Grunnlagsdata og reisevaneundersøkelser for ABM

Project name	**Grunnlagsdata og reisevaneundersøkelser for ABM**
Project no.	**1350058023**
Recipient	**Statens Vegvesen**
Document type	**Report**
Version	**01**
Date	**31.01.2024**
Prepared by	**Andre Uteng, Gina Gjelten, Martijn Hollestelle, Martin L. Sandbraaten and Olga Petrik**
Checked by	**Øyvind Lervik Nilsen**
Approved by	**Sindre Levinsen**
Description	**This report evaluates the existing data sources with the aim of developing an activity-based model (ABM).**

Contents

Sammendrag

Som en del av prosjektet "Grunnlagsdata og reisevaneundersøkelser for ABM", har Rambøll gjennomført en gjennomgang av transportmodelldataene som for øyeblikket er tilgjengelige i Norge. Formålet med denne gjennomgangen har vært å vurdere om de tilgjengelige dataene er egnet for å opprette et norsk ABM-rammeverk, eller om det er nødvendig med ytterligere data. Rambøll har blitt bedt om å gi råd om følgende alternativer:

1. Bruke dagens grunnlagsdata og reisevanedata slik de foreligger i dag.
2. Bruke dagens grunnlagsdata og reisevanedata, men med noen anbefalte mindre tilpasninger og tillegg.
3. Lage et optimalt datasett for grunnlagsdata og reisevaner («best practice»).

I denne rapporten har Rambøll kartlagt hvilke data som er nødvendige for å utvikle en ABM, og kategorisert de dataene som er tilgjengelige i dag i tre kategorier: (1) fullt tilgjengelige data, (2) delvis tilgjengelige data og (3) utilgjengelige data. Funnene fra gjennomgangen er sammenstilt i Tabell 1.

Tabell 1: Oversikt over tilgjengelige data.

Type data	Fullt tilgjengelig	Delvis tilgjengelig	Utilgjengelig
RVU data	• Sosioøkonomiske variabler med høy nøyaktighet. • Detaljerte data om husholdnings-sammensetning. • Data om bruk av transportmidler. • Informasjon om aktiviteter mellom reiser.	• Tilstrekkelig utvalgsstørrelse. • Informasjon om transportmiddel-attributter. • Kostnader relatert til alternative reisemåter (f.eks. Månedskort, ansatt-parkering, firmabil etc.).	• Reisedagbøker for alle husholdnings-medlemmer. • Reisedagbøker for barn under 13. • Bostedstype (enebolig vs. leilighet).
Geografiske sosioøkonomiske data	• Høyoppløselige geografiske data (sonestørrelse på omtrent 250x250 meter). • Demografiske data	• Sysselsettingsdata. • Førerkortdata. • Bilholdsdata.	
Reisetilbud	• Detaljert nettverk for kollektivtrafikk, gang- og sykkelveier. • Transportmodellnettverk for å muliggjøre bytte av transportmidler.		

Basert på tilgjengeligheten av de forskjellige datatypene, er konklusjonen vår at en ABM kan utvikles ved å bruke de eksisterende grunnlagsdataene og den norske reisevaneundersøkelsen (RVU) som det foreligger i dag. Rambøll mener også at det er mulig å lage en ABM som en utvidelse av den eksisterende RTM infrastrukturen i CUBE ved å bruke følgende justeringer:

- Implementere en algoritme for å fordele de eksisterende grunnkretsdataene (for nåværende og fremtidig situasjon) over tilsvarende celler på rutenett.
- Utvikle en applikasjon som kobler RTM-nettverket til rutenettets noder.
- (valgfritt) Utføre separate husholdningsundersøkelser (i motsetning til RVU, som er en individuell reiseundersøkelse) som er skreddersydd for å skaffe reisedagbøker for hele husholdningen.

1. Summary

As part of the project «Grunnlagsdata og reisevaneundersøkelser for ABM», Rambøll has performed a review of the transport model data currently available in Norway. The purpose of this review has been to assess if the available data are suitable for the establishment of a Norwegian ABM-framework, or if additional data is needed. Rambøll has been asked to advice among the following options:

1. Using the input- and survey data in their current form
2. Use the input and survey data in their current form, but with a recommended set of minor adjustments and addons.
3. Creating an optimal dataset of input-data and survey-data which follows the best practice.

In performing the task, Rambøll has mapped the data which are required for development an ABM, and categorized the currently available data using three categories: (1) fully available data, (2) partly available data and (3) unavailable data. The findings from the review are listed in Table 1.

Table 1: Overview of available data and required data-properties.

Data Type	Fully Available	Partly Available	Unavailable
Survey Data	• High endpoint accuracy and detailed socio-economic variables • Detailed household composition data • Data on mode usage • Information on activities between trips	• Sufficient sample size • Information on mode attributes • Data related to costs of travel alternatives (e.g., PT monthly card ownership, employer paying for car / parking)	• Travel diaries for all household members • Trip-diaries for ages <13 • Data on the residence type (detached houses vs. apartments)
Socio-economic spatial data	• High resolution spatial data (zone size of about 6 hectares in built-up areas) • Geographic specific demographic data	• Geographic specific employment data. • Zonal driver's license data • Car ownership	
Supply	• Detailed network for public transport, walking and cycling facilities. • Super and auxiliary network for easy mode transfers		

Based on the availability of the different datatypes, it is our conclusion that an ABM can be developed using the input and the Norwegian travel survey (RVU) in their current form. It is in Rambøll's opinion possible to create an operational ABM as an extension of the existing RTM infrastructure in CUBE using the following adjustments:

- Implement an algorithm for distributing the existing ward-data (for both current and future situation) across the corresponding grid cells.
- Develop an application that connects the RTM-network to the grid-nodes.
- (optional) Perform separate household surveys (as opposed to the RVU, which is an individual travel survey) that are tailored for acquiring household travel-diaries.

2. Introduction

Rambøll has been engaged by the Norwegian Public Road Authority (NPRA) to look at further developments of the activity-based model (ABM) based on previously conducted studies done on the topic. The first of the studies was done in 2022 where the future needs for transport models inside and outside city areas were evaluated[1]. For rural and inter-urban areas, the main goal is to evaluate socio-economic benefits and use this to prioritize between projects. For urban areas the transport system is more complex with several different measures to evaluate and an overall goal of achieving the zero-growth target. For the rural areas, the national (NTM) and regional (RTM) transport model in Norway is suitable for the overall goal of socio-economic profitability. This is due to less challenges regarding capacity and intermodal competition, and therefore the four-step model is sufficient for this purpose. For urban areas it is recommended to look further in developing ABMs by conducting pilots (Algers et al., 2005).

The second study done by Rambøll, was the development of a pilot called BRUTUS byRAMBOLL, for Trondheim. This was an activity- and agent-based model. In the report *Brutus model for Trondheim* (2023), Rambøll demonstrated how a disaggregated model differs from aggregated four-step models in terms of accuracy, functionality, and development process. The two main conclusions from the project are that Brutus can be used for urban transportation planning as a stand-alone model. Furthermore, agent-based models can complement the regional transport model (RTM) in modelling certain policy measures for which RTM is less suitable.

3. Background

Transport models can be categorised into four major model types[2]: trip-based four-step-models, tour-based four-step-models, Activity-based models (AcBMs) and agent-based models (AgBMs). The four-step models are aggregate models that simulate trips or tours as aggregate flows of traffic. As a result, the behaviour of the population is homogenous. In the trip-based four-step model, the unit of travel is an individual trip, while in tour-based four-step models, a tour (a chain of trips) serves as the unit of travel.

In contrast, ABMs (and agent-based models), simulate travels on an individual-level (Castiglione et al., 2015; Doherty & Mohammadian, 2011), allowing the model to consider personal characteristics (Zhang & Levinson, 2004). Activity-based models generate travel demand based on the desire for activities (Doherty & Mohammadian, 2011). However, before the traffic assignment step, the demand is often aggregated (Agriesti et al., 2022). A complete agent-based model doesn't necessarily model activities but can directly generate the population and distribute it individually onto the network. Consequently, not all agent-based models are activity-based, as they are not necessarily constrained by activities. Furthermore, activity-based models can be partially or fully agent-based.

The ABM modelling framework consists of the four main building blocks as presented in Figure 1. The first block represents all the model inputs. The data covers available transport infrastructure and services (road, transit, bicycle, and pedestrian networks), the spatial definition of the study area and land-use data, socio-demographic and mobility data, and policy scenarios.

[1] See also: Vurdering av framtidig transportmodellbehov i og utenfor byområder (2022) report by Ramboll.
[2] See also: Brutus model for Trondheim (2023) report by Ramboll.

All the data from the first building block are used to perform modelling in the second building block, providing the basis for the demand and supply models. The second block includes dividing the study area into zones/grids, generating synthetic population with attributed socio-demographic characteristics and mobility patterns, and assigning it to the zoning system. Additionally, the level of service for each transport alternative is usually calculated using the network characteristics.
The third building block represents demand and traffic assignment modelling steps. The demand model is calibrated using the travel survey for the study area and the inputs from the previous two building blocks. The trip patterns of each representative of the synthetic population and their spatial distribution can be sampled from the revealed preference travel survey or modelled (e.g., using multinomial logit model). The model generates the mobility of these individuals constrained to their generated residential location, land-use distribution in the study area, and the transportation network performance for the generated transport modes. The estimated mode choice model produces the probabilities used to assign a transport mode for each trip chain.

After development of the model for the baseline, a set of policy scenarios can be developed to be tested by the model. The outputs for each tested scenario normally include modal shares, travel times, average trip distance per different demand segments, and emissions.

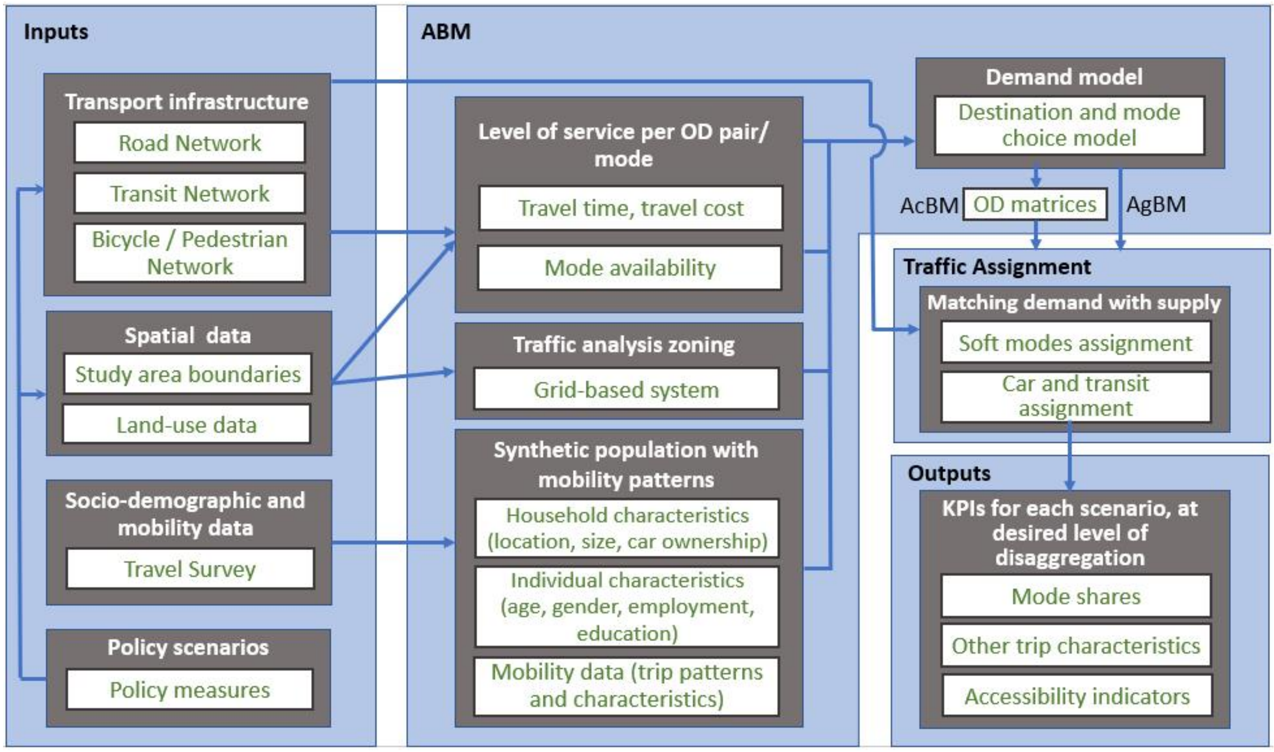

Figure 1: The four main building blocks of the AcBM/AgBM modelling framework.

Activity-based models provide a more natural and detailed representation of the transport system by using the personal preferences to model individual trip chains[3]. This has benefits in terms of what the model can analyse and simulate. For instance, RTM has a low level of detail regarding urban areas, where ABMs can be more precise. It also creates a more suitable platform for testing new mobility developments, such as on-demand public transport, car-sharing, micro-mobility, new cycling

[3] Travel forecasting resource: https://tfresource.org/topics/Benefits_of_Activity_Based_Models.html

infrastructure and so on (Bastarianto et al., 2023). A more detailed analysis with the possibility to evaluate more complex measures creates a good base for decision making in urban areas.

In this study, we will describe the requirements for input data in an ideal situation. Additionally, we will assess the quality of existing data sources, pinpointing any gaps that must be addressed before implementing an activity-based model. Finally, we will discuss potential model implementations and conclude with some remarks.

4. Requirements of ABMs

While the input data types required for ABMs are not drastically different from aggregate models, extra requirements exist for deploying an ABM ideally and capitalizing on their unique benefits (Vyas et al., 2018). We distinguish the main three data sources and outline their requirements in Table 2. Specific considerations compared to four-step models relate to spatial resolution, the need to generate a synthetic population, the modelling of activities sequenced in discrete trip chains and the possibility to model interaction between different decision makers.

Table 2: Listing of data requirements for main data sources to conduct an ABM.

Data type	Considerations	Requirements
Survey (Used for choice model estimation, synthetic population modelling and validation)	ABMs require a detailed and diverse travel survey with individual observations and a high geographical accuracy. Typically, a larger number of choice models is estimated from survey data due to diverse population segmentation. A higher geographical accuracy can help to better estimate sensitivity to independent variables in relation to the size variable of locations. Data collection for entire households is important for modelling of household interaction. A high level of detail requires robust data privacy processes. Often, short trips are underreported in surveys.	*Hard Requirements*: High trip endpoint accuracy, Individual observations with detailed socio-economic variables (age, employment status, income, car ownership, gender), Detailed household composition data, Sufficient sample size – including smaller population segments, Data on transportation modes combinations (can be especially useful for modelling and testing policy measures related to park & ride, ride & bike, kiss & ride, etc.), Information on activities between trips. *Nice to Have*: E-bike ownership, Work from Home related queries, Details of vehicles owned by households, work/office location information. New modes (e-scooters, e-bike, flying cars), dwelling type, how a person pays for travel (monthly public transport ticket, commute costs compensation). Travel diaries for all household members, including detailed information on trip purposes.
Socio-economic spatial data (Used for defining destinations and synthetic population modelling)	Higher resolution data is required due to runtime correlation with the number of decision makers, as opposed to the number of zones. ABMs are better suited to model short trips, usually represented as intrazonal in larger zonal systems. For synthetic population generation, detailed population dispersion data and the right socio-economic background variables (age, employment status, car ownership) are necessary.	*Hard Requirements*: High resolution spatial data (zone size of about 6 hectares in built-up areas), Geographic specific demographic data. *Nice to have*: Detailed specification of educational facilities, Car ownership distribution, Workforce and employment status data, Parking capacity data (relevant policy for municipalities). Ideally the land-use

		model (the ones used for future demographic and land-use forecasts) should be of the same spatial resolution / zoning system.
Supply (Used for making travel time matrices, accessibility metrics and assignment)	Detailed transport networks are required in ABMs to model the detailed routing required for modelling active mobility. Multimodal networks or networks with auxiliary modes allow for modelling intermodal trips.	*Hard requirements*: Detailed network for all modes, including walking and cycling, Super and auxiliary network for easy mode transfers, especially relevant for trips involving public transport or other mobility options. *Nice to Have*: Detailed network and environmental attributes to account for route choice preferences of different people. The details relevant for active modes are: data on slopes, quality of bike and walking ways. *Public transport comfort variables, etc*.

The next section investigates main sources available in Norway and assess these against the requirements listed in this section.

5. Benchmarking of existing sources

This section investigates the following sources:

- RVU travel survey
- SSB statistical grid
- Data from the RTM modelling framework, with a particular focus on assessing if the existing data can act as a basis for building an ABM.
- Alternative sources

5.1 RVU travel survey

The source used for this report is the travel survey questionnaire from 2021/2022 (conducted by Opinion) and the summarized questionnaire from the survey held in 2022. The questionnaire in the newest version is divided into eight sections with different information regarding demographic for each respondent and details on their travels for a specific date.

5.1.1 Demographic background variables

The demographic information collected by the respondent is the following:

1. Place of residence
2. Access to different means of transport
    - Driver's licence and access to vehicles
    - Ownership to car with car model year, car-type and fuel-type
    - Number of cars in household
    - Charging options for electric car
    - Access to bicycle, moped, MC, e-scooters
3. Profession
    - Status of profession, working conditions
    - Place of work
    - Place of school/studies
4. Household
    - Number of residents in household and their relations to the respondent
    - Age, driver's licence and professions per household member
    - Access for parking near household
    - Public transport near household
5. Background information
    - Education level, profession and industry
    - Person and household total income
    - Physical impairment factors that limit the use of different means of transport
    - Country of origin

For constructing a synthetic population, the survey contains the relevant data. It has specific socio-economic variables such as age, employment status, income, car-ownership and gender. Also, household composition information can be derived, except for gender.

5.1.2 Reported trips

The information regarding the respondents travel for a specific date is the following:

1. The day of the trips
    - Activities, and registration of up to fifteen trips
    - Access to a car on the day of trips

    For each trip:

  - Starting time, purpose of trip, starting point and end point for the trip (address or ward)
  - All transport modes used on the trip
  - Payment method used if public transport is used (single ticket, monthly, and so on)
2. Travel frequency
  - Frequency of different travel modes for the time of the year the survey was conducted
3. Long distance travels the latest month
  - Number of travels over 100km in or outside of Norway for the last 30 days
  - For each trip: weekday, purpose, start- and endpoint and transport mode
  - If applicable: overnight stays
4. Travel to work
  - Approximately length of trip and time use by the different modes.
  - Possibility for parking at workplace or near workplace.
  - If you get a compensation for travel expenses such as for a car, the tolls, parking, PT tickets and so on.

The survey provides a detailed description of travel, making it possible to reconstruct trips into detailed tours, possibly sub-tours and multimodal trips:

- All the different means of transport used by the respondent for each trip is stated in the survey. The transport mode used for most of the trip (in terms of distance travelled) is investigated further. For trips using public transport the walking time, waiting time, delay and if they had a seat or not are stated. For trips using a car for the longest amount the number of passengers, what car was used and the change between modes are described.
- The high endpoint accuracy needed by an ABMs a appears fulfilled with today’s survey if data is provided on address level.
- Data on the combination of transport modes are also important and included.
- The activities between trips are described as well with up to 46 different options.

Some limitations for the survey have been identified, too:

- Inclusion travel diaries for all the members of the household is be preferred, this is not the case today.
- For new modes of transport, the survey is specific in terms of what modes they have access to, but for each trip registered the modes does not match that question. For instance, you can answer that you have access to an e-scooter, but one cannot say that you used an e-scooter for a trip.

A difficulty in travel surveys is that people often leave out short trips. This is especially important for ABMS because these models try to capture such trips more accurately than aggregate models. The RVU documentation specifically mentions this limitation.

The data that can substantially benefit the model includes the information related to costs of the travel alternatives for the respondents. These do not imply the questions asking for the exact costs / ticket prices but include questions on who is providing and paying the car, paying the parking or PT card (e.g., employer), also if the person has any monthly card for PT, or a card with a reduced tariff (e.g., student, retired). If these data are missing, and the costs are calculated just proportionally to distance (for car), then the costs become highly correlated with travel times, and the model estimation becomes difficult mathematically and often leads to unreasonable cost coefficients (e.g., with negative sign). And missing data on the actual PT costs leads to unrealistic cost coefficient estimations, while these coefficients are highly important for testing cost-related transport policy measures, using the model. The RVU has most of the questions required for the estimation of the transport user costs, so it is important to keep such questions in future, or even add more questions / elaborate the existing ones, which would allow better estimation.

Additional information, which can benefit the model accuracy substantially, is the type of residence, such as a detached house versus an apartment. This variable has a strong explanatory power both in the car ownership model and in mode choice model. For example, in the Swedish National Agent-Based Model Sampers 4, a dummy variable for living in a villa is significant in the car ownership choice model, and in choice models for certain trip purposes and modes (e.g., car trips for shopping and recreational, bicycle trips for going to a school, etc.) Finally, it is a variable which can enforce the transport and land-use model integration. As the RVU has an address of each person, the value of this variable can be derived without adding a separate question into the survey.

5.1.3 Sample size

ABM usually requires more choice sub-models to be estimated compared with a four-step model. These are models for specific population segments and combinations of trip purposes. Also, other additional models are possible, e.g., activity type choice, choice of the departure time or sub-tour models. A good sample size is required to reach statistically significant estimations in each model.

Literature does not suggest information on the reasonable sample size. The main reason is that the sample size depends on many factors specific to the study area and its population, the model specification and the desired confidence level. Therefore, the literature suggests formulas for calculation of the required sample size depending on these factors (e.g., Habib, 2020). Since the sample size depends on the model specification, ideally, the variables to be included into the model, and all the choice alternatives, must be known before the calculation of the sample size. In reality, this is often hard to achieve, as the decision on the variables to be included and their acceptable statistical significance could be made during the model estimation. Additionally, there are some limitations in terms of budget and time, for collection of a sufficient number of the responses. Therefore, most of the existing models are estimated based on a smaller sample size, than it would be required ideally, strictly following statistical requirements for the sample representativeness.

There are some ways to enrich the survey samples using modern data collection methods, such as using smart phones (e.g., Nahmias-Biran et al., 2018), as well as methods of dealing with small sample sizes given the high spatial resolution of census data (Felbermaira et al., 2020).

Having a smaller sample size can save the budget on data collection but can lead to losses due to the higher uncertainty. Because of the computational complexity, the number of uncertainty analysis studies for transportation models, and especially for ABM, is fairly limited. The existing studies suggest that the project ranking in the cost-benefit analysis is not affected a lot by uncertainty in a transportation model (Asplund & Eliasson, 2016). Another kind of conclusion from uncertainty analysis study for the ABMs is that the variation mostly depends on the frequency of alternative being predicted from the choice process, being higher for less frequent alternatives. At the same time, the less frequent alternatives affect the overall indicators of interest less, due to their smaller effect in general. For example, the ABM model of Singapore relies on the around 30.000 individual response but the number of the respondents making a work-based subtour is less than 1000, leading to high variation in the model results related to the work-based subtours. At the same time, as very few people make those sub-tours, this has relatively little effect on the variations of total travel times, congestion level, etc. (Petrik et al, 2020).

The recommendation regarding the sample size could be, therefore, having a separate pilot-study, which will investigate these matters closely, based on literature review and, possibly, a pilot survey.

Another option is just to have the sample size as it is, accepting lower levels of confidence and statistical significance as a part of the uncertainty associated with the model.

5.2 SSB statistical grid

Activity-based models benefit from high resolution spatial data to accurately model trips. The ideal zone size for a statistical grid is approximately 250X250m meters when one wants to simulate urban traffic on a dense network without connectors obscuring traffic flows too much. We explored current availability of necessary data from Statistics Norway. It is feasible to obtain data at various grid levels—for example, sizes of 250x250m, 500x500m, and 1000x1000m—for a range of socio-economic variables. Accessible information such as population, dwellings, and buildings are publicly available. Procuring more detailed data distributed on a grid level requires an application[4] to Statistics Norway for access. So, it appears that the grid level data from Statistics Norway meet the hard requirement of the high-resolution spatial data. However, for more specific demographic data, like age and employment status, an application is necessary to get all the required data to take full advantage of ABMs modelling of populations and short trips. See Table 3 for a summary of the available data from Statistics Norway.

Table 3: Grid level data from Statistics Norway

Statistical variable	Accessibility
Population	Open
Dwellings	Open
Buildings	Open
Holiday houses	Open
Establishments	Open
Age	Apply to get access
Employment status	Apply to get access
Other demographic data	Apply to get access

The grid level data from Statistics Norway appears to meet most of the hard requirement of the high-resolution spatial data. However, rigorous data privacy processes may be required to make the model available to a wide-user basis.

5.3 Data from RTM modelling framework

5.3.1 Zonal data in RTM

Zonal data in RTM follows the ward structure. As identified earlier, a high-resolution land-use description is preferred to adequately model trips by active modes in urban areas. A clear benefit of zonal data in wards is that it meets data privacy requirements. This makes the model accessible to a wide range of stakeholders without needing data processing agreements. A key question is to what extent the ward structure is suitable, or what steps can be undertaken to (synthetically) increase the resolution of the wards to a spatial grid (as done in the Trondheim pilot).

Resolution: In densely populated urban places (densest areas in urban centres), the resolution of wards is often adequate. However, outside these centres (areas surrounding the centres), zone sizes expand considerably, which could limit insights into active mode trips, such as cycling or walking to schools. In rural areas, where zone size increases even further, the issue of accurately modelling flows and trip endpoints might be less pressing due to their clustered land-use patterns and lesser trip end generations.

[4] For public planning Statistics Norway require that the application is sent by the public authority.

Methods can be conducted which distribute socio-economic data from wards zones to a small zonal description, for example through combining ward data with the SSB statistical grid, or with information on building level. The Brutus Trondheim pilot shows that this possible, but a few limitations can be identified. Firstly, because significant errors caused by edge-cases are hard to detect without manual inspection. This may compromise the results for example for specific areas that generate a lot of trip ends. Secondly, differences in e.g., demographic composition in different areas of a ward may be averaged out. This may lead to more homogenous behaviour in the model when this happens on a large scale.

5.3.2 Car ownership model

RTM calculates car ownership on ward-level by modelling the shares of people falling into five different segments which describe their access to a car. Population in these wards are aggregated into 120 segments based on gender, age group, family type and number of adults in the household. For each segment, the car ownership model estimates the number of people falling in each group:

1. No drivers licence, no car (DBTP)
2. No drivers licence, car in household (FBTP)
3. Driver’s licence, no car (DBTF)
4. Driver’s licence, full car-access (at least one car per person with a driver’s licence) (FBTF)
5. Driver’s licence, limited car-access (fewer cars than persons with a driver’s licence) (GBTF)

Each share is calculated based on a function incorporating the following parameters:

a. number of persons divided by gender and family type.
b. number of persons divided by size of household and age group.
c. number of persons divided by gender and age group.

Thereafter, the first three segments are all aggregated into a single group which is not able to drive a car as they either cannot drive or do not own one.

As it was mentioned in the section related to the survey content, residence type variable (living in a detached house) normally has a very strong explanatory power for a car ownership models, so it can be considered for future version of the Norwegian car ownership model.

A model like this produces aggregated car ownership values for an area. This is different from an ABM, which models discretely if or how many vehicles each household owns. It may be possible to apply the same car ownership model on a synthetic population, but as of now it is hard to draw a conclusion without more insights in the mathematical description of the model. However - even if possible - this may not be the most natural implementation. For a synthetic population, it is more natural to model car ownership on household level based considering other characteristics too, such as household income, dwelling type and home location. It may be more suitable to develop a new car ownership model based on travel survey data and the vehicle register.

5.3.3 Networks

There are multiple networks available for the development of an ABM, however, the most likely approaches would consist of either using an open-source networks (such as the OpenStreetMap) or building the model on the network established for the regional transportation-models (RTM) based on NVDB (National Road Database).

The benefits from using the networks from open sources such as the OpenStreetMap is that they contain a high level of detail for the walking and cycling infrastructure, which is currently not present in RTM's network. However, data about actual travel times or value delay functions are usually not present. Nor is it safe to just assume that coverage is adequate across the whole country. If open-source networks were to be used, the work would also involve the creation of a methodology for editing the network and connecting it to the zonal data.

An approach built on using the RTM-network would therefore be easier to implement as many of the processes relating to network editing could be performed using the current model-infrastructure. Also, the implemented network in RTM is comprehensive for modelling public transport as well as access and egress networks. However, the drawback of the approach would be that the walking and cycling infrastructure in not completely represented within the RTM-network due to the case that changes to the walking and cycle infrastructure sometimes are not entered into NVDB. Adding infrastructure for walking and cycling into the database would be required.

5.4 Super and auxiliary networks

A super network is a type of network that enables mode switching. This means users can change their mode of transport - for instance, transitioning from walking to using public transport. Equally, an auxiliary network represents a simpler concept. It typically exemplifies the network utilized by individuals to reach public transport access points, such as stops or stations. Based on the currently available data from the previously mentioned sources, it's feasible to construct both types of networks and integrate them into an activity-based model.

5.5 Alternative spatial data sources

For grid level data, there do not seem to be any sources providing useful data. However, alternative data sources do exist, such as the data from Kartverket for day-cares, primary, and secondary schools. Although this data is publicly accessible, it is not distributed on a grid. Consequently, this point-of-interest data needs to be distributed on a grid. The same applies to data from other sources such as OpenStreetMap. But combining this with building data from Statistics Norway could be useful in determining the purpose of a trip. Drawback of such sources as mentioned here is that size variables are often missing (size of the school for example by number of students) which are strong indicators of how attractive a destination is.

6. Findings

Our review of the available data and data-sources up against the data-requirements for an ABM development in Norway, has found that the current data-sources in use provides most of the "must have" data. However, there are some shortcomings which will need be addressed if an ABM is to be realized.

Going through the different data-categories, it can be divided into the following three categories: (1) available data, (2) undetermined data and (3) unavailable data. The different datatypes which fall within each category is listed in table 3.

Table 4: Overview of available data and required data-properties.

Data Type	Fully Available	Partly Available	Unavailable
Survey Data	• High endpoint accuracy detailed socio-economic variables. • Detailed household composition data • Data on mode usage • Information on activities between trips	• Sufficient sample size • Information on mode attributes • Data related to costs of travel alternatives (e.g., PT monthly card ownership, employer paying for car / parking)	• Travel diaries for all household members • Trip-diaries for ages <13 • Data on the residence type (detached houses vs. apartments)
Socio-economic spatial data	• High resolution spatial data (zone size of about 6 hectares in built-up areas) • Geographic specific demographic data	• Geographic specific employment data. • Zonal driver's license data • Car ownership	
Supply	• Detailed network for public transport, walking and cycling facilities. • Super and auxiliary network for easy mode transfers		

In the following we provide a brief discussion regarding the availability of the different datatypes, and possible solutions for the different data-limitations that has been identified.

6.1 Survey data

Table 4 shows that the current survey data provide most of the necessary data for the development of an ABM. The only uncertainties are related to a) the properties of the sample regarding size (which may vary between regions/areas and as it is difficult to give an assessment of what constitutes the required sample size), and b) the lack of travel diaries for all household members (the survey only contains the travel diaries of the interviewees).

The uncertainty regarding the sample size is hard to assess, but potential shortcomings can be compensated for by aggregating data for larger or similar areas, and/or merging travel surveys for different years. Also, lower model significance for small population segments can potentially be acceptable if most decision makers in the model are served with statistically significant models.

The issue of the missing travel diaries for the entire household members is considered to have a more significant effect on the model properties. Acquiring household travel diaries would be beneficial as it would provide data on how households solve their combined travel-need, and how the trips of the different agents (bringing kids to school, giving partners a lift etc.) are integrated together.

Ideally, the issue of the missing household travel diaries could be solved through expanding the current travel survey to include questions regarding the travels of the entire household. However, as the travel survey already has issues with being highly time-consuming for the participants, we advise caution in expanding the current survey. As a part of the work of mapping the available data, we have considered the possibility of acquiring parts of the household travel-data through adding a few generalized questions regarding the other household-members to the RVU. However, the utility of incomplete information is likely to be limited, and we do therefore advice against such an approach.

Having considered the different options, and the data that is available in the current RVU, we conclude that an ABM can be developed using the existing data, and that the missing household diaries first and foremost will cause increased uncertainty in the model as the missing data would have to be replaced with assumptions.

Our assessment of the situation points towards the best approach being that of developing the ABM using the existing data (thus keeping the RVU as it is). Then, based on an assessment on the model behaviour, separate household travel-surveys can be performed at a later stage that are tailored for the ABM areas. Acquiring the household trip-diaries for specific areas separate from the RVU should provide a better trade-off between response rate and detail than expanding the current RVU, as the surveys can be tailored into capturing the most relevant data, and therefore be smaller in size regarding both sample size and number of questions.

6.2 Socio-economic spatial data

Regarding the Socio-economic spatial data, our review Table 4 shows that most of the data are available within the required level of detail (grid level), but that the data regarding employees per sector, driver’s licence status, and vehicle ownership are only partly available.

Firstly, regarding the employees per sector, the data for total number of employees is available on grid level, but the distribution of these across different sectors are not. It may be that the data per sector exists and can be acquired, however, the data may have to be corrected for the effects of staffing agencies, jobs attached to the company main address etc. It is therefore likely that it is more attractive to use the existing jobs-data per ward and distribute across the corresponding grids.

Regarding the data on driver licence and car ownership on grid level, the review has found that these are not easily accessible. The data exists, but the work involved with acquiring it on grid level is unclear, but it is likely to be substantial. Other methods may be suitable. One approach for acquiring a suitable dataset could be using car ownership figures produced by RTM. Car ownership can be estimated on household level. Any ward level car ownership information can be used to constrain the car ownership predictions of households living within the ward.

6.3 *Supply*

Our review of the available supply data has found the currently available data sources to be sufficient for the establishment an ABM. There are some potential issues related to the RTM-network (the walking and cycling network may at some areas be incomplete), but these can be solved through using different networks for the different types of traffic (RTM network for car traffic, street-map for walking and cycling etc.).

7. Recommendations

As part of the project «Grunnlagsdata og reisevaneundersøkelser for ABM», Rambøll has been asked to make a review of the currently available data, and based upon its findings make a recommendation for the future development of a Norwegian ABM. Specifically, Rambøll has been asked to advice among the following options:

1. Using the input- and survey data in their current form
2. Use the input and survey data in their current form, but with a recommended set of minor adjustments and addons.
3. Creating an optimal dataset of input-data and survey-data which follows the best practice.

Based on our review of the currently available data, it is our conclusion that the best approach consists of using the input and the Norwegian travel survey (RVU) in their current form, but with a recommended set of minor adjustments and addons (option 2). The current input and survey data is in other words found to be mostly suitable, and an operational ABM can be developed as an extension of the existing RTM infrastructure in CUBE using the following adjustments:

- Implement an algorithm for distributing the existing ward-data (for both current and future situation) across the corresponding grid cells.
- Develop an application that connects the RTM-network to the grid-nodes
- (optional) Perform separate household surveys that are tailored for acquiring household travel-diaries.

8. References

Agriesti, S., Roncoli, C., & Nahmias-Biran, B. H. (2022). Assignment of a synthetic population for activity-based modeling employing publicly available data. *ISPRS International Journal of Geo-Information, 11*(2), 148.

Algers, S., Eliasson, J. & Mattsson, LG. (2005). Is it time to use activity-based urban transport models? A discussion of planning needs and modelling possibilities. Ann Reg Sci 39, 767–789. https://doi.org/10.1007/s00168-005-0016-8

Asplund, D., and Eliasson, J. (2016). Does uncertainty make cost-benefit analyses pointless?. *Transportation Research Part A: Policy and Practice*, *92*, 195-205.

Bastarianto, F.F., Hancock, T.O., Choudhury, C.F., and Manley, E. Agent-based models in urban transportation: review, challenges, and opportunities. European Transport Research Review, 15 (2023). https://doi.org/10.1186/s12544-023-00590-5

Castiglione, J., Bradley, M., and Gliebe, J. (2015). Activity-based travel demand models: A primer, *Transportation Research Board*

Doherty, S.T., Mohammadian, A. (2011). The validity of Using Activity Type to Structure Tour-Based Scheduling Models. *Transportation*, 38 (1), 45-63.

Felbermaira, S., Lammera, F., Trausinger-Bindera, E., & Hebenstreita, C. (2020). Generation of a synthetic population for agent-based transport modelling with small sample travel survey data using statistical raster census data. *International Journal of Traffic and Transportation Management*, 2(02), 9-17.

Habib, K. N., El-Assi, W., & Lin, T. (2020). How Large is too Large? A Review of the Issues related to Sample Size Requirements of Regional Household Travel Surveys with a Case Study on the Greater Toronto and Hamilton Area (GTHA). arXiv preprint arXiv:2005.00563.

Nahmias-Biran, B. H., Han, Y., Bekhor, S., Zhao, F., Zegras, C., & Ben-Akiva, M. (2018). Enriching activity-based models using smartphone-based travel surveys. *Transportation Research Record*, 2672(42), 280-291.

Petrik, O., Adnan, M., Basak, K., & Ben-Akiva, M. (2020). Uncertainty analysis of an activity-based microsimulation model for Singapore. *Future Generation Computer Systems*, *110*, 350-363.

Vyas, G., Vovsha, P., Gupta, S., Givon, D., Birotker, Y., Mossek, A., ... & Bluer, E. (2018). Stepping Closer to ABM: Hybrid 4-Step Models (No. 18-02673).

Zhang, L., Levinson, D. (2004). Agent-based approach to travel demand modelling: Exploratory analysis. *Transportation Research Record*, 1898(1), 28-36.