Interpretable State-Space Model of Urban Dynamics for Human-Machine Collaborative Transportation Planning

Jiangbo Yu (jiangbo.yu@mcgill.ca**); Michael F. Hyland (**hylandm@uci.edu)

Abstract

Strategic Long-Range Transportation Planning (SLRTP) is pivotal in shaping prosperous, sustainable, and resilient urban futures. Existing SLRTP decision support tools predominantly serve forecasting and evaluative functions, leaving a gap in directly recommending optimal planning decisions. To bridge this gap, we propose an Interpretable State-Space Model (ISSM) that considers the dynamic interactions between transportation infrastructure and the broader urban system. The ISSM directly facilitates the development of optimal controllers and reinforcement learning (RL) agents for optimizing infrastructure investments and urban policies while still allowing human-user comprehension. We carefully examine the mathematical properties of our ISSM; specifically, we present the conditions under which our proposed ISSM is Markovian and a unique and stable solution exists. Then, we apply an ISSM instance to a case study of the San Diego region of California, where a partially observable ISSM represents the urban environment. We also propose and train a Deep RL agent using the ISSM instance representing San Diego. The results show that the proposed ISSM approach, along with the well-trained RL agent, captures the impacts of coordinating the timing of infrastructure investments, environmental impact fees for new land development, and congestion pricing fees. The results also show that the proposed approach facilitates the development of prescriptive capabilities in SLRTP to foster economic growth and limit induced vehicle travel. We view the proposed ISSM approach as a substantial contribution that supports the use of artificial intelligence in urban planning, a domain where planning agencies need rigorous, transparent, and explainable models to justify their actions.

Introduction

Many countries require planning authorities to compose comprehensive long-range transportation plans to receive approval or government funding for infrastructure projects and urban policies. Strategic transportation plans often have a 30- to 40-year planning horizon, and regions may update them every few years (with or without minor amendments between formal updates). Common visions and goals in regional transportation plans include improving sustainability, equity, accessibility, and resilience (Sciara and Handy, 2017).

Figure 1 displays a simplified interrelationship between a planning agency and the urban systems that the planning agency impacts but does not control. This interaction can be seen as a control feedback loop, where the metropolitan area represents the system, and the planning agency acts as the controller, striving to guide the system towards a desirable state or to optimize some performance measure(s).

This paper concentrates on the Strategic Long-Range Transportation Planning (SLRTP) phase, also known as sketch planning. This initial phase involves evaluating SLRTP decisions, such as funding allocations among modes and overall congestion management policies, laying the groundwork for more detailed regional transportation system planning in later phases. However, this process is susceptible to the mental models, conventions, and recent events that influence planners, analysts, researchers, and policymakers, leading to potential implicit cognitive biases and non-transparent decision-making.

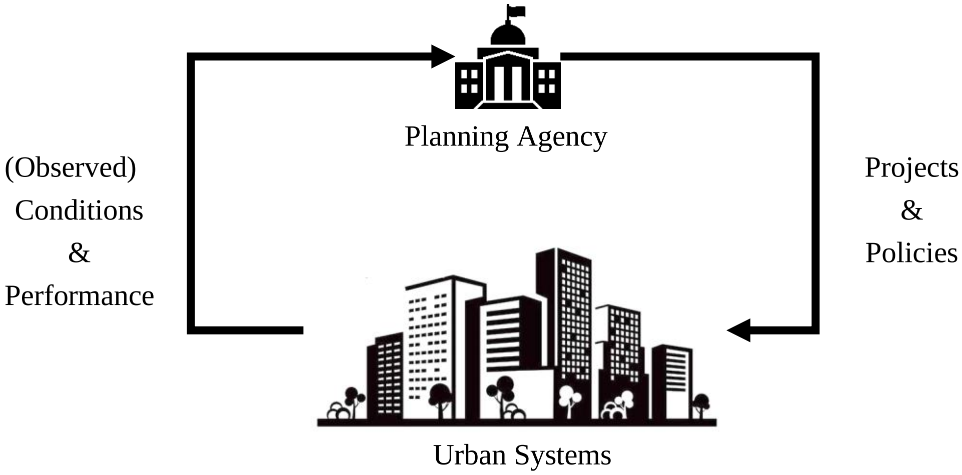


Figure 1: Simplified conceptualization of the interaction between a planning agency and the urban systems within the planning agency's domain.

Interpretable State-Space Representation of Urban Dynamics

Conceptual Framework

This subsection formulates the SLRTP as an interpretable state-space model (ISSM), where interpretable refers to the model's capacity to present its processes and outcomes transparently and understandably for human planners and decision-makers. Specifically, every state or action variable has a concrete and direct real-world mapping. For example, a state variable can be the total number of households or total lane miles of highways, both of which have immediate real-world counterparts and do not necessitate additional data mining or complex analytical interpretation for understanding. This approach ensures that the ISSM remains grounded in practical urban realities, facilitating more informed and effective SLRTP decisions and minimizing the model users' potential perception that the model is a black box.

Error! Reference source not found. illustrates a conceptualization of the SLRTP by adding additional details to Figure 1. The red arrow from the planning agency component represents the actions imposed on the urban system. This red arrow is further split into multiple sub-actions within the urban system component, such as infrastructure maintenance and pricing strategies. The dark arrows (within the urban system component) represent causal effects—their delays might vary from a few minutes (from travel patterns to accessibility) to a few years (from populations to housing prices). The blue arrow pointing from the urban system component to the planning agency component represents the information feedback from the urban system (e.g., system performance and public comments). More specifically, in **Error! Reference source not found.**, a decision-maker (planning agency) observes the system state at t (i.e., $\hat{\boldsymbol{s}}_t$) and the reward r_t (from a prior action $\boldsymbol{a}_{t-\tau}$) and then the agent decides on an action $\boldsymbol{a}_t$ (illustrated in red texts and arrows). The urban system (environment) evolves from $\boldsymbol{s}_t$ to $\boldsymbol{s}_{t+\tau}$, with the influence of $\boldsymbol{a}_t$ and, possibly, other stochastic factors. The decision-maker then observes the new system state $\hat{\boldsymbol{s}}_{t+\tau}$ and the latest reward $r_{t+\tau}$ and decides on the next action $\boldsymbol{a}_{t+\tau}$. We consider "no action" a possible action. The urban system then evolves from $\boldsymbol{s}_{t+\tau}$ to $\boldsymbol{s}_{t+2\tau}$. This process continues until the horizon year. Similarly, we have $\boldsymbol{a}_t = \boldsymbol{a}_{t_0+(h-1)\tau}, \forall t \in [t_0 + (h-1)\tau, t_0 + h\tau)$, where t_0 is the base year, $h \in \{1,2,\dots,H\}$ is the index of decision epochs, and $t_T = t_0 + (H-1)\tau$. H is the total number of planning intervals from t_0 to t_T. t_T is the planning horizon year. τ is the duration of each planning interval, say, four years.

We specify the urban system state with five main components:

- Transportation infrastructure and equipment (e.g., roadways, railways, buses, rail cars, shared bikes, transit stations)
- Socioeconomic conditions (e.g., households, employment, auto ownership)

- Real estate (e.g., housing units, business buildings, manufacturing facilities, warehouses)
- Natural environment (e.g., natural habits, terrains, soil conditions, hydrological conditions)
- Other factors such as politics, disruptive technologies, and new regulations (e.g., Americans with Disabilities Act (ADA)) that might influence people's (and businesses') accessibility to employers (labor) and propensity for real estate (development)

Note that the model we propose in this section is an urban systems model, not the model for the entire feedback system composed of the urban systems and the planning agency. Therefore, the action variables are exogenous to the proposed ISSM.

We define the following transition function to capture the dynamics of the system state variables:

$$\dot{\boldsymbol{s}}_t \equiv \frac{d\boldsymbol{s}_t}{dt} = M(\boldsymbol{s}_t, \boldsymbol{a}_t; \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t, \boldsymbol{\sigma}_t) \qquad \textbf{\textit{(0-1)}}$$

where $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^I$ captures the vector of a system state, $\boldsymbol{a} \in \mathcal{A} \subset \mathbb{R}^J$ represents the vector of decisions in the action space $\mathcal{A}$. $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are elasticity and base rate model parameter vectors, respectively. Each element of $\boldsymbol{\alpha}$ can be any positive real value (i.e., $\alpha_t \in \mathbb{R},$), while each element of $\boldsymbol{\beta}$ can only be positive real value (i.e., $\beta_t \in \mathbb{R}^+, \forall t$**)**. When we specify $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ as constant parameter vectors, we can simplify the notation by dropping the subscript t (i.e., $\boldsymbol{\alpha}, \boldsymbol{\beta}$). $\boldsymbol{\sigma}_{\boldsymbol{t}} \in \mathbb{R}$ is the parameter capturing the transition noise. Although we can capture noise as a stochastic process by assuming that $\boldsymbol{\sigma}$ follows, say, a Wiener process, in this paper, we model $\boldsymbol{\sigma}_t$ as a time-dependent function of t so that we can study the impact of the noise by varying this function during simulations. The mapping $M: \mathcal{S} \times \mathcal{A} \to \mathbb{R}^I$ captures the dynamics of the urban environment and governs how the state of the corresponding SLRTP model evolves from t to $t + dt$ with or without stochasticity. $t \in [t_0, t_T)$, or simply $t \in [0, T)$, is the time during a modeling period with a horizon at T, which we use to capture exogenous changes that influence urban dynamics. Note that one can treat time as a state variable with perfectly predictable dynamics. For example, we know that one year after 2023 is 2024. So, the MDP assumption is less restrictive than it first appears, as one can always incorporate historical information in the present state so that the evolution of the system from $\boldsymbol{s}_t$ to $\boldsymbol{s}_{t+dt}$ is only a function of $\boldsymbol{s}_t$ and $\boldsymbol{a}_t$, with independent stochasticity. However, it is still often useful to explicitly consider t. We further specify $M(\boldsymbol{s}_t, \boldsymbol{a}_t)$ as $M^+(\boldsymbol{s}_t, \boldsymbol{a}_t) - M^-(\boldsymbol{s}_t, \boldsymbol{a}_t)$, where $M^+(\boldsymbol{s}_t, \boldsymbol{a}_t) \geq 0$ represents the rate associated only with factors that increase the value of $\boldsymbol{s}_t$, while $M^-(\boldsymbol{s}_t, \boldsymbol{a}_t) \geq 0$ represents the rate associated only with factors that decrease the value of $\boldsymbol{s}_t$.

We start with the modeling of $M_i^-, \forall i \in \mathbb{I}$, where the critical component is the duration (delays) of the quantities in a state i. Let $D_t^i(\boldsymbol{s}_t, \boldsymbol{a}_t, t)$, $\forall i \in \{0,1, \dots, I\}$ represent the delay at t for the state variable i. In time-invariant cases, we simply write $D_t^i(\boldsymbol{s}_t, \boldsymbol{a}_t)$. We propose the following multiplicative form:

$$D_t^i\big(\boldsymbol{s}_t, \boldsymbol{a}_t, t; \alpha_t^{i,-}, \beta_t^{i,-}\big) = \beta^i \cdot \prod\nolimits_i \big(\tilde{s}_t^i\big)^{\alpha_t^{i,-}} \prod\nolimits_j \big(\tilde{a}_t^j\big)^{\alpha_t^{j,-}} \qquad \textbf{\textit{(0-2)}}$$

where $\tilde{s}_t^i$ and $\tilde{a}_t^j$ are the scaled values of s_t^i and a_t^j by $\breve{s}_t^i$ and $\breve{a}_t^j$. $\beta^i, \breve{\boldsymbol{s}}_t^i, \breve{\boldsymbol{a}}_t$ are the corresponding benchmark values and vectors, which are typically the base year values or some conventional values planning agencies use. This way, for any s_t^i, we can formulate M^- to determine the rate of decrease for the state variable i as follows:

$$M_i^-(\boldsymbol{s}_t, \boldsymbol{a}_t, t) = \frac{s^i}{D^i(\boldsymbol{s}_t, \boldsymbol{a}_t, t)}, \forall i \in \mathbb{I} \qquad \textbf{\textit{(0-3)}}$$

Note that $D^i \in \mathbb{R}^+$ and $0 \notin \mathbb{R}^+$. As we will show in Section **Error! Reference source not found.**, this approach preserves the desirable Markovian property of the dynamics of the state transitions.

The growth rate for the state variable i, M_i^+, is generally formulated as:

$$M_i^+(\boldsymbol{s}_t, \boldsymbol{a}_t, t) = \beta_t^{i,+} \prod\nolimits_i \big(s_t^i\big)^{\alpha_t^{i,+}} \prod\nolimits_j \big(a_t^j\big)^{\alpha_t^j}, \forall i \in \mathbb{I} \qquad \textbf{\textit{(0-4)}}$$

We define the reward of a given state as r_t, which measures the performance of a given state $\boldsymbol{s}_t$ without considering any return in the past or the future. We define $r\colon \mathcal{S} \mapsto \mathbb{R}$, and r and $\hat{r}$ are equivalent. We adopt a linear additive form for the mapping, as shown in Eqn. (***0-5***), to condense multiple performance metrics into a single composite quantity.

$$r(\boldsymbol{s}_t) = c + \sum\nolimits_{k\in\psi} w_k \cdot g^k(\boldsymbol{s}_t) \qquad \textbf{\textit{(0-5)}}$$

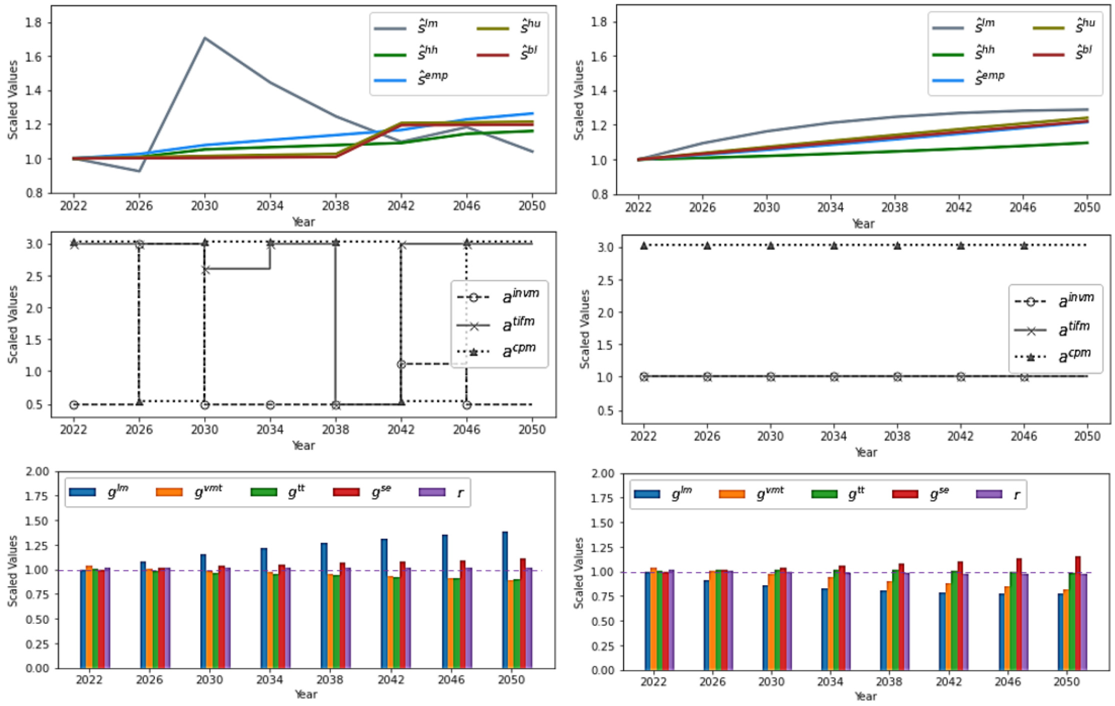
where $w_k \in [0,1]$, $t \in [t_0, t_T]$, $g^k(\boldsymbol{s}_t) = \left(\tilde{x}_t^k\right)^{\theta_k}$, $\theta_k \in \mathbb{R}$, and $\tilde{x}_t^k$ is the scaled auxiliary variable associated with performance measure $k \in \psi$. We might simplify $r(\boldsymbol{s}_t)$ as r_t when no confusion arises. Modelers and decision makers can utilize θ_k to capture the nonlinear effect of performance measure k in terms of its contribution to the composite function. If we want $\frac{\partial g^k}{\partial \theta_k} > 0$, then we set $\theta_k > 0$. If we further want $\frac{\partial^2 g^k}{\partial \theta_k^2} \leq 0$, then we set $\theta_k \in (0,1)$, otherwise we set $\theta_k > 1$. A similar principle applies to the case where $\frac{\partial g^k}{\partial \theta_k} < 0$. Thanks to these desirable properties and the linear additive property of $r(\cdot)$, we know that $r(\boldsymbol{s}_t') > r(\boldsymbol{s}_t)$, if and only if $\boldsymbol{s}_t' \succ \boldsymbol{s}_t$.

We can then obtain the cumulative rewards R_t (with discount factor $\gamma \in [0,1]$) from t to the planning horizon of an episode, as shown in Eqn. (***0-6***).

$$R_t = \sum\nolimits_{h=1}^{H} \gamma^{h-1} \cdot r_{t+h\cdot\tau} \qquad \textbf{\textit{(0-6)}}$$

Note that the immediate reward at t is counted from $t + \tau$, not from t, as any decisions made at t will not have any impact until $t + dt$ (or $t + \Delta t$ in a simulation, where Δt is the time step used for simulation). Furthermore, using a constant positive γ does not alter the preference rank in $\mathcal{S}$ (or $\hat{\mathcal{S}}$) because when $\gamma = 1$, we know that $\boldsymbol{s}_t' \succ \boldsymbol{s}_t$ (since $r(\boldsymbol{s}_t') > r(\boldsymbol{s}_t)$). Applying a positive γ other than 1 does not alter the preferential relationship.

Case Study with Reinforcement Learning



(a) **Policy Q**

(b) **Policy Q3**

Figure 9: Comparison between Policy Q (left column) and Policy Q3 (right column). For each column, the three rows of graphs show the time profiles of states, actions, and rewards, respectively. Both cases use the baseline weighting coefficients.