Creating and Calibrating POLARIS Models

Krishna Murthy Gurumurthy, Joshua Auld, Pedro Camargo, Jamie Cook, Taner Cokyasar, Felipe De Souza, Yantao Huang, Abdelrahman Ismael, Navjyoth Jayashankar, Nazmul Arefin Khan, Charbel Mansour, Gopindra S Nair, Aymeric Rousseau, Olcay Sahin, Hui Shen, Hyunseop Uhm, Omer Verbas, Griffin White, Jan Zill, Natalia Zuniga

Transportation and Power Systems Division, Argonne National Laboratory, Lemont IL USA

Submitted for presentation at the MATSim User Meeting 2025 to be held in Munich, Germany from June 12-13, 2025

EXTENDED ABSTRACT

Several agent-based models have been developed for transportation demand forecasting across the world. POLARIS (Auld et al., 2016) is one such tool that is an agent-based and activity-based model capable of simulating large metro areas and is the key focus of this abstract. Built wholly on a C++ codebase, simulation runtimes are moderate for large areas and multi-threading enables use of high-performance computing (HPC) architectures. For example, a 24-hour day simulation of a 25% Chicago metro model takes about 157 min. Other examples include MATSim (Horni et al., 2016), SimMobility (Adnan et al., 2016), MobiTopp (Mallig et al., 2013), each with strengths and weaknesses, and the choice of tool ultimately depending on the modeler's end goal.

In this presentation, the capabilities of POLARIS will be highlighted starting from model creation and calibration, and the availability of the federated toolset – namely Polaris-Studio and QPolaris that support POLARIS use in each of the steps. An overview of the topics is presented below.

Model Creation

POLARIS data models comprise both supply and demand components. Information that represent the regional supply such as links and nodes, transit service availability, and on-demand and shared mobility services, are created as a combination of SpatiaLite databases and parameter JSON files. Population synthesis input includes information for spatial distribution of households and persons, as well as information on the marginals that impact travel demand (number of households by household size, number of vehicles owned, etc.). Other demand models, such as econometric parameter data for mode, destination, timing, transit pass, activity generation, and other choices are also created in JSON formats.

Polaris-Studio, available on PyPI, supports a variety of steps involved in putting together a model. The semi-automated model builder requires a geographic polygon for the limits of the regional model required, and for the US context pulls together data from the Census, ACS, OpenStreetMaps, to populate the SQLite databases. Population synthesis inputs are also generated based on the list of census tract identifiers for the county or state, and the data year needed for the model. At the end of this process, a POLARIS model with most of the required data populated is created. For all choice models, default JSON files can be created with POLARIS as a starting point before calibration.

Model creation outside of the US is a semi-manual step. Integration with Eqasim (Hörl and Balac, 2021) is in the works and has been tested in France and is being expanded to other regions. A process to create models in Australia has been setup using the native model builder. These automated and semi-manual model builders are not perfect processes due to the variety in data sources across the world, and the constantly evolving nature of requirements that improve model robustness, so modeler checks are imperative to then obtain a final model for use with POLARIS.

Model Calibration

Once a POLARIS model is created (with supply and demand inputs) and the data passes Polaris-Studio's critical error checker, model calibration can be setup. Polaris-Studio provides functionality to setup a POLARIS convergence run. This is a series of iterations, with each iteration simulating a 24-hour day, and network states feeding into the next iteration to update demand outcomes. The calibration process is a variant of the convergence run, where demand parameters are updated to fit an expected target distribution of modes, distances, time of departure, and number of activities generated. In addition, key performance indicators (KPIs) for validation are also created when data is assembled by the modeler.

Calibrating an agent-based model is not straightforward especially since network state propagation into demand outcomes occurs over a series of iterations. Modifying parameters by too much or too little can have an impact on calibration. Calibration in POLARIS occurs at a variety of iterations now. At the beginning of the simulation, synthesized households have their home location decided based on the input data. However, workplace choice is a function of underlying network states and modal availability. Choice in workplaces impacts overall network state, and calibrating mode, timing, and activity generation outcomes need to be tiered with changing workplaces with sufficient gap in iterations to allow for network state to adapt to this change in demand.

DFW and Chicago Calibration and KPIs

An example of the calibration process for Dallas-Fort Worth (DFW) and Chicago are outlined here. DFW is home to about 7.5M people and the Chicago metro is home to 11M people over about the same land area of 10,000 sq. mi. Both these models were created based on data from the local metropolitan planning agencies, however, similar model inputs can be created from the automated model creation tool discussed earlier.

Table 1: Model Characteristics (Year 2019)

	DFW	**Chicago**
Population	7.51M	10.69M
# Households	2.75M	3.93M
# Vehicles	5.36M	6.44M
# Links : # Nodes	42,527 : 28,421	48,379 : 35,883

# Transit Agencies : Routes	3 : 160	4 : 345

Calibrated Outcomes

Mode shares were calibrated across trip purposes of home-based work, home-based other, and non-home based trips as most survey data is available in this format. Disaggregate trip data from POLARIS can be aggregated up to these distinct trip purposes to apply updates to the choice constants. Destination choice model was calibrated by matching average trip distance by trip purpose by synthesized agents compared to data. Travel times reported in surveys were confirmed with Google Maps data to be inflated, plot on model travel time compared to survey targets is consistently lower even though distances are on target.

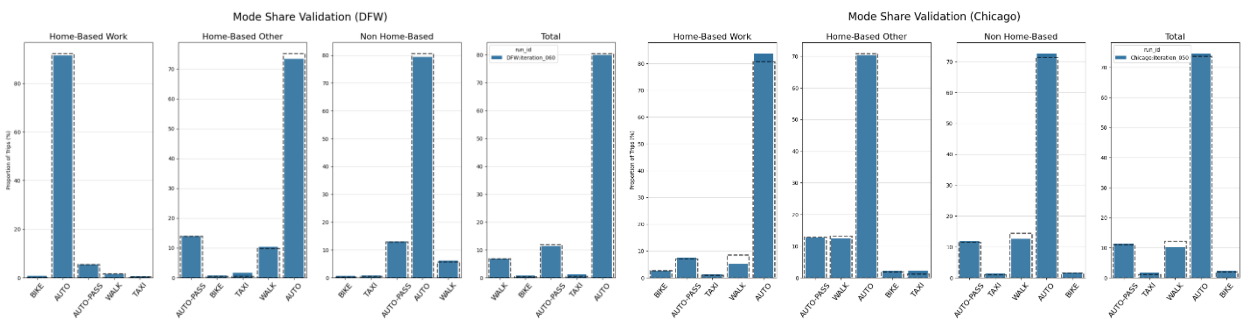


Figure 1 Mode Share Outcomes After Calibration

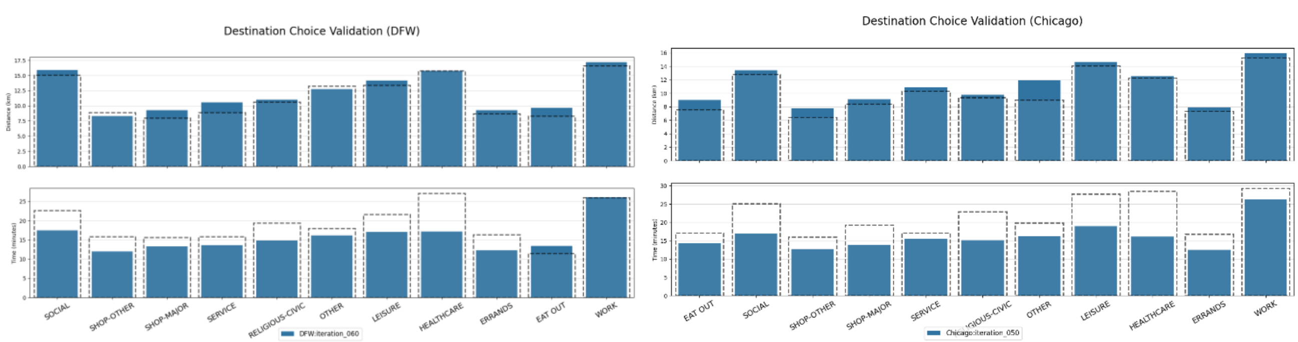
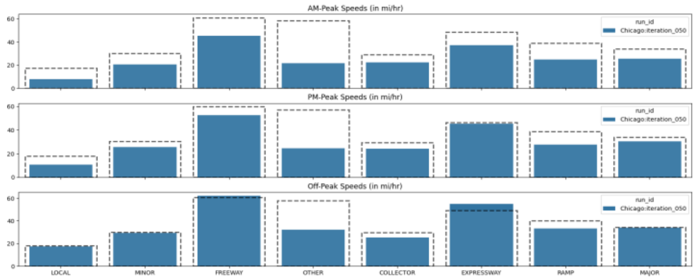


Figure 2 Trip Distance by Purpose After Destination Choice Calibration

Validations

Plots in the previous sub-section show the outcome of calibrating parameters and the closeness of fit on what is controllable. Link speeds and counts that occur from the assignment process and traffic flow conditional on the demand can only be minimally calibrated (using link capacity, for example). The plots here reflect the quality of POLARIS simulations



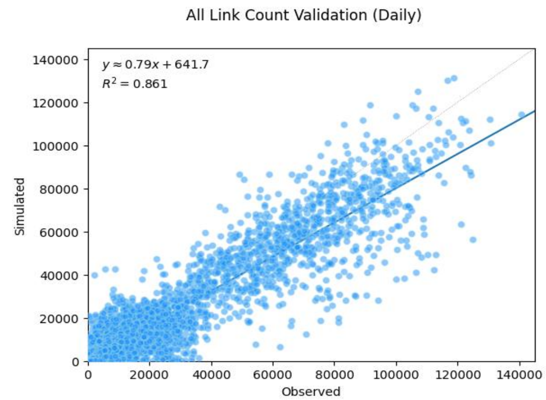


Figure 3 & 4 Speeds in Chicago Validated with INRIX data (***Left***) and Link Counts Validated for DFW (***Right***)

ACKNOWLEDGEMENTS

This report and the work described were sponsored by the U.S. Department of Energy (DOE) Vehicle Technologies Office (VTO) under the POLARIS Core Transportation Mobility, and supported by Energy Efficient Mobility Systems (EEMS) Program. The following DOE manager played important roles in establishing the project concept, advancing implementation, and providing ongoing guidance: Erin Boyd. The submitted manuscript has been created by the UChicago Argonne, LLC, Operator of Argonne National Laboratory (Argonne). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

REFERENCES

Adnan, M., Pereira, F.C., Azevedo, C.L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C., Ben-Akiva, M., 2016. SimMobility: A Multi-scale Integrated Agent-Based Simulation Platform. Presented at the Transportation Research Board 95th Annual MeetingTransportation Research Board.

Auld, J., Hope, M., Ley, H., Sokolov, V., Xu, B., Zhang, K., 2016. POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. Transp. Res. Part C Emerg. Technol. 64, 101–116. https://doi.org/10.1016/j.trc.2015.07.017

Hörl, S., Balac, M., 2021. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. Transp. Res. Part C Emerg. Technol. 130, 103291. https://doi.org/10.1016/j.trc.2021.103291

Horni, A., Nagel, K., Axhausen, K.W., 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press, London.

Mallig, N., Kagerbauer, M., Vortisch, P., 2013. mobiTopp – A Modular Agent-based Travel Demand Modelling Framework. Procedia Comput. Sci., The 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013), the 3rd International Conference on Sustainable Energy Information Technology (SEIT-2013) 19, 854–859. https://doi.org/10.1016/j.procs.2013.06.114