Institute of Transport Economics
Norwegian Centre for Transport Research

ENGLISH Summary

Towards activity-based demand modelling for the Greater Oslo Area

Using machine learning to predict travel mode choice and activity plans

TØI Report 2065/2024 • Authors: Stefan Flügel, Christian Weber, Simen Sørbøe Klommestein, Johan Korsmo, Anders Kielland • Oslo 2024 • 62 pages

This report documents the R&D efforts conducted as part of the PRELONG project between 2022 and 2024, focusing on generating activity plans for a synthetic population in the Greater Oslo Area. At the core of our approach are two neural network models that predict the characteristics of all trips made on a typical weekday. These machine learning models are trained on travel survey data from RUTER-MIS (2017–2024) and are applied to population and commuting data.

Transport modelling is evolving to address the challenges of urbanization, technological advancements, and complex travel behavior. Traditional four-step travel demand models, while foundational, often fail to capture the dynamic nature of travel patterns in urban areas.

Activity-based travel demand (ABDM) models offer a more detailed and dynamic representation of travel behavior, which accounts for the complexities of daily schedules, individual preferences and environmental constraints. This is achieved by: 1) The aggregated data structure is replaced with a disaggregated one, i.e. individuals (rather than population segments within geographic areas) are the unit of analysis. 2) Transport demand is derived from the demand to perform activities, with an emphasis on sequences of tours or all-day travel plans rather than individual (round) trips. 3) Time (clock time) is explicitly modelled (start/end times and activity planning are important).

To bridge the gap the towards activity-based and agent-based transport models, this report documents the generation of activity plans (first documented in Flügel 2022) using a machine learning approach, specifically deep neural networks.

Figure S1 gives a schematic overview of the main data, methods and concepts and how they are related in our overall approach to transport modelling. This report describes the four squares on the left side of the figure, which a specific focus on the demand model that in our case are two neural network models (model 1 and model 2).

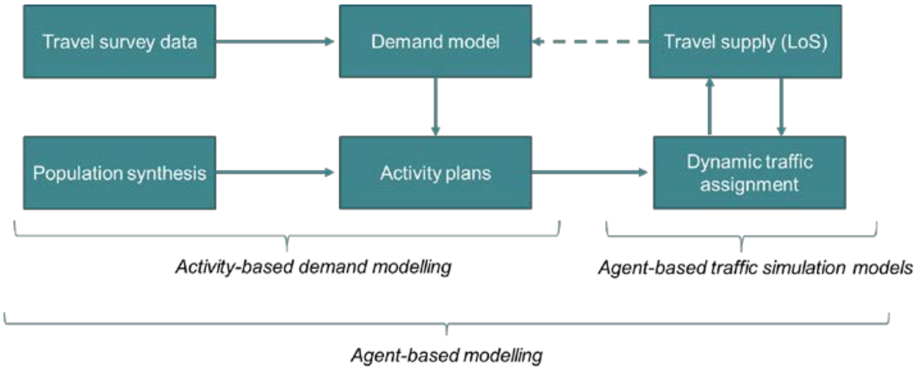

Figure S1: High level overview over related data and methods in this report.

As Figure S2 illustrates, model 1 does predict overall characteristics of the trip diary (activity plans): Start time of first trip, Start time of last trip, Number of trips and the Sequence of trips before and after the work activity. Model 2 predicts characteristics of trips: Travel mode, Purpose, End zone and Time duration since previous trip.

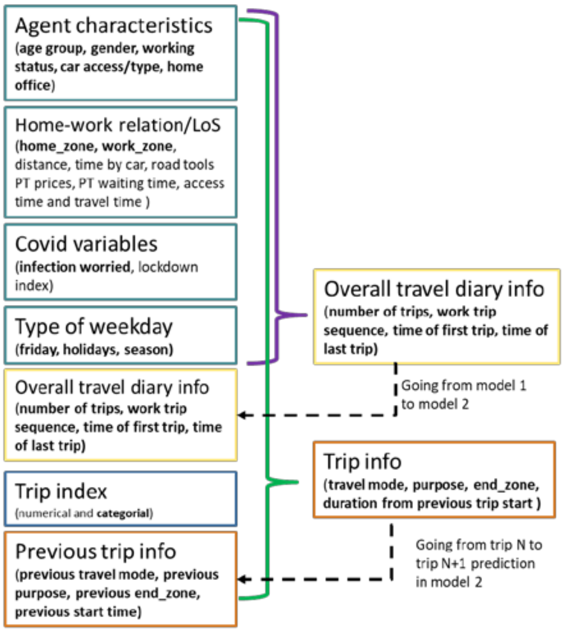

*Figure S2: Input and output variables of the two neural network models. The purple brackets indicate model 1 and the green bracket model 2. Categorical variables in **bold** font.*

The dotted arrows indicated that outputs are transformed into inputs when applying it to new data. Our approach captures correlations between multiple trip characteristics (trip purpose, travel mode, departure time, destination choice). For example, the output “purpose: picking up” will be strongly correlated with the output of “travel mode: car”.

The neural network models are trained on travel survey data, more specifically a random sample of 75% of the respondents in Ruter-MIS from January 2017 to September 2024. The models are validated against the remaining 25% of the sample (validation data). Overall the models seem to capture the correlation pattern on the validation data as exemplified by the next two figures.

Institute of Transport Economics, Gaustadalléen 21, N-0349 Oslo, Norway, E-mail: toi@toi.no www.toi.no

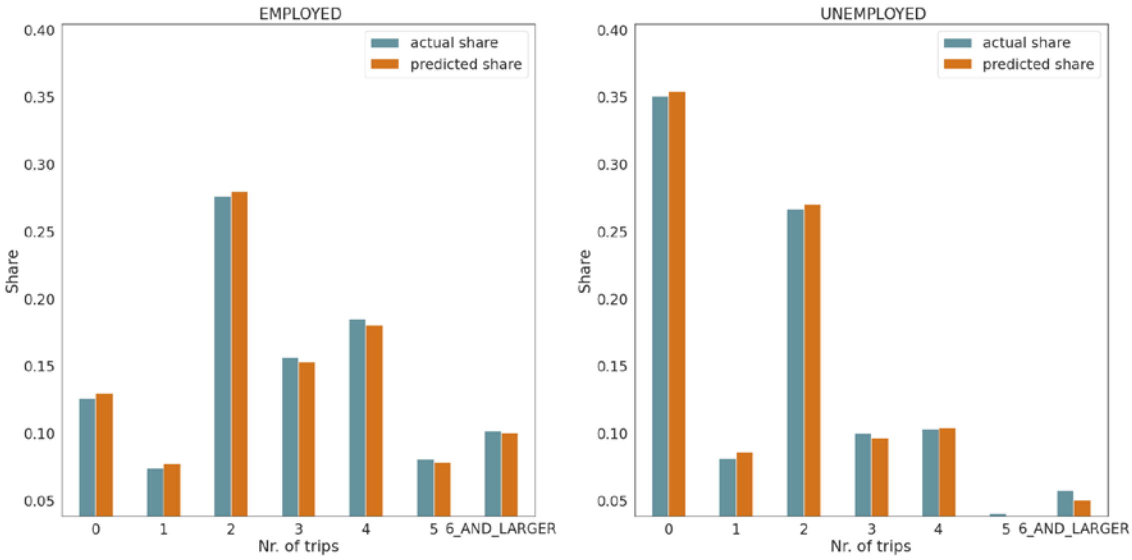

Figure S3: Actual share (in Ruter-MIS) and predicted share (from model 1) on the validation data for the Number of trips by occupation status.

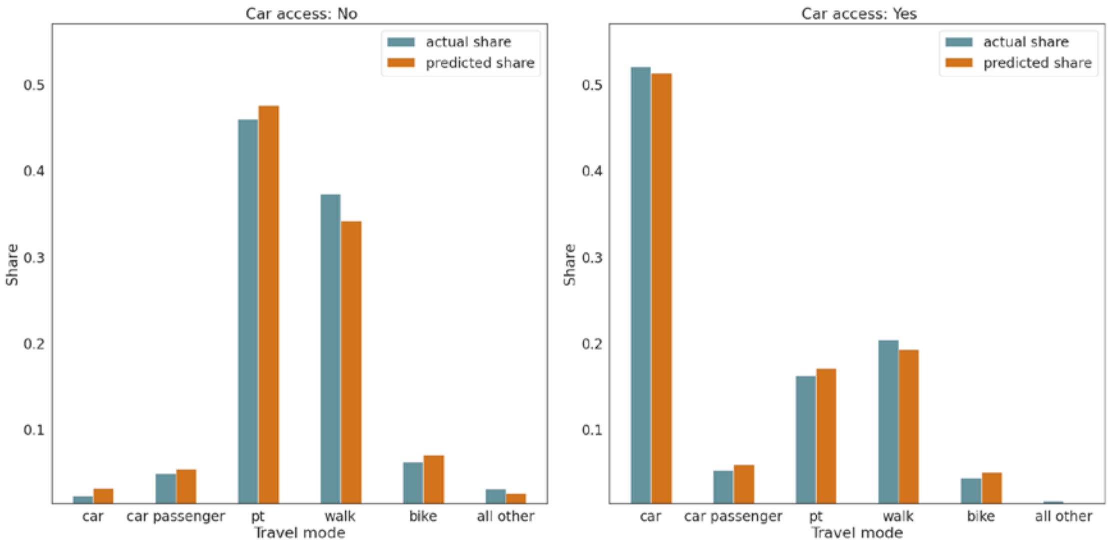

Figure S4: Actual share (in Ruter-MIS) and predicted share (from model 2) on the validation data for the travel mode choice; segmented by car access.

The models are than applied to a synthetic population that is assembled from various data sources (population data, commuting data, zonal data etc.).

Figure S5 shows the geographical location of different types of activities at 12 o'clock.

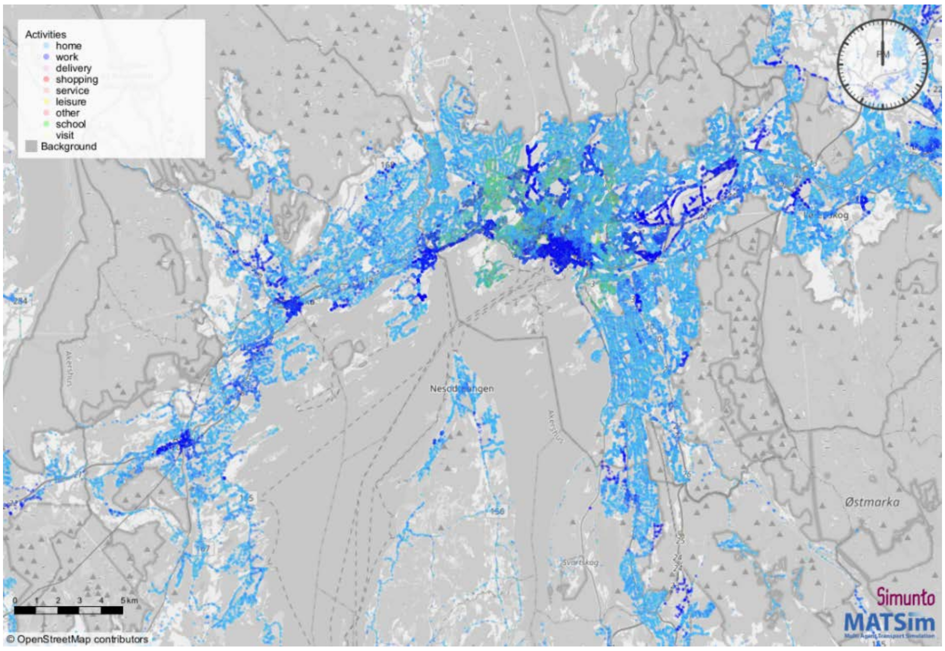

Figure S5: Geographical location of generated activities at timestamp 12:00:00.

The report includes some additional analysis regarding the effects of Level-of-Service on the prediction of travel model choice. We find that elasticities of demand, as implied by the neural network predictions, depend strongly on the model formulation and are in general on the lower side.

Our overall approach has several known limitations requiring further refinement. First, the coarse zonal system used for destination prediction results in imprecise non-work trip destinations, potentially leading to illogical or excessively long trips. Future versions should incorporate predictive travel distance models or geography-conditioned sampling to address these issues. Additionally, population synthesis could benefit from richer data inputs, such as microdata.no, and leveraging machine learning or generative AI to enhance agent characteristics like household type, education, and income.

Despite improvements since earlier versions, the model generates too many trips in the afternoon, likely due to the trip-by-trip prediction method. Exploring advanced architectures like recurring neural networks or attention-based models may help resolve this issue. Validation of synthetic activity plans remains challenging due to limited external data, with mobile data or MATSim-validated traffic simulations offering potential solutions. Lastly, the current backend implementation, relying on Python scripts and notebooks, is user-unfriendly, and a front-end interface is planned for 2025 to improve functionality and facilitate scenario testing.

Despite these challenges, the approach demonstrates significant promise. Its compatibility with dynamic transport models and applicability in agent-based modeling enables downstream applications, such as the MATSim transport simulation model. Moreover, the spatial-temporal data structure holds potential beyond transport, for example, in epidemiological simulations.

The method's adaptability for future synthetic populations (e.g., for 2030 or 2050) ensures relevance for long-term predictions, provided input data remains representative. Importantly, the approach can bridge future agent-based land-use models predicting long-term behavior and transport simulation models capturing detailed network-dependent behavioral adjustments, marking an important step toward advancing state-of-the-art transport models in Norwegian cities.