An Adversarial Variational Inference Approach for Travel Demand Calibration of Urban Traffic Simulators

Martin Mladenov[a], Sanjay Ganapathy Subramaniam[a], Chih-wei Hsu[a], Neha Arora[a], Andrew Tomkins[a], Craig Boutilier[a], Carolina Osorio[a,b,*]
[a]Google Research, Mountain View CA, USA
[b]HEC Montréal, Mountreal QC, Canada
* Corresponding author
{mmladenov, ganapathys, cwhsu, nehaarora, tomkins, cboutilier, osorioc}@google.com

ABSTRACT

This paper considers the calibration of travel demand inputs, defined as a set of *origin-destination matrices (ODs)*, for stochastic microscopic urban traffic simulators. The goal of calibration is to find a (set of) travel demand input(s) that replicate sparse field count data statistics. While traditional approaches use only first-order moment information from the field data, it is well known that the OD calibration problem is underdetermined in realistic networks. We study the value of using higher-order statistics from spatially sparse field data to mitigate underdetermination, proposing a variational inference technique that identifies an OD distribution. We apply our approach to a high-dimensional setting in Salt Lake City, Utah. Our approach is flexible—it can be readily extended to account for arbitrary types of field data (e.g., road, path or trip data).

CCS CONCEPTS

• **Applied computing → Transportation**; • **Theory of computation → Bayesian analysis**; • **Computing methodologies → Modeling and simulation**; **Neural networks**.

KEYWORDS

Variational inference, Simulation calibration, Urban travel demand

ACM Reference Format:

Martin Mladenov[a], Sanjay Ganapathy Subramaniam[a], Chih-wei Hsu[a], Neha Arora[a], Andrew Tomkins[a], Craig Boutilier[a], Carolina Osorio[a,b,*], [a]Google Research, Mountain View CA, USA, [b]HEC Montréal, Mountreal QC, Canada, * Corresponding author, {mmladenov, ganapathys, cwhsu, nehaarora, tomkins, cboutilier, osorioc}@google.com, . 2022. An Adversarial Variational Inference Approach for Travel Demand Calibration of Urban Traffic Simulators. In *The 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22), November 1–4, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3557915.3560940

1 INTRODUCTION

Differentiable neural network-based traffic simulators have attracted considerable attention in recent years due to their modeling power and ease of calibration. However, their yet unproven counterfactual robustness—the ability to accurately simulate a traffic scenario under interventions not present in the training data—limits their applicability to predicting the outcomes of counterfactual traffic policies and, hence, to traffic policy optimization. In this light, research on the calibration of mechanistic microscopic traffic simulators with their relatively well-established causal models remains an important avenue for research. In this work, we apply techniques introduced in the neural network community to the calibration of mechanistic traffic simulators, developing a general framework for using rich data sources to calibrate various types of input parameters.

We consider microscopic urban traffic simulators that provide a high-resolution representation of travel demand and of network supply. We apply our framework to the well-established problem of *origin-destination matrix (OD)* calibration, reflecting arguably, the most important and challenging simulation input, that of *travel demand.* A solution to the OD calibration problem is represented by an OD, a matrix reflecting traffic demand in which each entry denotes the expected travel demand that starts in a specific origin zone and ends in a specific destination zone. Its inputs include a set of field traffic data (e.g., segment speeds, segment counts), data statistics (e.g., first- or second-order moments), and an historical, possibly noisy OD (e.g., an outdated OD derived from travel survey or census data). The goal is to find an OD that minimizes some measure of divergence between field data statistics and their simulated counterparts, i.e., to identify an OD that allows the simulator to replicate observed traffic patterns.

When a simulation-based traffic model is used, the OD calibration is formulated as a high-dimensional continuous *simulation-based optimization (SO)* problem. For a metropolitan area the problem dimension (i.e., the number of non-zero OD entries) may be in the tens of thousands, inducing a difficult SO problem due to its dimensionality and its stochastic, simulation-based, non-differentiable objective function. Moreover, each estimation of the objective function requires one or more simulation calls, each with non-negligible compute times and costs.

In this paper, we consider the *static* OD calibration problem. We pay particular attention to the issues of observability and identifiability, and the inherent *underdetermined* (or underspecified) nature of this problem. This arises from the large number (typically, a continuum) of OD solutions that fit the field data equally well. In the classical setting studied here, field count data is spatially sparse, covering few network segments, and aggregated over coarse time

SIGSPATIAL '22, November 1–4, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9529-8/22/11.
https://doi.org/10.1145/3557915.3560940

periods (e.g., hourly vehicular counts). Moreover, not much information is extracted from field data, typically only first-order moment information. Hence, the problem is highly underdetermined.

Underdetermination has been a key concern since the early days of OD estimation, and various regularization approaches have been proposed [8]. Underdetermination is problematic in practice, since typically a single solution is found and used to carry out counterfactual analysis of a given transportation project (e.g., introduction of congestion pricing). Once an OD is chosen, little or no analysis is performed to evaluate the impact of changes to the OD on project performance. We focus on mitigating the impact of underdetermination by adopting a probabilistic approach, where the output is a probability distribution over possible OD solutions. This allows for robust counterfactual analysis that accounts for the existence of multiple solutions to the problem. .

While the degree of underdetermination can be reduced by increasing system observability (e.g.,by installing new types of sensors to collecting new forms of data, or sensors of the same type on additional segments), this is costly for city agencies that manage road networks. We instead study the added value of increasing the amount of information that is extracted from existing sensors.

We develop a novel, adversarial, variational-inference technique for this OD calibration problem. We exploit *higher-order information* from the field data to mitigate the impact of underdetermination. We formulate calibration as a distribution-matching problem whose aim is to reproduce both low and high-order moments of the field data. This contrasts with past work that matches only first-order moments. Our method is directly applicable to time-dependent settings with arbitrary types and amounts of field data. We test our framework on a Salt Lake City (Utah, USA) road network and show its ability to reduce the degree of underdetermination.

We propose a generative adversarial network (GAN) formulation of the calibration problem. To the best of our knowledge, GANs have so far been used to tackle traffic estimation and forecasting problems [7], their use for the calibration of simulators has not yet been explored. There is limited deep learning work for OD calibration [15, 16]. This paper explores the ability of GAN-based formulations to provide flexible calibration frameworks capable of accounting for various types of both field data and calibration parameters.

2 METHODOLOGY

2.1 General calibration framework

We integrate the RecSim NG simulation platform [10] with the SUMO (Simulation of Urban Mobility) road traffic simulator [9]. RecSim NG is an open source probabilistic programming platform, based on dynamic Bayesian networks, for specifying multi-agent behavioral models.[1] Its extensive modular library enables high-resolution general-purpose simulation for Bayesian learning and prediction, using automatic differentiation, program transformation, deep probabilistic programming, deep neural networks, and hardware accelerators via a TensorFlow [1] runtime. RecSim NG supports the learning of latent variables of generative models from data (e.g., an OD for a traffic simulator, recommender system simulations [10]). In this paper, we combine RecSim NG with GANs to formulate the OD calibration problem as a GAN problem. GANs have been used to generate synthetic traffic data [5] and predict traffic flow [18] without learning the distribution of real data, instead using a discriminator to distinguish real from generated data.

SUMO is an open-source stochastic microscopic road-traffic simulator. Our past work has used SUMO for calibration [4] and CO_2 emissions estimation [3]. Like most microscopic traffic simulators, SUMO uses non-differentiable demand and/or supply models. Though many of its disaggregate demand models (e.g., route choice, lane-changing, car-following) can be probabilistic, SUMO does not derive likelihood estimates. Hence, one cannot estimate how likely a given set of simulation inputs are to be consistent with field data. The integration of GANs and RecSim NG allows us to use differentiable optimization techniques to tackle high-dimensional OD calibration problems and provide a Bayesian analysis of the solutions. This is especially important in underdetermined problems: our approach yields a set of OD solutions that are consistent with the underlying field data, and defines a distribution (i.e., likelihood) over these solutions. Importantly, this integrated RecSim NG -SUMO framework can be extended readily to arbitrary: (i) simulation inputs (e.g., other demand and/or supply inputs), (ii) types of field data (e.g., static sensor data, probe data), and (iii) types of field data statistics (e.g., high-order moments).

2.2 Problem formulation

The dynamics of the simulated traffic network emerge from the behavior of a large number of individual agents (i.e., each actor or dynamic entity, such as vehicles and traffic lights). We assume that the behavior of this network can be described as a Markov chain with joint distribution P over a state space $\mathcal{S}$, defining the state and dynamics of the simulated traffic network A trajectory $\tau \in T$ of this stochastic process reflects a possible evolution of the network state over time. We assume a function $\phi : T \rightarrow \mathbb{R}^n$ defining the set of *observable features* for a trajectory τ.

A traffic simulator Q_θ with parameters θ is a model of P over the same state space $\mathcal{S}$. The calibration problem can be thus formulated as follows. Given a real-world distribution P, a parametric simulator Q_θ, and an observation function ϕ, find simulator parameters θ^* that maximize the similarity of the real-world observation distribution P^ϕ to the the simulated observation distribution Q_θ^ϕ. This instantiates the following optimization problem:

$$\theta^* = \text{argmin}_\theta \, D(Q_\theta^\phi, P^\phi) \,, \tag{1}$$

with:

- θ: parameters to calibrate (OD demands) [veh/hour]
- D: Divergence measure between distributions
- Q_θ^ϕ: θ-parametrized distribution of simulated segment counts
- P^ϕ: field data distribution of segment counts

We focus on calibration of OD parameters, hence θ is a vector representation of an OD. However, our formulation applies to the calibration of *arbitrary* simulation inputs.

Adversarial calibration. The nature of the optimization problem in Eq. (1) poses several challenges for microscopic traffic simulation. First, most traffic simulators (and SUMO in particular) are implemented as black-box Monte Carlo samplers from Q_θ, exposing only sample-level access to the underlying distribution. This precludes

[1] See `https://github.com/google-research/recsim_ng`.

the computation of various important quantities used in inference algorithms, such as log probabilities and their derivatives. Moreover, not all random choices are logged in the simulator output, resulting in an unknown number of latent variables (e.g., SUMO simulates a driver's lapse in judgement due to exhaustion as a random process). Second, the target distribution Q_θ^ϕ is non-differentiable w.r.t. its parameters (i.e., differentiation may be theoretically impossible or simply computationally intractable). While approaches to overcoming these issues generally involve rewriting the simulator code, we instead develop a method that operates within these restrictions at the expense of some statistical efficiency. Hence, our approach can be used with black-box commercial simulators, where one does not have access to the source code.

We define the *adversarial simulation calibration problem* as:

$$\theta^* = \operatorname{argmin}_\theta \sup_{g \in \Gamma} E_{X\sim P^\phi}\left[g(X)\right] - E_{X\sim Q_\theta^\phi}\left[f^*(g(X))\right], \quad (2)$$

where f^* is a convex function, Q_θ^ϕ is a *generator*, g is a *discriminator*, and Γ is a subset of function space $\{g : X \to \mathbb{R}\}$. Intuitively, the discriminator's goal is to produce high values (in expectation) on real data and low values on the synthetic data; that is, it tries to discriminate the real and synthetic distributions (which is impossible if the two are identical). Setting $\Gamma = \{g : X \to \mathbb{R}\}$, this objective minimizes a dual form of the f-divergence $D(P,Q) = E_Q[f(P/Q)]$ [11], with f^* being the Fenchel conjugate of f. If Γ is a Lipschitz function and f^* is the identity, it minimizes a dual form of the Wasserstein distance [2]. Since optimization over $\{g : X \to \mathbb{R}\}$ is intractable, we restrict this maximization over g to some parametric family. Furthermore, a smoothing regularizer might be added to prevent mode collapse on the discriminator side [14], resulting in the objective:

$$\theta^* = \operatorname{argmin}_\theta \sup_{\nu \in N} E_{X\sim P^\phi}\left[g_\nu(X)\right] - E_{X\sim Q_\theta^\phi}\left[f^*(g_\nu(X))\right] + R(\nu) \quad (3)$$

where $g_\nu, \nu \in N$ is a function with parameter vector ν. Eq. (3) lies between the f-GAN and the w-GAN formulations depending on the choices of discriminator and regularizer $R(\cdot)$. The design of convergent solvers for Eq. (3) is not straightforward. Several classes of convergent algorithms are known, we use the two-time-scale descent algorithm [6] due to its simplicity.

3 CASE STUDY

We use SUMO model of a network within Salt Lake City, Utah, USA. The road network and segment supply data (e.g. network topology, segment geometry) is derived from Google Maps.[2] To assess underdetermination and the distance of the true OD from the ODs produced by the algorithms, we use a synthetic OD as the ground truth (GT). We run 90 independent simulations using the GT OD to produce segment counts, which we treat as field data.

We use two OD calibration baselines: Simultaneous Perturbation Stochastic Approximation (SPSA) [17] and the metamodel approach of [4] (itself a simplification of [12, 13]). These baselines solve the classical OD calibration formulation, as defined in Eq. (1) of [13]. This formulation aims to match first-order moments of sensor counts, and does not use higher-order moment information. We use these methods as baselines to assess the added value of using higher-order, as we do in our adversarial optimization technique, in reducing the level of underdetermination.

[2]https://developers.google.com/maps/documentation/roads/overview

Method	Count nRMSE (%)	OD nRMSE (%)		
		Mean	Median	Minimum
Initial Points	33.7 ± 6.1	119.0	118.8	101.9
Adversarial	7.7 ± 1.5	72.4	72.1	62.1
Metamodel (regularization)	5.7 ± 0.5	111	46.7	46.7
Metamodel (no regularization)	7.3 ± 1.5	205.8	109.1	86.2
SPSA (regularization)	25.7 ± 6.1	117.6	118.1	96.1
SPSA (no regularization)	25.4 ± 6.3	118.6	119.5	100

Table 1: nRMSE summary of the best OD solutions for each method.

We consider a 62-dimensional instance. Each of the 62 ODs has three possible routes. Twelve segments are assumed to have field data. This problem instance embodies significant underdetermination: segments have various routes contributing to their traffic, hence the count data in the twelve segments with field data contains limited information about the underlying OD.

We allow each baseline to run multiple parallel simulations and optimization until convergence, with a maximum of 100 iterations. We run each of the baseline methods 100 times with a uniformly random initial point, and study the distribution of the performance of these 100 OD solutions. Additionally, we run the baseline methods with and without regularization based on a prior OD.

To quantify the fit of an OD, we use the nRMSE as defined in [4]. We allow a maximum of 100 iterations of simulation and optimization to all methods. The metamodel approaches converge within 15, the adversarial method converges in 50 and SPSA does not converge even after 100 iterations. Since our adversarial techniques yield an OD probability distribution, rather than a point OD, it is only run once, rather than 100 times. We use the prior OD as the initial point, but do not implement any form of regularization. For each initial point and each method, the point that has the least loss across all iterations is chosen as the best point.

Table 1 quantifies the nRMSE for each method. For each baseline method, nRMSE is the average across all 100 best solutions. For the adversarial method, we generate a sample of 250 ODs from the best OD solution (which is an OD distribution) and compute average nRMSE across the sample. Column 1 shows percentage nRMSE for the counts, i.e., it measures the distance between the ground truth counts and the simulated counts of the best OD solutions. Similarly, column 2 shows percentage nRMSE for the ODs themselves.

The SPSA methods, both with and without regularization, have similar performance. Their solutions offer only a slight improvement over the initial points. The metamodel method without regularization has substantially better fit to counts than SPSA, however its fit to ODs is much worse than that of the initial point. This is indicative of a highly underdetermined problem, where many different ODs with a good fit to counts can be found. The baseline with the best performance is the metamodel with regularization. Both its fit to counts and to ODs is substantially better than SPSA and to the unregularized metamodel approach. For 82% of the 100 runs, the average nRMSE for its ODs is 46.7%; however for the remaining 18% the average OD nRMSE is 400%. In other words, frequently the solutions proposed by the regularized metamodel method exhibit strong performance both in terms of fit to counts and fit to ODs. However, there is a non-negligible subset of starting points

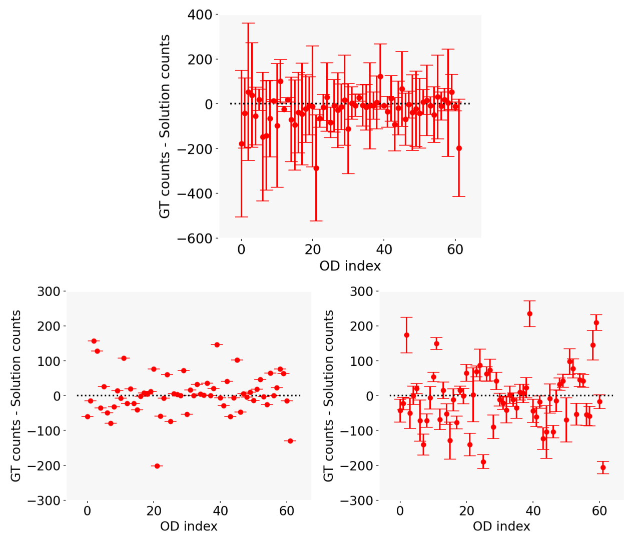

Figure 1: Comparison of best ODs identified by the metamodel with regularization (top and lower left plots) and the adversarial method (lower right plot).

for which the identified OD is *very far* from the ground-truth OD. Again, this is a consequence of the problem's underdetermination. Compared to the baselines, our adversarial method achieves a good fit to counts—comparable with the metamodel—while providing a significantly better mean fit to ODs. This demonstrates that the model uses second order information effectively to eliminate OD solutions that are far from the ground truth.

The top plot in Figure 1 considers the metamodel method with regularization. The x-axis is an OD index and the y-axis is the mean difference between the ground-truth OD and the solution OD. Error bars have a half-width of one standard deviation. This plot shows that average difference is not high, but there is substantial variability, highlighting the underdetermination of the problem. The lower left plot of Figure 1 considers the same method but only considers the 82 OD solutions mentioned above, which have a low nRMSE w.r.t. the OD. It shows that if we exclude the very poor solutions, the metamodel method performs very well. However, this performance is still reliant on the quality of the prior OD supplied to the optimization. The lower right plot of Figure 1 considers the sample of 250 ODs sampled from the adversarial method. It has the same y-axis range as the lower left figure, but unlike the lower left plot, it does not exclude any outlier ODs. The right plot shows that the average difference to GT ODs is small for the solutions derived by the adversarial approach. Moreover, the error bars are small, and there are no outlier ODs. This shows the added value of our adversarial approach in reducing underdetermination.

4 CONCLUSION

In this paper we investigated an adversarial variational inference approach to tackle demand calibration problems for stochastic traffic simulators. The results of our investigations indicate that our formulation contributes substantially to the mitigation of underdetermination by matching higher-order moments of the field data. Our approach extracts more information from existing field data, eliminating or reducing the need to deploy new sensors for accurate calibration. Importantly, it easily generalizes to the use of arbitrary field data statistics and types, such as individual trip times, spatio-temporal correlations and to the calibration of arbitrary demand parameters such as driver behavioral characteristics and preferences over routes. The proposed approach yields an OD probability distribution, as opposed to a single point estimate. We consider this an important step toward the routine adoption of uncertainty quantification in calibration and transportation optimization problems.

REFERENCES

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. 2015. TensorFlow: Large-scale Machine Learning on Heterogeneous Systems.

[2] M. Arjovsky, S. Chintala, and Léon Bottou. 2017. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia*, Vol. 70. 214–223.

[3] N. Arora, T. Cabannes, S. Ganapathy, Y. Li, P. McAfee, M. Nunkesser, C. Osorio, A. Tomkins, and I. Tsogsuren. 2021. Quantifying the sustainability impact of Google Maps: A case study of Salt Lake City. https://arxiv.org/abs/2111.03426.

[4] N. Arora, Y. Chen, S. Ganapathy, Y. Li, Z. Lin, C. Osorio, A. Tomkins, and I. Tsogsuren. 2021. An Efficient Simulation-Based Travel Demand Calibration Algorithm for Large-Scale Metropolitan Traffic Models. https://arxiv.org/abs/2109.11392.

[5] Y. Chen, Y. Lv, and F. Wang. 2020. Traffic Flow Imputation Using Parallel Data and Generative Adversarial Networks. *IEEE Transactions on Intelligent Transportation Systems* 21, 4 (2020), 1624–1630.

[6] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA*. 6626–6637.

[7] J. Jin, D. Rong, T. Zhang, Q. Ji, H. Guo, Y. Lv, X. Ma, and F. Wang. 2022. A GAN-Based Short-Term Link Traffic Prediction Approach for Urban Road Networks Under a Parallel Learning Framework. *IEEE Transactions on Intelligent Transportation Systems* (2022), 1–12.

[8] P. Krishnakumari, H. van Lint, T. Djukic, and O. Cats. 2020. A data driven method for OD matrix estimation. *Transportation Research Part C* 113 (2020), 38–56.

[9] P. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner. 2018. Microscopic traffic simulation using SUMO. In *21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2575–2582.

[10] M. Mladenov, C. Hsu, V. Jain, E. Ie, C. Colby, N. Mayoraz, H. Pham, D. Tran, I. Vendrov, and C. Boutilier. 2021. *RecSim NG: Toward Principled Uncertainty Modeling for Recommender Ecosystems.* Technical Report. arXiv:2103.08057 [cs.LG]

[11] S. Nowozin, B. Cseke, and R. Tomioka. 2016. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain*. 271–279.

[12] C. Osorio. 2019. Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Research Part C* 98 (2019), 186–206.

[13] C. Osorio. 2019. High-dimensional offline origin-destination (OD) demand calibration for stochastic traffic simulators of large-scale road networks. *Transportation Research Part B* 124 (2019), 18–43.

[14] K. Roth, A. Lucchi, S. Nowozin, and T. Hofmann. 2017. Stabilizing Training of Generative Adversarial Networks through Regularization. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA*. 2018–2028.

[15] L. Schultz and V. Sokolov. 2018. Practical Bayesian Optimization for Transportation Simulators. https://doi.org/10.48550/ARXIV.1810.03688

[16] D. Sha, K. Ozbay, and Y. Ding. 2020. Applying Bayesian Optimization for Calibration of Transportation Simulation Models. *Transportation Research Record* 2674, 10 (2020), 215–228.

[17] J. Spall. 2003. *Introduction to stochastic search and optimization: estimation, simulation, and control.* John Wiley & Sons, New Jersey, USA.

[18] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang. 2021. TrafficGAN: Network-Scale Deep Traffic Prediction With Generative Adversarial Nets. *IEEE Trans. Intell. Transp. Syst.* 22, 1 (2021), 219–230.