On the use of common random numbers in activity-based travel demand modeling for scenario comparison

Zhou, H.; Dorsman, J. L.; Mandjes, M.; Snelder, M.

DOI
10.1080/03081060.2023.2182784

Publication date
2023

Document Version
Final published version

Published in
Transportation Planning and Technology

Routledge
Taylor & Francis Group

Transportation Planning and Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gtpt20

On the use of common random numbers in activity-based travel demand modeling for scenario comparison

H. Zhou, J. L. Dorsman, M. Mandjes & M. Snelder

TRANSPORTATION PLANNING AND TECHNOLOGY
https://doi.org/10.1080/03081060.2023.2182784

Routledge
Taylor & Francis Group

Check for updates

On the use of common random numbers in activity-based travel demand modeling for scenario comparison

H. Zhou [a,b], J. L. Dorsman[b], M. Mandjes[b] and M. Snelder[a,c]

[a]Department of Sustainable Urban Mobility and Safety, Netherlands Organisation for Applied Scientific Research (TNO), The Hague, The Netherlands; [b]Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands; [c]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

ABSTRACT

Activity-based travel demand models provide a high level of detail when modeling complex travel behavior. Since stochastic simulation is used, however, this high level may induce large random fluctuations in the output, necessitating many model reruns to produce reliable output. This may become prohibitive in terms of computation time when comparing travel behavior between multiple scenarios, in which case each scenario requires its own simulation. To alleviate this issue, we study the use of common random numbers, which is a technique that reuses the same random numbers for choices made by travelers between scenarios. This ensures that any observed difference in output across scenarios cannot be attributed to mutual differences in drawn random numbers, eliminating an important source of random fluctuation. We demonstrate by a numerical study that common random numbers can greatly reduce the number of runs needed, and thus also the required computation time, to obtain reliable output.

ARTICLE HISTORY
Received 26 January 2022
Accepted 4 January 2023

KEYWORDS
Scenario comparison; common random numbers; activity-based modeling; travel demand

1. Introduction

In this paper, we study an efficient simulation method to compare the travel demand behavior under different traffic scenarios by means of activity-based travel demand modeling (ABM). Owing to its flexibility, robustness and high level of detail, ABM offers a highly suitable methodology to model complex travel behavior. In travel demand simulation, for each individual, ABM predicts what, where (destination), when (time) and for how long (duration) travel activities are conducted as well as which mode chain of transport is involved (Rasouli and Timmermans 2012). To achieve this high level of detail, an ABM may for example consist of different discrete choice models which successively make choices on e.g. destinations, time, duration, and mode. In most cases, these models adopt a simulation approach that makes choices based on (pseudo-)random numbers. The variability caused by the generation of these random numbers however

CONTACT H. Zhou han.zhou@tno.nl www.tno.nl Department of Sustainable Urban Mobility and Safety, Netherlands Organisation for Applied Scientific Research (TNO), The Hague 2595 DA, The Netherlands; Faculty of Science, University of Amsterdam, Amsterdam 1090 GE, The Netherlands

trigger random fluctuations in the output of the ABM (Vovsha, Donnelly, and Gupta 2008), also referred to as simulation error. This may cause the activity-based model to have to be rerun many times in order to get reliable results, which may not be feasible due to the excessive simulation effort required.

As mentioned above, ABM models travel behavior at a high level of detail, and thus provides output at a very low level of aggregation. In the literature, multiple studies have shown that the lower the level of aggregation of a model's output is, the more profound the issue of variability/simulation error becomes, underlining the fact that this is especially a complication for ABM. Indeed, while Veldhuisen, Timmermans, and Kapoen (2000) concluded that in the context of their RAMBLAS framework, a microsimulation model, the simulation error is negligible when studying the output at a highly aggregated level, Castiglione, Freedman, and Bradley (2003) confirmed this finding based on a case study on the travel demand in San Francisco but with the remark that simulation error becomes problematic when there are many choice alternatives or when some of these alternatives are very rare. To quantify these effects, Bao et al. (2015) used FEATHERS, a rule-based model, to determine the minimum number of runs required to obtain enough 'confidence' at different aggregation levels, i.e. the minimum number of runs required so that the output is sufficiently reliable. The results of this work indicate that the lower the aggregation level is of the desired output, the more model runs are required. We also mention (Horni, Charypar, and Axhausen 2011), where the random variability over multiple runs of MATSim is studied by analyzing travel demand on specific road links in a traffic network. The authors concluded that there is relatively little variability when regarding daily volumes, but that variability is significant when considering hourly volumes. This is in line with the fact that the lower the aggregation level of the results is, the more variability becomes an issue. It is worth noting that in traffic modeling, travel demand models and travel assignment models are often used in an alternating way, so that unreliable results in travel demand have direct ramifications for travel assignment. These ramifications have been studied in e.g. Vovsha, Donnelly, and Gupta (2008), Horni, Charypar, and Axhausen (2011) and Bekhor, Kheifits, and Sorani (2014).

While the studies mentioned above are mainly focusing on the model output of a single traffic scenario, the issue of simulation error is even more profound when comparing multiple traffic scenarios for the purpose of quantifying the difference between them. Especially when the actual difference between scenarios is not very large, much computation time may have to be spent to produce a reliable output for each of the scenarios, before any conclusion can be drawn regarding the difference. In this context it should be noted that when differences between scenarios are small, this does not necessarily mean that such differences are by definition irrelevant. For example, even when a certain scenario leads to only a 2% increase in the number of trips undertaken, in the regime of a highly loaded network this may have a significant impact on the level of congestion. In this paper, we discuss techniques to control the simulation-error issue in this multi-scenario context. More particularly, we study the use of the technique of *common random numbers* (CRN) to overcome the problem of requiring too many simulation runs to reliably estimate the simulation error. CRN is a celebrated technique stemming from the stochastic simulation community that attempts to induce a positive correlation between the outputs of different scenarios. It does so by using the same generated pseudo-random numbers for the same purposes across the simulation of different scenarios. When doing this, the observed difference in the model output for the different scenarios can then not be attributed to the fact that random numbers across scenarios differ, which is one of the sources of simulation error. This increases the likelihood that any difference in observed model output is a result of the intrinsically different features of the scenarios. This means that the number of required simulation runs will decrease, and as a result it induces less required computation time. For a more detailed explanation on CRN, see Glasserman and Yao (1992) and Section 9.7 of Ross (2013). In the context of activity-based travel demand modeling, the use of CRN in has been suggested before in e.g. Vovsha, Donnelly, and Gupta (2008) and has been implemented in CEMDAP (Pinjari et al. 2008) and MATSim (Horni, Nagel, and Axhausen 2016). As a result, we are not the first to implement CRN in an ABM. However, to the best of our knowledge, there has been no quantitative study on the added value of CRN in terms of required numbers of runs and computation time savings. This paper seeks to fill that gap. More particularly, in the rest of this paper, we aim to show how computation times can be shortened drastically by using CRN.

The contributions of this work can be summarized as follows. First, we demonstrate how to implement CRN in an activity-based travel demand model to compare multiple scenarios. We do this based on an extension of an activity-based travel demand model, which was recently developed by the authors Zhou et al. (2020). This extended ABM integrates a tour-based multimodal mode choice component in the activity-based model, which makes travel mode chain choices on each trip of a tour using a multinomial probit choice model. We however stress that the technique of CRN can- and in this paper will - also be applied to all other components of an ABM, also those based on a multinomial logit choice model, which we will also cover. Secondly, based on this extension, we demonstrate the potential of CRN when assessing multiple scenarios, in particular with regard to the question which scenario is more favorable (according to a given criterion). We do so by analyzing the variability of the difference between the model output of these scenarios, with and without the implementation of CRN, and we will conclude that the implementation of CRN greatly reduces this variability (measured in terms of the sample variance of the differences). As a result, to obtain reliable output, only a fraction of the earlier required simulation runs is still needed. For example, we show that, to keep the width of the confidence intervals of simulated indicators under a certain level, a reduction of the number of required simulation runs by a factor 100 is not an exception, which evidently has a drastic effect on the computation time needed. It turns out that CRN is especially worthwhile when the order of magnitude of the differences between the scenario outputs is either not too large, in which case traditional methods are already able to reliably indicate which scenario is favorable in an acceptable amount of time, or not too small, in which case the scenarios hardly differ in terms of the indicator studied anyway. We back all the aforementioned claims by studying numerical results of our extended model based on data of the metropolitan region Rotterdam-The Hague in The Netherlands stemming from various sources, such as Snelder et al. (2021) and Van de Werken (2018).

The remainder of this paper is organized as follows. Section 2 explains how CRN can be incorporated in an activity-based travel demand model. It furthermore introduces several ways of assessing the impact of CRN, such as by computation of the minimum required number of runs in order to keep confidence interval widths limited. In Section 3, we present the numerical study investigating the potential of CRN, which leads to the findings mentioned above. Finally, Section 4 presents conclusions and a brief discussion.

2. Method

An activity-based travel demand model makes, based on generated random numbers, choices at several levels. That is, from the perspective of a traveler, subsequent choices are made on long-term decisions, number of tours to be undertaken during a day, number of trips to be undertaken in each tour and finally the start time, duration, destination and mode of each trip. Since stochasticity is introduced on each of these choice levels, the final model output may be prone to large random fluctuations, which may lead to having to rerun the model many times and average the results to obtain reliable output. To address this issue, the use of CRN may be introduced at any of these levels. To explain this, we first give a description of the ABM that we use in the remainder of this paper to demonstrate the impact of CRN. This ABM is an extended model integrated within ActivitySim (Gali et al. 2008). Then, in Section 2.2, we demonstrate how to implement CRN at each level. More specifically, Section 2.2.1 describes how CRN is applied in our extended model based on a multinomial probit choice model, while Section 2.2.2 explains how CRN is applied in the other components of ActivitySim mainly using a multinomial logit model. Then, in Section 2.3, we describe several ways to assess the impact of CRN.

2.1. The activity-based travel demand model

To explain the implementation of CRN into an ABM, we take the extended ABM considered in our previous work (Zhou et al. 2020) as an example. This ABM makes choices sequentially through a series different choice components using random utility maximization theory. The first of these components makes long-term decisions on work and/or school locations for every individual of the synthesized population. Based on the output of this component, the next component determines the daily activity pattern of each individual while taking into account the interaction between the individuals in every single household. The following component then makes decisions for each individual on the number of tours undertaken, the duration of these tours and the main travel mode used in each tour. Afterwards, the subsequent component starts to make more detailed choices, namely on the number of stops to be made in each tour, including the departure time, duration and destination of each trip resulting as a result from the stops in the tour. We note that the tours may have a mandatory nature, such as work and study tours, but there are also non-mandatory tours, which most of the time have a more leisurely character. For the components mentioned so far, we have used the implementation of ActivitySim (Gali et al. 2008). The final component used in the model of Zhou et al. (2020) was however coded by the authors. This component makes mode choices for every trip undertaken by a traveling individual as part of the daily activity schedule. While the standard implementation in ActivitySim contains such a component, our customly coded component also allows for making multimodal mode choices, i.e. it can be decided that a traveling individual uses multiple modes within a single trip. This is an attractive feature when for example considering 'Mobility as a Service' (MaaS). To explain this concept, it is worth noting that we are witnessing the development of new transport technologies, such as connected vehicles using 5G, level 3, 4 or 5 automatic vehicles, and mobile app-based car-sharing or ride-sharing services. MaaS combines all these technologies and services, thus offering a tailored mobility package for individual travelers (see e.g. Jittrapirom et al. 2017 for more background). Therefore, when a traveler owns a MaaS subscription, this person has access to a shared car, a shared bike and a shared e-bike. Furthermore, the subscription enables the use of a shared taxi, minibus or other shared modes which are not used in conventional public transport (such as the bus, tram, metro and train). Finally, we mention that this component is based on a multinomial probit choice model.

2.2. CRN generation in ABM

In this section, we turn to the implementation of CRN. That is, in Section 2.2.1, we demonstrate how to implement CRN in a multinomial probit choice model, by taking the customly coded multimodal mode choice component as an example. Afterwards, in Section 2.2.2, we show how CRN can be implemented in a multinomial logit choice model, which is the used model for all the other, original components in the ActivitySim implementation. It is worth noting that for the numerical study, which we perform in Section 3, we apply CRN to all components of the model, according to the directions below.

2.2.1. Multinomial probit choice model

As an activity-based travel behavior model, ActivitySim produces random numbers to make choices on the level of individuals, households, tours and trips. For instance, at an individual level, random numbers are used to select the individual's choice on the location of work or school, while at a trip level, they are used to make choices for trip duration and destination. To apply CRN, one needs to make sure that across different scenarios, the same random numbers are used for the same choices, even when the scenario input is different. While ActivitySim incorporates an option to do this on all levels, implementation of CRN is perhaps best explained based on the extension we performed in Zhou et al. (2020).

Before explaining how to implement CRN, we therefore first treat the extension in detail. The extension concerns a customly coded component, which makes multimodal mode choices at a trip level by combining the access, main and egress modes, while making sure that the combination of these multimodal modes within a complete tour is feasible. For example, when dealing with a tour that consists of a home-work trip and work-home trip, and when the mode used for the outbound trip is the multimodal mode walk-car-bike (access-main-egress), the component will ensure that not just the bike as a unimodal mode is used for the inbound trip, since the car needs to be at home at the end of a tour. The model considers in total 32 feasible multimodal modes (consisting of access, main and egress modes) based on the following seven modes of transport: walk, bike, e-bike, car, car-passenger, demand-responsive-transport (DRT) and public transport. As described in our previous work (Zhou et al. 2020), each of these modes represents a specific mode category based on speed, weight, vehicle space per person occupied measured in passenger car units (PCU) and passenger capacity. For instance, 'bike' represents a category of modes which travel at a speed between 5 and 20 km/h, have a micro-modal nature, come with a vehicle space per person between 0.25 and 0.5 PCU and have no passenger capacity. As a result, not only the private bike fits this category. For example, this category also captured the shared bike, even though it may require some search time unlike the private bike. Whether or not this search time is incurred by the traveler depends on characteristics of the traveler, for instance concerning ownership of a MaaS subscription (governing whether shared services are used). Similar notes apply for the 'e-bike' and 'car' categories. The 'car-passenger' category is self-explanatory, while the 'DRT' category captures taxi services, ride-hailing services, et cetera.

To make choices in this component, a multinomial probit choice model is used. That is, a utility function is evaluated for each multimodal mode choice for every trip in the tour. This utility function incorporates two normally distributed error samples: one error sample is specific to every traveler/mode combination and models the traveler's personal mode preference (and is thus considered equal across trips), while the other error sample is specific to every traveler/mode/trip combination and models other random effects. Ultimately, the choice for all trip modes in the tour corresponds to the feasible mode combination with the highest aggregated utility. For the numerical study in this paper, we adopt the error parameters used in Zhou et al. (2020).

To apply CRN in this additional ABM component, one needs to make sure that across scenarios the same error samples are used for every traveler/mode combination and every traveler/mode/trip combination respectively. To make this happen, we use the notion of initial seeds. Every time the same initial seed is set in a random number generator (RNG), it will generate the same sequence of random numbers. Therefore, incorporating CRN for traveler/mode combinations errors can be done by associating with each traveler a seed. Then, every time a different scenario is considered, the RNG will still generate the same traveler/mode errors, independent of the actual scenario. For the traveler/mode/trip errors, this can be done at a trip level: we associate with each trip a seed, so that each time the trip is considered, the same traveler/model/trip errors are computed. This way, the errors between scenarios are maximally synchronized. Furthermore, this strategy has the additional computational advantage that, when a trip is not undertaken in a certain scenario, the required traveler/mode/trip errors will not be generated either, saving computation time.

2.2.2. Multinomial logit choice model

The multimodal mode choice component treated above incorporates a multinomial probit choice model. In travel demand modeling, however, multinomial logit choice models are also omnipresent. Unlike probit choice models, which assume error terms to be normally distributed, logit choice models assume error terms to be Gumbel distributed. By doing this, these models are able to directly assign a probability to each alternative without having to sample the error terms. A final choice is then made using a single uniformly distributed random number. We explain how this works through an example. Suppose there are three alternatives: A, B and C. Furthermore, let us suppose the multinomial logit choice model assigns probabilities 0.5, 0.3 and 0.2 to these alternatives, respectively. Then, when the single uniform random number happens to be smaller than 0.5, the choice for alternative A is made, alternative B is chosen when the random number is greater or equal to 0.5 and smaller than 0.8, and otherwise, alternative C will be the choice made.

In travel demand modeling, many choices for a traveling individual may be made this way within a single simulation experiment. To apply CRN in a multinomial logit choice model, one cannot reuse error samples between scenarios as before, since there are no error terms to be sampled anymore. Instead, the variability of results now stems from the uniformly distributed random numbers. Therefore, in multinomial logit choice models, when contemplating the same choice between scenarios, CRN will now be reusing the same uniform random number. For the sole choice mentioned in the previous paragraph, suppose that the uniform number sample would be 0.663, so that alternative B is chosen. Furthermore, suppose that as a result of a change in scenario, the probability of choosing alternative A would increase by 0.1, while the probability of choosing any of the other two alternatives lowers by 0.05 each. In this new scenario, the 'probability boundaries' move from 0.5 and 0.8 to 0.6 and 0.85. CRN would however again use the number 0.663, so that again scenario B would be chosen. By using this principle for every choice generated by a multinomial logit choice model between scenarios, different choices would only be made as a result of the probability boundaries shifting, making sure that different output is caused by the a difference in the nature of the scenario.

2.3. Assessing the impact of CRN

We now detail how we will assess the impact of CRN in this section. In particular, we will regard two scenarios, which we will describe in more detail in Section 3.2, in the metropolitan region Rotterdam-The Hague (MRDH) in The Netherlands. For this pair of scenarios, we record the differences of several indicators in each of the following two experiments, adopting the model parameters from Zhou et al. (2020) (unless specified otherwise):

(1) Base experiment: we run the model 30 times for both scenarios, and we do not apply CRN.
(2) Inclusion of CRN: we run the model 30 times for both scenarios while using the same initial seeds, and thus applying CRN.

Note that in our model, each run covers one day of activities. In both experiments, we opted to observe 30 simulation runs, as this number enables graphical presentation of all outcomes, while at the same time it allows us to make conclusions regarding the effectiveness of CRN. For any considered indicator, each of these experiments thus leads to 30 differences $X_1, \ldots, X_{30}$. These differences are obtained by subtracting the simulated indicator values found under the second scenario from those found under the first scenario. To assess the impact of CRN, we use several statistical measures. We now proceed to describe them, after which we remark on the assumptions required to use these measures and on why these assumptions are satisfied.

Sample variance S_X^2: We consider the sample variance of the differences

$$S_X^2 = \frac{1}{29} \sum_{i=1}^{30} (X_i - \bar{X})^2,$$

where $\bar{X}$ represents the sample mean of the 30 differences. The sample variance, which equals the square of the sample standard deviation S_X of the differences, is a proxy for the reliability of the results. From a mathematical perspective, it is to be expected that CRN leads to a smaller sample variance, since the indicators under both scenarios are now positively correlated. At the same time, the lower the variance of the differences are compared to the sample mean, the more reliable and representative the results are. To make the notion of reliability more precise, we build upon the sample variance to obtain the next two performance measures, in line with e.g. Wunderlich, Vasudevan, and Wang (2019).

Width of the confidence interval: Using basic statistics, assuming for the moment that modeling assumptions are correct, the real expected difference between the indicators corresponding to the two scenarios lies with 95% probability between the numbers $\bar{X} - t_{29,0.975}\, S_X/\sqrt{30}$ and $\bar{X} + t_{29,0.975}\, S_X/\sqrt{30}$, with $t_{29,0.975} \approx 2.045$ representing the 97.5% quantile of the student t-distribution with 29 degrees of freedom. The resulting interval

$$\left[\bar{X} - t_{29,0.975}\frac{S_X}{\sqrt{30}}, \bar{X} + t_{29,0.975}\frac{S_X}{\sqrt{30}}\right] \tag{1}$$

is therefore called the 95% confidence interval (CI). One ideally would like the CI to be as narrow as possible, which is why the width of the 95%-CI, being $2t_{29,0.975}\, S_X/\sqrt{30}$, is a good proxy for the reliability of the results after 30 simulation runs. We expect CRN to result in higher reliability, and thus a smaller confidence interval.

Required number of simulation runs: Another approach to assess reliability would be to pose the question of how many models runs would be minimally required so that the width of the 95%-CI is smaller than a fraction β of the actual expected difference μ. In case the actual variance of the difference between the indicators is given by σ^2, the width of the 95%-CI based on N runs is given by $2q_{0.975}\, \sigma/\sqrt{N}$, where $q_{0.975} \approx 1.96$ is the 97.5% quantile of the normal distribution. We are thus looking for the smallest number N for which $2q_{0.975}\, \sigma/\sqrt{N} \le \beta\mu$, which is given by

$$N_{\min} = \left\lceil \frac{4q_{0.975}^2\sigma^2}{\beta^2\mu^2} \right\rceil.$$

Since both μ and σ^2 are unknown parameters, we use $\bar{X}$ and S_X^2 to estimate these, yielding the formula

$$N_{\min} = \left\lceil \frac{4q_{0.975}^2 S_X^2}{\beta^2\bar{X}^2} \right\rceil. \tag{2}$$

It is easily argued that the lower $N_{\min}$, the lower the computation time that is required to obtain reliable output. In line with earlier statements, we generally anticipate the second experiment with CRN to have a much lower $N_{\min}$ than the first experiment without CRN. In the next section, we choose $\beta = 0.2$ for reporting the number of $N_{\min}$. Thus, the number reported is the minimum number of simulation runs required so that the width of the confidence interval does not exceed 20% of the average of the indicator values found. It is worth noting, however, that, although the value of β will affect the values of $N_{\min}$ themselves, the percentual reduction of $N_{\min}$ as a result of implementing

CRN is insensitive to the actual value of β considered. This is due to the fact that the numbers for both the base experiment and the experiment including CRN are calculated using the same value of β (cf. the denominator of the fraction in (2)).

It should be noted that the three measures discussed above are related. Therefore, in the following analysis, we may not compute every of these measures for every indicator that we consider. For example, when CRN decreases the required number of runs significantly, one can already conclude that the width of the confidence interval for a fixed number of runs will also be considerably smaller. Next to these measures, we will at times also perform a standard double-sided t-test on the differences. More particularly, for several indicators in the second experiment, we may test the null hypothesis that the expectation of the differences between scenarios equals zero. Should the t-test not reject this hypothesis with a confidence level of 95%, we induce that there is not an observed significant difference between the indicators under both scenarios. We will see that even though the implementation of CRN in those cases reduces the width of the 95%-CI and $N_{\min}$ significantly, the computation times remain infeasibly long.

Remark: The use of the measures introduced above requires several assumptions to be justified. For example, to have the sample variance form an unbiased estimator of the actual variance of the distribution of the differences, it should hold that the differences $X_1, \ldots, X_{30}$ are mutually statistically independent. By nature of our experiments, this is evidently the case. Furthermore, each of our experiments is based on 30 differences, so that the law of large numbers implies that the sample mean $\bar{X}$ and sample variance S_X^2 are highly accurate estimations of the actual mean μ and variance σ^2 of the distribution of $X_1, \ldots, X_{30}$. Similarly, to use the confidence interval and $N_{\min}$ as presented in (1) and (2), respectively, it is required that the sample mean $\bar{X}$ is (near-)normally distributed. Since this sample mean is based on 30 different observations, the central limit theorem indeed implies the fact that the distribution of the sample mean $\bar{X}$ is nearly normally distributed with mean μ and variance σ^2. The validity behind the Equations (1) and (2) are immediate consequences of this fact. In the absence of knowledge on the exact values of μ and σ^2, the sample mean $\bar{X}$ and sample variance S_x^2 are used in these equations, because of their known accuracy. For the confidence interval (1), we have opted to use the quantile $t_{29,0.975}$ rather than the corresponding quantile of the normal distribution to make the confidence interval slightly more conservative to offset for the fact that the sample mean $\bar{X}$ is not exactly normally distributed. An exact normal distribution would only be obtained by the sampling of an infinite number of differences, which is obviously infeasible.

3. Experimental results

In this section, we describe the results obtained from doing the two experiments described in the previous section. As mentioned before, to perform these experiments, we apply the model of Zhou et al. (2020) to data for the metropolitan region Rotterdam-The Hague (MRDH) in The Netherlands taken from various sources, which are also described in Zhou et al. (2020). We describe this data in more detail in Section 3.1, and we describe the scenarios considered in Section 3.2. Then, we perform the first two experiments presented in Section 2.3 using several indicators, namely for the average daily number of trips per day and per part of the day (Section 3.3), average travel distance per traveler per day (Section 3.4), modal split and average travel distance per mode (Section 3.5). All numbers reported in this section can also be found in Tables 1 and 2. Note that the output presented in these sections is the result of a series of subsequent choice components, involving both multinomial probit models and multinomial logit choice models. Thus, CRN is applied in each of these components as explained in Section 2.2.

3.1. Input data

The input data, on which the numerical experiments are based, include information on the population and land use as well as level-of-service data on travel times, distances and travel costs for each conceivable origin-destination trip pair in the MRDH area. This area covers the cities of Delft, Pijnacker, Nootdorp and Zoetermeer, being located between the two major cities of Rotterdam and The Hague. These data origin from various sources, which can be found in Zhou et al. (2020).

The population data that we use for this area have been synthesized in Snelder et al. (2021). The dataset consists of 278,698 individuals, making up 131,466 households. For our simulation, we randomly selected 10% of these households to base our simulation on, for the purposes of reducing the computational burden. Since these households are randomly selected, they form an unbiased representation of the population. As a result, hardly any impact is expected on the absolute values of the performance indicators. Within the selected households, 18% of the population is younger than 15 years old, 14% is between 15 (inclusive) and 25 years old, 26% is between 25 (inclusive) and 45, 27% is between 45 (inclusive) and 65 and finally, 15% of the population is older than or equal to 65 years old. For each individual, the synthesized data contains information on the home location, household composition, gender, whether the individual possesses a driving license and/or a student subscription for free public transport, level of education, income, migration background, types of owned bikes and/or vehicles, as well as urbanization level (which specifies the address density in the direct area of the individual's residence).

Concerning land use, the input data contains per traffic analysis zone information on the number of places of employment (offices, shops, etc.), number of education places, area (in m^2) and the average parking costs per hour.

3.2. Scenario descriptions

For the numerical study, we focus on the impact of *Mobility as a Service* on travel demand. To assess this impact, in the experiments we consider the following two scenarios, differing in terms of the adoption level of MaaS:

(1) In the first scenario, to be referred to as 'partial MaaS', 10% of the people younger than 15 or older than 65 have a MaaS subscription, while 20% of the remaining population also has a MaaS subscription. As a result, 16.5% of the overall population has a MaaS subscription.
(2) In the second scenario, 'full MaaS', 100% of the population has a MaaS subscription.

Table 1. Sample means and sample variances of the indicator differences between both scenarios.

Name	No-CRN Sample mean of the difference	CRN Sample mean of the difference	No-CRN Sample variance of the difference	CRN Sample variance of the difference
Daily number of trips	0.007	0.012	2.7×10^{-4}	8.8×10^{-6}
Daily number of trips <15 years	0.047	0.0382	1.3×10^{-2}	1.6×10^{-5}
Daily number of trips 15–25 years	0.009	0.0152	2.6×10^{-3}	2.8×10^{-5}
Daily number of trips 25–45 years	−0.01	0.0087	9.6×10^{-4}	3.7×10^{-5}
Daily number of trips 45–65 years	−0.0008	0.00086	1.2×10^{-3}	2.2×10^{-5}
Daily number of trips >= 65 years	0.0035	0.0055	1.4×10^{-3}	4.9×10^{-5}
Total trips in the morning peak	20	10	25,856	442
Total trips in the evening peak	15	29	31,014	1137
Total trips in the rest of day	161	297	153,169	5304
Travel distance per person	1.70	1.74	0.067	0.006
Travel distance <15 years	3.73	3.54	0.035	0.044
Travel distance 15–25 years	1.42	1.21	0.488	0.0156
Travel distance 25–45 years	1.25	1.57	0.333	0.0201
Travel distance 45–65 years	1.17	1.23	0.404	0.0299
Travel distance >= 65 years	1.35	1.35	0.287	0.0286
Travel distance by bike	0.29	0.29	0.001	0.0004
Travel distance by car	−0.38	−0.34	0.027	0.019
Travel distance by car-passenger	−1.66	−1.67	0.046	0.015
Travel distance by DRT	−4.1	−4.0	0.096	0.046
Travel distance by ebike	−0.67	−0.67	0.002	0.002
Travel distance by walk	0.14	0.14	0.0002	0.0001
Travel distance by public transport	−2.1	−2.1	0.269	0.063
Travel distance by multimodal modes	0.59	0.59	0.116	0.09
Trips by bike	−553	−493	96,503	39,332
Trips by car	−6766	−6706	38,698	16,885
Trips by car-passenger	−1815	−1820	11,781	6405
Trips by DRT	−2193	−2206	10,582	6221
Trips by ebike	2901	2934	12,208	8155
Trips by walk	1852	1854	31,704	12,617
Trips by public transport	466	475	11,319	3190
Trips by multimodal modes	6305	6298	15,261	6786

The main goal is to quantitatively assess the difference between both scenarios in terms of several indicators.

Table 2. Confidence interval widths of the indicator differences between both scenarios, as well as the required number of runs N_{min} to obtain reliable estimates.

Name	No-CRN Width of CI	CRN Width of CI	No-CRN N_{min}	CRN N_{min}
Daily number of trips	0.012	0.0022	2042	23
Daily number of trips <15 years	0.027	0.003	221	4
Daily number of trips 15–25 years	0.038	0.004	12384	47
Daily number of trips 25–45 years	0.023	0.0045	3383	188
Daily number of trips 45–65 years	0.026	0.0035	698,551	11,487
Daily number of trips >= 65 years	0.028	0.0052	42,613	629
Total trips in the morning peak	120	15.7	24,999	1780
Total trips in the evening peak	132	25.2	51,343	510
Total trips in the rest of day	292	54.4	2278	23
Travel distance per person	0.193	0.057	9	1
Travel distance <15 years	0.441	0.157	10	1
Travel distance 15–25 years	0.522	0.093	93	4
Travel distance 25–45 years	0.431	0.106	82	3
Travel distance 45–65 years	0.474	0.129	114	8
Travel distance >= 65 years	0.400	0.126	60	6
Travel distance by bike	0.02	0.02	5	2
Travel distance by car	0.12	0.1	74	65
Travel distance by car-passenger	0.16	0.09	7	2
Travel distance by DRT	0.23	0.16	3	2
Travel distance by ebike	0.04	0.03	3	2
Travel distance by walk	0.01	0.01	4	3
Travel distance by public transport	0.39	0.19	24	6
Travel distance by multimodal modes	0.25	0.22	127	101
Trips by bike	232	148	121	62
Trips by car	147	97	1	1
Trips by car-passenger	81	60	2	1
Trips by DRT	77	59	1	1
Trips by ebike	83	67	1	1
Trips by walk	133	84	4	2
Trips by public transport	79	42	20	6
Trips by multimodal modes	92	62	1	1

3.3. Average daily number of trips

The first indicator that we study is the average number of trips undertaken by a traveler during a day, which illustrates the impact of CRN particularly well. We first consider the number of daily trips for all age classes mentioned in Section 3.1 together. We then also regard results at a more detailed level, which distinguishes between the different age classes of the population.

3.3.1. Daily number of trips for all age classes combined

The study of the average number of trips undertaken by a traveler during a day turns out to illustrate the beneficial effects of CRN remarkably well. Figure 1(a) presents the average numbers of daily trips undertaken by a traveler in 30 different simulation runs for each of the two scenarios independently, using the base experiment where no CRN is applied. Although the mean value of the blue plot (i.e.partial MaaS) appears to be smaller than that of the orange plot (full MaaS), basing firm conclusions on this figure would be hard. Indeed, the sample variance of the difference of the left-hand figure, having value 0.00027, is still significant compared to the average of these differences, bearing the value of 0.007. In Figure 1(b), we plot the same quantities based on the same number of runs, but this time we *do* apply CRN, and thus introduce positive correlation. We consider the same simulation output as before for partial MaaS (and hence

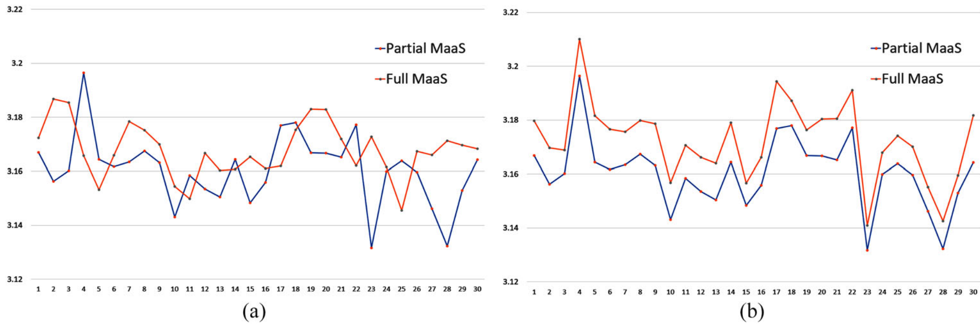

Figure 1. Average daily numbers of trips per traveler in 30 simulation runs. (a) Base experiment and (b) Inclusion of CRN.

obtain the same blue plot), but unlike before, we use the exact same error samples for the simulation of the average number of daily trips for full MaaS, leading to a different orange plot. The effect is clear: the level of variability observed in both plots is similar to that of the base experiment, but the plots move almost parallel and do not intersect anymore as a result of the synchronization. Specifically, the figure shows that the differences between the simulated number of trips are roughly constant. This is also illustrated by the fact that, although the average difference in the right-hand picture is 0.012 (which is similar to the average in the left-hand picture as expected), the sample variance of the differences in the right-hand picture now is only around the order of 8.8×10^{-6}, confirming the fact that CRN offers a lot more reliability.

Figure 1 confirms that by applying CRN the variability of simulated indicator value differences reduces dramatically, as these differences can now solely arise as a result of different scenarios. Therefore, one can now much more comfortably draw the conclusion that MaaS will probably lead to more trips per day per traveler. Indeed, if we deploy a *t*-test as explained in Section 2.3 on the 30 differences of the values plotted on Figure 1(b), it rejects the null hypothesis that the expected number of trips per traveler per day are equal under both scenarios. This supports such a conclusion, which can be explained by the fact that a MaaS subscription makes traveling more accessible and flexible, and therefore more attractive. If we were to deploy a *t*-test on the 30 differences of Figure 1(a), however, the *t*-test would not reject, which is in line with the fact that Figure 1 (b) sketches a more reliable picture.

To quantify the impact of CRN in practical terms, computation of $N_{\min}$ as provided in Section 2.3 leads to 2042 runs for the base experiment, while the inclusion of CRN brings this number down to 23. This means that, to obtain a confidence interval for the indicator difference that has a width less than 20% of the actual expected difference, the minimally required number of runs (and thus the minimally required computation time) is reduced by almost 99% when using CRN. Again, this shows the beneficial impact of CRN on numerical experiments.

3.3.2. Daily number of trips per age class

We now regard the daily number of traveler trips for all different age classes separately. The age class covering travelers between 15 and 25 years old sketches a similar picture as that of the aggregated level combining all age classes. That is, in the absence of CRN, it is hard to identify the influence of the varying natures of the scenarios on the difference of the average number of trips undertaken by a traveler based on 30 simulation runs, as shown in Figure 2(c).

In contrast, Figure 2(d), which includes CRN, again clearly shows that the possession of a MaaS subscription increases the traveling activity of young travelers. This is probably because MaaS provides convenient travel modes to do more activities. We obtain similar findings when regarding the age classes representing travelers younger than 15 years old, travelers between 25 and 45 years old, and travelers older than 65 years old, see Figure 2 (a,b,e,f,i,j). For each of these age classes, the double-sided t-test applied on the 30 observed differences rejects the null hypothesis that the expected difference in the number of trips undertaken by a traveler equals zero, and $N_{\min}$ in all of these four age classes is reduced by over 94% due to the use of CRN.

The age class concerning the population between 45 and 65 years old, however, shows an additional effect. For this case, Figure 2(g) again shows much more variability in the differences than Figure 2(h), in which the two plots are rather parallel to one another. However, even in the latter figure, the plots intersect a lot, so that no reliable conclusions can be drawn on which of the two scenarios leads to more trips in this age class. Indeed, the sample variance of the observed differences equals 0.000022, while the average of these differences only equals 0.00086 (see Table 1). As a result, the variability of the differences based on these 30 runs is still too large to produce a reliable conclusion. These findings are confirmed by performing a t-test, which does not at all reject the null hypothesis that there is no difference in indicator value between the two scenarios.

This is not to say that CRN in this case bears no effect. Indeed, the width of the 95%-CI based on the 30 runs reduces from 0.026 to 0.0035. While the latter number is still very large (relative to the quantity we wish to estimate, that is), the reduction is significant. Similarly, we find that by implementation of CRN, $N_{\min}$ is reduced from 698,551 to 11,487, yielding a decrease of required runs of about 98%. Yet, performing 11,487 simulation runs is prohibitive in terms of computation time. The conclusion to be drawn here is that CRN still is very effective, but that, at the same time, the actual difference between scenarios in this case is so small, that the effect of using CRN is too small to allow for a feasible computation time. This being said, one may wonder how important this unfeasible simulation would be, as the conclusion could be drawn that the difference between the two scenarios for this age class is negligible anyway. Intuitively, this is not surprising when considering the fact that most activities undertaken by these individuals are mandatory, since relatively a large number of trips undertaken by this age class is work-related. In other words, having a MaaS subscription may alter the travel mode used, but in the absence of such a subscription, no trips will be dropped in this age class.

Other related indicators are the daily number of trips during the morning peak, evening peak and the rest of the day. These are reported on in Tables 1 and 2 and also show a significant effect of CRN on required runtime.

3.4. Average travel distance per traveler

The next indicator that we consider is the average travel distance covered by a traveler within a day in kilometers. When performing the base experiment and the experiment

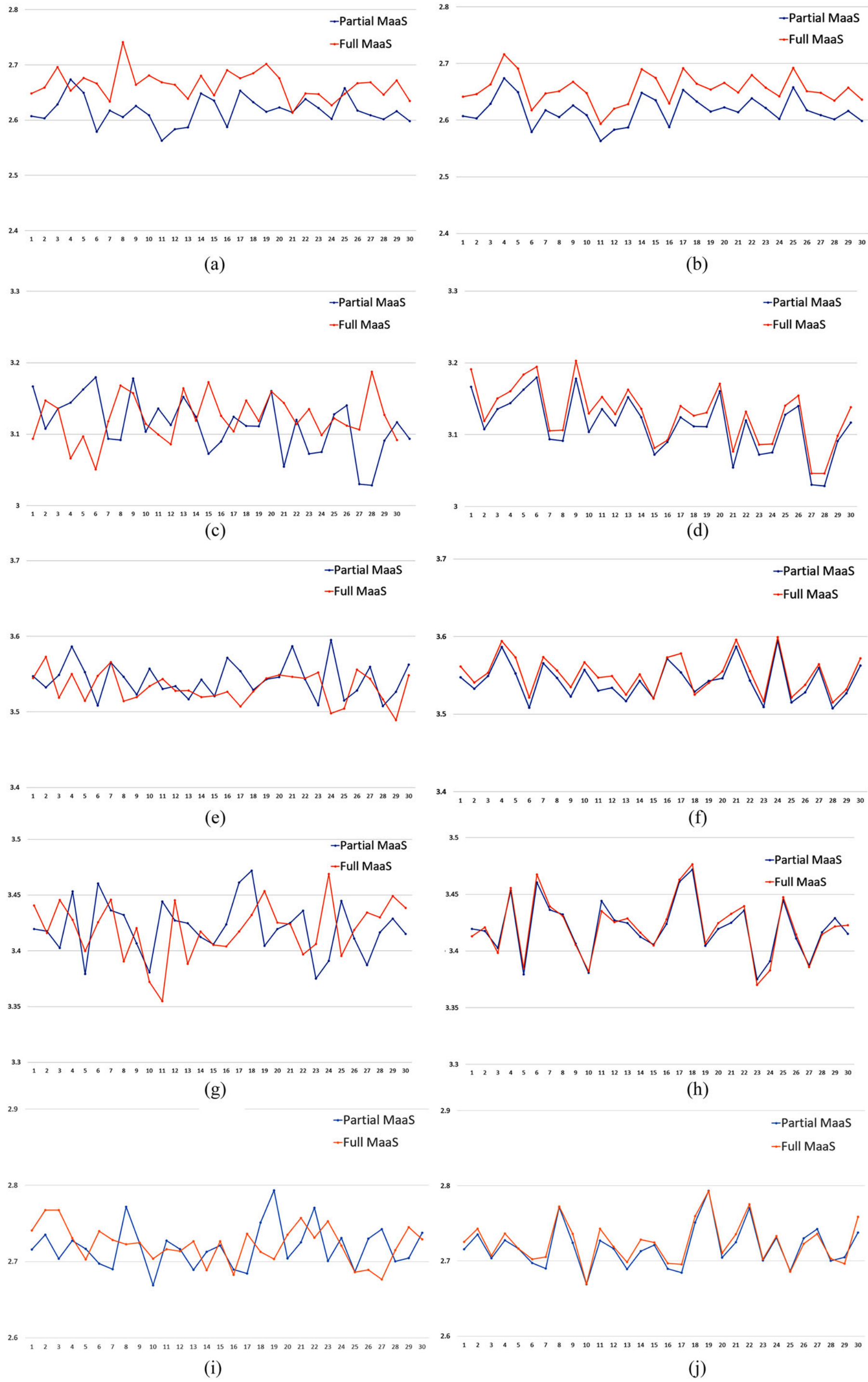

Figure 2. Average daily numbers of trips per traveler with an age $<$ 15 years ((a) and (b)), $\geq$ 15 and $<$ 25 years ((c) and (d)), $\geq$ 25 and $<$ 45 years ((e) and (f)), $\geq$ 45 and $<$ 65 years ((g) and (h)), and $\geq$ 65 years ((i) and (j)) in 30 simulation runs. (a) Base experiment. (b) Inclusion of CRN. (c) Base experiment. (d) Inclusion of CRN. (e) Base experiment. (f) Inclusion of CRN. (g) Base experiment. (h) Inclusion of CRN. (i) Base experiment and (j) Inclusion of CRN.

including CRN for all travelers, we obtain Figure 3. Again, the effect of CRN is clear as the differences in the right-hand picture are more constant than in the left-hand picture, and thus more firm conclusions can be drawn. However, one may argue that also based on the left-hand picture, although the blue and orange plot are less parallel than the right-hand picture, one can comfortably conclude that the more MaaS is available to travelers, the more they will travel in terms of distance covered.

The fact that the left-hand picture already is rather informative is due to the fact that the difference between the two scenarios is relatively large. In fact, it is so large, that the variability of the differences in the left-hand picture (the sample variance of the differences is 0.067) does not outweigh the actual difference (the sample mean of the differences is 1.7. Still, CRN has a substantial effect on computational complexity in this case. The widths of the 95%-CIs of the differences observed for the 30 runs of Figure 3(a,b) are 0.193 and 0.057, respectively, the latter thus offering a much higher level of reliability. Similarly, the value $N_{\min}$ for the base experiment would be 9, while this number equals 1 for the experiment which incorporates CRN. Therefore, although one may argue that CRN is not needed for this indicator, still a substantial amount of computation time can be saved when doing batch experiments. Tables 1 and 2 also report on the travel distance per traveler in the separate age classes considered in Section 3.3.2. These separate age classes show similar effects as the aggregate observations mentioned here.

3.5. Modal split

Finally, we inspect the so-called 'modal split' for travelers used. That is, we regard the (differences in) the percentage of trips that are undertaken by each mode (between scenarios). As mentioned in Section 2.2, next to the seven unimodal modes (walk, bike, e-bike, car, car-passenger, demand-responsive transport and public transport), we also consider 25 multimodal mode alternatives made up of a combination of single modes. In Figure 4 differences in mode use between the partial MaaS scenario and the full MaaS scenario are plotted. Each bar corresponds to one of the 30 simulation runs and represents the difference observed in the percentage of trips undertaken with the corresponding mode ('full MaaS minus partial MaaS'). For the purposes of this section, we

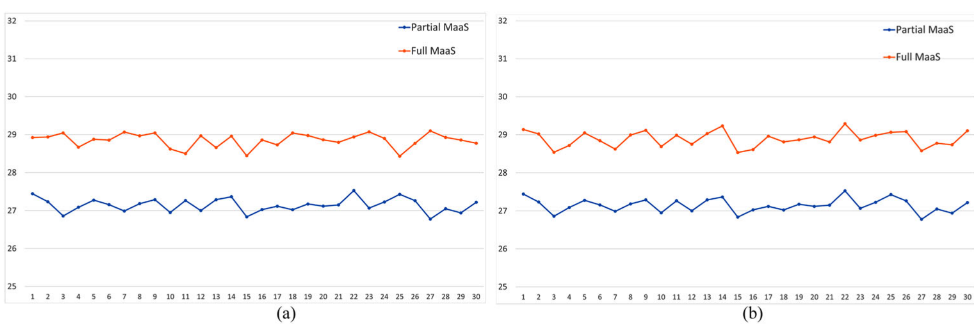

Figure 3. Average daily travel distance per traveler in kilometers in 30 simulation runs. (a) Base experiment and (b) Inclusion of CRN.

have combined all 25 multimodal modes in a single category 'multimodal'. Figure 4(a) shows the results observed in the base experiment, while Figure 4(b) shows them for the experiment including CRN.

As expected, the figures show that when moving from a partial MaaS to a full MaaS scenario, the e-bike and multimodal modes gain in popularity at the cost of the car modes and demand-responsive transport. In particular, in Figure 4(b), the share of the car mode dropped by 7.8% on average, while the e-bike was used in an additional 3.4% of the trips. Moreover, in an additional 7.3% of the trips on average, multimodal modes were chosen. Also the bike mode became less popular, which can be explained as follows. When travelers for example choose to use a personal bike for an outbound trip of a tour, typically it also needs to be chosen for an inbound trip. However, with access to MaaS, many more mode options are considered for use in each of the trips. As a result, if one of those modes is much more favorable than the bike for an inbound trip, then this also has automatic ramifications for the mode choice of the outbound trip, making the choice for the bike less obvious. The modes of walking and public transport do not have this issue, making them slightly more popular at the expense of the bike. The astute reader, however, will note that the effects impacting the bike mode should also impact the e-bike mode. However, for the e-bike mode, these effects are offset by the fact that a MaaS subscription makes the e-bike more popular.

What is intriguing is that Figure 4(a,b) hardly present mutually differing pictures, like in the previous sections. Indeed, in both the base experiment and the experiment including CRN, $N_{\min}$ is very small, almost never exceeding ten simulation runs for any mode. The only exception to this is the bike mode in the base experiment, for which $N_{\min} = 121$ which is indeed confirmed by the fact that the 'bike bars' in Figure 4(a) show a lot of variability.

The reason for this is that in this case variability of results is not so much an issue as it was in Sections 3.3 and 3.4. As mentioned in Section 2.2, mode choice is dictated by a multinomial probit choice model, and in many of these modes, the error terms only have a small impact on the ultimate mode choice. That is, the variability caused by the error terms in general is a lot smaller than the actual difference in 'observed utility' between the modes. As a result, the variability of the error terms do not have a lot of

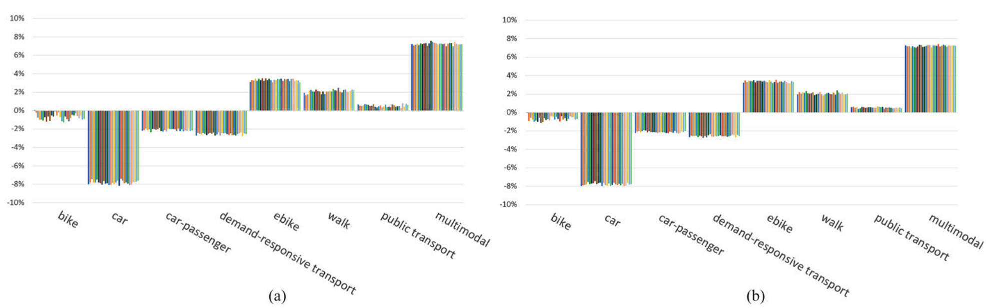

Figure 4. The difference in modal split between two scenarios: a positive value means that mode use is higher for the full Maas scenario than for the partial MaaS scenario. Each bar corresponds to one of the 30 simulation runs and represents the difference observed in the percentage of trips using the corresponding mode. (a) Base experiment and (b) Inclusion of CRN.

impact on e.g. whether or not the car is chosen. Since a similar effect also plays for the full MaaS scenario, the variability of the random number generation cannot trigger large fluctuations in terms of which modes are chosen. As a result, one cannot hope for a large impact of CRN in terms of $N_{\min}$ either. An exception to this is the bike mode, where the variability of the errors actually does impact the utilities of the bike modes enough to cause changes in the ultimate mode choice. Since variability is thus more of an issue here, CRN indeed has a larger impact.

In Tables 1 and 2, we also report on the difference in average travel distance per travel mode between the two scenarios. These show similar impacts of CRN as the modal split.

In summary, these findings show that when variability of pseudo-random number generation is not a problem, there is not much to solve for CRN either. As a result, the use of CRN overall does not have as much a dramatic impact as it had in the previous sections. That said, this does not mean that CRN does not have any merit at all. For example, the width of the computed 95%-CI in the percentage difference of trips that are taken with the car is reduced by 33.9% after implementation of CRN, which adds to the reliability of simulation results. As a final remark considering the modal split, it is reasonable that one would argue that, for the purposes of correct modeling, the variance parameters of the normal distributions underlying the error terms should have been estimated differently (i.e. taken to be a bit larger), so that error terms will have a larger influence on the mode choice. In that case, the variability of the errors will increase, and as a result, we expect the impact of CRN to be larger as well.

4. Conclusions and discussion

In this paper, we have considered the use of CRN in the context of activity-based travel demand modeling. More particularly, in this paper we showed how this technique can be applied to effectively simulate the impact of scenario changes to any indicator. By making sure that between scenarios the exact same pseudo-random numbers are used for the same purposes, any difference in output can in all likelihood be attributed to the change in scenarios. As a result, the variability of the simulated differences is much smaller, and therefore less simulation runs and less computation time is required to perform a simulation study with a meaningful output.

After we explained how CRN can be implemented in an activity-based travel demand model, we set out to demonstrate the potential of CRN in practice by studying the impact of MaaS on travel demand in the Metropolitan region Rotterdam-The Hague. In particular, we studied the question of whether a scenario in which the complete population has a MaaS subscription yields different indicators than the scenario in which only a small part of the population has such a subscription. We found that the implementation of CRN yields a computationally very efficient tool to answer this question in the affirmative: when everybody has a MaaS subscription, travelers will travel more, and use different (possibly multimodal) modes. That is, by implementing CRN in all components, such conclusions can be reliably drawn using up to 99% less simulation runs than the conventional setting without CRN. In a similar vein, when we compute the 95%-CI of a certain indicator difference based on a fixed number of runs, a simulation setting with CRN typically comes with a much smaller interval than a setting without. This shows that CRN can shorten the computation times required drastically, especially when considering the complete sequence of choice models/components in an ABM. It should be noted that it is not guaranteed that every scenario and indicator may lead to similar (or higher) amounts of speedup as (than) reported in this paper. Even in those cases, however, it is beneficial to implement CRN, because it will always lead to some amount of speedup, while the implementation efforts required are low.

By comparing various indicators, we also found that CRN is especially worthy of implementation when the differences between scenarios are moderate. In case they are very large, CRN still reduces computational complexity, but conventional simulation methods already can provide conclusions in reasonable time. In contrast, when differences are very small so that they are hardly observable, even though CRN reduces computational complexity also in this case, the setting with CRN still requires too many runs in order to decide which of the scenarios scores better. In such a setting, however, one can say that the difference in the nature of the scenarios hardly has an impact on the indicators.

Finally, we showed that the impact of CRN is related to the degree at which the variability of generated random numbers can cause random fluctuations in the model output. We saw that when this cannot occur to a large extent, CRN cannot be expected to yield up to 99% less required simulation runs as reported above.

In general, while in theory reliable results can still be had without the implementation of CRN by making sure that enough runs are performed, in practice the number of runs required may be practically infeasible. As can be deduced from Table 2, CRN may bring down the number of required runs to a manageable level, therefore enabling computational studies which might otherwise be impossible to perform. Since the implementational effort required for CRN is modest, and the required number of runs is guaranteed to come down, we therefore advocate the use of CRN when interested in indicator changes as a result from adopting different scenarios.

We now discuss opportunities for further research. Some of these opportunities concern the scenarios considered in this study. That is, we have considered the impact of CRN based on two scenarios that mainly differed in the penetration rate of MaaS subscription. The actual penetration rates considered, as well as other parameters, are chosen to reflect current and future reality as much as possible and are based on expert judgement. However, the scenarios might possibly be tuned to match reality even better, which is a topic for further study. Moreover, it would be possible to choose other scenarios to study the impact of CRN. That is, while CRN is guaranteed to lead to some level of speedup, the question of which characteristics of a scenario enable the obtained speedup levels of this paper and beyond, is another fruitful direction to pursue. In our subsequent work (Zhou et al. 2022), which is entirely devoted to a case study in the MRDH region, CRN is also used in other scenarios for other indicators having different measures such as different parking price, in-vehicle time of a specified travel mode.

Furthermore, it would be worth studying the potential of other variance reduction techniques developed in the stochastic simulation literature in the context of travel modeling. For example, we believe that the results in this paper may be further improved by implementing advanced sampling techniques such as hypercube sampling.

Next, we mention that, in this study, we have mostly focused on reducing simulation error by the introduction of CRN. However, there also exist other sources of error. That is, the results obtained in this study for the MRDH region may be biased as a result of for instance inaccuracies in the estimation of certain parameters and the used structure for the utility functions, even though they have been selected according to expert judgement. While these inaccuracies in principle do not impact the conclusion of this paper regarding the impact of CRN, they yield many opportunities for further research. In this regard, it is also important to refer to our follow-up work (Zhou et al. 2022), which is devoted to a case study in the MRDH region.

The final venue of further research that we highlight can be found in the direction of travel assignment models. Since this study only focused on the travel demand, any change of level-of-service output such as travel time of trips is not considered. This would require a connection of the current travel demand model with a travel assignment model. When this connection would be made, the impact of CRN on the whole model chain could be considered. Given the results in this paper, we would expect that CRN also has great potential when considering the complete model chain.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the NWO Gravitation programme NETWORKS [grant number 024.002.003] and TNO, the Netherlands Organisation for Applied Scientific Research (nederlandse organisatie voor wetenschappelijk onderzoek).

ORCID

H. Zhou http://orcid.org/0000-0001-8723-1060

References

Bao, Q., B. Kochan, T. Bellemans, D. Janssens, and G. Wets. 2015. "Investigating Micro-Simulation Error in Activity-Based Travel Demand Forecasting: A Case Study of the FEATHERS Framework." *Transportation Planning and Technology* 38 (4): 425–441. doi:10.1080/03081060.2015.1026102.

Bekhor, S., L. Kheifits, and M. Sorani. 2014. "Stability Analysis of Activity-Based Models: Case Study of the Tel Aviv Transportation Model." *European Journal of Transport and Infrastructure Research* 14 (4): 311–331. doi:10.18757/EJTIR.2014.14.4.3039.

Castiglione, J., J. Freedman, and M. Bradley. 2003. "Systematic Investigation of Variability Due to Random Simulation Error in an Activity-Based Microsimulation Forecasting Model." *Transportation Research Record: Journal of the Transportation Research Board* 1831 (1): 76–88. doi:10.3141/1831-09.

Gali, E., S. Eidenbenz, S. Mniszewski, L. Cuellar, and C. Teuscher. 2008. "ActivitySim: Large-Scale Agent Based Activity Generation for Infrastructure Simulation."

Glasserman, P., and D. D. Yao. 1992. "Some Guidelines and Guarantees for Common Random Numbers." *Management Science* 38 (6): 884–908. doi:10.1287/mnsc.38.6.884.

Horni, A., D. Charypar, and K. W. Axhausen. 2011. "Variability in Transport Microsimulations Investigated With the Multi-Agent Transport Simulation MATSim." *Arbeitsberichte Verkehrs-und Raumplanung* 692. doi:10.3929/ethz-a-006686329.

Horni, A., K. Nagel, and K. W. Axhausen. 2016. *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press. doi:10.5334/baw.

Jittrapirom, P., V. Caiati, A. M. Feneri, S. Ebrahimigharehbaghi, M. J. A. González, and J. Narayan. 2017. "Mobility as a Service: A Critical Review of Definitions, Assessments of Schemes, and Key Challenges." *Urban Planning* 2 (2): 13–25. doi:10.17645/up.v2i2.931.

Pinjari, A., N. Eluru, S. Srinivasan, J. Guo, R. Copperman, I. Sener, and C. Bhat. 2008. "CEMDAP: Modeling and Microsimulation Frameworks, Software Development, and Verification." In *Proceeding of the Transportation Research Board 87th Annual Meeting*, Washington, DC.

Rasouli, S., and H. Timmermans. 2012. "Uncertainty in Travel Demand Forecasting Models: Literature Review and Research Agenda." *Transportation Letters* 4 (1): 55–73. doi:10.3328/TL.2012.04.01.55-73.

Ross, S. M. 2013. *Simulation: Fifth Edition*. New York: Elsevier.

Snelder, M., Y. Araghi, B. Ashari, E. Charoniti, G. Klunder, R. Sterkenburg, and M. Van der Tuin, et al. 2021. "Rapport A: Methode Urban Tools Next II -- toelichting op gekozen aanpak voor parkeren, ketens en hubs, nieuwe mobiliteitsconcepten. Den Haag (The Netherlands)." https://publications.tno.nl/publication/34639979/i0dl3z/TNO-2021-R10644.pdf.

Veldhuisen, J., H. Timmermans, and L. Kapoen. 2000. "Microsimulation Model of Activity-Travel Patterns and Traffic Flows: Specification, Validation Tests, and Monte Carlo Error." *Transportation Research Record* 1706 (1): 126–135. doi:10.3141/1706-15.

Vovsha, P., R. Donnelly, and S. Gupta. 2008. "Network Equilibrium with Activity-Based Microsimulation Models: The New York Experience." *Transportation Research Record* 2054 (1): 102–109. doi:10.3141/2054-12.

Van de Werken A. 2018. *Verkeersmodel MRDH 2.0 Technische rapportage*. 001594.20181026.R1.02, Deventer (The Netherlands). https://mrdh.nl/project/verkeersmodel.

Wunderlich, K. E., M. Vasudevan, and P. Wang. 2019. *TAT Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software 2019 Update to the 2004 Version*. FHWA-HOP-18-036.

Zhou, H., J. Dorsman, M. Mandjes, and M. Snelder. 2020. "A Tour-Based Multimodal Mode Choice Model for Impact Assessment of New Mobility Concepts and Mobility as a Service." Ph.D. thesis. https://publications.tno.nl/publication/34639987/crS65z/zhou-2022-tour-based.pdf.

Zhou, H., J. Dorsman, M. Mandjes, and M. Snelder. 2022. "Sustainable Mobility Strategies and Their Impact: A Case Study Using a Multimodal Activity Based Model." Ph.D. thesis. https://publications.tno.nl/publication/34639988/iDbCfO/zhou-2022-sustainable.pdf.