THE JOURNAL OF TRANSPORT AND LAND USE http://jtlu.org
VOL. 12 NO. 1 [2019] pp. 893–913

JTLU

# Temporal sampling and service frequency harmonics in transit accessibility evaluation

**Brendan Murphy**
University of Minnesota

**Andrew Owen**
University of Minnesota

**Abstract:** In the context of public transit networks, repeated calculation of accessibility at multiple departure times provides a more robust representation of local accessibility. However, these calculations can require significant amounts of time and/or computing power. One way to reduce these requirements is to calculate accessibility only for a sample of time points over a time window of interest, rather than every one. To date, many accessibility evaluation projects have employed temporal sampling strategies, but the effects of different strategies have not been investigated and their performance has not been compared. Using detailed block-level accessibility calculated at one-minute intervals as a reference dataset, four different temporal sampling strategies are evaluated using aggregate sample error metrics as well as indicators of spatially clustered error. Systematic sampling at a regular interval performs well on average but is susceptible to spatially-clustered harmonic error effects which may bias aggregate accessibility results. A constrained random walk sampling strategy provides slightly worse average sample error, but eliminates the risk of harmonic error effects.

**Article History:**
Received: February 6, 2018
Revised version submitted: July 16, 2019
Accepted: October 14, 2019
Available online: December 5, 2019

## 1 Introduction

The concept of accessibility has been an active topic of research in fields related to transportation and land use for many decades (Hansen 1959; Ingram 1971). Early implementations of accessibility evaluation were encouraged by rapid advances in computing power in the 1960s and 1970s (McFadden *et al.* 1973), but were not able to provide detailed enough results, spatially and/or temporally, for use in many planning applications. Over the past decade, increased data availability and renewed interest in accessibility metrics have again encouraged the application of cutting-edge computing approaches to accessibility evaluation.

In the context of public transit networks, repeated calculation of accessibility at multiple departure times provides a more robust representation of local accessibility compared to accessibility metrics using a single departure time. However, these calculations can require significant amounts of time and/or computing power. One way to reduce these requirements is to calculate accessibility only for a sample of time points over a time window of interest, rather than every one. Many accessibility research projects already take this approach.


doi: 10.5198/jtlu.2019.1379


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.

Some benefits of temporal sampling are obvious: an accessibility calculation which is performed for every 5th minute requires only 20 percent of the computation effort as one which is performed for every minute. In addition to direct time and cost savings, making accessibility evaluation more tractable may increase its adoption in early-stage transportation or land use planning projects where planners benefit from quick feedback. Temporal sampling may also offer an opportunity to improve an accessibility evaluation in other dimensions, such as by increasing the spatial resolution or lengthening the temporal coverage, while maintaining the same overall project cost. However, all of these benefits come at the cost of decreased accuracy; additionally, this accuracy loss may not be distributed evenly over space or time. The goal of this analysis is to begin to identify and quantify those costs so that researchers and practitioners can make informed decisions when selecting parameters for accessibility evaluation projects.

This study explores four different methods of sampling minute-by-minute accessibility data from the Minneapolis–St. Paul metropolitan area. Section 2 discusses existing literature and highlights the specific contributions of this study's framework. The four sampling methods, which are evaluated against the source accessibility data, are described in detail in section 4. The sampled datasets are analyzed both statistically and spatially in section 5 to determine viability and appropriateness for practical implication of accessibility sampling in planning settings, including the assessment of spatial clustering of sampling errors. Implications and suggestions for implementation of the tested sampling procedures are discussed in section 6.

## 2 Background

Accessibility, as defined by Hansen (1959), measures the "opportunity for interaction" between people and places. Many specific implementations of accessibility metrics exist; Geurs and van Wee (2004) provide useful categorizations of approaches to measuring accessibility. All accessibility metrics depend on measuring, calculating, or modeling the cost of travel, often expressed in terms of time. Most of the accessibility metrics discussed in this analysis are *location-based* metrics: they incorporate (sometimes very detailed) calculations of what opportunities can be reached from a given location at a given time. As an example, a typical accessibility metric might indicate that if a traveler departs from a specific intersection at 8:10 AM, they could reach 15,000 jobs within 30 minutes.

Public transit networks, like all schedule-based transportation systems, have two important constraints: you can only depart from specific places, and you can only depart at specific times. The latter is an important consideration when evaluating accessibility of transit systems. Accessibility calculation relies on travel time calculation, and travel times vary over the course of the day. Thus, the selection of a departure time, or departure times, impacts the final accessibility metrics. Until relatively recently, implementations of accessibility measures for public transit typically focused on single time points, assigning the accessibility achieved at one departure time to a location or traveler. Some studies expanded this approach to include several time points, typically hourly, over the course of a day. These approaches are reviewed by Owen and Levinson (2015).

Since then, Geurs *et al.* (2015) have identified "temporal dynamics in accessibility" as a rapidly-expanding field, and this is clearly demonstrated in the literature. Owen and Levinson (2015) explicitly compares "continuous accessibility," calculated for every minute and then averaged over a time window, to single-departure-time metrics and demonstrates their ability to improve models of commute mode share. Similarly, Legrain *et al.* (2015) employ accessibility averaged continuously over time periods for mode share modeling; Farber *et al.* (2014), Widener *et al.* (2017), and Widener (2017) use accessibility calculated at every minute of the day to evaluate access to grocery stores, and compare access from areas with lower or higher transit connectivity. Kujala *et al.* (2018) demonstrate the importance of the consideration of departure times, wait times, and number of transfers in evaluating continuous accessibility via pareto-optimal routes in Helsinki, and propose efficiency metrics which incorporate the variables of wait times and number of transfers.

Alternatively, accessibility is evaluated at a sample of time points and then aggregated, typically by averaging. Karner (2015) demonstrates the application of a "highly resolved temporal metric" for transit accessibility in equity evaluation for federal grant programs, averaging 9 temporal samples selected from 15-minute bins over a two hour period. (This corresponds to the systematic sampling strategy evaluated below, with $f = 15$.) Tasic *et al.* (2014) and Lei *et al.* (2012) employ evaluation at multiple time points, but focus on access *to* transit facilities rather than accessibility *using* transit to reach destinations. Ding *et al.* (2015) expand the temporal dimension of transit accessibility to cases where use of transit is limited by capacity, investigating how supply and demand for transit service fluctuate over the day.

Notably, none of these studies robustly defend the choice of temporal sampling strategy, or compare it against other possibilities—yet each required a decision about what specific time points to use. In many cases, suitability seems to be implied simply because sample points are spread evenly over time. However, this explicitly violates the principle of probability sampling and may introduce sampling bias; this risk is higher for time-series datasets that exhibit cyclical patterns (Cryer and Kellet 1986) — which, as discussed below, is certainly the case with transit accessibility data (Figure 1).

In other cases, accessibility is evaluated continuously over time with the implication that more must be better—but perhaps equally meaningful results could be achieved with less computation effort. The time and computation resources required to implement detailed accessibility evaluations are rarely discussed, but they can be a significant barrier to implementation. In the *Access Across America: Transit 2015* project (Owen *et al.* 2016), the calculation of transit accessibility at 120 time points for 70,759 Census blocks in the Minneapolis-Saint Paul, MN statistical area was executed in parallel over many powerful computers; if run on a single high-end workstation, it would have taken approximately 30 hours to complete.

This computational complexity is not unreasonable for research projects or analysis programs undertaken by public agencies. But it is important to recognize that it may limit the impact that accessibility evaluation can have on transportation and land use planning. It is difficult to incorporate a calculation which may take many hours into sketch planning, scenario building, or alternatives analysis because any change to the model network requires an entirely new calculation. Interactive accessibility evaluation tools which provide a balance of fast enough and precise enough feedback in response to a network or land use change could dramatically change the way that planners interact with accessibility concepts. New algorithms for routing on transit networks (such as RAPTOR (Conway *et al.* 2017; Delling *et al.* 2014)) can improve response times; intelligent temporal sampling may also play a role.

## 3 Data

The accessibility data used in this analysis were calculated by the University of Minnesota's Accessibility Observatory as part of the *Access Across America: Transit 2016* project. For Census blocks in the Minneapolis-Saint Paul, MN metropolitan area, the data indicate the number of jobs that can be reached within various time thresholds using a combination of walking and transit. Travel times are calculated on a combined pedestrian and transit network using transit schedules effective in January 2016, and include walking, waiting, and on-vehicle trip components. Job counts and locations are based on estimates published by the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. Additional methodological details are presented in Owen *et al.* (2016). This analysis uses the data for the 30-minute travel time threshold, chosen to correspond roughly to the national average commute time of 25 minutes (American Association of State Highway and Transportation Officials 2015).

A critical feature of this dataset is that rather than assuming a single fixed departure time, the accessibility calculations were repeated for every minute from 7:00 AM to 9:00 AM, for a total of 120 accessibility observations at each origin block location. Figure 1 illustrates the data for a single block group: the accessibility value for each minute indicates the number of jobs that can be reached within

30 minutes if a traveler departs at exactly that time. Accessibility increases as transit vehicles approach nearby stops and stations, then drops after trips depart due to the wait the before the next trip. This particular block shows a very pronounced cyclical pattern, with accessibility peaks occurring every 10 minutes. The location is near a station on Minneapolis' Blue Line LRT route, which provides frequent service and a fast connection to downtown.

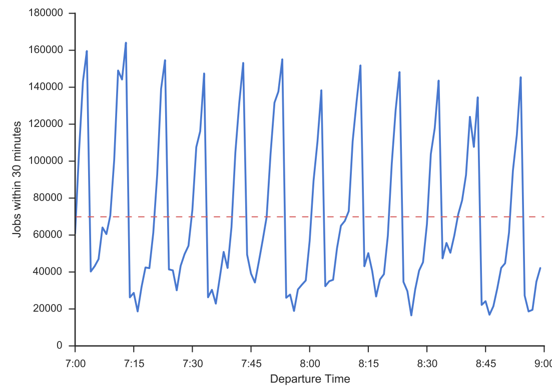


**Figure 1:** Accessibility plot for a single Census block (270531088002012, near the 38th St. Blue Line station in Minneapolis, MN). For each departure time between 7:00 and 9:00 AM, the accessibility value indicates the number of jobs that can be reached within 30 minutes by walking and transit. The dashed horizontal line indicates the average accessibility value over the entire time period.

The data cover the entire Minneapolis–Saint Paul statistical area defined by the U.S. Census Bureau, which includes those core cities plus many surrounding counties in Minnesota and Wisconsin. However, some of the suburban and most of the rural parts of this area are not served by a fixed-route transit system. From these locations, transit service has no impact on accessibility—the entire 30-minute time budget is consumed without ever reaching a stop or station. Therefore accessibility to jobs in these places, as represented in this dataset, is determined entirely by walking.

## 4 Methodology

Four sampling strategies are evaluated: systematic sampling, simple random sampling, hybrid systematic/random sampling, and random walk sampling. Because all of these sampling strategies contain a random element, each is repeated multiple times in a Monte Carlo approach and the results are averaged to provide an indicator of the overall performance of that strategy. In a single application of a sampling strategy, a specific sample pattern is generated and used to select, without replacement, a specified number of data points from the 7:00 AM–9:00 AM accessibility data for a block. The sample average is compared to the data average as an indicator of sample error. This methodology has been reproduced by Stepniak *et al.* (2019) in Poland, using the same four sampling strategies. The following sections describe each sampling strategy, and the performance evaluation methodology, in more detail.

### 4.1 Sampling Frequency

Each of the temporal sampling strategies explored in this analysis involves some variation on the concept of sampling frequency, which is expressed differently in each of the strategies. "Frequency" implies some degree of regularity, but the actual regularity of generated samples varies widely over these strategies. At the most fundamental level, the sampling frequency indicates the number of samples that will be selected from a window of time points: for a sampling frequency $f$, $\lfloor T/f \rfloor$ out of $T$ total time point will be selected. Each of the sampling strategies described below is evaluated at sampling frequencies of 2 through 30 minutes.

### 4.2 Simple Random Sampling

This sampling strategy is well-documented in the literature (Gupta and Shabbir 2008; Kadilar and Cingi 2005; Levy and Lemeshow 2011), and is the most straightforward: a specified number of sample times are selected at random, without replacement, from the time window. Because the selection is completely random, this strategy will often produce sampling patterns which are very unevenly distributed—a sample consisting of the first $\lfloor T/f \rfloor$ time points is just as likely as $\lfloor T/f \rfloor$ points evenly distributed over the time window.

### 4.3 Systematic Sampling

This sampling strategy involves selecting samples at a regular interval, defined by $f$, over the time window (see Weiss (1984), Brewer (1973), and Madow and Madow (1944) for discussions of broader applications and supporting theory). To avoid bias and give each data point an equal chance of being selected, the first sample in the sequence is selected randomly from the first $f$ time points. Thus, one application of this sampling strategy for $f = 5$ might produce the sequence 7:00, 7:05, 7:10, etc. while the next produces 7:03, 7:08, 7:13, etc. This strategy has the advantage that it is guaranteed to produce evenly-distributed samples. A disadvantage is that if the data itself is cyclical on a frequency which is a multiple of $f$, this strategy may produce a sample pattern which coincides only with peaks (or troughs), in which case the sample average could be significantly higher (or lower) than the data average.

### 4.4 Hybrid Sampling

This strategy begins with a systematic sample, starting with the first and selecting every $f$th sample thereafter. Next, a random offset based on the sampling frequency (a random integer in $[1, f]$) is applied to each sample point. This can also be regarded as a clustered sampling strategy, where sample candidates are clustered into bins with a width of $f$ and then a single random sample point is chosen from each cluster.

The motivation behind this strategy is to avoid the temporal clustering of samples that can occur in simple random sampling while also avoiding the potential harmonic effects of systematic sampling. While it does avoid clustering on a large scale, it does not enforce any minimum distance between sequential samples, with the result that the gap between samples varies from 1 to $2f$. In the (unlikely) worst case, this sampling strategy might produce a sample pattern consisting of $\frac{1}{2}\ \lfloor T/f \rfloor$ pairs of adjacent samples, with a distance of $2f$ between each pair. Because adjacent sample points are likely to have similar values, the sample error in this worst case has the potential to be similar to systematic sampling at half the frequency.

### 4.5 Constrained Random Walk Sampling

This sampling strategy is based on a random walk, where each sample point is a random distance from the previous (Kac 1947; Spitzer 1964). The first sample point is randomly chosen between 1 and $f$, and then each random "next step" is constrained by a function of the sampling frequency, in order to achieve a mix of randomness and temporal dispersion. To choose each next sample point, a random offset between $\lfloor f/2 \rfloor$ and $\lfloor f + (f/2) \rfloor$ is added to the previous sample point, so that on average the next sample point is $f$ greater than the previous.

One disadvantage to this sampling method is that it produces sample sets of varying size. For repeated trials, the average number of samples in each set approaches $\lfloor T/f \rfloor$. If a predictable number of samples is desired (for example to make computation times more predicable), the sample generation process can be filtered to discard sample patterns which do not contain exactly $\lfloor T/f \rfloor$ sample points.

### 4.6 Comparison of Sampling Strategies

Because the "ground truth" data are available in this analysis, each sampling outcome can be compared directly against the actual distribution, either by calculating the expected sampling outcome empirically or by estimating it using a Monte Carlo method. The simple random sampling, hybrid sampling, and random walk sampling strategies each incorporate a high degree of randomness and can produce extremely large numbers of specific sampling patterns. For example, simple random sampling of 24 data points from a population of 120 (a 10-minute sampling frequency over a 2-hour period) can produce over $10^{25}$ individual sampling patterns. Therefore, these sampling strategies are evaluated using a Monte Carlo method where the results of repeated samplings are averaged to estimate the expected outcome. In this analysis, each of these sampling strategies is repeated 1,000 times for each block and then averaged; this is repeated for each sampling frequency.

The systematic sampling strategy, on the other hand, produces far fewer specific sampling patterns—for example, 10 in the case of a 10-minute sampling frequency. Therefore it is feasible to calculate the result of all possible sampling outcomes at all sampling frequencies, and compare the averages to the actual distribution.

The performance of the various sampling strategy and frequency combinations are evaluated based on how well they estimate the true average accessibility value for each block. A normalized root mean square error (NRMSE) metric is calculated within the context of each block, where the average accessibility estimated over repeated trials of each sampling method is compared with the true average. To compare aggregate performance over many blocks with different accessibility scales, each result is normalized using the data range for that block. Thus, the average sampling error is expressed as a percentage of the range of the true data for that block (Equation 1).

$$\text{NRMSE} = \frac{\frac{1}{n}\sqrt{\sum_{t=1}^{n}\left(\hat{y}_t - y\right)^2}}{y_{max} - y_{min}} \tag{1}$$

$$\begin{aligned} \hat{y}_t &= \text{the average of sample set } t \\ y &= \text{actual average} \\ n &= \text{total number of sample sets} \end{aligned}$$

As noted above, the study area contains many blocks where transit service has no impact on accessibility because no stops or stations can be reached within the 30-minute travel time budget. In these locations, accessibility is constant over time (walking speeds are assumed to not vary by time of day), and so the average of every sample, regardless of strategy, will be equal to the actual average; the sampling error will always be zero. To avoid diluting the results from blocks where transit does have an impact on accessibility, 31,849 "transit-less" blocks are excluded from the analysis of sampling strategy performance, leaving 34,810 (52.2 percent) remaining blocks in the population.

### 4.7 Local Moran's *I* Clustering Analysis

Certain sampling methodologies or sampling frequencies may produce areas of high NRMSE values near transit service stations due to sampling at frequencies closely aligned with transit service frequencies, and such anomalies are statistically detectable using spatial autocorrelation analysis. As points of transit service tend to be fairly localized (i.e., transit service, and thus accessibility measurements and their associated sampling errors are spatially concentrated), a Local Moran's $I$ approach is chosen over simple Global Moran's $I$ or Global Geary's $C$ metrics. Local Moran's $I$ gives the following local indicator of spatial association (LISA—see Anselin (1995) for a discussion of these measures) statistic for each area of interest—in this case, Census blocks:

$$I_i = \frac{Z_i N}{\sum_i Z_i^2} \sum_j W_{ij} Z_j \tag{2}$$

where $Z_i$ is the deviation of NRMSE at Census block $i$ from the population mean, $N$ is the total number of Census blocks, and $W_{ij}$ is the weight associated with blocks $i$ and $j$ (typically 1 if the two blocks are neighbors, and 0 otherwise).

To test for the presence of clusters of Census blocks with higher (or lower) NRMSE values (areas of blocks with higher $I_i$ values), the variance and skewness of the distributions of block-level LISA statistics are reported. A combination of high variance and high right-skewness of the population of LISA statistics would indicate spatial clustering of blocks with high NRMSE values, indicating that a particular sampling methodology or frequency may produce biased estimates of accessibility.

## 5 Results and Discussion

Overall performance data for NRMSE statistics and spatial autocorrelation statistics, for each of the four methodologies and for all sampling frequencies tested, are reported in Table 1 and Table 3, respectively. Methodologies are evaluated and compared first by NRMSE statistics (subsection 5.1), and then by spatial autocorrelation statistics (subsection 5.2); computational speedup tradeoffs and methodology recommendations are outlined in subsection 5.3.

### 5.1 NRMSE Comparisons

The results for the simple random sampling strategy, presented in Figure 2(e) and Table 1, immediately demonstrate the role of increased randomness in sample selection. There are no harmonic error effects apparent, and standard deviation of NRMSE at the 5 and 10 minute sampling frequencies are significantly lower than for the systematic sampling strategy, indicating more consistent results. However, the overall performance of this strategy, as indicated by average NRMSE, is markedly worse. It is especially poor at shorter sampling frequencies (2–4), where its average sample error is over twice that of the the systematic sampling strategy.

Figure 3 shows the relative performance of all sampling strategies and frequencies. These charts clearly illustrate the variability in sample error at the 5- and 10-minute sampling frequencies when using the systematic sampling method. Overall, the random walk sampling strategy provides the best performance, as measured by low average sample error and low standard deviation of sample error, while also avoiding harmonic error effects. The systematic sampling method often performs very well, but appears to be strongly influenced by harmonic error effects that make its performance unpredictable.

Three of the four methods (all but systematic) exhibit grouping among certain sampling frequencies—that is, the effective sampling frequency is the same for some sets of input frequencies (e.g., 11 and 12 minutes, 21–24 minutes, etc.), due to the number of sample points $M$ being a stepwise function of input sampling frequency: $M(f) = \lfloor (120/f) \rfloor$. For the systematic method, even though the effective *number* of samples may be the same within a given group, the spacings between sample points are distinct, and thus sampling patterns for adjacent input sampling frequencies are unique for this method. For this reason, only data for the first input sampling frequency within each such group for the three affected methods are included in Table 1 and Table 3, and all data for the systematic method are included; others within each group are omitted for clarity. The effect in Figure 3(k) where NRMSE means decrease within each banded group for the hybrid method is due to increasing the upper bound of the random offset range of $[1,f]$ while $M$ remains constant within a group.

Figures 2(a) to 2(d) show a series of maps for the systematic sampling strategy. The most striking feature is the clear harmonic error effects, seen most clearly at the 5- and 10-minute sampling frequencies (Figures 2(a) and 2(c)) and to a lesser degree at the 15-minute sampling frequency (Figure 2(d)). The corridor of high error near the center of these maps corresponds to the Blue Line LRT route, which connects the major job centers of downtown Minneapolis, the Minneapolis–Saint Paul airport, and the Mall of America. During the 7:00–9:00 AM weekday morning window used in this evaluation, this route operates on a 10-minute frequency, seen clearly in Figure 1. When sampling at a 10-minute frequency, the samples often fall on the peaks or troughs of the local accessibility pattern, and therefore produce sample averages significantly higher or lower than the data average. When sampling at a 5-minute frequency, every other sample hits these peaks or troughs, producing the same effect but to a lesser degree. There are corresponding spikes in the standard deviation of the NRMSE at both the 5- and 10-minute sampling frequencies visible in Table 1, indicating increased dispersion in the sample results.

The hybrid sampling strategy (Figure 2(f) and Table 1) improves on the performance of the simple random sampling strategy in both average error and standard deviation of error, while also avoiding harmonic error effects at the 5- and 10-minute sampling frequencies. The random walk sampling strategy (Figure 2(g) and Table 1) provides a similar further improvement. However, at non-harmonic sampling frequencies the systematic sampling strategy provides a lower average sample error (but a greater standard deviation) than either the hybrid or the random walk strategy.

It is interesting to note the performance of the 7-minute sampling frequency with the systematic strategy. It produces a slightly better mean NRMSE (2.49 percent vs 2.56 percent) and a lower standard deviation (1.01 vs 1.32) than the 6-minute sampling frequency, despite having fewer sample points. This suggests that the 7-minute sampling frequency is not harmonic with the typical service frequency of transit routes in the area, and therefore provides a more accurate estimate of average accessibility.

**Table 1:** Performance Comparison of Sampling Strategies - NRMSE Statistics

| | Sampling Strategy | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Simple | | | Systematic | | | Hybrid | | | Random Walk | | |
| Sampling Frequency | Median | Mean | S.D. | Median | Mean | S.D. | Median | Mean | S.D. | Median | Mean | S.D. |
| 2 | 1.90% | 1.89% | 0.89 | 0.62% | 0.80% | 0.78 | 1.03% | 1.02% | 0.37 | 0.62% | 0.80% | 0.78 |
| 3 | 2.91% | 2.88% | 0.66 | 1.12% | 1.28% | 0.79 | 1.66% | 1.66% | 0.49 | 1.49% | 1.50% | 0.34 |
| 4 | 3.56% | 3.53% | 0.81 | 1.48% | 1.70% | 0.94 | 2.22% | 2.22% | 0.60 | 1.93% | 1.95% | 0.45 |
| 5 | 3.80% | 3.78% | 1.80 | 2.27% | 2.84% | 1.99 | 2.74% | 2.75% | 0.74 | 2.44% | 2.48% | 0.58 |
| 6 | 4.59% | 4.55% | 1.04 | 2.56% | 2.84% | 1.32 | 3.26% | 3.25% | 0.83 | 2.87% | 2.90% | 0.66 |
| 7 | 5.05% | 5.01% | 1.15 | 2.49% | 2.67% | 1.01 | 3.74% | 3.74% | 0.93 | 3.36% | 3.39% | 0.78 |
| 8 | 5.43% | 5.39% | 1.23 | 2.88% | 3.12% | 1.24 | 4.17% | 4.17% | 1.02 | 3.75% | 3.79% | 0.89 |
| 9 | 5.89% | 5.84% | 1.33 | 3.21% | 3.45% | 1.22 | 4.71% | 4.68% | 1.14 | 4.27% | 4.32% | 1.06 |
| 10 | 6.02% | 6.03% | 1.70 | 4.70% | 5.55% | 3.46 | 5.02% | 5.00% | 1.23 | 4.59% | 4.63% | 1.12 |
| 11 | 6.81% | 6.75% | 1.54 | 3.86% | 4.19% | 1.55 | 5.83% | 5.78% | 1.38 | 5.29% | 5.33% | 1.30 |
| 12 | - | - | - | 4.57% | 4.91% | 1.74 | - | - | - | - | - | - |
| 13 | 7.21% | 7.15% | 1.64 | 4.51% | 4.87% | 1.71 | 6.27% | 6.21% | 1.49 | 5.88% | 5.90% | 1.42 |
| 14 | 7.68% | 7.62% | 1.74 | 5.26% | 5.79% | 2.30 | 6.81% | 6.77% | 1.62 | 6.52% | 6.54% | 1.65 |
| 15 | 7.52% | 7.52% | 2.14 | 6.90% | 7.69% | 3.67 | 6.82% | 6.77% | 1.60 | 6.53% | 6.55% | 1.59 |
| 16 | 8.25% | 8.18% | 1.87 | 6.19% | 6.73% | 2.57 | 7.50% | 7.44% | 1.76 | 7.12% | 7.11% | 1.70 |
| 17 | - | - | - | 5.82% | 6.34% | 2.28 | - | - | - | - | - | - |
| 18 | 8.95% | 8.88% | 2.03 | 6.24% | 6.61% | 2.12 | 8.35% | 8.29% | 1.95 | 7.89% | 7.85% | 1.88 |
| 19 | - | - | - | 6.65% | 7.18% | 2.53 | - | - | - | - | - | - |
| 20 | - | - | - | 7.55% | 8.53% | 3.86 | - | - | - | - | - | - |
| 21 | 9.84% | 9.76% | 2.24 | 7.38% | 7.91% | 2.75 | 9.39% | 9.33% | 2.22 | 9.18% | 9.11% | 2.27 |
| 22 | - | - | - | 7.47% | 7.88% | 2.52 | - | - | - | - | - | - |
| 23 | - | - | - | 7.53% | 7.98% | 2.57 | - | - | - | - | - | - |
| 24 | - | - | - | 7.47% | 7.95% | 2.54 | - | - | - | - | - | - |
| 25 | 11.06% | 10.97% | 2.51 | 8.25% | 8.62% | 2.57 | 10.80% | 10.75% | 2.56 | 10.76% | 10.78% | 3.01 |
| 26 | - | - | - | 8.47% | 8.83% | 2.66 | - | - | - | - | - | - |
| 27 | - | - | - | 9.04% | 9.51% | 3.07 | - | - | - | - | - | - |
| 28 | - | - | - | 9.72% | 10.37% | 3.65 | - | - | - | - | - | - |
| 29 | - | - | - | 10.89% | 11.63% | 4.38 | - | - | - | - | - | - |
| 30 | - | - | - | 12.19% | 13.03% | 5.12 | - | - | - | - | - | - |

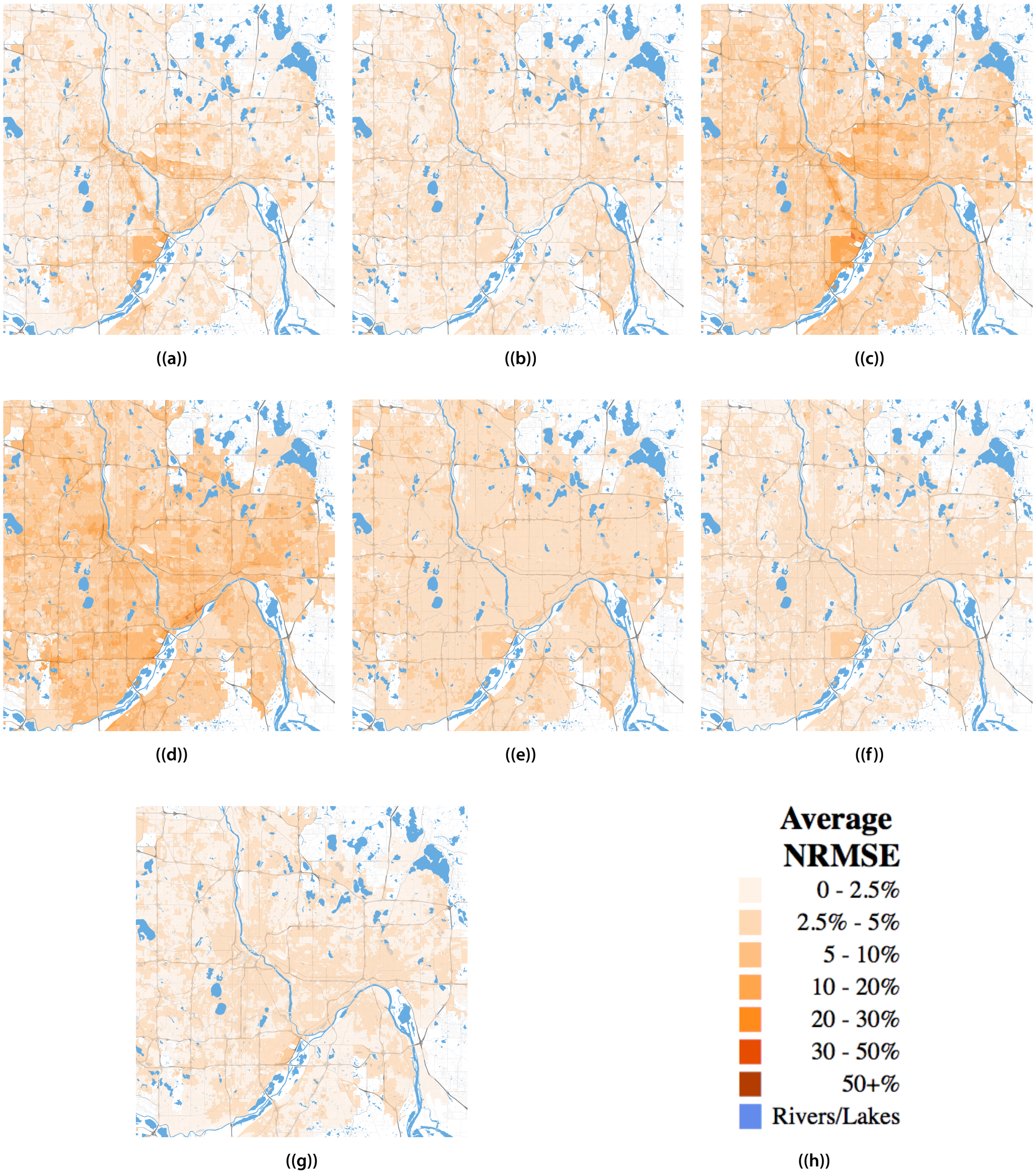


**Figure 2:** Series of maps showing block-level performance of the four sampling strategies implemented. ((a))–((d)) show the mean NRMSE values for sampling frequencies of 5, 7, 10, and 15 minutes for the systematic method—of these, all but ((b)), for a sampling frequency of 7 minutes, exhibit pronounced harmonic error effects at some locations associated with transit service stops. ((e)), ((f)), and ((g)) show the mean NRMSE values for the simple, hybrid, and random walk sampling methods, respectively, each at a sampling frequency of 5 minutes. Darker colors indicate greater average sample error, as shown in ((h)).

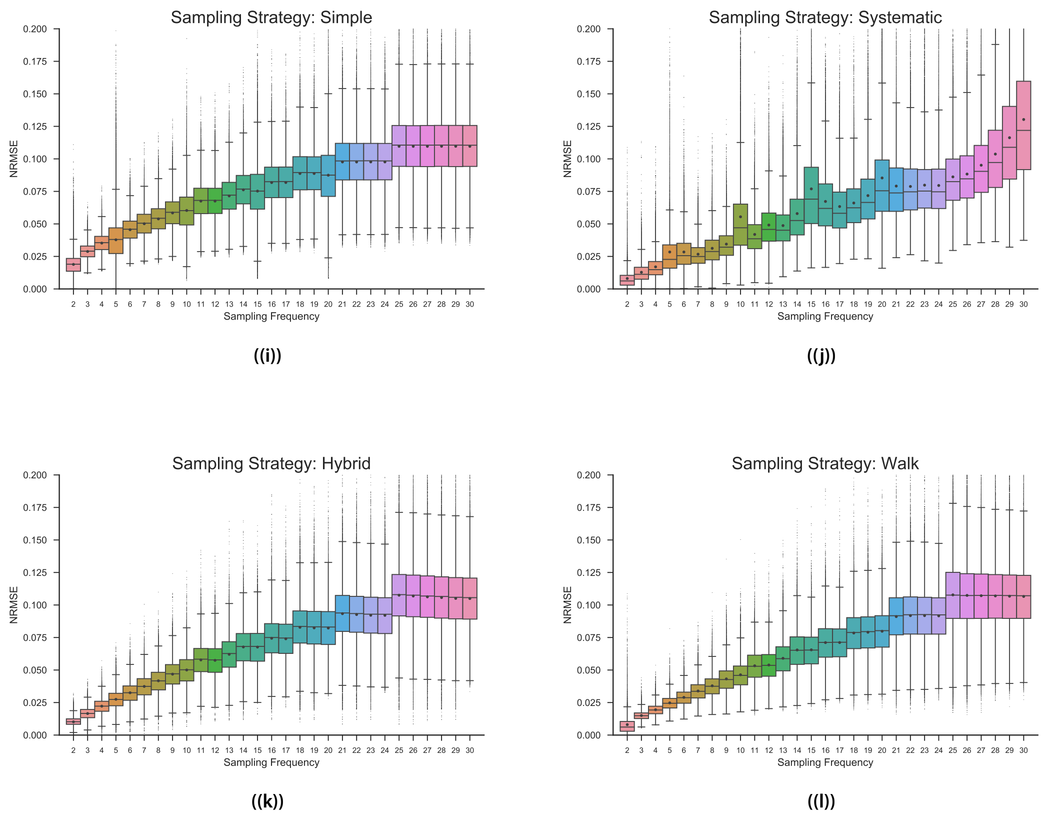


**Figure 3:** Box plots showing NRMSE performance for the four sampling strategies tested, over all blocks at each sampling frequency. Boxes show inter-quartile range (25th–75th percentile) with horizontal medial line; whiskers extend 1.5×IQR above and below. Outliers are plotted individually. Mean is indicated by a dot.

### 5.2 Spatial Autocorrelation Comparisons

To further illustrate that the systematic methodology suffers from unpredictable performance, Figures 4 and 5 show the variance and skewness, respectively, of block-population LISA statistics; the underlying data are reported in Table 3. Higher population variance suggests that a particular sampling method and frequency yields clusters of autocorrelation among block-level NRMSE values, since the presence of LISA outliers could suggest positive or negative autocorrelation; higher skewness values (right-skewness) suggest that LISA outliers lay on the right tail of the distribution of Local Moran's $I$ values. Clearly visible in Figure 4, the systematic sampling strategy yields very high relative variance and unpredictability in NRMSE across the geographic region, with peaks at sampling frequencies of 2, 5, 10, 15, and 20 minutes, corresponding directly to harmonic sampling errors. Figure 5 shows the systematic method yielding the highest right-skewness across the entire sampling frequency range, and also shows local peaks at sampling frequencies of 5, 10, 15, and 20 minutes, indicating the presence of clusters of blocks with high NRMSE values. The darker-colored areas in Figures 2(a), 2(c) and 2(d) correspond directly to the higher variance and right-skewness of LISA statistics for systematic sampling at frequencies of 5, 10, and 15 minutes. For example, with a sampling frequency of 5 minutes, the systematic method yields a LISA variance of 6.77 and skewness of 10.42, both local peaks; with a sampling frequency of 7 minutes, lower LISA variance and skewness values of 0.80 and 5.06, respectively, are obtained. All other methodologies show low variance and skewness among LISA statistics throughout the sampling frequency range, with the exception of constrained random walk at $f = 2$. At sampling frequency 2, the constrained random walk methodology collapses to the systematic methodology (since at $f = 2$, the random walk offset range of $\lfloor f/2 \rfloor$ to $\lfloor f + (f/2) \rfloor$ becomes $[1, 3]$ with an expected value of 2).

**Table 3:** Spatial Autocorrelation Comparison of Sampling Strategies - LISA Statistics

| | **Sampling Strategy** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Simple | | Systematic | | Hybrid | | Random Walk | |
| Sampling Frequency | Variance | Skewness | Variance | Skewness | Variance | Skewness | Variance | Skewness |
| 2 | 0.28 | 1.64 | 7.36 | 13.53 | 0.45 | 2.35 | 7.36 | 13.53 |
| 3 | 0.28 | 1.64 | 2.11 | 13.85 | 0.41 | 2.00 | 0.31 | 1.72 |
| 4 | 0.28 | 1.63 | 2.32 | 9.04 | 0.38 | 1.91 | 0.32 | 1.69 |
| 5 | 0.28 | 1.66 | 6.77 | 10.42 | 0.37 | 1.91 | 0.37 | 2.43 |
| 6 | 0.28 | 1.64 | 1.57 | 9.84 | 0.36 | 1.80 | 0.31 | 1.50 |
| 7 | 0.28 | 1.66 | 0.80 | 5.06 | 0.35 | 1.78 | 0.33 | 1.73 |
| 8 | 0.28 | 1.65 | 1.04 | 6.04 | 0.34 | 1.66 | 0.37 | 2.45 |
| 9 | 0.28 | 1.63 | 0.87 | 5.74 | 0.32 | 1.52 | 0.43 | 3.28 |
| 10 | 0.28 | 1.63 | 7.11 | 9.69 | 0.33 | 1.81 | 0.39 | 2.70 |
| 11 | 0.28 | 1.63 | 1.09 | 6.24 | 0.30 | 1.57 | 0.36 | 1.93 |
| 12 | - | - | 0.96 | 6.20 | - | - | - | - |
| 13 | 0.28 | 1.65 | 1.04 | 5.03 | 0.31 | 1.62 | 0.35 | 2.01 |
| 14 | 0.28 | 1.62 | 1.48 | 6.32 | 0.30 | 1.41 | 0.39 | 2.33 |
| 15 | 0.28 | 1.62 | 2.41 | 7.10 | 0.30 | 1.40 | 0.35 | 1.98 |
| 16 | 0.28 | 1.64 | 1.09 | 4.21 | 0.31 | 2.10 | 0.32 | 1.58 |
| 17 | - | - | 1.03 | 4.09 | - | - | - | - |
| 18 | 0.28 | 1.63 | 0.80 | 5.42 | 0.29 | 1.68 | 0.32 | 1.82 |
| 19 | - | - | 1.05 | 5.10 | - | - | - | - |
| 20 | - | - | 2.61 | 6.38 | - | - | - | - |
| 21 | 0.28 | 1.68 | 1.08 | 5.95 | 0.30 | 1.92 | 0.33 | 1.85 |
| 22 | - | - | 0.83 | 4.91 | - | - | - | - |
| 23 | - | - | 0.99 | 5.50 | - | - | - | - |
| 24 | - | - | 0.95 | 5.17 | - | - | - | - |
| 25 | 0.28 | 1.65 | 0.66 | 4.21 | 0.31 | 2.09 | 0.46 | 3.37 |
| 26 | - | - | 0.73 | 5.43 | - | - | - | - |
| 27 | - | - | 0.91 | 5.37 | - | - | - | - |
| 28 | - | - | 1.06 | 4.78 | - | - | - | - |
| 29 | - | - | 1.14 | 4.50 | - | - | - | - |
| 30 | - | - | 1.16 | 3.91 | - | - | - | - |

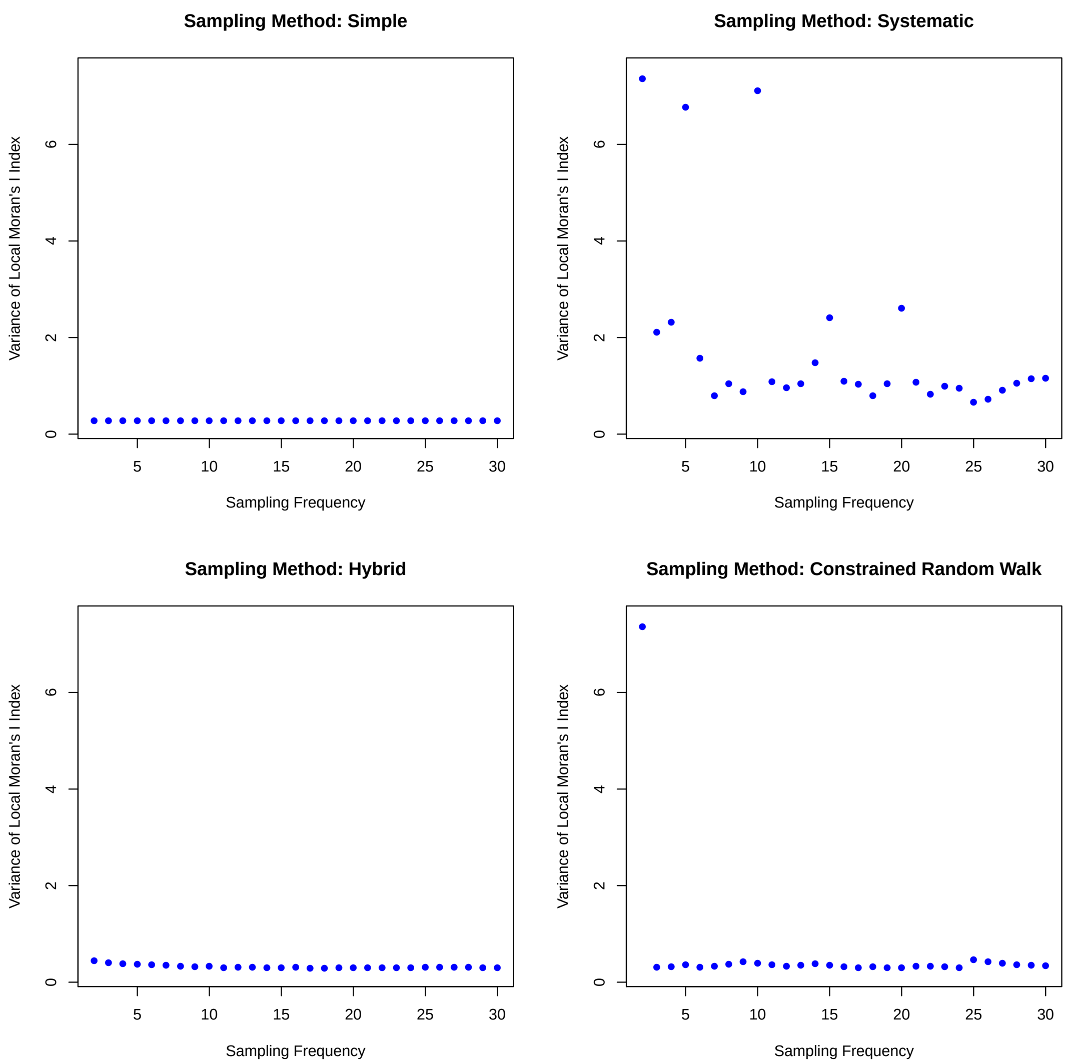


**Figure 4:** Series of plots showing population variance of Local Moran's $I_i$ statistics for the four different sampling methodologies, across the range of sampling widths.

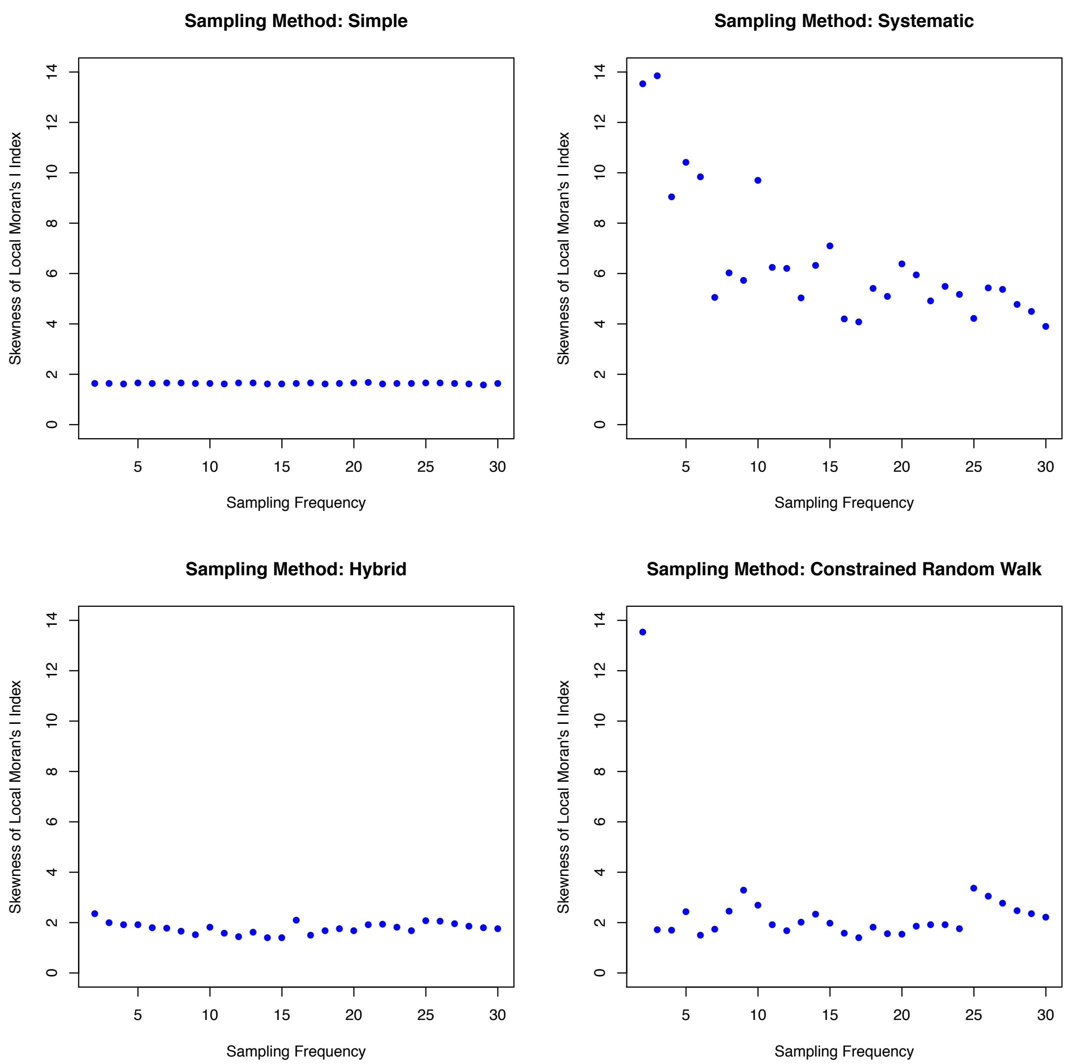


**Figure 5:** Series of plots showing distribution skewness of Local Moran's $I_i$ statistics for the four different sampling methodologies, across the range of sampling widths.

### 5.3 Computation Time Reduction and Methodology Recommendations

Recommendations for sampling strategy and sampling frequency can be set based upon chosen NRMSE tolerances; e.g., to guarantee, on average, $<=$ 2.5% error, the simple methodology only allows a frequency of 2 minutes. The systematic and hybrid methods allow frequencies of 2, 3, and 4 minutes under this error tolerance, while the random walk method allows frequencies of 2–5 minutes. For an NRMSE tolerance of $<=$ 5%, the simple method allows frequencies of 2–6 minutes; the systematic method allows frequencies 2–9; and the hybrid and walk methods allow frequencies of 2–10 minutes.

However, methods and frequencies yielding high variation in NRMSE values should also be avoided. Requiring that the Coefficient of Variation $\sigma/\mu <= 0.5$ is a reasonable tolerance; under this constraint, no viable sampling frequency is left for the systematic strategy with average error tolerance $<=$ 2.5%, while only frequencies of 6–9 minutes remain for the systematic strategy with average error tolerance $<=$ 5%. No other strategy-frequency combinations are eliminated by this constraint. Similarly, a constraint of avoiding spatial autocorrelation in NRMSE values could be imposed; choosing a framework which does not yield high variance and right-skewness in the block-level LISA statistics for NRMSE values would also result in eliminating many sampling frequencies under the systematic method, particularly those prone to eliciting harmonic effects. Table 5 gives an overview of which methodologies have low or high performance predictability (consistency throughout the sampling frequency range), and NRMSE and spatial autocorrelation statistics, based on the information in Table 3.

**Table 5:** NRMSE and LISA Summary Comparison of Sampling Strategies

| | **Sampling Strategy** | | | |
|---|---|---|---|---|
| | Simple | Systematic | Hybrid | Random Walk |
| Predictability | high | low | high | high |
| Mean of NRMSE | high | low | low | low |
| S.D. of NRMSE | high | high | low | low |
| Variance of LISA | low | high | low | low |
| Skewness of LISA | low | high | low | low |

Another consideration when choosing an appropriate sampling method and frequency is the overall computation time reduction obtained within the chosen framework. Figure 6 shows the tradeoff curves between mean NRMSE and computation speedup factor obtained, for each of the four sampling methods. The ratio of NRMSE to speedup factor is plotted across the range of sampling frequencies; diminishing returns are apparent. For example, moving from a sampling width of 5 to a sampling width of 10 for the simple method yields a 20.2 percent reduction in the NRMSE to speedup factor ratio, while moving from a sampling width of 10 to a sampling width of 15 only yields a 16.3 percent reduction in the NRMSE to speedup factor ratio. The systematic method once again shows spikes at sampling frequencies of 5, 10, and 15 minutes, clearly showing the error costs relative to speedup benefits associated with sampling at those frequencies. The diminishing returns for the three non-simple methods are lower than those for the simple method, indicating the reliability and consistency of the hybrid and constrained random walk methods in particular.

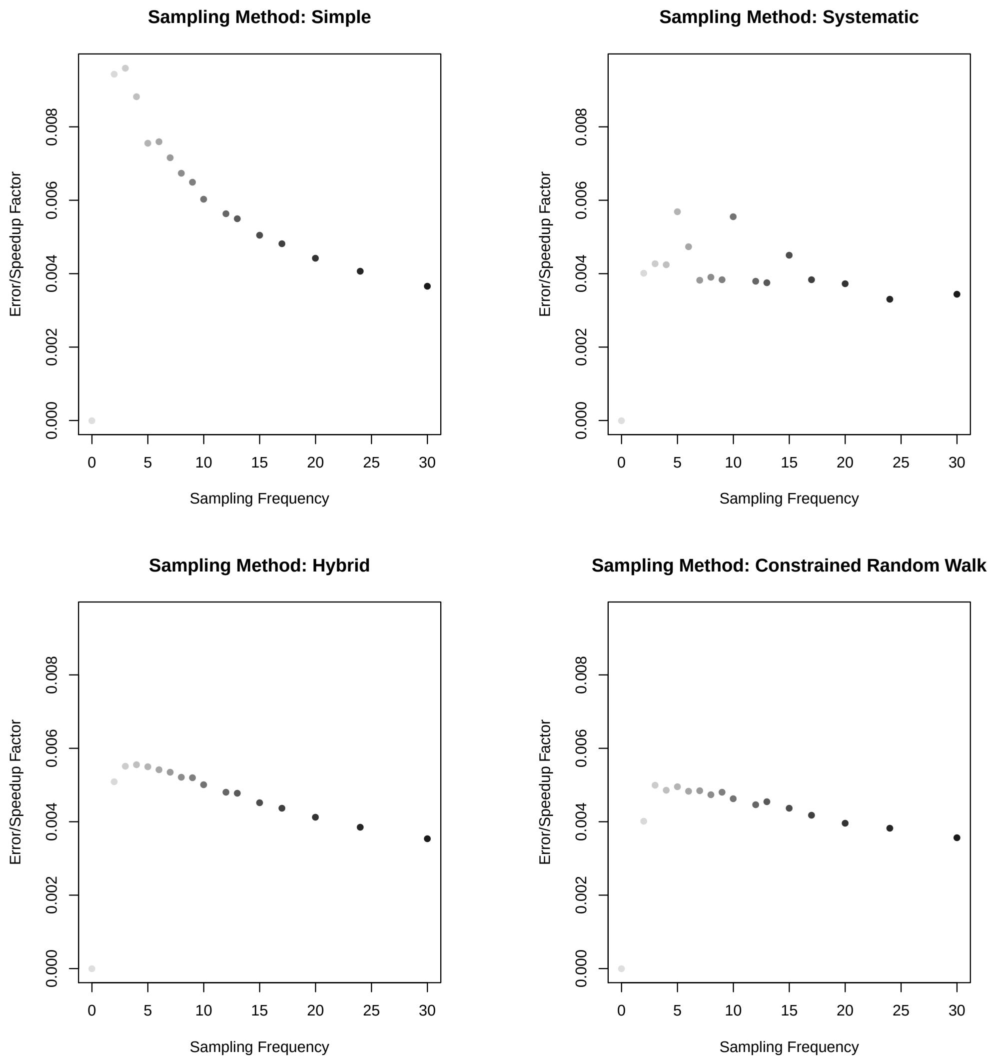


**Figure 6:** Series of plots showing the tradeoff between NRMSE and computational speedup, represented by the ratio of NRMSE to speedup factor (e.g., 2x, 3x, etc.), obtained via the four different sampling methodologies, across the range of sampling widths.

Computational speedup and NRMSE tolerances can be implemented in tandem to determine appropriate sampling methods. If for example a 5x speedup factor is desired (corresponding to a sampling frequency of 5 minutes) in tandem with a maximum NRMSE value of 2.5 percent, then only the constrained random walk method delivers the appropriate conditions. If a 10x speedup factor is desired with a maximum NRMSE of 5 percent, then both the hybrid and constrained random walk methodologies provide appropriate frameworks. However, it is important to note that while the ratios of NRMSE to speedup factor do decrease through the sampling frequency range for all methods, the NRMSE values themselves increase (Figure 3). It is this fundamental tradeoff between computation time speedups (and computer hardware limitations), and sampling error tolerances which informs the choice of sampling frequency and methodology.

## 6 Conclusion

It is clear that the selection of a temporal sampling strategy can have a significant impact on the results of accessibility evaluation, particularly if the chosen strategy does not avoid harmonic error interactions with the local transit network. Of the strategies compared in this analysis, a constrained random walk approach provided the best performance, as measured by sample error and variance, while avoiding harmonic error effects. However, it is important to note that this comparison relied on sample strategy performance in the context of a single transit network. It is possible that each strategy could perform better or worse if applied to the transit network in a different city, and additional research may be useful in finding a strategy that is generalizable.

It is also clear that a systematic sampling strategy can produce very erratic results compared to other methods, and that it is susceptible to harmonic error effects due to interactions with the local transit network. It may be wise to avoid this sampling strategy, particularly in applications with a goal of analyzing the spatial variation of accessibility. Because harmonic error effects are associated with nearby transit service, they are inherently clustered spatially and would bias any spatial analysis efforts. Sampling frequencies which are not simple divisors of 10- and 15-minute headway transit service (e.g., 3, 7, 9) may reduce the risk of harmonic error effects when a systematic sampling strategy is used. However, systematic sampling strategies yielded significantly higher population variances than those of the hybrid and random walk strategies, so the more robust frameworks may be more suitable for most transportation networks.

It may be interesting to explore sampling approaches which select different strategies and/or frequencies based on an analysis of the local transit service context. The systematic strategy gives the best performance in areas where harmonic error effects are not a concern; with minimal pre-processing of transit schedule data it may be possible to identify areas which are better suited to a particular sampling strategy or which would benefit from higher sampling rates. Transit networks can vary in terms of frequencies implemented, geographic layouts, service levels at different times of the day, and other characteristics, so application of this analysis framework to a different set of transit accessibility data may yield different results.

Temporal sampling strategies provide an attractive trade-off: an accessibility evaluation sampled at a 5-minute frequency using the constrained random walk strategy requires only 20 percent of the computing effort as one sampled at every minute, with an average sample error of only 2.5 percent. Perhaps this level of error tradeoff is acceptable to planning agencies looking to perform a large number of scenario evaluations, e.g., in designing a network of new bus rapid transit lines. However, sampling strategy and sampling frequency should be selected with an understanding of how they may influence the spatial patterns of accessibility results.

## 7 Acknowledgments

The authors appreciate the contributions of Yunis Adam and Haibing Jiang in data collection and processing.

## References

American Association of State Highway and Transportation Officials. 2015. Commuting in America 2013. Technical report, American Association of State Highway and Transportation Officials, http://traveltrends.transportation.org.

Anselin, L. 1995. Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2). ISSN 00167363. doi: 10.1111/j.1538-4632.1995.tb00338.x. 1011.1669.

Brewer, K. R. W. 1973. Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics*, 1:133–154.

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: Journal of the Transportation Research Board*, 2653:45–53. ISSN 0361-1981. doi: 10.3141/2653-06. URL http://trrjournalonline.trb.org/doi/10.3141/2653-06.

Cryer, J. D. and N. Kellet. 1986. *Time series analysis*, volume 101. Springer.

Delling, D., T. Pajor, and R. F. Werneck. 2014. Round-based public transit routing. *Transportation Science*.

Ding, Y., J. Zhou, Y. Li, *et al.* 2015. Transit accessibility measures incorporating the temporal dimension. *Cities*, 46:55–66.

Farber, S., M. Z. Morang, and M. J. Widener. 2014. Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, 53:149–159.

Geurs, K., A. Montis, and A. Reggiani. 2015. Recent advances and applications in accessibility modelling. *Computers, Environment and Urban Systems*, 49:82–85.

Geurs, K. T. and B. van Wee. 2004. Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, 12(2):127–140.

Gupta, S. and J. Shabbir. 2008. On improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics*, 35(5):559–566. ISSN 02664763. doi: 10.1080/02664760701835839.

Hansen, W. G. 1959. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76.

Ingram, D. R. 1971. The concept of accessibility: A search for an operational form. *Regional Studies*, 5(2):101–107.

Kac, M. 1947. Random walk and the theory of brownian motion. *The American Mathematical Monthly*, 54(7P1):369–391. doi: 10.1080/00029890.1947.11990189. URL https://doi.org/10.1080/00029890.1947.11990189.

Kadilar, C. and H. Cingi. 2005. Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters*, 19(2006):75–79. ISSN 08939659. doi: 10.1016/j.aml.2005.02.039.

Karner, A. 2015. Development of highly resolved spatial and temporal metrics of public transit accessibility 1 and their application to service equity analysis 2. In *Transportation Research Board 94th Annual Meeting*, 15-5677.

Kujala, R., C. Weckström, M. N. Mladenović, and J. Saramäki. 2018. Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys. *Computers, Environment and Urban Systems*, 67:41–54. ISSN 01989715. doi: 10.1016/j.compenvurbsys.2017.08.012.

Legrain, A., R. Buliung, and A. M. El-Geneidy. 2015. Who, what, when, and where: Revisiting the influences of transit mode share. *Transportation Research Record: Journal of the Transportation Research Board*.

Lei, T., Y. Chen, and K. Goulias. 2012. Opportunity-based dynamic transit accessibility in southern california: Measurement, findings, and comparison with automobile accessibility. *Transportation*

*Research Record: Journal of the Transportation Research Board*, (2276):26–37.

Levy, P. S. and S. Lemeshow. 2011. Chapter 3: Simple random sampling. In *Sampling of Populations: Methods and Applications: Fourth Edition*, pp. 43–81. Wiley. ISBN 9780470374597. doi: 10.1002/9780470374597.ch3. arXiv:1011.1669v3.

Madow, W. G. and L. H. Madow. 1944. On the theory of systematic sampling, I. *The Annals of Mathematical Statistics*, 15(1):1–24. URL https://www.jstor.org/stable/2236209.

McFadden, D., M. Johnson, and F. Reid. 1973. *Travel Demand Forecasting Study*. Institute of Urban and Regional Development, University of California.

Owen, A. and D. M. Levinson. 2015. Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, 74:110–122.

Owen, A., B. Murphy, and D. Levinson. 2016. Access across America: Transit 2015 methodology. Technical Report CTS 16-10, University of Minnesota Center for Transportation Studies, http://access.umn.edu/research/america/transit/2015.

Spitzer, F. 1964. *Principles of Random Walks*. Princeton, New Jersey: D. Van Nostrand, Inc.

Stepniak, M., J. P. Pritchard, K. T. Geurs, and S. Goliszek. 2019. The impact of temporal resolution on public transport accessibility measurement : Review and case study in Poland. *Journal of Transport Geography*, 75(July 2018):8–24. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2019.01.007. URL https://doi.org/10.1016/j.jtrangeo.2019.01.007.

Tasic, I., X. Zhou, and M. Zlatkovic. 2014. Use of spatiotemporal constraints to quantify transit accessibility: Case study of potential transit-oriented development in West Valley City, Utah. *Transportation Research Record: Journal of the Transportation Research Board*, (2417):130–138.

Weiss, A. A. 1984. Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, 26(3):271–281. ISSN 03044076. doi: 10.1016/0304-4076(84)90022-8. 0023.

Widener, M. J. 2017. Comparing measures of accessibility to urban supermarkets for transit and auto users. *Professional Geographer*, 69(3):362–371. ISSN 14679272. doi: 10.1080/00330124.2016.1237293.

Widener, M. J., L. Minaker, S. Farber, J. Allen, B. Vitali, P. C. Coleman, and B. Cook. 2017. How do changes in the daily food and transportation environments affect grocery store accessibility? *Applied Geography*, 83:46–62. ISSN 0143-6228. doi: 10.1016/j.apgeog.2017.03.018. URL http://dx.doi.org/10.1016/j.apgeog.2017.03.018.