

BISTRO: Berkeley Integrated System for Transportation Optimization

SIDNEY A. FEYGIN, Uber Technologies, Inc.

JESSICA R. LAZARUS and EDWARD H. FORSCHER, University of California, Berkeley;
Uber Technologies, Inc.

VALENTINE GOLFIER-VETTERLI, JONATHAN W. LEE, and ABHISHEK GUPTA,
Uber Technologies, Inc.

RASHID A. WARAICH and COLIN J. R. SHEPPARD, Lawrence Berkeley National Laboratory

ALEXANDRE M. BAYEN, Electrical Engineering and Computer Science, Berkeley;
Institute of Transportation Studies

The current trend toward urbanization and adoption of flexible and innovative mobility technologies will have complex and difficult-to-predict effects on urban transportation systems. Comprehensive methodological frameworks that account for the increasingly uncertain future state of the urban mobility landscape do not yet exist. Furthermore, few approaches have enabled the massive ingestion of urban data in planning tools capable of offering the flexibility of scenario-based design.

This article introduces Berkeley Integrated System for Transportation Optimization (BISTRO), a new open source transportation planning decision support system that uses an agent-based simulation and optimization approach to anticipate and develop adaptive plans for possible technological disruptions and growth scenarios. The new framework was evaluated in the context of a machine learning competition hosted within Uber Technologies, Inc., in which over 400 engineers and data scientists participated. For the purposes of this competition, a benchmark model, based on the city of Sioux Falls, South Dakota, was adapted to the BISTRO framework. An important finding of this study was that in spite of rigorous analysis and testing done prior to the competition, the two top-scoring teams discovered an unbounded region of the search space, rendering the solutions largely uninterpretable for the purposes of decision-support. On the other hand, a follow-on study aimed to fix the objective function. It served to demonstrate BISTRO's utility as a human-in-the-loop cyberphysical system: one that uses scenario-based optimization algorithms as a feedback mechanism to assist urban planners with iteratively refining objective function and constraints specification on intervention strategies. The portfolio of transportation intervention strategy alternatives eventually chosen achieves high-level regional planning goals developed through participatory stakeholder engagement practices.

At the time of writing, Lee and Gupta were full-time employees at Uber. Feygin, Lazarus, Forscher, Golfier-Vetterli, and Bayen contributed to this work partially as Uber contractors.

Authors' addresses: S. A. Feygin, 630 8th Street, #8, San Francisco, CA 94103; email: sid.feygin@gmail.com; J. R. Lazarus and E. H. Forscher, University of California, 116 McLaughlin Hall, Berkeley, CA 94720; email: {jlaz, Forscher}@berkeley.edu; V. Golfier-Vetterli, 55 Harrison St # 600 Oakland, CA 94607; email: vgolfi@ext.uber.com; J. W. Lee, University of California, 116 McLaughlin Hall, Berkeley, CA 94720; email: jonny@uber.com; A. Gupta, 55 Hawthorne St 4th floor, San Francisco, CA 94105; email: agupta@uber.com; R. A. Waraich, 8436 Fern Bluff Ave Round Rock, TX 78681; email: waraich.rashid@uber.com; C. J. R. Sheppard, 2913 Woodland Ct, Arcata, CA 95521; email: colin.sheppard@lbl.gov; A. M. Bayen, 642 Sutardja Dai Hall, Berkeley, CA 94720-1764; email: bayen@berkeley.edu.

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor, or affiliate of the United States government. As such, the United States government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for government purposes only.

© 2020 Association for Computing Machinery.

2157-6904/2020/06-ART38 \$15.00

<https://doi.org/10.1145/3384344>

CCS Concepts: • **Human-centered computing** → **Collaborative interaction**; • **Computing methodologies** → **Multi-agent systems**; **Agent/discrete models**; *Simulation tools*; • **Applied computing** → **Transportation**;

Additional Key Words and Phrases: Agent-based models, human mobility, system dynamics, big data, smart cities, digital decision support systems, urban informatics, computing with heterogeneous data, intelligent transportation systems

ACM Reference format:

Sidney A. Feygin, Jessica R. Lazarus, Edward H. Forscher, Valentine Golfier-Vetterli, Jonathan W. Lee, Abhishek Gupta, Rashid A. Waraich, Colin J. R. Sheppard, and Alexandre M. Bayen. 2020. BISTRO: Berkeley Integrated System for Transportation Optimization. *ACM Trans. Intell. Syst. Technol.* 11, 4, Article 38 (June 2020), 27 pages.

<https://doi.org/10.1145/3384344>

1 INTRODUCTION

As modern transportation systems undergo a period of intense technological change, researchers, practitioners, and policymakers are seeking to understand how long-term trends toward vehicle digitalization, automation, electrification, as well as the emerging sharing economy will shape the future day-to-day dynamics of human mobility in cities worldwide.

Likewise, rapid advances in computing as well as the advent of metropolitan scale data mining and pattern recognition (i.e., machine learning) have, for the first time, made it possible to characterize urban traffic flows based on movement traces from millions of individual travelers. These methods fuse passively collected spatiotemporal trajectories derived from smartphone data with static census data in order to map travel patterns to sociodemographic characteristics. Specialized traffic models can thereby be trained using feature-rich representations of human mobility.

While such predictive models have proven effective in nearcasting congestion events [1], purely data-driven methods often fail to generalize when applied to the task of forecasting the effect of transportation policy strategies on future demand—particularly when anticipated changes in urban sociodemographics, regulatory policy, and mobility technologies as well as interactions between these layers cannot themselves be predicted with a high degree of certainty [2]. In other words, the covariates of statistical models may not capture the numerous interacting physical and behavioral components (i.e., *network effects*) influencing transportation supply and demand.

Agent-based modeling and simulation (ABMS) techniques and software, on the other hand, are beginning to be used as flexible decision-support tools by planning groups, technologists, and regulatory agencies to facilitate forecasting the short- and long-term implications of transportation system interventions. That is, by simulating models of traveler decision-making in the context of a landscape of future plausible “states of the world,” stakeholders can better resolve uncertainty over how transportation system interventions may fare. For example, ABMS enables the investigation of the extent to which citizens may adopt on-demand autonomous vehicles, as well as how to sustain fleet operation [3, 4].

Even with these advances, selecting interventions that balance competing transportation system policy objectives remains a difficult and contentious process. Current methods to identify the best alternative under a given scenario often involves the use of Monte Carlo methods to evaluate different policy options. Such approaches can be extremely computationally expensive for large ABMS implementations and are unlikely to yield an optimal solution. Moreover, transportation researchers and practitioners often develop scenario-specific and/or geography-specific simulation models with limited integration of the efficient and highly generalizable methods developed in the *artificial intelligence* (AI) and *machine learning* (ML) communities.

The *Berkeley Integrated System for Transportation Optimization* (BISTRO) is an open source framework and software toolkit designed to address the increasingly complex problems arising in transportation systems worldwide. BISTRO does this through an innovative human-machine collaboration approach, using state-of-the-art optimization algorithms to efficiently automate the search for policy interventions that achieve good performance over diverse transportation system scenarios and stakeholder objectives. BISTRO includes an ABMS system and scenario development pipeline to build empirically-calibrated simulations of travel demand in metropolitan transportation systems. After a calibrated scenario has been developed, BISTRO enables use of state-of-the-art optimization algorithms to identify system interventions that best align with policy and planning objectives. Users can deploy BISTRO to enable distributed development of algorithms that rapidly optimize a feasible set of policy and investment decisions. Once one or more desirable solutions are found, BISTRO provides a suite of analysis and visualization tools to empower citizens, transportation system planners and engineers, private entities, and governments to better understand and collaborate on developing strategies that achieve equitable access to and sustainable use of current and emerging mobility services.

The rest of this article is organized as follows: Section 2 provides a brief overview of the current state of transportation planning, ABMS tools, and transportation optimization tools, respectively; Section 3 presents BISTRO, covering the system architecture, scoring function design, inputs, outputs, analysis capabilities, and performance characteristics; Section 4 details an initial pilot study, updates to BISTRO based upon the pilot, and a few algorithmic solution approaches; and Section 5 offers a short conclusion.

2 BACKGROUND

2.1 Transportation Planning and Policy Decision Support Systems

In the United States, the primary outcome of efforts to design, model, and communicate the impacts of proposed policy interventions and infrastructure investments on transportation systems are Metropolitan *Regional Transportation Plans* (RTPs) and State *Long-Range Transportation Plans* (LRTPs). These plans are forward-looking, long-term (20+ year time horizons) and have been a federally mandated task since the Federal-Aid Highway Act of 1962. Recently, state departments of transportation and *Metropolitan Planning Organizations* (MPOs)—the entities tasked with producing RTPs every four-to-five years—have been shifting toward performance-based planning and programming (PBPP) frameworks [5].

As part of RTP development using PBPP guidelines, an MPO or related agency will typically conduct a community engagement process to identify one or more *visions* or *goals* shared by stakeholders and the public describing a desired future state of the regional transportation system (e.g., safe roadways, accessible transit, environmental stewardship) [5–7]. Measurable *objectives* are defined together with quantitative *performance indicators* in order to evaluate the extent to which alternative *strategies* consisting of policy interventions or infrastructure investments could make progress toward achievement of a goal [5, 6]. Typically, the impact of these projects on *key performance indicators* (KPIs) of transportation system performance will be forecast using an analytic, data-driven model of travel demand.¹

¹Traditionally, the travel demand modeling process consists of four main steps: (1) trip generation to and from all analysis zones; (2) trip distribution (or matching origins and destinations, often using a gravity model); (3) assigning traveler mode choice based upon individual preferences and alternative characteristics; and (4) route assignment of trips onto physical network links; this is referred to as the four-step model. Many MPOs and other agencies are moving toward disaggregate, activity-based, or person-centric models of daily activity, rather than aggregate approaches operating on the zonal level [6].

The domain expertise necessary to understand the functionality of models of transportation demand may result in recommendations that are frustratingly opaque to public interpretation [7]. Explanations of the inner workings of the modeling process are not presented during public collaboration meetings, and questions regarding the validity of models have led to lawsuits over the lack of publicly available information on model specifics [8]. Often, contractual obligations or proprietary data formats used by the consulting firms tasked with developing planning software further restrict public access. The inability or unwillingness on the part of MPOs to promote flexible open source software licensing strategies discourages independent investigations (by, for example, academic or public interest groups) to verify that the theory, equations, algorithms, and data comprising a model match its software implementation. This lack of operational transparency not only subverts accountability, but also impedes the rapid transfer of innovative modeling methods and technologies across and even within agencies [9, 10].

To address criticism concerning model transparency and explainability, recent transportation and land-use planning organizations have begun to pilot software that visualizes the impacts of alternatives and provide for a public input via collaborative simulation platforms [6, 11–14]. An interesting example is the UrbanAPI project, which includes a web-based three-dimensional virtual reality visualization of the impacts of urban growth scenarios at the scale of individual neighborhoods [13]. While initial efforts have focused on the usability requirements of these tools, their broader impact has been limited due to narrowly defined geographic contexts or specialization for regional system objectives [15].

One project that does have broad support across several MPOs is ActivitySim, which is currently being developed by the Association of Metropolitan Planning Organizations as an open source activity-based travel modeling platform [16]. ActivitySim may be used in concert with synthetic households, synthetic persons, employment data, land use data, and network performance data (e.g., travel times by mode and time of day, costs, and transfers) to generate a full-scale model for a city's travel demand. ActivitySim may be seen as a complementary project to our own in that ActivitySim model outputs can be used as input data for a BISTRO scenario.

2.2 ABMS of Transportation Systems

Developing models that replicate how urban systems operate and evolve has been a major focus of transportation engineers, urban planners, and geographers. Trip-based methods, such as the traditional four-step model used by MPOs, are specified at aggregate geographic or categorical levels rather than at the level of individual decision-makers. This level of aggregation can limit the ability of such models to explain complex individual decisions. Some MPOs have begun to adopt a more behaviorally-descriptive activity-based approach [17]. In contrast to trip-based models, activity-based methods represent more comprehensive links between activity scheduling, mode choice, social interaction, and spatiotemporal constraints [17]. ABMS of transportation demand are capable of replicating observed *macroscopic* traffic patterns by simulating the *microscopic* decision-making behavior of a synthetic population of software agents as they execute their daily travel plans on a virtual model of the transportation system. When *calibrated* to ground-truth data by modifying only global parameters characterizing the embedded choice model, agent-based frameworks represent parsimonious descriptions of regional travel demand. Consequently, ABMS are capable of accurately capturing shifts in macro patterns when forecast changes in transportation infrastructure, policies, demographics, and vehicle ownership are introduced to the virtual travel environment [2].

In the past three decades, several multi-agent frameworks such as TTransportation ANalysis SIMulation System (TRANSIMS) [18], MultiAgent Transportation Simulation (MATSim) [3], Simulation of Urban Mobility (SUMO) [19], and Planning and Operations Language for Agent-based

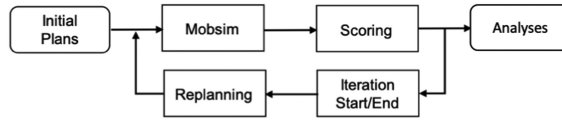


Fig. 1. Conceptual process of MATSim. It iteratively evaluates and mutates a proportion of agent plans until the utility of plans no longer improves. At this point, the system is said to have reached a *stochastic user equilibrium*. For further details, see Reference [3].

Regional Integrated Simulation (POLARIS) [11] have been developed and widely adapted for numerous applications in transportation and land use planning and research [2, 15, 20, 21]. MPOs often couple urban development simulation models² with microscopic agent-based transport models to better understand how predicted changes in population growth, land-use, real-estate development, and resource markets will co-evolve with changes in the transportation system [24, 25]. ABMS of transportation systems can take different population configuration files as inputs, giving planners and modelers the ability to simulate how long-term urbanization processes can be shaped by the daily transportation decisions of individuals (possibly in the presence of alternative policy interventions and new mobility technologies) [25, 26]. Increasingly, these open source platforms are enabling use of publicly available data to create transparent and replicable input preparation pipelines, reducing the cost and effort of a variety of urban planning tasks [27]. Since BISTRO primarily relies on BEAM and BEAM itself incorporates many aspects of MATSim, the remainder of this section takes a closer look at these two software frameworks' purpose, functionality, and computational characteristics.

MATSim. MATSim is an ABMS framework developed by teams at Eidgenössische Technische Hochschule (ETH) Zürich and Technische Universität (TU) Berlin [3]. MATSim enables simulating the travel behavior of millions of individual agents, representing a synthetic population of urban travelers. At the heart of MATSim is a co-evolutionary algorithm that iteratively executes, evaluates, and mutates (i.e., replans) the *daily* activity schedules of agents (see Figure 1). The end result of this process is an equilibrium between network supply and travel demand, resulting in realistic congestion patterns as agents compete for limited space on a virtual road network.

Besides road network and transit data, the key input to MATSim is the agent *population*. This file encodes a set of unrealized plans (one for each agent) consisting of the start times, types (e.g., “Home,” “Work,” “Shopping,” “School”), and locations of various significant activities.³ A mobility simulation (MobSim) executes these plans on a virtual road network. While, at first, agents only drive and/or walk to activities, additional modes may be introduced during plan mutation (explained below). Each agent's plan is then scored.⁴ Copies of evaluated plans are stored in a limited-size array, representing the agent's memory. At the start of the subsequent iteration, a portion of agents in the population are chosen to have a randomly selected mutation strategy applied to a plan drawn from each of their memories. Examples of plan mutations include changing activity start time, mode of travel for a tour (or subtour), and selecting a different route based on previously experienced link travel times. Optional extensions to the MATSim model further behavioral dimensions such as parking choice, group travel, and vehicle sharing.

²Examples of these include the *Integrated Transportation, Land Use, and Environment* model (ILUTE, [22]) and UrbanSim [23].

³The population file is often generated from the output of an activity-based travel demand model.

⁴This score can be interpreted as econometric utility. It is measured by a linear model that assigns negative value to time spent traveling and positive value to time spent at activities.

The algorithm converges once agents are no longer able to improve the score of plans in their memory, at which point MATSim produces a series of statistics and outputs describing the aggregate performance of system components as well as a snapshot of all events that occurred over the course of the simulation. Events can be processed to derive the actual paths and travel times realized by each agent and each vehicle, as well as other data reflecting simulation performance.

BEAM. The *Behavior Energy Autonomy and Mobility* (BEAM) framework is a multi-agent travel demand simulation framework developed at *Lawrence Berkeley National Laboratory* (LBNL) [28]. While the overall learning and traffic assignment mechanism is similar to MATSim's co-evolutionary algorithm, added functionality in BEAM is specifically focused on helping users understand the impacts of new and emerging travel modes on limited capacity resource markets. This section describes essential features of BEAM that led to its selection as the core simulation engine of BISTRO.

BEAM is closely integrated with the transit service capabilities of the R5 routing engine, which include *General Transit Feed Specification* (GTFS) file processing and routing based on multiobjective variations of the (Round-bAsed Public Transit Optimized Router) RAPTOR algorithm [29]. Transit may be combined with other modes modeled in BEAM such as autonomous vehicles, on-demand rides, e-bikes, and scooters, enabling agents to make realistic, multimodal mobility decisions. In order to provide agents with information about the time and monetary costs of different travel options, R5 computes the lowest generalized cost path (based on travel time estimates from the mobility simulation) for the corresponding mode(s) available to the agent for the trip.⁵

Unlike the replanning mechanism in MATSim, which only mutates plans *between* consecutive iterations, agents in BEAM are designed to adapt to changing conditions *during* an iteration according to what is known as a *within-day* or *online* model.⁶ Thus, agents can make unplanned and time-sensitive choices about how to maximize the score of their travel plans while competing for limited resources that vary in availability over time. For example, an agent that chooses a transit mode may be denied access to an overfull bus, requiring the agent to make a mid-trip change to their itinerary. The agent could then choose to wait for the next bus or hail a ride from the point of departure (if a driver is available nearby). The BEAM software architecture addresses the performance and complexity challenges of integrating new models of within-day dynamics by implementing agents as *actors*, as defined within the *actor-based model of concurrency*.⁷

Like MATSim, BEAM enables users to model realistic variations in travel preferences predicated on agent characteristics (which are themselves derived from sociodemographic statistics computed on census data and travel surveys). However, the mechanism for selection of travel alternatives differs significantly from MATSim's in order to model within-day decision-making. Specifically, the probability of selecting an alternative (route, mode, parking choice, and refueling decision) is represented in BEAM according to a multinomial logit model [32, 33]. That is, among several distinct travel schedules, agents are exponentially more likely to select the option that maximizes their enjoyment of important activities while reducing time and money spent traveling between

⁵For detailed information about the R5 router, see Ref. [30].

⁶While MATSim has a within-day mode, much of the functionality enabled by extension modules representing emerging mobility technologies assumes that replanning happens between iterations.

⁷Like objects in the object-oriented programming paradigm, actors encapsulate state and behavior. However, unlike the object model, actors do not share computer memory. Instead, each actor encapsulates its own thread of execution and interacts with one other actor using messages. An actor may send messages to other actors without blocking. Each actor processes messages synchronously in the order received; however, computation is scheduled asynchronously over multiple actors. Thus, the actor-based model of computation obviates the need for locking mechanisms commonly used to synchronize state among interdependent objects. Consequently, reasoning about agent behavior using actors can allow researcher developers to focus on implementing novel models and applications rather than debugging threads and locks [31]

activity locations. The choices made by BEAM agents and their corresponding scores comprise a BEAM plan, which is stored in memory following its execution. Thus, in the transit choice scenario described above, an agent's response to a full bus could incorporate multiple downstream choices within one iteration rather than wait for a score penalty to propagate an optimal response through the selected plan (i.e., using plan mutations that take place over the course of multiple iterations).⁸

2.3 Simulation-based Optimization of Transportation Systems

2.3.1 Optimization-based Formulation of the Planning Problem. The problem class solved by the BISTRO framework can be characterized as simulation-based optimization of large urban transportation systems. It can be symbolically formulated as an optimization problem:

$$\underset{\vec{d} \in \mathcal{D}}{\text{minimize}} \quad f(\vec{d}, \vec{x}; \vec{z}) \equiv \mathbb{E}[F(\vec{d}, \vec{x}; \vec{z})] \quad (1)$$

constrained by simulation outcomes and design constraints, i.e.,

$$\begin{cases} \vec{x} = B(\vec{d}; \vec{z}), \\ g(\vec{d}; \vec{z}) = 0, \end{cases} \quad (2)$$

respectively, where the objective, f , is defined as the expected value of a stochastic performance measurement function, F . The deterministic decision vector, \vec{d} , is chosen from a search space \mathcal{D} , which may be continuous, categorical, combinatorial, or conditional. In the BISTRO context, the decision variables, \vec{d} , are the user-defined inputs that control policy levers within the transportation system. For example, \vec{d} may specify the fare or vehicle types for specific public transit routes. The exogenous variables, \vec{z} , are the configuration inputs that determine the parameters of the population synthesis, the parameters of the transportation network, and the parameters governing supply of transportation services. The endogenous variables, \vec{x} , are the outcomes of the simulation run using \vec{d} and \vec{z} as input. The vector \vec{x} contains the details of agent and vehicle movements throughout the simulation run, such as mode choices, travel times, travel costs, and vehicle path traversals, which were realized during the simulation run, i.e., $\vec{x} = B(\vec{d}; \vec{z})$, where B represents the BEAM simulator. It is assumed that the iterative simulation process described in Section 2.2 has achieved stationarity.⁹

An important goal of BISTRO is to define objective functions that guide algorithms toward a range of solutions that represent *interpretable* and implementable policy decisions. A critical safeguard against unrealistic outcomes is implemented in BISTRO by translating business rules about inputs into mathematical constraint functions, g , parameterized by the decision vector, \vec{d} . Finally, F is computed as a convex combination of the score components that guide solutions toward the *system objective* (as defined in Section 2.1).¹⁰ The score components are evaluated from KPIs of the system performance, which are calculated using the simulation outputs, and relevant inputs.

⁸As in MATSim, the highest scoring BEAM plans are more likely to be re-evaluated and possibly selected for mutation. However, in contrast to MATSim, plans selected for mutation are cleared of all choices and re-evaluated within the context of the current iteration's transient state. Empirically, we find that this approach reduces the number of iterations needed to reach equilibrium.

⁹Individual optimization algorithms may relax this constraint in order to reduce compute time while potentially trading off reduced accuracy or increased stochasticity of simulation output statistics.

¹⁰Due to variable amounts of nondeterminism and stochasticity inherent in ABMS, given fixed \vec{d} and \vec{z} , the distribution of f can be approximated using n realizations of F as $\hat{f}(\vec{d}, \vec{x}; \vec{z}) = \frac{1}{n} \sum_{i=1}^n F_i(\vec{d}_i, \vec{x}_i; \vec{z})$. In practice, optimization usually proceeds with $n = 1$ in order to identify promising (i.e., close to optimal) subsets of \mathcal{D} ; however, when reporting final scores, one must carefully select n such that variability in output values is adequately captured.

Following Reference [34], we approximate the objective as the sample average of r independent realizations of F :

$$\hat{f}(\vec{d}, \vec{x}; \hat{z}) = \frac{1}{r} \sum_{i=1}^r F_i(\vec{d}, \vec{x}; \hat{z}) \quad (3)$$

The state space dynamics that govern the simulation-based optimization of an urban transportation system are highly complex and nonlinear, potentially containing several local minima. The lack of closed-form solutions to this class of optimization problem together with the computational expense associated with evaluating a single decision point make this class of problems particularly difficult to solve. In the next section, we describe several possible approaches to address this complexity.

2.3.2 Optimizing Complex Simulated Systems: Challenges and Approaches. As described in Section 2.2, agent-based micro- or meso-scale simulations of transportation systems model the interdependent choices of rational individuals as they navigate virtual representations of physical and human geographies. Calibrating such models to high-resolution GPS traces and other sensor data embedded in infrastructure makes them highly suitable for evaluating the outcomes of location-specific policy alternatives. The tradeoffs in accounting for the heterogeneous preferences of millions of agents are that (1) the simulation model is expensive to evaluate for different settings of \vec{d} , and (2) the complex relationship between network dynamics and agent behavior lead to stochastic, non-convex specifications of performance measure, F . Consequently, the efficient gradient-based methods used to optimize closed-form relaxations of mobility dynamics as well as data-driven models derived from historical movement patterns do not apply [34]. Instead, generalized *stochastic optimization* (SO) algorithms treat the simulator as a black box. Commonly used derivative-free SO approaches include grid search, random search, ranking and selection, meta-heuristic, and *metamodeling* techniques [35, 36].

Metamodeling algorithms encompass a broad class of simulation-based optimization approaches. These approximate F using a *surrogate model* Q that is less costly to evaluate. Flexible and computationally tractable representations such as polynomial splines are able to approximate any objective function; however, many simulation runs are still required to accurately fit the response surface of the underlying system [35, 36].

Sequential model-based optimization (SMBO) is a general metamodeling formalism that, given a history of previous evaluations, $\mathcal{H} = \{(\vec{d}_1, y_1), \dots, (\vec{d}_i, y_i)\}$, of observations $y_i = F(\vec{d}_i, \vec{x}_i; \vec{z})$ at sample points in \mathcal{D} , selects the optimal next point \vec{d}_{i+1} based on an approximation of F . To initialize SMBO, a small set of samples, $\{\vec{d}_1, \dots, \vec{d}_i\}$ from \mathcal{D} are selected using various experimental design techniques (e.g., random or Latin hypercube sampling). For each \vec{d}_i , evaluations of the expensive objective function F form an observation, which, together with \vec{d}_i are appended to a historical dataset \mathcal{H} . Once \mathcal{H} is initialized, SMBO then proceeds iteratively: First, a regression model, Q , is fitted to the current dataset, \mathcal{H} , yielding a surrogate model for F at the current iteration, which may be denoted Q_i . Based on Q_i , the next input \vec{d}_{i+1} to F is selected by optimizing an *acquisition function*, $\alpha : \mathcal{D} \mapsto \mathbb{R}$ over \mathcal{D} , which measures the utility gained from evaluating F at \vec{d}_{i+1} . Following evaluation of $F(\vec{d}_{i+1}, \vec{x}_i; \vec{z})$, \mathcal{H} is updated as $\mathcal{H} = \mathcal{H} \cup (\vec{d}_i, y_i)$. The SMBO process continues until a predetermined time or computation budget is exhausted.

SMBO techniques are typically distinguished by the forms of the surrogate model, Q , and the acquisition function, α . In Bayesian optimization (BayesOpt), a Gaussian Process (GP, [37]) is typically used to model a prior over Q , which, at each iteration, is updated using previously observed data \mathcal{H} to give a posterior predictive distribution $p(y | \vec{d}, \mathcal{H})$ [38, 39]. Several methods using GPs as surrogate models may be distinguished according to the form of the covariance

kernel parameterizing the GP [38, 40]. In lieu of GPs, BayesOpt algorithms have also used random forests [41] and *tree-based Parzen estimators* (TPE) [42, 43] as priors over Q . Acquisition functions are chosen to balance *exploration* and *exploitation* in the sample domain. The most common acquisition function used by these methods is based on an expected improvement criterion [44]; however, newer methods use variations on knowledge gradients [45, 46].

Many SMBO algorithms can run trials in parallel, which may yield reduced wall clock time (although the total number of trials required may be identical to that used by sequential implementations) [47]. One method to further reduce the running time of SMBO *trials* when model evaluations require an inner iterative loop to achieve stationarity (as in BISTRO) is to incorporate an early stopping rule for simulation evaluations that are likely to eventually be extremely suboptimal. An example of such an approach is “freeze-thaw Bayesian Optimization” [48].

Recent efforts in transportation science and operations research have also sought to develop tractable simplifications of scenario-based optimization of simulated urban transportation systems. One vein of research concentrates on deriving deterministic analytic equations describing system dynamics at equilibrium from static information, \vec{z} (e.g., network topology and bus schedules) to inform purely functional metamodels [49–51]. For example, Reference [34] combines a computationally tractable model of congested traffic based on queuing theory with a detailed local approximation using a linear combination of basis functions from a parametric family. Alternatively—and analogously to the “freeze-thaw” Bayesian optimization setting highlighted above—some approaches use information about the *process* by which the stochastic simulation achieves stationarity to develop techniques that rapidly evaluate different settings of the decision variable vector, \vec{d} , while avoiding the need to reach convergence [52, 53].

The present work refrains from prescribing a single best approach to solve optimization program Equations (1) and (2). Instead, the intent of BISTRO is to enable *replicable* future research in this area by providing a platform and problem setting that is generalizable across different planning contexts as well as approachable and of research interest to the ML/AI community. Problem characteristics such as the propensity for competing metrics to be present in system objectives, preemptive stopping of inner optimization loops, the high dimensionality of the search space, and the potential for hybridization of functional and physical metamodels are expected to provide challenging, scalable, and, critically, *explainable* solution approaches. Algorithms combining data-driven dimensionality reduction techniques as well as efficient experiment design can result in repeatable protocols to effectively constrain more general local and global search techniques. Following a presentation of the framework architecture, in Section 4, we empirically explore the effectiveness of some of these solution strategies as well as the interpretability of their outputs.

3 BERKELEY INTEGRATED SYSTEM FOR TRANSPORTATION OPTIMIZATION (BISTRO)

BISTRO is a new analysis and evaluation platform that works in concert with an ABMS (BEAM) to enable the open sourced development and evaluation of transportation optimization methods in response to given policy priorities. This section gives an in-depth description of the BISTRO framework and all of its major components, providing an overview of their purpose, use, and functionality as well as calling attention to the most novel aspects of its design.

3.1 System Architecture

As indicated in Section 2.3, BISTRO implements elements of the travel demand planning process coupled with components of an automated simulation-based optimization system. This section describes the high-level overall architecture of BISTRO (depicted in Figure 2), focusing on the

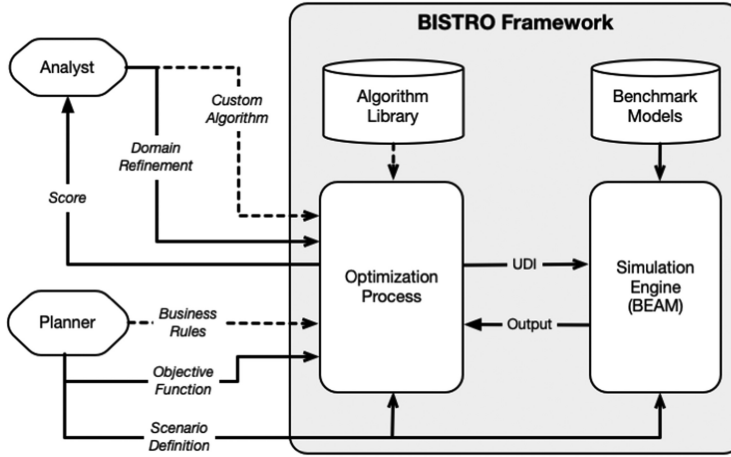


Fig. 2. BISTRO software architecture, illustrating how the optimization process modulates the flow of information between the BEAM simulation as well the two primary user types. The distinction between the *planner* and the *analyst* is critical in that we do not expect the analyst (an expert in applied ML/AI-based optimization methods) to have transportation or planning background, yet still they should be able to develop generalizable algorithms that can be used to optimize transportation system objectives set by the planning organization.

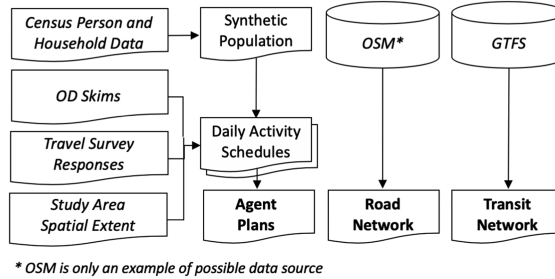


Fig. 3. Generation of fixed inputs. Italicized entities represent the necessary data to generate fixed inputs (in bold).

conceptual distinction between features relevant to scenario developers and those more appropriate to algorithm designers.

A BISTRO run environment is configured using a set of fixed input data defining the required transportation system supply elements (e.g., road network, transit schedule, on-demand ride fleet) and demand elements (e.g., synthetic population, activity plans, and mode choice function parameters). Precisely which aspects of the virtual transportation system should be represented in the simulation model depends on the strategic goals and system objectives defined as part of the planning and analysis process motivating a particular BISTRO use case. An example of the set of raw inputs and pre-processing steps is illustrated in Figure 3.

A boundary separates external, exogenously defined inputs from the BISTRO simulation optimization pipeline. Outside of the boundary, the *user-defined inputs* (UDIs) represent the investment, incentive, and policy levers applicable to and available for the study at hand. Concretely, algorithm developers encode solutions as numeric values that represent vector-valued variables controlling aspects of the initialization and evolution of the simulation. For example, a UDI that

alters frequency of buses on a route must specify a target transit agency, a route, a start time, an end time, and the desired headway.

While BISTRO maintains a library of available interventions compatible with BEAM, scenario designers, policy makers, and other stakeholders will often want assurance that infeasible, regressive, or otherwise undesirable input combinations are prevented from being selected as “optimal.” Together with syntactic and schematic validation of inputs, flexibly-defined *business rules* can effectively act as constraints on the search space—enhancing the interpretability and, thereby, the rhetorical and communicative value of BISTRO-derived solutions.

Just as UDIs from previously conducted BISTRO-based studies are actively maintained and made available to scenario designers, the BISTRO community contributes to a growing library of recommended optimization algorithms that, when evaluated across multiple BISTRO benchmark scenarios, demonstrate desirable performance characteristics. Thus, users lacking resources or expertise to develop optimization routines in-house can still benefit from what we anticipate will be cutting-edge research on algorithms and strategies to optimize the simulation of demand-responsive cyberphysical infrastructure.

Project owners of BISTRO deployments may work with stakeholders to develop representative models that will be used to benchmark optimization algorithms. Enabling a well-defined benchmark mechanism permits data on the performance of user-supplied algorithms to be compared. These comparisons can be used to assist in identification of design patterns and computational strategies that advance the state-of-the-art in simulation-based optimization of urban transportation systems.

3.2 Scoring Function Design

Transportation system intervention alternatives are scored in BISTRO based on a function of score components evaluated using KPIs of the simulation. KPIs for a given simulation should be selected in accordance with the operational, environmental, and social goals, or *system objectives* developed as part of the participatory planning process described in Section 2.1. BISTRO project planners may select KPIs to include in the scoring function from an existing library of options, or may choose to develop additional KPIs, as appropriate, for the goals and system objectives of the project. Additionally, the form of the scoring function may be designed by the analyst in consultation with the project planner.

3.2.1 Key Performance Indicators.

KPI overview. There are two general types of KPIs developed in BISTRO: (1) KPIs that measure the operational efficiency of the transportation system (e.g., *vehicle miles traveled* [VMT], vehicle delay, operational costs, revenues) and (2) KPIs that evaluate the experience of transportation system users (e.g., generalized travel expenditure, bus crowding experienced, accessibility). KPIs can be aggregated or disaggregated into score components to support particular policy objectives. For example, the accessibility KPI (detailed below) may be disaggregated by activity type, time period, mode used, and/or sociodemographics in order to evaluate the distributional equity of access provided across different opportunities at varying times of day and/or across population segments of concern.

In practice, any KPI that may be evaluated from the set of output variables (see Section 3.4) produced by a BISTRO simulation run may be included as a score component in the scoring function. However, careful consideration of candidate KPIs must include an evaluation of the sensitivity of the metric to the UDIs of interest as well as the efficiency of the KPI in providing the desired feedback regarding the optimality of outcomes of alternative UDI values. For example, *person miles traveled* (PMT) is a commonly used metric in transportation system performance measurement

to gauge the amount of mobility delivered by the system. Yet, PMT is highly invariant within a scenario in BISTRO due to the fact that agent plans are fixed. Thus, agents will make the same trips regardless of the UDI values and the miles traveled by each agent will only vary in so much as the networks available for each mode offer more or less direct paths to travel from the origin to destination of each trip.¹¹

Implemented KPIs. The following items represent categories of KPIs that have been developed and implemented in BISTRO at the time of publication of this article:

- (1) *Accessibility.* In an urban transportation planning setting, accessibility has often been defined as a measure of the ease and feasibility with which opportunities or points of interest can be reached via available modes of travel. It is quantified in BISTRO as the sum of the average number of points of interest (of a specific type of activity) reachable within a given duration of time, with functionality also provided to measure mode-specific accessibility.
- (2) *Trip Expenditure and Generalized Transportation Cost Burden.* The socio-demographic and spatial heterogeneity of travel behavior within BISTRO enables a variety of equity-focused impact analyses. Two such metrics have been implemented in BISTRO: average trip expenditure and average generalized transportation cost burden. While the former is the average monetary cost incurred by agents per trip, the latter is computed as the sum of the travel expenditures of the trip (costs of fuel and fares minus incentives, as applicable) and the monetary value of the duration of the trip (the product of the total trip duration and the population *average value of time* (VOT)), divided by the household income of the agent completing the trip. The monetary value of the trip duration is calculated by multiplying total duration by the population average VOT. Both KPIs may be disaggregated to emphasize particular equity goals (e.g., by socio-demographic groups, trip purpose, mode).
- (3) *Bus Crowding.* The *level of service* (LoS) experienced by public transit passengers has a direct influence on short- and long-term demand for public transit service. In addition to cost and travel time factors, the available capacity on a transit vehicle affects whether or not a passenger can board a public transit vehicle at their desired time as well as the level of comfort they experience during the trip. Though the LoS of public transit may be measured in BISTRO by any one of the factors mentioned, BISTRO includes a ready-made example of an LoS KPI related to passenger comfort: average bus crowding experienced. This metric is computed as the average overall transit legs of the total passenger-hours weighted by VOT multipliers corresponding to the load factor (the ratio of total passengers to the seating capacity) of the bus during the leg.
- (4) *Vehicle Miles Traveled (VMT) and Delay.* The BISTRO KPI library includes three examples of congestion score components that provide insight into the destination- or opportunity-independent level of mobility on a network, the overall network performance, and efficiency: total VMT by all motorized vehicles in the transportation system, total vehicle delay, and average vehicle delay experienced per passenger trip. Total vehicle delay is calculated as the sum over all path traversals of the difference between the realized duration and the free flow travel time of the traversal. Vehicle delay experienced per passenger trip is calculated as the total difference between the realized duration and free flow travel time of all legs of a trip completed by modes subject to congestion.
- (5) *Financial sustainability.* Most system interventions will have some impact on the flow of funds in or out of the transportation system, including a KPI that helps stakeholders

¹¹For example, a transit mode choice for a particular trip may result in more PMT than a walk mode choice for the same trip, as the sidewalk network may enable a more direct path.

understand the general financial impacts of such interventions is necessary. The financial sustainability metric provided in the BISTRO KPI library is the sum of all public transit fares collected minus all incentives distributed (if any) and all operational costs of the public transit system.¹² In the event that a BISTRO project does not alter public transit service, the operational costs may be omitted from the KPI, if desired.

- (6) *Environmental sustainability.* The environmental sustainability of a transportation system intervention may be measured as the local and/or global impacts to the system. In addition to VMT- and fuel efficiency-based estimates, BEAM enables estimation of emissions directly from the simulated fuel consumption, based on the realized speeds traveled by each vehicle throughout a simulation run.¹³ The VMT-based fine particulate emissions (PM_{2.5}) KPI captures local environmental sustainability via a mileage-based measure of air quality impacts based upon vehicle type. Additional local emissions KPIs may easily be included using the appropriate emissions factors.¹⁴ A *greenhouse gas* (GHG) emissions KPI allows the optimization to explicitly account for fuel-consumption-based global environmental sustainability.¹⁵

3.2.2 Scoring Function. The BISTRO scoring function serves as the objective function by which the UDIs are optimized. The selection and/or definition of the objective function in accordance with the project directives is considered to be the responsibility of the project planner. Herein, a general structure is defined to facilitate the creation of custom objective functions. Multiple project objectives (referred to here as *score components*) may be included in the scoring function—either as individual elements within a vector of scalar-valued score components to be minimized, or as parameters to a function that aggregates the objectives into a one-dimensional scalar score. The score components are computed as the normalized ratio of the value of the corresponding KPI in the given simulation run to the value of the same KPI in the *business-as-usual* (BAU) run.¹⁶ The improvement ratios are normalized using KPI values produced by a randomized sample of the UDI space, the size of which can be defined by the BISTRO project owner. This normalization (depicted graphically in Figure 4) accounts for differences in variance across KPIs, thus allowing the score components to provide meaningful feedback on the improvement achieved for each KPI relative to the distribution of the ratios of KPI to BAU produced by the random search. The composite score is thus a function of the normalized relative improvements of the candidate input to the BAU in each metric, as follows:

$$F(\vec{C}_s, \vec{K}, \vec{\sigma}, \vec{\mu}, \vec{\alpha}) = f(\vec{z}, \vec{\alpha}), \quad (4)$$

¹²The operational costs include the total costs of fuel consumed, and hourly variable costs of bus operations (see Table 3 for an example of operational costs). Hourly variable costs include estimated labor, maintenance, and operational costs. The rates for each of these factors is specified in the vehicle fleet configuration variables.

¹³For more information on the methodology followed to estimate fuel consumption, please refer to the BEAM documentation <https://beam.readthedocs.io/en/latest/index.html>.

¹⁴For more information on the methodology followed to develop this metric, please refer to the California Air Resources Board documentation, https://www.arb.ca.gov/cc/capandtrade/auctionproceeds/cc_i_emissionfactordatabase_documentation.pdf.

¹⁵It is important to note that the GHG emissions KPI will be correlated with VMT and fine particulate emissions. Thus, inclusion of all three KPIs creates a suite of environmental sustainability metrics that may apply disproportionate weight on environmentally-related objectives, which may or may not be desirable for certain policy agendas. Project planners may choose to apply scaling factors (as described in Section 3.2.2) to balance the influence of the environmental sustainability score components.

¹⁶In the BAU of a given scenario, the simulation is run without alteration from the initial configuration of that scenario.

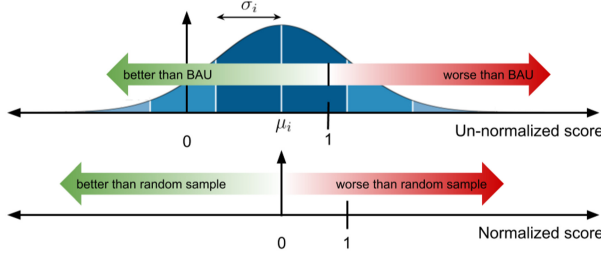


Fig. 4. A visual representation of the normalization procedure for a hypothetical score component, i . The ratio of the submission score and BAU score (depicted in the upper plot) is normalized by taking its z-score (depicted in the lower plot) relative to a random input sample.

where \vec{K} is the vector of all KPIs evaluated for a given set of inputs; \vec{C}_s ; $\vec{\mu}$ and $\vec{\sigma}$ are the vectors of normalization parameters; and \vec{z} is a vector of each KPI's z-scores, i.e.,

$$z_i = \frac{\frac{K_i(C_s)}{K_i(C_{BAU})} - \mu_i}{\sigma_i}, \quad (5)$$

for the i -th KPI. The value of the i -th score component in the BAU case is simply $K_i(C_{BAU})$.

The default objective is to minimize the composite score function, since an increase in many of the score components actually represents a scenario that is worse than the *status quo* (e.g., decreasing VMT over BAU results in a lower unscaled score than increasing VMT). To maintain consistency in this regard, the scoring function may include an additional parameter $\vec{\alpha}$ to allow for transformation of score components that are positively related to desirable outcomes (e.g., improvements in accessibility). For example, if the scoring function takes the form of a sum over all score components, the parameter $\vec{\alpha}$ may be used as a coefficient of each score component that determines whether the component will be summed or subtracted, as follows:

$$\alpha_i = \begin{cases} -1 & \text{if it is desirable for score component } i \text{ to increase} \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

This approach is distinct from the one typically used for *Cost-Benefit Analysis* (CBA) tasks in urban planning practice, in that it seeks to optimize an aggregate function of the relative improvements in each KPI rather than optimizing the net improvement from all KPIs. While CBA often draws skepticism due to the discretion inherent in the process of converting all KPIs into a common unit such as time or money so that the net value of costs and benefits can be computed, the approach taken in the BISTRO scoring function does not require any such assumptions to be made. Rather, each score component represents the relative improvement over the BAU that is achieved by a simulation run using a particular set of UDIs. Objective function designers may choose to apply additional scaling factors to the score components using the $\vec{\alpha}$ parameter vector.

3.3 Inputs

Preparation of fixed inputs. For each BISTRO study, a set of fixed inputs must be provided to BEAM. For a given study area, these typically include the road network, the transit schedule, and the demand profile. Depending on the system objectives, additional data may be necessary to fully configure the simulation. Figure 3 illustrates a schematic of the inputs for a typical simulation.

The *road network*, including the physical properties of its links and nodes, may be generated using *Open Street Maps* (OSM) data for the geography of interest. The *transit network* configuration follows the easily accessible *General Transit Feed Specification* (GTFS) format. *On-demand ride*

services (Transportation Network Companies (TNCs) such as Uber, Lyft, and Via), are modeled as a fleet of vehicles driven by agents that are exogenous to the population, or may be driven autonomously. The initial locations of the vehicles may be sampled randomly or from a specified distribution in accordance with appropriate data. The price of on-demand rides is fixed, consisting of a distance-based and a duration-based component. The size of the on-demand ride service fleet is a proportion of the total number of agents in the simulation, as determined by a configuration parameter. Driver repositioning behavior, when not currently driving to or serving a passenger, can be configured to follow one of several repositioning algorithms defined within BEAM.

At the start of the simulation, a *synthetic population* of virtual agents and households is generated such that the sociodemographic attributes of these virtual entities are spatially distributed in accordance with real-world census and/or location-based data from the city of interest. Each agent follows a daily plan consisting of several activities throughout the day. As illustrated in Figure 3, these *daily activity schedules* are generated based on *origin-destination* (OD) skims (matrices that provide the number of trips between zones), travel surveys, and zonal boundary spatial data.

Calibration. Prior to use in optimization runs, BEAM needs to be calibrated to empirical data by iteratively adjusting model parameters until a simulation outputs representing traffic patterns match their real-world counterparts with minimal error.¹⁷ Calibration of BEAM for usage in BISTRO should be limited to adjustments of behavioral parameters controlling microscopic decision-making (e.g., mode choice intercepts, prices, regulation-driven incentives/tariffs).¹⁸ While it is possible to adjust many additional BEAM parameters to reduce calibration error, this practice should be discouraged, as it may result in models that are overfitted to a state of the world represented by a particular ground-truth dataset, thereby limiting the calibrated simulation model's use in predictive contexts. In addition to the representativeness of ground truth data, the quality and quantity of input data (e.g., network data resolution or population spatial resolution), may influence the extent to which the model is able to achieve calibration end points.

Configuration of UDIs. BISTRO provides a library of possible inputs for scenario designers to adapt to specific use cases. The selection of UDIs is intended to be compatible with the system objective. UDIs may represent, for example, the investment (e.g., transit fleet mix modification, bus route modifications, parking supply, electric vehicle charge station locations, dynamic redistribution of e-bikes or on-demand vehicles), incentive (e.g., incentives to specific socio-demographic groups for selected transportation modes, road pricing/toll roads, fuel tax), or policy/operational (e.g., transit schedule adjustment, transit fare modification, parking pricing) levers applicable to the study at hand. The project owner may constrain the range of possible values upon which each UDI is valid by setting the corresponding input validation parameters and business rules. The example input file for bus scheduling shown in Table 1 defines alteration of the headway of a particular bus route during a particular service period (defined by its start and end times).

3.4 Output Analysis and Visualization

The *raw* outputs of a BEAM simulation include millions of events that reflect the microscopic actions of each agent as they make their way through the day. While this is a detailed history of what transpired during the day for each agent, it does not provide planners with explanatory insight

¹⁷BEAM calibration is typically targeted at mode split, volumetric traffic counts, and travel distance distributions. The choice of which target(s) to use may depend on regulatory requirements, literature recommendations, or precedent [2, 54].

¹⁸Often, due to computational constraints, a sub-sample of a full population is simulated. The capacity of physical resources (e.g., road network link carrying capacity, maximum transit occupancy, number of electric vehicle charging plugs per station) may need to be adjusted based on the size of the population sample. For evaluation purposes, the outputs of a sub-sampled simulation are often scaled back up to the full population.

Table 1. Example of Bus Frequency Adjustment Input File

route_id	start_time	end_time	headway_secs	exact_times
1,340	21,600	79,200	900	1
1,341	21,600	36,000	300	1
1,341	61,200	72,000	300	1

into how perturbations of input variables influence model behavior leading to changes in output responses. To simplify exploration of alternatives to purely black box optimization methods, the BISTRO platform provides a suite of tools that process and organize event data from BEAM simulation runs into a relational format, permitting relationships between each person's activities, trips, path traversals to be queried at different levels of detail.¹⁹ A Jupyter notebook then incorporates the post-processed outputs into a standard suite of multivariate analyses and visualizations, thereby facilitating interpretation and communication of the effect that policies have on system objectives.

3.5 Implementation Details and Performance Characteristics

Both BISTRO and BEAM are primarily implemented in Scala. Input files are read from a single directory and injected into the BEAM initialization routine. The system is containerized using Docker, which helps to facilitate OS-agnostic local and remote execution.

The runtime of BEAM depends on various inputs including population size, network resolution, transit network, ridehail fleet size, as well as available compute resources such as number of processors, memory, and so on. On a machine with 32 (2.5GHz Intel®Xeon®Platinum 8175) CPUs and 128GB RAM, the runtime for a 315,000 agent simulation of a San Francisco Bay Area scenario (representing a 25% sample of the approximate 2018 population) takes around 14 hours to complete 15 iterations. On a machine with identical compute resources, a 15,000 agent simulation for the the Sioux Faux scenario (see Section 4.1) takes approximately 30 minutes to complete 15 iterations.

Currently, the primary performance bottleneck in BEAM is routing. The routing engine generates millions of routes (reflecting multimodal options for agents to choose between) for a single simulation run. Some additional overhead considerations such as data availability, level of model resolution required, as well as the impact of augmented BEAM functionality must be balanced in light of available computational resources.

4 INITIAL PILOT STUDY AND LAUNCH

4.1 Sioux Faux

An agent-based model of transport supply and demand inspired by the real city of Sioux Falls, South Dakota²⁰ was adapted for the purpose of developing and testing example scenarios within BISTRO. To underscore that for these purposes, such scenarios were not developed to be true replicas of the city of Sioux Falls, this benchmark BISTRO scenario is referred to as *Sioux Faux*. The scenario configuration, input specification, and scoring function were designed to support strategic objectives of financial and environmental sustainability, reduced congestion, and improved equity, accessibility, and transportation system level of service.

¹⁹Further details on the output of the parser including an entity relationship diagram are available at http://bistro.its.berkeley.edu/assets/download/pdfs/General_System_Specification.pdf.

²⁰The "Sioux Falls" scenario is a commonly used benchmark in ABMS research; see <https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls>.

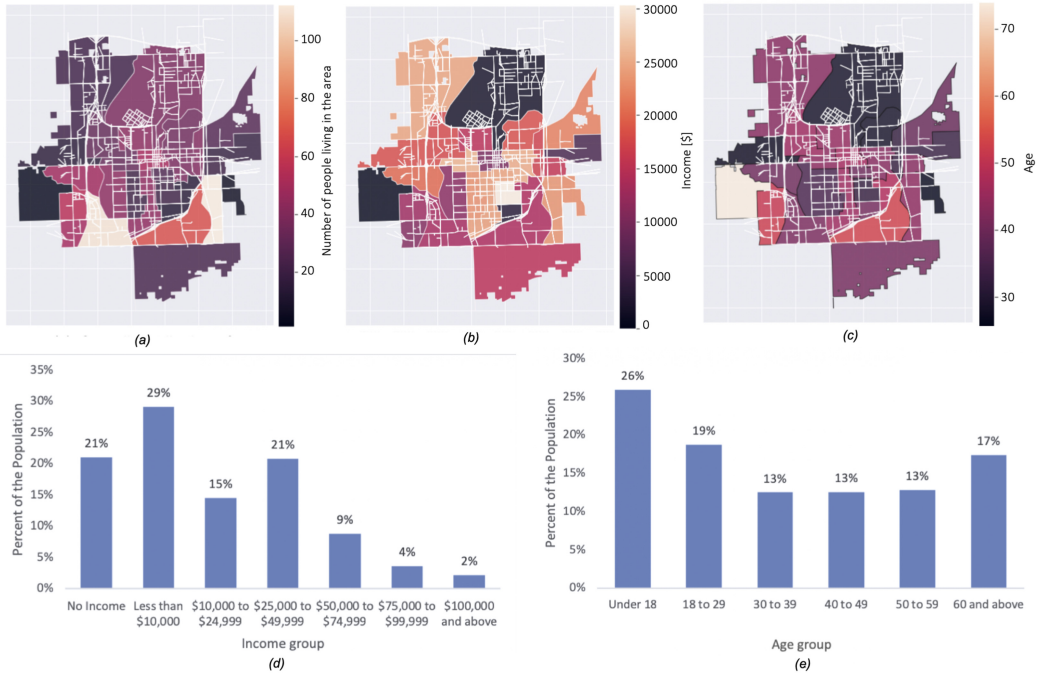


Fig. 5. Demographics of Sioux Faux. (a) Overall distribution of the population per census tract. (b) Distribution of the median population income per census tract. (c) Distribution of the median population age per census tract. (d) Overall population income distribution. (e) Overall population age distribution.

4.1.1 Scenario Configuration.

Population and plan synthesis. The synthetic population of Sioux Faux was generated using publicly-available census data for the city of Sioux Falls, South Dakota as inputs to the Doppelganger library,²¹ a state-of-the-art population synthesis framework developed in Python. Specific inputs to Doppelganger used to generate the Sioux Faux population included household and individual *Public Use Microdata Sample* (PUMS) data for South Dakota from the 2012-2016 (5-year) *American Community Survey* (ACS), which is conducted annually by the US Census.²² The *Public Use Microdata Area* (PUMA) for Sioux Falls constrains the state-wide survey data to our general area of interest. Population demographics derived from the synthetic population for Sioux Falls are shown in Figure 5.

An existing set of agent plans for Sioux Falls previously developed for MATSim simulations was used as the basis for the plans of our expanded Sioux Faux population.²³ After initial pilot testing to determine tradeoffs between population size, behavioral realism, and computational complexity, we took a 15% sub-sample of the full synthetic population (approximately 15,000 agents). We

²¹Doppelganger uses a novel Bayesian optimization approach combined with the hierarchical list-balancing algorithm developed as part of the PopSyn library [55]. For more information about the Doppelganger library, see <https://github.com/sidewalklabs/doppelganger>.

²²The 5-year PUMS comprises a 5% sample of the US population. It is computed as an aggregate of 1-year samples, which themselves aim to survey 1% of the US population.

²³More specifically, we modified the Sioux Falls 2016 scenario developed by Hörl ([4]), which is an update of a scenario prepared in 2014 by Chakirov and Fourie [56]. For more information on the Sioux Falls scenario, see <https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/901-1000/ab978.pdf>.

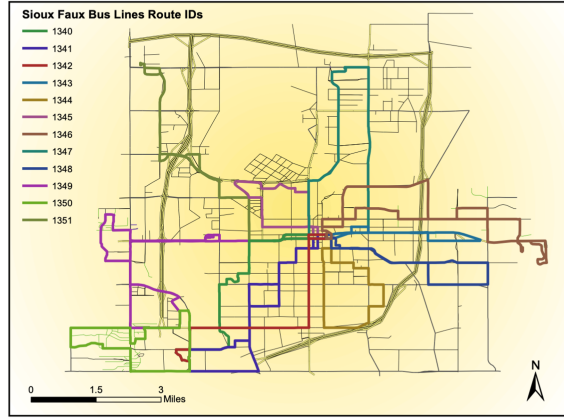


Fig. 6. Sioux Falls bus and road networks.

used a spatially-constrained sampling mechanism in order to allocate plans to agents in accordance with household locations and census tract household and individual attribute distributions. The subsampling mechanism also enforces realistic constraints on agent plans and behavior using predicates such as “agents under the age of 18 should not have a work activity” and “agents under the age of 16 should not be allowed to drive.”

Transportation Network. The Sioux Falls transportation network (Figure 6) includes a road network accessible to walking agents, personal vehicles, on-demand ride services (TNCs such as Uber and Lyft), and public buses providing fixed-route service.²⁴ The on-demand ride services implemented in this scenario include only single-passenger rides (e.g., UberX, Lyft Classic) from a fleet of on-demand ride vehicles that was distributed randomly across the road network at the start of each simulation run. Driving alone is the most frequently used mode, comprising approximately 75% of the miles traveled for the BAU scenario.

4.1.2 User-Defined Input Specification. For this initial pilot study, a set of four UDIs were investigated:

- (1) bus fleet vehicle composition
- (2) bus service frequency
- (3) bus fare
- (4) multimodal incentive program for on-demand rides and public bus trips

In the bus fleet for the BAU scenario, all vehicles were set to a default bus type. Optimization of the bus fleet vehicle composition and service frequency offers the opportunity to improve the level of bus service by better matching the bus type with specific demand characteristics of each route. Four types of buses (including the default) were considered (see Table 3), each with different technical properties (seating and standing capacity) and cost characteristics (cost per hour, cost per mile, fuel type, and fuel consumption rate).

A UDI was implemented to vary the bus schedule on each route, including the hours of service and the headway, or service frequency as shown in Table 1. Multiple service periods with varying headways on the same route were thus possible. The bus fare UDI allowed for the optimization of

²⁴The initial bus route scheduling is directly generated from the publicly available GTFS for Sioux Falls, which includes erratic headways across routes.

Table 2. Values Used for α_i in Each of the Subsequent Results Sections

KPI	KPI type	Contest	Post-Contest	New KPIs
accessible work locations	Accessibility	-1	-1	-
accessible secondary locations	Accessibility	-1	-1	-
accessible work locations by car	Accessibility	-	-	-1
accessible secondary locations by car	Accessibility	-	-	-1
accessible work locations by transit	Accessibility	-	-	-1
accessible secondary locations by transit	Accessibility	-	-	-1
average trip expenditure-work	LoS	1	1	-
average trip expenditure-secondary	LoS	1	1	-
average travel cost burden-work	Equity	-	-	1
average travel cost burden-secondary	Equity	-	-	1
average bus crowding experienced	LoS	1	1	1
total vehicle miles traveled	Congestion	1	1	1
average vehicle delay per passenger trip	Congestion	1	1	1
costs and benefits	Financial Sustainability	-1	-1*	-1*
total grams PM _{2.5} emitted	Environmental Sustainability	1	1	1
total grams GHG _e emitted	Environmental Sustainability	-	-	1

For score components that are positively related to desirable outcomes, negative α_i is provided to transform it consistent with a minimization problem.

*fixed KPI post-contest.

the fare on each route, segmented by passenger age groups. Finally, a multimodal incentive UDI was implemented to enable reimbursement for on-demand rides, walk to/from transit, or drive to/from transit trips to qualifying individuals based on age, income, or both.

4.1.3 Business Rules. In order to ensure that optimal solutions would be compliant with common policy and planning practices, four business rules were implemented: (1) there may be no more than five distinct bus service periods (this mimics a typical delineation of transit service provision: a.m. peak, midday, p.m. peak, evening, late night/early morning); (2) bus route headways may be no more than 120 minutes and no fewer than 3 minutes; (3) bus fares and mode incentives may not isolate a single age; and (4) ages for both fares and incentives may be specified in segments no smaller than five years in range and income for incentives may be assigned in segments no smaller than \$5,000 in range.

4.1.4 Scoring Function Design. The set of Sioux Faux UDIs have varying interconnected impacts on the operation of and access to public transit and on-demand ride service by agents. Thus, the scoring function upon which the inputs were optimized was designed to include a variety of metrics that relay feedback on the user experience and operational efficiency of the transportation system as a whole. Table 2 presents all of the KPIs used in Sioux Faux scenarios referenced in this text.²⁵

Five KPIs were developed to represent three main aspects of user experience: accessibility, travel expenditure, and transit passenger comfort. The accessibility and travel expenditure were both

²⁵Note that several of the KPIs in this table refer to two post-contest follow-on studies; see Section 4.2.

Table 3. Transit Vehicle Types Available for Sioux Faux Bus Fleet

Vehicle type, $c \in C$	a	b	c	d	e
BUS-DEFAULT	diesel	20048	89.88	37	20
BUS-SMALL-HD	diesel	18043.2	90.18	27	10
BUS-STD-HD	diesel	20048	90.18	35	20
BUS-STD-ART	diesel	26663.84	97.26	54	25

(a) Fuel Type, (b) Fuel Consumption Rate (J/m), (c) Operational Cost (USD/hr), (d) Seating Capacity, (e) Standing Capacity.

disaggregated by trip purpose such that score components for accessibility and travel expenditure to work and secondary activities were each included separately in the scoring function. Transit passenger comfort was measured as the average bus crowding experienced by bus passengers.²⁶

Four KPIs of operational efficiency were included to account for the congestion, environmental sustainability, and financial sustainability resulting from optimized inputs. Total VMT was included as a KPI for overall congestion while average vehicle delay per passenger trip served as a KPI of the average impact of congestion. The total amount of PM_{2.5} emitted served as a KPI of the environmental impact resulting from each simulation run. Finally, the financial sustainability KPI was included to incentivize outcomes with minimal impact to the bottom line of the transit agency by taking into account the operational costs, incentives distributed, and revenues collected from any combination of transit fleet mix, scheduling, fare structure and incentive program.

All metrics are aggregated according to the following function:

$$F(\vec{C}_a, \vec{F}, \vec{\sigma}, \vec{\mu}, \vec{\alpha}) = \sum_{i \in \mathbb{K}} \frac{\left(\frac{K_i(C_s)}{K_i(C_{BAU})} \right)^{\alpha_i} - \mu_i}{\sigma_i}, \quad (7)$$

where all variables are defined as described in Section 3.2.2, with the set of KPIs and corresponding α_i values as specified in Table 2. We executed 800 runs using randomly generated values of \hat{d} to produce the normalizing statistics (i.e., μ_i s and σ_i s in Equation (5)) for each metric.

4.2 Pilot Study Results

Contest participation and results. Over the course of 17 days, 487 people in teams of one to four (mostly consisting of engineers and data scientists with little to no domain expertise in transportation planning) effectively created nearly 1,000 different “city transportation plans” for the Sioux Faux scenario consisting of the UDIs described in section Section 4.1.2.²⁷

To be able to compare their results and scores with other participants, each team could submit up to five solutions per day and thus be ranked in a web-accessible leaderboard. While contestants trained algorithms online, final evaluation, leaderboard, and discussion boards were hosted by AICrowd.com.²⁸ Inputs from top teams were evaluated 5 times for 100 iterations each in order to achieve a consistent final score.

Figure 7 illustrates the evolution of submissions over time during the competition. Participation developed in two phases. During the first week, contestants became familiar with the BISTRO framework and the Sioux Faux transportation optimization problem. During the second phase, contestants continued to optimize their solutions.

²⁶ Average bus crowding in the Sioux Faux scenario was calculated as the average number of agent hours spent per transit trip in buses occupied above their seating capacity. This KPI has since been updated; see Section 3.2.1.

²⁷ Uber does not endorse any of the solutions presented.

²⁸ <http://www.aicrowd.com>.

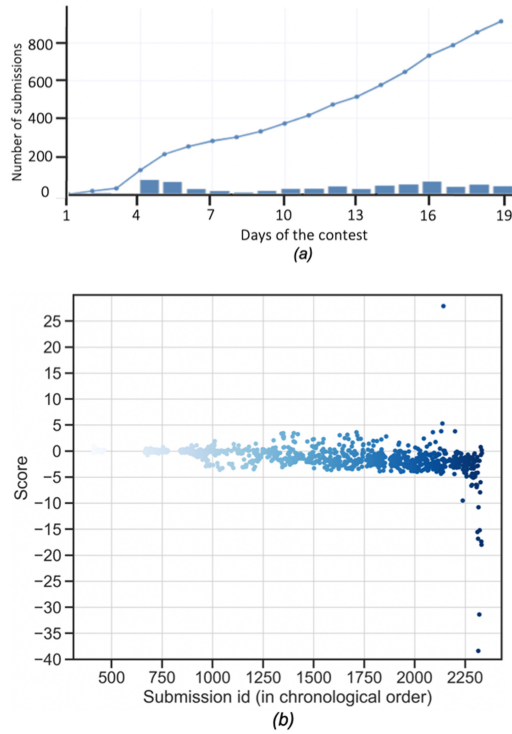


Fig. 7. Participation history. (a) Number of solutions submitted over time (per day and cumulative). (b) Evolution of scores over time (see Equation (7)).

Table 4. Proportion of Algorithmic Approaches Used, According to a Survey Conducted with Contestant Teams

Approach	Proportion
Bayesian optimization	34%
Evolutionary algorithms	28%
Gradient based	14%
Meta-heuristics	7%
Plackett-Burman design	3%
Hill climbing	3%
Other	10%

According to code submissions and a post-contest survey, the solutions that achieved the highest value of the objective function (averaged over five replications) followed similar strategies. Typically, they used domain-specific analysis to prune the large input space. For example, by sampling the age distribution within walking distance from a bus route’s stops, a reasonable bound on the components of the fare UDI could be assigned. Following these informed factor screening steps, contestants used a variety of algorithmic approaches to more efficiently search the lower dimensional design space. As shown in Table 4, the black-box global optimization techniques used during the Contest primarily incorporated variants of Bayesian optimization, genetic/evolutionary algorithms, gradient-based techniques, and meta-heuristics methods.

Most teams managed to improve their scores by three standard deviations better than a random search benchmark (i.e., with scores of approximately -3). Due to an insidious modeling deficiency, the financial sustainability score component could be optimized toward negative infinity. As such, any contributions from other score components would be relatively inconsequential. Two teams discovered input settings that took advantage of the lack of a lower bound on the financial sustainability score component and were thus able to reach extremely low scores of -30 or -40 . This experience underlines the importance of developing careful theory supplemented by judicious testing when designing objective functions.

Post-Contest and New KPIs follow-on studies and results. After the contest, two follow-on studies were conducted to interpret the solutions from the top algorithms in the context of improvements to the objective function. An initial set of improvements, hereby referred to in this text as the “Post-Contest” objective function study, simply addressed the unbounded financial sustainability score component as well as other minor problems discovered during the competition. The “New KPIs” objective refers to an expanded set of KPIs, summarized in Table 2. Two of the best-performing algorithms from the Contest—namely, Bayesian Optimization using TPE [42] and *Genetic Algorithms* (GA) [57]—were adapted and re-implemented to run BISTRO on the Sioux Faux scenario with both of the updated objective functions. As a baseline algorithmic benchmark, *random search* (RS) was performed for 800 trials using both objective functions. The GA assessment on both the “Post-Contest” and “New KPIs” objectives utilized five parallel evolutionary trajectories, each drawing a random sample of seeds from a larger gene pool. Both TPE and GA were run over identical design spaces for 1,400 trials on the “Post-Contest” and “New KPIs” studies.²⁹

Inputs corresponding to the trial yielding the top score for each algorithm were then simulated for 100 iterations with five replicates per trial. Figure 8 demonstrates that both GA and TPE produce input configurations that are superior to RS. We found that both GA and TPE achieve optimal solutions that reflect sensible yet distinct transportation system management strategies. For example, Figure 9 illustrates the relationship between activity start times (top subfigure) and bus utilization (bottom subfigure) as computed using outputs for the highest scoring solution for GA and TPE algorithms when using the “Post-Contest” objective function. This figure suggests that differences in optimal TPE and GA solutions for the “Post-Contest” objective arise from distinct transit usage patterns. Note that “Work” activity start times occur in the early morning (between 7:00 and 10:00 a.m.) and correspond to the highest period of utilization for buses under the optimal solution found for the TPE algorithm. In contrast, the GA solution ensures that buses are available during the evening peak commute time; that is, when agents travel back home from work and/or engage in secondary activities.

In the case of TPE, we found that the algorithm produced solutions that corresponded to sensible real-world policies. Figure 10 presents visualizations of input distributions for the top fifth percentile of TPE trials (corresponding to the top 70 of 1,400 evaluated solutions by score, as plotted on Figure 8). This figure illustrates that the values of components of \hat{d} for the best performing (lowest scoring) solutions using TPE occupy a narrow band in the design space. For example, in Figure 10(a), the highest scoring TPE input value sets evaluated in BISTRO under the “Post-Contest” objective suggest charging more expensive bus fares (\$8–\$10) for adult citizens (16–60) than for youth (1–15) (\$4–\$6) and elderly (60–120) (\$5–\$7). The low variance of the components of \hat{d} for these trial points is indicative of both objective function sensitivity to UDI definitions as well as robust algorithm convergence to a (locally) minimal score value. The corresponding bus types

²⁹Partial convergence criteria of 40 iterations were used during initial search, as this was determined to be sufficient for establishing a trajectory consistent with a fully relaxed state.

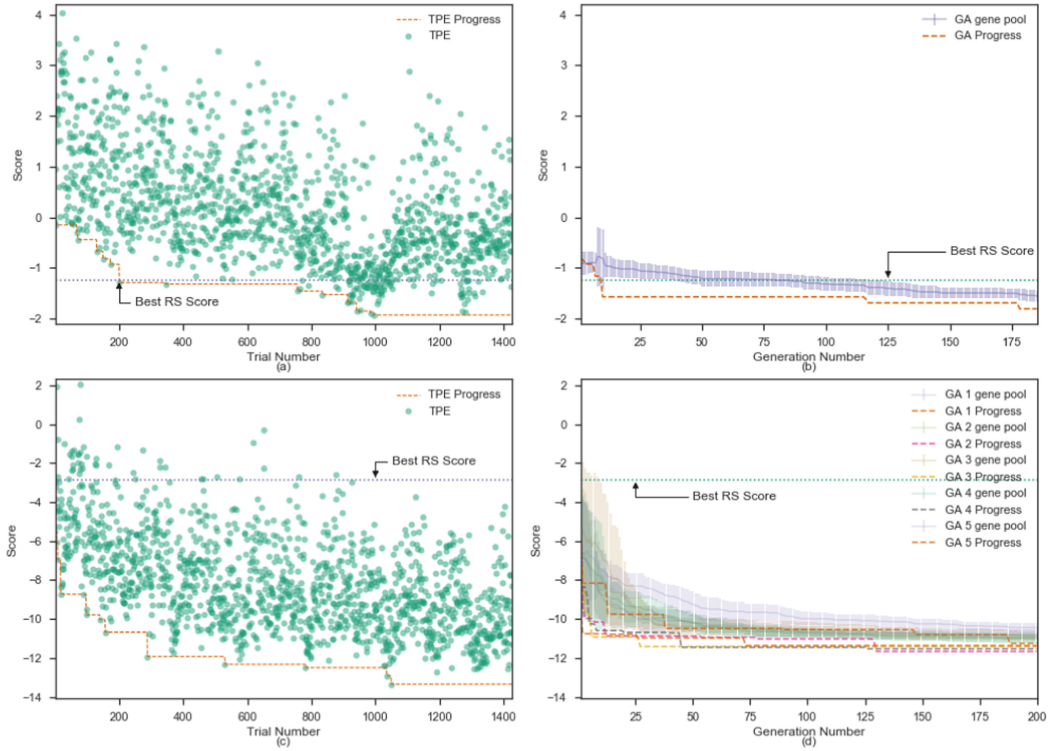


Fig. 8. Optimization of the Sioux Faux 15k scenario with TPE (left) and GA (right) using “Post-Contest” (top) and “New KPIs” (bottom) objective function settings. The dashed line(s) across the bottom of each denotes the best (lowest) score achieved by an algorithm within the first N trials. Individual trial scores (at 40 iterations) are shown for TPE plots, whereas one standard deviation ranges of current gene pools are displayed in the GA plots. For the “Post-Contest” objective, the TPE and GA algorithms surpass the best score from 800 RS trials of 40 iterations (-1.24) within 200 and 10 trials, respectively. For the “New KPIs” objective, both algorithms significantly outperform the best result (-2.84) of an 800 trial, 40 iteration RS almost immediately.

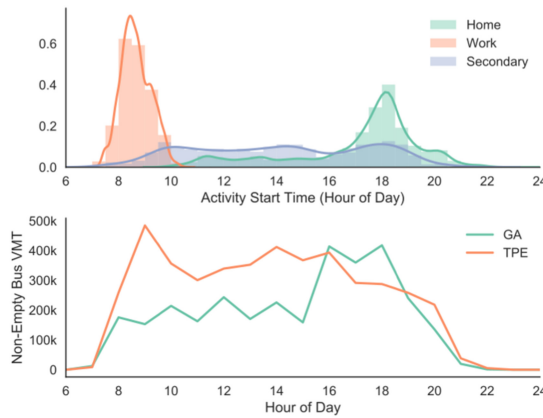


Fig. 9. Example of output analysis for the “Post-Contest” case study. The upper plot shows the various activity start times of agents by activity type. The lower plot shows non-empty bus VMT for two competing algorithms.

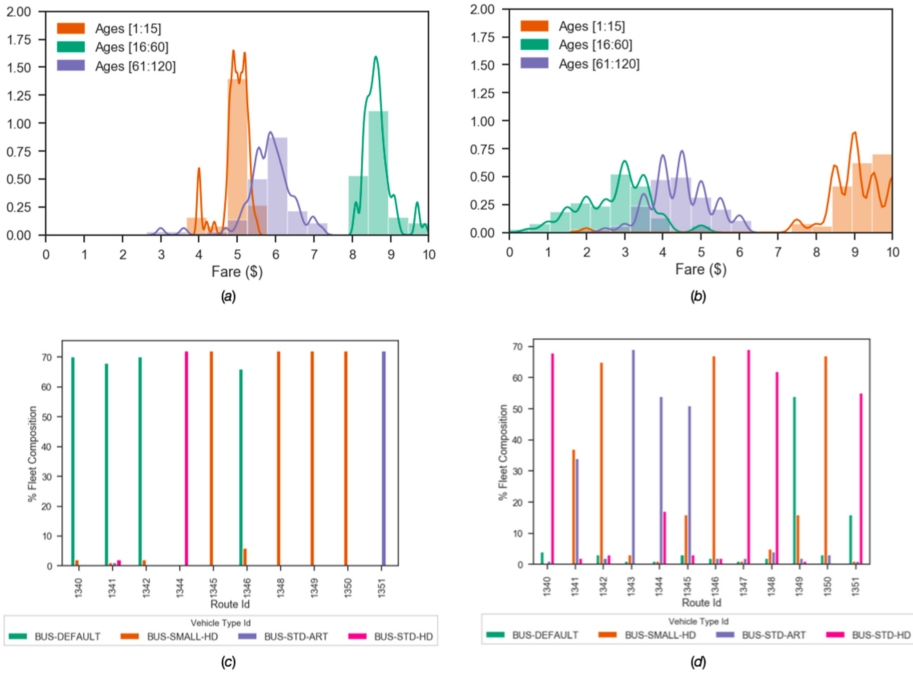


Fig. 10. Distributions of bus fare by age (top) and vehicle fleet mix by route (bottom) for inputs representing the best fifth percentile scores among trials run for Sioux Faux 15k scenario using the TPE algorithm, shown for “Post-Contest” (left) and “New KPIs” (right) objective functions.

on a given route suggested by these solutions are well-resolved and tend toward smaller vehicle models. In contrast, for near-optimal inputs evaluated using the “New KPIs” objective function, fares assigned to youth (\$4–\$10) are, on average, higher than those assigned to adults (\$0–\$4) and seniors (\$5–\$10). The corresponding bus types by route are also more diverse among optimal solutions, indicating that the objective is less sensitive to the VehicleFleetMix input when evaluated using the “New KPI” objective function. Using the “New KPIs” objective, GA (not shown) also finds a tight distribution of fares for top-performing solutions but contrarily finds diversity in its VehicleFleetMix solutions.

5 CONCLUSION

This article has presented the design, software architecture, and preliminary evaluation of BISTRO: a general-purpose transportation policy decision support tool and scenario-based optimization framework supported by empirically-driven agent-based models. When combined with sensible guidance from experienced planning professionals, BISTRO can be used to identify more holistic, empirically-driven approaches to urban transportation planning and management. In addition to overall system purpose, design, and software architecture, this work provides a concrete example of the process that BISTRO supports as implemented in the context of a scenario-based policy optimization “contest.” While many participants had little or no prior expertise in the transportation science and policy analysis methods typically used in urban planning practice, over a dozen teams developed algorithms that found inputs, which, when evaluated in the simulator, achieved scores that surpassed both random search as well as human judgment. The mixed results of the competition led us to conclude that the optimization-based search techniques enabled by BISTRO should

support an iterative approach that involves applying optimization algorithms to refinements of KPI specifications in order to better align objective functions with system goals.

Research conducted using BISTRO strives to meet the highest standard of reproducibility in computational experiments [58–60] as well as fact-based policymaking [61] by making all data, models, and algorithms freely available and open source.³⁰ One finding of post-contest reproducibility efforts was that different classes of algorithms appeared to converge to solutions that emphasized distinct policy strategies.

Our experience from this pilot study demonstrates that we have implemented a compelling platform to study human-in-the-loop design of expensive simulation-based optimization algorithms. This conclusion suggests that, in addition to its utility as a decision support system, BISTRO could serve as an exemplary testbed for multiple emerging streams of research (e.g., freeze-thaw, multi-objective, multi-task, and multi-fidelity optimization) in SMBO and associated meta-model-based optimization methods. Should BISTRO be widely adopted as part of the urban planning toolkit, innovative algorithms and new theory developed as part of inquiry in these sub-domains will have the added benefit of directly serving a humanitarian purpose.

ACKNOWLEDGMENTS

The authors wish to thank the Uber employees who participated in the contest and provided valuable feedback.

REFERENCES

- [1] Mogeng Yin. 2018. *Activity-based Urban Mobility Modeling from Cellular Data*. Ph.D. Dissertation. UC Berkeley. <https://search.proquest.com/docview/2101657212?accountid=49007>.
- [2] Hong Zheng. 2013. A primer for agent-based simulation and modeling in transportation applications. *Development* (2013), 75.
- [3] Andreas Horni, Kai Nagel, and Kay W. Axhausen. 2016. *The Multi-agent Transport Simulation MATSim*. Ubiquity Press London.
- [4] Sebastian Hörll. 2016. Implementation of an autonomous taxi service in a multi-modal traffic simulation using MATSim Master thesis in Complex Adaptive Systems. June (2016). DOI : <http://dx.doi.org/10.13140/RG.2.1.2060.9523>
- [5] Michael Grant, Janet D'Ignazio, Alexander Bond, Alanna McKeeman, et al. 2013. *Performance-based Planning and Programming Guidebook*. Technical Report. United States. Federal Highway Administration.
- [6] United States Department of Transportation. 2012. Trends in Statewide Long-Range Transportation Plans: Core and Emerging Topics. Retrieved from <https://www.planning.dot.gov/documents/State>.
- [7] Federal Highway Administration/Federal Transit Administration. 2019. Public Engagement: Case Studies and Notable Practices. Retrieved from https://planning.dot.gov/focus_caseStudies.aspx.
- [8] Gian-Claudia Sciara. 2017. Metropolitan transportation planning: Lessons from the past, institutions for the future. *Journal of the American Planning Association* 83, 3 (2017), 262–276. DOI : <http://dx.doi.org/10.1080/01944363.2017.1322526> arXiv:<https://doi.org/10.1080/01944363.2017.1322526>
- [9] Federal Highway Administration/Federal Transit Administration. 2016. *The Transportation Planning Process Key Issues: A Briefing Book for Transportation Decisionmakers, Officials, and Staff*. Technical Report. Washington, DC.
- [10] United States Government Accountability Office. 2009. *Report to the Ranking Member, Committee on Environment and Public Works: U.S. Senate—Metropolitan Planning Organizations: Options Exist to Enhance Transportation Planning Capacity and Federal Oversight*. Technical Report. Washington, DC.
- [11] Joshua Auld, Michael Hope, Hubert Ley, Vadim Sokolov, Bo Xu, and Kuilin Zhang. 2016. POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies* 64 (2016), 101–116.
- [12] Ebru Vesile Ocalir-Akunal. 2016. Decision support systems in transport planning. *Procedia Engineering* 161 (2016), 1119–1126. DOI : <http://dx.doi.org/10.1016/j.proeng.2016.08.518>
- [13] Zaheer Khan, David Ludlow, Wolfgang Loibl, and Kamran Soomro. 2014. ICT enabled participatory urban planning and policy development: The UrbanAPI project. *Transforming Government: People, Process and Policy* 8, 2 (2014), 205–229. DOI : <http://dx.doi.org/10.1108/TG-09-2013-0030>

³⁰ All code and data used and referred in the article is available at <http://bistro.its.berkeley.edu>.

- [14] Carlos Lima Azevedo, Ravi Seshadri, Song Gao, Bilge Atasoy, Arun Prakash Akkinipally, Eleni Christofa, Fang Zhao, Jessica Trancik, and Moshe Ben-Akiva. 2018. Tripod: Sustainable travel incentives with prediction, optimization, and personalization. In *Transportation Research Board 97th Annual Meeting*.
- [15] Maria Nadia Postorino and Giuseppe M. L. Sarné. 2016. Agents meet traffic simulation, control and management: A review of selected recent contributions. *CEUR Workshop Proceedings* 1664 (2016), 112–117.
- [16] Association of Metropolitan Planning Organizations. 2019. ActivitySim | An open platform for activity-based travel modeling. Retrieved from <https://activitysim.github.io/>.
- [17] Joe Castiglione, Mark Bradley, and John Gliebe. 2016. *Activity-Based Travel Demand Models: A Primer*. 7–8 pages. DOI : <http://dx.doi.org/10.17226/22357>
- [18] L. Smith, R. Beckman, and K. Baggerly. 1995. TRANSIMS: Transportation analysis and simulation system. In *Proceedings of the 5th National Conference on Transportation Planning Methods Applications—Volume II: A Compendium of Papers Based on a Conference Held in Seattle, Washington in April 1995 by the Transportation Research Board and Washington State Department of Transportation*. DOI : <http://dx.doi.org/10.2172/88648>
- [19] Michael Behrisch, Laura Bieker, Jakob Erdmann, and Daniel Krajzewicz. 2011. SUMO—simulation of urban mobility: An overview. In *Proceedings of the 3rd International Conference on Advances in System Simulation, SIMUL 2011*. ThinkMind.
- [20] Carlo Castagnari, Flavio Corradini, Francesco De Angelis, Jacopo de Berardinis, Giorgio Forcina, and Andrea Polini. 2018. Tangramob: An agent-based simulation framework for validating urban smart mobility solutions. *arXiv preprint arXiv:1805.10906* (2018).
- [21] Joseph Kamel, Reza Vosooghi, Jakob Puchinger, Feirouz Ksontini, and Göknur Sirin. 2019. Exploring the impact of user preferences on shared autonomous vehicle modal split: A multi-agent simulation approach. *Transportation Research Procedia* 37, September 2018 (2019), 115–122. DOI : <http://dx.doi.org/10.1016/j.trpro.2018.12.173>
- [22] Paul Salvini and Eric J. Miller. 2005. ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics* 5, 2 (2005), 217–234.
- [23] Paul Waddell. 2002. UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association* 68, 3 (2002), 297–314.
- [24] T. Nicolai and K. Nagel. 2012. Coupling transport and land use: Investigating accessibility indicators for feedback from a travel to a land-use model. In *Latsis Symposium*. 12–16.
- [25] Rolf Moeckel, Carlos Llorca Garcia, Ana Tsui Moreno Chou, and Matthew Bediako Okrah. 2018. Trends in integrated land use/transport modeling: An evaluation of the state of the art. *Journal of Transport and Land Use* 11, 1 (2018), 463–476. DOI : <http://dx.doi.org/10.5198/jtlu.2018.1205>
- [26] Thomas W. Nicolai and Kai Nagel. 2010. *Coupling MATSim and UrbanSim: Software Design Issues*. Technical Report. SustainCity Working Paper.
- [27] Dominik Ziemke, Kai Nagel, and Rolf Moeckel. 2016. Towards an agent-based, integrated land-use transport modeling system. *Procedia Computer Science* 83 (2016), 958–963.
- [28] US Department of Energy. 2018. *Energy Efficient Mobility Systems: 2018 Annual Progress Report*. Technical Report. US Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office. Retrieved from <https://www.energy.gov/sites/prod/files/2019/04/f62/VTO>.
- [29] Daniel Delling, Thomas Pajor, and Renato F. Werneck. 2014. Round-based public transit routing. *Transportation Science* 49, 3 (2014), 591–604.
- [30] Matthew Wigginton Conway, Andrew Byrd, and Marco van der Linden. 2017. Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and headway-based networks. *Transportation Research Record: Journal of the Transportation Research Board* 2653, 1 (2017), 45–53. DOI : <http://dx.doi.org/10.3141/2653-06>
- [31] Gul A. Agha, Prasannaa Thati, Reza Ziaei, H. Bowman, and J. Derrick. 2001. Actors: A model for reasoning about open distributed systems. *Formal Methods for Distributed Processing: A Survey of Object-oriented Approaches* (2001), 155–176.
- [32] Kenneth E. Train. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.
- [33] Moshe E. Ben-Akiva and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Vol. 9. MIT press.
- [34] Carolina Osorio and Michel Bierlaire. 2013. A simulation-based optimization framework for urban transportation problems. *Operations Research* 61, 6 (2013), 1333–1345.
- [35] Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente. 2009. *Introduction to Derivative-free Optimization*. Vol. 8. Siam.
- [36] Russell R. Barton and Martin Meckesheimer. 2006. Chapter 18 metamodel-based simulation optimization. *Handbooks in Operations Research and Management Science* 13, C (2006), 535–574. DOI : [http://dx.doi.org/10.1016/S0927-0507\(06\)13018-2](http://dx.doi.org/10.1016/S0927-0507(06)13018-2)

- [37] Carl Edward Rasmussen. 2003. Gaussian processes in machine learning. In *Summer School on Machine Learning*. Springer, 63–71.
- [38] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. 2015. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* 104, 1 (2015), 148–175.
- [39] Ian Dewancker, Michael McCourt, and Scott Clark. 2015. Bayesian Optimization Primer. (2015). DOI: <http://dx.doi.org/10.1016/j.jss.2013.05.010>
- [40] Peter I. Frazier. 2018. A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811* (2018).
- [41] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. 2011. Sequential model-based optimization for general algorithm configuration. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6683 LNCS (2011), 507–523.
- [42] James S. Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. 2011. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*. 2546–2554.
- [43] James Bergstra, Daniel L. K. Yamins, and D. Cox. 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the 30th International Conference on Machine Learning*. 115–123. <http://jmlr.org/proceedings/papers/v28/bergstra13.html>.
- [44] Donald R. Jones, Matthias Schonlau, and William J. Welch. 1998. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 13, 4 (1998), 455–492.
- [45] Peter Frazier, Warren Powell, and Savas Dayanik. 2009. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing* 21, 4 (2009), 599–613.
- [46] Jian Wu, Matthias Poloczek, Andrew G. Wilson, and Peter Frazier. 2017. Bayesian optimization with gradients. In *Advances in Neural Information Processing Systems*. 5267–5278.
- [47] Jasper Snoek, Hugo Larochelle, and Ryan Adams. 2012. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*. 2951–2959. DOI: [http://dx.doi.org/10.1016/s2468-2667\(17\)30214-1](http://dx.doi.org/10.1016/s2468-2667(17)30214-1)
- [48] Kevin Swersky, Jasper Snoek, and Ryan Prescott Adams. 2014. Freeze-thaw Bayesian optimization. *arXiv preprint arXiv:1406.3896* (2014).
- [49] Carolina Osorio and Linsen Chong. 2015. A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. *Transportation Science* 49, 3 (2015), 623–636. DOI: <http://dx.doi.org/10.1287/trsc.2014.0550>
- [50] Carolina Osorio and Kanchana Nanduri. 2015. Urban transportation emissions mitigation: Coupling high-resolution vehicular emissions and traffic models for traffic signal optimization. *Transportation Research Part B: Methodological* 81 (2015), 520–538. DOI: <http://dx.doi.org/10.1016/j.trb.2014.12.007>
- [51] Linsen Chong and Carolina Osorio. 2017. A simulation-based optimization algorithm for dynamic large-scale urban transportation problems. *Transportation Science* 52, 3 (2017), 637–656.
- [52] Ennio Cascetta. 1989. A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research Part B* 23, 1 (1989), 1–17. DOI: [http://dx.doi.org/10.1016/0191-2615\(89\)90019-2](http://dx.doi.org/10.1016/0191-2615(89)90019-2)
- [53] Gunnar Flötteröd. 2017. A search acceleration method for optimization problems with transport simulation constraints. *Transportation Research Part B: Methodological* 98 (2017), 1339–1351. DOI: <http://dx.doi.org/10.1016/j.trb.2016.12.009>
- [54] Gregory Bucci, Chris Calley, Michael Green, et al. 2018. *FHWA Research and Technology Evaluation: Agent-Based Modeling and Simulation*. Technical Report. United States. Federal Highway Administration. Office of Corporate Research.
- [55] Peter Vovsha, James E. Hicks, Binny M. Paul, Vladimir Livshits, Petya Maneva, and Kyunghwi Jeon. 2015. New features of population synthesis. *94th Annual Meeting of the Transportation Research Board* 250 (2015), 1–20.
- [56] Artem Chakirov and Pieter J. Fourie. 2014. Enriched sioux falls scenario with dynamic and disaggregate demand. *Arbeitsberichte Verkehrs-und Raumplanung* 978 (2014).
- [57] David E. Goldberg and John H. Holland. 1988. Genetic algorithms and machine learning. *Machine Learning* 3, 2 (1988), 95–99.
- [58] Roger D. Peng. 2011. Reproducible research in computational science. *Science* 334, 6060 (2011), 1226–1227.
- [59] Andrew Morin, Jennifer Urban, Paul D. Adams, Ian Foster, Andrej Sali, David Baker, and Piotr Sliz. 2012. Shining light into black boxes. *Science* 336, 6078 (2012), 159–160.
- [60] Geir Kjetil Sandve, Anton Nekrutenko, James Taylor, and Eivind Hovig. 2013. Ten simple rules for reproducible computational research. *PLOS Computational Biology* 9, 10 (10 2013), 1–4. DOI: <http://dx.doi.org/10.1371/journal.pcbi.1003285>
- [61] Justine S. Hastings, Mark Howison, Ted Lawless, John Ucles, and Preston White. 2019. Unlocking Data to Improve Public Policy. (03 2019). DOI: <http://dx.doi.org/10.31219/osf.io/28krq>

Received September 2019; revised January 2020; accepted February 2020