Delta t} = G_\theta(h_t),$$

where z_t denotes the input at time t, h_t is a hidden state summarizing past information, and G_θ maps the hidden state to a future prediction. The key idea is that the model carries information from previous time steps through the hidden state.

LSTM and GRU models improve the basic RNN structure by introducing gates. These gates control how information is stored, updated, and forgotten over time. This makes them more effective when the data contain long temporal dependence. For example, in some empirical urban applications, population movement may respond gradually to earlier changes in employment centers, transportation infrastructure, housing supply, or amenities. In such cases, LSTM or GRU models may be useful for capturing delayed responses and path dependence.

However, the population dynamics studied in this thesis are closer to a state-based spatial dynamic process. The basic assumption is that population movement over the next short time interval is mainly determined by the current population distribution and the current spatial factor fields. In other words, people make relocation decisions based on the current attractiveness of different locations, such as current job opportunities, commercial activity, amenities, and transportation accessibility. Past events may influence migration indirectly, but their effects are largely reflected in the current population density and the current spatial factors.

This idea is consistent with the one-step dynamic formulation used in this thesis. The model learns a local transition rule from the current state,

$$\lambda_t, W \quad \longrightarrow \quad \lambda_{t+\Delta t},$$

or equivalently,

$$\lambda_t, W \quad \longrightarrow \quad \frac{\partial \lambda}{\partial t}.$$

Under this formulation, long temporal memory is not the central modeling target. Instead, the main task is to learn the local spatial operator that maps the current population field and spatial factors to the next population change.

Standard temporal models also have an important limitation for the present problem. They are not naturally designed for two-dimensional spatial operators. If each spatial field is flattened before being passed into an RNN, LSTM, or GRU, the model loses the explicit two-dimensional grid structure. This is problematic because the dynamic mechanism studied in this thesis is local and spatial. Diffusion depends on neighboring population values, and advection depends on spatial gradients of the potential function. These mechanisms are more naturally represented by convolutional operations.

Another limitation is methodological clarity. The goal of this thesis is to compare factor sets and theory-guided dynamic structures. A highly flexible temporal model may improve prediction in some cases, but it may make it harder to interpret why the prediction improves. The improvement may come from temporal memory, from the selected spatial factors, or from the theoretical structure. This would weaken the comparison between black-box models and theory-guided models.

Spatio-temporal recurrent models, such as ConvLSTM, can partly address this issue by combining convolutional spatial processing with recurrent temporal memory. Such models preserve spatial locality while also learning temporal dependence. They are important candidates for future empirical work, especially when long continuous time-series data are available and when delayed responses are a central research question. However, they introduce additional complexity and are not necessary for the controlled simulation setting used in this thesis.

Therefore, RNN, LSTM, and GRU models are not selected as the main architecture. They are useful future extensions, but they are less suitable for the current objective of learning and comparing local spatial dynamic operators on a two-dimensional grid. In this thesis, the population transition is modeled primarily as a function of the current state rather than as a long-memory sequence prediction problem.

3.3.3 Convolutional Neural Network

A convolutional neural network is well suited to the present problem because the input variables are two-dimensional spatial fields. Each factor can be treated as one input channel, and the population density can be added as another channel when needed. The input tensor can therefore be written as

$$\Big(w_1(x), w_2(x), \ldots, w_n(x), \lambda_t(x)\Big).$$

This is similar to a multi-channel image, except that the channels represent spatial factors and population density rather than color values.

A convolutional layer can be written as

$$H^{(l+1)} = \sigma\left(K^{(l)} * H^{(l)} + b^{(l)}\right),$$

where $H^{(l)}$ is the feature map at layer l, $K^{(l)}$ is a convolutional kernel, $*$ denotes convolution, $b^{(l)}$ is a bias term, and σ is a nonlinear activation function. CNNs are widely used for grid-structured data because convolutional kernels learn local patterns through shared parameters [15, 16].

CNNs satisfy the main requirements of this thesis. First, they preserve the two-dimensional spatial structure of the data. The convolutional kernel operates locally on neighboring cells, which is consistent with the local nature of diffusion and advection. Second, parameter sharing makes CNNs more efficient than fully connected networks for large spatial grids. Third, CNNs can learn nonlinear spatial relationships without destroying the grid structure.

CNNs also fit the methodological objective of the thesis. The same basic CNN architecture can be used across different experiments, while the model structure changes around it. For example, in a black-box derivative-learning model, the CNN can directly learn the local time derivative. In a physics-aware model, the CNN can learn the potential function, and the population derivative is then produced through the corresponding dynamic operator. This allows different theoretical structures to be compared under a similar level of neural-network capacity.

The limitation of CNNs is that they do not explicitly model long temporal memory. A standard CNN uses the current spatial state and factor fields to make a one-step prediction or derivative estimate. However, this limitation is acceptable in this thesis because the main objective is to learn the local dynamic operator. Long-horizon behavior is then evaluated through recursive rollout and rolling-window prediction rather than through a recurrent temporal architecture.

3.4 Selected Architecture in This Thesis

Based on the comparison above, CNN is selected as the main neural-network architecture in this thesis. This choice follows the machine-learning principle that the architecture should match the input structure and prediction objective. Since the inputs are two-dimensional spatial fields and the target mechanism is local spatial dynamics, CNN provides a natural and efficient architecture.

The CNN used in the experiments consists of three convolutional layers with kernel size 3 and padding 1, so that the spatial grid size is preserved. The first two convolutional layers

are followed by ReLU activation functions. The hidden layers use 32 channels. The general structure is

$$\mathrm{Conv}_{3\times3} \to \mathrm{ReLU} \to \mathrm{Conv}_{3\times3} \to \mathrm{ReLU} \to \mathrm{Conv}_{3\times3}.$$

The number of input channels depends on the experiment. For black-box residual models, the input includes selected spatial factors and the current population density. For physics-aware models, the input includes selected spatial factors, and the CNN learns the potential function used by the dynamic operator.

To ensure fair comparison, the same basic CNN structure is used throughout the experiments. This is important because the thesis compares model structures rather than neural-network size. If different models used very different neural architectures, performance differences might be caused by different computational capacity rather than by factor selection or theoretical structure. By keeping the CNN architecture fixed, the experiments provide a clearer comparison of the proposed methodological steps.

Table 3.1: Comparison of candidate neural-network architectures.

Model	**Main advantage**	**Main limitation in this thesis**	**Decision**
MLP	General nonlinear approximation	Requires flattening and loses spatial neighborhood structure	Not selected
RNN	Captures sequential dependence	Not naturally designed for two-dimensional spatial operators	Not selected
LSTM / GRU	Better temporal memory than standard RNN	More complex and still requires spatial processing to preserve grid structure	Future extension
ConvLSTM	Combines spatial convolution and temporal memory	Useful for real time-series data but more complex than needed here	Future extension
CNN	Preserves grid structure, learns local spatial patterns, and remains computationally efficient	Does not explicitly model long temporal memory	Selected

Overall, CNN is selected because it provides the best balance between spatial suitability, computational simplicity, and methodological clarity. It preserves the two-dimensional grid structure, learns local spatial operators, and allows different model structures to be compared under a consistent neural-network architecture.

Chapter 4

Model Setup

This chapter uses simulated data to demonstrate the complete methodological procedure proposed in this thesis. The purpose of using simulation is not to replace real-world empirical analysis, but to create a controlled environment in which the true dynamic mechanism is known. At the same time, the simulated data are designed to be logically consistent with realistic urban population movement: population is attracted by job opportunities, commercial activity, amenities, and transportation accessibility, while the spatial pattern remains two-dimensional and locally dynamic.

4.1 Spatial Domain and Population Dynamics

The experiments are conducted on a two-dimensional spatial grid. Let $\Omega \subset \mathbb{R}^2$ denote the spatial domain, and let $\lambda(t, x)$ represent the population density at time t and location $x \in \Omega$. The population state evolves according to a PDE-based spatial dynamic mechanism.

The main theory-guided model considered in this thesis is the multi-factor advection–diffusion model:

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot \left(D \nabla \lambda - \lambda \nabla f(W) \right),$$

where $D > 0$ is the diffusion coefficient, $W = (w_1, w_2, w_3, w_4)$ is the vector of spatial factors, and $f(W)$ is the population potential function.

The diffusion term $D\nabla\lambda$ describes local spreading of population density. The advection term $-\lambda \nabla f(W)$ describes directed movement induced by the spatial gradient of the population potential. In this interpretation, population tends to move toward locations with higher spatial attractiveness, while diffusion prevents excessive concentration at a single location.

4.2 Input Factor Fields

The simulated system uses four main spatial factor fields:

$$w_1(x), \quad w_2(x), \quad w_3(x), \quad w_4(x).$$

They represent job opportunity density, commercial density, amenity density, and transportation accessibility, respectively. These factors are selected because they are common and intuitive drivers of residential location choice and urban population movement.

The four factor fields are defined as follows:

- $w_1(x)$: job opportunity density;
- $w_2(x)$: commercial density;
- $w_3(x)$: amenity density;
- $w_4(x)$: transportation accessibility.

Each factor is generated on the same 200×200 grid and normalized to the interval $[0, 1]$ using min–max scaling:

$$w^*(x) = \frac{w(x) - \min_x w(x)}{\max_x w(x) - \min_x w(x)}.$$

This normalization makes the numerical scales comparable across different spatial factors.

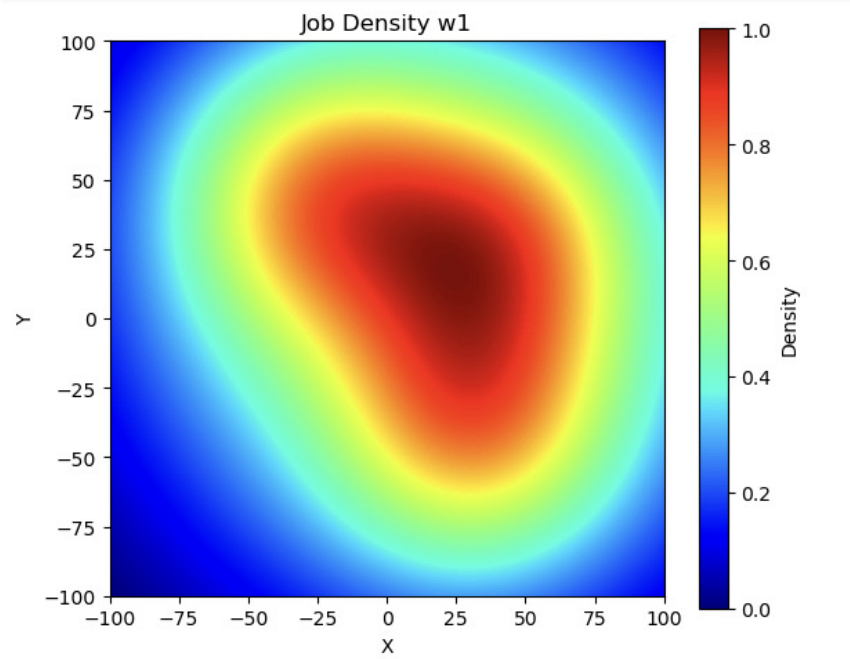


Figure 4.1: Job opportunity density w_1 is generated as a broad smooth field with major employment centers.

Job opportunity density represents the main economic driver of population movement. In the simulation, high-value regions indicate locations with stronger employment attractiveness.

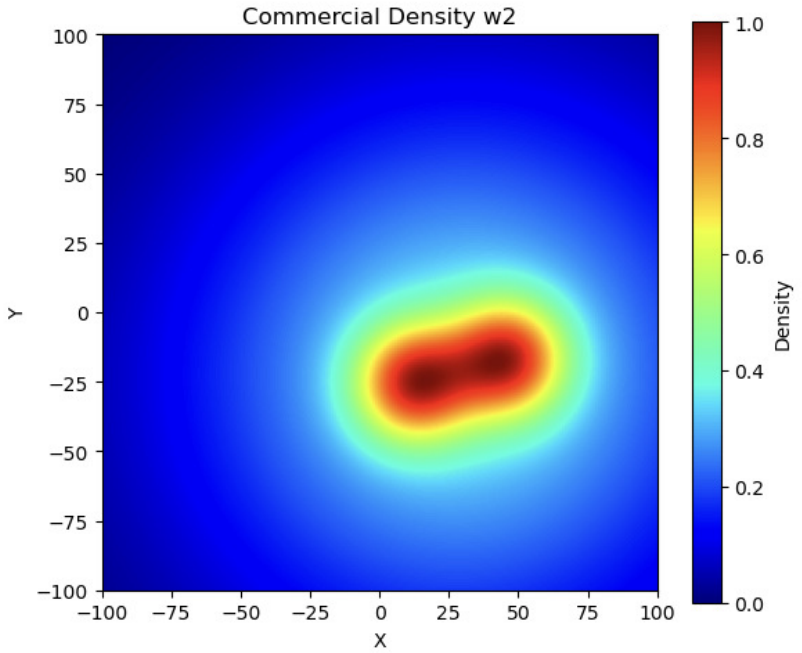


Figure 4.2: Commercial density w_2 represents concentrated commercial activity in the urban domain.

Commercial density provides additional local attractiveness beyond employment opportunities. It is designed as a more concentrated spatial field than job density.

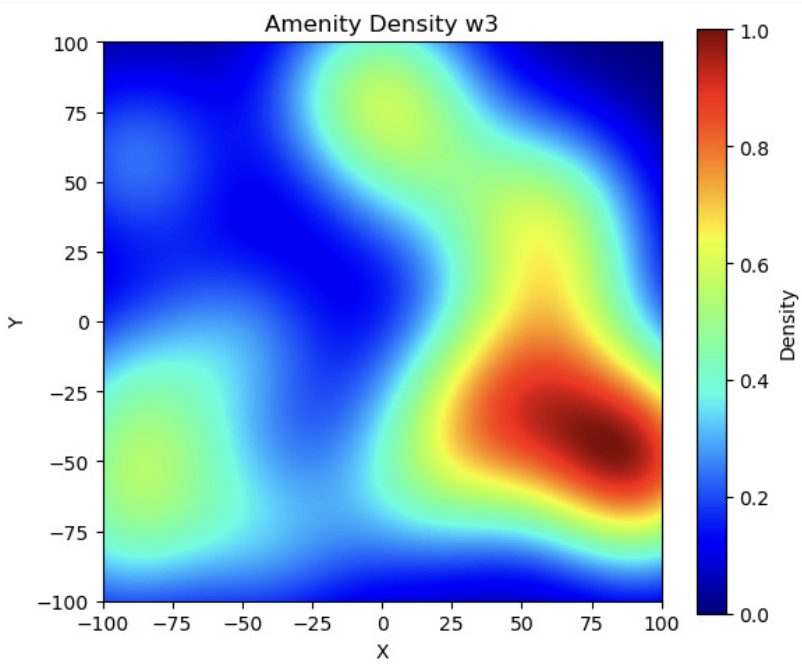


Figure 4.3: Amenity density w_3 is generated from multiple dispersed amenity centers.

Amenity density represents spatially distributed non-employment attractiveness. Its effect is not assumed to be purely linear in the true potential function.

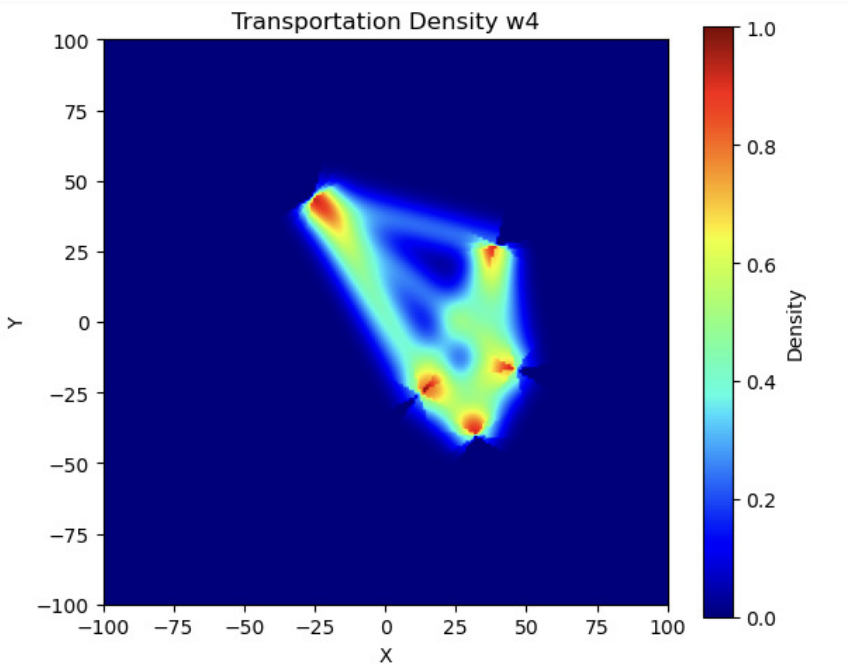


Figure 4.4: Transportation accessibility w_4 is represented by corridor-like structures connecting major centers.

Transportation accessibility is designed to reflect the idea that movement is affected by spatial connectivity. In the simulation, transportation does not act only as an independent linear factor, but interacts with other spatial conditions.

4.3 True Potential Function

The simulated population dynamics are generated from a known benchmark potential function:

$$f_{\text{true}}(w_1, w_2, w_3, w_4) = a_1 w_1 + a_2 w_2 + a_3 w_3 w_4.$$

The parameters used in the experiments are

$$D = 0.5, \qquad a_1 = 1.0, \qquad a_2 = 0.5, \qquad a_3 = 0.5.$$

This specification reflects three modeling assumptions. First, job opportunity density is treated as the most important driver of population movement, so its coefficient is the largest. This is consistent with the intuition of spatial economic models in which employment

opportunity or wage plays a central role in migration.

Second, commercial density provides additional explanatory power beyond job opportunity density, but its direct effect is weaker than the job-opportunity effect. Therefore, the coefficient of w_2 is lower than the coefficient of w_1.

Third, transportation accessibility is not assumed to affect population movement only as an independent additive factor. In real urban environments, the value of transportation may depend on the spatial context in which it appears. For example, accessibility near amenity centers, commercial centers, or employment centers may have a stronger effect than accessibility in isolation. To reflect this idea in a simple controlled form, the true function includes the product term $w_3 w_4$.

4.4 Behavior of the True Function

In this model, population is interpreted as a fluid-like quantity that moves in an attractive vector field induced by the population potential. More specifically, the diffusion term describes local spreading, while the advection term drives population toward locations with higher potential. Under the combined effect of these two mechanisms, the population distribution evolves over time and gradually converges to an equilibrium configuration.

From the perspective of urban development, cities often expand outward from an initially concentrated core area. To reflect this intuition in a simple and controlled way, this thesis uses a deliberately extreme initial condition: the initial population density is concentrated near the center of the spatial domain. Although real cities are usually more complicated, this stylized setting makes the subsequent diffusion, directional movement, and equilibrium behavior easier to observe.

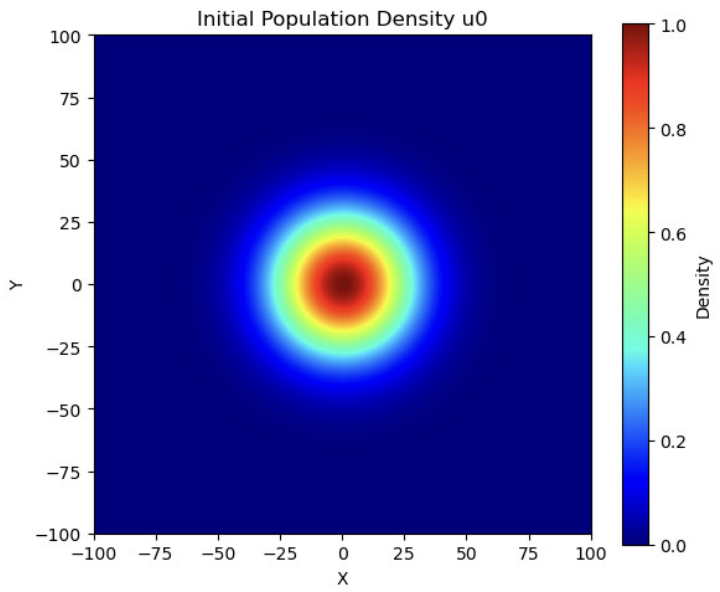


Figure 4.5: Initial population density u_0, concentrated near the center of the spatial domain.

Figure 4.5 shows the initial population density used in the simulation. This initial condition represents a city that starts from a relatively centralized core and then evolves under the influence of the spatial attractiveness field.

To further illustrate the dynamic behavior of the system, Figure 4.6 reports the absolute average population change over time. The curve decreases steadily, which indicates that the speed of population change becomes progressively smaller. In other words, the system adjusts rapidly at the beginning, then gradually slows down, and finally approaches equilibrium.

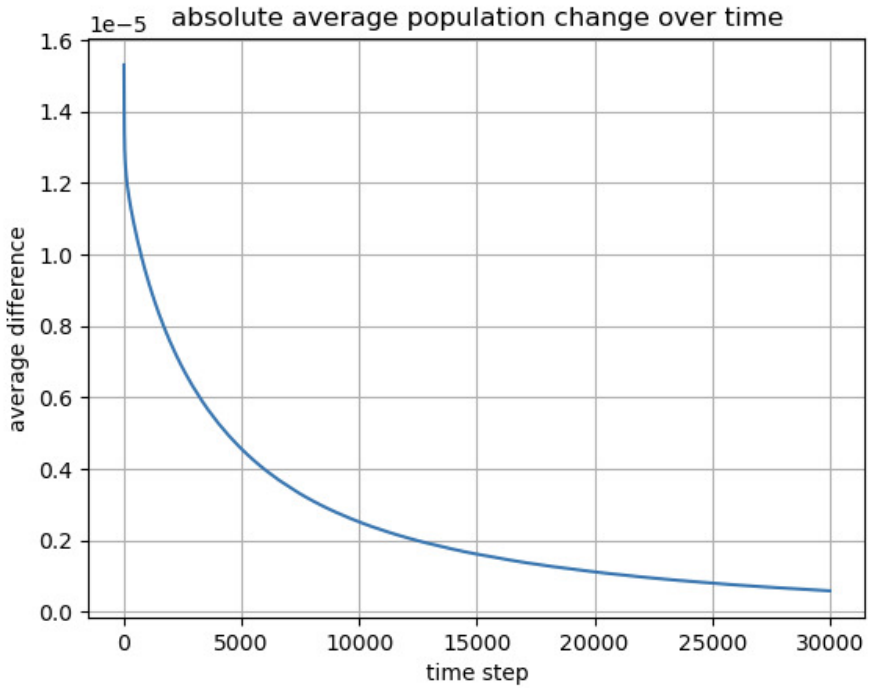
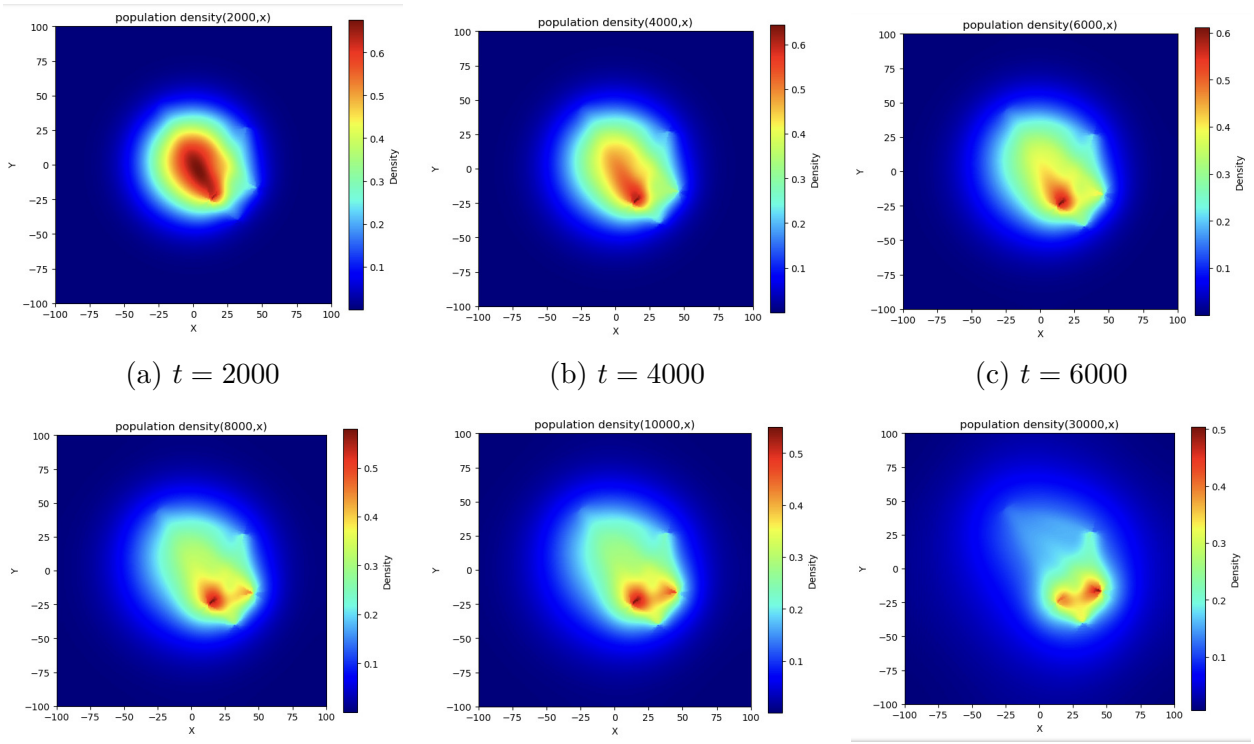


Figure 4.6: Absolute average population change over time. The decreasing curve indicates that the system gradually approaches equilibrium.

The next figure shows the population distribution at several representative time steps. These snapshots make the convergence process visually clear. At early stages, the population is still strongly influenced by the initial central concentration. As time increases, the distribution gradually shifts toward regions with higher spatial attractiveness and becomes smoother. By 30,000 time steps, the system is already very close to equilibrium.



(a) $t = 2000$ (b) $t = 4000$ (c) $t = 6000$

(d) $t = 8000$ (e) $t = 10000$ (f) $t = 30000$

Figure 4.7: Population density at different time steps. The distribution gradually moves under the attractive field and becomes increasingly stable. By $t = 30000$, the system is close to equilibrium.

Figure 4.7 shows that the population does not remain symmetric around the initial center. Instead, it gradually moves toward the more attractive regions implied by the true potential function. At the same time, diffusion smooths the overall distribution. This combination of directed movement and local spreading is exactly the behavior expected from the advection–diffusion structure.

Overall, these figures illustrate the intended behavior of the true model. The system starts from a highly centralized initial population distribution, then evolves through spatial movement and diffusion, and eventually approaches an equilibrium state. This makes the simulated data suitable for demonstrating the proposed methodology, because the dynamic process is both visually interpretable and logically consistent with stylized urban development.

4.5 Assessment of the Reasonableness of the Simulated Data

The reasonableness of the simulated data is assessed from two perspectives. First, the construction of the four spatial factors and the theory-guided dynamic model are examined against established urban and spatial-economic theories. Second, the resulting population distribution is compared with theoretical predictions and empirical population patterns reported for real cities.

4.5.1 Reasonableness of the Data-Generating Logic

The four spatial factors represent job opportunities, commercial activity, amenities, and transportation accessibility. Their inclusion is consistent with the residential-location literature, which identifies employment access, commercial and public services, amenities, and transportation accessibility as important determinants of household location choice [2, 3, 12, 26, 39, 23].

The spatial construction of these factors follows their intended urban interpretations. Job opportunities, commercial activity, and amenities are generated using several spatial clusters and surrounding areas of influence. Factor values are highest near their respective centres and gradually decrease with distance. The job-opportunity field contains relatively broad employment centres and a weak background level across the domain, whereas the commercial field contains more concentrated local cores. The amenity field consists of several dispersed local centres.

Transportation accessibility is constructed as corridor-like structures connecting the major job and commercial centres. Accessibility is highest along these connections and decreases with distance from them. This represents the role of transportation networks in linking major points of interest across an urban area. Similar factor fields could be constructed empirically using employment data, points of interest, and transportation networks [37, 18].

The central role assigned to job opportunities is consistent with New Economic Geography, in which employment, wages, and market access are major forces of worker migration and spatial agglomeration [13, 8]. The broader multi-factor setting is also consistent with evidence that residential decisions depend on several location characteristics rather than employment alone [26, 39].

Population evolution is generated by the multi-factor advection–diffusion model

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot \left(D \nabla \lambda - \lambda \nabla f_{\text{true}}(\mathbf{W}) \right).$$

The diffusion term represents local population spreading, while the advection term represents directed movement in response to spatial differences in attractiveness. This theory-guided structure is consistent with the continuous-space economic migration models of Mossay [19] and Ohtake [20]. Therefore, both the construction of the spatial factors and the mechanism generating population movement have reasonable urban and theoretical interpretations.

4.5.2 Reasonableness of the Resulting Population Distribution

The simulated population initially spreads from the central starting location and gradually moves toward more attractive regions. The resulting distribution is spatially uneven, contains several connected high-density areas, and remains smooth rather than collapsing into isolated grid cells.

These results are consistent with the theoretical conclusions of Mossay [19] and Ohtake [20]. Their continuous-space models show that directed migration toward economically attractive locations can generate population agglomeration, while diffusion smooths the population distribution and limits excessive local concentration. Depending on the spatial setting, the process may generate one dominant centre or multiple urban centres. The simulated results exhibit the same main features: movement toward attractive areas, spatial agglomeration, and smooth transitions between high- and low-density regions.

The resulting heat maps are also qualitatively comparable with empirical population patterns obtained from mobile-phone and geospatial data. Studies of Beijing identify strongly uneven population and activity distributions, with major concentrations and surrounding secondary centres [5, 35, 38]. Cellular-network studies of important locations and urban land use similarly show that residential, employment, and activity locations form spatial clusters rather than uniform distributions [11, 32]. For Foshan, Yang et al. [36] identify clustered home and work locations across multiple urban districts using mobile-phone signalling data.

This comparison is qualitative rather than a direct empirical validation. The simulated maps are not intended to reproduce the geography or population values of any particular city. Nevertheless, their main characteristics—major population concentrations, secondary centres, and decreasing density away from these centres—are consistent with both the theoretical models and the empirical urban patterns discussed above. The simulated data therefore provide a reasonable controlled benchmark for evaluating the methodological framework.

Chapter 5

Experimental Pipeline and Results

The pipeline contains five steps:

1. **Learnability test:** to examine whether neural networks can learn the simulated spatial population dynamics from the current population density and spatial factor fields.
2. **Factor-relevance and incremental inclusion tests:** to identify which spatial factors provide useful explanatory information and should be retained for later model-structure comparison.
3. **Theory-guided model-structure comparison:** to compare different candidate dynamic structures and select the model that best explains the observed population dynamics.
4. **Factor-structure test:** to examine how individual factors enter the selected model, especially whether they can be represented linearly.
5. **Rolling-window prediction evaluation:** to test whether the selected model remains accurate and stable when the prediction window moves forward, and whether updating the training window can improve prediction accuracy.

5.1 Dataset Split and Evaluation Metrics

The first $10,000$ time steps are used for model training and evaluation. The dataset is divided into training, validation, and test sets:

$$\text{training} : \text{validation} : \text{test} = 6000 : 2000 : 2000.$$

The first $6,000$ time steps are used for training, the next $2,000$ time steps are used for validation, and the final $2,000$ time steps are used for testing.

Two prediction horizons are considered:

$$T_{\text{short}} = 500, \qquad T_{\text{long}} = 2000.$$

The short-term horizon evaluates whether the model can accurately describe local dynamic behavior. The long-term horizon evaluates recursive error accumulation over a longer forecasting period.

The main evaluation metric is the percentage form of mean absolute error:

$$\text{MAE\%} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\widehat{\lambda}_i - \lambda_i|}{\lambda_i} \times 100\%,$$

where $\widehat{\lambda}_i$ is the predicted population density, λ_i is the simulated population density, N is the number of evaluated grid-time observations.

5.2 Step 1: Learnability Test

The first step examines whether the dynamic population process can be learned from the spatial factor fields and the current population density before imposing a specific economic-geography structure. This step is important because if the population dynamics cannot be learned even by a flexible neural-network model, then it would be difficult to justify more structured model-comparison experiments in later steps.

Two baseline learning methods are considered. Method 1.1 directly predicts the next population density, while Method 1.2 learns the local temporal change of population density. Both methods use the same spatial input factors, but they differ in how the population update is represented.

5.2.1 Method 1.1: Direct Prediction

Method 1.1 learns the one-step transition directly:

$$\lambda_{t+\Delta t}(x) = f_\theta\Big(w_1(x), w_2(x), w_3(x), w_4(x), \lambda_t(x)\Big).$$

This method is a black-box direct prediction model. It does not explicitly learn a temporal derivative and does not impose an advection–diffusion structure.

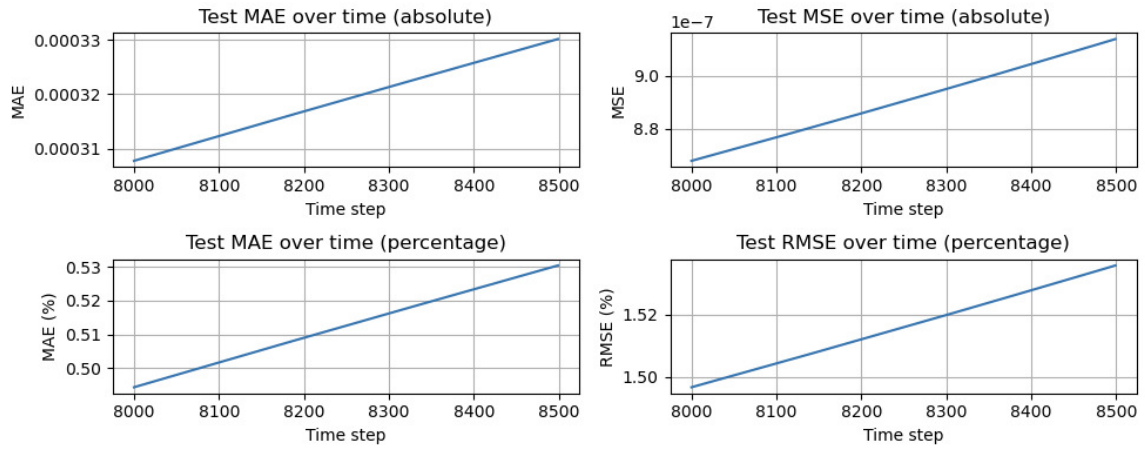


Figure 5.1: Short-horizon test error of Method 1.1 over the first 500 test steps.

Figure 5.1 shows that Method 1.1 maintains a low short-horizon error. The percentage MAE remains around 0.50%–0.53%, indicating that the model can learn the short-term population transition accurately.

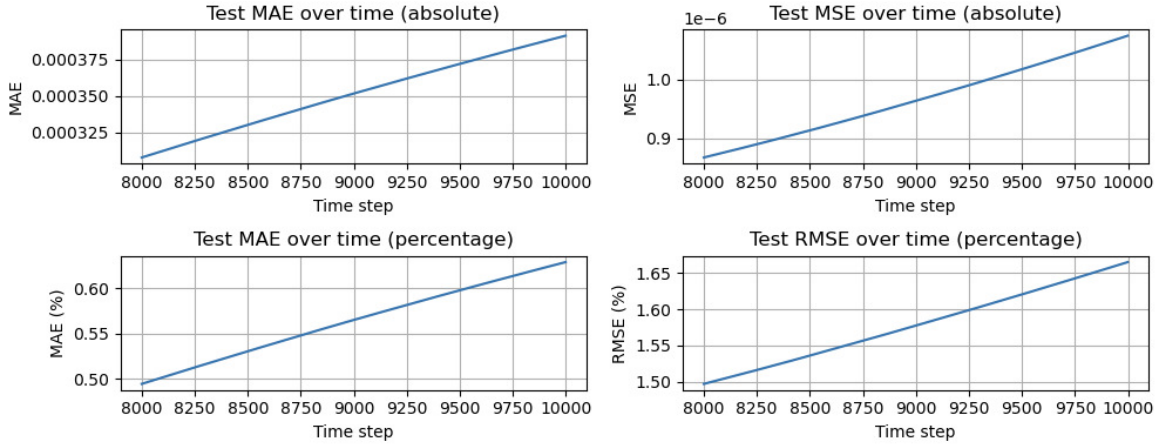


Figure 5.2: Long-horizon test error of Method 1.1 over the full 2000-step test period.

Figure 5.2 shows that the long-horizon error of Method 1.1 increases gradually but remains relatively stable. The percentage MAE increases from about 0.50% to about 0.63%, suggesting that direct prediction has good long-horizon stability in this experiment.

5.2.2 Method 1.2: Derivative Learning

Method 1.2 learns the local temporal change of population density:

$$\lambda_{t+\Delta t}(x) = \lambda_t(x) + f_\theta\Big(w_1(x), w_2(x), w_3(x), w_4(x), \lambda_t(x)\Big)\Delta t.$$

Equivalently, the model learns

$$\frac{\partial \lambda}{\partial t} = f_\theta(w_1, w_2, w_3, w_4, \lambda).$$

This method is closer to a dynamic evolution equation because it learns the population change rather than the next population state directly.

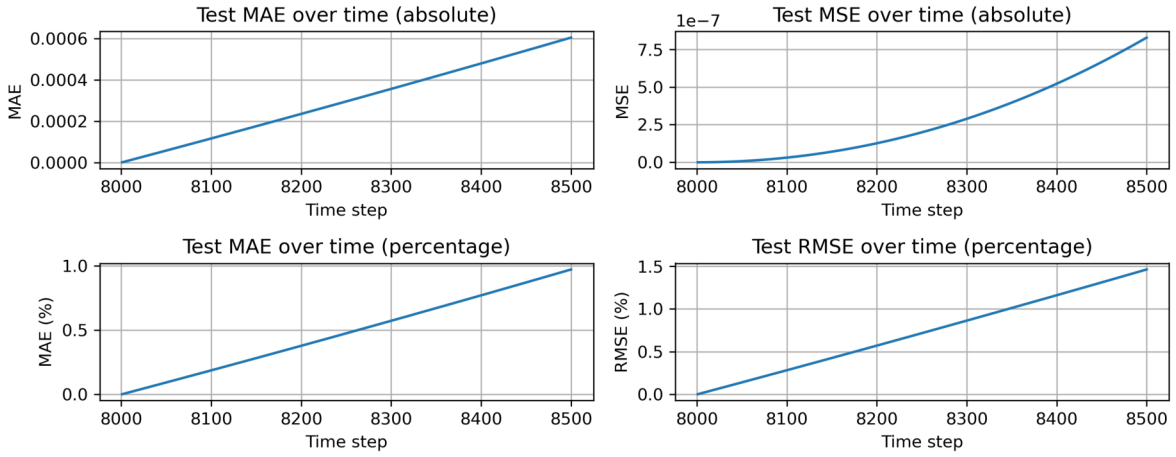


Figure 5.3: Short-horizon test error of Method 1.2 over the first 500 test steps.

Figure 5.3 shows that Method 1.2 performs well over the short horizon. The percentage MAE remains below about 1% during the first 500 test steps.

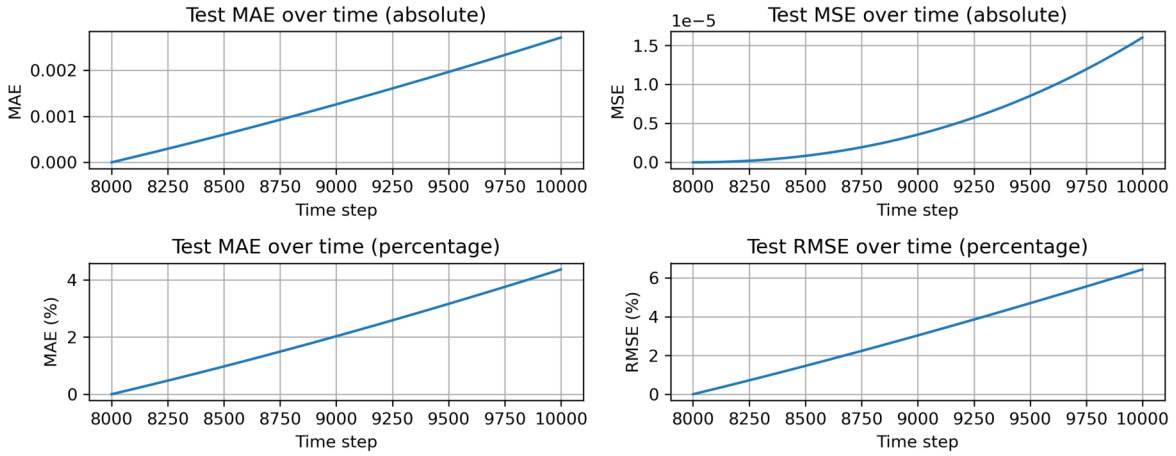


Figure 5.4: Long-horizon test error of Method 1.2 over the full 2000-step test period.

Figure 5.4 shows that the error of Method 1.2 grows more rapidly over the long horizon. This indicates that recursive derivative-based prediction is more sensitive to error accumulation when applied repeatedly over many time steps.

5.2.3 Results and Interpretation

Table 5.1: Step 1 results: short-horizon and long-horizon learnability test.

Method	Model type	Short-horizon MAE (%)	Long-horizon MAE (%)
Method 1.1	Direct prediction	0.5125	0.56
Method 1.2	Derivative learning	0.4792	2.08

Table 5.1 summarizes the main results. Both methods learn the short-horizon dynamics well, and their short-term errors are close. Method 1.2 achieves a slightly lower short-horizon MAE.

The long-horizon results show a different pattern. Method 1.1 remains more stable over the full 2000-step rollout, while Method 1.2 accumulates larger recursive prediction errors. This indicates that long-horizon recursive prediction should be interpreted separately from short-horizon dynamic learning.

Based on these results, Method 1.2 is selected as the black-box baseline for the later experiments. Although Method 1.1 has better long-horizon performance, its prediction error is already relatively high at the beginning of the test period. Therefore, shortening the prediction horizon through a rolling-window strategy does not provide much additional improvement for Method 1.1, because the model has limited room for improvement in short-window forecasting. In contrast, Method 1.2 is more useful for the methodological purpose of this thesis. Its short-horizon performance is slightly better, and its formulation is closer to a dynamic evolution model because it learns the local temporal change of population density. This makes Method 1.2 easier to compare with PDE-based and advection–diffusion structures in later steps. Therefore, Method 1.2 is used as the main black-box derivative-learning baseline, while Method 1.1 is retained as a direct-prediction reference model.

5.3 Step 2: Factor-Relevance and Incremental Inclusion Tests

Step 2 evaluates which spatial factors should be included in the learning model. The purpose is not only to improve prediction accuracy, but also to examine whether each candidate factor provides additional explanatory information for the population dynamics.

Since job opportunity density is the primary factor in this thesis, w_1 is used as the baseline factor. This choice follows the intuition of many spatial economic models, where

employment opportunity or wage attractiveness is one of the most important drivers of population movement. Starting from this baseline, two groups of experiments are designed.

The first group tests the marginal contribution of each additional factor. In this group, w_1 is always included, and one other factor is added at a time. This produces a sequence of two-factor models:

$$(w_1, \lambda), \quad (w_1, w_2, \lambda), \quad (w_1, w_3, \lambda), \quad (w_1, w_4, \lambda), \quad (w_1, w_5, \lambda).$$

where w_5 is a weak irrelevant factor. The purpose of this group is to examine how much each candidate factor improves the model when it is added to the primary job-opportunity factor.

The second group uses an incremental inclusion strategy. After a two-factor model is constructed, additional factors are added one by one to form three-factor, four-factor, and five-factor models. This group is designed to examine whether prediction performance improves as the model uses a richer multi-factor input structure. It also helps identify whether adding more factors always improves the model, or whether some factors introduce noise or structural complexity.

All Step 2 experiments use the derivative-learning black-box baseline selected in Step 1:

$$\lambda_{t+\Delta t}(x) = \lambda_t(x) + f_\theta(W, \lambda_t(x))\Delta t,$$

where W denotes the set of included spatial factors. Therefore, the difference among the Step 2 experiments comes from the input factor set rather than from the model form.

5.3.1 Step 2.1: Two-Factor Extension from Job Opportunity

The first group starts from job opportunity density alone and adds one additional factor at a time. This design directly tests whether commercial density, amenity density, transportation accessibility, or a weak irrelevant factor provides useful information beyond the job-opportunity baseline.

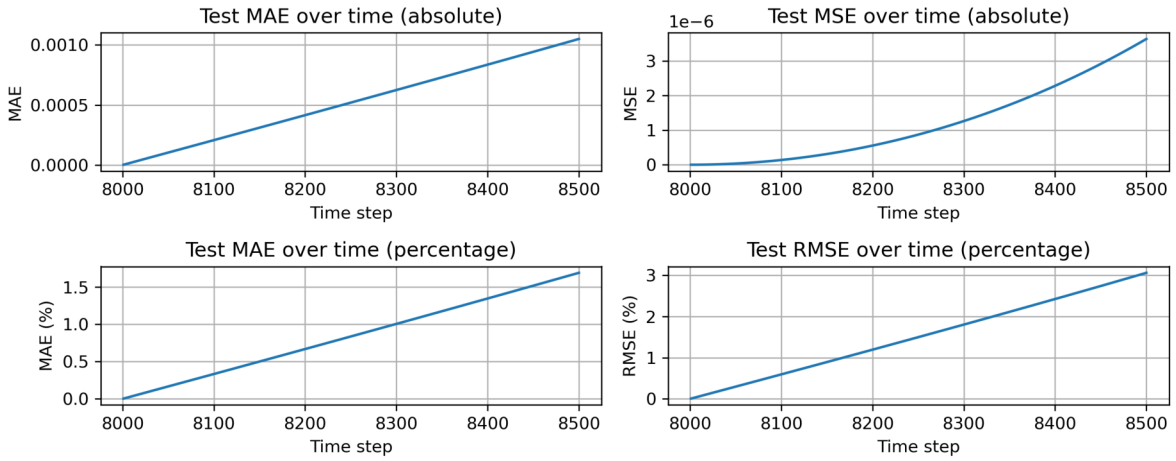


Figure 5.5: Short-horizon test error of Method 2.1.1 using job opportunity density only.

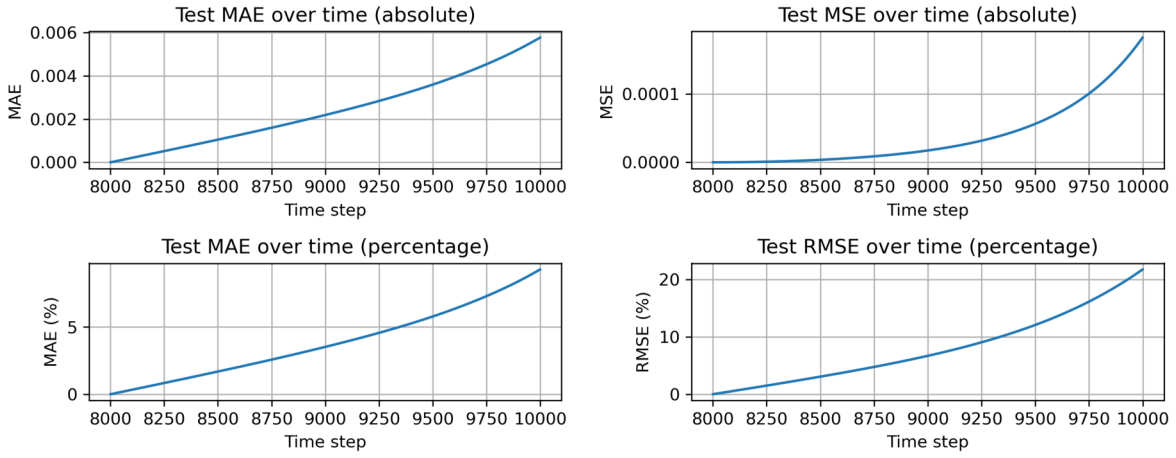


Figure 5.6: Long-horizon test error of Method 2.1.1 using job opportunity density only.

Figures 5.5 and 5.6 show the result of the job-only model. The short-horizon error remains moderate, but the long-horizon error increases substantially, suggesting that job opportunity alone is not sufficient to explain the full dynamic process.

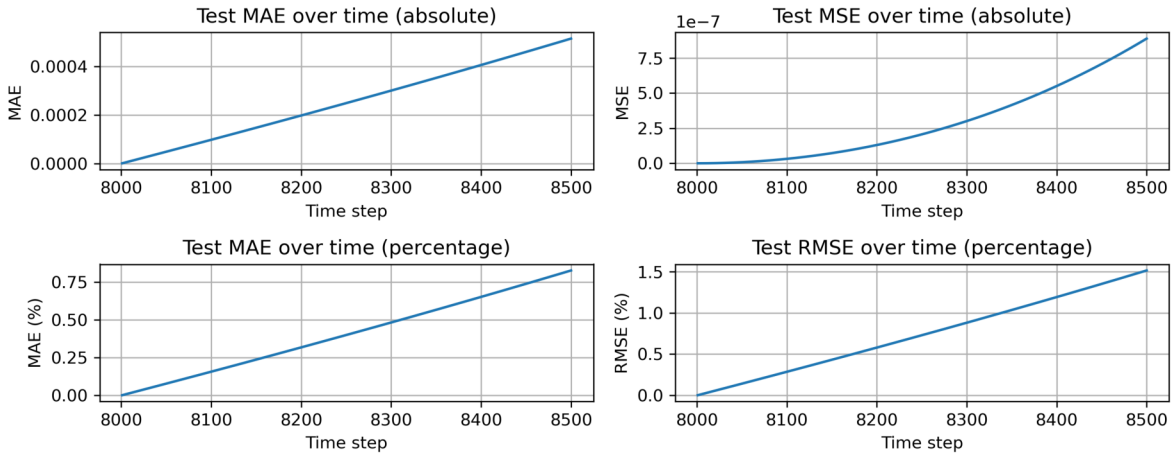


Figure 5.7: Short-horizon test error of Method 2.1.2 after adding commercial density.

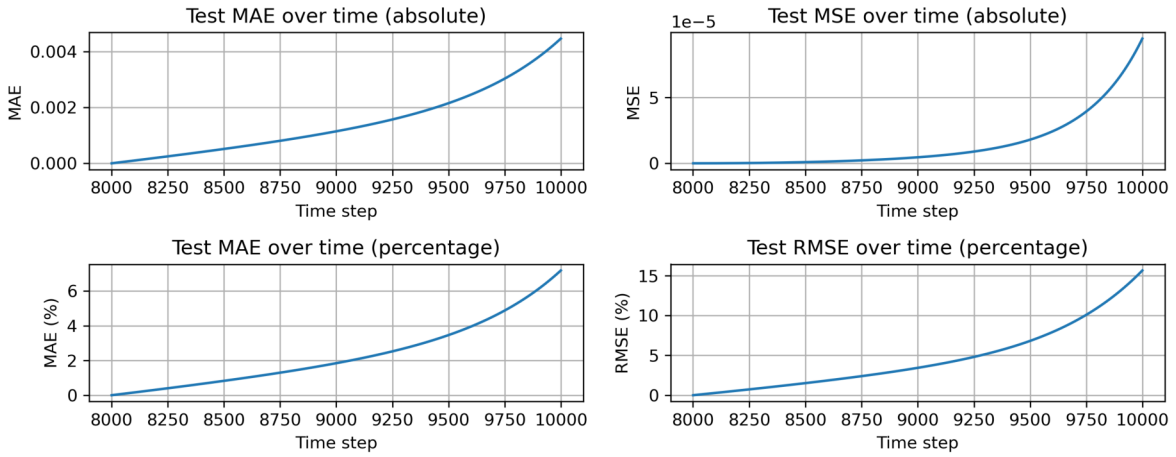


Figure 5.8: Long-horizon test error of Method 2.1.2 after adding commercial density.

Figures 5.7 and 5.8 show that adding commercial density improves the model relative to the job-only baseline. This indicates that commercial activity provides useful additional spatial information.

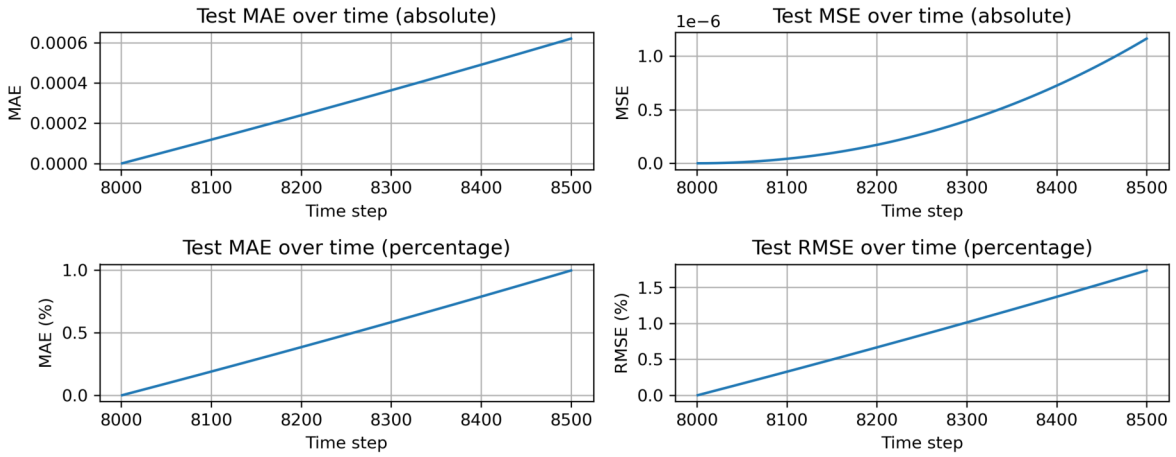


Figure 5.9: Short-horizon test error of Method 2.1.3 after adding amenity density.

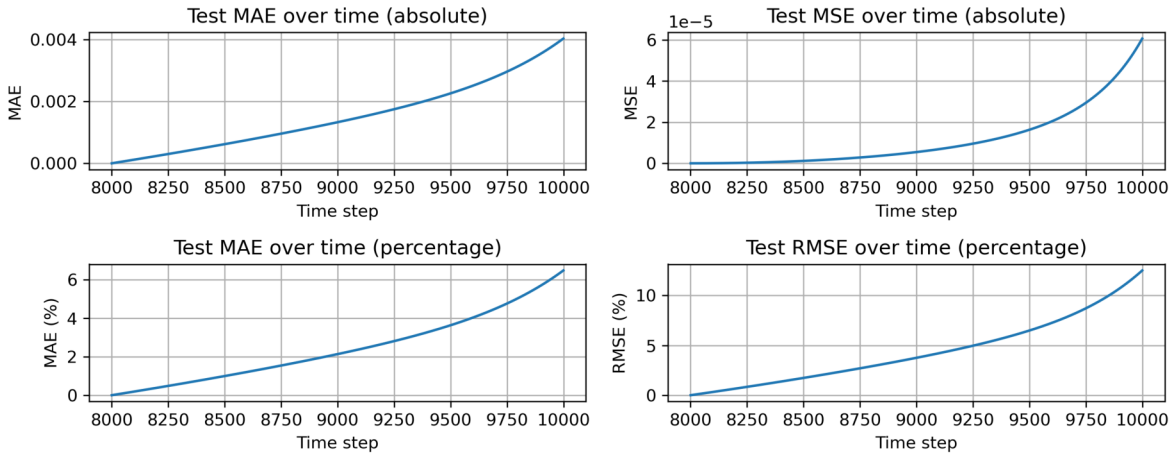


Figure 5.10: Long-horizon test error of Method 2.1.3 after adding amenity density.

Figures 5.9 and 5.10 show that amenity density also improves short-horizon prediction compared with the job-only model. However, the long-horizon error still accumulates over time.

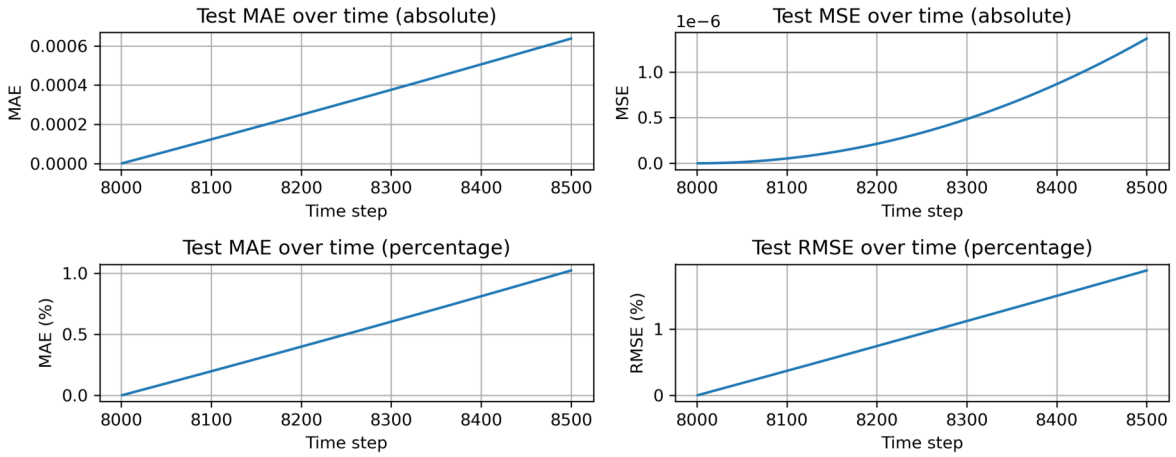


Figure 5.11: Short-horizon test error of Method 2.1.4 after adding transportation accessibility.

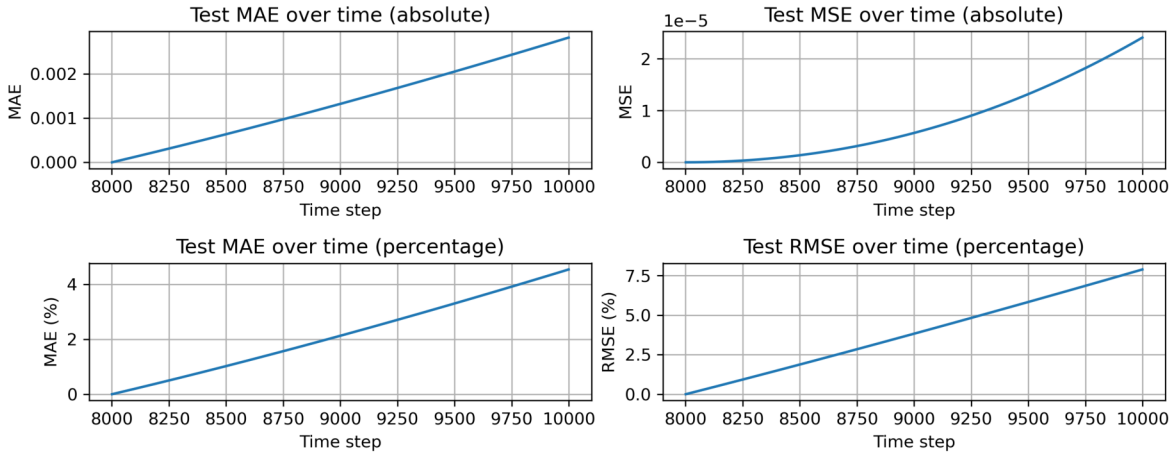


Figure 5.12: Long-horizon test error of Method 2.1.4 after adding transportation accessibility.

Figures 5.11 and 5.12 show that transportation accessibility provides additional information beyond job opportunity. Its long-horizon result is better than the job-only baseline, although recursive error accumulation is still visible.

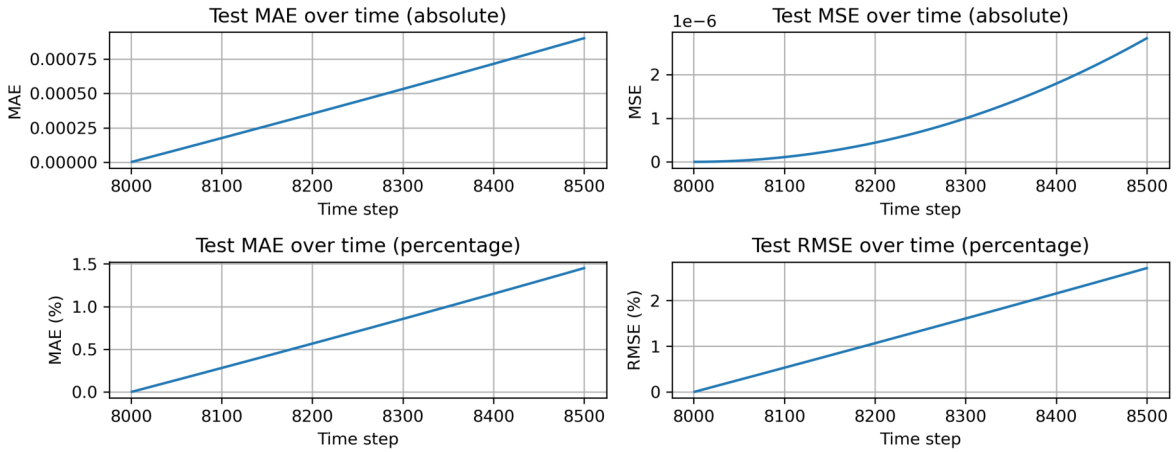


Figure 5.13: Short-horizon test error of Method 2.1.5 after adding the weak factor w_5.

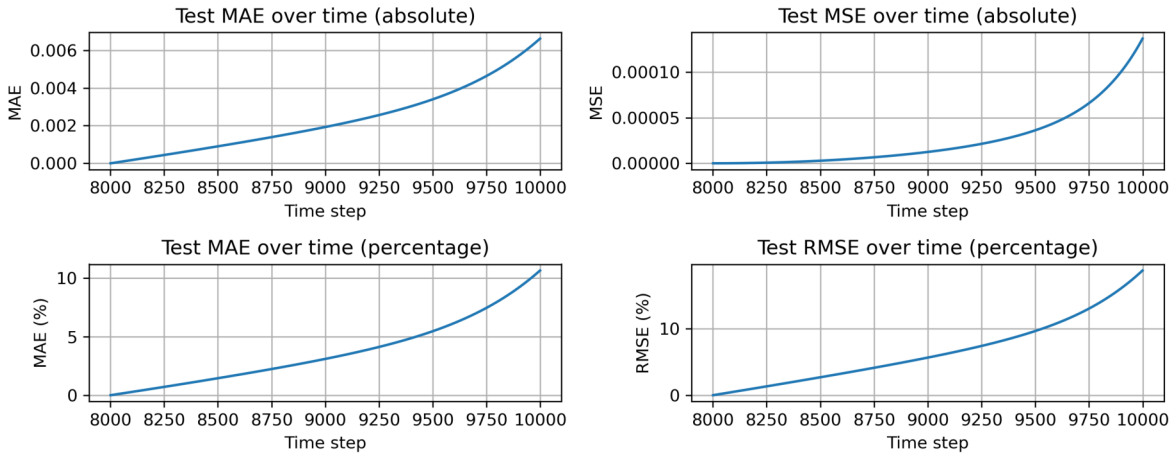


Figure 5.14: Long-horizon test error of Method 2.1.5 after adding the weak factor w_5.

Figures 5.13 and 5.14 show that the weak factor w_5 does not provide stable improvement. Its long-horizon error is much larger than the other two-factor models, suggesting that the framework can help distinguish useful spatial factors from weak or irrelevant candidates.

Table 5.2: Step 2.1 results: two-factor extension from job opportunity.

Experiment	**Model input**	**Short-horizon MAE (%)**	**Long-horizon MAE (%)**
Method 2.1.1	(w_1, λ)	0.8392	3.85
Method 2.1.2	(w_1, w_2, λ)	0.4048	2.33
Method 2.1.3	(w_1, w_3, λ)	0.4893	2.44
Method 2.1.4	(w_1, w_4, λ)	0.5058	2.17
Method 2.1.5	(w_1, w_5, λ)	0.7793	6.63

Table 5.2 summarizes the first group of factor-relevance tests. The job-only model has the largest short-horizon error among the meaningful factor combinations, which shows that job opportunity density alone is not sufficient. Adding commercial density produces the lowest short-horizon MAE, while adding amenity density or transportation accessibility also improves prediction relative to the baseline. In contrast, adding the weak factor w_5 does not provide reliable improvement and leads to the worst long-horizon error.

These results show that the proposed factor-relevance test can be used as a screening tool for deciding which spatial factors should be included in later analysis. By comparing the prediction errors after adding each candidate factor, the method helps identify whether a factor provides additional explanatory information beyond job opportunity density. The results suggest that commercial density, amenity density, and transportation accessibility should all be retained as candidate explanatory factors, while the weak factor w_5 should not be treated as a core factor. Therefore, this step provides a systematic basis for selecting the factor set before comparing theory-guided model structures.

5.3.2 Step 2.2: Incremental Multi-Factor Inclusion Test

The second group uses an incremental inclusion strategy. Starting from the job-only baseline, additional factors are added step by step to form two-factor, three-factor, four-factor, and five-factor models. This group is used to evaluate whether richer factor sets improve prediction performance and whether the model remains stable as more spatial information is included.

The incremental factor sets are:

$$(w_1, \lambda), \quad (w_1, w_2, \lambda), \quad (w_1, w_2, w_3, \lambda), \quad (w_1, w_2, w_4, \lambda), \quad (w_1, w_2, w_3, w_4, \lambda), \quad (w_1, w_2, w_3, w_4, w_5, \lambda).$$

The comparison between these models is useful because it evaluates whether adding more factors leads to monotonic improvement. In a multi-factor spatial system, this is not guaranteed. A factor may be relevant only in combination with another factor, or it may introduce additional noise if it is weakly related to the true dynamic process.

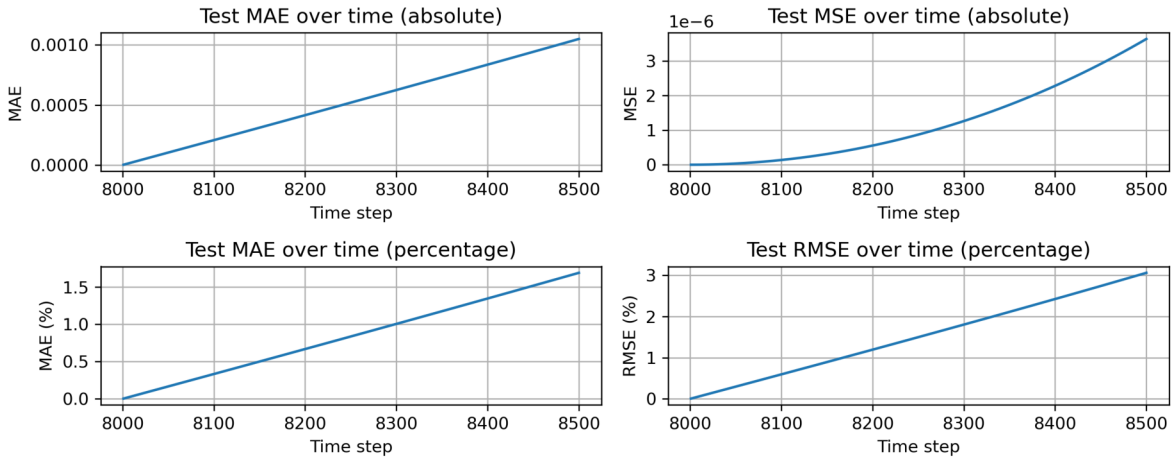


Figure 5.15: Short-horizon test error of Method 2.2.1 using job opportunity density only.

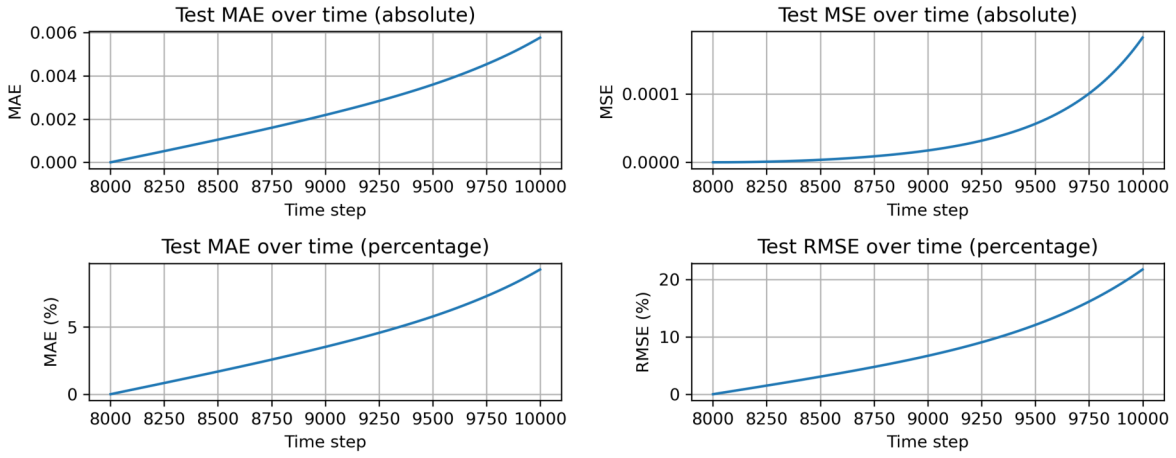


Figure 5.16: Long-horizon test error of Method 2.2.1 using job opportunity density only.

Figures 5.15 and 5.16 show the baseline result using only the job-opportunity factor. The long-horizon error increases clearly, confirming again that the primary factor alone is not enough for stable prediction.

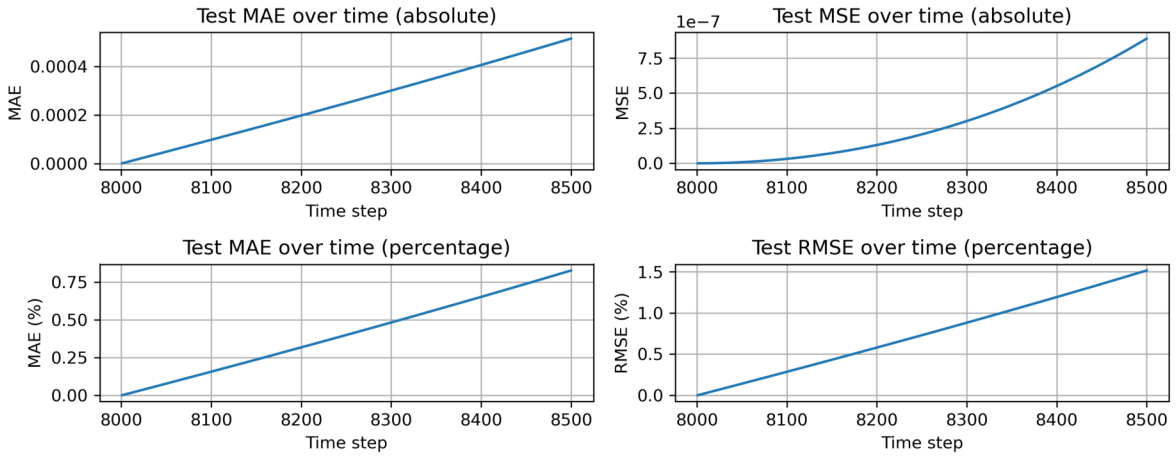


Figure 5.17: Short-horizon test error of Method 2.2.2 after adding commercial density.

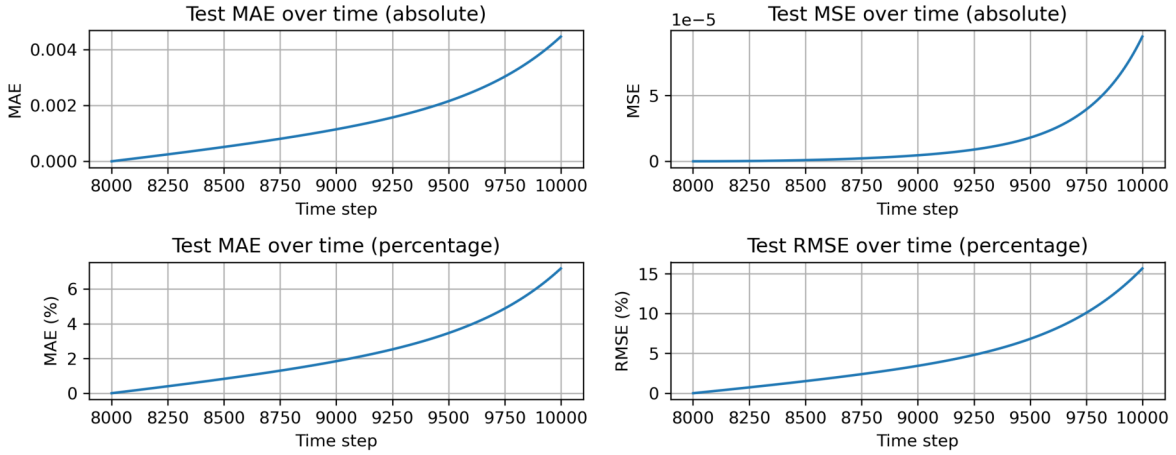


Figure 5.18: Long-horizon test error of Method 2.2.2 after adding commercial density.

Figures 5.17 and 5.18 show that adding commercial density improves performance compared with the job-only baseline. This supports the importance of commercial activity as a second explanatory factor.

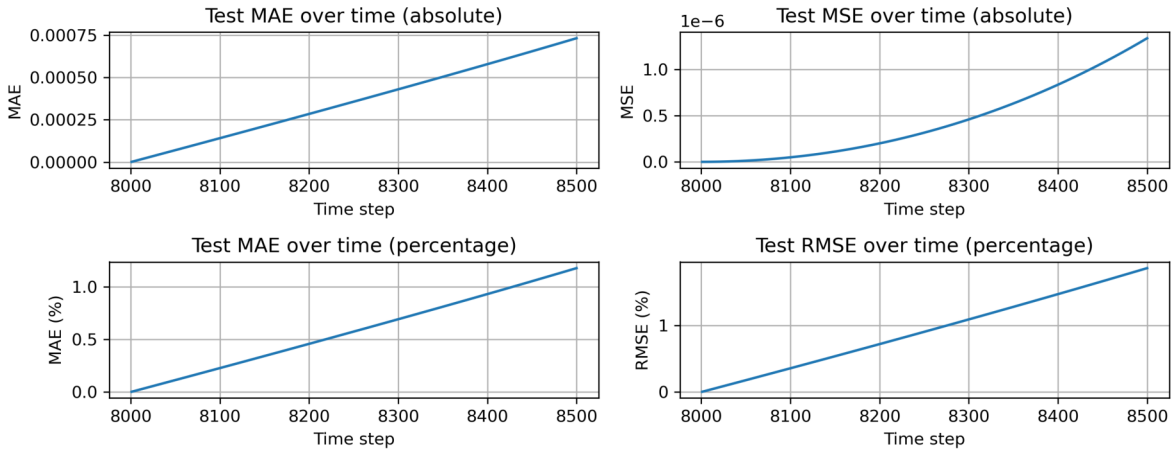


Figure 5.19: Short-horizon test error of Method 2.2.3 after adding amenity density to the job-commercial model.

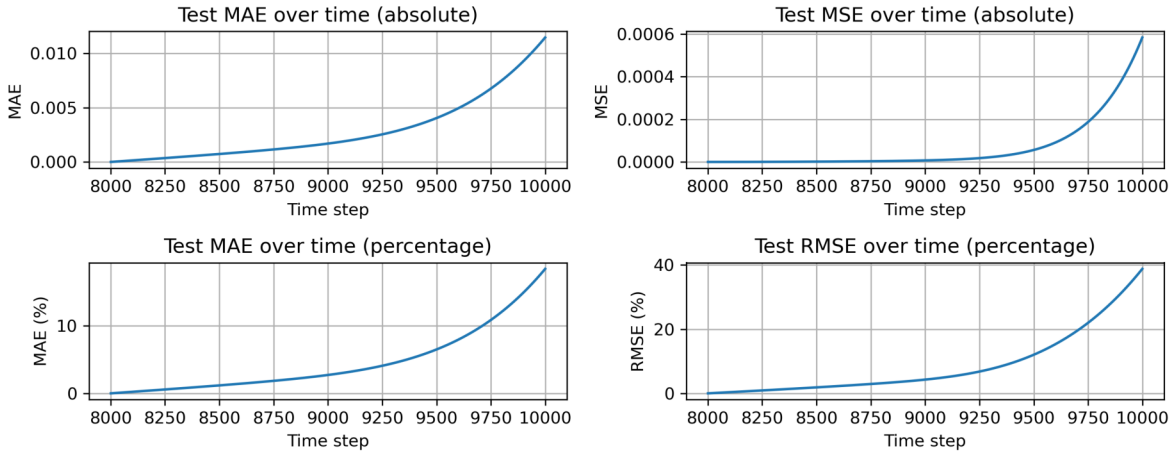


Figure 5.20: Long-horizon test error of Method 2.2.3 after adding amenity density to the job-commercial model.

Figures 5.19 and 5.20 show that adding amenity density does not necessarily improve prediction in this additive incremental setting. This suggests that the role of amenity density may not be well represented when it is added without the transportation factor.

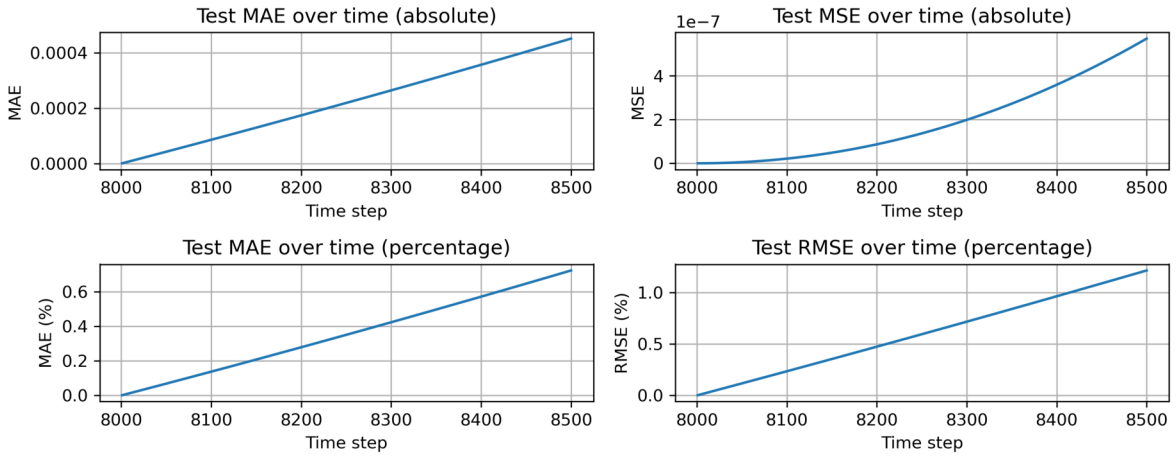


Figure 5.21: Short-horizon test error of Method 2.2.4 after adding transportation accessibility to the job-commercial model.

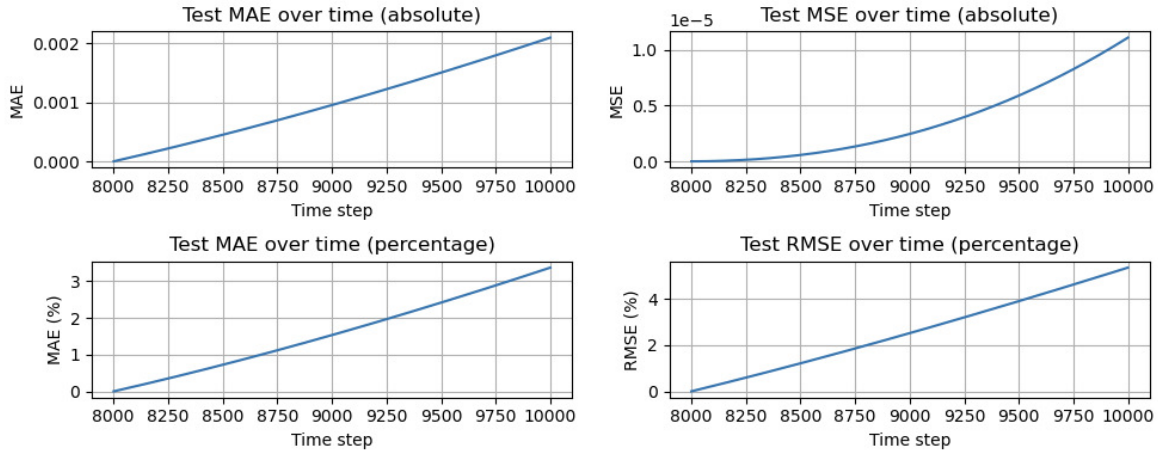


Figure 5.22: Long-horizon test error of Method 2.2.4 after adding transportation accessibility to the job-commercial model.

Figures 5.21 and 5.22 show that adding transportation accessibility to the job-commercial model gives a strong improvement. This indicates that transportation accessibility contains useful spatial information and can improve prediction when combined with the major economic factors.

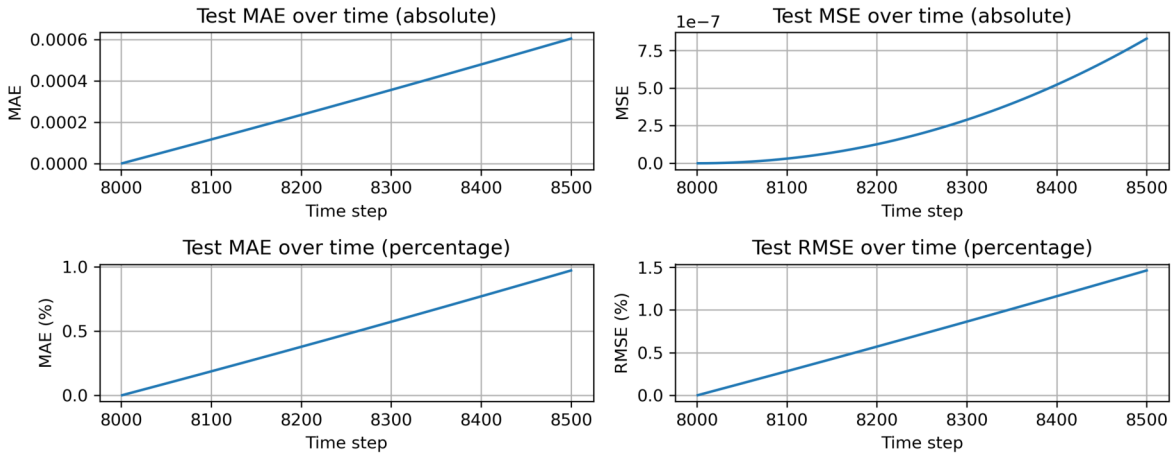


Figure 5.23: Short-horizon test error of Method 2.2.5 after adding amenity density to the job-commercial-transportation model.

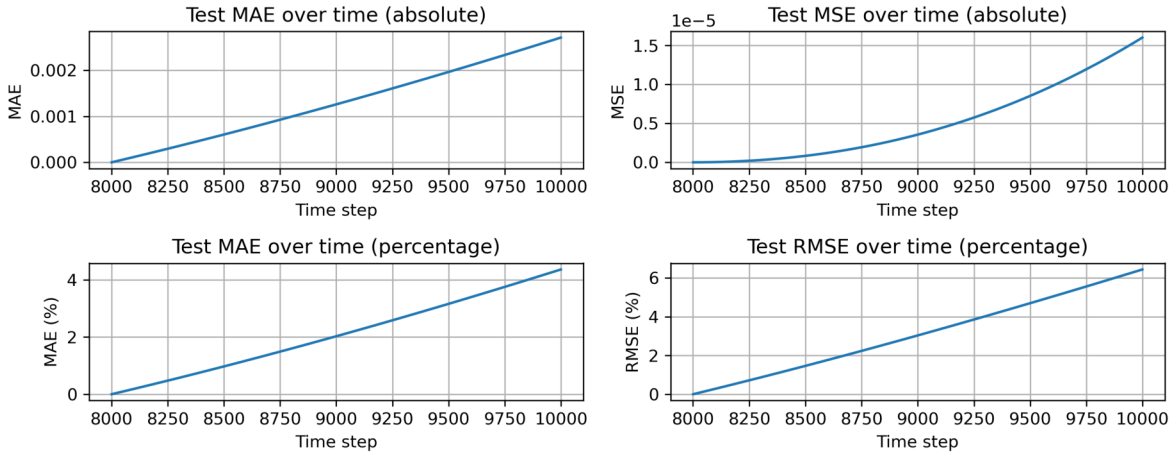


Figure 5.24: Long-horizon test error of Method 2.2.5 after adding amenity density to the job-commercial-transportation model.

Figures 5.23 and 5.24 show the performance of the full four-factor model. Compared with Method 2.2.4, adding amenity density does not further improve prediction accuracy in this incremental setting. This does not necessarily imply that amenity density is unimportant. Instead, it suggests that the contribution of amenity density may depend on the way it enters the dynamic process, which motivates the later factor-structure analysis.

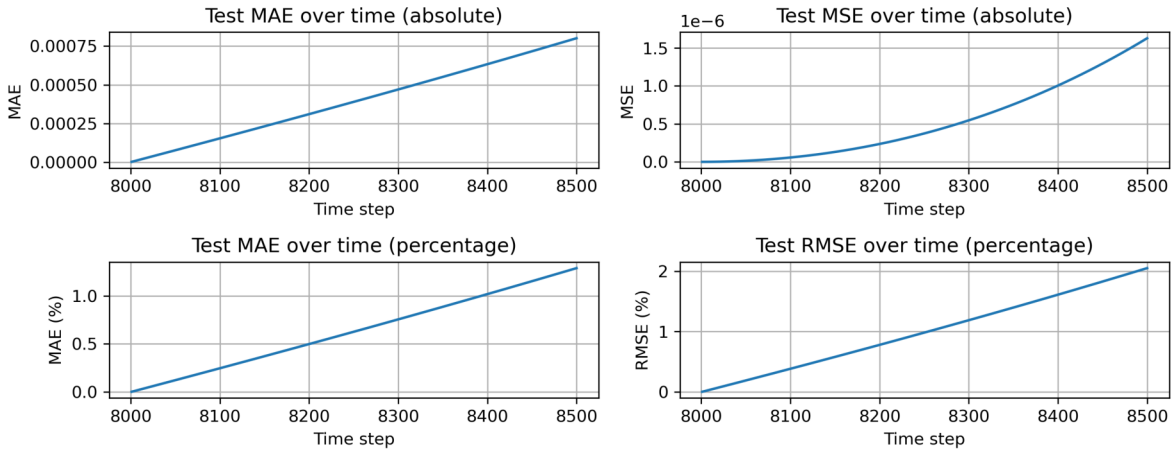


Figure 5.25: Short-horizon test error of Method 2.2.6 after adding the weak factor w_5 to the four-factor model.

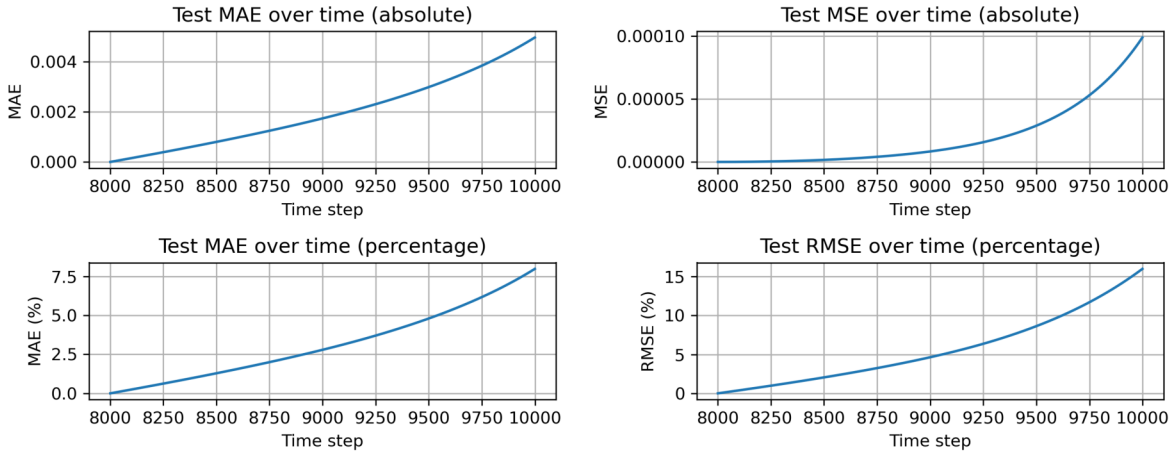


Figure 5.26: Long-horizon test error of Method 2.2.6 after adding the weak factor w_5 to the four-factor model.

Figures 5.25 and 5.26 show the result after adding the weak factor w_5 to the four-factor model. The short-horizon MAE is 0.63%, and the long-horizon MAE is 3.16%. Compared with Method 2.2.5, adding w_5 does not improve prediction performance and substantially increases the long-horizon error. Since w_5 is designed as a weak or irrelevant factor, this result is consistent with the expectation that non-informative inputs may increase model complexity without providing meaningful explanatory information.

Table 5.3: Step 2.2 results: incremental multi-factor inclusion test.

Experiment	**Model input**	**Short-horizon MAE (%)**	**Long-horizon MAE (%)**
Method 2.2.1	(w_1, λ)	0.8392	3.85
Method 2.2.2	(w_1, w_2, λ)	0.4048	2.33
Method 2.2.3	(w_1, w_2, w_3, λ)	0.5794	4.53
Method 2.2.4	(w_1, w_2, w_4, λ)	0.3554	1.58
Method 2.2.5	$(w_1, w_2, w_3, w_4, \lambda)$	0.4792	2.08
Method 2.2.6	$(w_1, w_2, w_3, w_4, w_5, \lambda)$	0.63	3.16

Table 5.3 summarizes the incremental inclusion results. The comparison again shows that w_1 alone is not sufficient and that adding w_2 substantially improves prediction. Among the meaningful three-factor models, (w_1, w_2, w_4, λ) achieves the best short-horizon and long-horizon performance.

However, these results should be interpreted carefully. The incremental inclusion strategy is more reliable when the true factor structure is close to a linear combination, such as

$$w_1 + w_2 + w_3 + w_4.$$

In such a case, adding one factor at a time can directly measure the marginal contribution of each factor. If a factor is useful, adding it should generally improve the prediction.

The situation becomes more complicated when the true structure contains a nonlinear or product term. In this simulation, the true potential function contains

$$w_3 w_4$$

rather than independent additive terms w_3 and w_4. Therefore, the true factor entering the model is closer to

$$w_1 + w_2 + w_3 w_4,$$

not

$$w_1 + w_2 + w_3 + w_4.$$

As a result, testing w_3 or w_4 separately after (w_1, w_2, λ) may not correctly represent their true contribution. One of the individual factors may have a spatial magnitude or distribution larger than the product $w_3 w_4$, while the other may have a weaker or incomplete overlap with the true product structure. This can cause the model to overestimate or underestimate the real effect of the nonlinear term. Therefore, the error may become larger or smaller than

expected even though the factor is related to the true function.

This explains why (w_1, w_2, w_3, λ) performs worse than (w_1, w_2, λ), while (w_1, w_2, w_4, λ) performs better. The result does not necessarily mean that w_3 is irrelevant or that w_4 alone is the correct explanatory factor. Instead, it shows that the effect of w_3 and w_4 is structural: they are designed to work together through the product term $w_3 w_4$. A simple incremental input test can detect useful factors when the relationship is approximately linear, but it may be misleading when the true relationship is nonlinear.

The four-factor model $(w_1, w_2, w_3, w_4, \lambda)$ performs reasonably well, but it does not dominate the best three-factor model in this black-box derivative-learning setting. This suggests that simply providing all factors to a black-box model does not guarantee that the model will efficiently recover the correct nonlinear structure. The five-factor model performs worse after adding w_5, especially in the long-horizon evaluation. Since w_5 is designed as a weak or irrelevant factor, this result should not be interpreted as meaningful explanatory evidence. Overall, Step 2 shows that factor selection should consider both prediction performance and the possible structural form through which factors enter the dynamic model.

5.3.3 Implications of Step 2

Step 2 should be understood as a factor-screening procedure rather than a final conclusion about the true importance of each factor. The purpose of this step is to provide a practical way to decide which candidate factors should be kept for later structural analysis. By comparing prediction errors after adding different factors, the framework can identify whether a factor contains useful information that should not be excluded too early.

The first group of experiments provides this initial screening result. When each candidate factor is added to the job-opportunity baseline, the models using w_2, w_3, and w_4 all improve the prediction relative to the job-only case, while the weak factor w_5 does not provide stable improvement. Therefore, the result suggests that w_1, w_2, w_3, and w_4 should all be retained for later consideration, whereas w_5 should not be treated as a core explanatory factor.

The second group of experiments further shows the limitation of simple incremental inclusion. This procedure works well when the true structure is close to a linear additive combination. In that case, adding one factor at a time can provide a reasonable test of whether that factor should be included. However, when the true function contains a nonlinear or product term, the incremental test may not correctly evaluate the contribution of each factor. In this simulation, the true potential contains $w_3 w_4$, so testing w_3 or w_4 separately does not fully represent their joint effect.

Therefore, Step 2 suggests that factor inclusion and factor structure are two different

questions. The first group helps decide which factors should be included in the candidate factor set. The second group shows that inclusion tests alone are not enough to determine how those factors enter the model. A factor may need to be included, but it may not enter the potential as an independent linear component.

For this reason, the later structural analysis focuses mainly on the form of factor entry. In particular, the following factor-structure test examines whether each retained factor can be represented linearly in the population potential, or whether it requires a more flexible nonlinear representation. This makes the pipeline more systematic: Step 2 first screens the factor set, and the later steps then test the appropriate structure of those factors.

Fortunately, This issue is also consistent with the modeling logic of many simplified spatial economic and residential-location-choice frameworks. In such models, location attributes such as employment opportunity, commercial activity, and amenities are often represented through additive or approximately linear components of a utility or attractiveness function. Transportation accessibility may be more complicated, because its effect depends not only on its own spatial distribution, but also on which locations it connects and which opportunities, amenities, or commercial centers become accessible through the transport network. Therefore, after Step 2 identifies the candidate factor set, the later analysis addresses this potential structure problem by comparing model structures and testing factor linearity.

5.4 Step 3: Theory-Guided Model-Structure Comparison

After the candidate factors are screened in Step 2, the next question is how these factors should be organized into a dynamic model structure. In the case studied in this thesis, Step 2 suggests that the four main spatial factors w_1, w_2, w_3, and w_4 should be retained for later consideration. Therefore, the corresponding black-box derivative-learning baseline is written as

$$\frac{\partial \lambda}{\partial t} = f_\theta(w_1, w_2, w_3, w_4, \lambda).$$

This model is the same four-factor derivative-learning model as Method 1.2, and it also appears as Method 2.2.5 in the incremental factor-inclusion test.

The purpose of Step 3 is to compare whether different theory-guided model structures can better explain the same dynamic process. The black-box model is flexible, but it does not explicitly specify the mechanism of population movement. Theory-guided models, by contrast, impose interpretable structures, such as advection–diffusion movement or wage-difference adjustment. Therefore, Step 3 tests whether the proposed methodology can select

the theoretical structure that best matches the observed data-generating process.

5.4.1 Method 3.0: Black-Box Baseline

The black-box baseline learns the local population change directly:

$$\frac{\partial \lambda}{\partial t} = f_\theta(w_1, w_2, w_3, w_4, \lambda).$$

This method uses the four factors selected from Step 2 and the current population density as inputs. It does not impose any specific theory-guided structure. Therefore, it provides a flexible benchmark for evaluating whether a structured model can improve prediction and interpretation.

5.4.2 Method 3.1: Multi-Factor Advection–Diffusion Model

The multi-factor advection–diffusion model follows the advection–diffusion structure introduced in the literature review. The original single-factor version can be written in two-dimensional divergence form as

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot (D \nabla \lambda - \lambda \nabla w), \qquad D > 0.$$

In this model, w is a single spatial attractiveness field.

In Method 3.1, the same advection–diffusion structure is retained, but the single factor w is replaced by a learned multi-factor potential:

$$w \quad \longrightarrow \quad f_\theta(w_1, w_2, w_3, w_4).$$

The resulting model is

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot (D \nabla \lambda - \lambda \nabla f_\theta(w_1, w_2, w_3, w_4)).$$

This formulation allows the selected factors from Step 2 to enter the population movement process through a learned spatial potential, while keeping the same diffusion–advection interpretation as the original model.

5.4.3 Method 3.2: Multi-Factor Wage-Difference Model

The wage-difference model is another theory-guided structure introduced in the literature review:

$$\frac{\partial \lambda}{\partial t}(t,x) = v\Big[w(t,x) - \bar{w}(t)\Big]\lambda(t,x), \qquad v > 0.$$

In this model, population changes according to whether the local attractiveness field $w(t,x)$ is above or below the average attractiveness level $\bar{w}(t)$.

In Method 3.2, the same modeling idea is extended to the multi-factor setting by replacing the single factor w with a learned multi-factor potential:

$$w \quad \longrightarrow \quad f_\theta(w_1, w_2, w_3, w_4).$$

The corresponding multi-factor wage-difference model is

$$\frac{\partial \lambda}{\partial t} = v\left[f_\theta(w_1, w_2, w_3, w_4) - \bar{f_\theta}\right]\lambda, \qquad v > 0.$$

This model keeps the original wage-difference logic, but allows the attractiveness field to be constructed from multiple spatial factors.

5.4.4 Method 3.3: Single-Factor Advection–Diffusion Model

Method 3.3 is the single-factor version of the advection–diffusion model:

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot \left(D\nabla\lambda - \lambda\nabla f_\theta(w_1)\right), \qquad D > 0.$$

Here, the spatial potential is restricted to one factor, namely job opportunity density w_1. This model is included to test whether the advection–diffusion interpretation remains effective when the potential is limited to a single dominant factor.

Compared with Method 3.1, Method 3.3 uses the same advection–diffusion structure but a more restricted potential function. Therefore, comparing Method 3.1 and Method 3.3 helps evaluate whether the multi-factor extension improves the explanatory capacity of the same theoretical structure.

5.4.5 Experimental Results

This subsection compares four methods. Method 3.0 is the black-box derivative-learning baseline. Method 3.1 is the multi-factor advection–diffusion model. Method 3.2 is the multi-factor wage-difference model. Method 3.3 is the single-factor advection–diffusion model.

Method 3.0: Black-Box Baseline

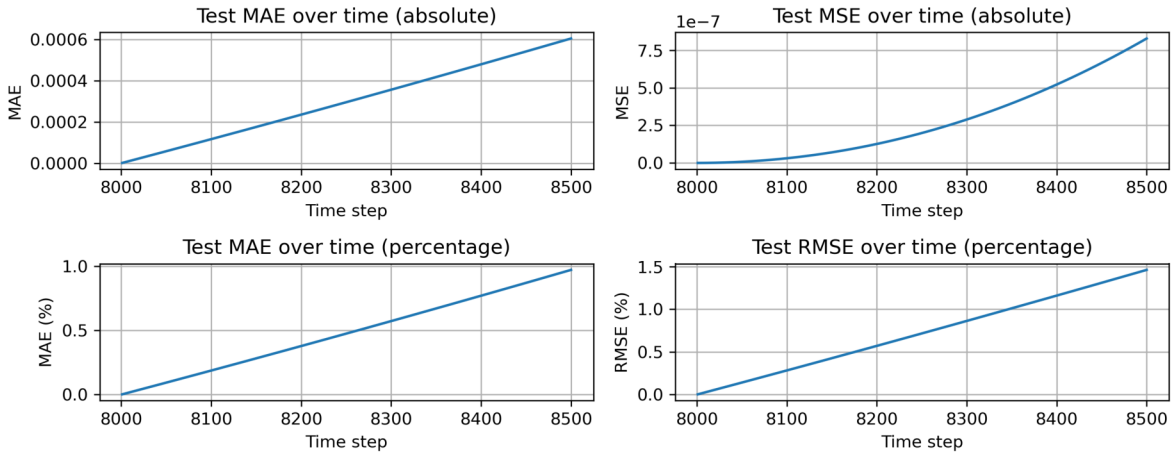


Figure 5.27: Short-horizon test error of Method 3.0, the four-factor black-box derivative-learning baseline.

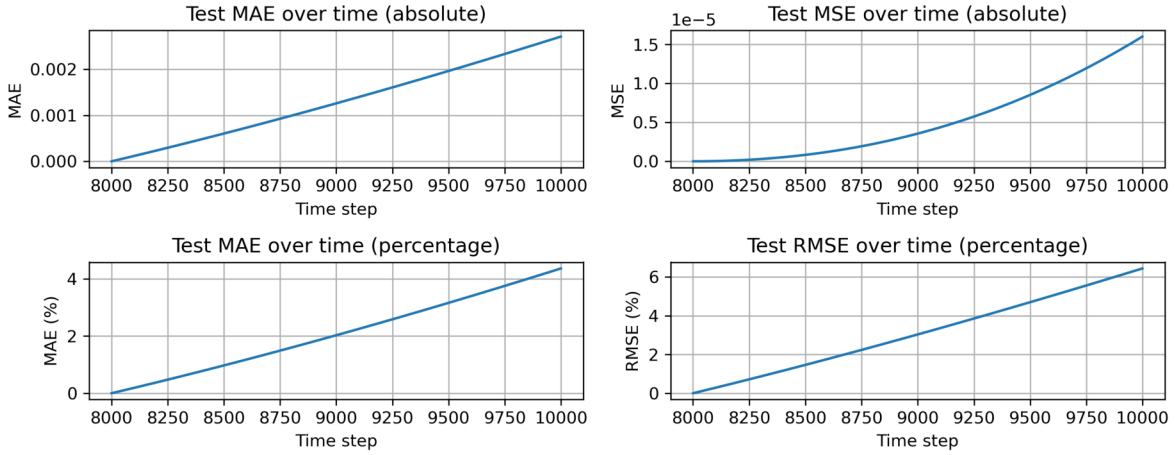


Figure 5.28: Long-horizon test error of Method 3.0, the four-factor black-box derivative-learning baseline.

Method 3.0 directly learns the local population change from the four selected factors and the current population density. It provides the black-box reference for evaluating whether a theory-guided structure improves performance.

Method 3.1: Multi-Factor Advection–Diffusion Model

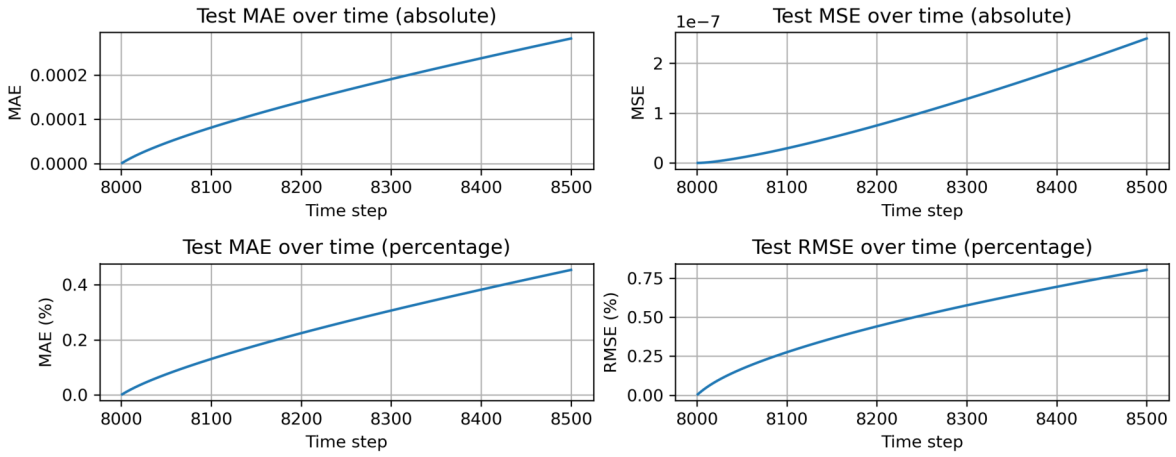


Figure 5.29: Short-horizon test error of Method 3.1, the multi-factor advection–diffusion model.

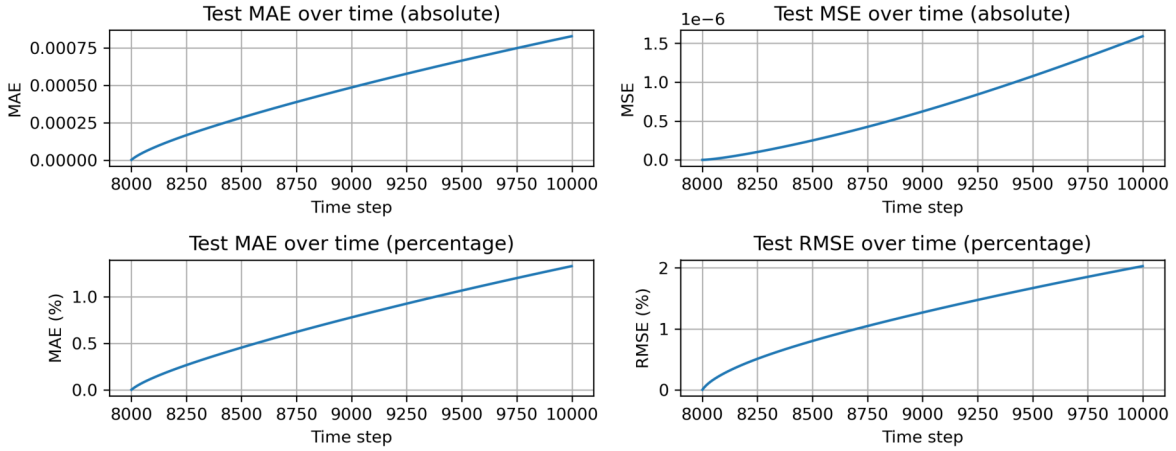


Figure 5.30: Long-horizon test error of Method 3.1, the multi-factor advection–diffusion model.

Method 3.1 learns a multi-factor potential and places it inside the advection–diffusion structure. It uses the same four factors as Method 3.0, but imposes a spatial movement mechanism.

Method 3.2: Multi-Factor Wage-Difference Model

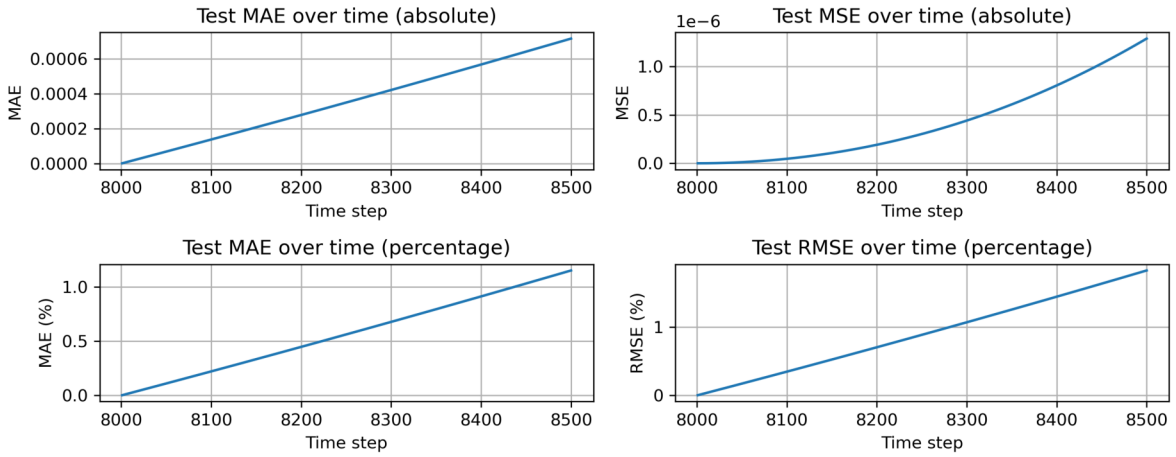


Figure 5.31: Short-horizon test error of Method 3.2, the multi-factor wage-difference model.

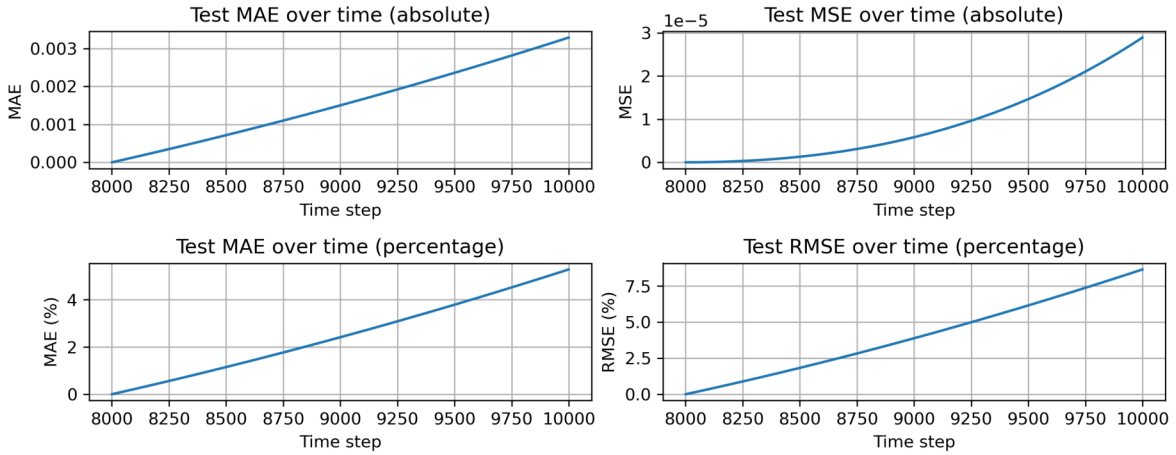


Figure 5.32: Long-horizon test error of Method 3.2, the multi-factor wage-difference model.

Method 3.2 uses the same learned multi-factor potential, but places it into the wage-difference structure. It updates population according to whether the local potential is above or below the spatial average potential.

Method 3.3: Single-Factor Advection–Diffusion Model

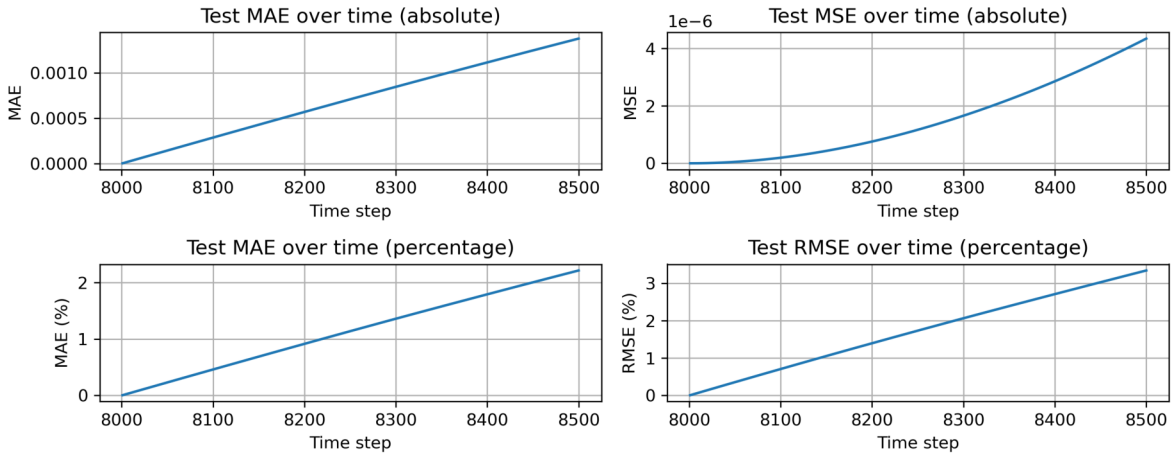


Figure 5.33: Short-horizon test error of Method 3.3, the single-factor advection–diffusion model.

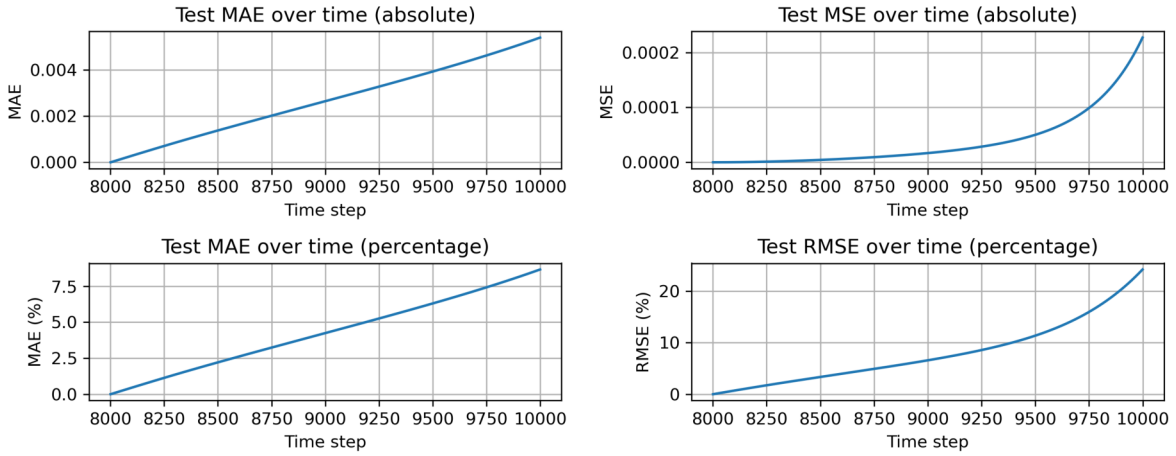


Figure 5.34: Long-horizon test error of Method 3.3, the single-factor advection–diffusion model.

Method 3.3 retains the advection–diffusion structure but uses only the job opportunity factor to construct the potential. It tests whether a single-factor advection–diffusion model is sufficient for the simulated data.

Table 5.4: Step 3 results: comparison of model structures using MAE.

Method	**Model structure**	**Short MAE (%)**	**Long MAE (%)**
Method 3.0	Black-box derivative-learning model	0.4792	2.08
Method 3.1	Multi-factor advection–diffusion model	0.2567	0.75
Method 3.2	Multi-factor wage-difference model	0.57	2.49
Method 3.3	Single-factor advection–diffusion model	1.13	4.27

Table 5.4 compares the four candidate model structures using short-horizon and long-horizon MAE. Method 3.0 provides the flexible black-box benchmark. Method 3.1 tests the multi-factor advection–diffusion interpretation. Method 3.2 tests the multi-factor wage-difference interpretation. Method 3.3 tests the single-factor advection–diffusion interpretation.

The results show that Method 3.1 achieves the best overall performance. Its short-horizon MAE is 0.2567%, and its long-horizon MAE is 0.75%. Compared with the black-box baseline, the multi-factor advection–diffusion model reduces the short-horizon error from 0.4792% to 0.2567% and the long-horizon error from 2.08% to 0.75%. This suggests that, after the relevant factors are retained, an appropriate theory-guided structure can improve both prediction accuracy and long-horizon stability.

The comparison between Method 3.1 and Method 3.2 shows that not every theory-guided structure fits the simulated dynamics equally well. Although the wage-difference model uses a meaningful attractiveness-difference mechanism, it does not explicitly model local spatial movement through gradients and flux. Its long-horizon MAE is 2.49%, which is much higher than the 0.75% long-horizon MAE of the multi-factor advection–diffusion model. This indicates that the advection–diffusion structure is more consistent with the simulated data-generating process.

The comparison between Method 3.1 and Method 3.3 further confirms the importance of using multiple factors within the same theoretical framework. Both methods are based on the advection–diffusion theory, but Method 3.1 constructs the potential from four selected spatial factors, while Method 3.3 restricts the potential to the single job-opportunity factor. Method 3.3 has a short-horizon MAE of 1.13% and a long-horizon MAE of 4.27%, which are much larger than the corresponding errors of Method 3.1. Therefore, within the same advection–diffusion theory, the multi-factor version performs substantially better than the

single-factor version.

The comparison between Method 3.3 and Method 3.0 is also important. Method 3.3 uses the correct type of theory-guided structure, namely advection–diffusion, but it uses only one factor. Method 3.0 does not impose the advection–diffusion structure, but it uses the four-factor input set selected from Step 2. Since Method 3.0 performs better than Method 3.3, this comparison suggests that factor selection has a stronger influence than theory selection when the factor set is incomplete. A correct theoretical structure alone is not sufficient if the model does not include the relevant explanatory factors.

Overall, Step 3 leads to three conclusions. First, the proposed pipeline can be used to select the theory-guided model structure that best matches the observed or simulated population dynamics. In this experiment, the multi-factor advection–diffusion model provides the best theoretical explanation among the tested candidates. Second, within the same advection–diffusion framework, the multi-factor version performs much better than the single-factor version, which supports the need for multi-factor modeling. Third, the comparison between Method 3.3 and Method 3.0 suggests that the factor set should be determined before selecting the theoretical structure. Therefore, the recommended order of the methodology is to first identify the relevant factors, and then compare which theory-guided model best explains the dynamics generated by those factors.

5.5 Step 4: Factor-Structure Test

After Step 3 selects the multi-factor advection–diffusion model as the most suitable theory-guided structure, the next question is how each factor enters the learned population potential. Step 2 identifies which factors should be retained, and Step 3 selects the dynamic structure. Step 4 further examines whether each retained factor can be represented as an independent linear component of the potential function.

Before presenting the restricted models, it is necessary to explain why Step 4 focuses on linearity testing. Step 2 has already served as a factor-screening stage. Its purpose is to decide which candidate factors should be retained for later analysis. The results suggest that w_1, w_2, w_3, and w_4 should all be considered, while the weak factor w_5 should not be treated as a core explanatory factor.

However, Step 2 also shows a limitation of simple factor-inclusion tests. Adding one factor at a time is useful when the underlying structure is close to a linear additive form. For example, if the true potential were approximately

$$f(w_1, w_2, w_3, w_4) = \beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3 + \beta_4 w_4,$$

then the marginal contribution of each factor could be evaluated relatively directly by adding it to the model. In this case, a factor that improves prediction after being added can be interpreted as providing additional explanatory information.

The situation is different when the true factor structure is nonlinear. In the present simulation, the true potential contains the product term

$$w_3 w_4,$$

so the relevant structure is closer to

$$f_{\text{true}}(w_1, w_2, w_3, w_4) = a_1 w_1 + a_2 w_2 + a_3 w_3 w_4.$$

Therefore, testing w_3 or w_4 separately in Step 2 cannot fully represent their joint effect. One factor may appear to improve prediction, while the other may appear less useful, even though the product of the two factors is part of the true data-generating mechanism.

For this reason, Step 4 does not simply ask whether a factor should be included. That question has already been addressed in Step 2. Instead, Step 4 asks how each retained factor enters the selected model structure. More specifically, it tests whether each factor can be represented as an independent linear component of the population potential, or whether it requires a more flexible nonlinear representation.

This focus is also consistent with the modeling logic of many simplified spatial economic and residential-location-choice frameworks. In such models, location attributes such as employment opportunity, commercial activity, and amenities are often represented through additive or approximately linear components of a utility or attractiveness function. Transportation accessibility may be more complicated, because its effect depends not only on its own spatial distribution, but also on which locations it connects and which opportunities, amenities, or commercial centers become accessible through the transport network. Therefore, after Step 2 identifies the candidate factor set, Step 4 focuses on linearity testing as a practical way to diagnose the structure of factor entry.

The baseline model is the multi-factor advection–diffusion model from Method 3.1:

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot \left(D \nabla \lambda - \lambda \nabla f_\theta(w_1, w_2, w_3, w_4) \right).$$

In this baseline model, the full potential function is learned flexibly by a neural network.

To test factor structure, restricted models are constructed. This design is related to the broader idea of partially linear modeling, where one component is restricted to a linear form while the remaining component is learned flexibly. In Methods 4.1–4.4, one selected factor

is forced to enter the potential linearly, while the remaining factors are learned flexibly by a neural network. The purpose is to examine whether each selected factor is compatible with a linear representation.

The four single-factor restricted potential functions are:

$$f(w_1, w_2, w_3, w_4) = \beta_1 w_1 + f_\theta(w_2, w_3, w_4),$$

$$f(w_1, w_2, w_3, w_4) = \beta_2 w_2 + f_\theta(w_1, w_3, w_4),$$

$$f(w_1, w_2, w_3, w_4) = \beta_3 w_3 + f_\theta(w_1, w_2, w_4),$$

$$f(w_1, w_2, w_3, w_4) = \beta_4 w_4 + f_\theta(w_1, w_2, w_3).$$

If imposing a linear form improves or maintains prediction accuracy, then the corresponding factor can be reasonably interpreted as approximately linear in the population potential. If imposing a linear form worsens prediction substantially, then the factor is likely to require a more flexible or nonlinear representation.

Based on the results of the single-factor linearity tests, an additional restricted model is introduced as Method 4.5. Since Methods 4.1 and 4.2 suggest that w_1 and w_2 are compatible with linear representations, while Methods 4.3 and 4.4 suggest that w_3 and w_4 are not well represented as independent linear components, Method 4.5 tests the combined structure

$$f(w_1, w_2, w_3, w_4) = \beta_1 w_1 + \beta_2 w_2 + f_\theta(w_3, w_4).$$

This model is used as a double-check of the factor-structure interpretation. It asks whether the model performs better when the two factors that appear linear are forced to enter linearly, while the remaining two factors are learned flexibly together.

5.5.1 Linear Test for w_1

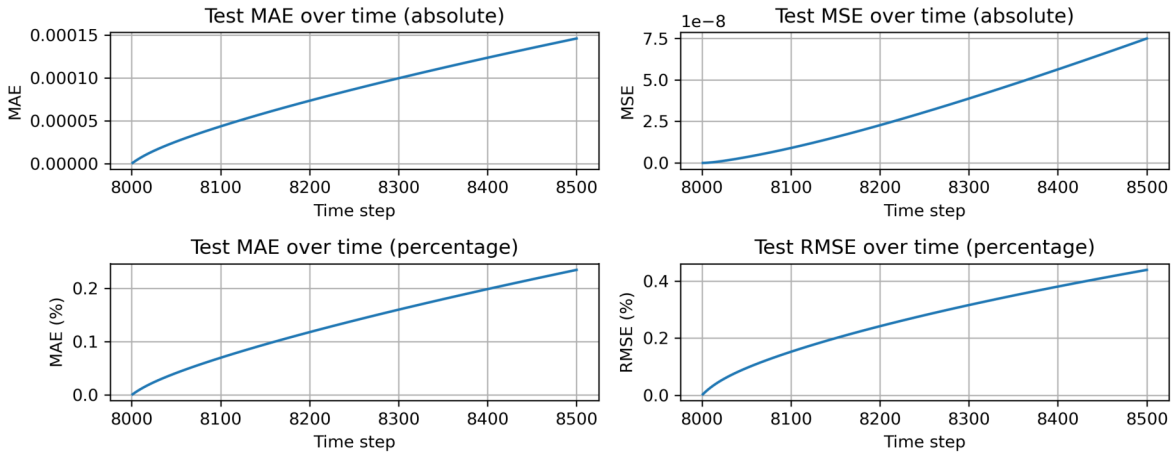


Figure 5.35: Short-horizon test error when w_1 is forced to enter the potential linearly.

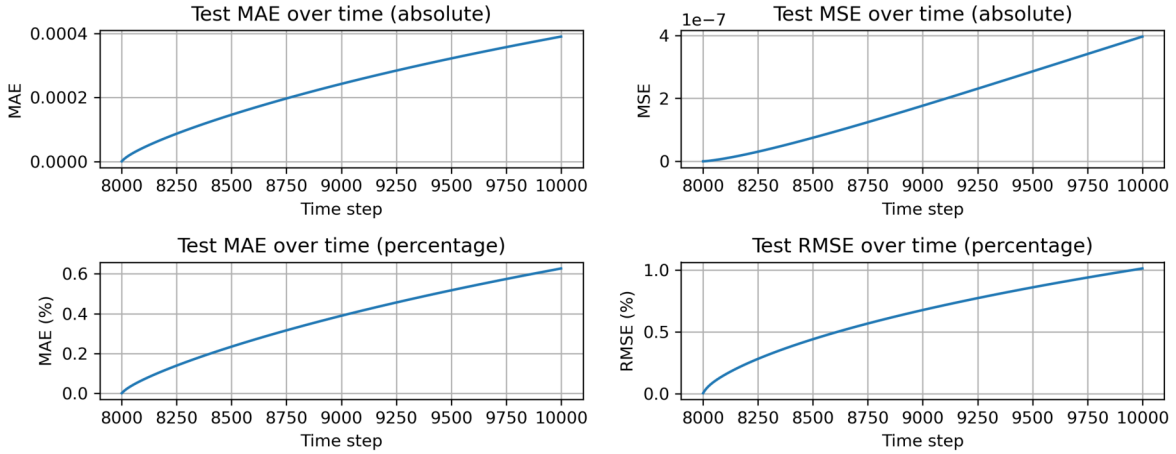


Figure 5.36: Long-horizon test error when w_1 is forced to enter the potential linearly.

This model tests whether job opportunity density can be represented as an independent linear component of the population potential.

5.5.2 Linear Test for w_2

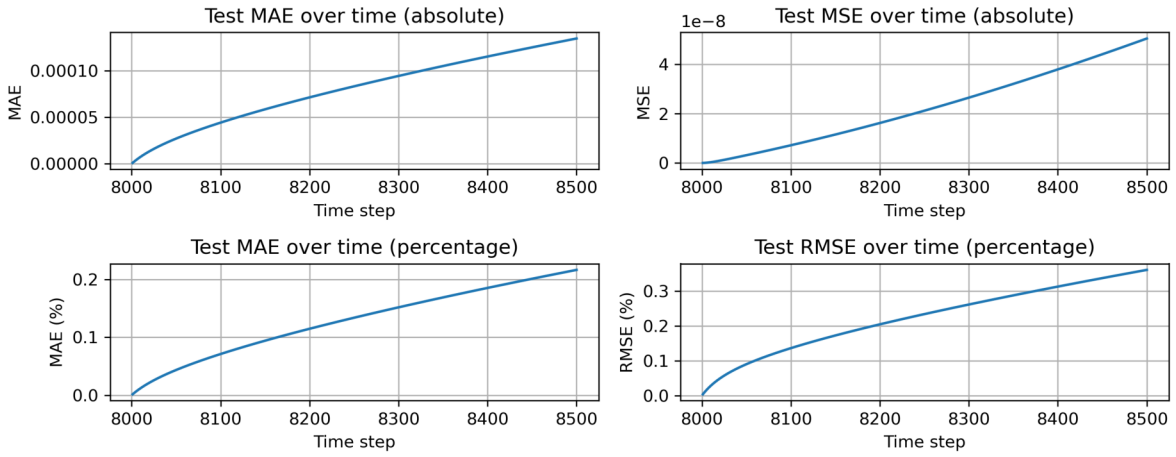


Figure 5.37: Short-horizon test error when w_2 is forced to enter the potential linearly.

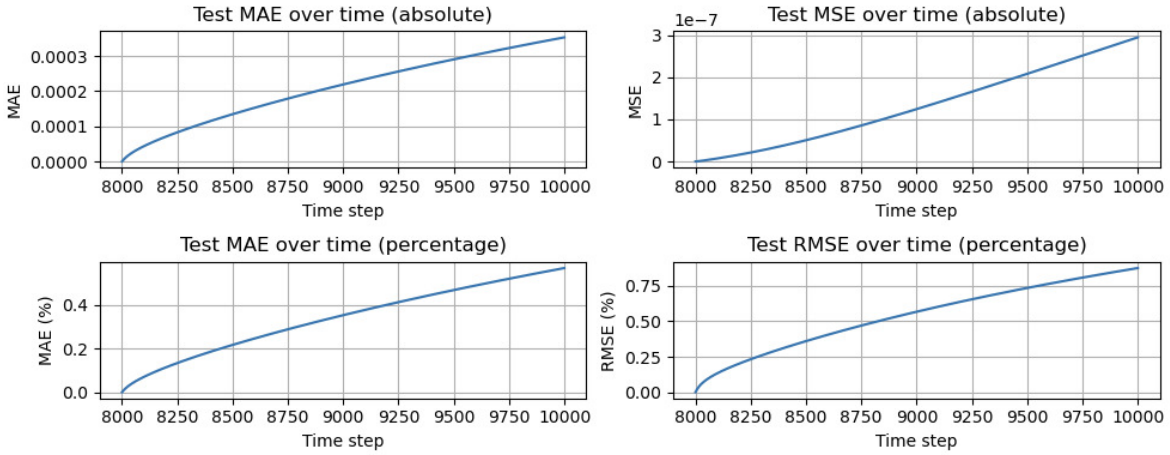


Figure 5.38: Long-horizon test error when w_2 is forced to enter the potential linearly.

This model tests whether commercial density can be represented as an independent linear component of the population potential.

5.5.3 Linear Test for w_3

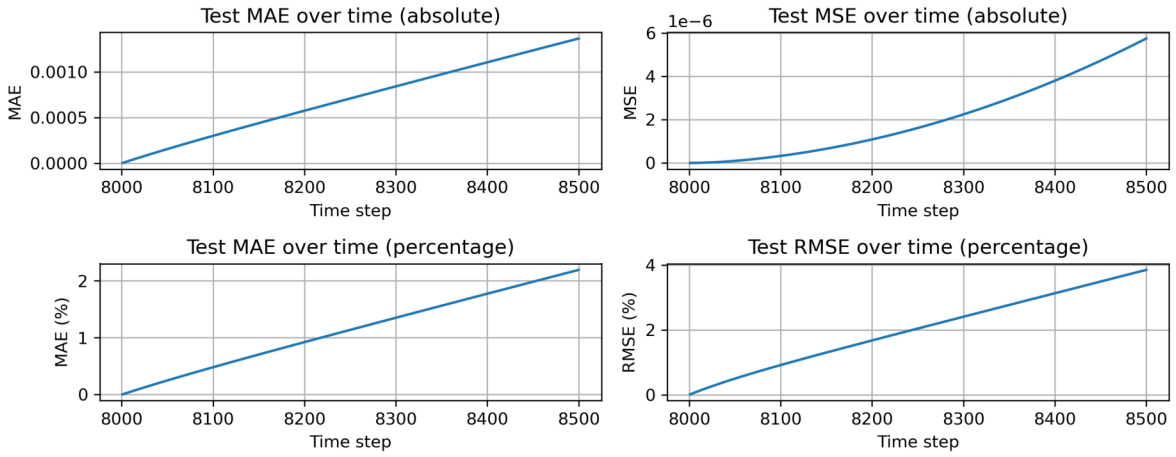


Figure 5.39: Short-horizon test error when w_3 is forced to enter the potential linearly.

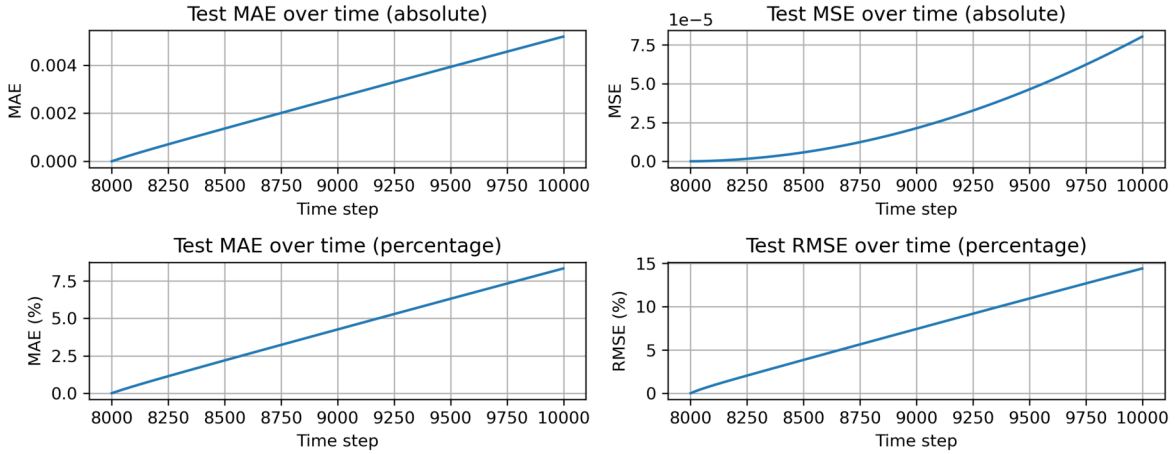


Figure 5.40: Long-horizon test error when w_3 is forced to enter the potential linearly.

This model tests whether amenity density can be represented as an independent linear component of the population potential.

5.5.4 Linear Test for w_4

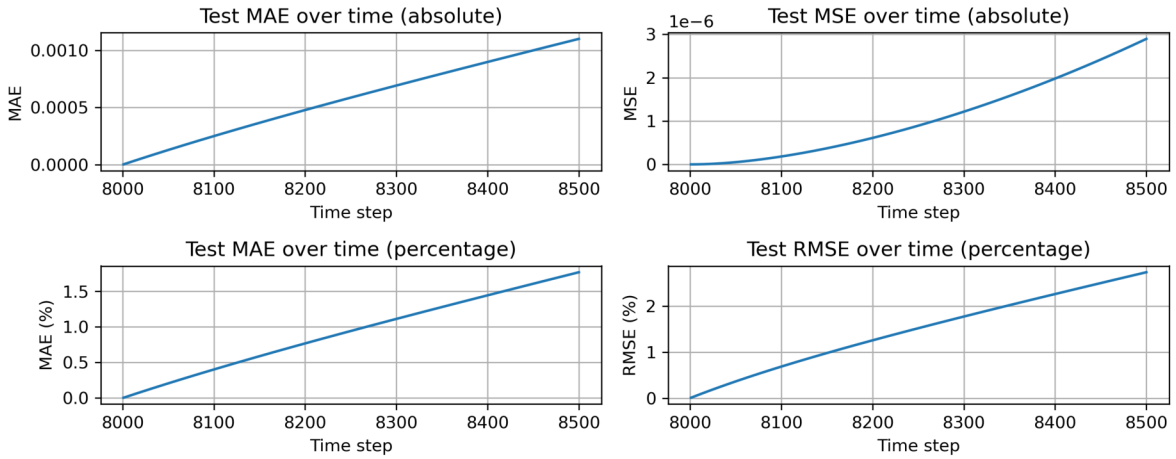


Figure 5.41: Short-horizon test error when w_4 is forced to enter the potential linearly.

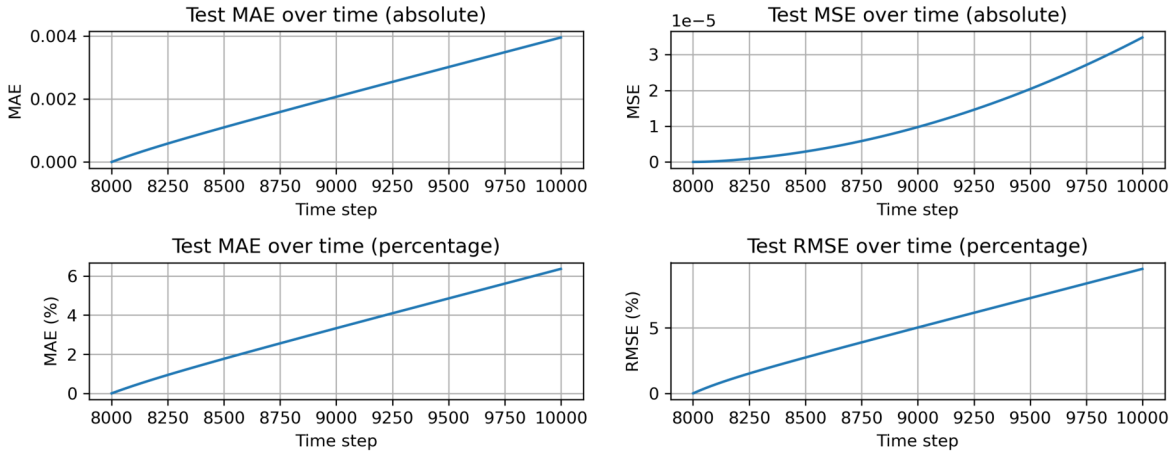


Figure 5.42: Long-horizon test error when w_4 is forced to enter the potential linearly.

This model tests whether transportation accessibility can be represented as an independent linear component of the population potential.

5.5.5 Combined Linear Test for w_1 and w_2

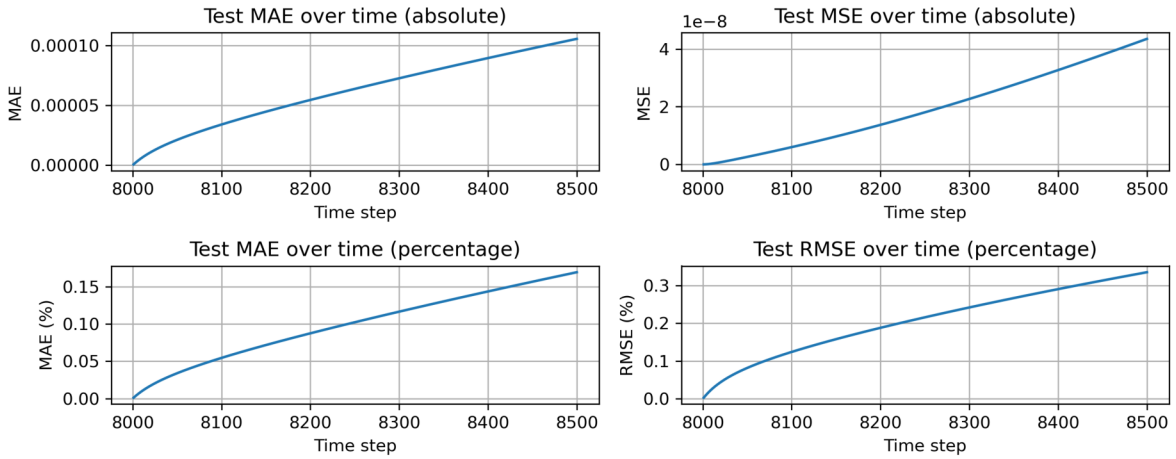


Figure 5.43: Short-horizon test error when w_1 and w_2 are forced to enter the potential linearly, while w_3 and w_4 are learned flexibly.

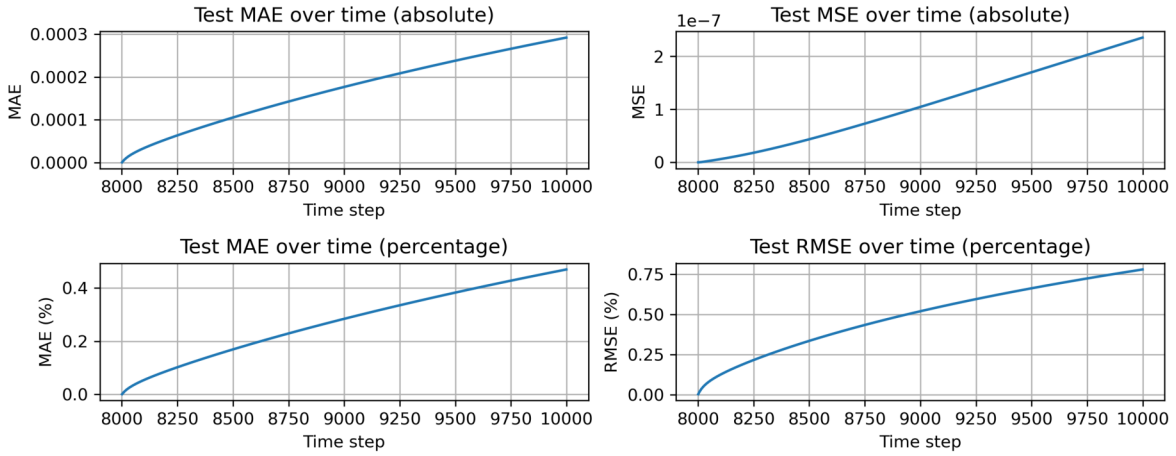


Figure 5.44: Long-horizon test error when w_1 and w_2 are forced to enter the potential linearly, while w_3 and w_4 are learned flexibly.

This model tests the combined partially linear structure

$$f(w_1, w_2, w_3, w_4) = \beta_1 w_1 + \beta_2 w_2 + f_\theta(w_3, w_4).$$

It is used to check whether the conclusions from the individual linearity tests are consistent when w_1 and w_2 are both treated as linear components and w_3 and w_4 are left to a flexible neural-network component.

Table 5.5: Step 4 results: factor-structure and linearity tests.

Model	**Potential structure**	**Short MAE (%)**	**Long MAE (%)**
Baseline	$f_\theta(w_1, w_2, w_3, w_4)$	0.2567	0.75
w_1 linear	$\beta_1 w_1 + f_\theta(w_2, w_3, w_4)$	0.1342	0.37
w_2 linear	$\beta_2 w_2 + f_\theta(w_1, w_3, w_4)$	0.1280	0.34
w_3 linear	$\beta_3 w_3 + f_\theta(w_1, w_2, w_4)$	1.1286	4.24
w_4 linear	$\beta_4 w_4 + f_\theta(w_1, w_2, w_3)$	0.9250	3.29
w_1, w_2 linear	$\beta_1 w_1 + \beta_2 w_2 + f_\theta(w_3, w_4)$	0.10	0.27

Table 5.5 summarizes the factor-structure test results. The baseline model uses a fully learned multi-factor potential and achieves a short-horizon MAE of 0.2567% and a long-horizon MAE of 0.75%.

The results show that imposing a linear structure on w_1 or w_2 improves prediction accuracy. When w_1 is forced to enter linearly, the short-horizon MAE decreases to 0.1342%, and the long-horizon MAE decreases to 0.37%. When w_2 is forced to enter linearly, the short-horizon MAE decreases to 0.1280%, and the long-horizon MAE decreases to 0.34%. This suggests that job opportunity density and commercial density are well represented as linear components of the population potential in this simulation.

In contrast, forcing w_3 or w_4 to enter the potential linearly leads to much larger errors. The w_3-linear model has a short-horizon MAE of 1.1286% and a long-horizon MAE of 4.24%. The w_4-linear model has a short-horizon MAE of 0.9250% and a long-horizon MAE of 3.29%. These results indicate that w_3 and w_4 are not well represented as independent linear components.

Method 4.5 further supports this interpretation. When both w_1 and w_2 are forced to enter linearly and the remaining part $f_\theta(w_3, w_4)$ is learned flexibly, the model achieves the best performance among all Step 4 experiments. Its short-horizon MAE is 0.10%, and its long-horizon MAE is 0.27%. This result confirms that treating w_1 and w_2 as linear components while allowing w_3 and w_4 to be represented flexibly is more consistent with the simulated dynamic structure than either a fully flexible potential or a model that forces w_3 or w_4 to enter linearly.

This result is consistent with the true potential function used in the simulation:

$$f_{\text{true}}(w_1, w_2, w_3, w_4) = a_1 w_1 + a_2 w_2 + a_3 w_3 w_4.$$

In the true function, w_1 and w_2 enter additively and linearly, while w_3 and w_4 enter through the product term $w_3 w_4$. Therefore, forcing w_3 or w_4 to enter independently and linearly

breaks the true nonlinear structure and leads to poorer performance. By contrast, Method 4.5 preserves the linear roles of w_1 and w_2 while leaving the joint effect of w_3 and w_4 to the flexible neural-network component.

Overall, Step 4 shows that the proposed pipeline can go beyond factor selection. Step 2 identifies which factors should be retained, and Step 3 selects a suitable dynamic model structure. Step 4 further diagnoses how the retained factors enter the selected structure. In this simulation, the method correctly identifies that w_1 and w_2 are compatible with a linear representation, while w_3 and w_4 require a nonlinear or interaction-based representation. The additional combined test in Method 4.5 further confirms this conclusion and shows that the partially linear structure $\beta_1 w_1 + \beta_2 w_2 + f_\theta(w_3, w_4)$ provides the best representation among the tested factor structures.

5.6 Step 5: Rolling-Window Prediction Evaluation

After Step 3 selects the multi-factor advection–diffusion model as the best theory-guided structure, and Step 4 further examines the internal factor structure of the learned potential, Step 5 evaluates whether the selected model remains accurate and stable under rolling-window prediction. This step is designed to test whether the model can maintain its predictive performance when the training and testing windows are moved forward over time.

The baseline model in this step is the multi-factor advection–diffusion model from Method 3.1:

$$\frac{\partial \lambda}{\partial t} = \nabla \cdot \left(D \nabla \lambda - \lambda \nabla f_\theta(w_1, w_2, w_3, w_4) \right).$$

This baseline is referred to as Method 5.0 in the rolling-window evaluation. It uses the same model structure as Method 3.1 and provides the reference performance for comparison.

The rolling-window evaluation uses four repeated experiments. In each experiment, the model is trained and validated on a shifted time window, and then evaluated on the following 500 time steps. The window design is summarized in Table 5.6. The notation 6000 : 2000 : 500 means that 6000 time steps are used for training, 2000 time steps are used for validation, and the following 500 time steps are used for testing.

Table 5.6: Step 5 rolling-window design.

Method	Training window	Validation window	Test window
Method 5.1	0–6000	6000–8000	8000–8500
Method 5.2	0–6500	6500–8500	8500–9000
Method 5.3	0–7000	7000–9000	9000–9500
Method 5.4	0–7500	7500–9500	9500–10000

This design evaluates the model in several consecutive prediction windows rather than only in one fixed test period. Therefore, it provides a more direct test of prediction stability over time. If the model performs well across all rolling windows, then the selected theory-guided structure is not only accurate in one test interval, but also robust across different temporal segments.

5.6.1 Method 5.0: Baseline Multi-Factor Advection–Diffusion Model

Method 5.0 is the baseline multi-factor advection–diffusion model, corresponding to Method 3.1 in Step 3. It achieves a short-horizon MAE of 0.2567% over the first 500 test steps and a long-horizon MAE of 0.75% over the full 2000-step test period.

Since each rolling-window experiment in Step 5 predicts the following 500 time steps, the short-horizon result is used as the main baseline reference. The long-horizon result is reported as an additional stability reference for longer recursive prediction.

5.6.2 Method 5.1: Rolling Window 1

Method 5.1 uses the first rolling-window split, with 6000 time steps for training, 2000 time steps for validation, and the following 500 time steps for testing. This setting is the same as the short-horizon evaluation of Method 5.0. Therefore, Method 5.1 is not plotted separately. Its average MAE is 0.26%, which is essentially the rounded version of the baseline short-horizon MAE of 0.2567%.

5.6.3 Method 5.2: Rolling Window 2

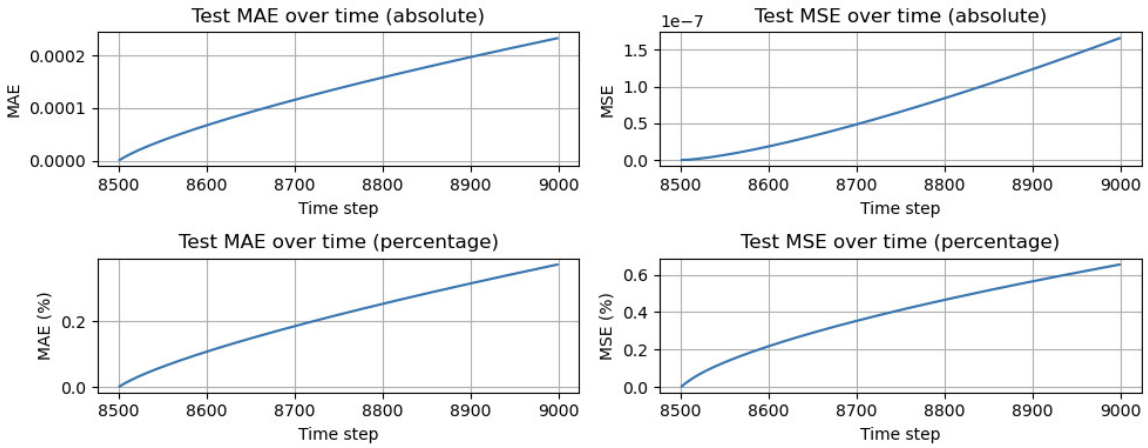


Figure 5.45: Rolling-window prediction error of Method 5.2.

Method 5.2 extends the training window to 6500 time steps, uses the following 2000 time steps for validation, and evaluates prediction on the next 500 time steps. The average MAE decreases to 0.21%.

5.6.4 Method 5.3: Rolling Window 3

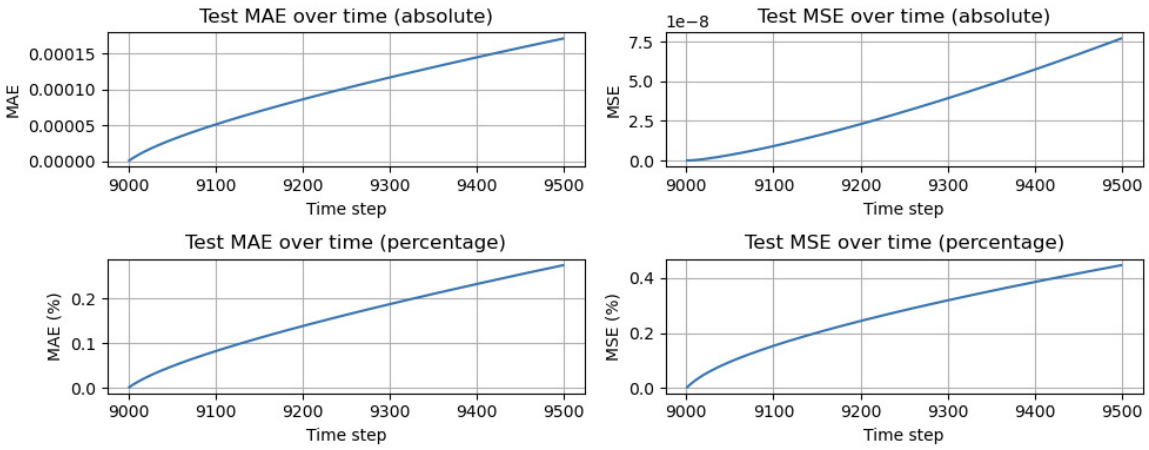


Figure 5.46: Rolling-window prediction error of Method 5.3.

Method 5.3 further extends the training window to 7000 time steps. It achieves the lowest prediction error among the rolling-window experiments, with an average MAE of 0.16%.

5.6.5 Method 5.4: Rolling Window 4

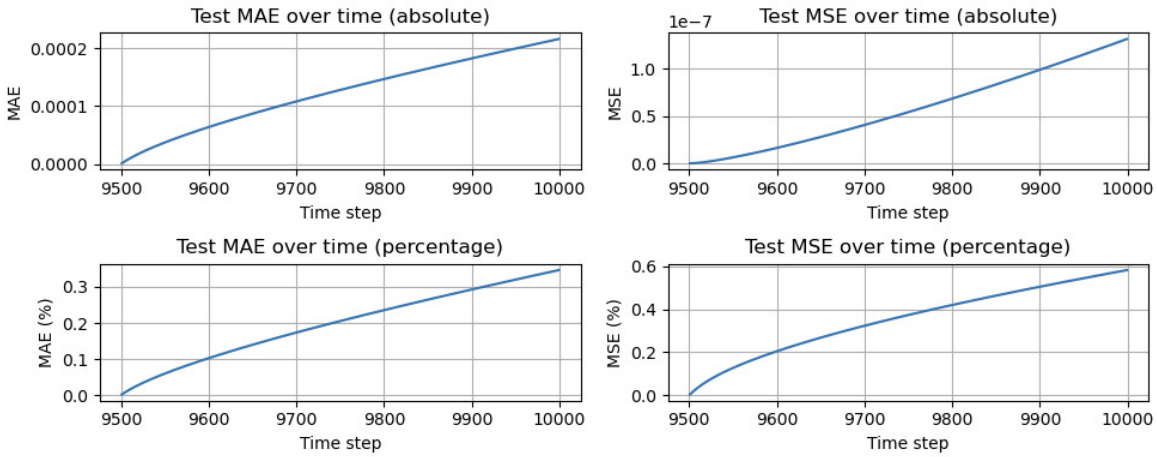


Figure 5.47: Rolling-window prediction error of Method 5.4.

Method 5.4 uses the longest training window, with 7500 time steps for training, 2000 time steps for validation, and the final 500 time steps for testing. The average MAE is 0.20%.

Table 5.7: Step 5 results: baseline and rolling-window prediction performance using MAE.

Method	**Evaluation setting**	**MAE (%)**
Method 5.0	Baseline long-horizon prediction, 6000 : 2000 : 2000	0.75
Method 5.1	Baseline short-horizon prediction, 6000 : 2000 : 500	0.2567
Method 5.2	Rolling-window prediction, 6500 : 2000 : 500	0.21
Method 5.3	Rolling-window prediction, 7000 : 2000 : 500	0.16
Method 5.4	Rolling-window prediction, 7500 : 2000 : 500	0.20

Table 5.7 summarizes the baseline and rolling-window prediction results. Method 5.0 reports the long-horizon baseline performance of the selected multi-factor advection–diffusion model over the full 2000-step test period. Method 5.1 reports the short-horizon baseline performance over the first 500 test steps. Since Method 5.1 uses the 6000 : 2000 : 500 setting, it serves as the direct reference for the later rolling-window experiments.

The rolling-window results show that updating the training window improves prediction accuracy. Compared with the short-horizon baseline in Method 5.1, Method 5.2 reduces the MAE from 0.2567% to 0.21%, Method 5.3 further reduces it to 0.16%, and Method 5.4 remains low at 0.20%. All rolling-window results are also much lower than the long-horizon baseline MAE of 0.75%.

These results indicate that the selected multi-factor advection–diffusion model remains stable when the prediction window moves forward over time. The MAE values of Methods 5.2–5.4 stay within a narrow range from 0.16% to 0.21%, and no window shows clear error accumulation. Therefore, rolling-window updating improves short-term prediction accuracy while maintaining prediction stability.

Overall, Step 5 supports two conclusions. First, rolling-window updating can improve prediction accuracy by allowing the model to use more recent training information. Second, the selected multi-factor advection–diffusion model remains stable across different temporal windows. This suggests that the proposed pipeline produces a model that is not only accurate in a fixed test period, but also robust under rolling-window prediction.

5.7 Summary of Main Findings

This chapter develops a step-by-step methodology for learning, selecting, and interpreting dynamic models of spatial population movement. The main purpose is not only to obtain accurate prediction, but also to build a systematic pipeline that connects factor selection, theory-guided model comparison, factor-structure diagnosis, and rolling-window evaluation.

Step 1 tests whether neural networks can learn the simulated population dynamics from spatial factor fields and population density. The results show that both direct prediction and derivative-learning models can achieve accurate short-horizon prediction. The derivative-learning model is especially useful as a baseline because it learns the local rate of population change, which is closer to the structure of continuous-time dynamic models. Therefore, Step 1 establishes that the simulated spatial population process is learnable by neural networks and provides a black-box benchmark for later comparison.

Step 2 uses factor-inclusion experiments to screen candidate explanatory factors. The results show that job opportunity density is the most important individual factor, while commercial density and transportation accessibility also improve prediction. The additional weak factor does not provide strong explanatory value. This step demonstrates the first methodological function of the pipeline: before selecting a theoretical model, the relevant factor set should be identified. At the same time, Step 2 also shows the limitation of simple inclusion tests, because nonlinear or joint factor effects may not be fully captured by adding

factors one at a time.

Step 3 compares different model structures using the retained factors. The results show that the multi-factor advection–diffusion model performs better than the black-box derivative model and the wage-difference model. This indicates that the proposed pipeline can be used to select the theory-guided model structure that best matches the data-generating process. The comparison between the multi-factor and single-factor advection–diffusion models further shows that, even within the same theoretical framework, the multi-factor specification performs substantially better than the single-factor version. Therefore, Step 3 supports both theory selection and multi-factor model extension.

Step 4 examines how the retained factors enter the selected model structure. Instead of asking whether a factor should be included, Step 4 asks whether each factor can be represented as an independent linear component of the learned potential. The results show that w_1 and w_2 are compatible with a linear representation, while w_3 and w_4 require a nonlinear or interaction-based representation. This is consistent with the true potential function used in the simulation, where w_1 and w_2 enter linearly and w_3 and w_4 enter through a product term. Thus, Step 4 demonstrates that the pipeline can diagnose factor structure, not only factor importance.

Step 5 evaluates the selected multi-factor advection–diffusion model under rolling-window prediction. The results show that rolling-window updating improves short-term prediction accuracy compared with the fixed short-horizon baseline, and the errors remain stable across different temporal windows. This indicates that the selected model is not only accurate in a fixed test interval, but also robust when the prediction window moves forward over time. Therefore, Step 5 confirms the temporal stability of the model selected by the previous steps.

Overall, the findings support the proposed methodology as a complete framework for interpretable dynamic population modeling. The pipeline first tests learnability, then screens relevant factors, compares theory-guided model structures, diagnoses factor-entry structure, and finally evaluates rolling-window stability. In this simulation, the best-performing and most interpretable model is the partially linear multi-factor advection–diffusion model, which combines interpretable linear components with a flexible learned potential structure. The results suggest that accurate spatial population modeling requires both an appropriate dynamic theory and a sufficiently rich set of explanatory spatial factors.

Chapter 6

Why Simulated Data? Real-Data Attempts and Challenges

This thesis uses simulated data not because real-world spatial data are irrelevant, but because the continuous time-series data required by this methodological framework are difficult to obtain in a unified and consistent form. The proposed pipeline requires repeated observations of fine-scale population density and related spatial factors over time, including job opportunity density, commercial activity, amenity distribution, transportation accessibility, and population movement.

Existing open data resources and empirical studies can provide important components of this information. However, they usually provide static snapshots, low-frequency observations, or isolated variables rather than a continuous spatial-temporal panel. Moreover, different open resources often use different spatial resolutions, temporal frequencies, data definitions, and update cycles. As a result, they cannot be directly matched with each other to form a consistent dynamic system for model training, rolling-window prediction, and factor-structure testing.

Therefore, the main challenge is not the complete absence of real-world data, but the lack of an integrated spatial-temporal data system that combines population dynamics, employment distribution, commercial and amenity indicators, transportation accessibility, and movement observations in a consistent way. For this reason, simulated data are used in this thesis as a controlled environment to demonstrate and validate the proposed methodology.

6.1 Global Human Settlement Layer

The Global Human Settlement Layer (GHSL) is one of the most useful open resources for studying large-scale human settlement patterns [4]. It provides global spatial information

related to human presence, including population grids, built-up surfaces, built-up volume, and settlement classification. Because GHSL is available at the global scale, it is especially valuable for cross-city or cross-country studies of urban settlement patterns.

For the purpose of this thesis, GHSL is useful because it can provide a gridded representation of population density and urban settlement structure. In principle, such data could be used to construct the population field $\lambda(t, x)$ in a real-world version of the model. It may also help define the spatial domain, identify urbanized regions, and compare settlement structures across different geographic contexts.

However, GHSL alone is not sufficient for the present research design. The main limitation is its temporal resolution. GHSL provides data on the built-up environment, including built-up surface, built-up volume, and residential versus non-residential function, as well as resident population and settlement classification based on the UN-recommended "degree of urbanisation" methodology. These products are provided at 5-year intervals starting from 1975. This is very useful for long-term urbanization analysis, because it allows researchers to compare population and settlement patterns across different historical periods. However, such a 5-year interval is too sparse for the continuous dynamic modeling framework developed in this thesis.

The model in this thesis requires repeated observations of population density and spatial factors over relatively fine time intervals, so that population movement, recursive prediction, and rolling-window evaluation can be studied as a dynamic process. A 5-year interval is not sufficient to observe short-term changes in population distribution or to train a model on continuous population evolution. In addition, GHSL mainly provides population and settlement-related information, but not the full set of explanatory spatial factors required by the model, such as job opportunities, commercial activity, amenity density, and transportation accessibility. Therefore, GHSL is valuable as a global population and settlement data source, but it cannot by itself support the complete multi-factor dynamic modeling framework proposed in this thesis.

6.2 Mobile-Phone-Based Home and Work Location Studies

Compared with GHSL, mobile-phone-based studies provide a more direct way to estimate population and job distributions at the urban scale. By observing repeated location traces, these studies can infer where users are likely to live and where they are likely to work. The inferred home locations can be aggregated to estimate residential population distribution,

while the inferred work locations can be aggregated to estimate job or working-population distribution. Therefore, this type of data is highly relevant to the framework proposed in this thesis, because it can potentially provide both the population field $\lambda(t, x)$ and the job-opportunity factor w_1.

Existing studies have already shown that this approach is technically feasible. For example, some studies use mobile phone signaling data to identify home and work locations, and others estimate working and living population distributions in large cities such as Beijing and New York [11, 32, 5, 35, 38]. Although many published results present population or job distributions at one representative time point or over a limited observation period, the underlying data collection technology can in principle support continuous time-series observations. If similar data were available repeatedly over time, it could be used to construct a dynamic spatial-temporal system for model training, prediction, and factor-structure testing.

For the purpose of this thesis, mobile-phone-based data are especially attractive because they can connect population distribution, job distribution, and movement behavior within the same empirical framework. Home-location inference can be used to estimate residential population density, work-location inference can be used to estimate job opportunity density, and movement trajectories can provide information about commuting flows and spatial interaction [37, 33, 36, 18]. In principle, this type of data is much closer to the continuous dynamic system required by the proposed methodology than low-frequency settlement datasets.

However, the main limitation is data accessibility and privacy protection. Mobile phone location data are highly sensitive because they are derived from individual movement traces. Even when the data are anonymized or aggregated, they may still raise concerns related to personal information protection, re-identification risk, commercial ownership, and data governance. For this reason, raw mobile-phone-based location data are rarely released as open public datasets. For the Canadian context, the main challenge is data availability. Based on the resources reviewed in this thesis, I have not identified an open Canadian dataset that provides continuous mobile-phone-based home and work location information at a fine spatial resolution and in a form directly usable for the proposed model.

Therefore, although mobile-phone-based methods are technically capable of supporting this research direction, the lack of accessible and privacy-compliant public data prevents their direct use in this thesis. This further supports the reason for using simulated data: the required empirical data structure is feasible in principle, but it is not currently available as an open, continuous, and integrated spatial-temporal dataset.

6.3 OpenStreetMap

OpenStreetMap (OSM) is a global open geographic database that provides detailed spatial information on roads, buildings, land use, public facilities, commercial points of interest, amenities, and transportation infrastructure. For this thesis, OSM is important because many of its features can be used to construct spatial factor fields in a real-world version of the proposed model [22, 21].

In particular, OSM can provide useful information for constructing the factors w_2, w_3, and w_4. Commercial points of interest, such as shops, restaurants, offices, and shopping centers, can be used to approximate commercial density w_2. Amenity-related features, such as schools, hospitals, parks, public services, and community facilities, can be used to construct amenity density w_3. Road networks, public transit stops, railway stations, and other transportation-related features can be used to construct transportation accessibility w_4.

Table 6.1: Examples of OSM data types and their relevance to spatial factor construction.

OSM data type	**Examples**	**Possible factor construction**
Commercial points of interest	Shops, restaurants, offices, malls	Commercial density w_2
Amenities	Schools, hospitals, parks, public services	Amenity density w_3
Road networks	Roads, intersections, highways, local streets	Transportation accessibility w_4
Public transit features	Subway stations, bus stops, train stations	Transportation accessibility w_4
Buildings and land use	Building footprints, residential areas, commercial areas, industrial areas	Supporting the construction and validation of spatial factor fields

As shown in Table 6.1, OSM is especially useful for constructing explanatory spatial factors related to commercial activity, amenities, and transportation accessibility. These data can be converted into gridded factor fields by counting or weighting OSM features within each grid cell, applying distance-based measures, or constructing network-based accessibility indicators. For example, commercial density may be estimated by the number or density of commercial points of interest in each grid cell, while transportation accessibility may be measured using road-network connectivity or distance to public transit facilities.

Therefore, OSM provides an important data foundation for future empirical applications of the proposed framework. Although the present thesis uses simulated factor fields, OSM suggests a practical way to construct several of these factors from real geographic data. In particular, it can support the empirical construction of w_2, w_3, and w_4 once the spatial resolution, feature classification rules, and aggregation methods are determined.

However, OSM data quality and completeness may vary across regions. Because OSM is a crowd-sourced geographic database, the density and accuracy of mapped features depend on local mapping activity, contributor behavior, and regional data imports. Some cities may have detailed road networks, buildings, and points of interest, while other areas may have incomplete or outdated records. Therefore, OSM-based factor construction would require careful preprocessing, validation, and consistency checks before being used in a dynamic population model.

Overall, OSM is valuable in this thesis because it shows that several explanatory spatial factors required by the proposed framework can be constructed from open geographic data. It is particularly relevant for commercial density, amenity density, and transportation accessibility, and therefore provides an important basis for extending the simulation-based methodology toward future real-data applications.

6.4 Why These Resources Are Not Yet Sufficient and Why Simulated Data Are Used

The resources and studies discussed above show that the empirical implementation of this thesis is technically feasible. GHSL can provide global gridded population and settlement information. OpenStreetMap can provide roads, public transit features, commercial points of interest, amenities, buildings, and land-use information. Mobile-phone-based home and work location studies show that population distribution, job distribution, and movement behavior can be estimated from repeated location traces. Together, these resources indicate that the data components required by the proposed methodology already exist in different forms.

However, these components are not yet organized into a complete and continuous dynamic system. The proposed framework requires repeated observations of population density, job opportunity density, commercial activity, amenity distribution, transportation accessibility, and population movement on a consistent spatial grid and over comparable time intervals. Existing resources usually provide only part of this structure. GHSL provides valuable population and settlement information, but its 5-year interval is too sparse for continuous

dynamic modeling. OSM provides frequently updated spatial infrastructure and points of interest, but it does not directly provide population movement, job distribution, or home–work dynamics. Mobile-phone-based studies can technically support continuous observation of home and work locations, but such data are rarely available as open public datasets because of privacy and data-access restrictions.

Therefore, the main challenge is not that the required data are impossible to obtain. Rather, the challenge is that the available resources are fragmented. They differ in spatial resolution, temporal frequency, variable definition, update mechanism, accessibility, and data-quality reporting standards [29, 28, 30, 27]. As a result, they cannot be directly matched with each other to form a ready-to-use spatial-temporal panel for model training, factor selection, theory-guided model comparison, factor-structure testing, and rolling-window prediction.

For this reason, this thesis uses simulated data as a controlled environment to demonstrate and validate the proposed methodology. Simulation allows all required variables to be constructed on the same spatial grid and observed over continuous time steps. It also makes the true data-generating mechanism known, which is essential for evaluating whether the proposed pipeline can identify relevant factors, select the correct model structure, and diagnose linear or nonlinear factor entry.

Using simulated data therefore serves a methodological purpose. It allows the thesis to test the complete pipeline under controlled conditions before applying it to real-world data. The goal is not to replace empirical data, but to clarify what kind of integrated data system would be needed for future applications. Once GHSL, OSM, mobile-phone-based mobility data, census data, employment data, and transportation data can be harmonized into a consistent spatial-temporal dataset, the same methodology could be applied to real urban population dynamics.

In this sense, the thesis also points to a future direction for data integration. Existing open resources and empirical studies already provide many of the necessary components, but they have not yet been combined into a unified dynamic system. By demonstrating the type of model that such data could support, this thesis aims to contribute to the development of future integrated datasets for multi-factor spatial population dynamics.

Chapter 7

Future Extensions

7.1 Temporal Distance and Transportation Accessibility

The current simulation represents transportation accessibility as a spatial density field. A future extension is to construct transportation accessibility using travel time or network distance. In real cities, physical distance and travel time may differ substantially because of road structure, public transit, congestion, and barriers [34, 14]. Replacing simple spatial proximity with temporal distance would make the model more realistic.

Another advantage of using temporal distance is that it may reduce the need to treat transportation accessibility as a separate explanatory factor. If travel time or network distance is used directly to measure the effective distance between locations and opportunities, then part of the transportation effect is already embedded in the spatial interaction structure of the model. In this case, transportation accessibility does not need to enter the potential function as an independent factor. This may make the factor structure closer to a linear additive form, because job opportunities, commercial activity, and amenities can be evaluated through their travel-time accessibility rather than through an additional transportation-density field. As a result, the model may become more interpretable and potentially more accurate.

7.2 Capacity Constraints and Upper Bounds

The present model does not impose explicit capacity constraints on local population density. In real urban systems, population growth is limited by housing supply, land-use regulation, infrastructure capacity, and zoning constraints. A future model can introduce a local carrying

capacity term:

$$\lambda(t, x) \leq K(x),$$

where $K(x)$ represents the maximum feasible population density at location x.

7.3 Boundary Conditions and Open Systems

The current simulation should not be interpreted as a strictly closed urban system. Although the model is computed on a finite spatial grid, population density is not forced to remain inside an artificial administrative boundary. In the numerical setting, particles or density mass can move toward and across the boundary of the computational domain. Therefore, the present model is closer to an open spatial system than to a closed one.

A closed-system interpretation becomes reasonable only when the computational grid is sufficiently large. If the spatial domain is enlarged until the population density near all outer boundaries converges to nearly zero, then boundary effects become negligible. In that case, the finite grid can be interpreted as an approximation of a larger closed system, because almost no population density reaches the artificial boundary.

However, real urban systems may require more careful treatment of boundary behavior. Population movement is not only driven by attractiveness gradients, but also by the availability of public resources and urban services. When population density in a peripheral region falls below a certain threshold, public facilities, commercial services, and transportation operations may no longer be sustainable. This can further reduce local attractiveness and cause additional out-migration.

Therefore, a future extension should introduce boundary constraints or threshold effects near low-density regions. These constraints may represent the fact that areas with population density below a minimum level cannot sustain the same level of public-resource support [9, 31]. Instead of treating the boundary as purely mathematical, the model should allow boundary behavior to depend on urban viability, minimum population density, and the sustainability of public services.

In this way, future work can distinguish between three cases: an open finite-domain model, a sufficiently large domain that approximates a closed system, and a threshold-constrained urban system where boundary behavior depends on minimum population density and public-resource sustainability. This would make the treatment of urban boundaries more realistic and better connected to the economic interpretation of population movement.

7.4 Interaction Between Population and Spatial Factors

In the current setting, the spatial factor fields are fixed during the simulation. In reality, population and spatial factors may evolve together. For example, population growth may attract more commercial activity, improve transportation investment, or change amenity provision. A future extension can model population and factors as a coupled system:

$$\frac{\partial \lambda}{\partial t} = \mathcal{M}_\lambda(\lambda, W),$$

$$\frac{\partial W}{\partial t} = \mathcal{M}_W(\lambda, W).$$

This would allow the model to describe feedback between population movement and urban development.

However, directly modeling the interaction between population density and spatial factors may introduce additional mathematical difficulty. In a fully coupled advection–diffusion system, if population density and spatial factors continuously respond to each other, the resulting feedback may make the stability of the system more difficult to control. Without additional constraints or damping mechanisms, the model may generate unstable dynamics or fail to converge to a stable equilibrium. For example, population growth may increase local commercial activity, which further increases attractiveness and attracts even more population. Without additional constraints, such feedback may lead to unrealistic concentration or instability.

One possible way to handle this difficulty is to use an alternating update scheme. In a short time interval, the spatial factors can be assumed to remain fixed, and the population density evolves according to the migration model. In the next short time interval, the population density can be treated as temporarily fixed, while the spatial factors are updated according to the new population distribution. Repeating these two steps would allow the model to approximate the feedback process between population movement and urban development while avoiding the instability of a fully simultaneous coupled system.

7.5 Advanced Neural Architectures and Algorithmic Improvements

The current thesis uses a relatively simple CNN architecture to keep the experimental comparison clear and consistent. Future work may test more advanced models, such as ConvL-

STM, graph neural networks, neural operators, or transformer-based architectures. These models may be useful when real data contain long temporal dependence, irregular transportation networks, or more complex spatial interactions.

Future work can also improve the computational implementation. The current code is designed mainly for clarity and reproducibility. Further improvements may include more efficient data loading, better memory management, GPU acceleration, optimized rollout procedures, hyperparameter tuning, learning-rate scheduling, and regularization.

Therefore, future extensions should include not only more advanced neural-network architectures, but also improvements in code structure and algorithmic efficiency. These improvements would make the framework more scalable and more suitable for future empirical applications.

7.6 Uncertainty and Stochastic Extensions

The current thesis focuses on deterministic population dynamics. A future extension is to introduce stochastic effects or uncertainty quantification. A stochastic version of the model may be written as

$$d\lambda = \mathcal{M}(\lambda, W)\, dt + \sigma(\lambda, W)\, dB_t,$$

where B_t represents a Brownian motion or spatial noise process.

The stochastic term can serve two purposes. First, it can absorb the influence of factors that are not explicitly included in the deterministic part of the model, such as housing prices, policy changes, local shocks, demographic variation, or other unobserved spatial conditions. Second, it can represent unbiased random movement that is not driven by a systematic attractiveness gradient. In this sense, the stochastic term captures residual uncertainty beyond the modeled spatial factors.

However, stochastic modeling introduces additional numerical and algorithmic difficulties. The neural-network models in this thesis are mainly trained for one-step prediction or one-step derivative learning. Once a stochastic noise term is added, recursive prediction becomes harder to control, because random perturbations may accumulate over time and may cause the rollout trajectory to become unstable. In numerical experiments, this can easily lead to exploding errors or unrealistic population values if the stochastic term is not carefully constrained.

Therefore, extending the framework to stochastic population dynamics would require further optimization of the algorithm. Possible improvements include more stable time-stepping schemes, stronger constraints on the noise term, better normalization, variance

control, regularization, and more robust rollout procedures. For this reason, stochastic extensions are treated as an important but separate direction for future work.

7.7 Empirical Application

The most important future extension is empirical application to real urban data. The value of the proposed framework is not limited to reproducing the simulated experiment. More importantly, it provides a methodology for testing and comparing existing theories of population migration under a common data-driven framework.

Once suitable empirical data are available, the same pipeline can be used to examine whether population movement is better explained by a certain theory-guided dynamic structure. In this way, the framework can help evaluate whether existing population migration theories are consistent with observed urban dynamics, rather than assuming one theory to be correct in advance.

The empirical version should not assume that the same true function used in the simulation is correct for real cities. Instead, the simulation should be treated as a methodological benchmark. Real-data analysis should be used to test which factors are relevant, which model structure best explains the observed dynamics, how factors enter the migration mechanism, and whether rolling-window updating improves prediction accuracy and stability.

This future direction also has implications for data construction. The proposed framework clarifies what kind of empirical data are needed for dynamic population modeling: repeated observations of population distribution, job distribution, commercial activity, amenities, transportation accessibility, and movement behavior on a consistent spatial grid. Therefore, the thesis not only proposes a modeling pipeline, but also provides guidance for future database development. A more complete and integrated spatial-temporal database would make it possible to test population migration theories more systematically and to apply the proposed framework to real urban systems.

Appendix A

Algorithm Details and Limitations

A.1 Algorithm Details

A.1.1 Computing Environment and Running Time

All experiments were conducted locally on an ASUS VivoBook laptop running Microsoft Windows 11 Home. The system was equipped with an AMD Ryzen 7 5800HS processor (8 physical cores and 16 logical processors) and 16 GB of installed memory. The implementation used Python 3.9.25, PyTorch 2.5.1, and SciPy 1.13.1. All experiments were executed on the CPU without GPU acceleration.

The spatial fields were represented on a 200×200 grid. Under this environment, training one method generally required approximately four hours. Running time varied across methods because different model structures used different input channels and required repeated calculation of spatial derivatives. Testing was performed through recursive rollout, and the exact short- and long-horizon testing times will be added after representative runs are recorded.

These running times are approximate and may vary with processor utilization, software versions, implementation details, and background system activity.

A.1.2 CNN Architecture and Training

All neural-network-based experiments use the same three-layer CNN so that performance differences can be attributed mainly to factor selection and model structure rather than to different network capacities. The architecture is

$$\mathrm{Conv}_{3\times3} \to \mathrm{ReLU} \to \mathrm{Conv}_{3\times3} \to \mathrm{ReLU} \to \mathrm{Conv}_{3\times3},$$

with padding 1 and 32 hidden channels. The number of input channels varies across experiments, but the hidden-layer size, kernel size, activation functions, and network depth remain fixed.

The models are trained using mean squared error between the predicted time derivative and the derivative estimated from consecutive population snapshots:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\frac{\partial\lambda_i}{\partial t}} - \frac{\lambda_i(t+\Delta t)-\lambda_i(t)}{\Delta t}\right\|^2.$$

Black-box models predict the derivative directly, whereas theory-guided models first learn the relevant spatial function and then obtain the derivative through the corresponding dynamic operator.

A.1.3 Evaluation, Numerical Constraints, and Reproducibility

Model performance is evaluated through recursive rollout using the mean absolute error percentage:

$$\mathrm{MAE}(\%) = \frac{\frac{1}{|\Omega|}\sum_{x\in\Omega}|\hat{\lambda}(t,x)-\lambda(t,x)|}{\frac{1}{|\Omega|}\sum_{x\in\Omega}|\lambda(t,x)|}\times 100.$$

MSE and RMSE percentages are also computed as supplementary diagnostics, while MAE percentage is used in the main comparison tables.

All generated population and factor fields are normalized and constrained to remain non-negative. To improve reproducibility, the spatial centres and random seeds are fixed, and deterministic PyTorch settings are used where possible. Although exact numerical agreement is not guaranteed across all hardware and software environments, these controls substantially reduce unnecessary variation between repeated runs.

A.2 Limitations of the Model

The model used in this thesis has three closely related limitations. Their common source is the recursive derivative-based prediction structure. At each time step, the model recalculates spatial derivatives from the current predicted population field. Spatial differentiation is sensitive to local perturbations, and an error introduced at one step may therefore be amplified before being propagated to subsequent steps.

First, the current model cannot effectively handle stochastic terms. When random disturbances are introduced into the population field, spatial derivatives are repeatedly calculated

from a noisy state. Since differentiation can amplify high-frequency variation, even relatively small stochastic disturbances may produce large local errors. These errors are then propagated through the recursive rollout and may quickly lead to unstable or unrealistic predictions.

Second, the model has relatively weaker performance in learning strongly nonlinear components. Nonlinear relationships may create sharper local variation in the learned spatial field. Small approximation errors in these nonlinear components can therefore be magnified when gradient and divergence operators are applied. The resulting derivative errors are subsequently carried into later predictions and accumulate over time.

Third, the model is trained using one-step prediction but is applied recursively for long-horizon forecasting. At each step, the predicted population field becomes the input for the next prediction, and spatial derivatives are recalculated from this predicted state. Consequently, a small error affects both the next input and the next derivative calculation. This creates a repeated amplification-and-accumulation process and explains why long-horizon errors can grow substantially faster than one-step errors. The rolling-window strategy used in this thesis reduces the length of each recursive rollout and incorporates more recent observations, but it does not completely eliminate this limitation.

Possible future solutions include multi-step training, in which the model is optimized over several recursively predicted steps rather than only the next step, and rollout-based loss functions that directly penalize error growth over a prediction trajectory. Periodic state correction using newly observed population data could prevent predicted states from drifting too far from the actual system. More stable numerical differentiation and integration schemes, spatial smoothing or regularization, and physical constraints on population values could also help control derivative-related error amplification. For datasets with meaningful delayed responses, temporal neural networks such as ConvLSTM could additionally be considered, although temporal architectures are distinct from multi-step training and are not required to implement it.

These limitations indicate that the current model is more reliable for deterministic, relatively smooth, and short-horizon dynamics than for stochastic, strongly nonlinear, or very long-horizon prediction.

A.3 Limitations of the Dataset

The dataset used in this thesis has several limitations. The most important limitation is that it is generated through a controlled simulation rather than collected from an observed urban system. Although the spatial factors and population patterns are designed to be

theoretically and qualitatively reasonable, the dataset does not represent the full complexity of a real city.

First, the four spatial factor fields are artificially constructed. Job opportunities, commercial activity, and amenities are represented by several spatial centres and their surrounding areas of influence, while transportation accessibility is represented by corridor-like connections between major centres. These constructions provide interpretable spatial patterns, but their locations, scales, and intensities are selected for the simulation rather than estimated from empirical observations. Consequently, they cannot fully capture the irregular spatial structure and heterogeneous local conditions of real urban areas.

Second, the simulated dataset includes only four explanatory factors. These factors provide a suitable controlled setting for demonstrating factor selection and model comparison, but real population movement may depend on a much larger collection of location characteristics. Relevant information may include housing supply and prices, schools, hospitals, public services, parks, land use, demographic composition, and different types of commercial and transportation facilities. The current dataset therefore represents only a limited subset of the points of interest and spatial conditions that may affect population distribution.

Third, the spatial factors remain fixed throughout the simulation. Population changes in response to job opportunities, commercial activity, amenities, and transportation accessibility, but these factors are not updated over time. In real cities, employment locations, businesses, public services, amenities, and transportation networks may change periodically. The current dataset therefore cannot represent population responses to newly opened facilities, infrastructure changes, business relocation, or other updates to the urban environment.

Fourth, the simulated observations are complete, smooth, and internally consistent. The population field and all spatial factors are available on the same grid at every time step, without missing observations or abnormal values. Real urban datasets may contain locally missing records, incomplete points of interest, measurement errors, inconsistent spatial coverage, and a small number of unusually high or low values. Because these problems are not included in the simulated dataset, the current experiments do not fully test the robustness of the framework under imperfect empirical observations.

A principal direction for improvement is to introduce real urban data. Population grids, employment records, points of interest, land-use information, and road or public-transit networks could be aligned to a common spatial framework and used to construct empirical factor fields. Unlike the current four-factor setting, real-data applications could include a larger number of points-of-interest categories and allow the factor-selection stage of the pipeline to determine which variables provide useful information.

The factor fields could also be updated periodically when new observations become avail-

able. For example, updated business, employment, transportation, and public-service data could be used to reconstruct the spatial inputs at regular intervals. This would allow the model to respond to changes in the urban environment rather than assuming that all factors remain fixed throughout the prediction period.

Finally, the data-processing procedure could be extended to handle locally missing or abnormal observations. Spatial interpolation or information from neighbouring grid cells could be used when a limited number of values are missing. Robust normalization, outlier detection, local smoothing, or robust loss functions could reduce the influence of a small number of abnormal values. Training experiments with artificially masked or perturbed inputs could also be used to evaluate whether the model remains stable when the input data are incomplete or contain measurement errors.

These limitations mean that the current dataset should be interpreted as a controlled benchmark for demonstrating the proposed methodology rather than as a complete empirical representation of urban population dynamics. Introducing real and periodically updated spatial data, expanding the range of points of interest, and improving robustness to local data-quality problems would provide the main path toward practical application.

Appendix B

Diffusion: Fick's Law

B.1 Diffusion: Fick's Law

Diffusion describes the process by which particles spread from regions of high concentration to regions of low concentration due to random motion. The fundamental laws governing diffusion were first formulated by Adolf Fick in 1855.

B.1.1 Fick's First Law

Fick's First Law states that the diffusive flux is proportional to the negative gradient of concentration. Mathematically, it can be written as

$$\mathbf{J} = -D\nabla C, \tag{B.1}$$

where $\mathbf{J}$ represents the diffusion flux (amount of substance per unit area per unit time), C denotes the concentration, D is the diffusion coefficient, and ∇C is the spatial gradient of the concentration. The negative sign indicates that diffusion occurs from regions of higher concentration to regions of lower concentration.

B.1.2 Fick's Second Law

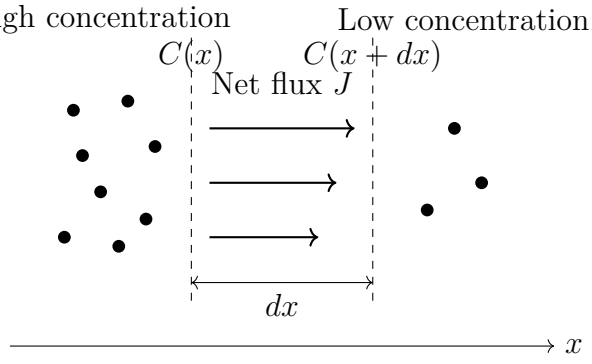
Fick's Second Law describes the time evolution of the concentration field. It is derived by combining Fick's First Law with the conservation of mass. The resulting partial differential equation is

$$\frac{\partial C}{\partial t} = D\nabla^2 C, \tag{B.2}$$

where ∇^2 denotes the Laplacian operator. This equation describes how the concentration changes over time due to diffusion.

Fick's Second Law is also known as the diffusion equation and forms the mathematical basis for modeling diffusion processes in physics, chemistry, biology, and many other fields.

B.2 Graphical Illustration of Fick's First Law



Consider two neighboring positions in a medium located at x and $x+dx$, as illustrated in the figure. The particle concentrations at these locations are denoted by $C(x)$ and $C(x+dx)$ respectively.

If the concentration at the left side is larger than that on the right side, i.e.

$$C(x) > C(x+dx),$$

then the number of particles moving from left to right due to random motion is greater than the number moving in the opposite direction. As a result, a net particle flux J appears in the positive x direction.

The magnitude of this net flux is proportional to the difference in concentration between the two locations:

$$J \propto C(x) - C(x+dx).$$

Dividing by the distance dx between the two points gives

$$J \propto -\frac{C(x+dx) - C(x)}{dx}.$$

Taking the limit as $dx \to 0$ leads to

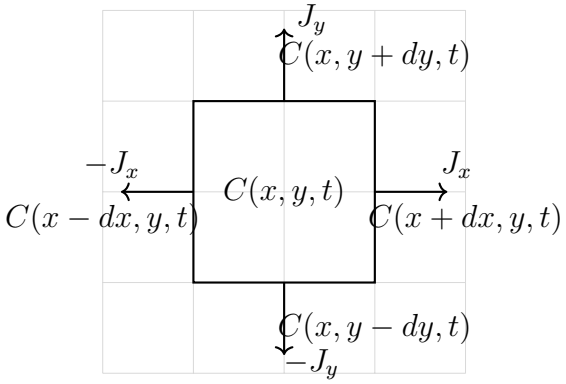
$$J = -D\frac{\partial C}{\partial x},$$

where D is the diffusion coefficient. In vector form, this relation can be written as

$$\mathbf{J} = -D\nabla C.$$

This result is known as **Fick's First Law**, which states that the diffusion flux is proportional to the negative gradient of the concentration field.

B.3 Graphical Illustration of Fick's Second Law



Consider a small region centered at (x, y) with concentration $C(x, y, t)$. Particles diffuse between this region and its neighboring regions located at

$$(x \pm dx, y), \quad (x, y \pm dy).$$

According to Fick's First Law, the diffusion flux in each direction is proportional to the concentration gradient:

$$J_x = -D\frac{\partial C}{\partial x}, \qquad J_y = -D\frac{\partial C}{\partial y}.$$

The change in concentration inside the small region results from the imbalance between incoming and outgoing fluxes. If more particles leave the region than enter it, the concentration decreases; if more particles enter, the concentration increases.

Using conservation of mass, the rate of change of concentration equals the negative divergence of the flux:

$$\frac{\partial C}{\partial t} = -\left(\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y}\right).$$

Substituting Fick's First Law into this relation yields

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right).$$

This equation is known as **Fick's Second Law in two dimensions**. It describes how the concentration field evolves over time due to diffusion.

Appendix C

Advection–Diffusion

C.1 Advection–Diffusion

Advection–Diffusion describes the transport of a scalar quantity, such as pollutant, heat, or particle concentration, under the combined effects of advection, which is transport by flow, and diffusion, which is spreading due to random motion.

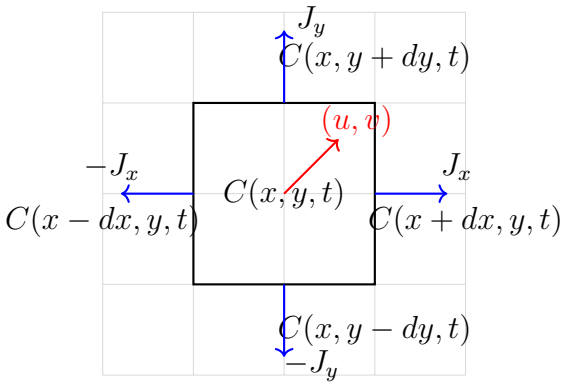
C.1.1 Governing Equation

In two dimensions, the advection–diffusion equation is

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right), \tag{C.1}$$

where $C(x, y, t)$ is the concentration, u and v are the velocity components in x and y directions, and D is the diffusion coefficient. The left-hand side represents advection, while the right-hand side represents diffusion.

C.1.2 Graphical Illustration



C.1.3 Explanation

Consider a small region centered at (x, y) with concentration $C(x, y, t)$.

- **Diffusion:** Particles flow between this region and its neighbors according to Fick's law. The diffusion flux in x and y directions are

$$J_x = -D\frac{\partial C}{\partial x}, \quad J_y = -D\frac{\partial C}{\partial y}.$$

  The net change in concentration due to diffusion is the divergence of the diffusion flux:

$$\left.\frac{\partial C}{\partial t}\right|_{\text{diffusion}} = -\left(\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y}\right) = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right).$$

- **Advection:** The scalar quantity is also transported by the flow with velocity (u, v), which moves the concentration along the direction of the flow. The contribution of advection to the rate of change is

$$\left.\frac{\partial C}{\partial t}\right|_{\text{advection}} = -\left(u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y}\right).$$

- **Combined effect:** Summing the contributions from diffusion and advection gives the full advection–diffusion equation:

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right).$$

This equation models how the concentration evolves over time under both flow-driven transport, advection, and spreading due to random motion, diffusion. The blue arrows in the

figure illustrate the diffusion flux, while the red arrow shows advection by the flow velocity (u, v).

Bibliography

[1] Takashi Akamatsu, Yuki Takayama, and Kiyohiro Ikeda. Spatial discounting, fourier, and racetrack economy: A recipe for the analysis of spatial agglomeration models. *Journal of Economic Dynamics and Control*, 36(11):1729–1759, 2012. doi: 10.1016/j.jedc.2012.04.010.

[2] Alex Anas, Richard Arnott, and Kenneth A. Small. Urban spatial structure. *Journal of Economic Literature*, 36(3):1426–1464, 1998.

[3] Isaac Bayoh, Elena G. Irwin, and Timothy C. Haab. Determinants of residential location choice: How important are local public goods in attracting homeowners to central city locations? *Journal of Regional Science*, 46(1):97–120, 2006. doi: 10.1111/j.0022-4146.2006.00434.x.

[4] European Commission, Joint Research Centre. Ghs population grid. https://human-settlement.emergency.copernicus.eu/ghs_pop.php, 2026. Global Human Settlement Layer (GHSL), accessed April 16, 2026.

[5] Zide Fan, Tao Pei, Ting Ma, Yanjun Du, Ci Song, Zhi Liu, and Chenghu Zhou. Estimation of urban crowd flux based on mobile phone location data: A case study of beijing, china. *Computers, Environment and Urban Systems*, 69:114–123, 2018. doi: 10.1016/j.compenvurbsys.2018.01.005.

[6] Masahisa Fujita and Tomoya Mori. Frontiers of the new economic geography. *Papers in Regional Science*, 84(3):377–405, 2005. doi: 10.1111/j.1435-5957.2005.00021.x.

[7] Masahisa Fujita and Hideaki Ogawa. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12(2):161–196, 1982. doi: 10.1016/0166-0462(82)90031-X.

[8] Masahisa Fujita, Paul Krugman, and Anthony J. Venables. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, MA, 2001.

[9] Elisabeth Gruber et al. Territorial cohesion under pressure? welfare policy and shrinking services in rural and peripheral regions. *Journal of Rural Studies*, 2019.

[10] John C. Hull. *Machine Learning in Business: An Introduction to the World of Data Science*. Self-published / independent publication, 2020.

[11] Sibren Isaacman, Richard Becker, Ramón Cáceres, Stephen Kobourov, James Rowland, and Alexander Varshavsky. Identifying important places in people's lives from cellular network data. In *Pervasive Computing*, pages 133–151. Springer, 2011. doi: 10.1007/978-3-642-21726-5_9.

[12] Tae-Kyung Kim, Mark W. Horner, and Robert W. Marans. Life cycle and environmental factors in selecting residential and job locations. *Housing Studies*, 20(3):457–473, 2005. doi: 10.1080/02673030500062335.

[13] Paul Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499, 1991. doi: 10.1086/261763.

[14] James Laird, Peter Mackie, and David Johnson. Commuting costs and their impact on wage rates. Technical report, Institute for Transport Studies, University of Leeds, 2006.

[15] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 255–258. MIT Press, 1995.

[16] Fei-Fei Li, Justin Johnson, Serena Yeung, et al. Cs231n: Convolutional neural networks for visual recognition – convolutional neural networks. https://cs231n.github.io/convolutional-networks/, 2026. Stanford University course notes.

[17] Fei-Fei Li, Justin Johnson, Serena Yeung, et al. Cs231n: Convolutional neural networks for visual recognition – recurrent neural networks. https://cs231n.github.io/rnn/, 2026. Stanford University course notes.

[18] Guowei Luo, Jiayuan Ye, Jinfeng Wang, and Yi Wei. Urban functional zone classification based on poi data and machine learning. *Sustainability*, 15(5):4631, 2023. doi: 10.3390/su15054631.

[19] Pascal Mossay. Increasing returns and heterogeneity in a spatial economy. *Regional Science and Urban Economics*, 33(4):419–444, 2003. doi: 10.1016/S0166-0462(02)00041-8.

[20] Kensuke Ohtake. Pattern formation by advection-diffusion in new economic geography. *Decisions in Economics and Finance*, 48(2):515–549, 2025. doi: 10.1007/s10203-025-00533-w.

[21] OpenStreetMap Wiki contributors. Downloading data. https://wiki.openstreetmap.org/wiki/Downloading_data, 2025. OpenStreetMap Wiki, accessed April 16, 2026.

[22] OpenStreetMap Wiki contributors. Map features. https://wiki.openstreetmap.org/wiki/Map_features, 2025. OpenStreetMap Wiki, accessed April 16, 2026.

[23] Muntahith Mehadil Orvin and Mahmudur Rahman Fatmi. A residential location search model based on the reasons for moving out. *Transportation Letters*, 16(6):566–580, 2024. doi: 10.1080/19427867.2023.2222990.

[24] Gianmarco I. P. Ottaviano, Takatoshi Tabuchi, and Jacques-François Thisse. Agglomeration and trade revisited. *International Economic Review*, 43(2):409–436, 2002. doi: 10.1111/1468-2354.t01-1-00021.

[25] Michael Pflüger. A simple, analytically solvable, chamberlinian agglomeration model. *Regional Science and Urban Economics*, 34(5):565–573, 2004. doi: 10.1016/S0166-0462(03)00043-7.

[26] Patrick M. Schirmer, Michael A. B. van Eggermond, and Kay W. Axhausen. The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2):3–21, 2014. doi: 10.5198/jtlu.v7i2.740.

[27] Statistics Canada. Non-response variability guidelines. https://www.statcan.gc.ca/en/statistical-programs/document/5065_D3_T9_V1-eng.pdf, 2006. accessed April 2026.

[28] Statistics Canada. Data quality, concepts and methodology. https://www150.statcan.gc.ca/n1/pub/11-526-x/2013001/technote-notetech1-eng.htm, 2013. accessed April 2026.

[29] Statistics Canada. Concepts, definitions and data quality. https://www.statcan.gc.ca/en/statistical-programs/document/2101_D3_T2_V115, 2020. accessed April 2026.

[30] Statistics Canada. 2021 census data quality guidelines. https://www12.statcan.gc.ca/census-recensement/2021/ref/98-26-0006/982600062021001-eng.cfm, 2022. accessed April 2026.

[31] Erik Sundqvist et al. A study of regional political leadership in sweden and the downward spiral of unemployment, shrinking tax bases, and withdrawal of public services. *Regional Studies*, 2023.

[32] Jameson L. Toole, Michael Ulm, Dietmar Bauer, and Marta C. González. Inferring land use from mobile phone activity. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, 2012.

[33] Jameson L. Toole, Serdar Çolak, Bradley Sturt, Lauren P. Alexander, Alexandre Gonçalves Evsukoff, and Marta C. González. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58:162–177, 2015. doi: 10.1016/j.trc.2015.04.022.

[34] Jos van Ommeren and Mogens Fosgerau. Workers' marginal costs of commuting. *Journal of Urban Economics*, 65(1):38–47, 2009. doi: 10.1016/j.jue.2008.08.001.

[35] Wenlai Wang, Tao Pei, Jie Chen, and Ci Song. Population distributions of age groups and their influencing factors based on mobile phone location data: A case study of beijing, china. *Sustainability*, 11(24):7033, 2019. doi: 10.3390/su11247033.

[36] Yingkun Yang, Chen Xiong, Junfan Zhuo, and Ming Cai. Detecting home and work locations from mobile phone cellular signaling data. *Mobile Information Systems*, 2021: 5546329, 2021. doi: 10.1155/2021/5546329.

[37] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 186–194, 2012. doi: 10.1145/2339530.2339561.

[38] Xi Zhao et al. Mapping hourly population dynamics using remotely sensed and geospatial data: A case study in beijing. *GIScience & Remote Sensing*, 58(6):717–732, 2021. doi: 10.1080/15481603.2021.1935128.

[39] Barry Zondag and Marits Pieters. Influence of accessibility on residential location choice. *Transportation Research Record*, 1902(1):63–70, 2005. doi: 10.1177/0361198105190200108.