



Coordinated flow model for strategic planning of autonomous mobility-on-demand systems

Jiangbo (Gabe) Yu & Michael F. Hyland

To cite this article: Jiangbo (Gabe) Yu & Michael F. Hyland (05 Sep 2023): Coordinated flow model for strategic planning of autonomous mobility-on-demand systems, Transportmetrica A: Transport Science, DOI: [10.1080/23249935.2023.2253474](https://doi.org/10.1080/23249935.2023.2253474)

To link to this article: <https://doi.org/10.1080/23249935.2023.2253474>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 05 Sep 2023.



[Submit your article to this journal](#)



Article views: 165



[View related articles](#)



[View Crossmark data](#)

Coordinated flow model for strategic planning of autonomous mobility-on-demand systems

Jiangbo (Gabe) Yu ^a and Michael F. Hyland ^{a,b}

^aInstitute of Transportation Studies, University of California, Irvine, CA, USA; ^bCivil and Environmental Engineering, University of California, Irvine, CA, USA

ABSTRACT

High-quality strategic planning of autonomous mobility-on-demand (AMOD) systems is critical for the success of the subsequent phases of AMOD system implementation. To assist in strategic AMOD planning, we propose a dynamic and flexible flow-based model of an AMOD system. The proposed model is computationally fast while capturing the state transitions of two coordinated flows (i.e. co-flows): the AMOD service fleet vehicles and AMOD customers. Capturing important quantity dynamics and conservations through a system of ordinary differential equations, the model can economically respond to a large number and a wide range of scenario-testing requests. The paper illustrates the model efficacy through a basic example and a more realistic case study. The case study envisions replacing Manhattan's existing taxi service with a hypothetical AMOD system. The results show that even a simple co-flow model can robustly predict the systemwide AMOD dynamics and support the strategic planning of AMOD systems.

ARTICLE HISTORY

Received 27 February 2023

Accepted 18 August 2023

KEYWORDS

On-demand mobility; strategic planning; autonomous vehicles; state transition; co-flow; smart mobility; system dynamics

1. Introduction

Autonomous mobility-on-demand (AMOD) services are similar to existing ride-hailing/mobility-on-demand (MOD) services like those operated by Uber, Lyft, and Didi, except that AMOD vehicles are driverless. AMOD services present unique opportunities for central control, which can greatly enhance operational efficiency. As human drivers may not be completely compliant nor process instructions consistently, this paper assumes full vehicle automation is a necessary condition for central control of an AMOD service (i.e. MOD services with drivers are not interchangeable with AMOD services with driverless vehicles). Additionally, given the speculative nature of large-scale AMOD services (compared to large-scale MOD services that currently exist at scale), the modelling needs and focus for AMOD and MOD services are different, with AMOD services facing significant uncertainties that need to be systemically addressed during strategic planning.

CONTACT Jiangbo (Gabe) Yu  jiangby@uci.edu  Institute of Transportation Studies, University of California, 4000 Anteater Instruction & Research Building, Irvine, CA 92697, USA

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

To grapple with the challenges and opportunities of AMOD services, research in recent years has focused on analyzing the operations, management, and regulations of AMOD services. However, AMOD operators and public agencies lack strategic planning tools for launching or making major adjustments to AMOD systems. Such tools might help an AMOD operator choose the top 10 cities from 300 candidate cities for field testing, piloting, or initially implementing an AMOD service. Simultaneously, the tool might provide reasonable ranges for important service design variables for the 10 cities, including fleet size, service area, and pricing. Such a tool can also help a public agency determine whether to approve a business request from an AMOD operator or apply for state funding to procure AMOD vehicles for their communities.

The goal of this paper is to develop a strategic planning tool for AMOD systems with the following characteristics:

- (i) Computationally fast, i.e. the model requires only a few seconds to run a full-day scenario in a regular edge device, enabling the evaluation of a wide range of scenarios and allowing (near-) real-time interactions with the modelling tool during meetings, discussions, and collaborations among decision-makers and analysts.
- (ii) Parsimonious, i.e. the model quickly provides reasonable estimates for key performance metrics such as wait time, revenues and costs, and empty vehicle miles travelled.
- (iii) Policy sensitive, i.e. the model captures the relationship between system performance measures and changes in important service design parameters such as fleet size, service area size, and pricing.
- (iv) Dynamic, i.e. the model captures fundamental temporal dynamics.
- (v) Theoretically sound, i.e. the model captures causal chains between key system variables and parameters, and the conservation of physical quantities (vehicles and customers) in AMOD systems.

To achieve the goal of enhancing the understanding and planning of AMOD (Automated Mobility on Demand) systems, this paper introduces a novel modelling framework. Unlike previous flow-based MOD models (Nourinejad and Ramezani 2020; Xu et al. 2021), which have primarily focused on operational decisions using high-resolution models, this paper's contribution lies in the development of a co-flow model specifically designed for strategic planning purposes. The proposed framework also captures coordinated flows of vehicles and customers, which contrasts with the approach proposed by Jin, Martinez, and Menendez (2021) who use a 'fluid' to represent trips, rather than vehicles and customers separately.

In short, what sets the modelling approach in this paper apart from the existing literature is its ability to preserve key logical and causal relationships in dynamic AMOD systems by explicitly and jointly considering the state transitions of the service fleet vehicles and the customers. This integrative approach allows for a comprehensive understanding of the system, yielding valuable insights into critical causal relations, temporal dynamics, decision scopes, and model boundaries. By employing a state-space representation, our model enables quantitative analysis, providing decision-makers with a robust tool for understanding and optimizing the complex dynamics of AMOD systems. Through an illustrative example, we explore the mechanics of a simple model instantiated from the proposed

framework. Through a concrete case study, the effectiveness and computational efficiency of the co-flow model framework are empirically validated. The results demonstrate the model's ability to capture and represent the intricate dynamics of an AMOD system accurately with low computational burden, making it a practical tool for decision-makers in strategic AMOD planning.

The present paper does not attempt to maximize the application domain of the modelling framework. That is, we propose a modelling framework specifically for AMOD services, even though extensions or modifications to the framework could permit modelling and analysis of a broader range of mobility services. With data on human drivers, the modelling framework could capture driver behaviour and represent MOD services where the operator does not have the complete control over the vehicles (i.e. non-centralized or semi-centralized control). However, we will leave such scope-broadening and generalisations for future research.

The remainder of the paper is structured as follows. Section 2 reviews literature related to modelling AMOD operational, managerial, and regulatory problems, including relevant literature on modelling MOD services with human drivers. Section 3 introduces the mathematical formulation of the modelling framework with the aid of a stock-flow diagram. Section 4 and 5 describe how the co-flow model considers delays, decision variables, and their impacts. Section 6 presents a simple illustrative example to help readers gain a general understanding of how to use an instantiated co-flow model. Section 7 presents a case study where we examine the benefits and costs of deploying AMOD services in Manhattan (New York City) in different scenarios. We compare the outputs with an agent-based model (ABM) that was developed using taxi request data for studying hypothetical AMOD services. Section 8 concludes with the limitations and future research directions for the proposed models.

2. Literature review

2.1. Existing strategic planning models for mobility-on-demand services

Several research studies propose and develop strategic level planning models for mobility services. These studies aim to improve the planning capacity of transportation agencies (Inturri et al. 2019; Masoud et al. 2017; Nam et al. 2018; Steiner and Irnich 2020) and within-/cross-city operators (Rath et al. 2023) related to MOD services. The present paper contributes to the existing literature by focusing on preserving the dynamics and quantity conservation of AMOD system models with low computational burden to facilitate the strategic planning of AMOD services.

The present paper shares a similar overarching goal with Rath et al. (2023), namely, strategic planning and selecting cities for novel mobility services that have limited observations. However, the framework in Rath et al. (2023) involves estimating market equilibrium conditions (if they exist), generating synthetic scenarios, and fitting forecast models, for selecting MOD deployment strategies across cities. Conversely, the present paper focuses on reducing computational burden as much as possible while still satisfying the needs of AMOD strategic planning as described in Section 1. We believe the two approaches are complementary and can help address strategic planning problems.

2.2. *Modelling methods related to strategic planning of mobility-on-demand services*

When developing a model for AMOD system evaluation, choosing a suitable modelling method is critical. Some existing studies on modelling centrally controlled MOD systems involving human drivers implicitly or explicitly assume that human drivers are completely obedient, do not misunderstand instructions, and always respond with predictable deliberation time. In such cases, the corresponding models are also applicable to the present paper's scope on AMOD systems modelling. Popular MOD modelling methods include high-resolution, high-fidelity¹ (typically dynamic) ABMs, network flow-based models (Chen and Levin 2019; Ke et al. 2020), queuing models (Pinto et al. 2020; Sayarshad and Chow 2015), and lower resolution and lower fidelity (typically steady state) analytical, or statistical models. These models may or may not be intended for strategic planning in the MOD system engineering process, but we believe they could be candidate modelling paradigms or methods for the strategic planning of MOD systems. In the rest of the section, we review studies that utilise these popular modelling methods. We then discuss fluid – or flow-based models (Jin, Martinez, and Menendez 2021; Nourinejad and Ramezani 2020; Xu et al. 2021) and their differences from our proposed co-flow approach.

ABMs of MOD systems tend to suffer from considerable development workload and heavy computational burden to run even a single scenario (Azevedo et al. 2016; Djavadian and Chow 2017; Fagnant and Kockelman 2014; Hyland and Mahmassani 2018; Inturri et al. 2019; Javanshour et al. 2021; Javanshour, Dia, and Duncan 2019; Masoud et al. 2017; Narayan et al. 2021; Oh et al. 2020; Syed, Dandl, and Bogenberger 2021). When using ABMs, planners and analysts have to spend much effort on collecting data and specifying the detailed variables and logic rather than on the actual strategic planning activities. The ABM method also faces significant calibration challenges due to the data requirements. Hence, ABMs often fail to produce straightforward insights into the relationship between system decision variables and system performance. As a final point, in time – and resource-constrained strategic planning, it is often better to run a large number of scenarios using a lower resolution, lower fidelity modelling method than just a few scenarios with an ABM.

Network flow models (Chen and Levin 2019; Ke et al. 2020) and network queuing models (Pinto et al. 2020; Sayarshad and Chow 2015) typically evaluate MOD services with explicit topologies of highways, requiring a significant amount of additional *microscopic* detail that increases model development, calibration, validation, and run time. Additionally, many of these models postulate the existence of equilibrium condition(s) and/or other simplified assumptions on which analysts evaluate system performance.

In contrast, aggregate, analytical, equilibrium-oriented, and statistical approaches (Acheampong et al. 2020; Bilali et al. 2020; Rahimi, Amirgholy, and Gonzales 2018; Yang and Wong 1998; Zha, Yin, and Xu 2018) tend to be neither data-intensive nor computationally expensive. These models have the potential to be used by AMOD businesses and public agencies in their strategic planning. But these models do not provide information on the dynamics of the AMOD system. For example, it is challenging for these models to provide insights into the impact of an irregular demand surge or a major disruptive event at 2 pm on the service performance at 5 pm. Moreover, as decision-makers cannot visualise how the 'trajectory' of the system performance 'evolves' from 2 pm to 5 pm with these models,

decision-makers might find them unsatisfactory for certain decisions or too risky to trust the model outputs.

Several studies in the literature propose modelling MOD vehicles and customers as flows (Jin, Martinez, and Menendez 2021; Nourinejad and Ramezani 2020; Xu et al. 2021). Jin, Martinez, and Menendez (2021) develop an intuitive and mathematically tractable quad-compartment model that builds on the concept of state transitions. Nourinejad and Ramezani (2020) develop a dynamic and non-equilibrium-based model that calculates the customer-vehicle match rate directly rather than modelling the wait time first. The authors denote that the average wait time is approximately the ratio between the waiting riders and the matching rate. They model matching rate using a Cobb–Douglas function for which the two elasticity variables for the number of idle vehicles and unassigned passengers (γ_1 and γ_2) need to be calibrated. However, if either of these parameters is calibrated to a value other than unity, then the units of the matching rate and boarding rate do not have a physical interpretation. Xu et al. (2021) model the spatial–temporal characteristics of a region, including vehicles and customers moving among regions. Moreover, their model explicitly considers customer pick-up time (after being matched) in their model. Interestingly, Xu et al. (2021) model state transitions using a discrete event approach – the number of vehicles (customers) at a given state at $t + w$ is determined by the number of vehicles (customers) in the precedent state at t , where w is the state-transition delay (e.g. 20 min travel time).

Our proposed co-flow model in this paper offers several advantages over the existing flow-based models in the literature. Our co-flow model considers customers and vehicles as separate but coordinated flows, whereas Jin, Martinez, and Menendez (2021) consider only MOD trips in their model. The differentiation in our model between passenger flows and vehicle flows captures the reusable nature of vehicles – after a vehicle drops off a traveller, the vehicle can transition to a repositioning state or idle state. The proposed co-flow model in this paper allows greater flexibility in modelling operational policies at the macroscopic level (e.g. repositioning strategies) and analyzing various public policies (e.g. maximum price and fleet size regulations). We also propose a delay-oriented approach for modelling state transitions in the co-flow model that overcomes the unit-varying concern of the Cobb–Douglas function adopted by Nourinejad and Ramezani (2020). Additionally, unlike Xu et al. (2021), our co-flow model incorporates a continuous, rather than discrete, delay modelling method permitting a state-space representation that is fully continuous and differentiable. In addition to the explicit consideration of vehicle flows and customer flows, unit tractability, and the preservation of the Markov property, the co-flow model has the novel capability of adapting to a range of resolutions, fidelities, and boundaries based on decision needs and data availabilities, which is absent from the existing literature.

The idea of explicitly modelling the coordinated nature of service resources and entities to be serviced as flows is not new. For example, Forrester (1969) considers households and housing units in a coordinated manner so that the change of housing occupancy reflects the underlying ratio between households and housing units. Some researchers have considered the coupled nature of commodities with vehicles in static assignment. For example, Guélat, Florian, and Crainic (1990) consider constrained vehicle (e.g. trucks and trains) inventory when assigning heterogeneous products to transportation networks. Additionally, Chow, Ritchie, and Jeong (2014) relax the coupling of the commodity flows and

commodity vehicle flows in the approach by Gu  lat, Florian, and Crainic (1990) to capture traffic flows more realistically in a network.

3. Modelling framework for strategic planning of AMOD systems

3.1. Definitions and nomenclature

- Stock variable (or simply stock): a variable that changes its value through accumulating and releasing (or leaking) quantities that represent a certain feature (e.g. total vehicles in a certain state) of the system under study. Stock variables change values through their associated inflows and outflows.
- Flow variable (or simply flow): a variable representing the flow (as rate) between stocks within a system boundary or across the model boundary and thereby into or out of the system; changes in stocks over time. Flows represent activities (e.g. matching vehicles with customers), in contrast to stocks, which represent states of the system. An inflow variable (into a certain stock variable) represents flows into a stock, while an outflow variable represents flows out of a stock. The change rate of a stock variable at time t is the summation of all the inflows and outflows at that moment.
- Auxiliary variable: a model variable that can change its value instantaneously at $t + dt$ based *solely* on the values of other model variables at t . This type of variable has no ‘memory’ of its historical value (values in previous simulation steps during a simulation) and is usually an intermediate variable to facilitate the expression of a functional dependency of a flow variable to a corresponding stock variable.
- Exogenous variable: a variable that is influenced by factors outside the system boundary. Exogenous variables may be explicit or implicit. For example, this paper explicitly assumes that the average time to drop off customers is exogenous as the paper considers the fleet size and service activities do not have a significant impact on the overall traffic congestion. On the other hand, the weather is an implicit exogenous variable as the paper does not have any variable that represents weather. An exogenous variable can be converted to an endogenous one by making it a function of other variables within the system boundary.
- Endogenous variable: a variable that is internal to the system boundary. It is assumed to be (causally) influenced by variable(s) within the system boundary. An endogenous variable can be converted to an exogenous variable by making it a constant or only dependent on time. This is appropriate when such a conversion does not influence model behaviour significantly for a given study need. For example, when the AMOD vehicles are only a small portion of the overall fleet in the traffic network, we can simply use a constant or a lookup table that only depends on time-of-day to capture traffic speed.
- $X_t^{(i)}$: number of AMOD service vehicles in state i at time t , where $i \in \mathbb{I}$ and $t \in [0, T]$. $\mathbf{X}_t$ is the state vector whose element i is $X_t^{(i)}$. $\mathbb{I} = \{1, 2, \dots, i, i', \dots, l\}$. $\mathbf{X}_t \equiv \mathbf{X}(t)$. $\hat{\mathbf{X}}_t$ represents the corresponding observed or target state vector.
- $x_t^{(i,i')}$: rate of AMOD service vehicles changing from state i to i' at time t . where $i, i' \in \mathbb{I}$ and $t \in [0, T]$.

- $Y^{(j)}$: number of customers in state j at time t , where $j \in \mathbb{J}$ and $[t \in [0, T]$. $\mathbf{Y}_t$ is the state vector whose element j is $Y_t^{(j)}$. $\mathbb{J} = \{1, 2, \dots, j, j', \dots, J\}$. $\mathbf{Y}_t \equiv \mathbf{Y}(t)$. $\hat{\mathbf{Y}}_t$ represents the corresponding observed or target state vector.
- $y_t^{(jj')}$: rate of customers changing from state j to j' at time t . where $j, j' \in \mathbb{J}$ and $t \in [0, T]$
- $\mathbf{M}_t$: A state vector formed by concatenating $\mathbf{X}_t$ and $\mathbf{Y}_t$ in the same dimension. It contains $I + J$ elements. It is used when describing the dynamics of the whole AMOD system. $m_t^{(l)}$ is the l^{th} element, $l \in \{0, 1, \dots, I + J\}$. $\hat{\mathbf{M}}_t$ represents the corresponding observed or target state vector.
- $\mathbf{a}$: Decision variable vector (English letter a , not Greek letter α).
- $\boldsymbol{\beta}$: Parameter vector.
- $f: \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^{\dim(\mathbf{a})} \rightarrow \mathbb{R}^I$ maps the system state and action variables to the transition rates of vehicle states.
- $g: \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^{\dim(\mathbf{a})} \rightarrow \mathbb{R}^J$ maps the system state and action variables to the transition rates of customer states.
- $h: \mathbb{R}^{I+J} \times \mathbb{R}^{\dim(\mathbf{a})} \rightarrow \mathbb{R}^I$ maps the system state and action variables to the transition rates of vehicle states.
- $\bar{h}: [0, T] \rightarrow \mathbb{R}^{I+J}$ is the solution of $\dot{\mathbf{M}} = h(\mathbf{M}, \mathbf{a}; \boldsymbol{\beta})$ such as $\mathbf{M}_t = \bar{h}(t)$.
- $\psi: \mathbb{R}^{\dim(\mathbf{M})} \times \mathbb{R}^{\dim(\mathbf{a})} \rightarrow (0, 1]$. Matching efficiency
- $k_{\min}$: the minimum of the number of idling vehicles and the number of unmatched waiting customers.
- $q: \mathbb{R}^+ \cup \{0\} \rightarrow \mathbb{R}^+$ is a continuous, decreasing, and inverse s-shape function with $q(0)$ is the maximal allowable value (greater than 1), $q(1) = 1$, and $\lim_{x \rightarrow +\infty} q(x) = 0$.
- $\varepsilon: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ captures the elasticity of matching efficiency to repositioning time (effort)
- t_0 : initial time. When it is a subscript and no confusion arises, it is simply denoted as 0.
- T : modelling horizon. For example, $T = 24$ th hour (i.e. 1440th minute) from the starting time.
- τ : time step (for simulation). For example, $\tau = 1$ minute. When no confusion arises, the subscript t may be dropped to simplify the notation.

3.2. Co-flow model idea

This subsection conceptualises the co-flow model with the aid of a stock-flow diagram in Figure 1 that shows the model structure of the proposed co-flow model. In addition to the text description of each variable in the diagram, we also include mathematical notations. A stock variable (represented by a rectangular box) in the diagram can be intuitively understood as a tank or reservoir that stores fluids. In each time step, the stock level changes based on the difference between the inflow and outflow rates/variables. In a sense, the stock variables contain the memory of their previous time step. A chain of stocks can be understood as a sequence of tanks with fluids flowing among them or from/to external sources.

The lower stock chain in Figure 1 captures customer fluids that correspond to the process customers experience when using an AMOD service, while the upper stock chain captures vehicle fluids that correspond to the process in which AMOD service vehicles are operated. In other words, we use one stock chain to capture the state transitions of aggregate riders

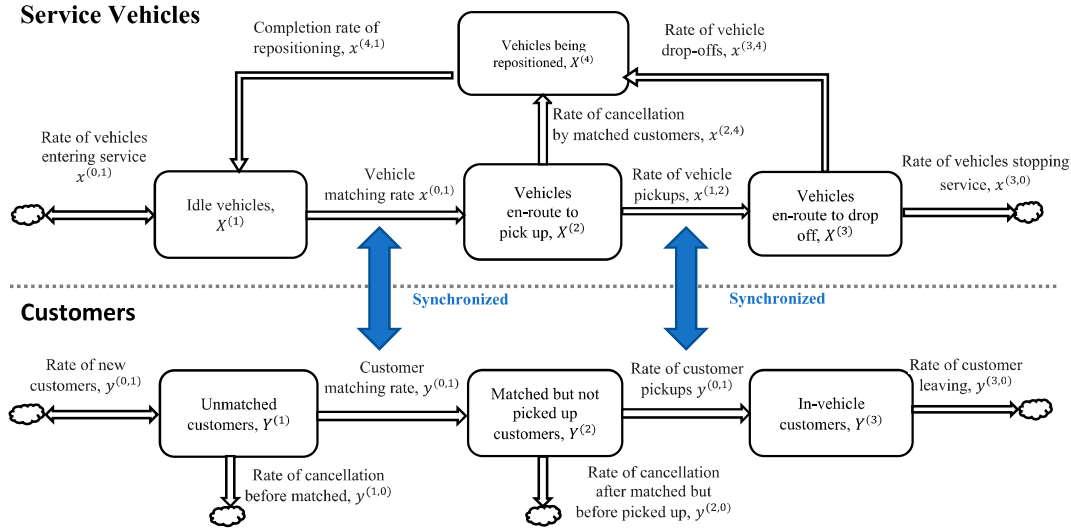


Figure 1. Example co-flow model of the AMOD system, where the top stock chain represents service vehicles, and the bottom stock chain represents the customers.

and use another stock chain to capture the state transitions of aggregate service vehicles. These two stock chains are paired chains that jointly and consistently capture the dynamics of the AMOD system operation through the customer stocks and flows as well as service vehicle stocks and flows.

The vehicle stocks include idle vehicles, $Y^{(1)}$, vehicles en-route to pick up, $Y^{(2)}$, vehicles en-route to drop-off, $Y^{(3)}$, and repositioning vehicles, $Y^{(4)}$. The customer stocks include unmatched customers, $X^{(1)}$, matched customers that are waiting for pickup, $X^{(2)}$, and in-vehicle customers, $X^{(3)}$.

In the chain of customer stocks, the rate of change of unmatched customers, $Y^{(1)}$, is the difference between the rate of newly entering customers, $y^{(0,1)}$, and the customer matching rate, $y^{(1,2)}$, assuming no customers leave the system before being matched. Similarly, the rate of change of matched but not yet picked up customers, $Y^{(2)}$, is the difference between the customer matching rate, $y^{(1,2)}$, and the rate of customer pickups, $y^{(2,3)}$. Finally, the rate of change for the in-vehicle customers who are en-route to their destinations, $Y^{(3)}$, is the difference between the rate of customer pickups, $y^{(2,3)}$, and the rate of customer arrivals (i.e. drop-offs), $y^{(3,0)}$. Similar ideas apply to vehicle flows, $x^{(i,i')}, \forall i, i' \in \mathbb{I}, i \neq i'$, and vehicle stocks, $X^{(i)}, \forall i \in \mathbb{I}$.

The flow variables describe the change of stock variables. For example, the change in fleet size for vehicles en-route to pick up at a given moment depends on the number of vehicles just matched with customers (inflow) minus the number of vehicles that just picked up customers (outflow). The flow from the stock of idle vehicles to the stock of vehicles en-route to pick up must equal the flow from the stock of unmatched customers to the stock of customers that have been matched but have not been picked up.

Some might wonder why there is no 'pipeline' connecting from the repositioning vehicles stock variable $X^{(4)}$ to the vehicles en-route to pick up stock variable $X^{(2)}$ in Figure 1. This is because this subsection only provides one possible instantiation of the proposed co-flow modelling framework. The following sub-section generalises this simple example and provides a mathematical formulation of the co-flow model.

3.3. State-space representation of the co-flow model

The co-flow model can be represented mathematically using a state-space representation, which facilitates rigorous analysis of system properties and proper simulations. Let $\mathbf{X}_t \equiv \mathbf{X}(t)$ denote the state of the AMOD service vehicle stock variables at time t . When no confusion arises, we use $\mathbf{X}$ instead of $\mathbf{X}_t$ to simplify the notation. $\mathbf{X} \in \mathbb{R}^I$. A similar principle applies to other symbols. Equation (1) displays the rate of change of the stock variables as a function of several endogenous variables and exogenous factors:

$$\dot{\mathbf{X}} \equiv \frac{d\mathbf{X}}{dt} = f(\mathbf{X}, \mathbf{Y}, \mathbf{a}; \beta) \quad (1)$$

where $f(\cdot)$ captures how the system $\mathbf{X}_t$ changes to $\mathbf{X}_{t+dt}$; $\mathbf{a}$ is a vector of decision variables (we will discuss how to model the impact of $\mathbf{a}$ on a co-flow model in details in Section 5), and β represents the parameter vector.

The state vector, $\mathbf{Y} \in \mathbb{R}^J$, contains the J stock variables for the customers. Equation (2) describes the dynamics of the customer stock variables:

$$\dot{\mathbf{Y}} \equiv \frac{d\mathbf{Y}_t}{dt} = g(\mathbf{X}, \mathbf{Y}, \mathbf{a}; \beta) \quad (2)$$

where $g(\cdot)$ captures how the system $\mathbf{Y}_t$ changes to $\mathbf{Y}_{t+dt}$.

In Equations (1) and (2), we assume that all states, such as total number of customers waiting to be matched, are fully observable for the operator. We can further simplify the notation using Equation (3), where $\mathbf{M} \equiv (\mathbf{X}, \mathbf{Y})$, $\dot{\mathbf{M}} \equiv \frac{d\mathbf{M}}{dt}$, and $h(\cdot)$ captures how the overall AMOD system changes from $\mathbf{M}_t$ to $\mathbf{M}_{t+dt}$. $m_t^{(l)}$ is the l^{th} element, $l \in \{0, 1, \dots, I + J\}$.

$$\dot{\mathbf{M}} = h(\mathbf{M}, \mathbf{a}; \beta) \quad (3)$$

Equations (1) and (2) clearly illustrate the coordinated nature of the customer and AMOD service vehicle stock chains within the model. So, we further describe the modelling framework by specifying Equations (1) and (2). Equation (4) describes the dynamics of all J stock variables that $\mathbf{Y}$ captures, where $\delta^{(kj)} = 1$ if there is a connection from the stock k to j ; otherwise, $\delta^{(kj)} = 0$. In the chain of customer stocks illustrated in Figure 1, for example, the rate of change for customers matched but not picked up is the cumulative difference between the inflow rate (i.e. the customer matching rate) and the outflow rate (i.e. the rate of customer pickups).

$$\dot{Y}_t^{(j)} = \sum_{\forall k \in \mathbb{J}, k \neq j} \delta^{(kj)} y_t^{(kj)} - \sum_{\forall m \in \mathbb{J}, m \neq j} \delta^{(jm)} y_t^{(jm)}, \forall j \in \mathbb{J} \quad (4)$$

Related to Equation (4), Equation (5) describes the cumulative system dynamics of $Y_T^{(j)}$ starting from the initial condition for the stock variables, $Y_{t_0}^{(j)}$, to the current time T ($T \geq t_0$).

$$Y_T^{(j)} = Y_{t_0}^{(j)} + \int_{t_0}^T \left(\sum_{\forall k \in \mathbb{J}, k \neq j} \delta^{(kj)} y_t^{(kj)} - \sum_{\forall m \in \mathbb{J}, m \neq j} \delta^{(jm)} y_t^{(jm)} \right) d\tilde{t}, \forall j \in \mathbb{J} \quad (5)$$

Equation (6) parallels Equation (4) except it captures AMOD vehicle system dynamics of all I possible vehicle states. A parameter $\eta^{(li)} = 1$ if there is a connection from the stock l to i ;

otherwise, $\eta^{(l,i)} = 0$.

$$\dot{X}_t^{(i)} = \sum_{\forall l \in \mathbb{I}, l \neq i} \eta^{(l,i)} x_t^{(l,i)} - \sum_{\forall v \in \mathbb{I}, v \neq i} \eta^{(i,v)} x_t^{(i,v)}, \forall i \in \mathbb{I} \quad (6)$$

Similar to Equation (5), Equation (7) describes the cumulative process of the stock variables for vehicles, $X^{(i)}$, from their initial state, $X_{t_0}^{(i)}$, to the current time T , $T \geq t_0$.

$$X_T^{(i)} = X_{t_0}^{(i)} + \int_{t_0}^T \left(\sum_{\forall l, l \neq i} \eta^{(l,i)} x_{\tilde{t}}^{(l,i)} - \sum_{\forall v, v \neq i} \eta^{(i,v)} x_{\tilde{t}}^{(i,v)} \right) d\tilde{t}, \forall i \in \mathbb{I} \quad (7)$$

When there is no ridesharing, the matching rate for waiting customers, $x^{(1,2)}$, and that for idling vehicles, $y^{(1,2)}$, should be the same, $x^{(1,2)} = y^{(1,2)}$.

Equation (8) shows that if the AMOD operator can only match idle vehicles to unmatched customers, then the matching rate is constrained by the minimum of the number of idle vehicles and the number of unmatched customers. That is, we assume that the AMOD will not allocate non-idle or non-existent vehicles to waiting customers. We define the minimum of $X^{(1)}$ and $Y^{(1)}$ as an auxiliary variable, $k_{\min}$. Matching efficiency (i.e. the percentage of customers that are successfully matched to vehicles) is rarely 100% in practical operation. Hence, we introduce a function $\psi(\mathbf{M}, \mathbf{a}) \in (0, 1)$, which measures matching efficiency as a percentage per unit time. The corresponding flow rates can be formulated as.

$$x^{(1,2)} = y^{(1,2)} = k_{\min} \cdot \psi(\mathbf{M}, \mathbf{a}) \quad (8)$$

MOD operators can control the function $\psi(\mathbf{M}, \mathbf{a})$, and they typically decrease the value when idle vehicles are located far from customers. Furthermore, $\psi(\mathbf{M}, \mathbf{a})$ would be higher as the ratio of vehicles to customers increases, and even as the total number of vehicles (independent of the number of customers) increases due to network effects. However, if the AMOD operator needs to serve a large volume of requests, this may slow down the operator's computing servers, causing $\psi(\mathbf{M}, \mathbf{a})$ to decrease. In this case, however, it is still possible for the absolute transition rates to increase while $\psi(\mathbf{M}, \mathbf{a})$ decreases.

In simulations, we can specify a finite time step, τ , to approximate Equation (4), using the Euler integration method as shown in Equation (9).

$$Y_{t+\tau}^{(j)} \approx Y_t^{(j)} + \left(\sum_{\forall k, k \neq j} \delta^{(k,j)} y_t^{(k,j)} - \sum_{\forall m, m \neq j} \delta^{(j,m)} y_t^{(j,m)} \right) \cdot \tau, \forall j \in \mathbb{I} \quad (9)$$

The same method can be applied to Equation (6). Selecting the appropriate step is beyond the scope of the paper, but we find that a step size between one to five seconds is practically reasonable in the context of simulating regional AMOD system dynamics.

We have established the relationships among stocks and flows in the co-flow model, but we have not yet provided the formulations for the flows, except for $x^{(1,2)}$ and $y^{(1,2)}$. In Section 4, we introduce a unified approach that uses negative exponential distributions for capturing a variety of delays in state transitions with flexible resolutions and fidelities. The delays serve as the key auxiliary variables for capturing the flows in the system (except for input flows).

3.4. Existence and uniqueness

Section 3.3 shows that we can describe the dynamics of an AMOD system in the form of Equation (3) and potentially find a solution. By a solution of Equation (3) at some interval $t \in [0, T]$, we mean the case where we can find $\bar{h} : [0, T] \rightarrow \mathbb{R}^{I+J}$ such as $\mathbf{M}_t = \bar{h}(t)$. That is, we can describe the dynamics as (only) a function of t . The associated initial value problem consists of Equation (3) and $I + J$ initial conditions. Although the overall model structure shown in Equation (4) and Equation (5) is a linear formulation, the transition rate $x_t^{(i,j')}$ and $y_t^{(j,j')}$ may or may not be linear. However, we still have the following proposition.

Proposition: Let $h_1, h_2, \dots, h_{I+J}$ in Equation (3) be continuous functions having continuous partial derivatives $\frac{\partial h_1}{\partial m^{(1)}}, \frac{\partial h_2}{\partial m^{(2)}}, \dots, \frac{\partial h_{I+J}}{\partial m^{(I+J)}}$ in the definition domain containing the initial condition point $t = 0$ and $(m_0^{(1)}, m_0^{(2)}, \dots, m_0^{(I+J)})$. Then there exists a solution $\bar{h} : [0, T] \rightarrow \mathbb{R}^{I+J}$ on $t \in [0, T]$ satisfying $\mathbf{M}_t = \bar{h}(t)$. Furthermore, the solution is unique.

Proof: According to the existence and uniqueness theorem proven by Coddington and Levinson (1984) for systems of ordinary differential equations, we know the existence and uniqueness of the solution for $t \in (-T, T)$, where $T > 0$. As $[0, T]$ is a subset of $(-T, T)$, we know the theorem applies to $t \in [0, T)$. Define $\delta = \lim_{t \rightarrow T^-} T - t$. We can apply Equation (3) to $\mathbf{M}_{T-\delta}$ to obtain $\mathbf{M}_T$. Hence, the proof is complete. ■

In Section 4, we propose one flexible approach for modelling the transition rates in a way that ensures the continuous partial derivatives. However, the formulation of $k_{\min}$ in Equation (8) might be a strict one, as defining $k_{\min} = \min(X^{(1)}, Y^{(1)})$ will cause the violation of the existence of continuous partial derivatives in the proposition above. Therefore, we use $k_{\min} = X^{(1)} \cdot (1 - \exp(-Y^{(1)}/X^{(1)}))$ when $X^{(1)} > 0$ and $k_{\min} = 1$ when $X^{(1)} = 0$ to ensure that the continuous partial derivatives exist. As illustrated in Figure 2, this specification of $k_{\min}$ ‘softens’ the original min function to ensure the existence of continuous partial derivatives over both $X^{(1)}$ and $Y^{(1)}$. For example, when $X^{(1)} = 1$ and $Y^{(1)} = 500$, $k_{\min} \approx 0.99$; when $X^{(1)} = 500$ and $Y^{(1)} = 1$, $k_{\min} \approx 0.99$; when $X^{(1)} = Y^{(1)} = 1$, $k_{\min} \approx 0.632$; when $X^{(1)} = Y^{(1)} = 500$, $k_{\min} \approx 316$.

3.5. Model estimation methods

Numerous methods exist to estimate and calibrate dynamic models. The most common approaches aim to optimise or minimise descriptive statistics, often called goodness-of-fit measures, which evaluate the point-by-point agreement between the model outputs and target data series. The target data series can come from empirical sources or other models (e.g. high-resolution, high-fidelity agent-based models, decision-makers’ mental models). In the proposed framework’s context, these metrics quantify the error between the model outputs and the target data at specific time points where data are available. These statistics then average the relevant time horizon, with or without the inclusion of discount factors. Commonly used statistics include likelihood (or log-likelihood) functions, R-square, mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and Theil’s inequality statistics. For instance, we can employ MAE as a goodness-of-fit criterion to estimate the parameter vector β , which we can solve using nonlinear

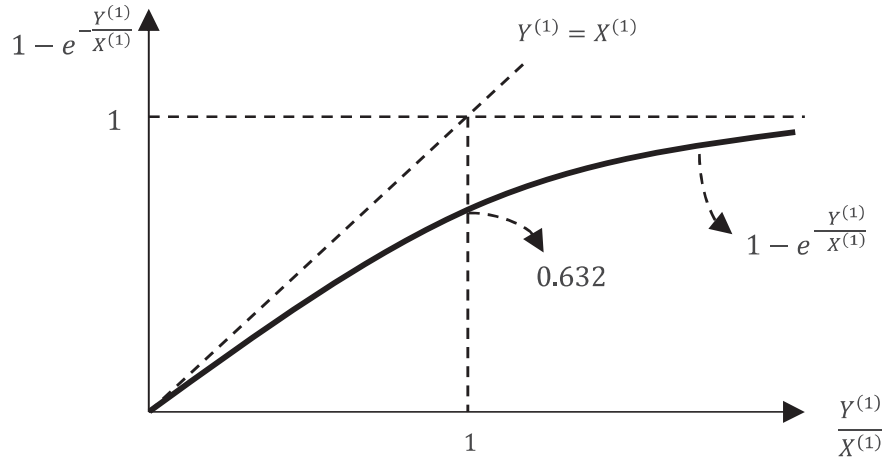


Figure 2. Illustration (not to scale) of $1 - e^{-\frac{Y^{(1)}}{X^{(1)}}}$ when varying the ratio of $Y^{(1)}$ to $X^{(1)}$ ($X^{(1)}, Y^{(1)} > 0$).

optimisation algorithms such as Bayesian Optimization, Bayesian Inference, Expectation-Maximization (EM), Simultaneous Perturbation Stochastic Approximation (SPSA), Simulated Annealing, Genetic Algorithms, and Inverse Optimization (Burton and Toint 1992; Chow, Ritchie, and Jeong 2014). When dealing with highly complex state-space models involving numerous time-dependent parameters or when continuous online estimation is necessary, methods such as Particle Filtering and Kalman Filtering (and Smoothing) are relevant. When an instantiated model is relatively simple, we can use grid search methods to explore the impact of different combinations of model parameters on the predetermined goodness-of-fit measures and choose the best combination of parameters.

4. Modelling delays

There are multiple types of delays in the AMOD system, such as the delay that occurs when (i) the AMOD operator matches customers with vehicles; (ii) customers deliberate whether they should take the offer from the AMOD operator; (iii) the AMOD operator or individual vehicles decide whether to accept a customer request; (iv) service vehicles are currently on their way to pick up customers; and (v) service vehicles are on their way to drop off customers.

To capture these delays, we propose a unified delay modelling approach that uses the negative exponential distribution as an elementary mechanism for modelling state transitions. This approach can capture a wide range of delay distributions and maintain differentiable state transitions.

Delay variables may be endogenous, meaning that they can be a function of other variables in the model. For instance, a larger volume of unmatched customers may contribute to reduced average pick-up delays for upcoming matched customers when given a certain fleet size of idling vehicles. Additionally, jointly considering non-idle AMOD service vehicles with non-AMOD vehicles may be useful for obtaining overall network density and speeds, which could impact vehicle matching, pick-up, and drop-off rates (Beojone and Geroliminis 2021).

4.1. Negative exponential distributions of delays

When we assume that the content in a stock loses its sequence (i.e. perfect mix with lost memory in terms of the sequence of customer ‘arrivals’ into the stock) and a given average delay, we can simply divide the total stock by the average delay to obtain the flow at each time moment. The distribution of such a delay is reverse-J shaped and is often called the negative exponential distribution.

Travel time (TT) can be directly given as an exogenous variable or obtained through dividing the distance by the network speed ($spd_t > 0$), as shown in Equation (10). As indicated by the subscript t , spd_t is time dependent.

$$TT_t^w = \frac{dist_t^w}{spd_t} \quad (10)$$

where $w \in \mathbb{P}$, and $\mathbb{P} = \text{pickup } (p)$, dropoff (d), and reposition (r), as shown in Figure 1. To simplify the case, we assume that the average driving speed is the same for all trip purposes. In addition to the speed being time-dependent, the speed may also be a function of other system variables. For example, when the AMOD fleet size is not negligible relative to the total inventory of the active vehicles in the network, the number of en-route AMOD vehicles might influence the network speed. In such cases, the impact of AMOD system dynamics influences the network speed, which in turn affects the AMOD system dynamics.

We can use TT_t^w and Equation (11) as the average delay to calculate flows (exit rates from the corresponding stock variables) for customers at t .

$$y_t^{(j,j')} = \frac{Y_t^{(j)}}{TT_t^{w(\mathbb{P},j)}} \quad (11)$$

We abuse the notation a bit by using $(w(\mathbb{P},j))$ to represent the activity (specified in $\mathbb{P}$) that corresponds to $Y^{(j)}$. A similar method applies to the vehicle flows (existing rates from the corresponding stock variables) for service vehicles:

$$x_t^{(i,i')} = \frac{X_t^{(i)}}{TT_t^{w(\mathbb{P},i)}} \quad (12)$$

We have specified the formulation for the matching rate, where $\psi(\mathbf{M}, \mathbf{a})$ may be interpreted as the percent of $k_{\min}$ being matched from t to $t + dt$, but they together can also be interpreted from the delay perspective: if $X^{(1)} \leq Y^{(1)}$, the average delay is $\frac{1}{\psi(\mathbf{M}, \mathbf{a})}$ for $X^{(1)}$ and $\frac{Y^{(1)}}{X^{(1)} \cdot \psi(\mathbf{M}, \mathbf{a})}$ for $Y^{(1)}$; if $X^{(1)} > Y^{(1)}$, the average delay is $\frac{X^{(1)}}{Y^{(1)} \cdot \psi(\mathbf{M}, \mathbf{a})}$ for $X^{(1)}$ and $\frac{1}{\psi(\mathbf{M}, \mathbf{a})}$ for $Y^{(1)}$.

Equations (11) and (12) together imply a so-called negative exponential delay of fluid in a corresponding stock. The flow of existing customers being picked up is the total customers waiting to be matched (at time t) divided by the average pick up time. This implies that the corresponding stock variable does not have any memory about when each customer (or vehicle) enters the stock. In simulations, if a customer (vehicle) enters at $t = 5$ min and $TT^p = 10$ min and the time step is 1 min, then we know that 1/10 of this customer (vehicle) exits from the stock in a minute, and then 1/10 of what was left exits in the next minute.

Note that delays may be exogenous or influenced by a combination of controllable and non-controllable factors. For example, when the AMOD operator’s policy has little influence

on the customers' destination choices and fleet size is small compared to the overall en-route vehicles, we can assume π_t^d and π_t^p are exogenous. For another example, the rate of vehicles becoming idle after repositioning status may be mainly determined by the policy of the AMOD operator, especially when all the service vehicles are autonomous and in central control. In this paper, the pick-up time is captured as a function of a base value for pickup time, $\tilde{\pi}^p$, the number of idle vehicles, and the number of waiting and customers, through a multiplicative form. That is,

$$\pi_t^p = \tilde{\pi}^p \cdot q(\tilde{X}_t^{(0)}; \beta_{i;p}) \cdot q(\tilde{Y}_t^{(0)}; \beta_{w;p}) \quad (13)$$

where $q: \mathbb{R}^+ \cup \{0\} \rightarrow \mathbb{R}^+$ is a continuous, decreasing, and inverse s-shape function with $q(0)$ is the maximal allowable value (greater than 1), $q(1) = 1$, and $\lim_{x \rightarrow +\infty} q(x) = 0$. $\beta_{i;p}$ and $\beta_{w;p}$ are both expected to be positive to capture the elasticity of pick-up time with respect to the scaled idling fleet size and scaled waiting customers, respectively.

4.2. Chained or higher-order negative exponential distributions of delays

Some delay distributions are clearly not negative exponential. Repositioning might have a delay distribution that has one or multiple mounded modes. In these cases, we can use a sequence of delays with each component delay following an exponential distribution to capture more complex delay distributions. Chaining stocks to better capture delay distributions is known as high-order delay modelling in system dynamics (Sterman 2018).

One way to increase a delay's order is to decompose a single stock into multiple general stocks so that the expected shape of the delay distribution can be matched statistically. In a scenario where delays are clearly not negatively exponentially distributed, one can split the corresponding pair of stocks. For example, when in-vehicle travel time is distributed differently than the negative exponential distribution, one can split $X^{(3)}$ (and $Y^{(3)}$) into more chained stocks to capture delays with higher orders. The outflow rate of each stock can be calculated using the stock level divided by the overall delay as well as by the order of the delay. Figure 3 shows a series of examples of such a split, the order of which does not require each stock to have a tangible meaning but rely on empirical observations.

Figure 4 shows the corresponding outflow rate profiles (of the last stock in each stock chain) for the same delay time (10 s) with different delay orders, where the average delay for each stock is D/m and m is the total order of delay. When the order is one, it is equivalent to negative exponential distribution, which has a mode at the left-most side of the support and gradually reduces to zero. When the order is second or higher, the distribution starts from zero with zero probability and gradually increases to a peak before turning back to zero. With fixed total average delay, the higher the order, the more the delay distribution resembles a conveyor in the First-in-First-Out (FIFO) or the First-Come-First-Serve (FCFS) process.

Multi-modal delay can be approximated using multiple stock chains with different average delays and orders. When each sub-process follows an exponential distribution with the same total mean delay, it is equivalent to the Erlang distribution (Ravindran 2007). Note that when an operator changes their operational strategy such that the delay of one process is changed, the delay of other processes might be influenced as well. For example, when an operational strategy increases the average vehicle repositioning time, the average distance to pick up might decrease, as intended.

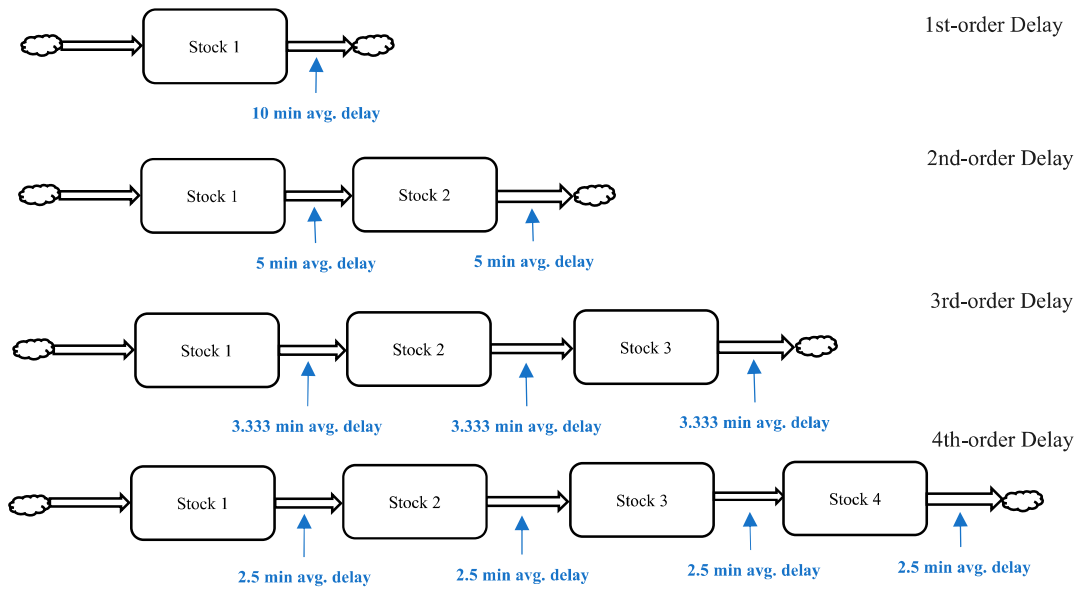


Figure 3. The delay generated by a n -stock chain for an initial unit pulse is equivalent to a n -order Erlang distribution.

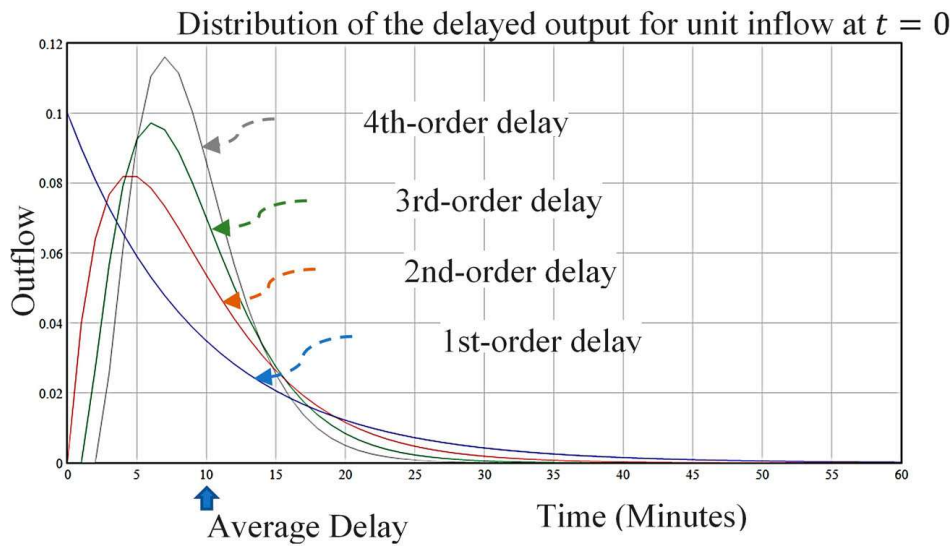


Figure 4. The delay generated by a n -stock chain for an initial unit pulse is equivalent to a n -order Erlang distribution.

The order of delays (number of stocks in a stock chain) may be estimated based on observed data, but it may also have explicit meaning. Using the high-fidelity AMOD service ordering process in (Yu and Hyland 2020) as an example, the ordering process that users experience can be considered explicitly through the stock-flow diagram by disaggregating the stock 'Waiting to be matched' as a chain of stocks shown in Figure 5. Note that customers may reject the service during the process and can be modelled as outflows from stocks (not shown in the figure). Individual vehicles may also leave the service fleet (permanently or temporarily) and can be modelled in a similar manner.

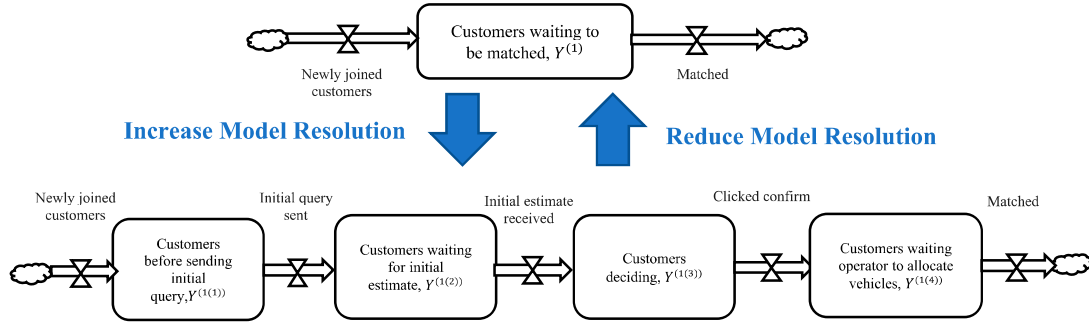


Figure 5. Disaggregation of ‘Customers waiting to be matched’ into component subprocesses and their associated delays

5. Decision variables

Section 3 introduced the vector of decision variables $\mathbf{a}$. This section specifies three major decision variables: fleet sizing, repositioning effort, and pricing. Each of these variables affects other model variables in different ways. While $\mathbf{a}$ is typically exogenous for a given co-flow model, it can become endogenous in effect when strategic planning decision-makers adjust $\mathbf{a}$ based on evaluation results for another round of simulation. Additionally, $\mathbf{a}$ can be specified as a function of co-flow model variables to allow it to adapt to changing system states during simulations. However, this level of detail is outside the scope of the present paper.

5.1. Fleet sizing

We specify fleet sizing through setting the initial values of $X_t^{(i)}$ (i.e. $X_0^{(i)}$). The total fleet size is therefore $\sum_{i \in \mathbb{I}} X_0^{(i)}$. We define a decision variable, the fleet sizing multiplier, $a^{fm,(i)}$ so that the fleet size is some multiplier of the baseline fleet size $\tilde{X}_0^{(i)}$, as shown in Equation (14). That is, the total fleet size is $a^{fm,(i)} \cdot \sum_{i \in \mathbb{I}} X_0^{(i)}$.

$$X_0^{(i)} = a^{fm,(i)} \cdot \tilde{X}_0^{(i)}, \forall i \in \mathbb{I} \quad (14)$$

5.2. Pricing

The impact of pricing multiplier, a^{pm} , on demand (customer entering rate) is modelled as.

$$x_t^{(0,1)} = \tilde{x}_t^{(0,1)} \cdot (a^{pm})^{\beta^{pr:d}} \quad (15)$$

where $\tilde{x}_t^{(0,1)}$ is the baseline demand (customer entering rate), $a^{pm} > 0$ is the pricing multiplier, and $\beta^{pr:d} < 0$ is the elasticity of demand with respect to the pricing multiplier. As $\beta^{pr:d}$ is negative, the increase in pricing leads to a decrease in demand with a diminishing effect, as intended. Note that we do not assume customers instantly respond to the pricing multiplier. Rather, we model the impact of a pricing multiplier on the general adaptation of customers to the pricing change over a period.

5.3. Repositioning effort

We measure repositioning effort using an average repositioning time multiplier, a^{rm} , with respect to the current average repositioning time made by human taxi drivers. The impact of a^{rm} on ψ is modelled as $\psi_t = \tilde{\psi}^{\varepsilon(a^{rm}; \beta^{r:\psi})}$, where $\tilde{\psi}$ is the baseline ψ_t value. We define $\varepsilon : R^+ \rightarrow R^+$ to capture the elasticity of matching efficiency to repositioning time (effort). We specify $\varepsilon(a^{rm}; \beta^{r:\psi})$ as $(a^{rm})^{\beta^{r:\psi}}$, such that:

$$\psi = \tilde{\psi}^{(a^{rm})^{\beta^{r:\psi}}} \quad (16)$$

As $\beta^{r:\psi}$ is negative, ε is a decreasing function of a^{rm} . As $\psi_0 \in (0, 1]$, the decreased ε leads to increase of ψ , which is a desired property since the additional relocation efforts should lead to increase of ψ . The diminishing return of the repositioning effort is also captured by Equation (16) as $\frac{\partial^2 \psi}{\partial (a^{rm})^2}$ is negative. The cases where vehicles are matched with customers during relocations are also captured through the diminishing return of repositioning time.

We capture the impact of a^{rm} on TT^P through a multiplicative form $TT^P = TT_0^P \cdot (a^{rm})^{\beta_{r:p}}$. Combining with Equation (13) in a multiplicative form, we have.

$$TT^P = TT_0^P \cdot (a^{rm})^{\beta_{r:p}} \cdot q(\tilde{X}_t^{(0)}; \beta_{i:p}) \cdot q(\tilde{Y}_t^{(0)}; \beta_{w:p}) \quad (17)$$

As $\beta_{r:p}$ is negative, the increase of a^{rm} leads to decrease of TT^P with diminishing return. When the AMOD fleet size is relatively small compared to the overall fleet size in a traffic network, we can assume that travel times are only determined by travel distance and, therefore, exogenous when destination choice is exogenous. However, when needed, one can capture the endogenous congestion effect using a similar multiplicative form.

6. Illustrative example

In this section, we use a simple, fictitious example to explore the mechanics of a model instantiated from the proposed framework. Suppose an AMOD fleet serving a hypothetical urban area with two possible vehicle states, namely, idling ($i = 1$) and serving customers ($i = 2$). Customers have two possible states, namely, waiting to be matched ($j = 1$) and being serviced ($j = 2$). In this simple example, we do not explicitly differentiate between customers that are matched with vehicles (but not yet picked up) and customers that have been picked up and are en-route to their destinations. We can then formulate the dynamics of the AMOD system as follows.

$$\dot{X}^{(1)} = X^{(2,1)} - X^{(1,2)} + a^{(0,1)} \quad (18)$$

$$\dot{X}^{(2)} = X^{(1,2)} - X^{(2,1)} \quad (19)$$

$$\dot{Y}^{(1)} = Y^{(0,1)} - Y^{(1,2)} = \beta^{nc} - Y^{(1,2)} \quad (20)$$

$$\dot{Y}^{(2)} = Y^{(1,2)} - Y^{(2,0)} \quad (21)$$

We formulate the transition rates in Equation (18)–(21) as follows.

$$x^{(2,1)} = X^{(2)}/tt_2 \quad (22)$$

$$x^{(1,2)} = k_{\min}/tt_1 \quad (23)$$

$$y^{(0,2)} = \beta^{nc} \quad (24)$$

$$y^{(1,2)} = k_{\min}/tt_1 \quad (25)$$

$$y^{(2,0)} = Y^{(2)}/tt_2 \quad (26)$$

$$k_{\min} = X^{(1)} \cdot (1 - \exp(-Y^{(1)}/X^{(1)})) \quad (27)$$

We can write Equation (18)–(27) in a standard state-space representation as follows (all the states are observable).

$$\frac{d\mathbf{M}}{dt} = \begin{bmatrix} \dot{X}^{(1)} \\ \dot{X}^{(2)} \\ \dot{Y}^{(1)} \\ \dot{Y}^{(2)} \end{bmatrix} = \begin{bmatrix} X^{(2)}/tt_2 - X^{(1)} \cdot (1 - \exp(-Y^{(1)}/X^{(1)}))/tt_1 + a^{(0,1)} \\ X^{(1)} \cdot (1 - \exp(-Y^{(1)}/X^{(1)}))/tt_1 - X^{(2)}/tt_2 \\ \beta^{nc} - X^{(1)} \cdot (1 - \exp(-Y^{(1)}/X^{(1)}))/tt_1 \\ X^{(1)} \cdot (1 - \exp(-Y^{(1)}/X^{(1)}))/tt_1 - Y^{(2)}/tt_2 \end{bmatrix} \quad (28)$$

The system is in equilibrium when $\frac{d\mathbf{M}}{dt} = 0$. As we have four unknown variables ($X^{(1)}, X^{(2)}, Y^{(1)}, Y^{(2)}$), we can specify their initial conditions (and specified parameters and decision variables) to obtain an analytical solution. We can linearise the system around the equilibrium point to further study the sensitivities of the system performance to initial conditions, parameters, and decision variables. However, as a common problem with linearisation, under what conditions such linearity assumption is sufficient tends to be unclear. Therefore, we adopt a simulation approach with $\tau = 0.1$ (i.e. 0.1 s per simulation step) for further investigation.

In the remainder of this section, we provide numerical solutions of Equation (28) with four different sets of initial conditions, parameters, and decision variables ($X_0^{(1)}, Y_0^{(1)}, \beta^{nc}, tt_1, tt_2$, and $a_t^{(0,1)}$) to provide insights into the model system. Specifically, we aim to illustrate the impact of different inputs on the performance of this hypothetical AMOD system.

In the first scenario, we specify the total (initial) fleet size ($X_0^{(1)}$) as 800 vehicles, $\beta^{nc} = 10$ customers/min, $tt_1 = 15$ min, $tt_2 = 35$ min, $a_t^{(0,1)} = 0$ vehs/min. Figure 6 shows the simulation results through stack plots of vehicles and customers. As we can see, after around 1000 simulation steps (100 min), the system reaches an equilibrium. However, we also notice that when the number of waiting customers is large, there is still a noticeable portion of vehicles being in the idling state. This is mainly due to modelling pick up time (tt_1) as an exogenous variable independent from the fleet size and the number of unmatched customers, instead of modelling it according to Equation (17).

In the second scenario, as Figure 6 shows a significant idling fleet throughout the simulation, we halve the fleet size; the results are displayed in Figure 7. Figure 7 shows that although there is an improvement in vehicle utilisation rate (reduced portion of $X^{(1)}$), the fleet cannot satisfy the customer inflow rate. Therefore, the total number of customers in the AMOD system grows larger and larger without converging to an equilibrium condition.

In the third scenario, we pivot from the second scenario via adding one vehicle per minute (10 simulation steps) from Step 2000 to Step 6000 – Figure 8 displays the results.

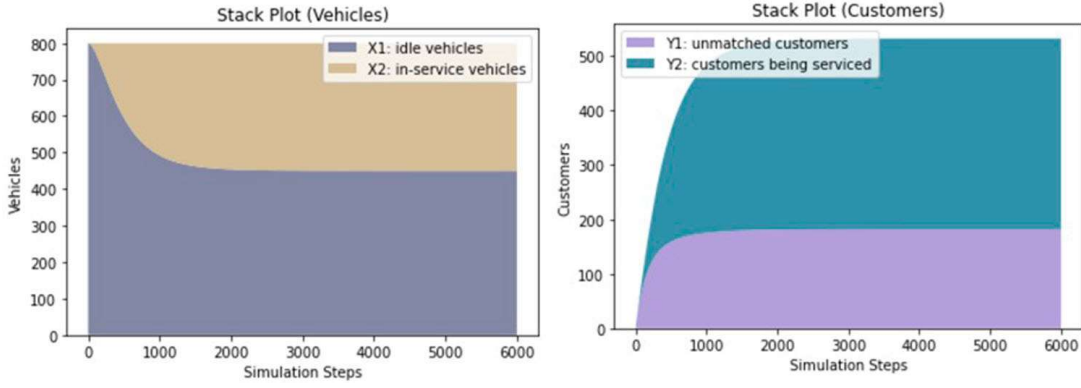


Figure 6. Scenario where $X_0^{(1)} = 1000$, $X_0^{(2)} = 0$, $Y_0^{(1)} = 0$, $Y_0^{(2)} = 0$, $\beta^{nc} = 10$, $tt_1 = 15$, $tt_2 = 35$, $a^{(0,1)} = 0$, $\tau = 0.1$. Left: stack plot of the number of idling and in-service vehicles. Right: stack plot of the number of waiting and in-service (matched or in-vehicle) customers.

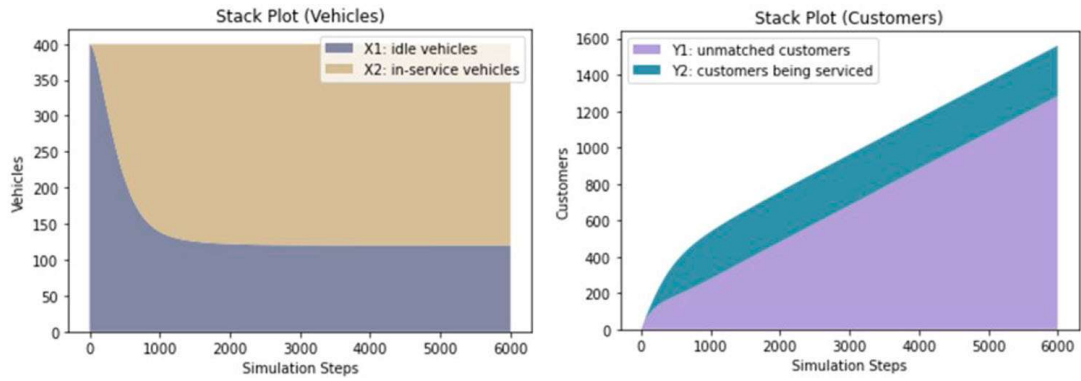


Figure 7. Scenario where $X_0^{(1)} = 400$, $X_0^{(2)} = 0$, $Y_0^{(1)} = 0$, $Y_0^{(2)} = 0$, $\beta^{nc} = 10$, $tt_1 = 15$, $tt_2 = 35$, $a^{(0,1)} = 0$, $\tau = 0.1$. Left: stack plot of the number of idling and in-service vehicles. Right: stack plot of the number of waiting and in-service (matched or in-vehicle) customers.

Note that the number of customers in the system ($Y^{(1)}$ and $Y^{(2)}$) starts to decline after the fleet size increases to a level where the total service rate surpasses the incoming customer rate. Although the ‘injection’ of additional vehicles starts at Step 2000, the total number of waiting customers and the total number of customers being served keep increasing until around Step 2800 and Step 3100, respectively. This latency from the effort of increasing vehicle size to the improvement (reduction of waiting customers) is also illustrated through the vertical dotted lines in Figure 8.

In the final scenario, we pivot from the third scenario via modifying parameters such that β^{nc} is 20 customers/min, tt_1 is 5 min, and tt_2 is 25 min. The results are shown in Figure 9. Compared to Figure 8, Figure 9 shows that $X^{(1)}$ (the number of idling vehicles) has a steeper decline initially and lower increasing rate after the ‘injection’ of new service vehicles, as expected. There is a higher portion of $Y^{(1)}$ (the number of waiting/unmatched customers), which suggests that the impact of the increased new customer entry rate is larger than the joint impacts of the reduced matching time (tt_1) and the reduced service time (tt_2). The

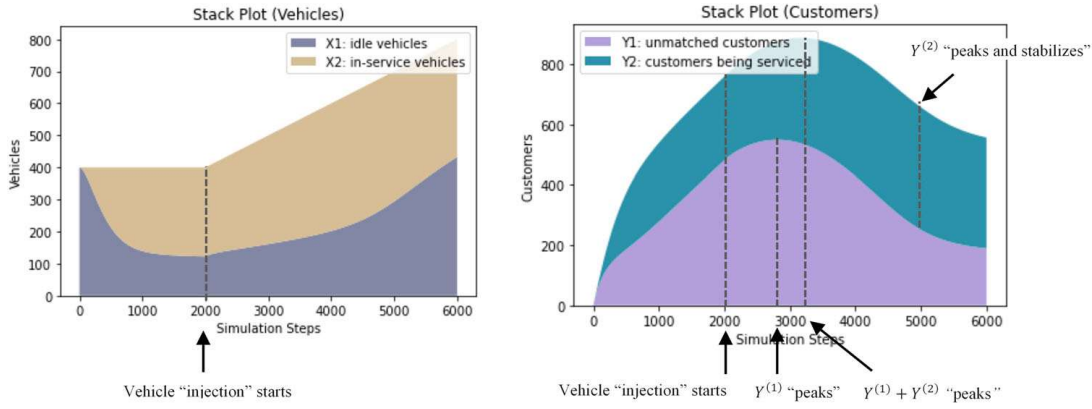


Figure 8. Scenario where $X_0^{(1)} = 400$, $X_0^{(2)} = 0$, $Y_0^{(1)} = 0$, $Y_0^{(2)} = 0$, $\beta^{nc} = 10$, $tt_1 = 15$, $tt_2 = 35$, $\tau = 0.1$, and $a^{(0,1)} = 1.0$ after the simulation step 2000 (0 otherwise). Left: stack plot of the number of idling and in-service vehicles. Right: stack plot of the number of waiting and in-service (matched or in-vehicle) customers.

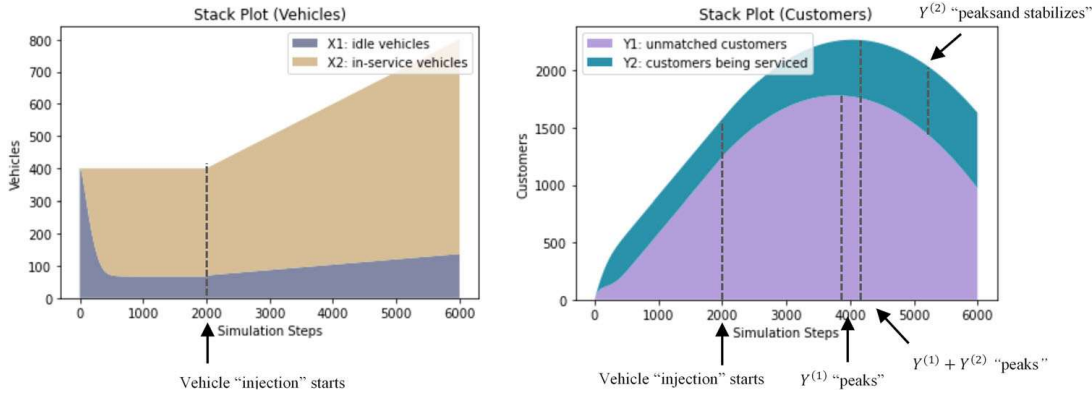


Figure 9. Scenario where $X_0^{(1)} = 400$, $X_0^{(2)} = 0$, $Y_0^{(1)} = 0$, $Y_0^{(2)} = 0$, $\beta^{nc} = 20$, $tt_1 = 5$, $tt_2 = 25$, $\tau = 0.1$, and $a^{(0,1)} = 1.0$ after the simulation step 2000 (0 otherwise). Left: stack plot of the number of idling and in-service vehicles. Right: stack plot of the number of waiting and in-service (matched or in-vehicle) customers.

latency from the effort of increasing vehicle size to the improvement (reduction of waiting customers) is greater than that in Figure 8 when comparing the vertical dotted lines in Figures 8 and 9.

7. Case study

This section presents a case study in which we combine the formulations in the previous three sections and instantiate a co-flow model for strategic planning of a hypothetical AMOD system in the City of New York's Manhattan borough (around 22.82 mi²).

7.1. Background and scope

Suppose the City of New York is interested in collaborating with an AMOD service provider to deploy centrally controlled AMOD vehicles to serve the travel demand that is currently

served by human-driven taxis in the Manhattan borough. At the current time, the city wants to evaluate the feasibility and viability of introducing an AMOD service. In this case study, we use the three key decision variables specified in Section 5: AMOD fleet size, pricing policy, and average repositioning time.

In addition to the three selected strategic decision variables, several key parameters that are uncontrollable and exogenous to the AMOD service need to be considered. These parameters include the elasticity of demand with respect to pricing, the sensitivity of matching efficiency to repositioning effort, the contribution of repositioning effort to pick-up time, the base customer request rate, and the impact of the densities of idling vehicles and unmatched waiting customers on pick-up time.

7.2. Model baseline and scenario specification

To evaluate the performance of the AMOD service, we modelled the entire Manhattan borough as a single area with homogenous customers and service vehicles. As will be demonstrated in Section 7.4, this approach effectively captures the overall dynamics of the AMOD operation. We also experimented with considering multiple traffic analysis zones instead of treating Manhattan as a single area, but we did not find a significant improvement in modelling accuracy or policy implications for strategic planning. However, for scenarios that require higher model resolution and fidelity, it is possible to increase the level of detail and consider multi-region cases.

We obtain baseline waiting time and drop-off time parameters based on the empirical taxi demand data and the Census Transportation Planning Product (CTPP) 2012–2016 data, a special tabulation of the American Community Survey data for transportation planning. In this data by PUMA5 geographies in Manhattan, the mean travel time by taxi cabs for workers (who did not work at home) residing in Manhattan is 20.8 min. For the pick-up travel time, we consider the overall size of Manhattan and assume a perfect grid network, so we multiply the distance (and average travel time) by 1.414. As human taxi drivers have experience in terms of request patterns, the taxis were not completely evenly distributed; rather, they follow the demand pattern to a certain degree. So, we multiply the distance by a factor of 0.8. We assign the initial states of all the service vehicles as idling at midnight. We do not consider vehicles that are leaving and joining the service vehicle fleet in this case study, though the idling state implicitly considers the vehicles that are not in service or accepting any service. For the function q in Equations (13) and (17), we use a lookup table with $q(0) = 2.5$, $q(0.3) = 2.4$, $q(0.6) = 2.1$, $q(0.8) = 1.7$, $q(1.0) = 1.0$, $q(10) = 0.25$, $q(100 \text{ or larger}) = 0.1$. The co-flow model obtains other values using linear interpolation. Table 1 specifies the base model structure and parameters.

Figure 10 shows the empirical taxi request data (ride request rates by time of day) from Apr. 4 (Monday) and Apr. 5 (Tuesday) of 2016. We call them Day 1 and Day 2, respectively. We use Day 1 as the base demand. To further facilitate the model development and analysis in Section 6.3, we smooth out the noise of the base demand using a moving average such that each time point is an average of the 15 min before and 15 min after that time point from the empirical data of Day 1. In Section 6.4, we will return to the original empirical data (without smoothing) from both days.

One challenge of strategic planning of the hypothetical AMOD service is parameter uncertainty. This is mainly because we are proposing a service that has little to zero

Table 1. Variables in the co-flow model.

Variable	Description	Type*	Unit	Need to specify	Initial/Base Value
$X^{(0)}$	Potential service vehicles	Stock	Vehicles	No	–
$X^{(1)}$	Idle Vehicles	Stock	Vehicles	Yes	4000
$X^{(2)}$	Vehicles en-route to pick up	Stock	Vehicles	No	0
$X^{(3)}$	Vehicles en-route to drop off	Stock	Vehicles	No	0
$X^{(4)}$	Vehicles repositioning	Stock	Vehicles	No	0
$x^{(0,1)}$	Newly joined vehicles (negative if leaving)	Ex Flow	Vehicles/Second	No	0
$x^{(1,2)}$	Vehicles matched with customers	En Flow	Vehicles/Second	No	0
$x^{(2,3)}$	Vehicles that picked up customers	En Flow	Vehicles/Second	No	0
$x^{(3,4)}$	Vehicles finished dropping off customers	En Flow	Vehicles/Second	No	0
$x^{(4,1)}$	Vehicles relocating	En Flow	Vehicles/ Second	No	0
$Y^{(0)}$	Potential customers	Stock	Customers	Yes	–
$Y^{(1)}$	Customers waiting to be matched	Stock	Customers	No	0
y^2	Matched customers waiting to be picked up	Stock	Customers	No	0
$y^{(3)}$	In-vehicle customers	Stock	Customers	No	0
$y^{(0,1)}$	Newly joined customers	Ex Flow	Customers/Second	No	0
$y^{(1,2)}$	Customers matched with vehicles	En Flow	Customers/Second	No	0
$y^{(2,3)}$	Customers picked up by vehicles	En Flow	Customers/Second	No	0
$y^{(3,0)}$	Customers served	En Flow	Customers/Second	No	0
k	Minimum of $X^{(1)}$ and $Y^{(1)}$	En Auxiliary	Customers (Vehicles)	No	–
ψ	Percent of minimum of idle service vehicles and unmatched customers to be paired in a time step.	Ex Auxiliary	Percent/Second	Yes	65
TT^p	Average travel time to pick up customers	Ex Auxiliary	Second	Yes	130
TT^d	Average travel time to drop off customers	Ex Auxiliary	Second	Yes	600
TT^r	Average reposition time	Ex Auxiliary	Second	Yes	60
τ	Time step (for simulation)	–	Unitless	Yes	1.0

*Ex: Exogenous; En: Endogenous.

observed data. So, we derive or infer these parameters using empirical data of existing taxi services specified in the previous subsection and professional judgment. To better understand what information we should collect, and which variables or parameters have minor impacts on strategic planning, we conducted sensitivity analyses, and the results are presented in the next subsection. Table 2 outlines the ranges for the three decision variables and the six behavioural/environmental parameters, along with their corresponding baseline values.

The model is coded in Python 3.8.5, aided by a PySD package (version 3.9.0). Using a computer with Intel(R) Core(TM) i7-1065G7 CPU @ 1.30 GHz 1.50 GHz and RAM 16.0 GB (15.7 GB usable), one simulation of a whole day takes around 20 s (including the input request data loading time).

7.3. Results and analysis

This subsection analyzes the impact of the three decision variables and six behavioural/environmental parameters on four selected performance metrics: the customer wait time to be matched, the customer wait time to be picked up (after matched), the percentage (or portion) of empty vehicle distance, and vehicle utilisation (in terms of whether the vehicles are

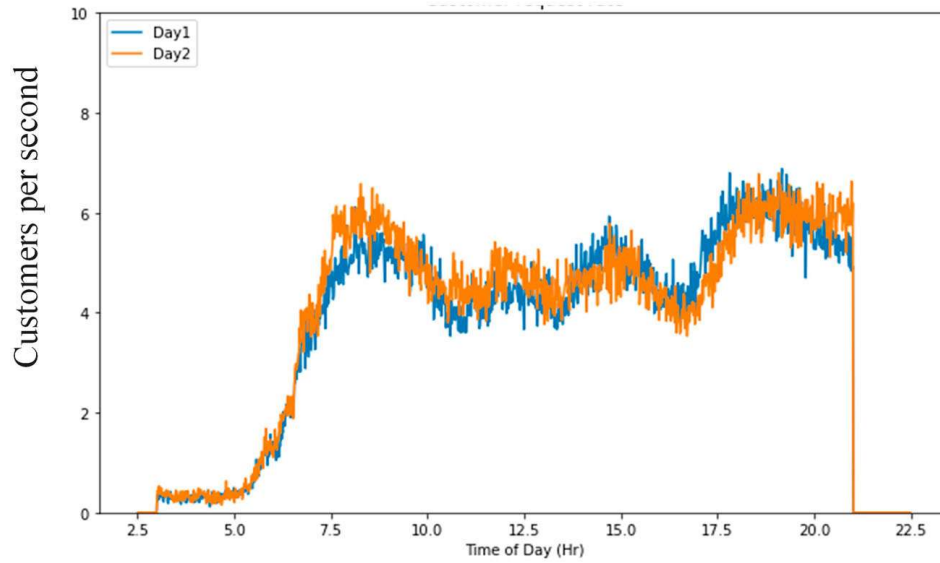


Figure 10. Manhattan taxi request rates: Apr. 4 (Monday – Day 1) and Apr. 5 (Tuesday – Day 2)

Table 2. Summary of variables that need to be initialised, calibrated, or externally provided.

Variable type	Notation	Description	Base value	Test range
Decision variables	a^{fm}	Total fleet size multiplier	1.0	[0.5, 2.5]
	a^{pm}	Pricing multiplier	1.0	[0.2, 3.5]
	a^{rm}	Repositioning time multiplier	1.0	[0.1, 2.5]
Behavioural/Environmental parameters	$\beta^{pr:d}$	Elasticity of demand with respect to pricing multiplier	−0.2	[−1.5, 0]
	$\beta^{r:\psi}$	Sensitivity of matching efficiency to repositioning time	1.0	[0.1, 1.5]
	$\beta^{r:p}$	Elasticity of pick-up time with respect to repositioning time multiplier	−0.2	[−1.0, −0.1]
	β^{dm}	Demand multiplier	1.0	[0.7, 2.5]
	$\beta^{i:p}$	Elasticity of pickup time with respect to idling fleet size	0.025	[0, 0.05]
	$\beta^{w:p}$	Elasticity of pickup time with respect to number of customers waiting to be matched	0.025	[0, 0.05]

en-route to pick-up and drop-off customers). In Section 6.3.1, we examine the impact of the decision variables on the system performance under random parameters. We then show the change in the AMOD system dynamics in response to different scenarios in Section 6.3.2 and provide decision implications in Section 6.3.3.

7.3.1. Performance by decision variables under uncertainty

We simulate 1250 scenarios, where the values of the six parameters are randomly generated from the triangular distributions. To be more specific, as we are uncertain about the exact values of the behavioural and environmental parameters specified in Table 2 and an AMOD operator does not have direct control over them, we assume that these parameters follow independent, symmetrical triangular distributions whose lower and upper bounds are specified in Table 2. We use Monte Carlo simulations to obtain the distributions of the four selected metrics for each combination of the three decision variables. Based on the

central limit theorem, when we increase simulation samples, the distribution of a metric θ should asymptotically approach a normal distribution with the standard deviation of (sample) mean $\bar{\theta}$ and (sample) variance $\frac{\sum_{n=1}^N (\theta_n - \bar{\theta})^2}{N-1}$.

We select 27 combinations of decision variables (including their scenario IDs) and show their performance metrics in Table 3. The main evaluation results include the (sample) means and the (sample) standard deviations (std.).

We plot the four metrics of the 27 selected scenarios in Figures 11–14, respectively. In Figure 11, we observe the general pattern where higher fleet sizes and higher pricing multiplier leads to lower waiting time for matching with customers. However, within each fleet size, we see that the impact of the repositioning effort is nonlinear. When the fleet size is 3500 and $a^{pm} = 0.8$, the highest wait time occurs when $\alpha^{rm} = 0.6$, but when the fleet size is 4000 and $a^{pm} = 1.0$, the highest wait time occurs when $\alpha^{rm} = 1.4$. This rank reversal (i.e. the ordinal ranking of different alternatives of α^{rm} changes when a^{pm} and a^{fm} change) indicates that the impact of repositioning effort depends on the relationship between customer demand and vehicle supply. Although the rank reversal seems robust given the relatively large simulation samples, it is indeed possible that the ranking reversal is caused by sampling errors, and we leave it for future research. Besides, we observe that the standard deviations (due to random parameters) tend to decrease with increases in fleet size. This is consistent with our intuition that more vehicle supply tends to increase the robustness (reliability) in terms of responding to a surge of demand.

In Figure 12, we observe the general pattern where lower repositioning effort tends to lead to higher wait time for picking up (matched) customers, while the impact of higher fleet size and higher pricing multiplier are mixed/insignificant. One possible explanation for this mixed/insignificant pattern could be the already low average wait time to be picked up, as compared to the average wait time to be matched. The impact of repositioning effort might become more significant when the vehicle fleet small relative to demand or when we study a much wider range of customer request intensities.

In Figure 13, we can observe that higher fleet size and greater repositioning effort tend to lead to higher portion of empty distance, though their effects are moderate. Indeed, optimising fleet size and repositing effort can be seen as making a tradeoff between higher operating cost for shorter average customer wait time when jointly examined with Figures 7 and 8. Besides, we observe that the standard deviations of the estimated portion of empty distance (due to random parameters) tend to slightly increase with the increase of fleet size. The impact of pricing multiplier appears insignificant.

In Figure 14, as expected, we observe that higher fleet size, higher pricing multiplier, and higher repositioning effort tend to decrease fleet utilisation. In study, a vehicle is being utilised when it is either in the state of picking up customers or in the state of dropping off customers. Figure 10 shows that the standard deviations of the estimated portion of fleet utilisation (due to random parameters) tend to increase with the increase of fleet size.

Overall, Table 5(a) and Figures 11–14 show clear tradeoffs among customer wait time and operational cost. Higher fleet size leads to lower wait time to be matched but not necessary wait time to be picked up. Higher fleet size also tends to increase empty travel distance and decrease fleet utilisation, as expected. Higher pricing leads to shorter waiting time but fewer customers entering the system. Repositioning effort tends to have smaller impact than the other two decision variables in general, especially when vehicles are insufficient

Table 3. Performance of the 27 scenarios (decision variables varying across scenarios) under random parameters.

d^m	d^{pm}	d^m	Scenario ID	Avg. wait time to match (std.)	Avg. time to pick up (std.)	Avg. % empty distance (std.)	Avg. % fleet utilisation (std.)
0.875 (3500 vehs)	0.8	0.6	1	87.05 (24.17)	2.35 (0.30)	0.27 (0.02)	0.82 (0.01)
		1	2	83.36 (46.18)	2.05 (0.33)	0.27 (0.01)	0.78 (0.04)
		1.4	3	68.47 (37.06)	2.00 (0.25)	0.28 (0.01)	0.75 (0.05)
		0.6	4	39.38 (43.66)	2.57 (0.37)	0.26 (0.01)	0.79 (0.04)
	1.0	1	5	37.21 (43.12)	2.48 (0.31)	0.28 (0.01)	0.74 (0.06)
		1.4	6	39.71 (33.92)	2.17 (0.31)	0.28 (0.02)	0.73 (0.04)
		0.6	7	25.90 (23.31)	2.81 (0.31)	0.27 (0.02)	0.78 (0.04)
		1	8	24.96 (22.46)	2.35 (0.37)	0.27 (0.02)	0.76 (0.03)
1.000 (4000 vehs)	0.8	1.4	9	41.13 (26.58)	2.15 (0.26)	0.28 (0.01)	0.75 (0.02)
		0.6	10	26.93 (24.01)	2.84 (0.46)	0.27 (0.02)	0.77 (0.06)
		1	11	25.19 (31.33)	2.46 (0.42)	0.28 (0.02)	0.76 (0.02)
		1.4	12	32.66 (32.52)	2.18 (0.26)	0.28 (0.01)	0.73 (0.06)
	1	0.6	13	12.72 (25.47)	2.87 (0.34)	0.27 (0.02)	0.73 (0.05)
		1	14	17.20 (21.52)	2.63 (0.35)	0.28 (0.02)	0.70 (0.08)
		1.4	15	19.19 (23.05)	2.28 (0.32)	0.29 (0.02)	0.70 (0.06)
		0.6	16	10.58 (15.52)	2.96 (0.37)	0.28 (0.02)	0.74 (0.06)
	1.2	1	17	6.67 (11.61)	2.65 (0.31)	0.28 (0.02)	0.68 (0.07)
		1.4	18	7.72 (10.41)	2.37 (0.28)	0.29 (0.01)	0.68 (0.07)
		0.6	19	9.26 (5.40)	2.83 (0.30)	0.27 (0.02)	0.73 (0.06)
		1	20	5.76 (5.19)	2.70 (0.28)	0.28 (0.02)	0.67 (0.09)
1.125 (4500 vehs)	0.8	1.4	21	8.05 (6.04)	2.45 (0.23)	0.29 (0.01)	0.66 (0.07)
		0.6	22	6.43 (7.84)	2.88 (0.36)	0.27 (0.02)	0.72 (0.10)
		1	23	2.97 (6.22)	2.53 (0.37)	0.27 (0.02)	0.66 (0.08)
		1.4	24	5.76 (8.61)	2.56 (0.34)	0.3 (0.02)	0.66 (0.09)
	1.0	0.6	25	1.78 (2.24)	2.82 (0.43)	0.27 (0.02)	0.70 (0.06)
		1	26	1.83 (3.19)	2.40 (0.25)	0.27 (0.01)	0.64 (0.09)
		1.4	27	1.41 (2.05)	2.40 (0.22)	0.29 (0.01)	0.63 (0.06)
		0.6					
	1.2	1					
		1.4					
		0.6					
		1					
		1.4					

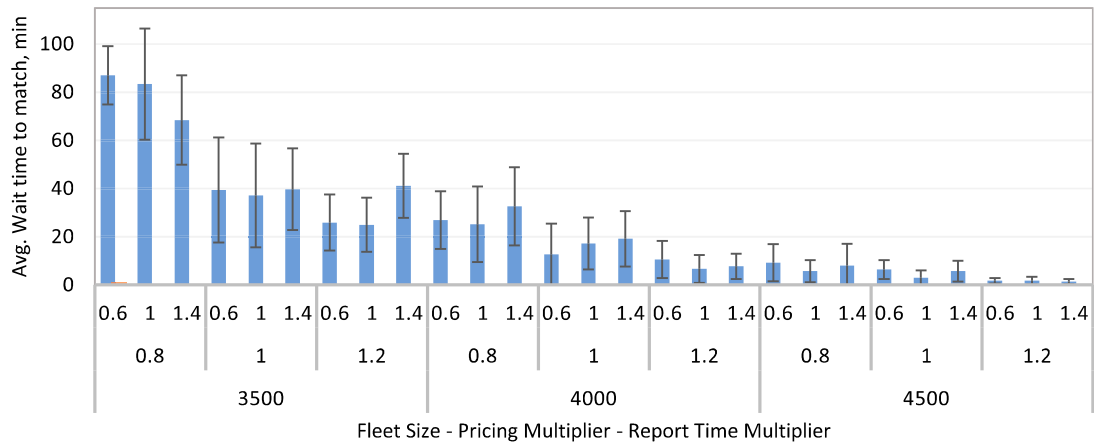


Figure 11. Simulated mean customer wait time to match (sample standard deviation displayed with bars).

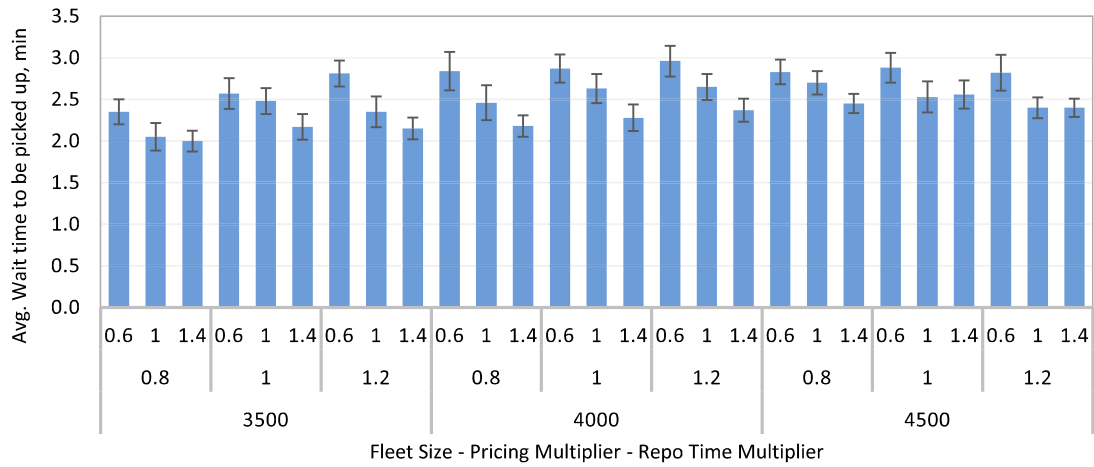


Figure 12. Simulated mean customer wait time to be picked up, after matched (sample standard deviation displayed with bars).

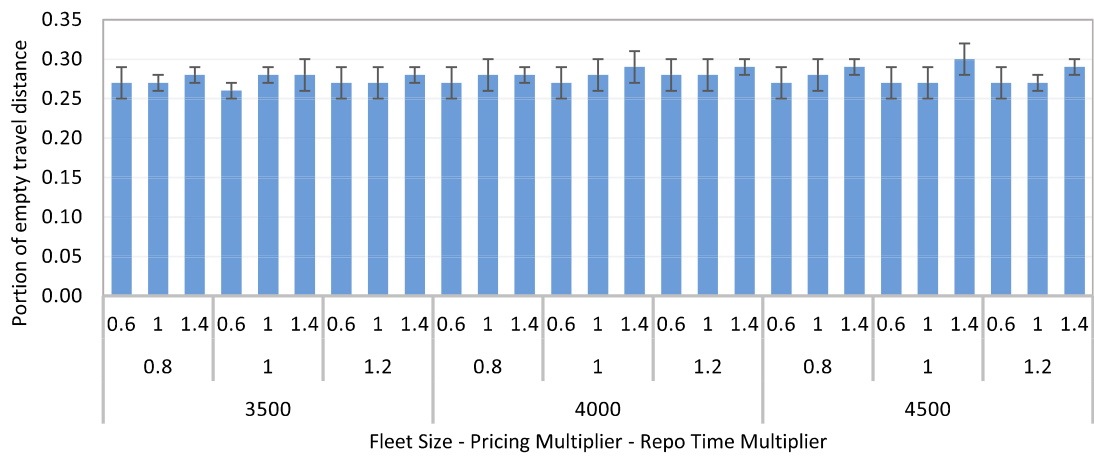


Figure 13. Simulated mean of the portion of empty travel distance (sample standard deviation displayed with bars).

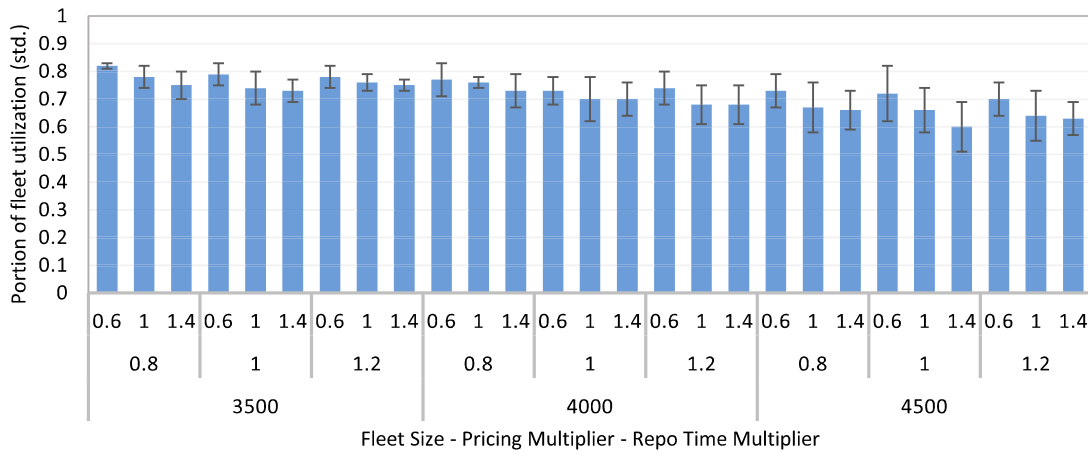


Figure 14. Simulated mean of fleet utilisation (sample standard deviation displayed with bars).

relative to the customer requests. Additionally, repositioning efforts seem to also show non-linear relationships with various system performance metrics. The optimal combination of the three decision variables depends on the specific weights that the strategic planning decision-makers determine for the various performance metrics.

7.3.2. Regression analysis of performance by decision variables and parameters

So far, we have examined the impact of decision variables under uncertain parameters. We are also interested in the joint impact of decision variables and parameters. As we find that the impact of decision variables is in general monotonic, we use linear regression to model the relationship between decision variables and uncertain parameters on performance metrics, based on the results of all 1250 simulations.

We estimate four linear regressions (one for each performance metric). The regression results are summarised in Table 4. As the decision variables and parameters have different units, we scale them in terms of their standard deviation.

The adjusted R^2 value and the correlation coefficient ('Multiple R') show that linear regressions produce high goodness-of-fits for the four performance metrics. The signs of the coefficients show reasonable directions. For example, the positive coefficient of a_3 for the average wait time to pick up suggests that increasing repositioning effort tends to reduce the wait time for matched customers, which is consistent with the intuition. For another example, that $\beta^{i:p}$ and $\beta^{w:p}$ are negative for the average percentage of empty distance is consistent with our intuition as the increase of idling fleet size and unmatched customers tend to reduce the pick-up and drop-off distance, and hence shorter empty distance.

Larger scaled regression coefficients indicate greater importance in two senses: (1) the associated parameter or decision variables impacts the corresponding performance metrics more, and therefore (2) modellers or decision-makers should put more effort into reducing the uncertainty associated with a parameter's impact on performance. Consistent with the finding from the previous section, repositioning does not influence the performance metrics as significantly as the other two decision variables, especially when the fleet is a scarce resource.

Table 4. Regression results.

	Avg. wait time to match			Avg. time to pick up			Avg. % empty distance			Avg. % fleet utilisation		
	Coef [^]	SD [~]	Scaled [*]	Coef	SD	Scaled	Coef	SD	Scaled	Coef	SD	Scaled
Intercept	101.81	7.26	14.02	2.17	0.17	13.03	0.25	0.01	43.81	0.78	0.01	58.71
a_1	−0.06	0.00	−80.26	0.00	0.00	54.07	0.00	0.00	40.95	0.00	0.00	−39.04
a_2	−44.90	1.93	−23.31	0.76	0.04	17.27	0.02	0.00	14.61	−0.05	0.00	−14.09
a_3	3.30	1.92	1.72	−0.87	0.04	−19.80	0.00	0.00	3.10	−0.06	0.00	−17.51
$\beta^{pr:d}$	−15.11	9.00	−1.68	0.15	0.21	0.72	0.00	0.01	0.37	0.01	0.02	0.50
$\beta^{r:\psi}$	1.43	1.34	1.06	−0.04	0.03	−1.18	0.00	0.00	−1.28	0.00	0.00	−0.88
$\beta^{r:p}$	−4.17	13.13	−0.32	−0.25	0.30	−0.84	−0.01	0.01	−1.15	−0.02	0.02	−0.81
β^{dm}	209.43	4.48	46.77	−3.86	0.10	−37.64	−0.11	0.00	−32.96	0.25	0.01	30.82
β^{ip}	−269.26	13.47	−19.98	−14.56	0.31	−47.12	−0.79	0.01	−75.20	−0.59	0.02	−24.15
$\beta^{w:p}$	2.06	12.98	0.16	−0.26	0.30	−0.89	−0.01	0.01	−0.82	−0.02	0.02	−0.89
Adjusted R^2	0.88			0.85			0.87			0.74		
Corr Coef ^{&}	0.94			0.92			0.93			0.86		
Total SD	19.18			0.44			0.01			0.03		
Observations	1250			1250			1250			1250		

[^]: Coefficients.[~]: Standard error.^{*}: Scaled coefficient by standard error.[&]: Correlation Coefficient.**Table 5(a).** Vehicle-centric data summary.

Day	Total vehicles	Repo	Avg Idle Vehicles (std)	Avg En-Route Vehicles (std)	Avg Drop-Off Vehicles (std)	Avg Unmatched Users (std)	Avg Matched Users (std)	Avg In-Vehicle Users (std)
4 April 2016	4,000	No	1,521 (1,352)	430 (304)	2,051 (1,116)	397 (883)	429 (304)	2,049 (1,117)
4 April 2016	4,000	Yes	1,482 (1,279)	396 (237)	2,049 (1,098)	460 (1,038)	396 (237)	2,049 (1,098)
5 April 2016	4,000	No	1,397 (1,295)	399 (226)	2,140 (1,130)	465 (1,035)	399 (226)	2,140 (1,130)
5 April 2016	4,000	Yes	1,443 (1,358)	417 (306)	2,140 (1,145)	465 (1,012)	417 (306)	2,140 (1,145)

Table 5(b). Customer-centric data summary.

Day	Total Demand	Repo	Avg Request Rate Per Sec (std)	Avg Match Time in Sec (std)	Avg Pick-Up Time in Sec (std)	Avg IVTT in Sec (std)
4 April 2016	263,152	No	3.66 (2.13)	102 (198)	226 (262)	560 (373)
4 April 2016	263,152	Yes	3.66 (2.13)	119 (233)	234 (274)	560 (373)
5 April 2016	269,859	No	3.75 (2.17)	117 (232)	245 (296)	571 (378)
5 April 2016	269,859	Yes	3.75 (2.17)	117 (239)	230 (284)	571 (378)

7.3.3. Dynamics of system behaviors

The ability to examine the impact of strategic planning decisions on the within-day dynamics of AMOD operations is a key differentiator of our model compared with existing models in the literature. In this subsection, we first compare the dynamics under different decision variables given baseline parameter values, and then we study the impact of random parameters given a selected set of decision variables.

Figure 15(a) shows the impact of (only) changing total fleet size on the dynamics of idling vehicles with baseline values for all other decision variables and parameters. The pattern is consistent with our intuition that when the total fleet size is small, the idling vehicles tend

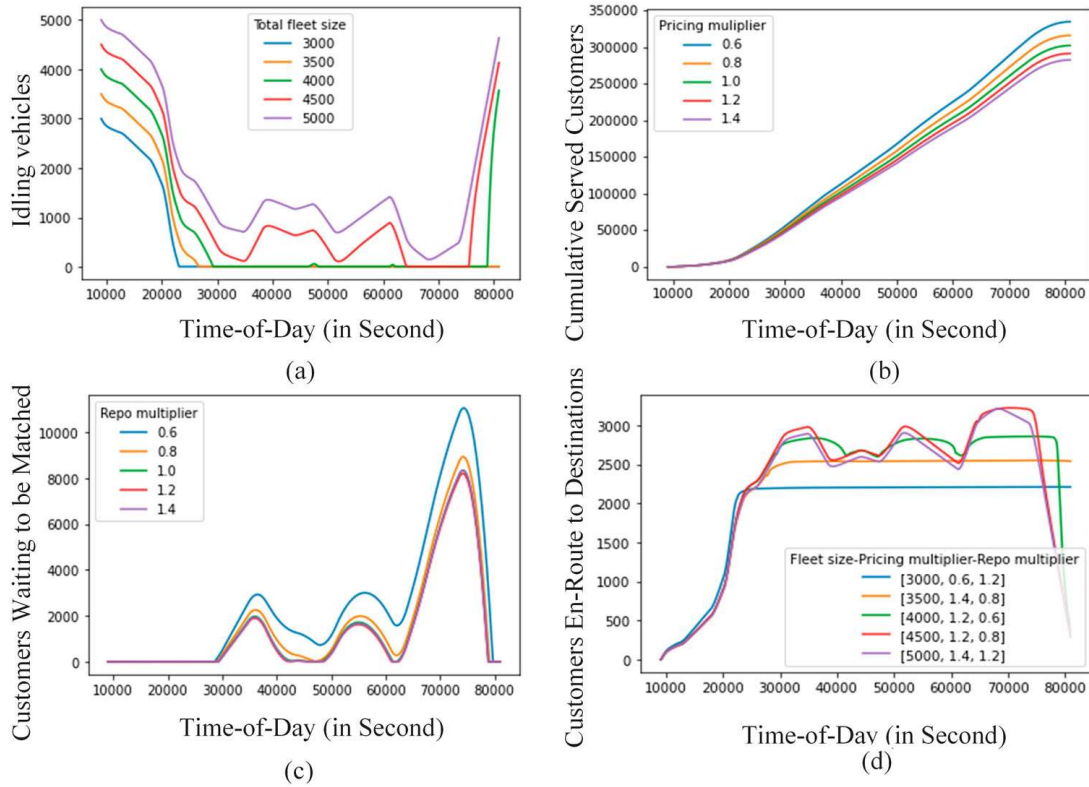


Figure 15. Impact of decision variables on within-day AMOD system dynamics. (a) Total fleet size on idling fleet size. (b) Pricing multiplier on cumulative customer served. (c) Repositioning multiplier on customers waiting to be matched. (d) Joint impact of three decision variables on customers en-route to destination.

to be low. When the total fleet size is over 4000, the system starts to have idling vehicles in the afternoon peak period.

Figure 15(b) shows the impact of (only) changing the pricing multiplier on the dynamics of the cumulation of served customers. The observed pattern aligns with our intuition that as the pricing multiplier increases (indicating higher prices for customers), the cumulative number of customers tends to decrease.

Figure 15(c) illustrates the impact of (only) varying the repositioning multiplier on the dynamics of the customers waiting to be matched. The pattern in the graph indicates that the effort has diminishing effectiveness in reducing customers waiting to be matched.

In Figure 15(d), we observe the impact of simultaneously perturbing the total fleet size, pricing multiplier, and repositioning multiplier on the number of customers en-route to their destination. The results indicate that when the fleet size is low (3000 and 3500), vehicles are consistently scarce resources that do not vary with the fluctuation of customers, regardless of the values of the other two decision variables. Conversely, when the fleet size is high (4000, 4500, 5000), the vehicles begin to experience periods of slackness. The observed pattern is consistent across different values of the pricing and repositioning multipliers.

Next, we fix the base decision setting and explore the impact of behavioural and environmental variables on the dynamics of the AMOD service. Combining the time profiles in the present subsection and the performance metrics in the previous subsections, we find that

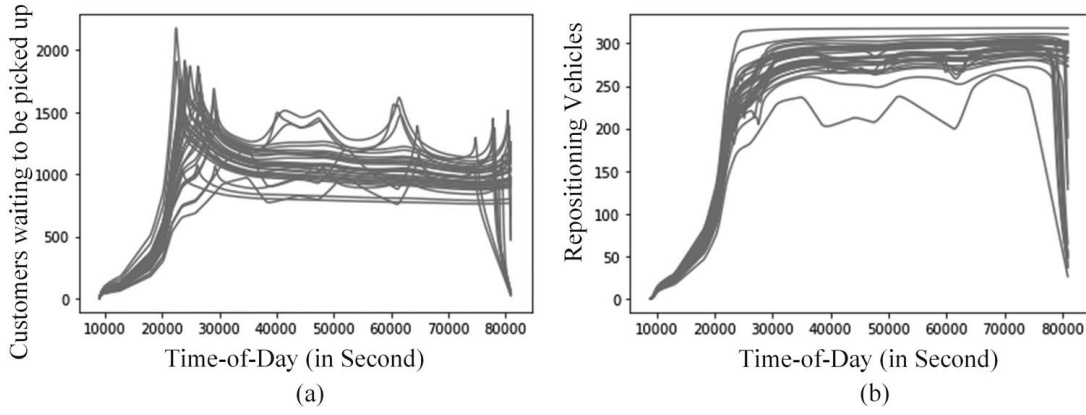


Figure 16. Impact of behaviour and environmental parameters on the within-day AMOD system dynamics (30 simulation instances). (a) Customers waiting to be picked. (b) Vehicles being repositioned.

the decision setting of 4000, 1.2, and 1.0 for fleet size, pricing multiplier, and repositioning efforts, respectively, achieves a reasonable balance among the four selected performance metrics. To understand the impacts of uncertain parameters on the dynamics of selected stock variables, we can visualise the simulated time profiles of a set of decision variables with random parameters. As a demonstration, we show 30 simulation instances through the two plots in Figure 16: the left is the number of customers waiting to be picked up (after matched), and the right is for the number of vehicles in the state of repositioning. The figure shows that the overall shapes of the simulated time profiles in each graph remain stable despite the variation of parameters.

7.3.4. Discussion

Overall, the evaluation results demonstrate a trade-off between reducing customer wait time and operational costs. However, the relationship between each decision variable and its impact on system performance is not necessarily linear due to the complex interactions with other decision variables and parameters over time during a simulation period. This nonlinearity highlights the value of using a dynamic system model for evaluation.

The results also demonstrate that the model can capture a wide range of scenarios, which allows decision-makers to explore different scenarios and identify areas where further data collection could reduce uncertainty about selected performance metrics. However, it is worth noticing that the model primarily provides strategy-level evaluation and insights into how changes in decision variables and parameters affect the time profile shapes of the system states.

7.4. Comparison with agent-based modeling approach

To gain an understanding of how well a co-flow approach can capture the system-level AMOD service performance, we compare the model results with those from a high-resolution, high-fidelity ABM developed in an earlier study by Dandl et al. (2019) for operational analysis. Although we know that the outputs from the ABM assume 4000 as the fleet size (without within-day fleet size adjustment), and the model covers the same area and timeframe to those in this case study, the two models do not have perfect one-on-one

correspondence as to most variables. Additionally, the high-resolution, high-fidelity ABM does not explicitly consider the impact of pricing. Therefore, we first explore the outputs from the ABM. Then we examine which of the scenarios (if any) in Section 6.4 will produce outputs similar to those from the ABM.

Tables 5(a) and 5(b) summarise the outputs of four ABM model runs. The first two scenarios use the customer demand data from 4 April 2016, with and without repositioning ('Repo'), while the third and fourth scenarios use the customer demand data for 5 Apr 2016, with and without repositioning. The ABM produces vehicle and passenger (location-time) trajectories in units of seconds. The trajectory data also includes the status of each vehicle and customer at each time step of the simulation. Hence, it is easy to obtain the number of idle vehicles, en-route pickup vehicles, en-route drop-off vehicles, and repositioning vehicles at every time step from the ABM, which are the key information as to comparing the co-flow mode outputs with those from the ABM.

We use two main types of goodness-of-fit measure. The first type, ε_A , is developed based on the mean absolute percent error (MAPE),

$$\varepsilon_A = \frac{1}{N} \sum_{t_s \in \mathbb{T}_s} \left(\sum_{i \in \mathbb{I}} \frac{|\hat{X}_{t_s}^{(i)} - X_{t_s}^{(i)}|}{X_{t_s}^{(i)} + 1} + \sum_{j \in \mathbb{J}} \frac{|\hat{Y}_{t_s}^{(j)} - Y_{t_s}^{(j)}|}{Y_{t_s}^{(j)} + 1} \right) \quad (13)$$

where t_s represents the sampled time steps from set $\mathbb{T}_s$. Here we sample system states every 30 min (1800 s). $N = 200$ is the total number of samples – 40 sampled time steps and each step measures five state variables. Since we do not consider share rides, $X^{(2)}$ and $X^{(3)}$ are always equal to $Y^{(2)}$ and $Y^{(3)}$, respectively. Therefore, we only considered $X^{(1)}$, $X^{(2)}$, $X^{(3)}$, $X^{(4)}$, $Y^{(1)}$ in calculating MAPE to avoid double counting.

The second type, ε_B , is root mean square error (RMSE) based, which penalises large errors more than small ones. To obtain ε_B , we first calculate a Chi-square based cost function as:

$$\chi^2 = \sum_{t_s \in \mathbb{T}_s} \left(\sum_{i \in \mathbb{I}} \frac{(\hat{X}_{t_s}^{(i)} - X_{t_s}^{(i)})^2}{X_{t_s}^{(i)} + 1} + \sum_{j \in \mathbb{J}} \frac{(\hat{Y}_{t_s}^{(j)} - Y_{t_s}^{(j)})^2}{Y_{t_s}^{(j)} + 1} \right) \quad (14)$$

where $\mathbb{I}$ is the complete set for all the vehicle-related stock variables, and $\mathbb{J}$ is the set for all the customer-related stock variables. $X^{(i)}$ and $\hat{X}^{(i)}$ are the modelled and observed vehicle-related variable values, respectively. Similar meanings apply to $Y^{(i)}$ and $\hat{Y}^{(i)}$, but for customer-related variable values. To avoid dividing by zero, we add one unit to each stock variable. This does not influence the monotonicity of χ^2 as a measure of goodness of fit. The relative weight for each term can be further added when, say, a researcher cares for the goodness-of-fit of certain time intervals more than others. Additionally, autoregressive correction can also be applied to further improve the goodness-of-fit metric.

We further process the evaluated Chi-square-based cost function to obtain the root mean-square error (RMSE) measure. We can obtain the adjusted root mean square error (RMSE) of a parametric approximation through the following equation:

$$\varepsilon_B = \sqrt{\frac{\max(\chi^2 - df, 0)}{df(N - 1)}} \quad (15)$$

Table 6. Goodness-of-Fit measures for training and validation data sets.

		Day 1 demand		Day 2 demand	
		No reposition	With reposition	No reposition	With reposition
Model	ε_A (MAPE based)	0.378	0.280	0.430	0.323
	ε_B (RMSE based)	0.586	0.547	1.183	1.256
	ε_{B_0} (RMSE based)	0.581	0.543	1.174	1.246
Constant	ε_A (MAPE based)	37.70	36.84	36.02	38.98
	ε_{B_0} (RMSE based)	16.13	14.36	16.46	16.34

where $df = N - k$, and k is the number of parameters. As a reference, the unadjusted RMSE, which does not consider the number of parameters, is displayed in Equation (16).

$$\varepsilon_{B_0} = \frac{\sqrt{\max(\chi^2 - N + 1, 0)}}{N - 1} \quad (16)$$

As discussed in Section 3.5, numerous methods (e.g. Bayesian optimisation) are available to estimate the model parameters. As the model in the case study is relatively simple (i.e. it includes only a few parameters), we adopt a grid search approach to minimise MAPE wherein we constrain each parameter to be within the range defined in Table 2. We selected the parameters that minimised the error defined in the form of MAPE in Equation (13) though we also reported RMSE defined in Equation (14). We find that the following co-flow model scenario produces similar outputs, including the system dynamics: decision variables a^{fm} , a^{pm} , and a^{rm} are 1.0 (i.e. 4000 vehicles), 1.0, and 0.8, respectively and the six parameters ($\beta^{pr:d}$, $\beta^{r:\psi}$, $\beta^{r:p}$, β^{dm} , $\beta^{i:p}$, $\beta^{w:p}$) are -0.3 , 1.0 , -0.2 , 1.0 , 0.0025 , 0.025 , respectively. Note that, as the three decision variables are known from the ABM, the grid search in these three dimensions are confined by the three fixed values. In the present example, we use variables such as total number of en-route vehicles and total number of customers waiting to be matched for the estimation, which we believe are readily available by any MOD operators, so we believe that one can easily apply the same estimation approach using empirical data.

To gain more confidence about the ability of the co-flow model in obtaining reasonably accurate system dynamics produced by the ABM, we evaluate the co-flow model's system dynamics against the ABM's system dynamics on April 5th (Day 2), where we use the same co-flow model decision variables and parameters calibrated based on Day 1. What varies in the co-flow model from Day 1 to Day 2 is the input data for $y^{(0,1)}$.

Table 6 displays the goodness-of-fit measures for both Day 1 (training day) and Day 2 (testing day). The results show that the co-flow model performs significantly better than the constant daily average prediction.

Figures 17 and 18 visualise the stacked time series data for the customer-related stock variables – total unmatched customers, total (matched) customers to be picked up, and total customers in service vehicles. At each time step, the sum of these three stock variables provides the number of customers in the AMOD service system. The figures show that the co-flow model captures the dynamics generated by the ABM with high accuracy but with significantly fewer model parameters. Although some variabilities from the ABM are smoothed out, the key patterns, such as the three peaks in the morning (around 8 am), afternoon (around 3 pm), and evening (around 8 pm) for all three customer states, are preserved.

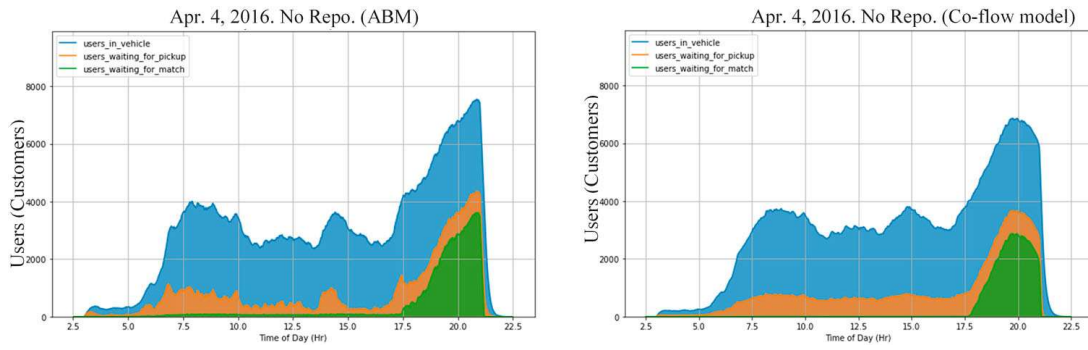


Figure 17. Stacked time profile of the *customer* stock variables in the case *without* vehicle repositioning for the *training* dataset, for the ABM (left) and co-flow model (right).

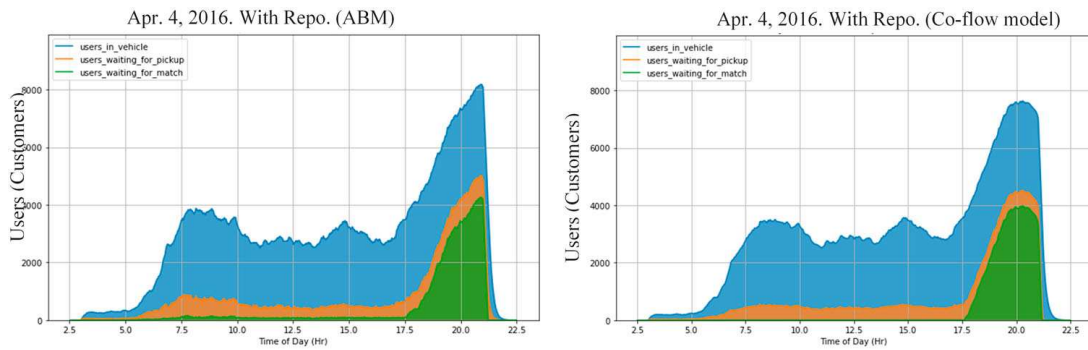


Figure 18. Stacked time profile of the *customer* stock variables *with* vehicle repositioning for the *training* dataset, for the ABM (left) and co-flow model (right).

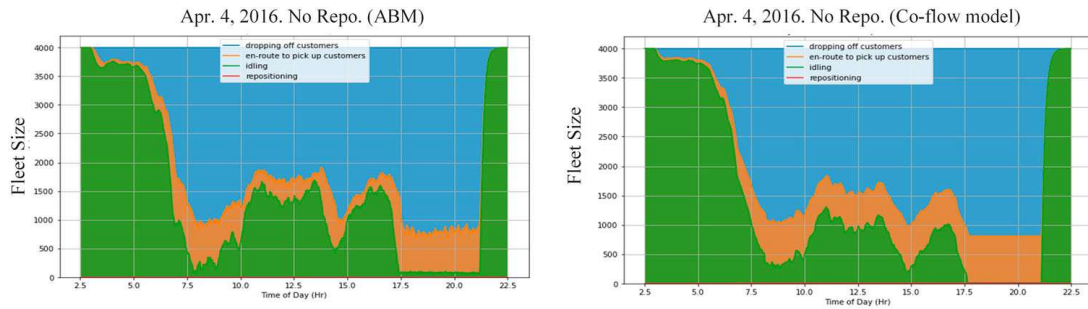


Figure 19. Stacked time profiles of the *service vehicle* stock variables *without* vehicle repositioning, for the *training* data set, for the ABM (left) and co-flow model (right).

Figures 19 and 20 visualise the stacked time series data for the AMOD vehicle-related stock variables – idle vehicles, en-route pickup vehicles, en-route drop-off vehicles, and repositioning vehicles. Like Figures 17 and 18, the minor variations are smoothed out in the co-flow model, but the key patterns are preserved. Note that the vehicles in repositioning state in Figure 20 (zero vehicles in the repositioning state in Figure 19) are generally higher than those in the ABM. The differences may be further reduced by increasing the resolution and fidelity of the co-flow model to capture more operational details.

Figures 21–24 are the counterparts of Figures 16–20, respectively, except Figures 21–24 are for the validation datasets (Day 2). Similar to the results in Table 6, we do not observe

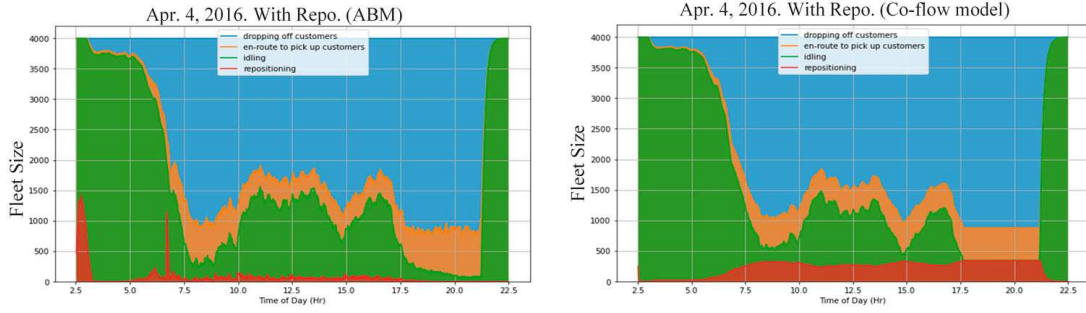


Figure 20. Stacked time profiles of the *service vehicle* stock variables *with* vehicle repositioning, for the *training* data set, for the ABM (left) and co-flow model (right).

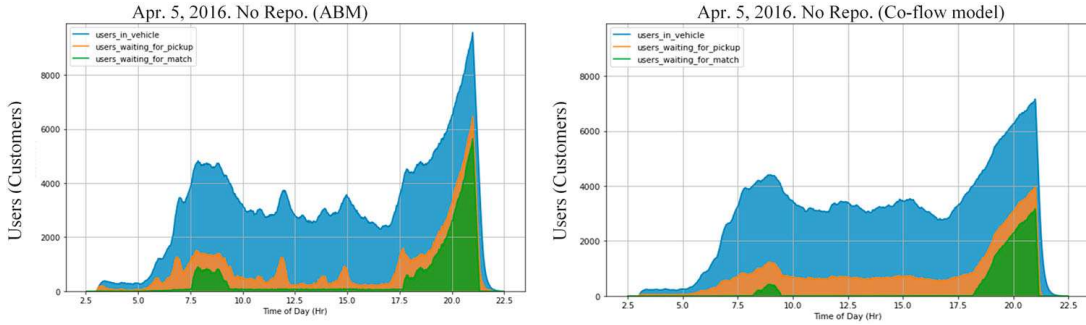


Figure 21. Stacked time profiles of the *customer* stock variables *without* vehicle repositioning for the *validation* dataset, for the ABM (left) and co-flow model (right).

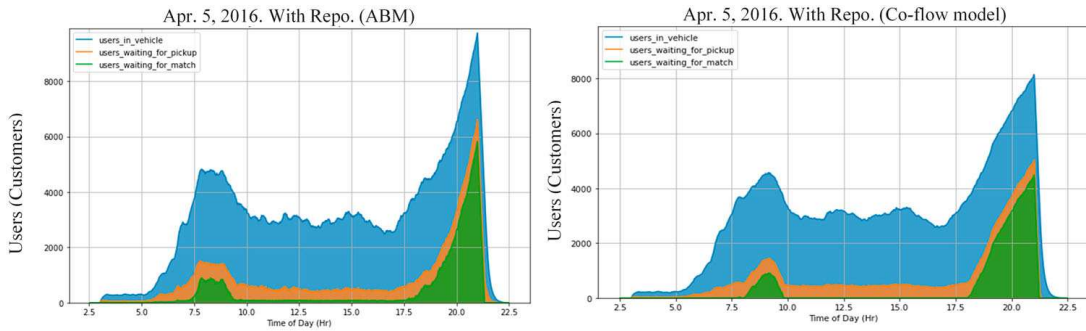


Figure 22. Stacked time profiles of the *customer* stock variables *with* vehicle repositioning for the *validation* dataset, for the ABM (left) and co-flow model (right).

model underfitting or overfitting issues. The results also show that even using a simple co-flow structure and the negative exponential distributions of delays for each state transition, we can robustly explain the systemwide dynamics for both customers and vehicles produced by the ABM simulator.

The comparison in this subsection shows that the co-flow model can provide sufficiently accurate outputs for strategic planning despite not explicitly modelling the operational details of the AMOD services. The co-flow model fails to capture the minor variations of the AMOD system dynamics, but we consider it sufficient for high-level strategic planning.

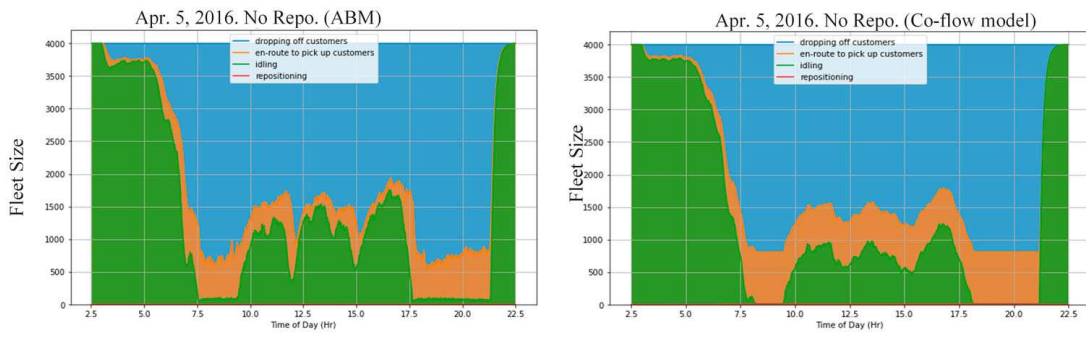


Figure 23. Stacked time profiles of the *vehicle* stock variables *without* vehicle repositioning for the *validation* dataset, for the ABM (left) and co-flow model (right).

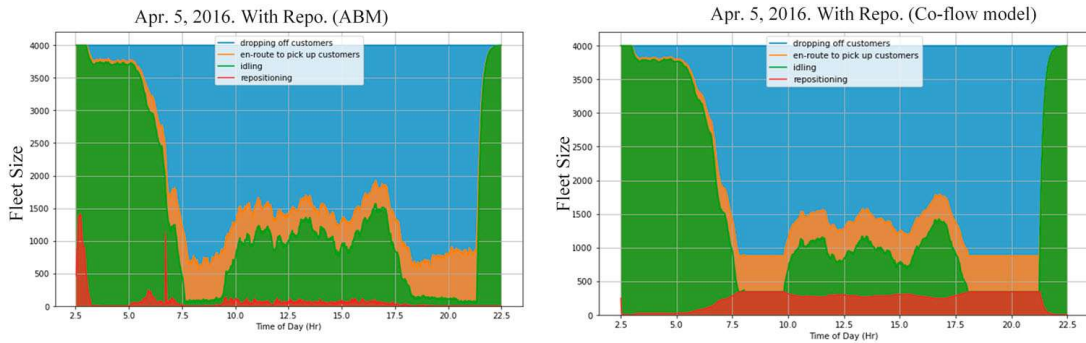


Figure 24. Stacked time profiles of the *vehicle* stock variables *with* right repositioning for the *validation* dataset, for the ABM (left) and co-flow model (bottom).

In practice, we suggest regularly comparing the co-flow model outputs with both higher-resolution, higher-fidelity models, and the empirical data (after the AMOD system is implemented) to improve understanding of the conditions where the co-flow model provides sufficiently accurate information about key system dynamics. In this sense, high-resolution, high-fidelity operational models and the proposed approach for strategic planning are complementary. On the one hand, a co-flow model can guide the development of an ABM. On the other hand, after the AMOD system is implemented, the ABM and the empirical data can provide useful feedback to the co-flow model about its *ex-ante* forecasts for further improvement. This way, the decision-makers can use the improved co-flow model for the next round of strategic planning.

8. Concluding remarks

High-quality strategic planning of AMOD systems is critical for the success of the subsequent phases such as service area permitting, fleet procurement, and depot capacity acquisition. This paper presents a novel and flexible modelling approach for strategic planning of AMOD systems using a system of ordinary differential equations that has a fluid/flow analogy. A key contribution of the proposed co-flow modelling framework, aside from the separate but coordinated vehicle and customer stocks and flows, is that the proposed approach for modelling state transition delays effectively captures the systemwide AMOD

dynamics. Additionally, the co-flow model preserves properties of the state space representation and Markov Decision Process model requirements. Informed by a comparison with a high-resolution, high-fidelity ABM, the proposed co-flow model explains, with high accuracy, the macroscopic system dynamics of an AMOD fleet serving taxi requests in Manhattan. Hence, the proposed co-flow model represents a valuable and reliable tool for strategic planning of AMOD services.

The proposed co-flow approach contributes to existing AMOD system modelling methods as it is (i) dynamic, rather than static or steady-state based; (ii) computationally inexpensive wherein the problem complexity is independent of fleet size and the number of customer requests, unlike ABMs; and (iii) and capable of capturing important conservations of different states of vehicles and customers, unlike most empirical/statistical models. The proposed modelling approach provides credible systems level results that allow for immediate insights into strategic planning decisions. These strategic planning level insights should allow AMOD companies and analysts to narrow their tactical and operational level decision spaces for further, detailed, analysis.

This study does have its limitations. The co-flow model application in this study assumes that the whole service region is a single pair of stock chains. Clearly, multiple pairs of interconnected stock chains will be necessary to model more heterogeneous service areas than Manhattan, New York City. While having separate stock flows for different subregions will increase computational burden, the impact is relatively minor given the aggregate nature of the co-flow model framework. In the case study, we also recognise that, although the request patterns from the two days have clear differences, they do have some similarities because they are back-to-back weekdays. We leave testing of a wider range of demand patterns for future research. Other future research includes the concrete demonstration that the proposed co-flow modelling framework can provide flexibility as to model boundaries (e.g. expanding the boundary by converting exogenous variables into endogenous ones), model resolutions (e.g. increasing the fidelity by splitting a chain of state variables into multiple chains to explicitly consider different zones or subnetworks), and model resolutions (e.g. increasing the fidelity by using more detailed state variables in a stock chain), as it is common for decision-makers to adjust the study boundary and resolution *during* strategic planning. It would also be worthwhile for future research to examine the transferability of a co-flow model to other geographies and service areas.

Note

1. In this research, 'resolution' refers to the level of spatial or temporal detail associated with a model attribute or model attributes in general. Conversely, 'fidelity' refers to the inclusion or exclusion of model attributes, physical processes and/or behavioural process in a model.

Acknowledgement

The authors appreciate the support of NSF CMMI# 2125560 'SCC-IRG Track 1: Revamping Regional Transportation Modeling and Planning to Address Unprecedented Community Needs during the Mobility Revolution.

'We appreciate the anonymous reviewers' comments, including Reviewer 2's recommendation to include an Illustrative Example section.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Science Foundation [grant number CMMI# 2125560].

ORCID

Jiangbo (Gabe) Yu  <http://orcid.org/0000-0002-4525-4640>

Michael F. Hyland  <http://orcid.org/0000-0001-8394-8064>

References

- Acheampong, Ransford A., Alhassan Siiba, Dennis K. Okyere, and Justice P. Tuffour. 2020. "Mobility-on-demand: An Empirical Study of Internet-based Ride-hailing Adoption Factors, Travel Characteristics and Mode Substitution Effects." *Transportation Research Part C: Emerging Technologies* 115 (June): 102638. <https://doi.org/10.1016/j.trc.2020.102638>.
- Azevedo, Carlos Lima, Katarzyna Marczuk, Sebastián Raveau, Harold Soh, Muhammad Adnan, Kakali Basak, Harish Loganathan, et al. 2016. "Microsimulation of Demand and Supply of Autonomous Mobility on Demand." *Transportation Research Record: Journal of the Transportation Research Board* 2564 (1): 21–30. <https://doi.org/10.3141/2564-03>.
- Beojone, Caio Vitor, and Nikolas Geroliminis. 2021. "On the Inefficiency of Ride-sourcing Services towards Urban Congestion." *Transportation Research Part C: Emerging Technologies* 124 (March): 102890. <https://doi.org/10.1016/j.trc.2020.102890>.
- Bilali, Aledia, M. A. A. Rathore, U. Fastenrath, and K. Bogenberger. 2020. "An Analytical Model to Evaluate Traffic Impacts of On-demand Ride Pooling; An Analytical Model to Evaluate Traffic Impacts of On-demand Ride Pooling." *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6. <https://doi.org/10.1109/ITSC45102.2020.9294287>.
- Burton, D., and Ph. L. Toint. 1992. "On an Instance of the Inverse Shortest Paths Problem." *Mathematical Programming* 53 (1–3): 45–61. <https://doi.org/10.1007/BF01585693>.
- Chen, Rongsheng, and Michael W. Levin. 2019. "Dynamic User Equilibrium of Mobility-on-demand System with Linear Programming Rebalancing Strategy." *Transportation Research Record: Journal of the Transportation Research Board* 2673 (1): 447–459. <https://doi.org/10.1177/0361198118821629>.
- Chow, Joseph Y.J., Stephen G. Ritchie, and Kyungsoo Jeong. 2014. "Nonlinear Inverse Optimization for Parameter Estimation of Commodity-vehicle-decoupled Freight Assignment." *Transportation Research Part E: Logistics and Transportation Review* 67 (July): 71–91. <https://doi.org/10.1016/j.tre.2014.04.004>.
- Codington, E. A., and N. Levinson. 1984. *Theory of Ordinary Differential Equations*. Malabar, FL: Krieger.
- Dandl, Florian, Michael Hyland, Klaus Bogenberger, and Hani S. Mahmassani. 2019. "Evaluating the Impact of Spatio-temporal Demand Forecast Aggregation on the Operational Performance of Shared Autonomous Mobility Fleets." *Transportation* 46 (6): 1975–1996. <https://doi.org/10.1007/s11116-019-10007-9>.
- Djavadian, Shadi, and Joseph Y.J. Chow. 2017. "An Agent-based Day-to-day Adjustment Process for Modeling 'Mobility as a Service' with a Two-sided Flexible Transport Market." *Transportation Research Part B: Methodological* 104 (October): 36–57. <https://doi.org/10.1016/j.trb.2017.06.015>.
- Fagnant, Daniel J., and Kara M. Kockelman. 2014. "The Travel and Environmental Implications of Shared Autonomous Vehicles, Using Agent-based Model Scenarios." *Transportation Research Part C: Emerging Technologies* 40 (March): 1–13. <https://doi.org/10.1016/j.trc.2013.12.001>.
- Forrester, Jay W. 1969. *Urban Dynamics*. Pegasus Communications, Inc.
- Guélat, Jacques, Michael Florian, and Teodor Gabriel Crainic. 1990. "A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows." *Transportation Science* 24 (1): 25–39. <https://doi.org/10.1287/trsc.24.1.25>.

- Hyland, Michael, and Hani S. Mahmassani. 2018. "Dynamic Autonomous Vehicle Fleet Operations: Optimization-based Strategies to Assign AVs to Immediate Traveler Demand Requests." *Transportation Research Part C: Emerging Technologies* 92 (July): 278–297. <https://doi.org/10.1016/j.trc.2018.05.003>.
- Inturri, Giuseppe, Michela Le Pira, Nadia Giuffrida, Matteo Ignaccolo, Alessandro Pluchino, Andrea Rapisarda, and Riccardo D'Angelo. 2019. "Multi-agent Simulation for Planning and Designing New Shared Mobility Services." *Research in Transportation Economics* 73 (March): 34–44. <https://doi.org/10.1016/j.retrec.2018.11.009>.
- Javanshour, Farid, Hussein Dia, and Gordon Duncan. 2019. "Exploring the Performance of Autonomous Mobility On-demand Systems under Demand Uncertainty." *Transportmetrica A: Transport Science* 15 (2) Taylor and Francis Ltd: 698–721. <https://doi.org/10.1080/23249935.2018.1528485>.
- Javanshour, Farid, Hussein Dia, Gordon Duncan, Rusul Abduljabbar, and Sohani Liyanage. 2021. "Performance Evaluation of Station-based Autonomous On-demand Car-sharing Systems; Performance Evaluation of Station-based Autonomous On-demand Car-sharing Systems." *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2021.3071869>.
- Jin, Wen Long, Irene Martinez, and Monica Menendez. 2021. "Compartmental Model and Fleet-size Management for Shared Mobility Systems with for-hire Vehicles." *Transportation Research Part C: Emerging Technologies* 129 (August) Pergamon: 103236. <https://doi.org/10.1016/J.TRC.2021.103236>.
- Ke, Jintao, Hai Yang, Xinwei Li, Hai Wang, and Jieping Ye. 2020. "Pricing and Equilibrium in On-demand Ride-pooling Markets." *Transportation Research Part B: Methodological* 139 (September): 411–431. <https://doi.org/10.1016/j.trb.2020.07.001>.
- Masoud, N., D. Nam, J. Yu, and R. Jayakrishnan. 2017. "Promoting Peer-to-peer Ridesharing Services as Transit System Feeders." *Transportation Research Record: Journal of the Transportation Research Board* 2650 (1): 74–83. <https://doi.org/10.3141/2650-09>.
- Nam, Daisik, Dingtong Yang, Sunghi An, Jiangbo Gabriel Yu, R. Jayakrishnan, and Neda Masoud. 2018. "Designing a Transit-feeder System Using Multiple Sustainable Modes: Peer-to-peer (P2P) Ridesharing, Bike Sharing, and Walking." *Transportation Research Record: Journal of the Transportation Research Board* 2672 (8): 754–763. <https://doi.org/10.1177/0361198118799031>.
- Narayan, Jishnu, Oded Cats, Niels van Oort, and Serge Paul Hoogendoorn. 2021. "Fleet Size Determination for a Mixed Private and Pooled On-demand System with Elastic Demand." *Transportmetrica A: Transport Science* 17 (4) Taylor and Francis Ltd: 897–920. <https://doi.org/10.1080/23249935.2020.1819910>.
- Nourinejad, Mehdi, and Mohsen Ramezani. 2020. "Ride-sourcing Modeling and Pricing in Non-equilibrium Two-sided Markets." *Transportation Research Part B: Methodological* 132 (February): 340–357. <https://doi.org/10.1016/j.trb.2019.05.019>.
- Oh, Simon, Ravi Seshadri, Carlos Lima Azevedo, Nishant Kumar, Kakali Basak, and Moshe Ben-Akiva. 2020. "Assessing the Impacts of Automated Mobility-on-demand through Agent-based Simulation: A Study of Singapore." *Transportation Research Part A: Policy and Practice* 138 (August): 367–388. <https://doi.org/10.1016/j.tra.2020.06.004>.
- Pinto, Helen K.R.F., Michael F. Hyland, Hani S. Mahmassani, and I. Ömer Verbas. 2020. "Joint Design of Multimodal Transit Networks and Shared Autonomous Mobility Fleets." *Transportation Research Part C: Emerging Technologies* 113 (April): 2–20. <https://doi.org/10.1016/j.trc.2019.06.010>.
- Rahimi, Mahour, Mahyar Amirgholy, and Eric J. Gonzales. 2018. "System Modeling of Demand Responsive Transportation Services: Evaluating Cost Efficiency of Service and Coordinated Taxi Usage." *Transportation Research Part E: Logistics and Transportation Review* 112 (April) Elsevier Ltd: 66–83. <https://doi.org/10.1016/j.tre.2018.02.005>.
- Rath, Srushti, Bingqing Liu, Gyugeun Yoon, and Joseph Y.J. Chow. 2023. "Microtransit Deployment Portfolio Management Using Simulation-based Scenario Data Upscaling." *Transportation Research Part A: Policy and Practice* 169 (March): 103584. <https://doi.org/10.1016/j.tra.2023.103584>.
- Ravindran, Ravi. 2007. *Operations Research and Management Science Handbook*. Edited by Ravi Ravindran.

- Sayarshad, Hamid R., and Joseph Y.J. Chow. 2015. "A Scalable Non-Myopic Dynamic Dial-a-Ride and Pricing Problem." *Transportation Research Part B: Methodological* 81 (November): 539–554. <https://doi.org/10.1016/j.trb.2015.06.008>.
- Steiner, Konrad, and Stefan Irnich. 2020. "Strategic Planning for Integrated Mobility-on-demand and Urban Public Bus Networks." *Transportation Science* 54 (6): 1616–1639. <https://doi.org/10.1287/trsc.2020.0987>.
- Sterman, John. 2018. "System Dynamics at Sixty: The Path Forward." *System Dynamics Review* 34 (1-2): 5–47. <https://doi.org/10.1002/sdr.1601>.
- Syed, Arslan Ali, Florian Dandl, and Klaus Bogenberger. 2021. "User-Assignment Strategy Considering Future Imbalance Impacts for Ride Hailing." 2441–2446. <https://doi.org/10.1109/ITSC48978.2021.9564559>.
- Xu, Zhengtian, Yafeng Yin, Xiuli Chao, Hongtu Zhu, and Jieping Ye. 2021. "A Generalized Fluid Model of Ride-Hailing Systems." *Transportation Research Part B: Methodological* 150 (August): 587–605. <https://doi.org/10.1016/j.trb.2021.05.014>.
- Yang, Hai, and S. C. Wong. 1998. "A Network Model of Urban Taxi Services." *Transportation Research Part B: Methodological* 32 (4): 235–246. [https://doi.org/10.1016/S0191-2615\(97\)00042-8](https://doi.org/10.1016/S0191-2615(97)00042-8).
- Yu, Jiangbo, and Michael F. Hyland. 2020. "A Generalized Diffusion Model for Preference and Response Time: Application to Ordering Mobility-on-Demand Services." *Transportation Research Part C: Emerging Technologies* 121 (December): 102854. <https://doi.org/10.1016/j.trc.2020.102854>.
- Zha, Liteng, Yafeng Yin, and Zhengtian Xu. 2018. "Geometric Matching and Spatial Pricing in Ride-Sourcing Markets." *Transportation Research Part C: Emerging Technologies* 92 (July): 58–75. <https://doi.org/10.1016/j.trc.2018.04.015>.