Amendment letter

Luyu Liu and Harvey J. Miller[[1]](#footnote-1)\*

Department of Geography and Center for Urban and Regional Analysis

The Ohio State University, Columbus, Ohio, USA

We really appreciate more useful comments from reviewer 1 and we are also glad to see we solved all the questions from reviewer 2. In the following paragraphs, we will thoroughly respond the new comments and questions.

**Reviewer 1**  
  
The authors have addressed all of my comments. The contribution of the paper is now more clearly articulated. I think, some minor improvements could be made to further enhance links with the journal’s themes.

1. The distinction between non-real time (or static) and real-time measures offers an interesting angle. I wonder if authors could expand on this further, also to highlight how these new data sources are transforming the urban transport sector, and what opportunities, obstacles and challenges this entails.

Response: Real-time measures, or measures/data sources with higher temporal accuracy, have many advantages over their counterparts with lower temporal accuracy. We listed two most prominent points and incorporated them into section 2.2, subsection “Real-time measures”, paragraph 2:

1. The measured time is closer to the actual time. This is very important when it comes to temporal analysis. If the recorded time is not the actual time or has large difference, the results can have considerable systematic error during the analysis.
2. The measures have higher trueness. Not only is the recorded time more accurate, we are also more confident in the measured value. Unless the measured time and the actual event time share the exact same distribution and mean, which is very unlikely to happen, it is generally true that higher temporal measures have higher trueness.

In short, real-time measures have higher authenticity.

In terms of challenges, there are several additional requirements to obtain real-time big data:

1. Real-time measures requires more responsive in-situ censors, which take additional technological and economic costs. The censors should be installed on the bus (in-situ) and the total cost for entire system coverage is considerably high; the censors should have multiple functionalities, such as detecting location, determine which stop the bus is at, and calculate time.
2. Real-time measures requires real-time data, which needs data supports such as standard format and data streaming pipeline. These infrastructures will take more costs.
3. For high-frequency events such as buses’ departure and arrival time at each stop, to keep up with the events’ occurrence, the data will have a very large volume and high velocity. We discussed this challenge in the Big Data section.
4. Real-time data usually requires the administrative support to be able to connect the censors with the bus.

We incorporated this part into the section 2.2, subsection “real-time measures”, paragraph 2.

1. One point that would be worth clarifying is the notion and value of temporal ‘accuracy’. In which way do static measures systemically deviate from actual time? Or is the issue rather temporal precision? I believe more discussion of the new, dynamic measures would be valuable to increase the noticeably expanded yet still limited engagement with the journal’s theme.

Response: we also considered the same issue when we were conceptualizing the concept in the paper, and finally we decided to use “accuracy” instead of “precision”. According to the ISO definition, accuracy describes a combination of both types of observational error, including random and systematic, while precision only covers random errors. In other words, accuracy refers to closeness of the measurements to a specific value, while precision refers to the closeness of the measurements to each other.

As for the temporal accuracy, the error it measures includes systematic error: due to the delay of measuring, the time of measured value is always later than the time of actual event. This is not a random error since it is one-sided. This is the main reason why we decided to name it temporal accuracy.

1. On p.15 ln.8, small data are still associated with ‘deliberate’, but as the examples of AVL or GTFS show, these are big yet deliberate.

Response: We apologize for the neglect, since all “deliberate” should be removed in the last draft. We deleted the word in this section, for we want to shift the focus from “deliberate versus byproduct” to “small versus big”.

1. The stated drawbacks of small data could be revisited slightly given that such data sources are often carefully sampled (p.15).

Response: This is a good point. We added this to the text in section 2.1, paragraph “Small data”.

1. P.31 ln 13, I wasn’t sure if the authors refer to the frequency of the ‘generating’ service or the ‘receiving’ service. I would expect correlations between ‘transfer risk’ and the receiving service. Perhaps the authors could clarify which type of service they refer to.

Response: This is a good question. We are using the average frequency of both generating and receiving services. We believe the correlation results between transfer risk and the receiving services is close, since at each stop each receiving trip in a transfer can also be a generating trip in another transfer. For example, transfer A consist of generating trip *a* (route 1) and receiving trip *b* (route 2); while transfer B consists of generating trip *b* andreceiving trip c (an later trip for route 1); and for most trips (except for the earliest and latest one), there are at least two counterparts: one with itself as generating trip and the other with itself as receiving trip. Therefore, compared with calculating the headway/frequency of generating/receiving trips alone, we are merely counting twice if we calculate the total average, and the average will stay the same. Figure 1 shows the relationship between all transfer trips and only receiving trips. The figure proves that the value of two headways calculations are extremely close. Therefore, we can say it will not change the results, which is especially true for hourly and daily headways correlation analysis since the data points is very few.

Figure 1 the relationship between hourly headway of all transfer trips and only receiving trips.

1. P.31 ln 38, I couldn’t follow why ‘transfer risk’ increases at night. I would expect that it reduces, in particular due to less traffic.

Response: Yes, this is surely anti-intuitive. However, this will make more sense when we consider that the bus trips in operation are also changing. During the night, the percentage of transfers on the High Street (the major north-south thoroughfare in Columbus, indicated by a red circle in Figure 2 in the main text) is increasing since other bus trips involved with transfers will not run in late hours. As we introduced in the paper, the risk of missed transfer or transfer risk is relatively high on this road. Since these highly missed buses on the High Street are basically only buses in operation during the late night hours, the risk of missed transfers is surely higher.

1. On ‘transfer risk’, I still find the term somewhat misleading. Perhaps a more accurate term could be found, such as ‘risk of failed transfers’.

Response: Per reviewer 1’s suggestion, we changed the term from “transfer risk” to “risk of missing transfers” and changed “TR” to “RoMT”. Please see the corresponding highlighted words in the paper. We believe the new term will be less misleading for new readers.

**Editor comments:**

1. In P. 18, line 8, the comment says that “not clear how that could be measured or why of interest in a node-specific context”.

Response: According to Guo and Wilson (2011), transfer cost comes from two sources: the opportunity cost of extra time or money spent on transfers that otherwise could be spent on work or leisure, and the disutility of the transfer itself, the transfer penalty. Accordingly, we changed this sentences to “they develop an index that measures each transfer node’s effectiveness based on additional transfer time and fare and apply it to the London Underground system.”

1. In P. 18, line 42, the comment says that “this is clumsy, can it be better expressed”.

Response: we changed the start of this paragraph to “On-board questionnaires can be considered as the first real-time measure, however, they are not widely used to study transfers due to the difficulties of acquiring abundant data. After the emergence of big data, automated big data real-time measures based on first-hand information with *high velocity* and *large volume* create new opportunities for transfer studies.”

1. In P. 29, line 48, the comment says that “Suggest you use "ellipse" [rather than "circle"] and just "rectangle", and delete "red" and Green" - unless you intend to pay for colour figures in hard copy. ”

Response: we made the changes according to this very useful suggestion. We also made several changes on the graphs so that they can be more distinguishable in greyscale.

1. In the last paragraph of the paper, several comments pointed out that this paragraph could be improved.

Response: we rewrote the whole paragraph and now it should be better. We start from potential future research directions and end the paragraph by listing the limitations of the papers.

1. \* Corresponding author e-mail: [miller.81@osu.edu](mailto:miller.81@osu.edu)

   First author e-mail: [liu.6544@osu.edu](mailto:liu.6544@osu.edu) [↑](#footnote-ref-1)