

Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning

Matthew Wigginton Conway

Conveyal & School of Geographical Sciences and Urban Planning, Arizona State University

Andrew Byrd **Michael van Eggermond**

Conveyal Future Cities Laboratory, Singapore-ETH Center

Abstract: Accessibility is increasingly used as a metric when evaluating changes to public transport systems. Transit travel times contain variation depending on when one departs relative to when a transit vehicle arrives, and how well transfers are coordinated given a particular timetable. In addition, there is necessarily uncertainty in the value of the accessibility metric during sketch planning processes, due to scenarios which are underspecified because detailed schedule information is not yet available. This article presents a method to extend the concept of “reliable” accessibility to transit to address the first issue, and create confidence intervals and hypothesis tests to address the second.

Article History:

Received: October 31, 2016

Accepted: August 30, 2017

Available online: July 23, 2018

1 Introduction

Accessibility is becoming a popular metric in public transport planning exercises (e.g., Palmateer, Owen, & Levinson, 2016; APPM Management Consultants & Goudappel Coffeng, 2016). One of the challenges with using accessibility in public transport sketch planning is variation and uncertainty. Transit travel times can vary significantly depending on the exact departure time (due to waiting for transit vehicles at the first stop or at transfers). The literature on “reliable” travel time and accessibility metrics is relatively well developed, but there is little research extending this concept to the public transport context, which is surprising given the magnitude and regularity of variation in transit travel time.

In addition, rapid-turnaround sketch planning and geodesign activities are becoming prevalent in planning (e.g., Walker, 2010; Stewart & Zegras, 2016). These types of exercises present a number of challenges for the use of accessibility. In sketch planning, scenarios are often underspecified, with quick, rough plans rather than detailed descriptions. Furthermore, sketch planning requires computations to be performed very rapidly, with results needed in seconds or (at worst) minutes, limiting the complexity of the algorithms used. This paper presents a technique to address both the need for “reliable” metrics which account for variation, and a statistical approach to ensure that these results are not simply due to random error. These techniques are used in Conveyal’s sketch planning system

Copyright 2018 Matthew Wigginton Conway, Andrew Byrd and Michael van Eggermond.

doi: [10.5198/jtl.u.2018.1074](https://doi.org/10.5198/jtl.u.2018.1074)

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – NonCommercial License 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

The Journal of Transport and Land Use is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies. This paper is also published with additional sponsorship from WSTLUR.

to provide near-real-time feedback on transport network changes. A case study of an expansion of the Singapore Mass Rapid Transit (MRT) system is presented in the penultimate section of this paper.

2 Literature Review

Including variation and uncertainty in the definition of accessibility is not a new phenomenon. Hall (1983) theorized that modes with less certain travel times cause their users to allow a “safety margin” on top of the expectation of travel time, which increases the effective travel time. Owen and Levinson (2014) and Farber and Fu (2017) have taken the variation in public transit travel times into account by computing the average number of opportunities reachable given random departure during a time window (for instance, the morning peak commute). Farber, Morang and Widener (2014) use the percent of the day that supermarkets are accessible within a given travel time as their metric, thus creating a measure not of travel time but of variation.

Chen et al. (2013, 2017) develop reliable metrics for automobile travel time and accessibility given traffic variation, estimating the accessibility given a desired probability of on-time arrival. For example, one might want to ensure that they arrive on time 90 percent of the time, given stochastic variations in travel time. In sketch planning, we frequently do not know the exact schedule of new or modified lines, but only their frequencies, making it difficult to quantify waiting time at the start of the trip and at transfers. A common approach to dealing with this, which can be termed *half-headway*, involves assuming that the waiting time for any boarding, either at the start of the trip or after a transfer, involves waiting for half of the headway of the vehicle (e.g., Farber & Grandez, 2017).

However, the half-headway method cannot handle the case of multiple competing routes. There may be multiple routes running along the same street (or even along parallel streets), and a clever system user will not wait an average of half the headway of any individual route, but will instead take the first vehicle that arrives on any route. Additionally, the travel times reported to each location are likely not all possible simultaneously; most likely it is not possible to schedule the network such that every transfer wait is exactly half of the headway of a line. Instead, we adopt the method of calculating accessibility metrics given sketch plans of public transport systems described in Conway, Byrd and van der Linden (2017). This method uses the full timetables of the existing transit system, and thus has information about the transfer performance and existence of competing lines. When there are lines that do not have timetables present in the scenario, a Monte Carlo approach is taken, generating many possible timetables from the set of all timetables that could exist given the description of the network. Thus, the full distribution of travel times is explored, including the effects of competing lines.

Owen and Levinson (2014) and Farber and Fu (2017) include travel time variation over a possible departure time window in their accessibility results by computing the travel time given departure at each minute during the time window. The same approach is used in Conway, Byrd and van der Linden (2017), so therefore the output of our algorithm is travel times given departure at each minute during the time window.

Given travel times, or in our case a list of travel times, from each origin to each destination, we must create a single accessibility number for each origin. There are numerous ways to summarize travel times into a single accessibility metric (for a summary, see Geurs & van Wee, 2004/2006 and El-Geneidy & Levinson, 2006). Perhaps the simplest is the cumulative accessibility metric, which simply counts the number of opportunities within a particular travel time cutoff. More complex formulations are also possible, giving more distant opportunities a lower but non-zero weight in the metric relative to nearby opportunities.

When computing accessibility measures for public transport systems, we are faced with the problem of not having a single travel time for each O-D pair, but rather having many possible travel times depending on exactly when you depart. Owen and Levinson (2014) address this by computing a cumulative accessibility metric at each possible minute of departure, and taking the arithmetic mean of

those values. We term this metric *average instantaneous accessibility*: instantaneous because the accessibility is calculated independently at each instant, and average because those values are averaged.

3 Computing Accessibility

Our method consists of several steps. We first compute travel times to all destinations, accounting for the uncertainty in proposed transit lines that have not yet received schedules. We then choose a travel time cutoff and a desired level of reliability for the accessibility metric. Finally, we evaluate uncertainty in our estimates using a bootstrapping process. This process is depicted visually in Figure 1, and is described in the following sections.

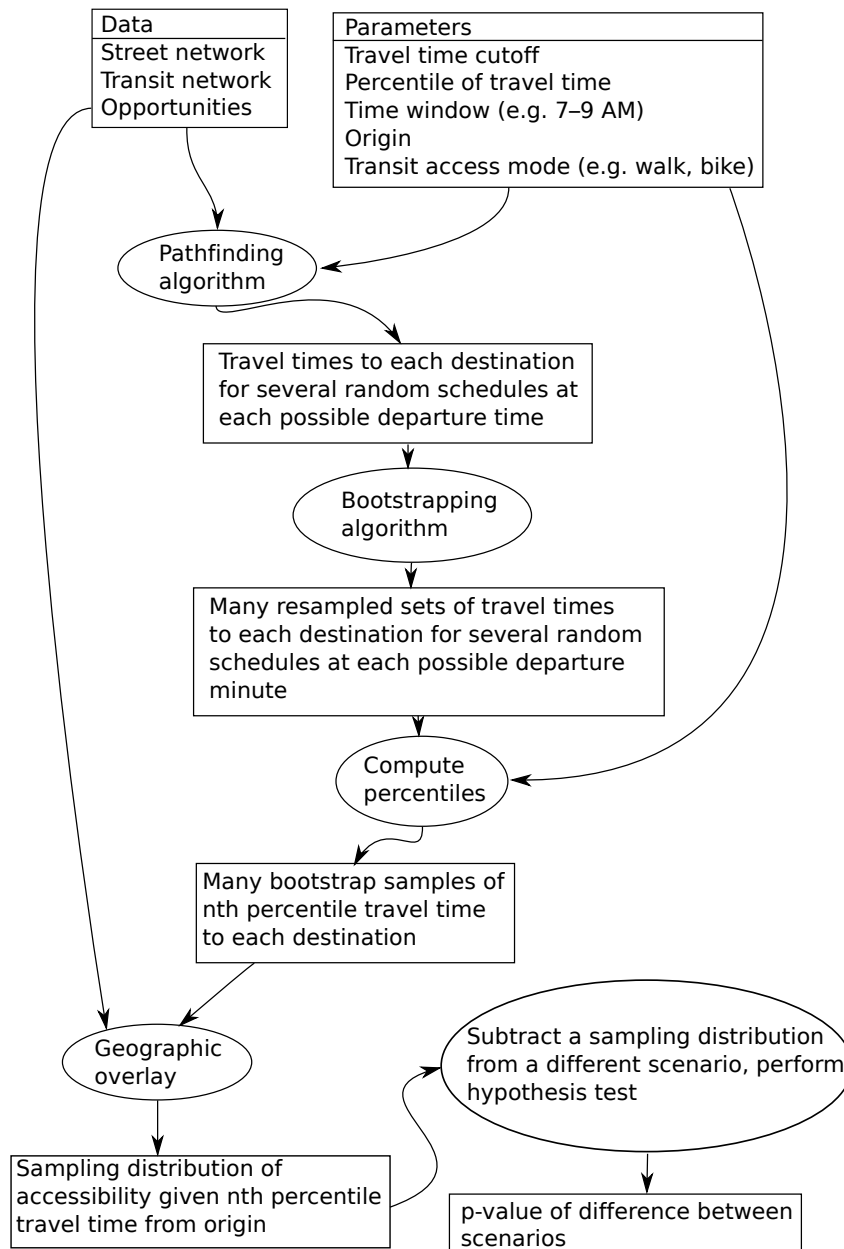


Figure 1: Process diagram of our accessibility computation framework

For ease of interpretation, we have chosen to use a cumulative accessibility metric; it is far easier to grasp the idea of n opportunities being within a certain travel time, rather than a weighted average

metric of many opportunities which do not count equally. This is a particular advantage in the sketch planning context, where the results likely need to be communicated to non-experts. The methods for quantifying reliability and uncertainty described in this paper, however, depend only on travel time and can be applied to any accessibility metric that takes as input an O-D matrix of travel times.

In order to compute a cumulative accessibility metric, we need to summarize our list of travel times for all departure times and a random sample of possible timetables to a single travel time for each O-D pair. Using the aforementioned average instantaneous accessibility is attractive for a number of reasons. As the mean of many instantaneous accessibility values, there are no issues with undefined or infinite data; accessibility values are defined and finite at every location and every departure time (although they may be zero). Other metrics that instead take the mean of travel time (for instance, that used in the case study section of Conway, Byrd, & van der Linden, 2017) have issues with locations that are reachable for part of the time window (i.e., have defined travel time) and later become unreachable (i.e., have undefined travel time) given constraints such as maximum walk distance and maximum overall trip duration. Such problems could occur, for example, at a location that is served by an express bus that runs once, but which is no longer accessible by transit after that bus has stopped running for the day, producing undefined or effectively infinite travel times over part of the departure time window. There is no straightforward way to deal with undefined values in the context of a mean.

However, average instantaneous accessibility presents a concern regarding the fungibility of opportunities (Conway, Byrd, & van der Linden, 2017). The average accessibility metric considers the number of opportunities accessible at each minute, but does not differentiate between opportunities accessible at different minutes. For instance, if 200,000 jobs in one city are accessible within 60 minutes one-quarter of the time, and 100,000 different jobs in another city are accessible within 60 minutes another quarter of the time (for instance, due to infrequent commuter rail service), the average instantaneous accessibility metric would show that 75,000 jobs are accessible within 60 minutes, when in actuality no single job is accessible within 60 minutes at a majority of possible departure/arrival times. Given this fungibility concern, we wish instead to calculate the average travel time to each destination, and include only destinations that have an average travel time of less than our cutoff in the accessibility figure. With opportunities that are not perfect substitutes, this makes sense: we are looking at opportunities which *independently* have an average travel time less than the cutoff. This is an opinionated metric; infrequent transit which requires users to build their lives around the transit schedule will not perform well given the assumptions of this metric. Frequent transit systems which allow relatively stable travel times at all potential departure times will fare much better under this metric.

We use the median travel time, rather than the arithmetic mean travel time, because it does not exhibit the aforementioned problems with destinations that are unreachable for part of the time window. A departure time or timetable for which a destination is not reachable at all will have the same effect on the indicator value as a departure time or timetable for which that destination is reachable, but the travel time exceeds the travel time budget. This is because the median is affected only by the relative positions of travel times above and below the cutoff, not their values (Conway, Byrd, & van der Linden, 2017, 46). We term this approach *accessibility given median travel time*. This is represented by the “compute percentiles” step in Figure 1; it is followed by a geographic overlay step which sums all of the opportunities with median travel times less than the cutoff.

4 Accounting for Travel Time Variation Using Reliable Accessibility Metrics

The direct interpretation of this metric (accessibility given median travel time) is that a user of the transit system will be able to reach a certain set of opportunities within the travel time budget at least half of the time during the time window (using many random timetables when there are headway-based lines). While this may be appropriate for certain types of activities, such as arriving at a coffee shop (or even arriving at work, if an employee has a flexible work schedule), it is woefully inadequate for other types of trips, for example a job interview, a trip to the airport, or even a daily trip to the workplace for

many employees (Chen et al., 2013, 1506). Chen et al. (2017) create a “reliable” accessibility metric, counting only the opportunities reachable with a certain probability given variation in travel time. Reliable metrics are useful even when system users make the same trip every day on a deterministic, scheduled network, because the system provides access to many opportunities which start at times uncorrelated with the transit network, and therefore some travelers effectively wait at their destination for their activity to start, even if they left their origin at a time that allowed them to catch a transit vehicle with minimal waiting.

It is straightforward to generalize our accessibility given median travel time metric to be a similar “reliable” accessibility metric by replacing the median travel time with another percentile of travel time, for instance the 75th or the 90th. We term this metric *accessibility given x percentile travel time*. The interpretation of this metric is straightforward: the accessibility given 75th percentile travel time indicates that at least 75 percent of the time a user will arrive at their chosen destination within the travel time budget, giving us a reliable accessibility metric. This metric retains the property of correctly handling travel times that are undefined or infinite when certain destinations are unreachable for part of the time window since, like the median, percentiles are based only on positional information.

5 Evaluating Uncertainty Due to Schedule Randomization

For lines described with headways rather than complete timetables, we use a Monte Carlo technique to generate a sample of timetables (Conway, Byrd, & van der Linden, 2017), because exhaustively calculating the impact of all possible timetables is intractable. One concern with using a random sample of possible timetables is that each run of the model will yield different results, and we must properly choose the sample size to keep this instability within acceptable tolerances. When lines are specified with headways rather than full timetables, we cannot produce an exact travel time as we do not know how long the user will have to wait to board the vehicle or at transfers. In theory it would be possible to provide a stable result by computing the travel time using every possible timetable at every possible minute of departure, but this is computationally infeasible. The Monte Carlo approach we use instead yields a random sample from these travel times. In order to avoid concluding that a particular change to the transit system has a positive or negative effect on accessibility when in fact the change in indicator values is due only to timetable sampling error, we would like to create confidence intervals on the accessibility given median (or other percentile) travel time. Additionally, we wish to perform hypothesis tests to ensure that differences observed between transit network scenarios are in fact due to changes in the transit network, rather than due to random variation in the results. In order to compute these confidence intervals and perform hypothesis tests, we would like to know the sampling distribution of the accessibility given median or a reliable percentile of travel time, i.e., the distribution of the results of repeatedly performing the accessibility calculation described previously.

The sampling distribution described in this section quantifies the uncertainty present in the accessibility given median (or percentile) travel time and random schedules. If we run the analysis again with the same inputs, we would expect the results to follow this distribution. Thus, we can use this sampling distribution to evaluate the level of uncertainty in the results due to the method we are using, and to test whether results from two different networks differ due to a systematic effect as opposed to random chance. This sampling distribution does *not* encompass the distribution of outcomes that might be seen once a particular change to the transport network is implemented and given a specific timetable. For computational reasons, our method for evaluating travel time given random timetables mixes many possible timetables together to produce a single median travel time value. However, most agencies in the developed world use timetables, even if the schedule is presented to the public as frequencies. Once the changes to the transport network are implemented, there will thus generally only be a single timetable, which may have much better or much worse than median performance from a given location, depending on transfer timing. Some possible techniques to address this inconsistency are explored in the “Further Research” section below.

Computing the sampling distribution of the accessibility given median (or percentile) travel time can be accomplished with the bootstrap (Efron, 1979), a general method for empirically approximating the sampling distribution of a statistic of interest. It works by sampling *with replacement* (i.e., permitting the same observation to be sampled more than once) from a set of observations to create a series of additional datasets *of the same size* (termed *bootstrap samples*) that resemble datasets drawn from the complete population (in our case, the population of all possible schedules fitting scenario constraints). The statistic of interest is then computed for each of these bootstrap samples, and these additional values of the statistic (termed *bootstrap replications*) can then be used to estimate a sampling distribution. This sampling distribution can then be used to perform hypothesis tests and estimate confidence intervals (Efron & Tibshirani, 1993). This is the “bootstrapping algorithm” step in Figure 1.

In order to compute a sampling distribution of accessibility given median travel time, we take many bootstrap samples from the original Monte Carlo timetable draws, and compute a bootstrap replication of accessibility given median travel time for each of those bootstrap samples. Since the uncertainty is in the travel times (rather than in the accessibility value directly), each bootstrap sample consists of many travel times to each destination, sampled randomly from all of the travel times we have found for that destination using different randomized timetables and departure times. That random sample is then used to compute an accessibility given median or percentile travel time value.

The main challenge with applying the bootstrap in this context is that the naïve bootstrap requires *independent* data (Lahiri, 2003, ch. 2), which our samples are not; they are a time series with repeating elements due to the headways of the lines in the network, and the travel times to destinations served by the same transit line are correlated. Since the introduction of the bootstrap, many researchers have devised methods to bootstrap dependent data (many are described in Lahiri, 2003). Many of these methods use an understanding of the data’s dependence structure to create samples which internally have the same dependence structure as the data, but are independent of each other. While none of the methods defined in the literature appear to be directly applicable to our use case, we can use our understanding of our data’s dependence structure to devise a sampling technique that constructs independent bootstrap samples.

There are two main sources of dependence in our accessibility calculations. The first is that the travel times from a particular combination of departure time and randomly selected timetable will have correlated effects on many destinations (for example, if those destinations are all served by the same transit line, and that line requires a long wait at that particular combination of departure time and timetable). In the most extreme case, a long wait may be encountered at the origin, moving the entire network further away in time. When we calculate the point estimate of accessibility given median travel time, each random timetable is used to compute the travel time to all destinations from the origin point of interest; new random timetables are not generated for each destination. In order to reproduce this dependence, we use travel times from the same randomly sampled set of random timetables and departure times for the n th bootstrap sample at every destination. If we were instead to sample randomly from the travel times to each destination independently, we would understate the variation in the results because more timetables would be included in each sample and therefore we would tend to have fewer bootstrap samples where many of the chosen timetables happen to provide exceptionally good or exceptionally poor travel times.

The second source of dependence is the fact that we associate each random timetable with one specific departure time. In order to make the algorithm computationally tractable, we distribute the number of requested random timetables over the individual minutes of the departure time window. For example, if 1000 random timetables are requested for an analysis of the morning peak, 7:00 AM to 9:00 AM, we will perform 9 searches at each minute of departure within that window, each using a different randomized timetable. Thus, 1080 random timetables are evenly distributed among all the minutes in the departure time window (we round up to ensure that the same integer number of timetables are generated for each departure time).

Random timetables that share the same departure time will have travel times (and therefore accessibility values) that are more correlated with each other than with those for different departure minutes (due to the fixed, timetabled component of the network). In addition, there is correlation between adjacent departure minutes; the travel time to a particular destination given departure at 9:14 AM can be no more than one minute longer than the travel time given departure at 9:15 AM; waiting one more minute at the origin is always an option. This effect applies at both the origin and transfer points. There are also periodic components due to the headways of the transit lines. Other researches have used Fourier transforms to determine these periodic dependence components, albeit for a single O-D pair (Farber & Fu, 2017).

These issues can be addressed using bootstrap techniques designed for hierarchical data. Rather than treating the travel times from all departure minute/random timetable combinations as if they were independent, we instead sample from the travel times given a random timetable *at each departure minute individually*, yielding a sample where all departure times are equally represented, meaning any dependence structure between departure times is preserved. This is similar to the technique described by Davison and Hinkley (1997, 100f), for balanced hierarchical data. However, rather than performing a two-stage sampling process (sampling first from departure times, then from travel times given each departure time and a random timetable), we use all the departure times in each bootstrap sample, and then randomly sample from the travel times computed for each of those departure times, yielding a sample with equal representation of all possible departure times. Our approach better reflects the process of generating the point estimate of accessibility, which also uses all possible departure times.

In this way, all systematic variation due to the portion of the network with known timetables is encapsulated within the individual bootstrap samples, and the sampling distributions represent only the error caused by the random selection of timetables for lines specified with headways. We can then compute a confidence interval using this distribution, for example by using the percentile method (Efron & Tibshirani, 1993, ch. 13).

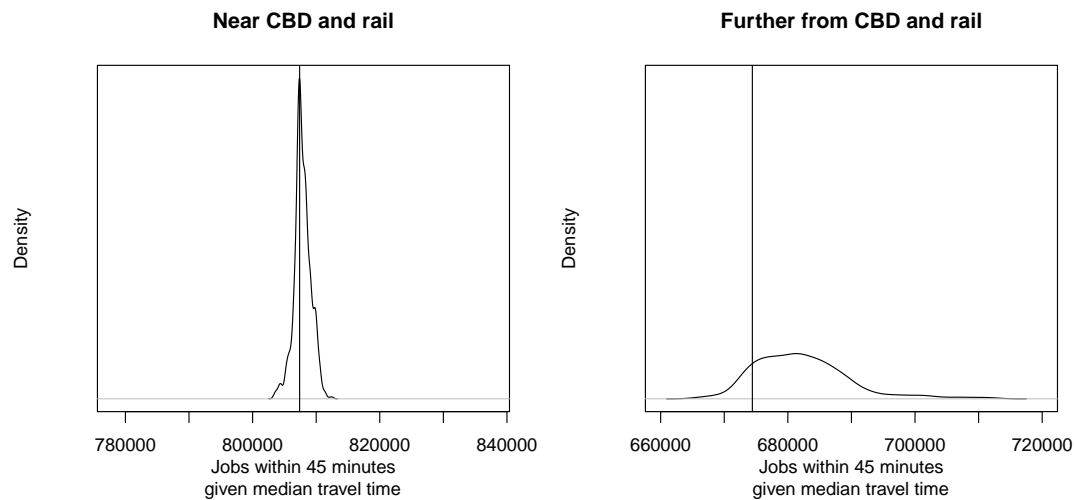


Figure 2: Contrasting bootstrapped sampling distributions for accessibility at origin points with different characteristics.

The bootstrapped sampling distributions of accessibility given median travel time for two locations in Singapore are shown in Figure 2. The first is a location near the central business district (CBD) and served by a proposed frequent rapid transit line. It has a very narrow sampling distribution; there is minimal variation in travel times due to the sampling process. When vehicles run this frequently, all random schedule draws are relatively similar, and the value of the indicator is well defined. Contrast

this to the less central location on the right of Figure 2, served by less frequent buses, where there is much more uncertainty in the results. The vertical lines in both plots represent the accessibility values that were calculated from the original (non-bootstrapped) set of travel times. As expected, they fall in the dense parts of the sampling distributions.

In order to produce defensible sampling distributions, it is important that enough random timetables are used to generate the input to the bootstrapping method, and that enough bootstrap iterations are used to give stable results. Further research is needed to determine these values for different types of transit systems.

6 Comparing Scenarios Probabilistically

6.1 Confidence Intervals on Change in Accessibility Due to a Scenario

Accessibility numbers are minimally useful in and of themselves; they are more useful as relative values (for comparing scenarios) rather than absolute numbers (Geurs & van Wee, 2004/2006). Since our accessibility calculations contain uncertainty, we wish to use the sampling distributions we have computed to compare results from two different transport scenarios (one of which may be the existing transit service) and determine whether a given change is statistically significant, or consistent with random chance.

We first perform the analysis as described above for each scenario, choosing an appropriate percentile of travel time to ensure destinations are reliably reachable. Then at each origin point of interest, we subtract the accessibility sampling distribution for one scenario from that for the other scenario, yielding a distribution of the accessibility differences we might encounter if the analysis were re-run and the comparison repeated many times. The difference of two random variables is a convolution, which can be computed empirically: since subtraction is fast on modern computer hardware, we can exhaustively subtract every bootstrap replication of accessibility for the second scenario from each bootstrap replication of accessibility for the first scenario without resampling. The bootstrap replications from the two scenarios are independent, so no special sampling techniques are needed to ensure that the sampling distribution of the difference is correct.

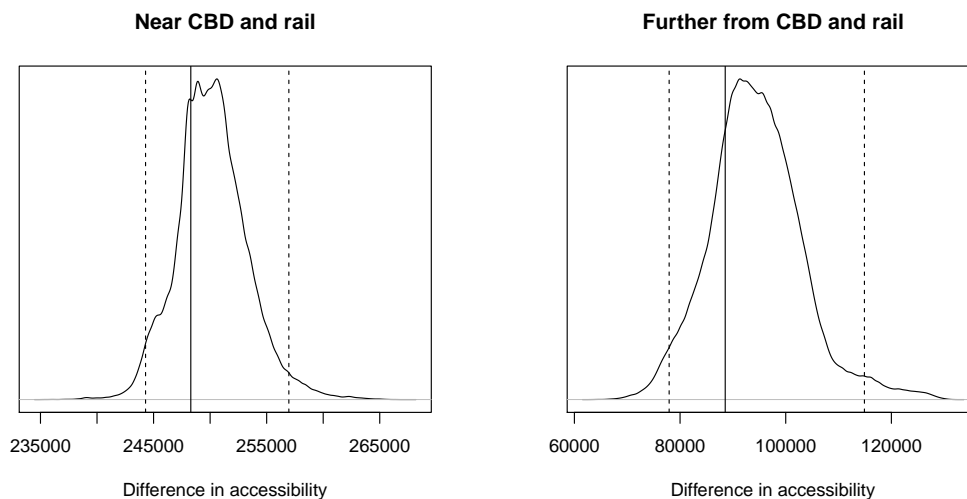


Figure 3: Contrasting bootstrapped sampling distributions for differences in accessibility at origin points with different characteristics.

From this distribution of accessibility differences, we can compute a bootstrap confidence interval on the difference between the two scenarios at each separate origin point using the percentile method

described in Efron and Tibshirani (1993, ch. 13). In this method, we find two specific percentiles of the sampling distribution of differences, giving us the endpoints of the confidence interval. For instance, to compute the 95 percent confidence interval, we would take the 2.5th and 97.5th percentiles of the sampling distribution of the differences in accessibility. The confidence interval then allows us to state the sign and approximate magnitude of change in accessibility given median or percentile travel time due to transit network changes described in the scenario. Figure 3 shows the sampling distribution of accessibility differences due to the changes in the transport network described in the case study section, at the two origin points described in the previous section. The dark vertical lines represent the differences in the point estimate of accessibility, and 95 percent confidence intervals are indicated by the dashed lines. As described above, this confidence interval only characterizes the change in accessibility given median travel time with random departure time and timetables; it does not represent the full range of possible rider experiences given a single timetable.

6.2 Hypothesis Tests on Change in Accessibility Due to a Scenario

It is also possible to perform a hypothesis test to evaluate whether reported changes in accessibility are due to changes in the transport network described in a scenario, or simply due to sampling error in the randomized timetables. We do this by computing a p -value that the sampling distributions of accessibility at a particular origin point from the two scenarios differ. Our analysis software uses this method to display scenario comparison results on a map: the difference in the point estimate of accessibility is displayed at each origin point, but locations with a p -value greater than a chosen cutoff are hidden to mask random noise, highlighting changes that are very likely to be the result of the scenario itself.

We can use the percentile method to produce this p -value from the sampling distribution of differences described above. We construct a confidence interval in which one end of the interval falls exactly at 0, exclusive (Efron & Tibshirani, 1993, 214ff). We determine whether the low or high end of the confidence interval should be placed at zero based on the sign of the (non-bootstrapped) estimate of accessibility. If the estimated value is positive, the low end of the confidence interval will be placed at zero; if it is negative the high end will be placed at zero. The confidence level of this confidence interval is the complement of the p -value of the (two-tailed) hypothesis test. For example, if this happens to be an 85 percent confidence interval, the p -value is 0.15. We prefer two-tailed tests as they make no assumptions about the direction of change (if a proposed transit improvement had a statistically significant *negative* impact, we'd certainly want to know about it!).

Statistical significance testing effectively masks much of the noise in the results, but does not completely eliminate it. When using a 95 percent confidence level, for example, 5 percent of the time we can expect to get a statistically significant result when in the change was not due to the scenario. This problem is exacerbated when many p -values are computed as a part of an analysis (Wasserstein & Lazar, 2016). If we compute the change in accessibility and an associated p -value for thousands of locations within a region, it is almost inevitable that a few changes will appear significant even in areas where change is due to noise rather than the scenario. We suggest using high thresholds for statistical significance in order to minimize the number of locations that appear significant when there is in fact no change. Even with high thresholds for statistical significance, when an analysis contains (say) 100,000 departure points, it is highly likely that some will show a statistically significant effect when there is in fact no change.

There is a simple visual check which can help to determine whether a particular observed change is due to the scenario, or simply due to random variation. Accessibility metrics are spatially autocorrelated, and generally vary smoothly across space (the opportunities that are reachable from one cell are, for the most part, the same as the opportunities that are reachable from the adjacent cell). Thus, when there are a number of statistically significant cells clustered together, they are likely to represent

a change due to the scenario; isolated statistically-significant cells are more likely to be due to random variation.

Formal methods do exist for controlling error in the presence of multiple testing. The best known multiple testing corrections are the Bonferroni and Šidák tests, which consist of reducing the critical value at which you consider a particular change statistically significant in order to control the probability of any cell being falsely deemed significant (Abdi, 2007). The Bonferroni test works by dividing the desired p -value of any error existing in the study area by the number of tests performed (the Šidák test is similar but slightly more mathematically complex). Thus, if one wants to limit the probability of *any* false positives to, say, 0.05, when testing accessibility in 100 cells, one would consider a cell statistically significant if it had a p -value of $0.05/100 = 0.0005$ or less.

Clearly, given the large number of cells in a typical accessibility analysis (this analysis contained over 6000 cells, and covers a relatively small area), the critical values will be so low that there will be few if any statistically significant cells. The p -values that are deemed significant will also be based on the extreme tails of the distribution of differences, and thus the cells considered significant are likely to vary a large amount from one run to the next.

Part of the problem is that when there is dependence between tests, the Šidák and Bonferroni approximations provide a lower bound on the appropriate critical value to control the probability that there are any falsely-significant cells (Abdi, 2007). Clearly, our tests are highly dependent, since accessibility varies geographically. Anselin (1995, 96) finds a similar problem with using the Bonferroni correction with a spatially autocorrelated statistic, and notes that it is not clear how to best perform multiple testing correction in such a situation. da Silva and Fotheringham (2015) developed a method to estimate the appropriate correction factor for Geographically Weighted Regression, which also has many spatially correlated tests; future research could use similar techniques to correct our results.

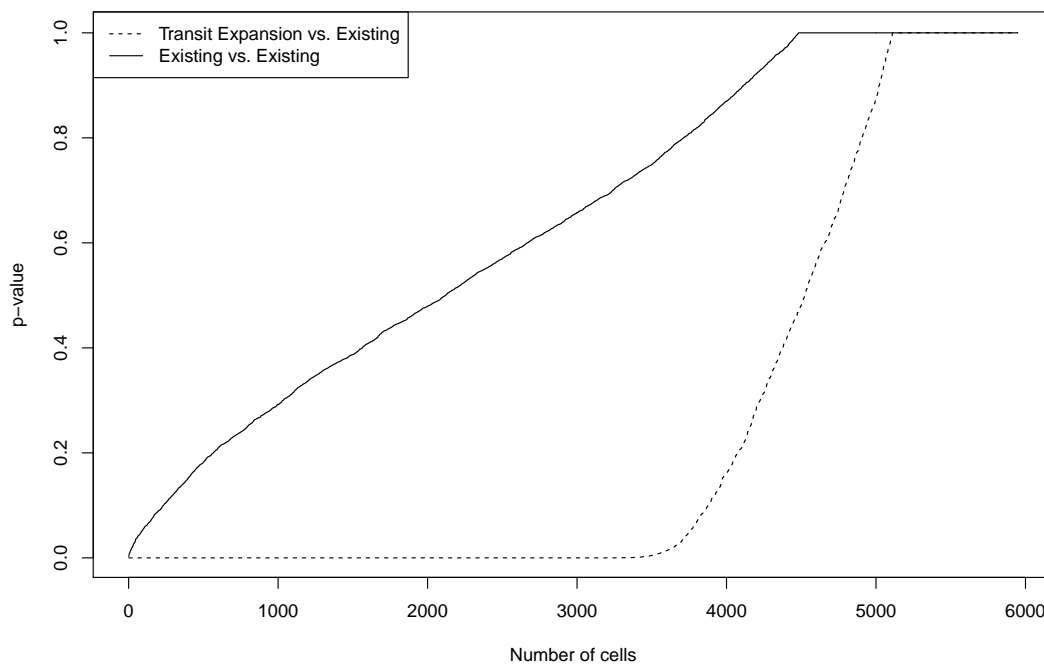


Figure 4: Cumulative distribution of p -values when comparing two analyses of the same transport network (solid line) and a transit expansion to the existing transit network (dashed line).

In an empirical analysis, our method finds more cells than expected with no statistically significant change. The solid line in Figure 4 shows the cumulative distribution of p -values for all origins when evaluating the difference between two analyses of access to jobs using the same transport network and land use data. In this case, we would expect to find that 5 percent of the cells have a p -value less than 0.05, but we see that the number is smaller than that. This is due to a large number of cells with a p -value of 1. There are two ways these can occur. The simplest is that there is no variation between the two runs of the algorithm, so the difference is exactly 0, with no variation, which happens in areas where there is no transit. Figure 4 excludes lakes, the sea, and other areas with an accessibility value of 0, but there may be locations where some jobs are accessible on foot (or via timetabled transit with no travel time variation between analysis runs, although this type of transit does not exist in this particular network). The more likely way is that the point estimate of accessibility lies on the opposite side of zero as the majority of the density (not unlikely when the expectation of the distribution of accessibility differences is zero). In this case, the percentile interval is undefined. Suppose the point estimate is larger than zero, while the majority of the density lies below zero. The percentile method calls then for constructing a confidence interval with its left end at zero; no such confidence interval exists, as a majority of the density of the distribution lies below zero. Our software thus assumes no change and returns a p -value of 1. The dotted line, representing the cumulative distribution of p -values from comparing a transit expansion to the baseline, shows a smaller number of cells with a p -value of 1, as expected because this expansion has a positive impact on accessibility and thus fewer locations have a difference in accessibility of 0, therefore the point estimate and the majority of the density will more often lie on the same side of 0.

There is also a very slight sigmoid shape to the distribution of p -values; the magnitude and direction of this sigmoid shape appears to depend on the ratio of number of random schedules used to the length of the time window. In this example and the case study given below, this ratio is very large. This network has no scheduled component, so we treated the time window as having a length of 1 minute, and all 1000 random schedules used were treated as coming from that minute. With smaller numbers of random draws relative to time window length, the sigmoid shape can be in the opposite direction, and larger, resulting in more statistically significant changes than expected. It may be possible to reduce this effect by using more advanced techniques than the percentile method, such as bias-corrected confidence intervals. These techniques are discussed in the “Further Research” section below.

7 Case Study: Singapore MRT Expansion

We applied this method to a case study in Singapore, where several new MRT (grade-separated metro) lines will be built in the coming years. A scenario describing this network expansion has characteristics that we know *a priori*: the changes can only improve accessibility figures, and should do so in areas near new stations. The results of our method include random error from the Monte Carlo process in addition to any true change in accessibility. Since under this scenario we already know that all change not due to random variation is positive, it allows us to test our method for filtering out statistically insignificant changes in accessibility. Results judged to be significant should be overwhelmingly positive, while negative results should be filtered out as insignificant.

7.1 Input Data

Given the lack of publicly available data sources describing the public transport supply and work locations in Singapore, we opted to use processed data from a MATSim implementation that is based on public transport smart card data (Fourie et al., 2016). MATSim is an open source agent-based transport demand model and has been implemented for cities and regions worldwide (e.g., Horni et al., 2016).

In this MATSim implementation, a full day of public transport smart card transactions recorded in Singapore was transformed into both transport supply and demand. A regression model of bus speed between stops that is dependent both on the level of demand and network topology was implemented, and a model of bus dwell time at stops that is dependent on the ridership of the bus and its configuration was used. MATSim utilizes an evolutionary algorithm to reach a steady state. The same day is simulated many times, with a fraction of the agents modifying their plans after each iteration. The output of MATSim is stored in an events file, that contains events such as “bus enters stop” and “bus leaves stop.” The events from a simulation that reached steady state were exported to the more commonly used General Transit Feed Specification (GTFS) format that can be read by our analysis software.

While MATSim output is a series of events at exact times, the Singapore transit network (both bus and rail) is experienced by its users as a pure headway-based network, with vehicles arriving at a set headway but not at the same time from one day to the next. Therefore, when converting the MATSim output to the GTFS format that is read by our analysis software, we used stop-to-stop travel times produced by the MATSim model accounting for traffic congestion, but replaced the exact times of arrival and departure with nominal headways from the initial MATSim network description. Our analysis method then calculated accessibility using many randomly selected timetables meeting these headway and travel time constraints, as described above.

The MATSim network is a snapshot of the Singapore network at the time the smart card data was collected. Using Conveyal’s map-based scenario editor, several rail lines were added or extended to reflect the expected state of the system in the mid-2020s.

Reflecting the Tuas West Extension that is planned to open in 2017, the East-West MRT line was extended four stations past its current western terminus at Joo Koon into the Tuas industrial area. The planned Thomson–East Coast Line (expected to open in five stages from 2019 through 2024) is entirely absent from the MATSim model, as is the Downtown line (around half of which is in operation as of this writing). The full length of both these lines was added to our scenario using station location maps available from the government’s one-stop data portal (data.gov.sg).

This particular MATSim model contains only stages 1–3 of the Circle Line; while there are other implementations of the MATSim model, this is the only one that was available for this project. In our scenario, the line was extended up to and beyond its current terminus at HarbourFront to form a complete loop, which is planned for around 2023. Finally, the North-South line was extended one stop to its current terminus at Marina South Pier. An average speed of 40 km/h was applied to all new rail lines and segments, based on observed travel times over the whole length of existing lines. Headways on new lines were set to four minutes to reflect peak service; while some lines run more often than this at peak, this was seen as a reasonable estimate.

The bus network is also likely to change in the coming years, especially with so many changes to the rail network. These changes are not anticipated by our scenario. While their absence does have an impact on the quality of the results, our intent here is to demonstrate our probabilistic scenario comparison methodology rather than focus on the Singapore results themselves.

In order to compute accessibility, we also need data on opportunity locations. Workplaces in Singapore were derived from public transport smart card data processed in MATSim (Chakirov & Erath, 2012). Subsequently, work-related trips were distributed to individual buildings by means of iterative proportional fitting and inflated by mode share per TAZ as observed in the Singapore Household Interview Travel Survey (Ordóñez Medina & Erath, 2013). We binned these data into a regular grid. The center points of these grid cells also served as the origin points for our accessibility analysis.

7.2 Results

The raw output from our Monte Carlo accessibility calculations is not filtered to exclude statistically insignificant changes. Figure 5 shows such output for the change in number of jobs accessible in a me-

dian travel time under 45 minutes, holding employment data at present levels. This analysis used transit headways and travel times for the morning peak (at 8 AM) and randomly generated 1000 timetables meeting these constraints. In these raw results we see a large amount of accessibility increases near transit. This is the expected effect of MRT expansion, the significant “signal” we want to present in our final visualization. Negative changes are also found at a large number of locations around the fringes of the city. These must be artifacts of the Monte Carlo approach: our scenario only adds lines or extends them in ways that do not affect travel or wait times on the existing segments. We deduce that the scenario cannot cause any travel times to increase, and therefore cannot cause accessibility to decrease.

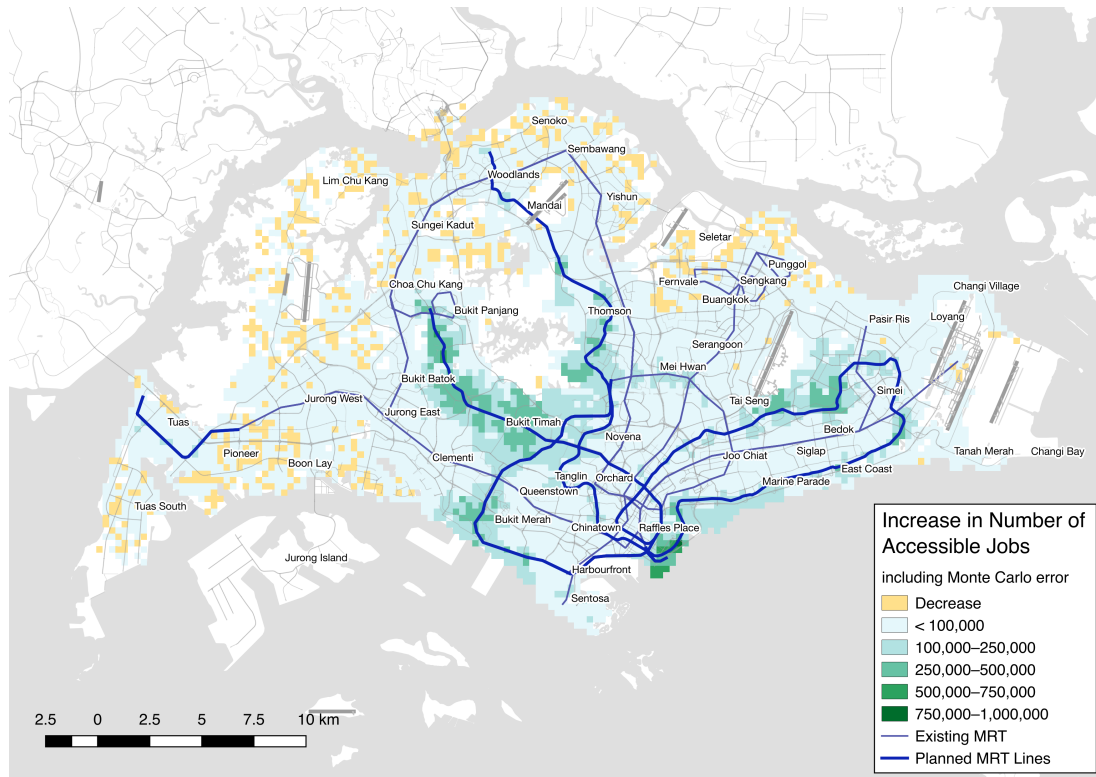


Figure 5: Raw output from employment accessibility calculations, including undesirable noise causing spurious negative values.

We can avoid misinterpreting this noise in the results by applying the probabilistic scenario comparison methods described above, isolating only the statistically significant changes in accessibility. Figure 6 shows the change in job access due to the expansion of the MRT for only locations with a change that is statistically significant at the $p < 0.02$ level.

This visualization of our final filtered results is coherent with the expected impact of our scenario. There are significant gains in accessibility throughout Singapore, with the largest increases in access occurring along the new or extended MRT lines (shown as darker, heavier blue lines, with the lighter lines representing the existing MRT lines). New lines provide a large increase in access to those living nearby, and network effects yield smaller benefits at locations throughout the city from which the new lines can be reached via transfers.

All reported decreases in the raw data are known to be random noise rather than a systematic effect of transit expansion, and all but a few of these have been effectively masked. The few scattered decreases that remain are a result of our chosen confidence level: masking using p -values does not entirely eliminate error, but it hides a large proportion of locations where noise dominates reliable information about changes in accessibility. Some areas of increase have also disappeared, indicating

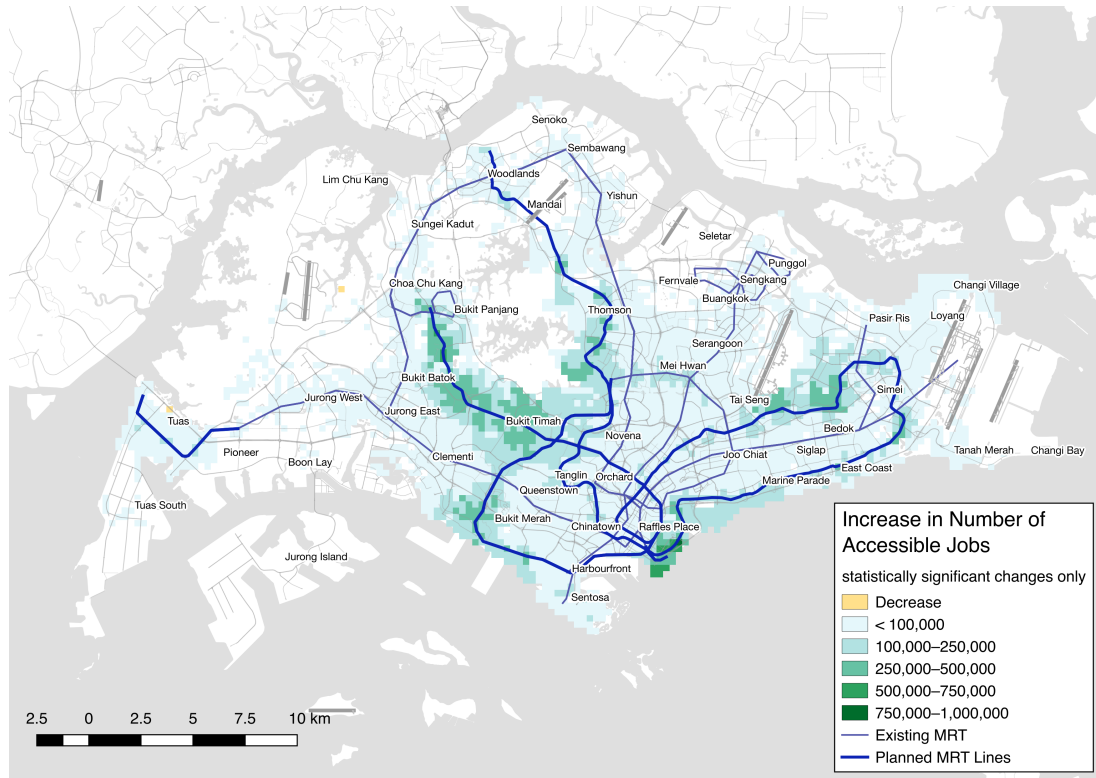


Figure 6: Change in job accessibility due to planned MRT expansion, statistically significant ($p < 0.02$) changes only.

that those particular increases are also likely attributable to timetable sampling, rather than systematic effects of the scenario.

Of course, we can also perform the analysis using a different percentile of travel time, in order to achieve more reliable results, as they are called in the literature. Figure 7 shows how accessibility indicator values change when 85th percentile travel time is used instead of median travel time (with the MRT expansion described above present in both calculations). As expected, accessibility values change less in the center of the network, where travel times are more reliable and commute times to major job centers are shorter. In the periphery, variability is higher, with the reduction in job access due to increased reliability demands exceeding 30 percent in some areas away from the MRT. This is as expected; peripheral locations have longer commutes to job centers (closer to our 45 minute cutoff) as well as less reliable travel times due to longer headways and additional transfers. This shows the value of using reliable metrics; if travelers need to arrive at a specific time, the use of an accessibility metric using average travel time may significantly overstate access.

8 Further Research

There are several avenues for further refinement of this methodology. Its main shortcoming is that it understates the uncertainty in accessibility estimates when headway-based lines in a scenario are eventually operated with a single fixed timetable. This is because our accessibility metric is based on median travel times over many departure times (which is not problematic) but simultaneously over many timetables, and thus does not capture extreme travel times that occur only when a single timetable with particularly well or poorly timed transfers is combined with a particular range of departure times. Travel time variation due to different departure times is not properly isolated from travel time variation due to different timetables. The apparent solution to this problem is to compute the accessibility distri-

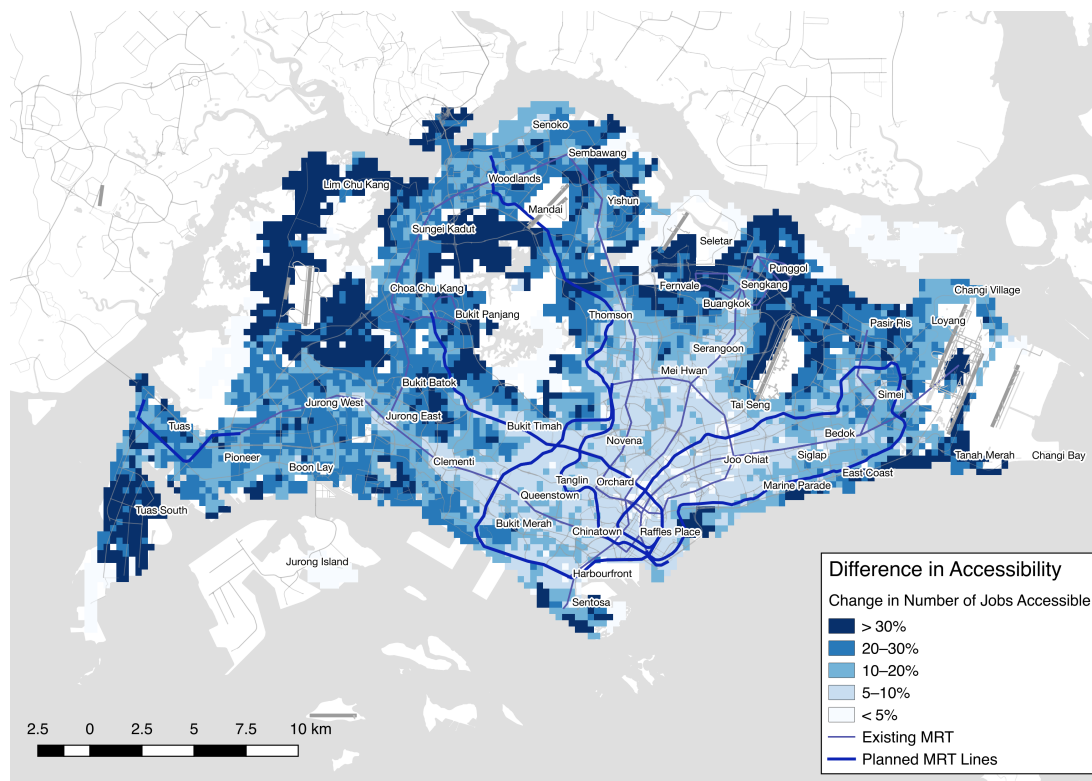


Figure 7: Change in reliably accessible jobs when using 85th percentile rather than median travel time.

butions by combining each randomly selected timetable with all departure times in the window (rather than one particular departure time, as described above). This should yield wider distributions including more extreme accessibility values that occur only under unusual timetables. However, achieving this additional degree of accuracy requires far more computation, which is likely incompatible with the speed requirements of interactive sketch planning.

It should be possible to more quickly approximate the results of this optimal technique by combining each timetable with a random sample of departure times within the time window, i.e., extend the Monte Carlo method to both variables instead of just one (see also Owen & Jiang, 2015). Additional sampling error would be introduced by using only a subset of the departure times, but that uncertainty could be addressed with carefully chosen sample sizes and the bootstrap techniques described above.

A second shortcoming of our method is that it does not take into account any uncertainty in the input data. Zhao and Kockelman (2002) observe that travel demand models frequently propagate only point estimates forward from one stage to the next, without variance information, which leads to unknown levels of uncertainty in results. Much of our input data contains uncertainty. Our land use inputs frequently come from the US Census Bureau's LEHD LODES, which is a simulated population (Spear, 2011), or from regional land use model outputs. In addition, the scenario inputs may be misspecified (travel times on new routes, for example, are often estimates, and may not be reflective of the final travel times). Finally, our method takes the timetables of the network as truth, but in reality there is deviation from these timetables that causes day-to-day variation in accessibility (Stewart, 2017, ch. 8). Our bootstrap method is extensible to account for these types of exogenous variation, by sampling from the distributions of input variables as well when computing results. However, this might require a yet-more-complex sampling strategy for deriving independent bootstrap samples.

Another avenue of further research is to evaluate the uncertainty in our estimates of uncertainty. For example, when we compute the p -value of the change at a particular location, that value is based on several random processes (the random selection of timetables for the two transit scenarios involved,

and the randomization in the bootstrap itself). Thus, that p -value has a sampling distribution of its own, and areas that show a marginally significant change in one run of the analysis might be marginally insignificant in the next. This second-level result would not be presented to the users of a sketch planning system, but rather used to calibrate that system, ensuring that a sufficient number of bootstrap iterations and random schedules are being used. Thus far, the values for these parameters have mostly been based on general practices and rules of thumb. Techniques to quantify uncertainty in bootstrap estimates exist (e.g., the jackknife-after-bootstrap, Efron & Tibshirani, 1993), but may require architectural changes to our processing pipeline to retain the relevant data until it is needed.

We could also improve our method for calculating confidence intervals from bootstrapped sampling distributions. While the percentile method is attractive due to its ease of computation, it has limitations. Notably, it cannot handle situations where no monotonic transformation exists from the empirical sampling distribution to the normal distribution (Efron & Tibshirani, 1993, 176). The bias-corrected and accelerated confidence interval (BCa) represents an improvement (ch. 14). However, its computation requires the original bootstrap samples, rather than simply the bootstrap replications of the differences; since our samples are produced by a complex process of generating travel time distributions and hierarchical sampling, the work needed to adapt these more accurate confidence intervals to our technique is significant, and is thus left as further research.

9 Conclusion

This article has detailed the construction of a reliable transit accessibility metric using particular percentiles of travel time to guarantee a desired probability of on-time arrival. Because a Monte Carlo technique is used to evaluate the effects of transit lines which do not have explicit timetables, there is random error in the results, which we quantify using a bootstrap method to estimate a confidence interval on the result. This technique allows us to perform probabilistic comparisons, evaluating whether an observed change in accessibility is due to a change in the transit network, or simply due to random error. We applied this technique to a scenario with predictable characteristics describing rapid transit expansion in Singapore, and found that it significantly masks random noise in the results.

This technique presents a number of avenues for further research. The most notable is improving our calculation of uncertainty to account for the fact that, once implemented, the public transport network will likely use a single timetable, rather than a different timetable each day. In addition, given the focus on uncertainty, it would be useful to include uncertainty from other sources (e.g., land use models used as input data) in our outputs.

10 Acknowledgements

The research in this paper makes use of derived data describing the transit network and the employment opportunities and is governed by a non-disclosure agreement with Singapore's Land Transport Authority (LTA). The software Conveyal Analysis is open source; thus all methods are available to the public.

The authors wish to acknowledge the helpful comments of three anonymous reviewers in revising this paper. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the Singapore Government. Street network data used to compute and visualize these results is © OpenStreetMap contributors and is available under an open license.

References

- Abdi, H. 2007. The Bonferonni and Šidák Corrections for Multiple Comparisons. In N. Salkind, ed., *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage. URL <https://www.utdallas.edu/~herve/Abdi-Bonferroni2007-pretty.pdf>.
- Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2):93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x. URL <http://doi.wiley.com/10.1111/j.1538-4632.1995.tb00338.x>.
- APPM Management Consultants and Goudappel Coffeng. 2016. Toekomstbeeld OV: Pilot Zuidelijke Randstad. Technical report. URL https://mrdh.nl/system/files/vergaderstukken/I.02.3_Bestuurlijke%20Rapportage%20Toekomstbeeld%20OV%20Zuidelijke%20Randstad_1.pdf.
- Chakirov, A. and A. Erath. 2012. Activity identification and primary location modeling based on smart card payment data for public transport. In *13th International Conference on Travel Behaviour Research*. Toronto.
- Chen, B. Y., Q. Li, D. Wang, S.-L. Shaw, W. H. K. Lam, H. Yuan, and Z. Fang. 2013. Reliable space–time prisms under travel time uncertainty. *Annals of the Association of American Geographers*, 103(6):1502–1521. doi: 10.1080/00045608.2013.834236. URL <http://www.tandfonline.com/doi/abs/10.1080/00045608.2013.834236>.
- Chen, B. Y., H. Yuan, Q. Li, D. Wang, and S. L. Shaw. 2017. Measuring place-based accessibility under travel time uncertainty. *International Journal of Geographical Information Science*, 31(4):783–804. doi: 10.1080/13658816.2016.1238919. URL <http://www.tandfonline.com/doi/abs/10.1080/13658816.2016.1238919>.
- Conway, M. W., A. Byrd, and M. van der Linden. 2017. Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and headway-based networks. *Transportation Research Record*, 2653:45–53. doi: 10.3141/2653-06. URL <http://trrjournalonline.trb.org/doi/10.3141/2653-06>.
- da Silva, A. R. and A. S. Fotheringham. 2015. The multiple testing issue in geographically weighted regression. *Geographical Analysis*, 48(3):233–247. doi: 10.1111/gean.12084. URL <http://doi.wiley.com/10.1111/gean.12084>.
- Davison, A. C. and D. V. Hinkley. 1997. *Bootstrap methods and their application*. Cambridge, England: Cambridge University Press.
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26. URL <http://www.jstor.org/stable/2958830>.
- Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman and Hall/CRC.
- El-Genedy, A. M. and D. M. Levinson. 2006. Access to Destinations: Development of Accessibility Measures. Technical Report MN/RC-2006-16, University of Minnesota. URL <http://conservancy.umn.edu/handle/11299/638>.
- Farber, S. and L. Fu. 2017. Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis) investments over time. *Computers, Environment and Urban Systems*. doi: 10.1016/j.compenvurbsys.2016.10.005. URL <http://www.sciencedirect.com/science/article/pii/S0198971516302861>.
- Farber, S. and M. Grandez. 2017. Transit accessibility, land development and socioeconomic priority: A typology of planned station catchment areas in the Greater Toronto and Hamilton Area. *Journal of Transport and Land Use*, 10(1):1–24. doi: 10.5198/jtlu.2017.980. URL <https://www.jtlu.org/index.php/jtlu/article/view/980>.
- Farber, S., M. Z. Morang, and M. J. Widener. 2014. Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, 53(C):149–159. doi: 10.1016/j.apgeog.2014.06.012. URL <http://dx.doi.org/10.1016/j.apgeog.2014.06.012>.

- Fourie, P. J., A. Erath, S. A. Ordóñez Medina, A. Chakirov, and K. W. Axhausen. 2016. Using smart card data for agent-based transport simulation: the case of Singapore. In J.-D. Schmoecker and F. Kurauchi, eds., *Public transport planning with smart card data*. Boca Raton, FL: CRC Press.
- Geurs, K. T. and B. van Wee. 2006. Accessibility measures: a literature review. In *Accessibility, land use and transport: Accessibility evaluation of land-use and transport developments and policy strategies*, chapter 2. Eburon. Reprinted from *Journal of Transport Geography*, 12(2), 2004, 127–140, URL <http://igitur-archive.library.uu.nl/dissertations/2006-0613-200112/>.
- Hall, R. W. 1983. Travel outcome and performance: The effect of uncertainty on accessibility. *Transportation Research Part B: Methodological*, 17(4):275–290. doi: 10.1016/0191-2615(83)90046-2. URL <http://linkinghub.elsevier.com/retrieve/pii/0191261583900462>.
- Horni, A., K. Nigél, and K. W. Axhausen. 2016. *The multi-agent transport simulation MATSim*. London: Ubiquity Press.
- Lahiri, S. N. 2003. *Resampling methods for dependent data*. New York: Springer.
- Ordóñez Medina, S. A. and A. Erath. 2013. Estimating dynamic workplace capacities by means of public transport smart card data and household travel survey in Singapore. *Transportation Research Record*, 2344:20–30. doi: 10.3141/2344-03. URL <http://trrjournalonline.trb.org/doi/10.3141/2344-03>.
- Owen, A. and H. Jiang. 2015. Temporal Sampling Intervals and Service Frequency Harmonics in Transit Accessibility Evaluation. Unpublished.
- Owen, A. and D. Levinson. 2014. Access Across America: Transit 2014 Methodology. Technical report, University of Minnesota. URL <https://pdfs.semanticscholar.org/82ca/8271f5eca0706222c258053a7a94f73a15cd.pdf>.
- Palmateer, C., A. Owen, and D. M. Levinson. 2016. Accessibility evaluation of the Metro Transit A-Line. Technical report, University of Minnesota. URL <http://conservancy.umn.edu/handle/11299/180900>.
- Spear, B. 2011. Improving employment data for transportation planning. Technical Report NCHRP 08-36, Task 098, American Association of State Highway and Transportation Officials. URL [http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36\(98\)_FR.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(98)_FR.pdf).
- Stewart, A. F. 2017. *Advancing accessibility: public transport and urban space*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA. URL <http://hdl.handle.net/1721.1/111444>.
- Stewart, A. F. and P. C. Zegras. 2016. CoAXs: A Collaborative Accessibility-based Stakeholder Engagement System for communicating transport impacts. *Research in Transportation Economics*, pp. 423–433. doi: 10.1016/j.retrec.2016.07.016. URL <http://dx.doi.org/10.1016/j.retrec.2016.07.016>.
- Walker, J. 2010. Should we redesign our bus network? How? URL <http://humantransit.org/2010/05/basics-should-we-redesign-our-bus-network-and-how.html>.
- Wasserstein, R. L. and N. A. Lazar. 2016. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133. doi: 10.1080/00031305.2016.1154108. URL <http://dx.doi.org/10.1080/00031305.2016.1154108>.
- Zhao, Y. and K. M. Kockelman. 2002. The propagation of uncertainty through travel demand models: An exploratory analysis. *The Annals of Regional Science*, 36(1):145–163. doi: 10.1007/s001680200072. URL <http://link.springer.com/10.1007/s001680200072>.