The Journal of Transport and Land Use jtlu.org
Vol. 19 No. 1 [2026] pp. 475-508
https://doi.org/10.5198/jtlu.2026.2892

JTLU

Land use supply at the parcel level in the metropolitan-scale area: Framework, application, and behavioral insights

Jiaxin (Jason) Zhang, University of Toronto, moonjiaxin.zhang@mail.utoronto.ca
Mwendwa Kiko, University of Toronto, mwendwa.kiko@mail.utoronto.ca
Eric J. Miller, University of Toronto, eric.miller@utoronto.ca

Abstract: Land-use (LU) supply is a critical component of integrated urban systems, but its representation remains underdeveloped in many metropolitan-scale models. Existing frameworks have not jointly achieved parcel-level spatial resolution, explicit integration of policy-based availability constraints, behaviorally interpretable decision rules, and inter-sectoral disaggregation of developer preferences. This paper develops a metropolitan-scale, parcel-level framework for LU supply modelling that addresses all four gaps simultaneously. The framework adopts the land parcel as the decision-making unit, constructs heterogeneous feasible choice sets governed by zoning by-laws and capacity constraints, and employs a panel-data mixed logit model as its behavioral engine. The framework is demonstrated through a case study of commercial real estate development in the Greater Toronto-Hamilton Area (GTHA), Canada, from 2015 to 2023, covering retail, industrial, office, and mixed-use development. Results show that developers respond more consistently to market dynamics than to static market levels, suggesting that market inertia plays a critical role in shaping new supply. The model also reveals substantial inter-sectoral heterogeneity: retail, industrial, and office developers follow fundamentally distinct locational logics. Furthermore, both spatial and temporal sensitivity analyses suggest that developer behavior is scale-dependent, with shorter-term effects being particularly influential. Overall, the study demonstrates that parcel-level, policy-constrained LU supply modelling is both feasible and informative under limited data conditions and offers methodological insights for future research.

Keywords: parcel-level modelling, land-use supply, commercial built space, panel-data mixed logit, inter-sectoral behavioral heterogeneity

1 Introduction

Land-use (LU) supply is a foundational component of integrated urban systems. Spatial distribution of new real estate buildings directly influences travel demand, employment accessibility, and the long-term sustainability of metropolitan regions (Waddell et al., 2007). However, compared to the demand side, the supply-side representation in most integrated urban models remains underdeveloped (Miller, 2018). Many operational models still rely on coarse spatial units, which mask intra-zonal dynamics and homogenize driving factors across heterogeneous parcels (Pan et al., 2010).

Although computational LU researchers have made non-trivial contributions toward advanced modelling, existing frameworks have not jointly achieved the four capabilities required for a comprehensive supply-side tool: 1) parcel-level spatial resolution operable at metropolitan scale; 2) explicit integration of legislated policy constraints into individual-specific choice sets; 3) behaviorally interpretable decision rules that reveal why development occurs, not merely where; and 4) inter-sectoral disaggregation that

Copyright 2026 Jiaxin (Jason) Zhang, Mwendwa Kiko & Eric J. Miller
ISSN: 1938-7849 | Licensed under the Creative Commons Attribution – Noncommercial License 4.0
Received April 14, 2026; Revised October 22, 2026; Accepted November 2, 2026; Available online November 19, 2026

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

uncovers the heterogeneous preferences among different types of developers. Each of these gaps might be partially addressed in isolation, but a study that holistically addresses all of them is missing.

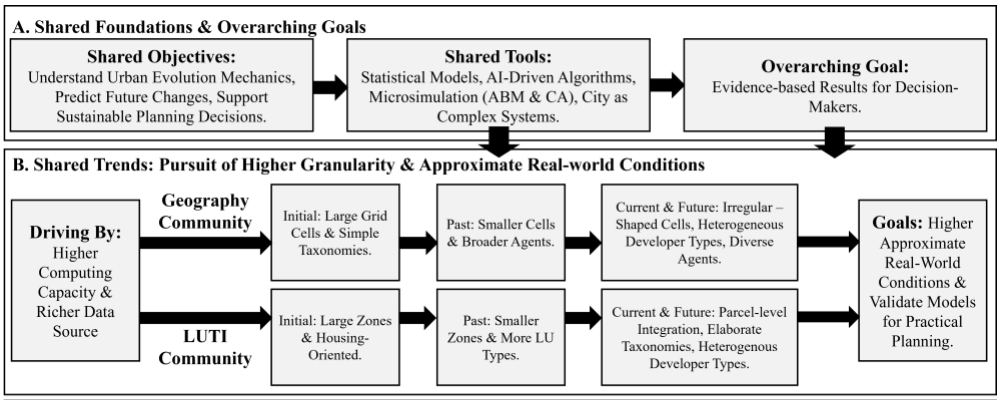
This paper proposes a metropolitan-scale, high-spatial-resolution framework for LU modelling that addresses all four gaps simultaneously. The framework adopts the land parcel as the fundamental decision unit, constructs heterogeneous, feasible choice sets governed by real-world constraints, and employs a panel-data mixed logit model as its econometric behavioral engine. The framework is demonstrated through a case study of commercial real estate development (retail, industrial, office, and mixed-use) in Canada's Greater Toronto-Hamilton Area (GTHA) from 2015 to 2023. Key findings reveal that developers respond more consistently to market inertia than to static levels, that retail, industrial, and office developers follow fundamentally distinct locational logics, and that both spatial and temporal scale choices can influence model estimation.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature and identifies research gaps; Section 3 presents the conceptual framework; Section 4 details the GTHA case study; Section 5 reports and discusses results; and Section 6 addresses limitations, future directions, and concluding remarks.

2 Literature review

2.1 Interdisciplinary perspectives on land-use dynamics

Understanding LU dynamics is a shared interest of several research disciplines, particularly computational LU geography and Land Use-Transport Interaction (LUTI) modelling. This review builds on the work of both communities, which share many similarities (Figure 1). They both increasingly rely on computational tools such as microsimulation models to explain what drives LU change, to develop forecasting tools, and to assist better decision-making. They also follow a common trend toward finer granularity. For example, spatial units have recently been downscaled from the large grid cells in the 1990s (Clarke & Gaydos, 1998) to the land-parcel level in more recent studies (Xu et al., 2023).



C. Divergences: Primary Objectives, Methods, and Spatial Scales

Aspects	Geography Community (Landscape Focus)	LUTI Community (Human Activity Focus)
1. Scope & LU Definition	Land Use & Land Cover Change (LULC); Includes Specific Land Cover Types (Grass, Forest, Water).	"Land Use" Only (Human Utilisation); General Natural Features, e.g., "Open Space".
2. Focus & Primary Objective	Focus on Land Surface Changes; Evolution of "Landscape".	"Human Activities" & Real Estate; Focus More on the Development of Built Spaces.
3. Methodological Preference	AI-Driven Algorithms, ABM, CA Models; Prioritizes Higher Prediction Accuracy.	Econometric Methods as the Cornerstone; Prioritizes Balanced Behavioural Interpretability & Prediction Accuracy.
4. Spatial Scale	Broader Spectrum: Single City -> National -> Global.	Single City, City Cluster, Metropolitan Region.
5. Integration Goal	Optimise Standalone LU Model.	Seamless Integration with Transport-side Modules.

Figure 1. Interdisciplinary convergence and divergence

The main difference between the two fields lies in their emphasis (Figure 1). Geographers frequently model both "land uses" (residential or commercial) and "land covers" (grass, forest, and water) (Guan et al., 2011), focus on the evolution of landscapes, and give priority to predictive accuracy. Therefore, they embrace AI-driven methods more actively (e.g., Gharaibeh et al., 2020; Islam et al., 2018; Sajan et al., 2022). In contrast, LUTI modellers define LU by human utilization (Hunt et al., 2008), focus on the development decisions that shape real-estate markets, and are less willing to sacrifice behavioral interpretability, therefore keep econometric modelling as the methodological core (e.g., Waddell et al., 2007). The two fields are complementary, and this paper leverages contributions from both communities.

2.2 Methodological foundations

This paper conceptualizes the city through the lens of complexity theory (Batty, 2013). Under this view, microsimulation approaches such as Agent-based Modelling (ABM) and Cellular Automata (CA) have emerged as dominant methods (Guan et al., 2011; Jjumba & Dragićević, 2012). ABMs represent individual stakeholders and their interactions explicitly. They are a suitable choice when such interactions (e.g., among planners, landowners, and developers) are important for the research question (Jjumba & Dragićević, 2012). CA can be viewed as a fixed-topology subset of ABM, wherein the "agents" are stationary cells. In CA, each piece of land can be treated as a single cell that represents an individual decision unit (Tobler, 1979). This representation is suitable for LU supply and is also more computationally efficient than ABM. Therefore, CA is a more practical foundation for the metropolitan-scale and high-resolution modelling that this paper targets.

Many leading CA frameworks, such as SLEUTH (Clarke & Gaydos, 1998) and PLUS (Xu et al., 2022), have supported large-scale studies (Jantz & Goetz, 2005), including applications that extend to global settings (Zhou et al., 2019). Some LUTI-oriented frameworks adopt a similar structure for their LU supply models, such as PECAS (Hunt et al., 2008) and UrbanSim (Waddell et al., 2007). Although not all of these frameworks explicitly identify themselves as CA models, this paper refers to their shared modelling logic as CA-like. Under this logic, each piece of land is treated as a stationary agent characterized by its surrounding geospatial context, such as accessibility, market conditions, and neighborhood activities. The agent's attributes and context are then translated into a probability-based decision, such as development or inaction. In classical CA, this mechanism is commonly termed the transition rule (e.g., Ding et al., 2022), whereas this paper uses the broader term "decision rule" to emphasize its behavioral interpretation.

2.3 Research trends and gaps

2.3.1 Demands for higher spatial resolution

The choice of spatial resolution is critically important in LU modelling (Pan et al., 2010). Despite the movement toward finer spatial granularity (Huang et al., 2024), most large-scale studies still use coarse zones as their basic spatial unit, which can be categorized into two types. The first is regular-shaped zones, such as grid pixels ranging from 30m to 500m (Darrel Jenerette & Potere, 2010; Wang et al., 2018; Xu et al., 2022; Xu et al., 2023). The second is irregular-shaped zones, typically derived from administrative boundaries, such as census zones and transportation analysis zones

(Farooq, 2011; Jjumba & Dragićević, 2012; Martinez, 1996). Coarse resolution creates at least two problems. First, it neglects intra-zonal dynamics, since a large zone containing several LUs is usually labelled by its dominant use or simply as mixed-use, which masks transitions inside it. Second, it treats driving factors as homogeneous, since aggregating many parcels into one zone averages their different attributes and obscures local variation.

Therefore, although high-resolution models come with greater computational burdens and are more data-intensive, leading regional integrated modellers have attempted to tackle this problem over the past decades. Among the substantial number of models developed, to the authors' knowledge, UrbanSim and PECAS are the only two leading frameworks that have succeeded in this endeavor. However, UrbanSim has recently shifted toward commercialization, which inevitably limits the transparency of its methodological details. While its website (https://www.urbansim.com/) and a model report (Waddell et al., 2018) state the achievement of parcel-level simulation, the specific approach and overall model performance remain unknown.

PECAS's attempt at parcel-level modelling traces back to 2008, documented in a Transportation Research Board conference paper (Hunt et al., 2008). Possibly constrained by conference submission word limits, the documentation of methodological and validation details remains incomplete. In a more recent publication (Wang et al., 2020), the model's uncertainty analysis is comprehensively presented through a case study of the Greater Baltimore Region. Overall, compared to the extensive modelling efforts in other dimensions (Nguyen et al., 2025), the path toward high-resolution LU modelling remains significantly unclear and demands considerably greater attention.

2.3.2 Challenges to establish policy-based availability constraints

Land development is a process strictly regulated by government policy, such as zoning by-laws (City of Toronto, 2018), and not every parcel is permitted to accommodate every type of building use. In LU models, however, such policy constraints can be relatively relaxed when coarse resolution is adopted or when regular-shaped spatial units (grid cells) are used. Both situations create a context where a single spatial unit may contain multiple LU types, or multiple parcels regulated by different policies. It is therefore unsurprising that much of the literature (Huang et al., 2024) has not incorporated policy constraints. Although many irregular-shaped, coarse-resolution models have acknowledged the necessity of incorporating policy constraints (Waddell et al., 2003), how to achieve this is frequently under-discussed.

When downscaling to the parcel level, however, policy constraints become a nontrivial consideration. Every parcel typically has its own specific regulations (City of Toronto, 2018), making ambiguous or approximate representations unacceptable. However, computational modelling requires policy constraints in digital format (ESRI shapefiles), whereas many local municipalities within a large region only provide text-based policy documents. As a result, parcel-level modelling practice is largely restricted to cities with available digital policy data, which is rarely achievable in countries like Canada, where many local municipalities lag behind in the digitization of planning policies (Zhang, 2026). In contrast, available evidence suggests that the PECAS Baltimore case study did not face this problem (Wang et al., 2020).

2.3.3 Insufficient behavioral interpretability

The decision rule is a core methodological focus in CA-like studies. In early studies, it was defined by relatively simple mechanisms, such as neighborhood similarities and

suitability assessments (White et al., 1997). Since the early 2000s, modellers have increasingly incorporated additional techniques to enhance model performances, such as Markov Chains (Guan et al., 2011; Hyandye & Martz, 2017; Wang et al., 2018), spatial regression model (Feng & Tong, 2018; Ku, 2016), and System Dynamics approaches (Lauf et al., 2012; Xu et al., 2016).

Since the late 2010s, AI-driven algorithms have been increasingly used. Commonly adopted methods include decision trees, random forests, and support vector machines (Liang et al., 2021; Yang et al., 2008). Neural networks, in particular, have drawn significant attention over the past decade (Gharaibeh et al., 2020; Islam et al., 2018; Sajan et al., 2022). Moreover, integrations can extend beyond a single method. For example, Xing et al. (2020) combined recurrent neural networks with convolutional neural networks to capture long-term temporal dependencies. Similarly, Huang et al. (2024) developed multi-network architectures. Notably, hybrid frameworks can also span different modelling paradigms. Gharaibeh et al. (2020) fused CA for spatial interactions, Markov Chains for empirically grounded temporal transition rates, and ANN for learning nonlinear influences, thereby maximizing the strengths of each approach simultaneously.

While acknowledging these substantial contributions, the primary focus of LU studies has been on improving prediction accuracy, which is nontrivial for valid scenario-based analysis. However, most of the aforementioned frameworks lack the ability to comprehensively identify the factors driving LU change and the behavioral logic underlying development decisions. This gap has drawn attention from discrete-choice-oriented modellers (Zhao & Peng, 2012). However, these studies either rely almost solely on market-signal attributes with few geospatial or locational factors considered (Hunt et al., 2008), or depend heavily on locational factors (Zhao & Peng, 2012) with fewer market-signal drivers involved. We suspect that this asymmetric focus is at least partially caused by resource limitations. This study aims to move beyond this constraint by not only conducting comprehensive behavioral interpretations but pursuing an in-depth understanding of how the choice of spatial and temporal scales influences development decisions.

2.3.4 Insufficient inter-sectoral comparison

A final gap concerns the limited ability of existing LU models to compare development behavior across detailed LU sectors. LU development is a long-term process involving various stakeholders. Even from a developer's perspective alone, developers of residential, retail, industrial, office, and institutional projects may respond to different market and locational conditions (Zhao & Peng, 2012). However, partly due to the primary focus on accuracy and the relatively coarse resolution discussed above, existing models often have limited ability to reveal these differences (Guan et al., 2011; Jantz & Goetz, 2005; Wang et al., 2018; White et al., 1997; Xu et al., 2022).

Although exceptions exist, such as Waddell et al. (2007), Jjumba and Dragićević (2012), and Zheng et al. (2015), many still aggregate detailed activities into broad classes. For example, while "commercial" use is frequently distinguished from "residential" and "open space", most of the studies cited above aggregate "retail," "office," and "industrial" into a single consolidated category, with some exceptions only treating "industrial" as a separate category. This aggregation limits the ability to identify whether different commercial sectors follow different development logics. A sector-explicit framework is therefore critical to deepening the understanding of the land development process.

2.4 Literature review synthesis and research objectives

Accordingly, this paper aims to develop a metropolitan-scale LU supply framework that addresses these gaps by modelling parcel-level development decisions through zoning-constrained, behaviorally interpretable, and sector-explicit choice sets. To clarify how the research gaps identified above are carried forward through the remainder of the paper, Table 1 links each gap to the corresponding framework component, case-study implementation, and empirical assessment. The next section presents the framework, beginning with its core assumptions and overall structure.

Table 1. Alignment between research gaps, framework design, and empirical implementation

Research Gap	Framework Component	Case Study Component	Results & Findings
Higher Spatial Resolution ***(Section 2.3.1)***	Land parcel adopted as the fundamental decision unit within a high-resolution environment. *(Section 3.1.1 & 3.2.1)*	The GTHA case study adopts provincial assessment parcels as the modelling units and integrates commercial development records, zoning information, market signals, and spatial attributes into a parcel-level spatiotemporal database. *(Section 4.1 & 4.2)*	Embedded in all model estimation results in *Section 5*
Policy-based Availability ***(Section 2.3.2)***	Feasible choice set defined by legislated permissions. *(Section 3.2.2 & 3.2.3)*	GTHA-wide parcel-level digital zoning dataset gives each parcel a heterogeneous, zoning-constrained choice set. *(Section 4.2)*	Embedded in all model estimation results in *Section 5*
Behavioral Interpretability ***(Section 2.3.3)***	RUM-based econometric discrete-choice decision rule, chosen over black-box AI for behavioral transparency. *(Section 3.1.2, 3.1.3 & 3.2)*	Panel-data mixed logit estimated with random coefficients after screening a large number of candidate variables. *(Section 4.2 & 4.3)*	*Section 5.1 & 5.3*
Inter-Sectoral Comparison ***(Section 2.3.4)***	Sector-explicit choice set *(Section 3.2.2)*	Retail, Industrial, Office, Mixed and NO modelled as separate alternatives. *(Section 4.1 & 4.3)*	*Section 5.2*

3 Conceptual framework

The framework views LU supply as a dynamic parcel-level decision process. It relies on two fundamental assumptions. First, cities are viewed as complex adaptive systems whose shape, morphology, built form, and flows of information and energy arise from the bottom-up behaviors of individual agents rather than from top-down control (Batty, 2013). Second, these agents are assumed to be rational decision-makers who select the alternative that maximizes their utility. Although utility is a theoretical construct, it provides a quantitative summary of the perceived benefits of different choices under specified influencing factors, consistent with Random Utility Maximization (RUM) theory (Train, 2009). Collectively, these two assumptions lead to a microsimulation-

based, choice-driven representation of land development, which is based on two interconnected components: a high-resolution environment that defines where development can occur, and a behavioral engine that determines what is developed and when.

The complete summary and logic of the proposed framework is provided in Figure 2, which presents in four parts. The first two parts of the diagram constitute the high-resolution environment, while the third part illustrates the behavioral engine. The fourth part is not a working component of the current framework, but an "IDEAL" data standard that represents the data quality expected of a more fully developed framework in the future, which will be discussed at the end of this section.

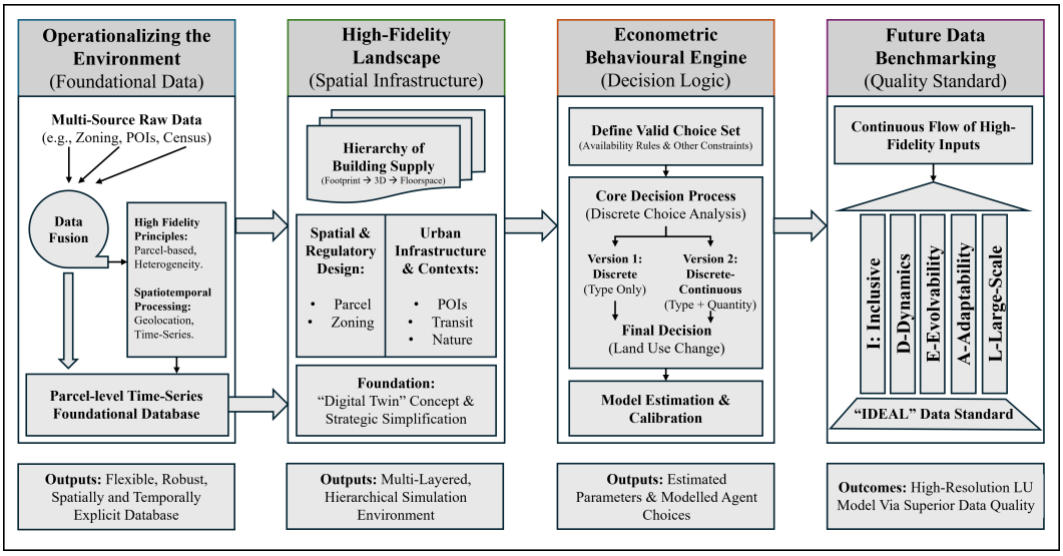


Figure 2. Overview of the proposed framework. Note: The fourth panel presents an aspirational future data standard and is not currently operational in the presented framework.

3.1 Operationalizing the high-resolution environment

The first component is the high-resolution environment, which represents the physical and regulatory setting in which development takes place. LU supply is defined as a unit of physical stock, regardless of the number or diversity of commercial tenants it houses. Accordingly, individual business establishments are conceptualized as demand-side agents that occupy the supply generated by this framework. Such an environment should be resolved at the spatial scale where development is actually decided and regulated. This requirement raises three questions: at what unit, with what data, and through what derived attributes.

3.1.1 The fundamental spatial unit

As the first design decision, the selection of an appropriate high-resolution spatial unit is guided by four principles. First, the unit should respect by-law boundaries, which are the administrative borders separating distinct regulatory regimes to ensure no single unit spans two jurisdictions with different policies. Second, it should align with real-world planning units, which are the units of decision-making in the real estate market, such as the land parcel in Canada. Third, it should limit LU diversity to ensure clarity in classification, with a single LU category dominating the available space. Finally, it should balance fidelity and cost, rather than pursuing the finest possible resolution, to

optimize the trade-off between computational demand and analytical precision. Following these principles, the land parcel is recognized as the fundamental unit of this framework.

3.1.2 Data requirements and temporal structure

With the parcel adopted as the basic unit, the data attached to it should next be assembled. Data collection is a big challenge in urban research, and higher resolution further raises the bar for data quality. Recognizing this data-intensive nature, this framework proposes a number of baseline data requirements (Figure 3). Raw data from various sources can carry distinct information and vary in spatiotemporal scale and format. Consequently, although labor- and time-intensive, data preparation (cleaning, manipulation, filtering, and integration) is critical, and the example workflow is illustrated in Figure 3.

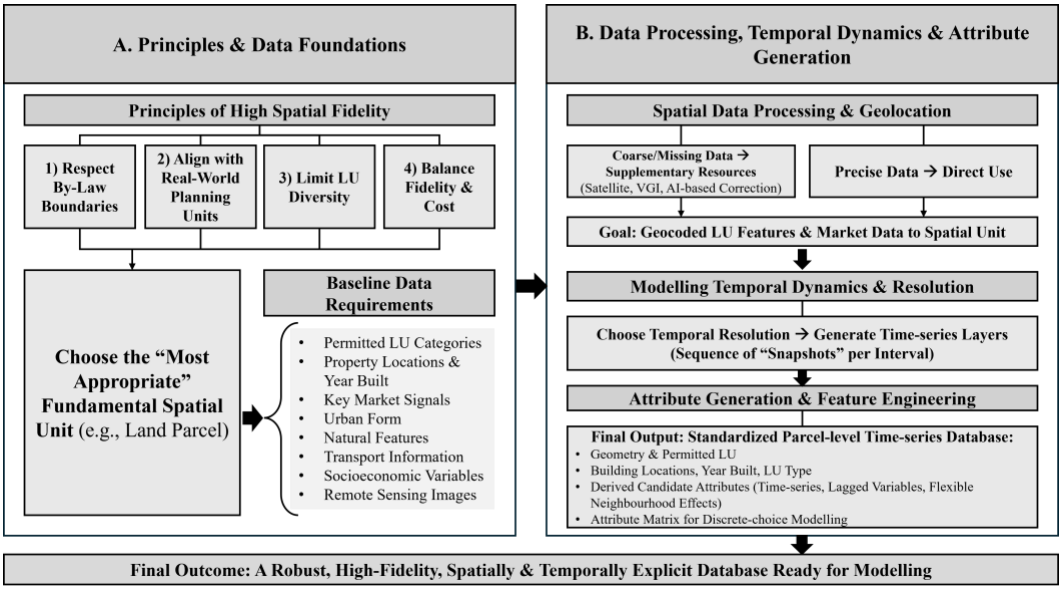


Figure 3. Data collection, processing & feature engineering

Beyond their content, these inputs also vary over time. Land development is inherently a spatiotemporal process, and time is often discretized in LU models, leading to what are termed "quasi-dynamic" models (Coulombel, 2010). It is suggested that temporal resolution should be chosen first and aligned with the research objectives and data availability. Common choices include monthly, quarterly, annual, five-year, and even ten-year increments (Farooq, 2011; Jantz & Goetz, 2005; Wang et al., 2018; Xu et al., 2022), with one year being the most common among operational models (Waddell et al., 2018; Wang et al., 2020). This framework generates a sequence of time-series layers, each showing the location and type of every feature built on an annual basis. Once the targeted LU features are processed, the procedure is repeated for other variable types to assemble a parallel spatiotemporal database of market signals, locational characteristics, demographic variables, and so on.

Among these inputs, building data can particularly be supplied at different levels of fidelity. The current framework targets the baseline situation where every building's location and year built are known but floorspace data are unavailable. Therefore, it omits the consideration of building size and represents each building as a single unit of supply,

which captures the type of development rather than its quantity. However, more sophisticated objectives, such as joint discrete-continuous choice modelling, may require building size information. For this purpose, three hierarchical levels of building-data quality can be accommodated. These range from useable floorspace, to detailed 3D building representations, to a simpler combination of building footprint and height.

3.1.3 Attribute generation and feature engineering

Once the spatial unit and its supporting data are in place, the final step in building the environment is to turn them into model-ready attributes. The land parcel is defined as the decision-maker which is characterized by the relevant market and locational attributes. Time-series market signals are re-projected onto parcel boundaries, and the dataset is then updated to record every parcel's geometry and coordinates, policy constraints, and the building characteristics it encompasses. These data inform the key components of the decision rule, which in this framework corresponds to the discrete-choice model, including the availability of alternatives, the actual choice made, and each decision-maker's choice set. The dataset is then completed by deriving the full set of candidate attributes that potentially drive land development.

In generating these attributes, and given imperfect data availability, priority should be given to time-series variables whenever possible. For instance, if annual updates of transit-station locations are available, related measures should be recalculated for each year rather than treated as static. With geocoded data at high resolution and explicit temporal tagging, the framework also offers the flexibility to investigate spatial and temporal lags. Temporal lags are customized to capture trends over varying horizons. For instance, new industrial supply completed one year (short-term), three years (mid-term), or five years (long-term) before the decision year can be computed subject to data availability. Similarly, neighborhood effects can be defined using adjustable buffer distances (100m to 1km), reflecting interactions at different spatial scales. The capacity to flexibly define "neighborhood" boundaries is critically important in urban studies, as varying definitions can lead to distinct outcomes (Pan et al., 2010). Overall, these steps yield a prepared environment, based on which the behavioral engine acts to generate development decisions.

3.2 The econometric behavioral engine

The second component is the behavioral engine, which translates the prepared environment into development decisions. Discrete choice models are particularly effective for decoding agent behavioral patterns and establishing correlations between attributes and choices. Therefore, they remain the dominant methodological core in behavioral LU studies. While AI-driven approaches frequently demonstrate superior predictive accuracy, they are less suitable for this study. One key limitation that machine learning and AI-driven approaches faced was the lack of clear procedure for dealing with the situation where not all alternatives are available in every observation. Though this scenario rarely arises in traditional classification problems, it is a central part of discrete choice modelling in general, and of this framework in particular. There are techniques such as logit masking that can and have been used in contexts such as continual learning (Liang et al., 2025) and which could potentially be applied here. But examples of such use-cases were few and from relatively niche fields, which contributed to the decision not to test any AI models in this application of the framework. Nevertheless, the framework architecture is modular and permits the substitution of the econometric core with AI algorithms, should future objectives demand it and the capabilities of the algorithms permit it. Overall, operationalizing the engine requires three fundamental elements: the

decision-maker, the alternatives among which it chooses, and the rules that determine which alternatives are available.

3.2.1 Characterizing the decision-maker

Land development involves diverse stakeholders, such as landowners, developers, and builders (Farooq et al., 2013). Theoretically, these roles are distinct and clearly separated. However, in practice, functional boundaries are often blurred. Landowners may assume developer functions by retaining assets to maximize long-term returns rather than opting for an immediate sale. Similar role overlaps are observed among builders. These overlaps complicate the precise definition of agents, and many state-of-the-art LU models therefore adopt a simplified definition to address this issue (Wang et al., 2020).

Following this practice, this framework characterizes the agent as the individual possessing decision-making authority over the development of a particular land parcel. The primary criterion is not the agent's formal identity, but rather their authority to make final development decisions. Thus, the model assumes that fundamental preconditions have already been fulfilled, including clear land ownership, secured financial resources, and a completed feasibility assessment based on legislative regulations and other relevant constraints. For maximum flexibility, decision-maker is further limited to managing a single spatial unit (parcel), and their decision-making responsibility is restricted solely to this assigned unit. This structure reflects a CA-like principle in which land is subdivided into distinct cells, each independently determining its development state (Waddell et al., 2007; Wang et al., 2020). The framework therefore does not assign multiple land units to a single developer for coordinated decision-making (see Appendix 1 for the limitations of the "location choice" strategy). Such coordination, however, could easily be accommodated by extending the framework.

Each decision-maker is assigned a set of characteristics that define its status, knowledge, and behavioral constraints within the simulation. All decision-makers remain active at every time-step and must make exactly one choice. Each decision-maker also has an explicit choice set, which may differ across parcels. They retain a memory of historical development records, both on their own parcel and within the surrounding neighborhood, such as annual changes in LU. This allows the development process to be path dependent. The initialization of each decision-maker, including its pre-simulation development status, is based on empirical data and retained throughout the simulation. In terms of knowledge, decision-makers can access urban, transportation, and environmental attributes across the study area. They are also informed by the attributes, contexts, and decisions of other agents. However, following Tobler's first law of geography, the influence of these external conditions declines with distance (Tobler, 1979). Finally, decision-makers are aware of the official planning regulations relevant to their allocated land to ensure that simulated developments comply with these requirements.

3.2.2 The discrete choice architecture

Having characterized the decision-maker, the set of development choices available to it should then be specified. Decision-makers address three core questions: where, when, and what type of building to develop. The "where" and "when" decisions are implicitly modelled, as the decision of "what type" automatically determines the location and timing of the event, since one of the alternatives is the outside good. The choice set is composed of development alternatives. While a single LU classification scheme is not enforced, the inclusion of "No or Other Development" (NO) and "Mixed" is recommended. "NO" is the outside good in the model and represents a decision to either

postpone development or to select a development type not explicitly defined in the model. "Mixed" captures developments that combine multiple LUs. Modelling "Mixed" development presents significant challenges (Kausar et al., 2024), as it demands data of exceptional resolution. Few datasets reliably capture both the location and the functional composition of mixed-use properties. Consequently, modellers must establish specific criteria to identify mixed-use instances.

A fine-grained definition of "Mixed" improves realism but can undermine data sufficiency for model estimation. Mathematically, if there are "N" single LUs, the number of possible "Mixed" combinations is $\sum_{k=2}^{N}\binom{N}{k}$. For instance, four base types (apartment, retail, office, and industry) produce 11 combinations. Some of these may never be observed, but a fully detailed taxonomy creates a large choice set with many small groups. This fragmentation often leads to class imbalance and risks convergence failure during model estimation. Therefore, a careful selection of classifications is advocated. To address limited data quality and avoid the pitfalls of overly detailed "Mixed" classifications, a "project" concept is introduced. A "project" comprises all development actions occurring in the same timestep on the same parcel, shifting the focus from multi-use buildings to project-based groupings. This concept simplifies data requirements. A "project" does not always involve multiple LUs: if identical developments occur on the same parcel within a single timestep, they are also grouped into one "project."

3.2.3 Defining the feasible choice set (availability rules)

These alternatives form a universal choice set, but not all of them are available to every parcel. Two principal factors determine land availability in this framework (Figure 4). The first is legislated permissions. Each spatial unit's permitted LU types are derived from government-issued regulations, such as zoning codes or master plans. Although every jurisdiction issues controls on land development, the level of government responsible varies. In Canada, municipalities hold zoning authority (City of Toronto, 2018), whereas Japan and Singapore, for example, use a national "Master Plan" to prescribe zone categories (City Planning Division, 2016; Urban Redevelopment Authority, 2025). Modelling therefore demands a thorough understanding of each study area's planning regime. This framework applies only where formal planning systems are functioning. Furthermore, it is most suitable where LU plans are given greater weight in the planning process and "other material considerations" do not play a dominant role. The second is capacity constraints. Parcel size, the share of developable land, and development history further constrain availability. Land shapes and sizes often differ by use. For example, residential lots are typically smaller than industrial sites. Large parcels may contain undevelopable segments, such as waterbodies, which require recalculation of net developable land. Finally, parcels already fully built are limited to redevelopment only and are thus considered saturated and excluded from the active agent pool.

As discussed in Section 2.3.2, digitizing legislated permissions remains a practical obstacle. Defining capacity constraints also requires flexibility tailored to case-specific contexts. Factors such as study-area size, spatial resolution, study-period length, temporal interval selection, and data availability all influence how availability constraints are operationalized. Although challenging, the feasibility of using AI-driven approaches to help build the dataset of policy constraints has been demonstrated in Zhang (2026). Specifically, parcel-level ZBL shapefiles in municipalities without digital ZBLs can be generated probabilistically by leveraging training data from municipalities with available digital ZBLs, thereby completing the regional ZBL shapefile in a systematic way. Additionally, a TensorFlow Keras classifier has been developed to build the capacity

constraint database based on remote sensing images and multi-source LU data as ground truth. With the rapid progress of AI-driven technologies, more feasible approaches will emerge to extract valuable information from text-based data sources, such as ZBL documents, and more automated text-to-digital conversion tools will be developed to support more nuanced computational LU studies.

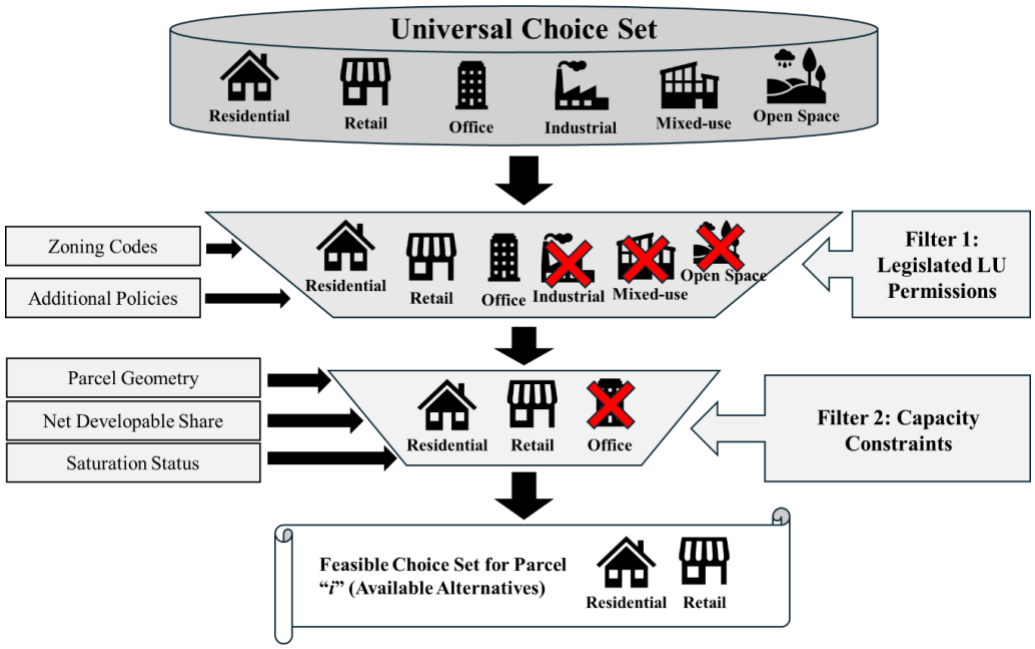


Figure 4. Apply availability rules to define the feasible choice set for each agent

3.3 Transferability and expected "IDEAL" data infrastructure

Model transferability has become an increasing concern in the urban modelling community (Rosier et al., 2022). This raises a practical question: what does it take to apply the framework to a new region? As discussed, the behavioral engine is adaptable to several computational approaches, whether econometric or AI-driven, so the choice of method is not itself a constraint. The main concern is instead assembling the right data. Four kinds of input are required as a baseline. Parcel geometries serve as the decision units, and a time-stamped record captures what was built, where, and when. The land-use permissions attached to each parcel define the feasible choice sets, while a set of explanatory variables represents the candidate drivers of LU supply. The model is then estimated on these local data, as in any new application. Among these inputs, parcel geometries and most explanatory variables are commonly available from open or governmental sources or are already widely used in urban research (e.g., Gharaibeh et al., 2020; Huang et al., 2024). Two inputs, however, are generally harder to obtain.

The first is parcel-level LU permissions. This framework is best applied to study areas where digital policy data are available. Because such data are typically owned and managed by government agencies, acquiring them may require active engagement with the relevant authorities. Where this is not feasible, a coarser definition of permissions can serve as a fallback, at the cost of some spatial precision. For example, commercial and residential development might be permitted in a dense urban core, while industrial development is excluded. The second is comparable market data, such as time-series prices, vacancy, and leasing activity, which are often proprietary and costly to obtain even for academic use. Although some LU supply literature have shown that such data

can be acquired in other study areas (e.g., Hunt et al., 2008), this paper acknowledges that they remain a scarce and difficult-to-access source. In summary, these two issues mean that the framework is most readily replicated in large, data-rich cities, typically in developed countries with mature data infrastructure and public agencies willing to collaborate and share data.

More fundamentally, data constitute the foundation of computational urban models. Beyond the two working components introduced above, the framework therefore also defines the data quality that a more fully developed version would ideally require. This is captured by an "IDEAL" dataset standard (Figure 2): the data should be inclusive across a broad spectrum of urban domains, dynamic in capturing temporal progressions, evolvable to derive new insights through data fusion, adaptable to various disciplinary applications, and large-scale in spatial and temporal coverage. These characteristics should be treated as a future benchmark rather than a strict requirement for applying the present framework. This is also why the framework aims to remain as operational as possible under limited-data conditions, while benefiting from improvements in data infrastructure over time. The following section demonstrates the framework under such conditions, applying it to commercial LU supply in the GTHA.

4 Case study

4.1 Study scope: Study area, commercial land use, and choice set

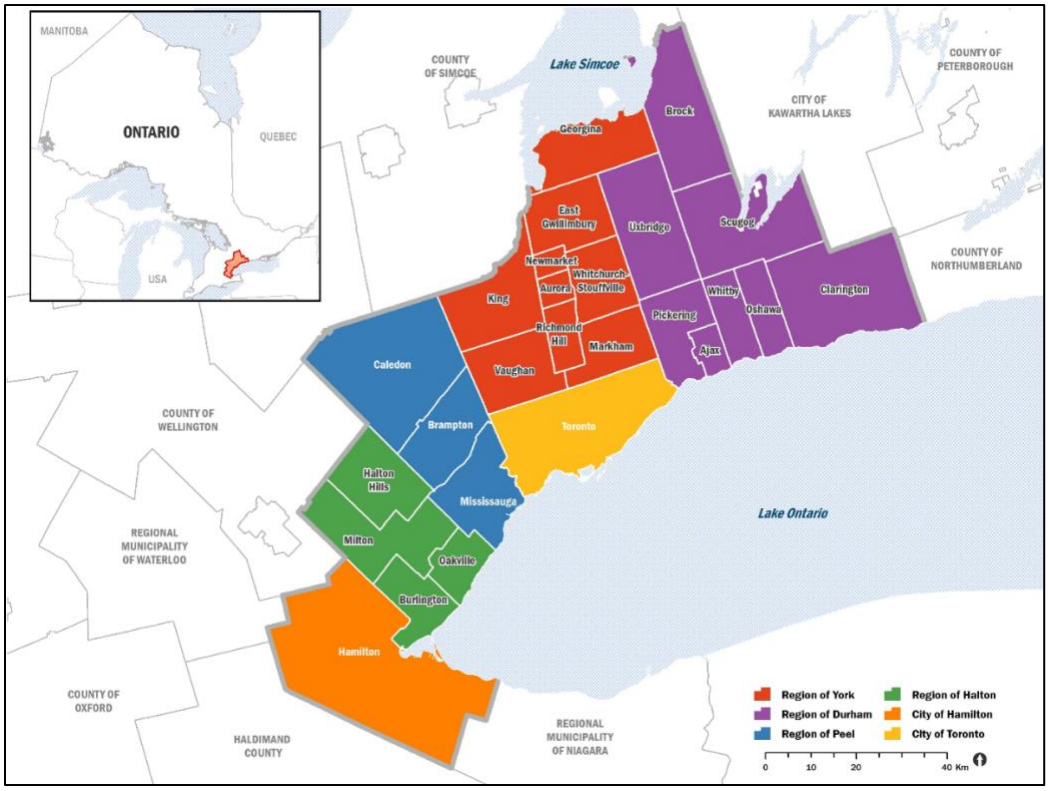


Figure 5. Municipal boundaries and regions in the GTHA (Critchley, n.d.)

The GTHA is Canada's largest metropolitan region (Figure 5). It represents a dense concentration of commercial activity, contributing 21.8% of the national GDP (Statistics Canada, 2021). Empirical analysis (based on DMTI Group's data) indicates that in 2022, it accommodated approximately 20% of all retail, 19.5% of all industrial, and 17.7% of

all office establishments in Canada. The City of Toronto serves as the core urban area, containing 47.5% of retail, 38.4% of industrial, and 50.1% of office locations (based on CoStar Group's data). Additionally, the GTHA supported over a quarter (25.9%) of Canadian employment in these sectors (Statistics Canada, 2025a, 2025b). Overall, it offers a rigorous environment to demonstrate the practicability of the proposed framework.

The case study focuses on commercial LU supply. Commercial LU includes retail, industrial, and office properties, which collectively form the majority of the non-residential urban fabric. Studies on residential LU supply in the GTHA can be found in Zhang and Miller (2025). Existing literature in the GTHA frequently relies on coarse resolution, such as "Transportation Analysis Zones" or "Dissemination Areas" (Farooq & Miller, 2012), but this study adopts the provincial assessment parcel as the spatial unit. Empirical analysis shows that the median parcel area is approximately 450 m² and that each parcel contains an average of 1.07 buildings. The parcel level therefore captures the micro-scale urban fabric well.

Each parcel's baseline choice set contains Retail, Industrial, Office, Mixed, and None or Other Development (NO) (Figure 6). The "Mixed" category is defined to encompass combinations involving only Retail, Industrial, and Office uses. While this definition theoretically allows for four distinct combinatorial types, they are consolidated into a single "Mixed" alternative. This consolidation helps avoid small classes, mitigate class imbalance, and ensure estimation convergence. Furthermore, although the baseline choice set comprises five categories, every parcel's actual choice set is heterogeneous and dictated by zoning regulations (City of Toronto, 2018). For instance, the "Retail" alternative is removed from a parcel's choice set if it is not permitted by zoning.

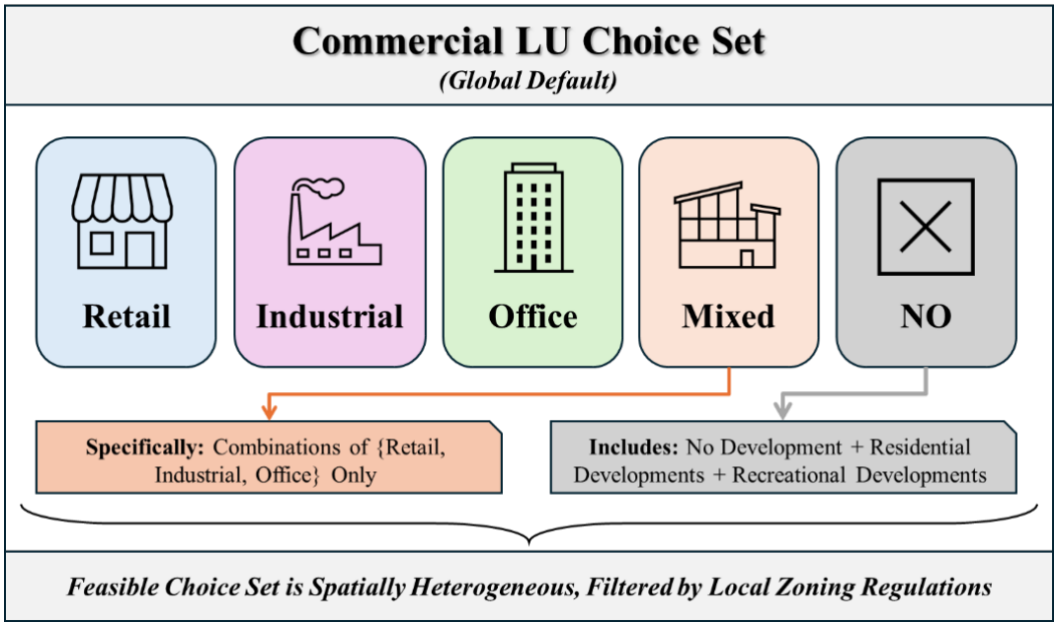


Figure 6. Choice set used by this case study

4.2 Spatiotemporal database

This study spans a nine-year temporal horizon from 2015 to 2023, aligned with the maximum data availability (2013 to 2023). The initial two years of data were reserved for attribute initialization to support investigations of temporal lag. The core commercial LU information (building-level details of building type, precise location, and year built) was

sourced from a leading urban data collection firm (CoStar Group). The dataset is further enriched with longitudinal market-signal indicators, including but not limited to prices, stock availability, and operating costs (Table 2). The raw market-signal data were originally at a meso-scale; a spatial interaction model was applied to downscale relevant metrics to the parcel level[1] based on the assumption that developers across the GTHA are aware of market conditions in other parts of the region. Additionally, the zoning regulations of parcel-level permitted LU were retrieved from Zhang (2026), which systematically produced a GTHA-wide digital zoning dataset. Within this dataset, 17 municipalities' zoning data were obtained from the government and 9 municipalities' zoning was generated using AI-driven approaches (Figure A2 in Appendix). Finally, substantial data from open sources, university libraries, and Scholars GeoPortal (https://geo1.scholarsportal.info/) were collected to generate relevant variables driving LU change (Table 2).

Table 2. The complete list of candidate variables for model tuning

Feature Type	Description	Spatial Scale			Data Source	Data Availability
		1km Buffer	100m Buffer	Parcel		
Shape Indices	Parcel Size			*	Scholars GeoPortal	University Students in Ontario, Canada
	Parcel Perimeter			*		
	Area of Convex Hull			*		
	Circularity (Polsby-Popper score)			*		
	Convexity			*		
	Fractality			*		
	Rectangularity			*		
Accessibility	Distance to the nearest bus stop			*	1) Scholars GeoPortal 2) OpenStreetMap	Partially Open to University Students in Ontario, Canada; Partially Open to All
	Distance to the nearest airport			*	Scholars GeoPortal	University Students in Ontario, Canada
	Distance to the nearest leisure open space			*		
	Distance to the nearest metro station			*	1) Scholars GeoPortal 2) OpenStreetMap	Partially Open to University Students in Ontario, Canada; Partially Open to All
	Distance to the nearest waterbodies			*	Scholars GeoPortal	University Students in Ontario, Canada
	Distance to the nearest wetlands			*		
	Distance to the waterfront shoreline			*		
	Distance to open space			*		
	Auto accessibility to industrial cluster	*			ILUTE Group	Not Open
	Auto accessibility to office cluster	*				
	Auto accessibility to retail cluster	*				
Land Use	Density of the count of office real estates (with lag effects)		*	*	CoStar Group	
	Density of the count of retail real estates (with lag effects)		*	*		
	Density of the count of industrial real estates (with lag effects)		*	*		
	Density of the count of flex real estates (with lag effects)		*	*		
	Density of waterbodies	*		*	Scholars GeoPortal	University Students in Ontario, Canada
	Density of wetlands	*		*		
	Density of open space	*		*		
	Density of lands used for public service	*		*	OpenStreetMap	Open
	# Residential Addresses	*	*	*	Scholars GeoPortal	

1 See Appendix Section 3 for the methodological details.

Feature Type	Description	Spatial Scale			Data Source	Data Availability
		1km Buffer	100m Buffer	Parcel		
	# high-density residential buildings	*	*	*		
	Density of developed lands	*	*			
	# buildings of all type		*			
	Designated greenbelt region			*		University Students in Ontario, Canada
	the corresponding land cover of the parcel			*		
	the corresponding land use of the parcel			*		
	Land-use mix	*				
	Commercial supply mix	*			CoStar Group	Not Open
Transportation	Density of highways	*	*		Scholars GeoPortal	University Students in Ontario, Canada
	Density of railways	*	*			
	Density of major and local roads	*	*			
	Density of trails	*	*			
	# of transit stops	*	*		1) Scholars GeoPortal 2) OpenStreetMap	Partially Open to University Students in Ontario Canada; Partially Open to All
Point of Interests	# of Agriculture, Forestry, and Fishing EPOIs	*	*		Scholars GeoPortal	University Students in Ontario, Canada
	# of Mining EPOIs	*	*			
	# of Construction EPOIs	*	*			
	# of Manufacturing EPOIs	*	*			
	# of Transportation EPOIs	*	*			
	# of Wholesale EPOIs	*	*			
	# of Retail EPOIs	*	*			
	# of Finance EPOIs	*	*			
	# of Service EPOIs	*	*			
	# of Public Administration EPOIs	*	*			
	# Business Establishments	*	*			
Slope	Average slope indices	*				
	Max slope indices	*				
	Min slope indices	*				
Socio-Economic	Population density	*			Transportation Tomorrow Survey	Not Open
	Density of registered vehicle	*				
	Density of Employments	*				
	Density of Houses	*				
	Density of Apartments	*				
	Density of Townhouse	*				
	Market Signals (Rent/Sale Price, Cap Rate, # Sales/Leasing Deals, Cost)	*			CoStar Group	Not Open
	Change of Market Signals	*				
Others	Whether locates within the agricultural ecumene			*	Scholars GeoPortal	University Students in Ontario, Canada
	Whether and what type of population ecumene			*		
	Whether and what class of population ecumene			*		
	Whether locates within Commercial/Industrial/Retail/Residential Zone			*		

Figure 7 illustrates the data processing architecture used to integrate multi-source data into a unified modelling database. The process begins with the acquisition of raw spatial and tabular inputs, as summarized in Table 2. These inputs are then cleaned and synthesized through a series of procedures, including the rectification of geometry errors, the geocoding of civic addresses, and the downscaling of meso-level market indicators. Finally, the processed datasets are spatially integrated onto the parcel shapefile, which produces a consolidated database in which multi-source information is consistently attached to the fundamental modelling unit.

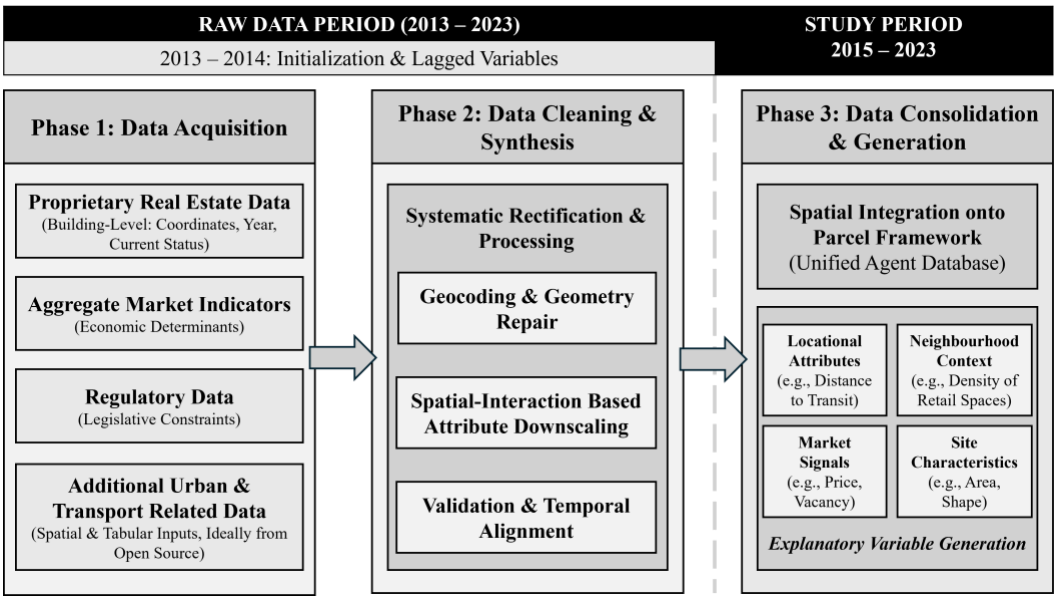


Figure 7. Workflow of data preparation for the case study

Based on this database, a comprehensive set of explanatory attributes was constructed to support an in-depth investigation of the drivers of new LU development. As shown in Table 2, the candidate attributes cover several thematic domains. Locational attributes capture accessibility and proximity conditions, such as distances to the nearest bus stop, metro station, and airport. Neighborhood context variables describe the surrounding urban fabric through aggregate measures such as population density and LU entropy. Market-signal variables represent economic conditions related to demand and supply, including spatial price gradients and sector-specific vacancy rates. Site characteristics further describe the geometric and physical properties of individual parcels, such as parcel area and slope.

This study further examines how the choice of spatial and temporal scales influences development decisions. The parcel-level configuration provides a suitable basis for this purpose, as it allows spatial context and lagged development activity to be measured around each individual parcel. For the selected transportation and urban form attributes, two types of measurements using two buffer distances were computed: 100m (immediate micro-environment) and 1km (broader neighborhood context) (Table 2). Additionally, three temporal lag scales (1-year, 3-year, and 5-year lag) were defined and calculated the corresponding measurements of lagged supply for each alternative for every parcel. These measurements allow the model to examine whether development decisions are more strongly shaped by immediate or broader spatial contexts, and by short-term or longer-term development histories.

4.3 Panel-data mixed logit architecture

Figure 8 illustrates the primary modelling flow in this case study. Following the conceptual framework, each land parcel, as the decision-maker, chooses the alternative that maximizes its utility, where utility is quantified by a set of factors that drive the decision of land development (Equation 1). At each timestep, if the parcel's decision differs from NO, the model evaluates whether the parcel has been fully developed. If fully developed, the parcel exits the model and is excluded from subsequent simulation. If not, it remains eligible and proceeds to participate in the next decision cycle. Empirical

observations show that most parcels "choose" NO in the majority of timesteps and thus develop no commercial use and stay in the market.

In fine-resolution spatiotemporal modelling, it is highly likely that the correlation in unobserved error terms is greater between decisions made by the same agent than between decisions made by different agents. This heightened correlation violates the fundamental assumption of independence and identical distribution (i.i.d.) of errors in standard Multinomial Logit Models (MNL). In addition, it is believed that there is variation in the tastes of decision-makers, and thus in their response to certain attributes of the choice situation. Therefore, this case study employs a Random Parameter Panel-Data Mixed Logit Model, which relaxes the i.i.d. assumption and is thus more appropriate for longitudinal applications (Train, 2009).

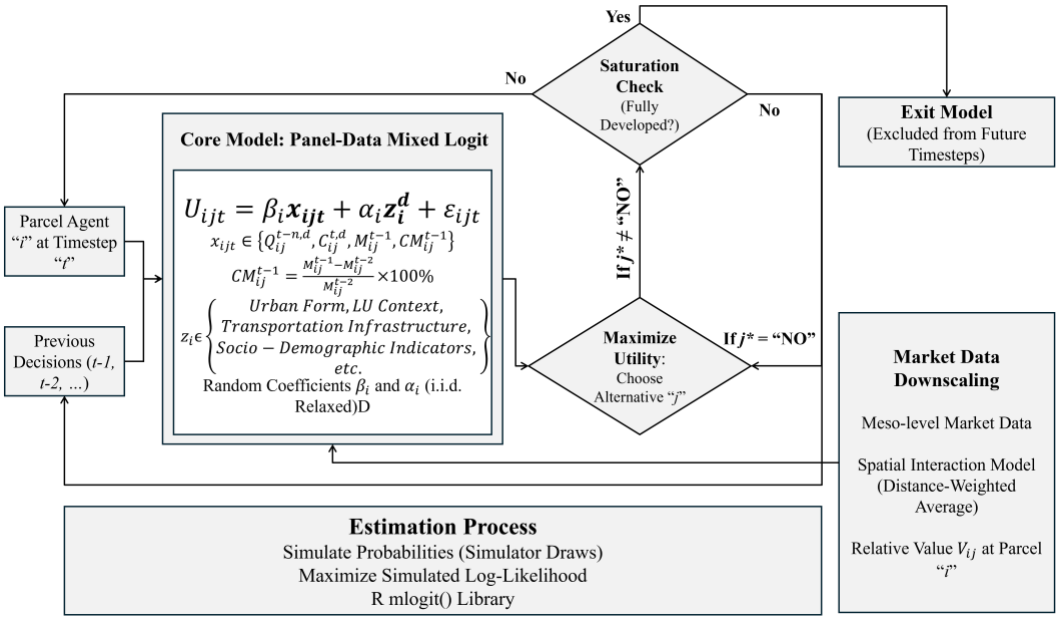


Figure 8. Panel-data mixed-logit model architecture

Essentially, it is assumed that each parcel agent i chooses the alternative j in year t that maximizes its utility U_{ijt}:

$$U_{ijt} = \delta_j + \beta_i x_{ijt} + \alpha_j z_i^d + \varepsilon_{ijt} \quad (1)$$

Where:

- δ_j is the alternative-specific constant of alternative j
- x_{ijt} is the vector of alternative-specific attributes for alternative j at time t for parcel i
- z_i^d is the generic attribute vector for parcel i, which includes measures of urban form, LU context, transportation infrastructure, and socio-economic and demographic indicators, all calculated based on the area within distance d of the parcel
- β_i is the coefficient vector of the alternative-specific attributes, which may vary across parcels

- α_j is the alternative-specific coefficient vector corresponding to generic attributes; δ_{NO} and α_{NO} are normalised to zero, with NO as the reference alternative
- ε_{ijt} is a Type-I Extreme Value (Gumbel) error term

Conditional on β_i, the probability that agent i selects alternative j in year t is:

$$P_{ijt|\beta_i} = \frac{\exp(\delta_j + \beta_i x_{ijt} + \alpha_j z_i^d)}{\sum_{k \in C_{it}} \exp\left(\delta_k + \beta_i x_{ikt} + \alpha_k z_i^d\right)} \tag{2}$$

Where C_{it} is a subset of the universal commercial LU choice set for parcel i in year t.

The unconditional probability of the sequence of choices made by parcel i over its T_i choice occasions is then:

$$P_i = \int \prod_{t=1}^{T_i} \prod_{j \in C_{it}} \left(P_{ijt|\beta_i}\right)^{b_{ijt}} f\left(\beta_i \mid \theta_\beta\right) \mathrm{d}\beta_i \tag{3}$$

Where:

$$b_{ijt} = \begin{cases} 1, if\ parcel\ i\ chose\ alternative\ j\ at\ time\ t \\ 0, otherwise \end{cases} \tag{4}$$

and $f\left(\beta_i \middle| \theta_\beta\right)$ is the density, parameterized by θ_β, from which β_i is drawn.

Simulator draws were applied to approximate P_i. For each parcel i, the model drew R independent samples β_{ir}, where r = 1, …, R, from $f\left(\beta_i \middle| \theta_\beta\right)$, holding each draw fixed across all choice occasions of parcel i, to get the simulated probability as:

$$P_i' = \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} \prod_{j \in C_{it}} \left(P_{ijt|\beta_{ir}}\right)^{b_{ijt}} \tag{5}$$

Finally, the model is estimated by maximizing the simulated log-likelihood (SLL):

$$SLL = \sum_i \ln P_i' \tag{6}$$

See Train (2009) for extensive formulations and estimation procedures of the panel-data mixed logit model. In the final specification, the coefficient of "Change of Sale Price" is random, following a normal distribution, and all other coefficients are fixed. The model was estimated using 200 Halton draws (Bhat, 2001). As a benchmark, a multinomial logit (MNL) model with the same specification was also estimated. The mixed logit improves the LL from -4,914.71 to -4,913.07, which is significant under a likelihood-ratio test (LR = 3.27, p = 0.035, applying the boundary correction for testing a variance at zero). However, it is necessary to further explain the configuration of x_{ijt} in Equation 1 as follows:

$$x_{ijt} \in \left\{Q_{ij}^{t-n,d}, C_{ij}^{t,d}, M_{ij}^{t-1}, CM_{ij}^{t-1}\right\}, CM_{ij}^{t-1} = \frac{M_{ij}^{t-1} - M_{ij}^{t-2}}{M_{ij}^{t-2}} \times 100\% \quad (7)$$

Where:

- Q_{ij}^{t-n} is the quantity of type j projects developed in the nth year prior to t that lie within a distance d of parcel i
- $C_{ij}^{t,d}$ is the cumulative quantity of all type j projects developed up to (but not including) year t that lie within distance d of parcel i
- M_{ij}^{t-1} is the column vector of market attributes (e.g., price, inventory, maintenance cost, total sales) for all type j projects in year t-1 relative to parcel i
- M_{ij}^{t-2} is the equivalent market vector for year t-2
- CM_{ij}^{t-1} is thus the percentage change in the market attributes between t-2 and t-1

The final model specification was determined through iterative attribute screening. In practice, different attribute combinations were tested and evaluated based on model fit, statistical significance, domain knowledge, and the interpretability of each coefficient, following standard procedures in econometric estimation (Train, 2009). Forward selection was used for this process, as it is more sensitive to attribute interrelationships and often yields deeper insights compared to backward selection, despite being more labor-intensive.

Following a comparison of multiple state-of-the-art coding packages (Molloy et al., 2021), R's mlogit (Croissant, 2020) library was selected to operationalize the proposed mixed-logit model. Data preparation and consolidation required numerous specialized spatial analytical tools, such as ArcGIS Pro, GeoPy, and GeoPandas. Access to restricted source datasets is governed by their respective licensing and data-sharing agreements, as described in the Data Availability statement. The model code is openly available and can be retrieved from the project GitHub repository at https://github.com/zha12097/JTLU_LU_Supply_2026.

5 Results and discussion

The final model specification was constructed by screening over 60 candidate attributes (Table 2). The final parameter estimates are presented in Tables 3 and 4. Results for indicators identifying the regional municipality to which a parcel belongs can be found in the Appendix.

5.1 Market dynamics and supply inertia

Results show that over the study period, recent developments in the neighborhood tended to encourage further development (# projects developed 1 year ago; Table 3). In contrast, the cumulative number of projects completed before the start of simulation (one candidate variable listed in Table 2) was not retained during variable selection because it was not statistically significant. There is no direct empirical evidence to indicate that developers actively avoided established agglomerations. One plausible explanation is that traditional commercial centers (downtown areas) have less potential for new development. These areas command higher land values and often require costly redevelopment rather than new construction. Since the current case study is unable to capture redevelopment due to data limitations, this issue requires further investigation.

Table 3. Generic parameter estimates (1-year-lag "primary" model)

Generic Parameters	Estimate	Std. Error	Sig.	Clustered Std. Error	Clustered Sig.
# Projects Developed 1 Year Ago	0.2143	0.0286	***	0.0365	***
Rent Price	0.0644	0.0149	***	0.0259	**
Change of Sale Price	0.0358	0.0090	***	0.0147	**
Change of Cap Rate	1.3962	0.3066	***	0.4714	**
Change in # Leasing Deals	0.0024	0.0008	***	0.0011	**
# Sales Listed	0.0052	0.0013	***	0.0025	*

*Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Clustered standard errors are clustered by CoStar submarket with t-tests on 13 degrees of freedom. Change of Sale Price is the random parameter (normally distributed; 200 Halton draws).*

Rent or sale price serves as a proxy for expected revenues, where higher values increase the incentive for development. Table 3 confirms that higher rent prices stimulate development activity, whereas the "level" of sale prices was not retained because it was not statistically significant. Instead, "changes" in sale price (price growth inertia) positively influence new supply, which raises at least two conclusions. First, although prior studies lacking sale price data used rent price as a proxy for sale price (Farooq, 2011), the two metrics capture distinct demand-side influences and should be treated separately when both are available. Second, when the level and change of the same market attribute were tested, the change was retained for sale price and cap rate, whereas only rent was retained as a level. This indicates that developers respond more consistently to price dynamics than to static price levels. Therefore, greater attention to market inertia is recommended, which is rarely addressed in the literature. The predominance of significant inertia-driven attributes over contemporaneous market levels (Table 3) further supports this argument.

Market levels describe the prevailing market conditions around each parcel, whereas market inertia captures their longitudinal dynamics. Our analysis indicates that developers prioritize future-oriented market trends. Locations with rising net returns on investment (cap rates; Table 3) and consistently increasing leasing activity (# leasing deals; Table 3) attract more development. A "hot" market is therefore interpreted as one experiencing sustained growth rather than one with high transaction volumes in a single year.

To assess the robustness of the market-signal findings, we conducted three additional diagnostic analyses. First, because the market signals vary at the submarket-by-year level, Table 3 also reports cluster-robust standard errors. All generic parameters remain significant at the 5% level, except "# Sales Listed" (at the 10% level). This suggests that the main market-signal findings are not driven by within-submarket correlation. Second, we examined multicollinearity among all numeric explanatory variables included in the final model, with particular attention to the market-level and market-change variables. The full variance inflation factor (VIF) results are reported in Table A1 (Appendix). All variables have VIF values below 10, a commonly used threshold for identifying problematic multicollinearity (O'brien, 2007). This suggests that the distinction between static market levels and market-change variables is not driven by severe collinearity. Third, we evaluated the sensitivity of the model results to alternative spatial downscaling specifications for the price variables. In addition to the inverse-squared distance-weighting approach used in the main specification, we tested linear and exponential distance-decay functions, including multiple parameter values for the exponential specification (Östh et al., 2018). The full sensitivity results are provided in Table A2

(Appendix). Across these alternative downscaling methods, the direction of the market-signal effects remains consistent, although the level of statistical significance varies for some variables (Table A2). Overall, the reported effects are reasonably robust to changes in the construction of the downscaled price measures.

5.2 Inter-sectoral behavioral heterogeneity

Our model is capable of revealing distinct behavioral characteristics among developers (Table 4). Each development type exhibits a particular locational logic and reflects differentiated strategies.

Table 4. Alternative-specific parameter estimation results (1-year-lag model)

Alternative-Specific Parameters	Retail		Industrial		Office		Mixed	
	Estimate	Sig.	Estimate.	Sig.	Estimate	Sig.	Estimate	Sig.
Parcel Size	1.0872	*	3.0647	***	-2.3030		3.0535	
Commercial Supply Mix	-0.3472		0.5097		0.9912	**	0.5045	
# Business Establishments	-0.0288		-0.5174	***	-0.0608	**	-0.0735	
# Residential Addresses	-0.1283		1.9811	**	-0.5796		-3.2175	*
# Buildings of All Type (100m Buffer)	0.0028	**	0.0042		0.0011		-0.0068	
Density of Developed Lands	0.0040		0.0039		0.0187	***	-0.0086	
Density of Open Space	-0.0701		-0.3531	**	-0.1892		-0.0637	
Density of Major Local Roads	-0.0116		-0.0398	***	-0.0088		-0.0005	
Density of Railways	-0.0239	***	-0.0134		0.0072		0.0116	
Density of Railways (100m Buffer)	-0.1386		0.0679		-0.3428	**	0.2335	
Auto Accessibility to Retail Cluster	8.5777		8.4793		-56.3533	***	5.0382	
Auto Accessibility to Industry Cluster	-14.3330	**	2.5430		-27.8171	***	-4.9684	
Auto Accessibility to Office Cluster	-0.5678		-10.7733	**	4.3602		13.3163	
Distance to the Nearest Bus Stop	-0.2179	***	-0.0552		-0.1436		-0.1578	
Distance to the Nearest Metro Station	-0.0627	**	-0.0539		-0.0790		-0.1379	
Distance to the Nearest Airport	0.0005		-0.0172		-0.0467	**	-0.0024	
Distance to Open Space	-0.4751	***	0.1184		-0.3900		-0.1935	
Commercial Zone	0.9576	***	0.5085	*	1.3599	***	1.5248	**
Industrial Zone	-0.1531		1.2471	***	0.4944	**	0.9945	**
Retail Zone	1.6892	***	-0.2604		0.4060		0.7695	*
Residential Zone	1.1612	***	0.7002		1.0379	***	0.3660	

*Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Alternative-specific constants and regional dummies are reported in Table A3 in the Appendix.*

5.2.1 Locally responsive developer (retail)

Retail developers favor larger parcels and areas containing a dense concentration of nearby buildings within a 100 m buffer, which suggests sensitivity to surrounding activity and pedestrian flows. Their developments most often occur within commercial and residential concentrated zones. This highlights the importance of proximity to consumers. Meanwhile, they tend to avoid zones dominated by heavy industrial use (negative coefficient of density of railways and accessibility to industry clusters). Collectively,

results show retail developers as pragmatic actors targeting moderately dense, transit-accessible neighborhoods (negative coefficients of distance to the nearest bus stop and metro station). They prioritize convenience and local accessibility and seek opportunities to fill gaps along partially developed commercial-residential edges.

5.2.2 Agglomeration-driven developer (industrial and mixed)

Industrial developers show the largest positive response to parcel size, confirming a strong preference for large, contiguous spaces suitable for low-density structures and logistics operations. Negative parameters for business-establishment density and open-space density suggest that industrial developers avoid dense business areas and open-space settings. Instead, industrial projects align closely with established industrially zoned land. This pattern emphasizes path dependence and agglomeration by extending traditional industrial belts outward into adjacent but undeveloped parcels. Mixed developments tend to avoid areas with dense residential addresses (negative coefficient of "# Residential Addresses"). They also exhibit an orientation towards established commercial and industrial zones. Given the small number of Mixed developments, these patterns should be interpreted with caution.

5.2.3 Balance-driven developer (office)

Office developers respond positively to commercial, residential, and industrial zones, higher densities of developed land, and a more diverse commercial supply mix. This reflects a preference for established, built-up environments shared with other urban uses. Meanwhile, office development is negatively associated with business-establishment density and with accessibility to retail and industrial clusters, which suggests that office projects favor nodes distinct from the main concentrations of other commercial activity. Relative to Hamilton, office development is more likely in Toronto Downtown and Midtown and in York South (Table A3). Office developers therefore balance mixed, built-up urban settings with separation from retail and industrial clusters, in both the established core and major suburban employment nodes.

5.3 Scale dependency and sensitivity

5.3.1 Spatial scale

Attributes computed at different buffer distances allow comparison of the model's sensitivity to near- versus far-neighbor effects. In this study, all variables not explicitly labelled as "100 m-buffer" represent the coarser "1 km-buffer" context (Table 4). Although most neighborhood attributes were considered at both scales (Table 2), the 1 km measures were retained in most cases, suggesting that commercial developers are more responsive to the broader 1 km neighborhood than to immediate micro-surroundings. However, this aggregate tendency masks important variations across sectors. Retail and office projects display sensitivity to short-range characteristics, whereas industrial development responds more consistently to macro-level spatial structure. Although most attributes respond only at the macro-scale, there still exists one (density of railway) that is significant at both scales. These findings highlight the persistent challenge of scale choice in spatial modelling. The selection of analytical units could shift both coefficient magnitude and significance, which suggests that "optimal" scale is not universal but case specific. High-resolution frameworks, such as the one proposed here, therefore provide a means to assess this sensitivity systematically rather than rely on arbitrary modeler judgement.

5.3.2 Temporal scale

Temporal lag analysis further clarifies how recent versus older developments shape current supply decisions. Table 5 compares model responses using one-, three-, and five-year lag configurations. The three models estimated over 2018 to 2023 use the same data window[2], which allows the effect of lag length to be compared directly. The two one-year-lag models further provide a comparison between the full 2015 to 2023 period and the shorter 2018 to 2023 window.

The number of projects developed in the previous year shows a strong and significant positive effect, reflecting clear short-term clustering of adjacent developments. As the lag increases, both coefficient magnitude and significance decline, which indicates a temporal decay in influence. As the lagged development counts are positively correlated in our data, longer-lag coefficients may partly absorb the effect of more recent development, suggesting that the observed decline is likely to be conservative. While daily mobility processes adapt over short intervals, urban LU change unfolds over much longer horizons. The observed positive-decay pattern suggests that short- and medium-term clustering effects dominate, but very long-term dynamics remain uncertain: 1) the lag impact could either asymptotically approach zero or 2) eventually reverse sign if built-out saturation discourages new projects. Because the study horizon (due to data availability) limits testing beyond five years, future research using extended datasets or continuous-time simulation is needed to characterize the full temporal-decay function.

Table 5. Comparison of market-signal estimates across estimation windows and 1-year, 3-year, and 5-year lag settings

	1-Year-Lag, 2015-2023			1-Year-Lag, 2018-2023			3-Year-Lag, 2018-2023			5-Year-Lag, 2018-2023		
Variables	**Est.**	**Std. Err**	**Sig.**	**Est.**	**Std. Err**	**Sig.**	**Est.**	**Std. Err**	**Sig.**	**Est.**	**Std. Err**	**Sig.**
# Projects Developed 1/3/5 Years Ago	0.2143	0.0286	***	0.1965	0.0432	***	0.0407	0.0233	*	0.0066	0.0162	
Rent Price	0.0644	0.0149	***	0.0482	0.0210	**	0.0510	0.0209	**	0.0498	0.0210	**
Change of Sale Price (mean)	0.0358	0.0090	***	0.0663	0.0177	***	0.0621	0.0177	***	0.0624	0.0177	***
Change of Sale Price (S.D.)	0.0279	0.0104	***	0.0287	0.0150	*	0.0339	0.0134	**	0.0360	0.0130	***
Change of Cap Rate	1.3962	0.3066	***	2.9615	0.5571	***	3.0949	0.5603	***	3.1654	0.5607	***
Change in # Leasing Deals	0.0024	0.0008	***	0.0044	0.0012	***	0.0045	0.0012	***	0.0046	0.0012	***
# Sales Listed	0.0052	0.0013	***	0.0026	0.0034		0.0022	0.0034		0.0022	0.0034	

*Notes: *** p < 0.01; ** p < 0.05; * p < 0.10. All four models share the same specification (Tables A3 to A6); Change of Sale Price is the random parameter (normally distributed; 200 Halton draws). The 2015-2023 and 2018-2023 one-year-lag models differ only in estimation window; the three 2018-2023 models differ only in lag length. The 1-Year-Lag model for 2015–2023 is based on 2,059 parcels (13,469 choice occasions) and has a log-likelihood of −4,913.1. The three models for 2018–2023 are each based on 1,551 parcels (7,739 choice occasions), with log-likelihoods of −2,996.6, −3,004.8, and −3,006.4 for the 1-Year, 3-Year, and 5-Year lag settings, respectively.*

The estimated S.D. of Change of Sale Price also varies across estimation windows and lag settings (Table 5), indicating that the estimated heterogeneity is sensitive to temporal settings. These results also highlight an important issue in the behavioral interpretation of LU models. Although the directions of the major market-signal effects remain broadly

2 2018 being the first year in which a five-year lag can be computed from data beginning in 2013.

consistent, their magnitudes and estimated heterogeneity may vary across lag settings and between the two estimation periods. Therefore, parameters estimated from one temporal setting deserve more cautious interpretations when applied to another period. This is particularly relevant for long-term urban modelling, where behavioral parameters may be applied beyond the period from which they were estimated.

Collectively, these sensitivity analyses suggest that developer behavior is scale-dependent, and that both estimated effects and heterogeneity can vary across spatial and temporal specifications. Recognizing and explicitly modelling these sensitivities is therefore essential for accurate, transferable urban LU studies.

5.4 Validation of simulation performance

Econometric discrete-choice models lack a universally mandated evaluation protocol. This paper follows several commonly used evaluation methods derived from two highly cited sources: Train (2009) and McFadden (1977). Specifically, the final model specification yielded a log-likelihood of -4,913.07 and a McFadden-R2 (ρ_c^2) of 0.194, with a statistically significant likelihood-ratio test ($\chi 2 = 2{,}366.5$, $p < 0.001$). This result indicates a significant improvement in log-likelihood over the null model and is consistent with a reasonable fit for a discrete-choice model of this kind (McFadden, 1977). Additionally, the null log-likelihood underlying the reported ρ_c^2 is calculated from the observed aggregate choice shares rather than equal shares across alternatives. Therefore, the dominance of the NO alternative has already been reflected in this benchmark.

Table 6. Probability-weighted prediction-success matrix, in-sample (A) vs. out-of-sample (B)

6A: Probability-weighted prediction-success matrix, in-sample (1-year-lag model, 2015 to 2023)

			Average Predicted Probability (%)						
Observed Choice	**Observation Count (N)**	**Observed Share (%)**	**NO**	**Retail**	**Industrial**	**Office**	**Mixed**	**Prediction-Success Index (PSI)**	**Normalized PSI (PSI*)**
NO	12,079	89.68	90.02	4.70	3.09	1.61	0.58	0.0050	0.0483
Retail	672	4.99	84.00	13.66	0.63	1.34	0.37	0.0855	0.0900
Industrial	416	3.09	87.00	1.15	10.15	0.96	0.73	0.0674	0.0696
Office	224	1.66	85.86	4.21	2.13	6.99	0.82	0.0530	0.0539
Mixed	78	0.58	83.96	4.60	6.36	2.79	2.30	0.0168	0.0169
All Choice Occasions	13,469	100.00	89.52	5.03	3.19	1.67	0.59	–	–

Notes: PSI follows McFadden (1977). PSI = PSI / (1 − observed share), i.e., PSI relative to its maximum attainable value, which makes alternatives with very different base rates comparable. ρ² is 0.194*

6B: Probability-weighted prediction-success matrix, out-of-sample (1-year-lag model, 2015 to 2023)

			Average Predicted Probability (%)						
Observed Choice	**Observation Count (N)**	**Observed Share (%)**	**NO**	**Retail**	**Industrial**	**Office**	**Mixed**	**Prediction-Success Index (PSI)**	**Normalized PSI (PSI*)**
NO	12,066	89.68	89.79	4.76	3.19	1.64	0.63	0.0045	0.0436
Retail	672	4.99	84.37	13.15	0.67	1.39	0.42	0.0798	0.0840
Industrial	416	3.09	87.20	1.17	9.82	1.00	0.81	0.0617	0.0637
Office	223	1.66	85.82	4.51	2.26	6.55	0.86	0.0476	0.0484
Mixed	77	0.57	84.24	4.76	6.63	2.78	1.59	0.0087	0.0087
All Choice Occasions	13,454	100.00	89.34	5.06	3.28	1.69	0.63	–	–

Notes: PSI follows McFadden (1977). PSI = PSI / (1 − observed share), i.e., PSI relative to its maximum attainable value, which makes alternatives with very different base rates comparable. Two parcels, each the only case of its development type within a market-area dummy, were kept in every training set, since holding them out would leave that coefficient unidentified; their 15 choice occasions are therefore excluded from 6B. ρ² is 0.158*

However, as a likelihood-based measure of model fit, the ρ_c^2 does not directly reflect how predicted probabilities are distributed across observed choices. Train (2009) explicitly advises against the percent-correctly-predicted measure for discrete-choice models, because these models generate a probability for each feasible alternative rather than a single deterministic prediction. Accordingly, this paper adopts the "prediction-success matrix" developed by McFadden (1977). As presented in Table 6, the matrix should be interpreted as a probability matrix rather than a conventional classification table in which each observation is assigned a single predicted choice. Table 6 reports both in-sample evaluation results and out-of-sample performance from a five-fold parcel-level holdout test. Parcels were assigned to five folds, with each parcel's complete 2015–2023 sequence retained within a single fold, and each held-out fold was predicted using the model estimated on the remaining four folds.

For example, in the out-of-sample results, among the 672 choice occasions in which Retail development was observed, the model assigned Retail an average predicted probability of 13.15%, compared with its aggregate predicted probability of 5.06% across all choice occasions. Although NO retains the highest average probability, the probability assigned to Retail when Retail is actually observed is therefore more than twice its overall predicted probability. A similar pattern is observed across all four active development alternatives. Therefore, the model systematically assigns a higher probability to an alternative when it is actually observed. The PSI results (Equation A6) further support this conclusion (McFadden, 1977). All alternatives have positive PSI values in both the in-sample and out-of-sample evaluations, while Retail, Industrial, and Office also achieve higher normalized PSI values (PSI*; Equation A7) than NO (Table 6). This indicates that the model contains discriminatory information for the relatively infrequent development outcomes rather than merely reproducing the dominant NO outcome. Mixed is the weakest-performing active development alternative, which is consistent with the limited number of observed Mixed developments.

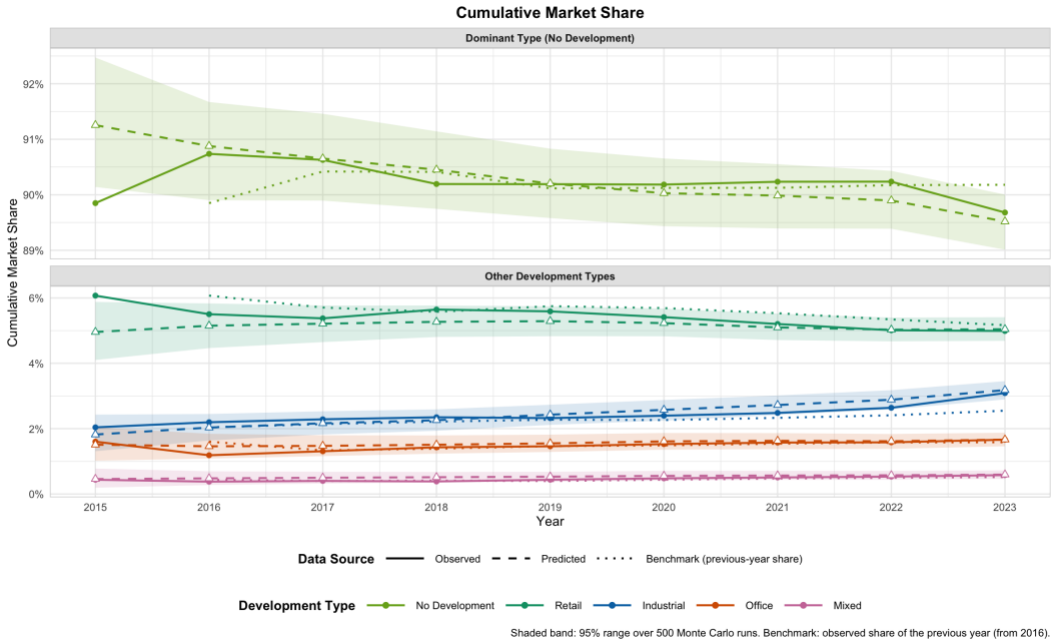


Figure 9. Longitudinal comparison of cumulative market share by development type in the GTHA (2015–2023). Upper section: the "No Development" category; lower section: Active Commercial Development Types. Shaded bands show the 95% range over 500 Monte Carlo runs; the benchmark is the observed share of the previous year (results generated from the 1 year-lag model).

To further assess the model's reliability, commercial development trajectories over the 2015 to 2023 study period were simulated. For each year, the simulation takes the observed lagged supply, market conditions, and choice sets as given and draws one alternative for each choice occasion according to the model-predicted probabilities. This procedure is repeated 500 times to produce a mean simulated share and a 95% simulation interval for each development type. As a naive benchmark, the observed share of each development type in the previous year is used. This benchmark, like the simulation, relies only on information available up to the previous year. Figure 9 shows that the major discrepancy occurs at the beginning of the simulation. In 2015, the model predicts a higher cumulative share of NO and a lower share of Retail than observed. The gap, however, narrows in the following years. From 2016 onward, the observed cumulative shares of all development types fall within the 95% simulation intervals, and by 2023 the simulated and observed shares are close. Over 2016-2023, the simulation also performs better than the naive benchmark, reducing Root-Mean-Square Error (RMSE) by 33% for cumulative shares (0.17 compared with 0.26) and by 51% for annual shares (0.75 compared with 1.51). These results suggest that the model captures the overall evolution of development shares over the study period reasonably well.

6 Limitation, future direction and conclusion

6.1 Data quality constraints

Despite the illustrative insights provided by this application, several key limitations derived from data quality must be acknowledged that constrains both model specification and interpretation. Specifically, reliable information on actual commercial floorspace was lacking and therefore the simplified assumption was adopted that each commercial-use building corresponds to a single unit of supply. Although this proxy is common in urban

and GIS research, it masks heterogeneity in building size, height, and functional intensity. To achieve a more holistic understanding of urban development, future work should integrate enriched built-form data (floorspace) rather than relying solely on building counts.

Data limitations also prevented incorporating residential LU. Although housing data are generally more accessible than commercial datasets, it remains challenging to obtain residential records that simultaneously offer high spatial precision, consistent temporal coverage, and accurate geocoded coordinates with built-year metadata for a large-scale context. These attributes should also be aligned with the retail, industrial, and office data. In practice, researchers can usually find such detailed data for a single LU type, but assembling a unified dataset covering all major LU categories with adequate quality is challenging.

Additionally, the available data did not contain information on which parcels shared a single developer. This made it impossible to incorporate correlations in the development choices on parcels sharing a developer, which ultimately affected the model's goodness of fit and explanatory power. Future research should seek to identify potential databases with this information and integrate them into the flexible framework that has been developed here, and whose panel data mixed model would be easily able to handle this extension.

Another limitation concerns the zoning data used to define the feasible choice sets. Some municipalities lacked official digital zoning data, and permitted uses were therefore generated using AI-driven approaches (Zhang, 2026). Although this dataset was systematically produced and documented, misclassified permitted uses could lead to inaccurate feasible choice sets for some parcels. As official digital zoning data become available, future research could update the feasible choice sets and assess the sensitivity of the results to such misclassification. Moreover, developments are dated by their year of completion rather than the year in which the development decision was made. Because commercial construction can span more than one year, the short-term lag effects reported here may partly reflect the development pipeline. Similarly, although market signals enter the model with a one-year lag, new supply may also influence local market conditions, so the estimated market-signal effects should be interpreted as associations rather than causal effects.

Finally, redevelopment could not be modelled in this study given data limitations. While a comprehensive case study is presented targeting new development, future research should focus on demonstrating a redevelopment-inclusive case study to complete the figure. In summary, to address these gaps, public agencies at all levels are strongly encouraged to invest in systematic collection and open dissemination of high-quality LU data. Private data providers are similarly urged to facilitate access for non-commercial scientific studies through collaborative partnerships.

6.2 AI-empowered future: Limitations and potentials

The boom in AI offers new opportunities to enhance the efficiency of the framework. In North America, as well as in many Commonwealth countries, municipalities routinely publish their zoning by-law documents online which provides useful source information. The primary challenge lies in converting these text-based regulations into structured, digitized geospatial data for modelling and GIS analysis. Large Language Models present a promising solution by automating the extraction and geocoding of zoning rules, thereby reducing labor-intensive manual digitization. Similarly, developments in image pattern recognition enable the efficient generation of time-series LU datasets from satellite or drone imagery (Li et al., 2016). By integrating cutting-edge AI tools, this research lays

the groundwork for urban studies with high spatial resolution, and we anticipate that continued contributions from the scientific community will drive this vision forward.

6.3 Concluding remarks

In summary, this paper developed and demonstrated a metropolitan-scale, parcel-level framework for LU supply modelling that integrates behaviorally interpretable decision rules with policy-based availability constraints. The GTHA case study confirms that this integration is both feasible and informative under limited data conditions. Three substantive findings carry implications beyond the case study per se. First, developers respond more consistently to market dynamics than to static market levels, which suggests that LU supply modelling should incorporate longitudinal market trends in addition to cross-sectional snapshots. Second, commercial development is not a monolithic process. Retail, industrial, and office developers follow fundamentally distinct locational logics, which highlights the need for sector-explicit modelling rather than consolidated "commercial" categories. Third, both spatial and temporal scale choices affect parameter estimates and significance, which reinforces that scale sensitivity should be explicitly tested rather than simply assumed.

The framework is designed to be transferable beyond the GTHA and beyond commercial LU. Its modular architecture accommodates different data environments, LU taxonomies, and behavioral engines. In the future, improvements in data infrastructure, particularly AI-assisted digitization of planning regulations and satellite-derived LU time series, will progressively lower the barriers to parcel-level modelling in data-scarce regions. Strengthening the supply-side representation at this resolution can contribute to a more robust foundation for next-generation LU models.

Acknowledgments

We acknowledge that this project is funded by the iCity 2.0 Ontario Research Fund – Research Excellence project. We acknowledge our data provider, CoStar, for their generous donations of data and the contributions by Prof. Murtaza Haider for his assistance in helping us access these valuable data. Finally, we disclose that Claude Opus 5.5 was used solely during the drafting process to improve grammar and language; the authors take full responsibility for the final content.

Declaration of interest statement

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of data availability

The source code developed for this study has been made available through an open-access repository. Full independent reproduction of the empirical results is not currently possible without obtaining third-party access to the CoStar dataset and, for some municipalities, the relevant zoning layers. The commercial real estate data were obtained from CoStar Canada under a restrictive licensing agreement and therefore cannot be redistributed by the authors. The other licensed inputs listed in Table 2 (e.g., Scholars GeoPortal, ILUTE Group, and TTS data) and the DMTI data used in Section 4.1 are subject to the terms of their respective providers. Some municipal zoning layers are also subject to data-sharing restrictions, while others are publicly available. Upon reasonable request, the authors will assist researchers in coordinating access to restricted municipal

zoning data and will share model-ready zoning data derived from publicly available sources without additional restrictions.

Author contribution

The authors confirm their contribution to the paper as follows: Conceptualization: J. Zhang; Data Curation: J. Zhang, E. J. Miller; Methodology: J. Zhang, M. Kiko; Investigation, Formal Analysis, Coding: J. Zhang, M. Kiko; Funding Acquisition, Resources, and Supervision: E. J. Miller; Writing – Original Draft: J. Zhang; Writing – Review & Editing: J. Zhang, E. J. Miller, M. Kiko.

Appendix

Appendix available as a supplemental file at https://doi.org/10.5198/jtlu.2026.2892.

References

Batty, M. (2013). *The new science of cities*. MIT Press.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, *35*(7), 677–693. https://doi.org/10.1016/S0191-2615(00)00014-X

City of Toronto. (2018). City of Toronto: Zoning by-law (Part 1). https://www.toronto.ca/wp-content/uploads/2018/07/97ec-City-Planning-Zoning-Zoning-By-law-Part-1.pdf

City Planning Division. (2016). Urban land use planning system in Japan. City and Regional Development Bureau, Ministry of Land, Infrastructure and Transport. https://www.mlit.go.jp/common/001050453.pdf

Clarke, K. C., & Gaydos, L. J. (1998). Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science: IJGIS*, *12*(7), 699–714. https://doi.org/10.1080/136588198241617

Coulombel, N. (2010). *Residential choice and household behavior: State of the art.* Ecole Normale Supérieure de Cachan. https://www.sustaincity.ethz.ch/publications/WP_2.2a_Residential_Choice_and_Household_Behavior.pdf

Critchley, J. (n.d.). *GTHA municipalities*. Jonathan Critchley. https://jonathancritchley.ca/gtha.html

Croissant, Y. (2020). Estimation of random utility models in R : The mlogit package. *Journal of Statistical Software*, *95*(11), 1–41. https://doi.org/10.18637/jss.v095.i11

Darrel Jenerette, G., & Potere, D. (2010). Global analysis and simulation of land-use change associated with urbanization. *Landscape Ecology*, *25*(5), 657–670. https://doi.org/10.1007/s10980-010-9457-2

Ding, Y., Cao, K., Qiao, W., Shao, H., Yang, Y., & Li, H. (2022). A whale optimization algorithm–based cellular automata model for urban expansion simulation. *International Journal of Applied Earth Observation and Geoinformation*, *115,* 103093. https://doi.org/10.1016/j.jag.2022.103093

Farooq, B. (2011). Evolution of urban built space: Markets and decisions [Doctoral dissertation, University of Toronto]. University of Toronto TSpace. https://hdl.handle.net/1807/29718

Farooq, B., & Miller, E. J. (2012). Towards integrated land use and transportation: A dynamic disequilibrium based microsimulation framework for built space markets. *Transportation Research Part A: Policy and Practice*, *46*(7), 1030–1053. https://doi.org/10.1016/j.tra.2012.04.005

Farooq, B., Miller, E. J., & Haider, M. (2013). A multidimensional decisions modeling framework for built space supply. *Journal of Transport and Land Use*, *6*(3), 61–74. https://doi.org/10.5198/jtlu.v6i3.350

Feng, Y., & Tong, X. (2018). Dynamic land use change simulation using cellular automata with spatially nonstationary transition rules. *GIScience and Remote Sensing*, *55*(5), 678–698. https://doi.org/10.1080/15481603.2018.1426262

Gharaibeh, A., Shaamala, A., Obeidat, R., & Al-Kofahi, S. (2020). Improving land-use change modeling by integrating ANN with Cellular Automata-Markov Chain model. *Heliyon*, *6*(9), Article e05092. https://doi.org/10.1016/j.heliyon.2020.e05092

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T., & Hokao, K. (2011). Modeling urban land use change by the integration of cellular automaton and Markov model. *Ecological Modelling*, *222*(20), 3761–3772. https://doi.org/10.1016/j.ecolmodel.2011.09.009

Huang, C., Zhou, Y., Wu, T., Zhang, M., & Qiu, Y. (2024). A cellular automata model coupled with partitioning CNN-LSTM and PLUS models for urban land change simulation. *Journal of Environmental Management*, *351*, Article 119828. https://doi.org/10.1016/j.jenvman.2023.119828

Hunt, J. D., Abraham, J. E., de Silva, D., Zhong, M., Bridges, J., & Mysko, J. (2008, January). Developing and applying a parcel-level simulation of developer actions in Baltimore. In *Transportation Research Board 87th Annual Meeting* (No. 08-2593)

Hyandye, C., & Martz, L. W. (2017). A Markovian and cellular automata land-use change predictive model of the Usangu Catchment. *International Journal of Remote Sensing*, *38*(1), 64–81. https://doi.org/10.1080/01431161.2016.1259675

Islam, K., Rahman, Md. F., & Jashimuddin, M. (2018). Modeling land use change using cellular automata and artificial neural network: The case of Chunati Wildlife Sanctuary, Bangladesh. *Ecological Indicators*, *88*, 439–453. https://doi.org/10.1016/j.ecolind.2018.01.047

Jantz, C. A., & Goetz, S. J. (2005). Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science: IJGIS*, *19*(2), 217–241. https://doi.org/10.1080/13658810410001713425

Jjumba, A., & Dragićević, S. (2012). High resolution urban land-use change modeling: Agent iCity approach. *Applied Spatial Analysis and Policy*, *5*(4), 291–315. https://doi.org/10.1007/s12061-011-9071-y

Kausar, A., Zubair, S., Sohail, H., Anwar, M. M., Aziz, A., Vambol, S., Vambol, V., Khan, N. A., Poteriaiko, S., Tyshchenko, V., Murasov, R., Ejaz, F., & Khan, O. I. (2024). Evaluating the challenges and impacts of mixed-use neighborhoods on urban planning: An empirical study of a megacity, Karachi, Pakistan. *Discover Sustainability*, *5*(1), Article 24. https://doi.org/10.1007/s43621-024-00195-5

Ku, C.-A. (2016). Incorporating spatial regression model into cellular automata for simulating land use change. *Applied Geography*, *69*, 1–9. https://doi.org/10.1016/j.apgeog.2016.02.005

Lauf, S., Haase, D., Hostert, P., Lakes, T., & Kleinschmit, B. (2012). Uncovering land-use dynamics driven by human decision-making – A combined model approach using cellular automata and system dynamics. *Environmental Modelling & Software: With Environment Data News*, *27*, 71–82. https://doi.org/10.1016/j.envsoft.2011.09.005

Li, M., Stein, A., Bijker, W., & Zhan, Q. (2016). Urban land use extraction from very high resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, 192–205. https://doi.org/10.1016/j.isprsjprs.2016.10.007

Liang, G., Chen, Z., Su, S., Zhang, S., & Zhang, Y. (2025). A masking, linkage and guidance framework for online class incremental learning. *Pattern Recognition*, *160*, Article 111185. https://doi.org/10.1016/j.patcog.2024.111185

Liang, X., Guan, Q., Clarke, K. C., Liu, S., Wang, B., & Yao, Y. (2021). Understanding the drivers of sustainable land expansion using a patch-generating land use simulation (PLUS) model: A case study in Wuhan, China. *Computers, Environment and Urban Systems*, *85*, Article 101569. https://doi.org/10.1016/j.compenvurbsys.2020.101569

Martinez, F. (1996). MUSSA: Land use model for Santiago City. *Transportation Research Record*, *1552*(1552), 126–134. https://doi.org/10.3141/1552-18

McFadden, D. (1977). *Quantitative methods for analyzing travel behaviour of individuals: Some recent developments* (Discussion paper No. 474). Cowles Foundation for Research in Economics, Yale University. https://elischolar.library.yale.edu/cowles-discussion-paper-series/707

Miller, E. J. (2018). Viewpoint: Integrated urban modeling: Past, present, and future. *Journal of Transport and Land Use*, *11*(1), 387–399. https://doi.org/10.5198/jtlu.2018.1273

Molloy, J., Becker, F., Schmid, B., & Axhausen, K. W. (2021). mixl: An open-source R package for estimating complex choice models on large datasets. *Journal of Choice Modelling*, *39*, Article 100284. https://doi.org/10.1016/j.jocm.2021.100284

Nguyen, N. A., Schweizer, J., & Rupi, F. (2025). Large-scale activity-based demand generation modeling: A literature review and exploration of potential approaches. *Transportation Engineering*, *20*, Article 100329. https://doi.org/10.1016/j.treng.2025.100329

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Östh, J., Reggiani, A., & Nijkamp, P. (2018). Resilience and accessibility of Swedish and Dutch municipalities. *Transportation*, *45*(4), 1051–1073. https://doi.org/10.1007/s11116-017-9854-3

Pan, Y., Roth, A., Yu, Z., & Doluschitz, R. (2010). The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, *34*(5), 400–408. https://doi.org/10.1016/j.compenvurbsys.2010.03.003

Rosier, J. F., Taubenböck, H., Verburg, P. H., & van Vliet, J. (2022). Fusing Earth observation and socioeconomic data to increase the transferability of large-scale urban land use classification. *Remote Sensing of Environment*, *278,* 113076. https://doi.org/10.1016/j.rse.2022.113076

Sajan, B., Mishra, V. N., Kanga, S., Meraj, G., Singh, S. K., & Kumar, P. (2022). Cellular automata-based artificial neural network model for assessing past, present, and future land use/land cover dynamics. *Agronomy*, *12*(11), Article 2772. https://doi.org/10.3390/agronomy12112772

Statistics Canada. (2021). Table 36-10-0468-01: Gross domestic product (GDP) at basic prices, by census metropolitan area (CMA) (x 1,000,000). [Dataset]. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610046801

Statistics Canada. (2025a). Table: 14-10-0202-01 [Dataset]. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410020201

Statistics Canada. (2025b). Table: 14-10-0384-01 [Dataset]. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410038401

Tobler, W. R. (1979). Cellular geography. In S. Gale & G. Olsson (Eds.), *Philosophy in geography* (Theory and decision library, Vol. 20) (pp. 379–386). Springer. https://doi.org/10.1007/978-94-009-9394-5_18

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Urban Redevelopment Authority. (2025). Master plan. Urban Redevelopment Authority. https://www.ura.gov.sg/Corporate/Planning/Master-Plan

Waddell, P., Boeing, G., Gardner, M., & Porter, E. (2018). *An integrated pipeline architecture for modeling urban land use, travel demand, and traffic assignment*. arXiv. https://doi.org/10.48550/arxiv.1802.09335

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics*, *3*(1), 43–67. https://doi.org/10.1023/A:1022049000877

Waddell, P., Ulfarsson, G. F., Franklin, J. P., & Lobb, J. (2007). Incorporating land use in metropolitan transportation planning. *Transportation Research Part A: Policy and Practice*, *41*(5), 382–410. https://doi.org/10.1016/j.tra.2006.09.008

Wang, C., Wang, Y., Wang, R., & Zheng, P. (2018). Modeling and evaluating land-use/land-cover change for urban planning and sustainability: A case study of Dongying city, China. *Journal of Cleaner Production*, *172*, 1529–1534. https://doi.org/10.1016/j.jclepro.2017.10.294

Wang, W., Zhong, M., Zhang, Y., Li, Y., Ma, X., Hunt, J. D., & Abraham, J. E. (2020). Testing microsimulation uncertainty of the parcel-based space development module of the Baltimore PECAS Demo Model. *Journal of Transport and Land Use*, *13*(1), 93–112. https://doi.org/10.5198/jtlu.2020.1454

White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning & Design*, *24*(3), 323–343. https://doi.org/10.1068/b240323

Xing, W., Qian, Y., Guan, X., Yang, T., & Wu, H. (2020). A novel cellular automata model integrated with deep learning for dynamic spatio-temporal land use change simulation. *Computers & Geosciences*, *137,* 104430. https://doi.org/10.1016/j.cageo.2020.104430

Xu, L., Liu, X., Tong, D., Liu, Z., Yin, L., & Zheng, W. (2022). Forecasting urban land use change based on cellular automata and the PLUS model. *Land*, *11*(5), 652. https://doi.org/10.3390/land11050652

Xu, Q., Zhu, A.-X., & Liu, J. (2023). Land-use change modeling with cellular automata using land natural evolution unit. *Catena*, *224,* 106998. https://doi.org/10.1016/j.catena.2023.106998

Xu, X., Du, Z., & Zhang, H. (2016). Integrating the system dynamic and cellular automata models to predict land use and land cover change. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 568–579. https://doi.org/10.1016/j.jag.2016.07.022

Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, *34*(6), 592–602. https://doi.org/10.1016/j.cageo.2007.08.003

Zhang, J. (2026). Decoding urban evolution: A parcel-level blueprint for the ILUTE 2.0 land use module [Doctoral dissertation, University of Toronto]. University of Toronto TSpace. https://hdl.handle.net/1807/153905

Zhang, Y., & Miller, E. J. (2025). Exploring housing supply dynamics and temporal modelling of residential construction activities. *International Journal of Housing Markets and Analysis*, *18*(6), 1522–1552. https://doi.org/10.1108/IJHMA-04-2024-0049

Zhao, L., & Peng, Z.-R. (2012). LandSys: An agent-based cellular automata model of land use change developed for transportation analysis. *Journal of Transport Geography*, *25*, 35–49. https://doi.org/10.1016/j.jtrangeo.2012.07.006

Zheng, H. W., Shen, G. Q., Wang, H., & Hong, J. (2015). Simulating land use change in urban renewal areas: A case study in Hong Kong. *Habitat International*, *46*, 23–34. https://doi.org/10.1016/j.habitatint.2014.10.008

Zhou, Y., Varquez, A. C. G., & Kanda, M. (2019). High-resolution global urban growth projection based on multiple applications of the SLEUTH urban growth model. *Scientific Data*, *6*(1), Article 34. https://doi.org/10.1038/s41597-019-0048-z