

Empirically quantifying city-scale transportation system resilience to extreme events

Brian Donovan¹, Daniel B. Work^{2,*}

Abstract

This article proposes a method to quantitatively measure the resilience of transportation systems using GPS data from probe vehicles such as taxis. The granularity of the GPS data necessary for the method is relatively coarse; it only requires coordinates for the beginning and end of trips, the metered distance, and the total travel time. The method works by computing the historical distribution of pace (normalized travel times) between various regions of a city and measuring the pace deviations during an unusual event. Periods of time containing extreme deviations are identified as events. The method is applied to a dataset of nearly 700 million taxi trips in New York City, which is used to analyze the city transportation infrastructure resilience to Hurricane Sandy. The analysis indicates that Hurricane Sandy impacted traffic conditions for more than five days, and caused a peak delay of two minutes per mile. Practically, it identifies that the evacuation announcements coincided with only only minor disruptions, but significant delays were encountered

*Corresponding author

¹B. Donovan is a graduate student in the Department of Civil and Environmental Engineering, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA.

²D. Work is an Assistant Professor in the Department of Civil and Environmental Engineering, and Coordinated Science Laboratory, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA. Email: dbwork@illinois.edu

Preprint submitted to Elsevier

February 7, 2017

during the post-disaster response period. Since the implementation of this method is very efficient, it could potentially be used as an online monitoring tool, representing a first step toward quantifying city scale resilience with coarse GPS data.

1. Introduction

1.1. Motivation

In recent years, many quantitative and qualitative methods have been proposed to determine resilience of transportation infrastructure (Matherly and Langdon (2014); Faturechi and Miller-Hooks (2014); Konstantinidou et al. (2014a,b)). When disasters and other extreme events occur, critical infrastructure may fail, incurring large human, economic, and environmental costs. This is especially relevant for urban transportation infrastructure, since it is crucial for city evacuations and emergency services in post-disaster environments. Empirical methods are needed to quantitatively monitor the transportation infrastructure in terms of its ability to withstand and recover from such events.

The goal of this article is to develop and implement a method for measuring resilience of city-scale transportation networks using only publicly available GPS data (e.g., available from taxis). The technique is designed with the following characteristics. First, the method can be applied at the city-scale, or larger. Because extreme events such as hurricanes have the ability to affect an entire city, it is important to examine impacts at a high-level city view, rather than the level of individual vehicles or streets. Second, the method measures network performance quantitatively, in terms of recovery time and peak pace deviations. Recovery time and peak performance degradation are standard quantities of interest in the resilience literature (Aven (2011); Haines (2009b)). While travel times are a natural performance measure for transportation networks, we instead use *pace*, or travel time per mile. This normalization accommodates the varied length of taxi

trips within a city. Third, the method accommodates inherent variability in traffic conditions and data. The estimate of the state of traffic in a city contains noise and errors due to data availability and many unmodeled human factors. As a result, the method evaluates events that cause statistically significant disruptions, in order to separate the signal from the noise. Finally, the method is computationally tractable. Since taxi trips occur very frequently in large cities, the amount of data available for analysis is large. In order to be tractable, the computation should be $O(N)$, where N is the number of taxi trips, and ideally require only one pass through the raw data. Of practical significance, these single-pass algorithms could also be used to process the data in a realtime stream.

1.2. Related Work

In recent years, the study of resilience has gained popularity in the systems engineering community. Haines (2009a,b, 2011) gives a framework for assessing resilience, which focuses on modeling a system and the possible outcomes of various events. He asserts that a resilient system should suffer only slight degradation during an event, then rapidly recover. Reed et al. (2009) note that the quality of service abruptly drops during an event, then exponentially decays back to typical values. They suggest that an appropriate resilience measure is the integral of this exponential curve. Authors in the related field of risk analysis emphasize the importance of unknown factors while assessing resilience (Aven (2011); Kaplan and Garrick (1981)).

Though there is no precise consensus on the definition of resilience, *peak disruption* and *recovery time* are consistently discussed quantities. In other words, peak disruption measures how far the quantity of interest deviates

from typical values, and recovery time measures how long it takes to return to typical values. Most of these works also emphasize that resilience must be measured with respect to a given event and quantity of interest. For example, one case study used the number of functioning nodes in a power grid as the quantity of interest, assessing resilience against hurricanes and minor events (Ouyang et al. (2012)).

Several authors have proposed quantities of interest for transportation systems. Omer et al. (2013) proposed a method which measures the resilience of a road-based transportation network in terms of travel times between cities. Chang and Nojima (2001) evaluated a post-earthquake transportation network in terms of accessibility and coverage. This is partly based on an accessibility metric devised by Allen et al. (1993), which considers travel times between various regions of a city. Thus, travel time is a standard quantity on which to measure resilience. This article will use the related quantity of pace, or travel time per mile. A comprehensive set of measures for transportation system resilience can be found in the review by Faturechi and Miller-Hooks (2014). The interested reader is directed to the related reviews on evacuation modeling (Murray-Tuite and Wolshon (2013)) and post disaster planning and management (Konstantinidou et al. (2014a,b)) for a more complete picture of transportation system resilience, evacuation, and post-disaster response.

A distinct set of studies use large amounts of data to extract useful information about urban systems. The works most closely related to resilience are the studies by He and Liu (2012) and Zhu et al. (2010), which measure the effect of the I-35W bridge collapse in Minneapolis in 2007. Geroliminis

and Daganzo (2008) use loop detector data, combined with 500 GPS vehicles to extract macroscopic traffic properties from an urban-scale transportation network. Other works use GPS traces of mobile devices to analyze movement patterns of crowds during typical days and atypical events (Calabrese et al. (2010, 2011)). Castro et al. (2012) present a method for inferring current and future traffic states from taxi GPS data. Zheng et al. (2011) propose a method that tracks taxi trips between various regions of a city and identifies flawed urban planning, while Zhan et al. (2016a) empirically measure the (in)efficiency current taxi systems. Another study measures temporal patterns in the density of taxi pickups and dropoffs to identify the social function of various city regions (Qi et al. (2011)). They point out that unusual output can be used to detect events like holidays. Chen et al. (2012) specifically focuses on identifying anomalous taxi trajectories, in order to detect fraud or special events. Ferreira et al. (2013) created a graphical querying tool which can be used to count taxi trips between arbitrary geometrical regions as a function of time. They noted the drop in the frequency of taxi trips during Hurricane Sandy and Hurricane Irene, pointing out that the Irene-related drop was more significant, but the Sandy-related drop was longer lasting. By examining pace, we confirm that Hurricane Sandy had a longer recovery time, but find the contrasting result that Hurricane Sandy also has a more significant peak disruption.

1.3. Outline and Contributions

The contributions of this work are as follows. In Section 2, a method is proposed to use taxis as pervasive city-scale resilience sensors. This method detects unusual events and measures them in terms of peak disruption and

recovery time. It introduces paces between regions of the city as the key performance measure, and it uses the historical pace distribution to detect and measure extreme events. In Section 3, the method is applied to a four-year dataset from New York City to identify and compare properties of events such as Hurricane Sandy. Of practical significance, the analysis identifies minor atypical traffic is observed pre-Sandy, contrasted with the extreme conditions observed post-Sandy. Conclusions and future work is summarized in Section 4. As a technical contribution, all data (Donovan and Work (2014)) and source code³ used in this analysis is made publicly available.

2. Methodology

2.1. Overview

The proposed technique to measure city-scale resilience of the transportation network in response to various events by examining taxi trip data is constructed in three steps. In section 2.2, individual taxi trips are aggregated by origin-destination pairs in order to measure typical paces between various regions of the city. This aggregation technique makes it possible to extract city-scale features at various points in time, since it is difficult to measure resilience from individual trips. Section 2.3 imposes a one-week periodic pattern on the paces. Defining the variance across weeks as a probability distribution allows for a description of typical behavior that accounts for noisy day-to-day variations. Finally, Section 2.4 uses these distributions to quantify how typical or atypical the pace is at a particular point in time.

³source code available at: <https://github.com/Lab-Work/gpsresilience>.

Atypical times (e.g., the 5% most unlikely points in time) are flagged as events, and they are examined in more detail.

2.2. Extraction of Time-Series Features from Aggregated Trips

In the first stage of analysis, trips are grouped by their geographic locations and times of occurrence. More specifically, the city is divided into k regions, for example defined by census tracts or neighborhoods. This allows each taxi trip to be labeled as belonging to one of k^2 unique origin-destination pairs. Time is discretized into hours, so a large sample of trips can be gathered at any point in time. The start zone, end zone, and departure time are used to partition all of trips into subsets. The variable $T_{i,j,t}$ denotes the set of all trips from zone i to zone j at time t , where t is an absolute time in hours with the convention that $t = 0$ refers to the first hour of the dataset:

$$T_{i,j,t} = \{r | o(r) \in z(i), d(r) \in z(j), \lfloor s(r) \rfloor = t\}. \quad (1)$$

In (1), $o(r)$ is the origin of trip r , $d(r)$ is the destination of trip r , $z(i)$ is the geographic region of zone i , and $\lfloor s(r) \rfloor$ is the start time of trip r rounded down to the hour using the *floor* operator $\lfloor \cdot \rfloor$. The start time of the trip $s(r)$ is recorded by the taximeter, and is typically reported to the nearest minute. It is assumed that i and j are both in $\{0, 1, \dots, k-1\}$. Once these subsets of trips are defined, macroscopic traffic features can be extracted from them. Of particular interest is the expected travel time between two regions. However, travel times of individual vehicles between the same two regions are not equal, due to the varying lengths of trips that connect the regions. Much of the travel time variation can be accounted for by normalizing against the

trip distance before averaging the trip, which results in a quantity called the *average pace* (Daganzo (1997); Saberi et al. (2014)).

On a single road segment, the average pace of traffic is equal to the inverse of the Edie's generalized average speed (Edie (1963); Daganzo (1997)), which satisfies the relation that the flow is the product of the density and speed. When derived at the network scale, the pace $P(i, j, t)$, of taxis from zone i to zone j at time t is computed as the ratio of the total duration of all trips to the total distance of all trips as follows:

$$P(i, j, t) = \frac{\sum_{r \in T_{i,j,t}} u(r)}{\sum_{r \in T_{i,j,t}} l(r)}, \quad (2)$$

where $u(r)$ is the travel time of trip r and $l(r)$ is the metered length of trip r . When calculated according to (2), the pace is simply the inverse of the speed as defined for the macroscopic (network) fundamental diagram (Saberi et al. (2014)).

Note also that the average pace (2) can be interpreted as a weighted average of the individual trip paces:

$$P(i, j, t) = \frac{\sum_{r \in T_{i,j,t}} u(r)}{\sum_{r \in T_{i,j,t}} l(r)} = \frac{\sum_{r \in T_{i,j,t}} l(r) \frac{u(r)}{l(r)}}{\sum_{r \in T_{i,j,t}} l(r)} = \frac{\sum_{r \in T_{i,j,t}} l(r) p(r)}{\sum_{r \in T_{i,j,t}} l(r)}, \quad (3)$$

where $p(r) = \frac{u(r)}{l(r)}$ is the pace of trip r .

For a fixed value of t , all k^2 distance-weighted average paces collectively form the *mean pace vector*, denoted $\mathbf{a}(t) \in \mathbb{R}^{k^2}$. This vector is a function of time, and contains the k^2 pace values at a particular point in time. Specifi-

cally, the n th element is given by

$$\mathbf{a}(t)_n = P \left(\left\lfloor \frac{n}{k} \right\rfloor, n \bmod k, t \right), \quad (4)$$

where $n \in \{0, 1, 2, \dots, k^2 - 1\}$.

It is desirable to use pace as the performance metric instead of the more traditional measure of vehicle counts, since the goal is to measure traffic conditions during extreme events. If the flow of vehicles between two regions drops significantly, it is difficult to determine whether this is due to increased congestion or decreased demand. However, an increase in pace indicates congestion, while a decrease in pace indicates decreased demand. Although the pace of taxis might be a biased estimate of the pace of all vehicles, it is assumed that if taxis are stuck in traffic jams, so are the other vehicles around them. In summary, we follow a standard assumption that the taxi travel times are sufficiently representative to infer the true traffic conditions on the roadway (Zhan et al. (2013)). Practical validation of this assumption is difficult due to limited urban sensing infrastructure or the public availability of large GPS datasets for personal vehicles.

2.3. Identification of City-Scale Typical Behavior

The mean pace vector, $\mathbf{a}(t)$, has a strongly periodic weekly pattern. During rush hour, the pace is high, especially in dense downtown regions, and at night the pace is low. On weekends, the rush hour is less extreme. However, the mean pace vector has some variance around this periodic pattern, so it is viewed as a distribution conditioned on time. For example, the mean pace vector for all Tuesdays at 3pm will be different, with larger derivations during an unusual event. To facilitate this grouping, the *reference set* Q_t is

defined for all times t . This set contains all of the mean pace vectors which occur at the same point in the periodic pattern as $\mathbf{a}(t)$, except for $\mathbf{a}(t)$ itself. Intuitively, when deciding how typical the traffic data is at time t , that data should not be used as part of the definition of typical. Since there are 168 hours in a week, the reference set can be defined as

$$Q_t = \{\mathbf{a}(h) | h \equiv t \pmod{168}, h \neq t\}. \quad (5)$$

The reference set Q_t makes it possible to compute the expected value of the mean pace vector $\mu(t)$ as well as the covariance matrix $\Sigma(t)$. This covariance matrix is important because it quantifies the noisy day-to-day fluctuations in the mean pace vector, outside of the event at hand, and how the dimensions correlate. The time-dependent sample mean and covariance matrices can be defined as:

$$\begin{aligned} \mu(t) &= \frac{1}{|Q_t|} \sum_{\mathbf{a} \in Q_t} \mathbf{a} \\ \Sigma(t) &= \frac{|Q_t|}{|Q_t|-1} \sum_{\mathbf{a} \in Q_t} \left(\frac{\mathbf{a}\mathbf{a}^\top}{|Q_t|} - \mu(t)\mu(t)^\top \right). \end{aligned} \quad (6)$$

Note the presentation of the nonstandard but equivalent calculation of the sample covariance matrix is used for computational benefits when working with datasets too large to fit into memory, as discussed later in Section 3.2. If an independence assumption is desired, the diagonal components of these matrices can be extracted. However, it is likely that many of the k^2 dimensions of $\mathbf{a}(t)$ are highly correlated, so the full covariance matrix is used for the remainder of the analysis. For example, trips that start or end in the same region often have highly correlated paces. Together, $\mu(t)$ and $\Sigma(t)$ define the first two moments of a probability distribution, conditioned on time.

2.4. Detection of Deviations from Typical Behavior

Intuitively, $\mu(t)$ captures the expected traffic conditions at a particular point in time. If the observed traffic conditions are significantly far from this expectation, then those conditions are classified as an extreme event. The covariance matrix $\Sigma(t)$ is also considered; if there is typically very little deviation from $\mu(t)$, then a large deviation is even more extreme. In one dimensional cases, this is typically addressed by standardizing the data via a z-score. In higher dimensions, the generalized z-score is called the Mahalanobis distance Mahalanobis (1936). For this analysis, the Mahalanobis distance for an observed mean pace vector is viewed as a function of the time that the observation occurred:

$$M(t) = \sqrt{(\mathbf{a}(t) - \mu(t))^\top \Sigma(t)^{-1} (\mathbf{a}(t) - \mu(t))}. \quad (7)$$

This time-dependent Mahalanobis distance serves as an outlier score for observations at various points in time. Note that it normalizes the deviations in each dimension by the corresponding variances, and also considers correlations between dimensions. The Mahalanobis distance is a natural way of measuring outliers in multivariate normal data, and it has shown to be useful even when the data is not normal (Warren et al. (2011)). In fact, the multivariate generalization of Chebyshev's inequality gives an upper bound on the probability of observing a Mahalanobis distance greater than some fixed value (Navarro (2014)). In other words, it is unlikely to observe a datapoint with a high Mahalanobis distance, regardless of the distribution. So, when $M(t)$ rises above a given threshold, an unusual event is detected. The event is declared complete when $M(t)$ returns to a value lower than the threshold.

Event Detection – Threshing

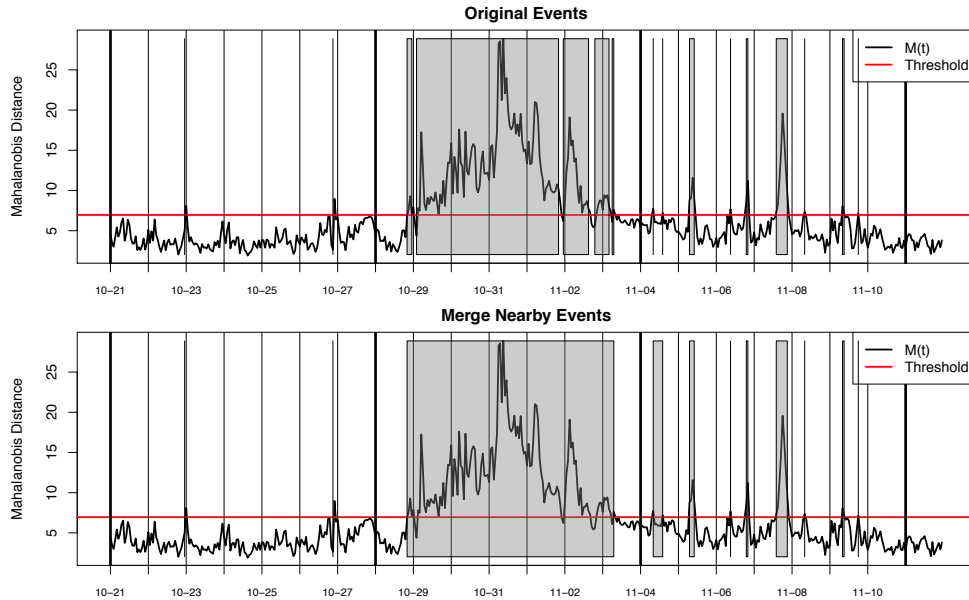


Figure 1: Demonstration of event detection. Events are detected when $M(t)$ goes above the threshold, but thrashing often occurs. The top graph shows that this thrashing causes events to be divided into several pieces. For this reason, events with fewer than six hours between them are merged, as shown in the bottom graph.

In this work, the choice of the threshold is the 95% quantile of $M(t)$, but this value can easily be lowered to detect smaller events or raised to detect only the most severe events.

The function $M(t)$ is a fairly noisy, which means that it can occasionally *thrash* over the threshold. In other words, $M(t)$ may rise above the threshold, then immediately drop back below it, effectively breaking the event into two pieces. To prevent this, consecutive events separated by fewer than six hours are merged. Figure 1 illustrates this process.

Once the recovery time of an event is computed, other properties can be computed. For example, it is possible to compute the maximum pace deviation, or the slowest type of trip during the event. Thus, each event can be described with a set of meaningful statistics. Comparisons between various events make it possible to describe which types of events the city can easily endure, and where there is room for improvement. For longer-lasting events like Hurricane Sandy, it is possible to examine different stages of the event in greater detail.

3. Application to Hurricane Sandy with New York City Taxi Data

In this section, the previously described methodology is applied to a dataset of New York City taxi trips. This dataset, which was obtained through a *Freedom of Information Law* (FOIL) request, covers four years of operation and details nearly 700 million trips. Many events are detected within this four year span and compared quantitatively. Special attention is given to the Hurricane Sandy and some interesting properties are discovered.

3.1. The Dataset

The New York City data used in this work details every yellow taxi trip that occurred in the city between 2010 and 2013 inclusive. The dataset contains 697,622,444 trips, which requires 180 GB to store in a PostgreSQL database. Each trip records the pickup and dropoff dates, times, and coordinates, as well as the metered distance reported by the taximeter. The trip data also includes fields such as the taxi medallion number, fare amount, and tip amount which are not relevant to the present analysis. Table 1 contains a small subset of this data for reference, and the full dataset is made publicly

| pickup datetime | dropoff datetime | duration (sec) | distance (mi) | pickup lon | pickup lat | dropoff lon | dropoff lat |
|------------------------|------------------------|-------------------|------------------|---------------|---------------|----------------|----------------|
| 2013-05-01 00:02:11 | 2013-05-01 00:14:28 | 737 | 2.9 | -74.00 | 40.74 | -74.01 | 40.71 |
| 2013-05-01 00:02:12 | 2013-05-01 00:12:31 | 618 | 1.8 | -74.00 | 40.73 | -73.98 | 40.72 |
| 2013-05-01 00:02:12 | 2013-05-01 00:07:39 | 326 | 1.3 | -73.97 | 40.76 | -73.96 | 40.77 |
| 2013-05-01 00:02:13 | 2013-05-01 00:04:35 | 141 | 0.6 | -73.99 | 40.75 | -74.00 | 40.75 |
| 2013-05-01 00:02:14 | 2013-05-01 00:04:09 | 115 | 0.5 | -73.98 | 40.75 | -73.99 | 40.74 |

Table 1: A small subset of the data used in this analysis. Each row corresponds to an occupied taxi trip.

available (Donovan and Work (2014)) in CSV format with this manuscript. The dataset only includes trips occupied by a passenger, and does not include (empty) roaming trips in search of passengers.

Since the data contains only two coordinates for each trip, the quality is lower than GPS data collected by traffic monitoring companies, which contain the location of vehicles at periodic rates up to once per second. In spite of the lack of intermediate waypoints in the data, the methods presented in this

article are able to analyze the city-scale resilience of the traffic conditions.

As with many large datasets, the taxi trip data contains errors. Some errors are easy to identify, such as pickup or dropoff coordinates of $(0, 0)$, or trips with metered distances in excess of one million miles. Other errors are identified by examining the consistency of redundant information contained in each trip. For example, the *winding factor* is computed as the metered distance divided by the straight-line distance between the pickup and dropoff coordinates. Trips with a winding factor less than one are geometrically impossible, although values slightly less than one are permissible due to errors in the GPS positioning data.

Other trips are discarded because they are not informative to the analysis of traffic conditions. For example, a winding factor of five or more is technically possible, but it indicates that the driver did not proceed directly to the destination. These trips are ignored because their paces might be corrupted by the traffic conditions in other unknown regions. Table 2 shows the thresholds used in each filter as well as the quantity of trips that violate the threshold. Figure 2 shows the distribution of the trip data for each filtering metric. Overall, 11% of the data is discarded in this step. Other data reconciliation steps not detailed here (e.g., unit conversions) allowed additional erroneous data records to be cleaned and used in the analysis.

3.2. Computational Issues

Due to the size of the dataset, an efficient software implementation of the analysis is crucial. This section discusses the algorithmic and practical aspects of the analysis, using the New York City taxi dataset as an example. In this way, concrete figures can be used for quantities like runtime or

| Feature | Lower Threshold | Upper Threshold | Portion of Errors |
|----------------------------------|-----------------|-----------------|-------------------|
| Latitude (Degrees) | 40.65 | 40.9 | 3.61% |
| Longitude (Degrees) | -74.05 | -73.85 | 3.91% |
| Straightline Distance (Miles) | 0 | 8 | 3.18% |
| True Distance (Miles) | 0 | 15 | 1.97% |
| Winding Factor | 0.95 | 5 | 6.04% |
| Duration (Minutes) | 1 | 60 | 0.89% |
| Pace (Min/Mi) | 0.667 | 60 | 0.45% |

Table 2: The error thresholds that are applied to various features of the dataset. The last column shows the percentage of trips that violate each threshold. Some trips may violate multiple feature thresholds.

Data Filtering

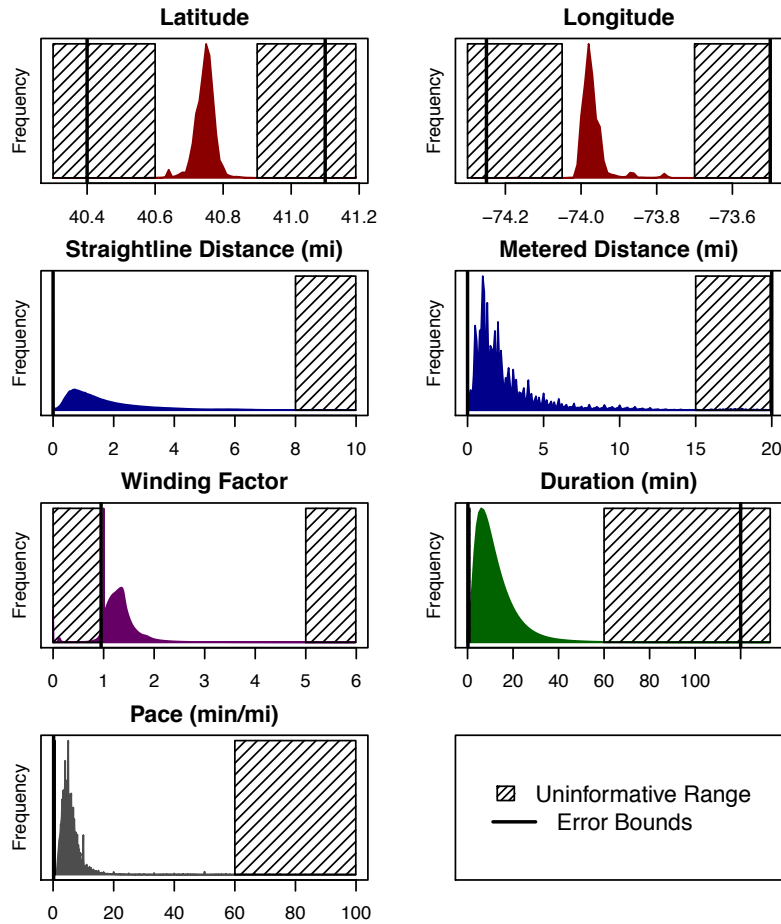


Figure 2: Distributions of individual features of taxi trips. Simple thresholds are used to filter trips that contain errors, or are otherwise uninformative. Note that the *straightline distance* is the Euclidean distance between start and end coordinates, while the *metered distance* is the value returned by the odometer. The *winding factor* is the metered distance divided by the straightline distance. A winding factor less than 1 is geometrically impossible, and a large value indicates that the taxi did not proceed directly to its destination.

data size. More general concepts like time complexity do not depend on the dataset.

The overarching concern when performing any analysis on the dataset is due to its size (i.e., 180GB), which is too large to store in memory of most commodity computers available today. Consequently all calculations must be performed in an *online* or *streaming* setting, where only a small subset of the data is loaded into memory at a given time. For larger datasets, a fully distributed computing framework may be required.

The first step described in Section 2.2 is the most computationally expensive. All of the 697,622,444 individual trips are aggregated into 35,064 mean pace vectors, and consequently $t \in \{0, \dots, 35,063\}$. Recall that there are a total of $4 \times 365 \times 24 + 24 = 35,064$ hours in a 4 year dataset (including a leap day). Since the trip data is sorted chronologically, it is possible to compute these mean pace vectors in a single pass. Recall from (2) that the mean pace computation involves the sum of trip durations and the sum of trip distances. Thus, these two sums are initialized to zero for each of the origin-destination trips. Each time a trip is read from the file, the relevant sums are incremented. The error filtering from Section 3.1 can also be performed at this stage, so an additional pass of the dataset is not required. When the start hour of the current trip (rounded) is greater than the start hour of the previous trip, the sums are complete for the previous hour. The mean pace vector is computed by division and output, then the sums are reset to zero. Thus, the computation is one large loop over the entire dataset. A short pseudocode is given in Algorithm 1. Note that NUM_TYPES is the total number of trip types (i.e., k^2).

Algorithm 1 Online Mean Pace Vector Extraction

```
prev_hour := -1                                ▷ Start at beginning of time
sum_duration := zeros(NUM_TYPES)              ▷ Initialize sums to 0
sum_distance := zeros(NUM_TYPES)             ▷ Initialize sums to 0
for all trip ∈ chronological_trips do        ▷ Loop over all trips
  while trip.hour > prev_hour do           ▷ If previous hour is complete:
    output(prev_hour,  $\frac{\text{sum\_duration}}{\text{sum\_distance}}$ )    ▷ Output mean pace vector
    sum_duration := zeros(NUM_TYPES)         ▷ Reset sums to 0
    sum_distance := zeros(NUM_TYPES)         ▷ Reset sums to 0
    prev_hour += 1                           ▷ Advance to next hour
  end while
  if trip.isValid() then                   ▷ Data filtering
     $i \leftarrow \text{category}(\text{trip.pickup}, \text{trip.dropoff})$     ▷ Determine trip type
    sum_duration[i] += trip.duration         ▷ Update distance sum
    sum_distance[i] += trip.distance         ▷ Update duration sum
  end if
end for
```

Since each trip is accessed only once, the computation is $O(N)$, where N is the total number of trips. The computation of each hour time-slice is independent, making it possible to employ parallel processing if the data is partitioned ahead of time. The analysis was implemented in python and run on an 8-core 2.5GHz machine with 24GB of ram. The extraction of all 35,064 mean pace vectors took about 75 minutes, using roughly 40MB of RAM for each of the eight processes. The fact that the runtime is much shorter than the real timespan of the dataset combined with the single-pass property means that this preprocessing could be performed in realtime. In other words, this system could realistically collect trips as they occur, update the relevant sums, then output the mean pace vector at the end of the hour.

The remaining computations involve mean pace vectors instead of raw trip data. They also have linear time complexity and are much faster than the preprocessing. Recall from (5) and (6) that, at a particular hour, the mean and covariance need to be computed for *all hours in the periodic pattern except that hour*. The naive implementation of this calculation has a quadratic time complexity, since each mean pace vector must be compared against every other mean pace vector. However, it is possible to compute all of these quantities in linear time. Instead of directly computing the mean of all values except $\mathbf{a}(t)$, the sum of all values including $\mathbf{a}(t)$ is computed up front. Then, in the loop, $\mathbf{a}(t)$ is subtracted from this sum. Formally, the *inclusive reference set*, Q_{t+} , is defined in a similar way to (5), except that it includes the mean pace vector $\mathbf{a}(t)$. In other words,

$$Q_{t+} = \{\mathbf{a}(h) | h \equiv t \pmod{168}\} = Q_t \cup \{\mathbf{a}(t)\}. \quad (8)$$

Unlike the reference set from (5), the inclusive reference set is identical for values of t that occur at the same point in the periodic pattern. Thus, Q_{t+} and the sum of all vectors in Q_{t+} only need to be computed once. To compute the sum of all vectors in Q_{t+} *except* $\mathbf{a}(t)$, it is sufficient to subtract $\mathbf{a}(t)$ from this sum. Thus, the mean computation can be written as

$$\mu(t) = \frac{1}{|Q_t|} \sum_{\mathbf{a} \in Q_t} \mathbf{a} = \frac{1}{|Q_{t+}| - 1} \left(\left(\sum_{\mathbf{b} \in Q_{t+}} \mathbf{b} \right) - \mathbf{a}(t) \right). \quad (9)$$

A similar technique is used for the sum of outer products in the covariance computation (6). This method avoids redoing most of the addition in each iteration, allowing for a significant improvement on large datasets. Once $\mu(t)$ and $\Sigma(t)$ are computed, $M(t)$ can be computed in constant time. Thus, the entire operation runs in linear time. On the same machine, this computation ran in less than 30 seconds, producing the timeseries of $M(t)$. Again, this operation would be feasible in a real-time system. However, it is worth noting that it may be desirable to re-generate old values of $M(t)$ in light of new information.

Once $M(t)$ is generated, the event detection described in Section 2.4 can also be performed in linear time. Events and spaces between events are stored as a linked list, where each node contains the start time and end time. Scanning through $M(t)$ chronologically, a new node in the linked list is generated each time $M(t)$ crosses above or below the threshold. Then, to remove short spaces between events, this linked list is iterated upon. Each time a non-event node of less than the desired duration is discovered, that node and its two neighbors are replaced with one larger node. On the same machine as the previous computations, it took less than one second to perform the

event detection.

3.3. Extraction of Pace Features

The map of New York City is first split into four large regions, shown in Figure 3. For the remainder of the analysis, the zones will be referred to in the following way: *Upper Manhattan* (U), *Midtown* (M), *Lower Manhattan* (L), and *East of the Hudson River* (E). Note that the Eastern region is connected only by bridges and tunnels and thus problems with this infrastructure will tend to increase travel times between this region and others. Specifically relevant to Hurricane Sandy is the Lower Manhattan region, since it experienced severe flooding and power outages. Choosing four large regions in this way satisfies the first goal outlined in Section 1.1 because it defines meaningful city-scale properties. Instead of looking at every street in New York in detail, it defines large areas with key geographic and infrastructural properties. The travel times between these regions reflect the overall performance of city-scale transportation infrastructure. It is worth noting that the methodology allows for an arbitrary choice of regions, which can be chosen manually or automatically via graph clustering techniques. The present implementation chooses zones manually to aid interpretation of the results. The regions are also selected to study areas where the taxis typically operate, and consequently some regions (e.g., Staten Island) are not included in the analysis due to the lack of taxi trips to/from the region.

Recall that a taxi can take one of 16 possible trips between these regions. Aggregating these trips by type and hour as in Section 2.2 produces the 16-dimensional mean pace vector, $\mathbf{a}(t)$, at all points in time. Figure 4 shows three typical weeks of mean pace vectors (as identified by visual inspection of

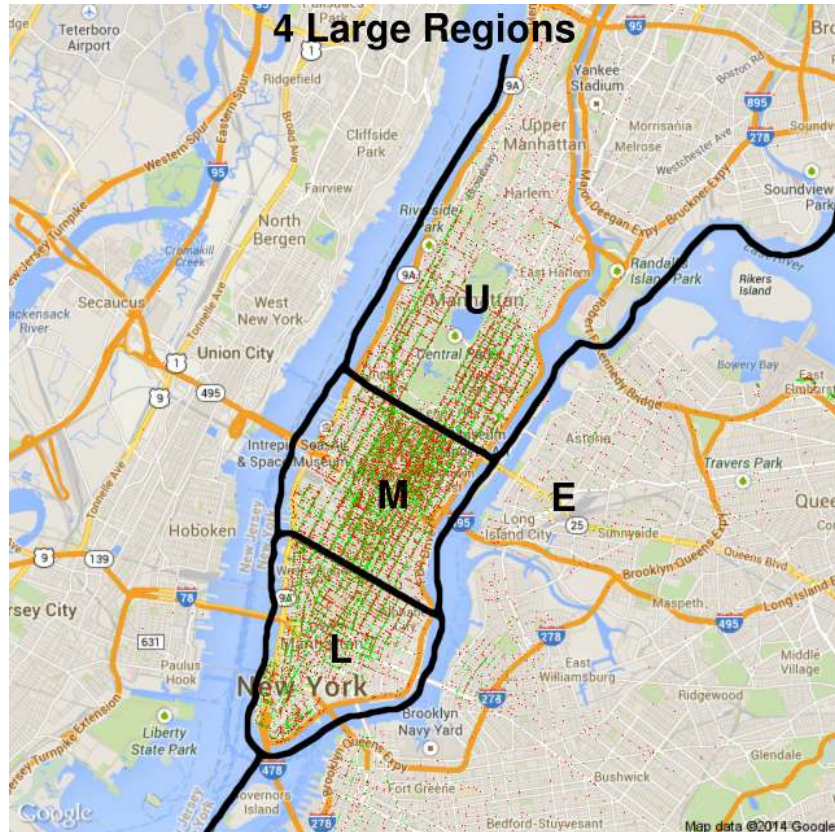


Figure 3: Division of New York City into four large regions denoted $U, M, E,$ and L . A random sample of 0.01% of the taxi trips in 2012 are shown. Pickups locations are marked in green, and the corresponding dropoffs are marked in red. The majority of trips occur in Manhattan, with especially high concentration in the Midtown region.

the Mahalanobis distance time series), revealing the expected weekly periodic pattern assumed in the analysis. Note the data from these three weeks are only used for illustration purposes and not in any formal analysis.

Several elements of the pace vector over the four year dataset are estimated with an extremely large number of trips (e.g., 10% of the elements in the pace vector time series have more than 3200 trips), while 80% of the elements have more than 225 trips in the given hour. At the extreme tail, we find 97% of the elements are estimated with more than 50 trips in the hour, and 99.9% of the elements in the pace vector time series have at least 10 trips.

3.4. Analysis of Events

As detailed in Section 2.3, the expected behavior is generated for all times t according to $\mu(t)$ and $\Sigma(t)$. An interesting way to view the mean pace vector $\mathbf{a}(t)$ is by standardizing it, element by element, producing the *standardized pace vector*. The i th element of this vector is given by

$$\mathbf{S}(t)_i = \frac{\mathbf{a}(t)_i - \mu(t)_i}{\sqrt{\Sigma(t)_{i,i}}}. \quad (10)$$

Intuitively, the standardized pace vector tells how many standard deviations away from the mean the pace of each category of trips is at time t . In other words, it is possible to identify the trips that are going slower or faster than expected, and how significant this difference is. Figure 5 shows the standardized pace vector during the week of Hurricane Sandy. This figure gives some intuition on the behavior of various regions of the city during and after the hurricane. It also includes labels indicating the occurrences of

Mean Pace Vector – Three Typical Weeks

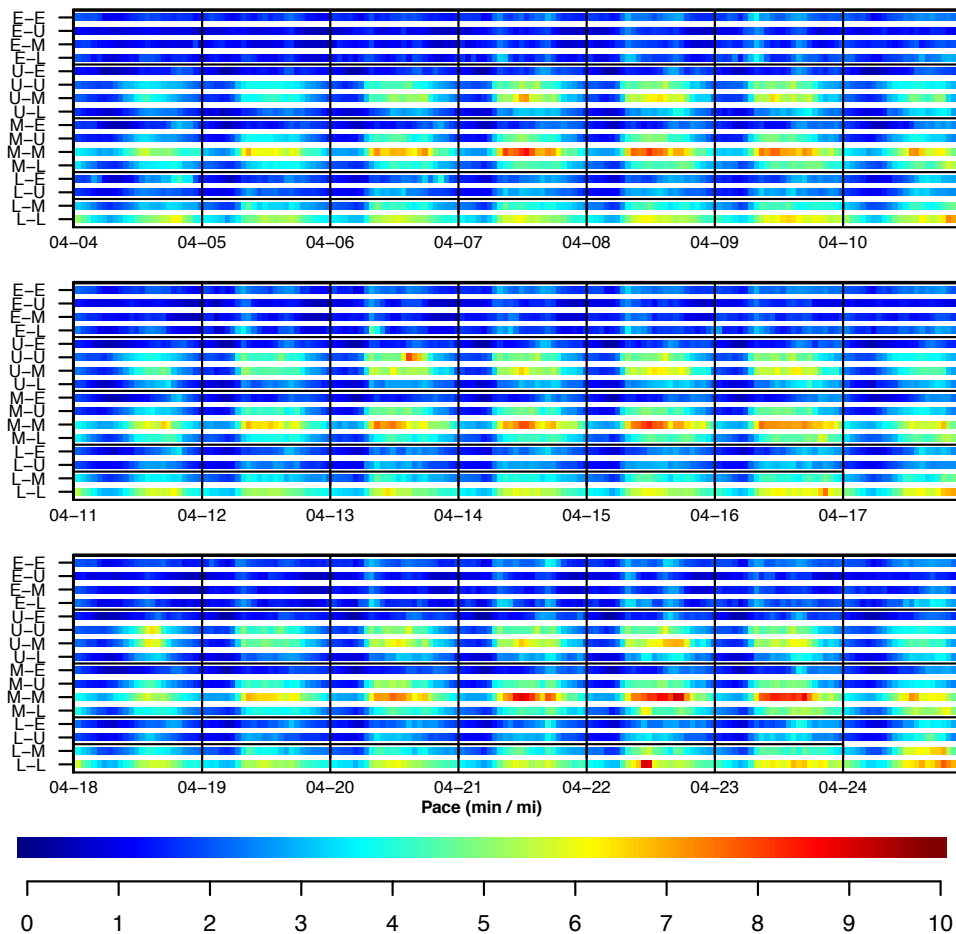


Figure 4: The mean pace vector, $\mathbf{a}(t)$ for three typical weeks, starting on April 4, 2010. A periodic pattern is observable, with high paces during rush hour.

various phases of the event, obtained from a post-Sandy study by Kaufman et al. (2012). The most notable finding is that the slowest traffic occurred on Wednesday October 31st, almost two days after the hurricane struck land. On this day, some airports, buses, and commuter rails attempted to resume normal service, but much of the infrastructure was still damaged (Kaufman et al. (2012)).

However, it is even more surprising that Midtown-to-Lower Manhattan and Lower Manhattan-to-Lower Manhattan travel times are significantly *lower* than expected during this time. The pace of these trips remains almost five standard deviations below the mean until Saturday the 3rd, despite the severe flooding and power outages in Lower Manhattan. The fact that a hurricane can actually make traffic move faster in some areas of the city indicates that the usage of the infrastructure changed. It is likely that the hurricane decreased demand on the transportation network in Lower Manhattan until the infrastructure recovered.

This standardized pace vector gives a meaningful interpretation of unusual travel times between various regions of the city. However, it fails to account for correlations between these typical travel times i.e., the off-diagonal elements of $\Sigma(t)$. In contrast, the Mahalanobis distance $M(t)$ considers the full covariance matrix. As described in Section 2.4, events are detected when $M(t)$ goes above a threshold for a significant period of time. Figure 6 shows this process, along with the average pace of all taxis. Note Figure 6 also indicates that the disruption caused by Hurricane Sandy actually begins in the late evening of October 28, nearly 24 hours before the storm hit the city. Figure 5 gives further insight as to why the Mahalanobis distance is extreme

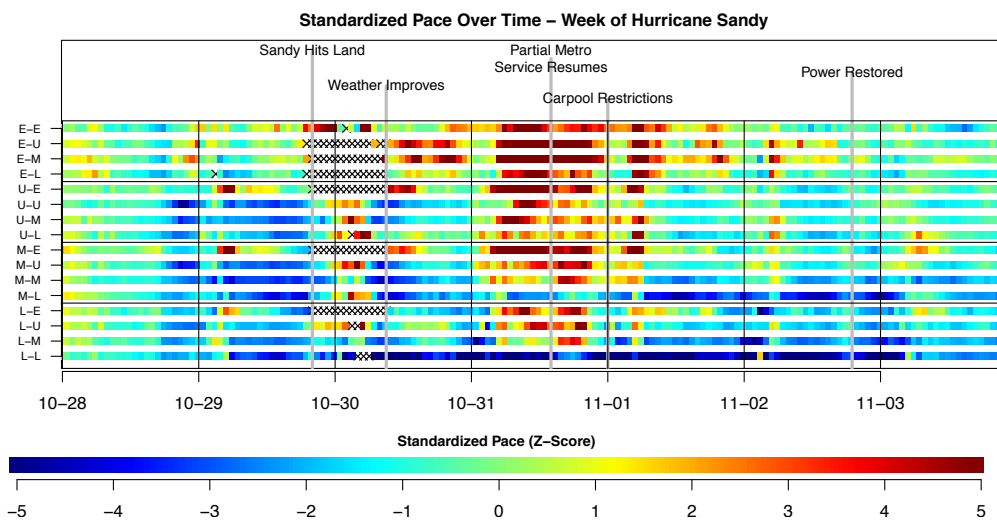


Figure 5: The standardized pace vector (i.e., vector of z-scores) during the week of Hurricane Sandy, 2012. Labels are included to show the times of specific phases of the event (Kaufman et al. (2012)). An average week would have values of zero everywhere, but significant deviations are shown during the week of Hurricane Sandy. Missing data (hours where there are less than five occurrences of a given trip) are marked with black Xs.

during this period, namely that the traffic conditions were generally much faster than would be typically expected during those times.

Table 3 shows the top ten events, sorted by duration. At the top of the list is Hurricane Sandy, taking over five and a half days for travel times to return to normal, which is more than three times the recovery time of Hurricane Irene. The longer duration of Hurricane Sandy agrees with the results of Ferreira et al. (2013), which showed that the total number of Manhattan taxi trips returned to normal more quickly during Hurricane Irene than Hurricane Sandy. At its worst, Sandy added over two minutes to each mile driven by taxis in the city, while Irene added less 40 seconds. This is in contrast to the results of Ferreira et al. (2013), which showed that the peak drop in the number of taxi trips was greater during Hurricane Irene. The blizzard of December 2010, while shorter, added four minutes of travel time to each mile at its peak.

Note it is difficult to evaluate the accuracy of the results in Table 3, since the true severity of each event is not known. If a training set of events is available, one could raise or lower the detection threshold until the desired balance between type I and type II errors is reached.

4. Conclusions and future work

This analysis has shown that it is possible to detect and measure the effects of unusual events on transportation infrastructure using only taxi GPS data. This is a first step toward assessing and improving city-scale resilience. Of key importance, the method is extremely low cost, because it does not require the installation of any additional sensors beyond GPS data streams

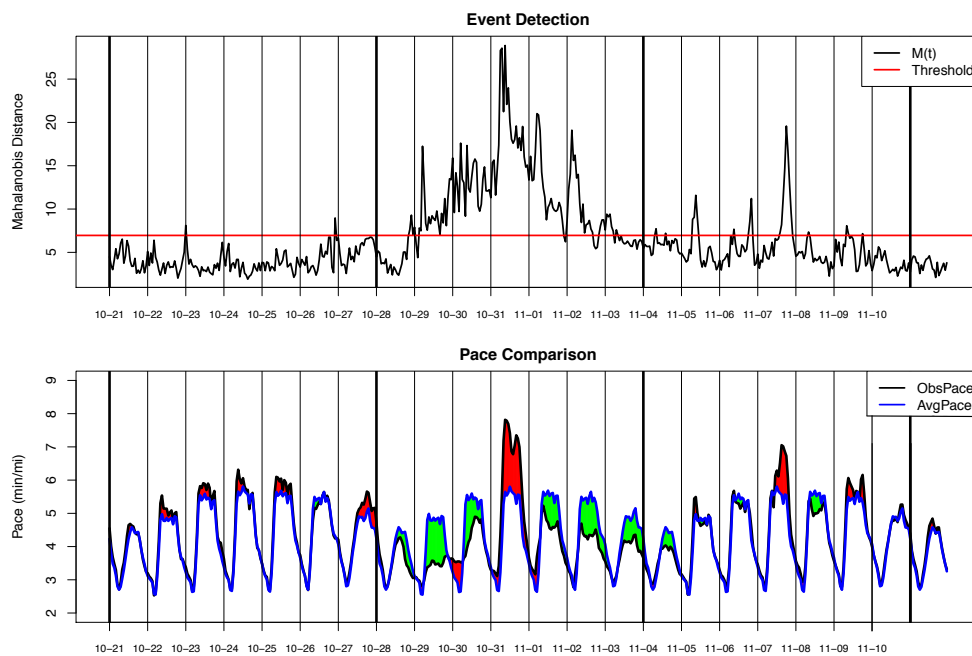


Figure 6: Probabilistic detection and measurement of the event Hurricane Sandy. The Mahalanobis distance, $M(t)$, is plotted in the top figure and events are detected when it goes below the threshold. For comparison, the average pace of all taxis in the city is plotted below and compared to the expected value. Green areas indicate that travel times are low, but red indicates that they are unusually high.

| Event | Start Time | Duration (hours) | Max (min/mi) | Min (min/mi) | Worst Trip |
|-----------|---------------------|---------------------|-----------------|-----------------|------------|
| Sandy | 2012-10-28 21:00:00 | 132 | 2.25 | -1.6 | E → M |
| Blizzard | 2010-12-26 13:00:00 | 112 | 4.41 | 0 | M → M |
| Blizzard | 2011-01-31 08:00:00 | 49 | 2.04 | 0 | E → E |
| Irene | 2011-08-27 13:00:00 | 43 | 0.64 | -1.66 | E → E |
| Unknown | 2013-10-12 03:00:00 | 33 | 1.09 | 0 | E → L |
| Blizzard | 2013-02-08 06:00:00 | 26 | 1.54 | -0.58 | E → E |
| Blizzard | 2010-02-10 06:00:00 | 24 | 0.67 | -1.01 | E → E |
| New Years | 2012-12-31 15:00:00 | 20 | 1.42 | -2.66 | E → M |
| Memorial | 2011-09-09 08:00:00 | 19 | 1.66 | 0 | U → U |
| Blizzard | 2011-01-28 02:00:00 | 18 | 2.57 | 0 | L → L |

Table 3: Comparison of New York City transportation infrastructure resilience to the 10 longest events. The duration in hours, and the maximum/minimum pace deviation in minutes/mile is given for each event. Note that a positive number indicates a delay while a negative indicates a decreased pace. The final column indicates which of the 16 trips most frequently had the highest standardized pace during the event. Labels for events (the first column) are determined manually.

available from GPS equipped fleet vehicles such as taxis, or personal navigation services. This method proposes computing *origin-destination paces*, or average travel time per mile between various regions of the city. The effects of various events are quantified by the sizes and durations of pace deviations from typical values. Importantly, this measurement is done in a probabilistic way, so significant events can be distinguished from random day-to-day fluctuations.

The main benefit of the method is that it is quantitative, and can presumably be applied to any city with large volumes of GPS trip data. The main limitation of the method is that it only provides a partial view of the resilience of the infrastructure. For example, it does not indicate the actual state of the physical infrastructure (e.g., if an individual road segment is damaged or destroyed), or the demand placed on the infrastructure (e.g., the number of people in the city desiring travel over the infrastructure). Similarly, the method only indicates the magnitude and duration of the disruption, it does not indicate why a particular event recovers faster or slower than another. Regardless, it does represent a practical step towards building comprehensive real-time analytics for measuring infrastructure resilience.

The proposed method is applied to a dataset from New York City, and Hurricane Sandy is analyzed in detail. The analysis shows this was the longest event in the four year dataset, and one of the most severe in terms of peak pace deviation. At its worst, Hurricane Sandy caused over two minutes of delay per mile, but actually resulted in *faster* traffic for most of its duration. Most interestingly, the spike in delay occurred two days after the hurricane struck. This post-disaster traffic congestion was extreme, suggest-

ing that more traffic management might be necessary following an event. The analysis of an extreme event like Hurricane Sandy demonstrates the ability of the proposed method to capture and describe atypical city-scale properties of the transportation network.

Several extensions to the methodology are anticipated in our future work. In this work, the origin-destination regions are manually selected to aid interpretation. Automatic region identification is potentially possible by leveraging recent advances in road graph partitioning (e.g., the patented algorithm of Dellinger et al. (2011) to emphasize travel times of trips crossing connection-critical infrastructure. It is also possible to apply the outlier-detection methods to other types of paces. For example, instead of measuring paces between various origin-destination zones, one may desire to compute approximate paces on each link of the network graph. Algorithms exist which can estimate link travel times (for example Hunter et al. (2009); Zhan et al. (2013); Santi et al. (2014); Zhan et al. (2016b)), but they are computationally expensive for large datasets. If the same outlier-detection methods are applied to link-level pace data, it is possible to examine whether such a heavy computation is necessary. If the results are unchanged, the simpler method presented in this article may be sufficient. Otherwise, an higher resolution approach may provide additional insight, for example by explicitly including the network structure from the road graph.

We also note the current method does not allow direct comparison of some holidays that occur on a fixed date, such as Independence Day (4 July) in the US, because it appears on different days of the week in the dataset. Practically, identifying an extreme holiday compared to the typical

holiday would require significantly more data, since only four observations of the holiday are present in the current dataset. Regardless, extensions that directly exploit the temporal structure of the traffic patterns is an interesting future direction.

We emphasize that the method as presented is a static analysis and does not account for changes like gradual increases in travel demand or changes in the infrastructure capacity. As more data becomes available, it may both be possible and necessary to de-trend the data. Alternatively, old data can be discarded, and the method could be rerun using the most recent years only. Such techniques are conceptually straightforward extensions that may require additional practical engineering effort to implement them in a computationally efficient framework.

Finally, application of the methodology across multiple cities would enable resilience comparisons both across event types within a city, as well as the same event type across different cities. We note that our preliminary efforts to acquire taxi data in other large cities in the US have been initially unsuccessful, for example due to the fact that the data is stored by the payment processors that manage the credit card readers in the taxis, and not the city. It is however currently possible to acquire suitable taxi datasets in other parts of the world (Zhang et al. (2015); Yuan et al. (2010)).

Acknowledgments

This work was supported by the National Science Foundation under Grant No. CNS-1308842.

References

- Allen, W. B., Liu, D., Singer, S., 1993. Accessibility measures of US metropolitan areas. *Transportation Research Part B: Methodological* 27 (6), 439–449.
- Aven, T., 2011. On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. *Risk Analysis* 31 (4), 515–522.
- Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C., 2011. Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Transactions on Intelligent Transportation Systems* 12 (1), 141–151.
- Calabrese, F., Pereira, F. C., D. Lorenzo, G., Liu, L., Ratti, C., 2010. The geography of taste: analyzing cell-phone mobility and social events. In: *Pervasive Computing*. Springer, pp. 22–37.
- Castro, P. S., Zhang, D., Li, S., 2012. Urban traffic modelling and prediction using large scale taxi GPS traces. In: *Pervasive Computing*. Springer, pp. 57–72.
- Chang, S. E., Nojima, N., 2001. Measuring post-disaster transportation system performance: the 1995 Kobe earthquake in comparative perspective. *Transportation Research Part A: Policy and Practice* 35 (6), 475–494.
- Chen, C., Zhang, D., Samuel Castro, P., Li, N., Sun, L., Li, S., 2012. Real-time detection of anomalous taxi trajectories from GPS traces. In: *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Vol. 104 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics*

- and Telecommunications Engineering. Springer Berlin Heidelberg, pp. 63–74.
- Daganzo, C. F., 1997. Fundamentals of transportation and traffic operations. Vol. 30. Pergamon Oxford.
- Delling, D., Goldberg, A. V., Razenshteyn, I., Werneck, R. F., 2011. Graph partitioning with natural cuts. In: Proceedings of the 2011 IEEE International Parallel & Distributed Processing Symposium (IPDPS). pp. 1135–1146.
- Donovan, B., Work, D. B., 2014. New York City taxi data (2010–2013). <http://dx.doi.org/10.13012/J8PN93H8>.
- Edie, L. C., 1963. Discussion of traffic stream measurements and definitions. Proceedings of the 2nd International Symposium on the Theory of Traffic Flow, 139–154.
- Faturechi, R., Miller-Hooks, E., 2014. Measuring the performance of transportation infrastructure systems in disasters: A comprehensive review. *Journal of Infrastructure Systems* 21 (1), 04014025.
- Ferreira, N., Poco, J., Vo, H. T., Freire, J., Silva, C. T., 2013. Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips. *IEEE Transactions on Visualization and Computer Graphics* 19 (12), 2149–2158.
- Geroliminis, N., Daganzo, C. F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42 (9), 759–770.

- Haimes, Y. Y., 2009a. On the complex definition of risk: A systems-based approach. *Risk Analysis* 29 (12), 1647–1654.
- Haimes, Y. Y., 2009b. On the definition of resilience in systems. *Risk Analysis* 29 (4), 498–501.
- Haimes, Y. Y., 2011. Responses to Terje Aven’s paper: On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. *Risk Analysis* 31 (5), 689–692.
- He, X., Liu, H. X., 2012. Modeling the day-to-day traffic evolution process after an unexpected network disruption. *Transportation Research Part B: Methodological* 46 (1), 50–71.
- Hunter, T., Herring, R., Abbeel, P., Bayen, A., 2009. Path and travel time inference from GPS probe vehicle data. *NIPS Analyzing Networks and Learning with Graphs*.
- Kaplan, S., Garrick, B. J., 1981. On the quantitative definition of risk. *Risk Analysis* 1 (1), 11–27.
- Kaufman, S., Qing, C., Levenson, N., Hanson, M., 2012. Transportation during and after Hurricane Sandy. Tech. rep., Rudin Center for Transportation, NYU Wagner Graduate School of Public Service.
- Konstantinidou, M. A., Kepaptsoglou, K. L., Karlaftis, M. G., 2014a. Transportation network post-disaster planning and management: A review part I: Post-disaster transportation network performance. *International Journal of Transportation* 2 (3), 1–16.

- Konstantinidou, M. A., Kepaptsoglou, K. L., Karlaftis, M. G., 2014b. Transportation network post-disaster planning and management: A review part II: Decision-making and planning of post-disaster operations. *International Journal of Transportation* 2 (3), 17–32.
- Mahalanobis, P. C., 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)* 2, 49–55.
- Matherly, D., Langdon, N., 2014. A guide to regional transportation planning for disasters, emergencies, and significant events. Tech. Rep. 777, National Cooperative Highway Research Program.
- Murray-Tuite, P., Wolshon, B., 2013. Evacuation transportation modeling: An overview of research, development, and practice. *Transportation Research Part C: Emerging Technologies* 27, 25 – 45, selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII).
- Navarro, J., 2014. Can the bounds in the multivariate Chebyshev inequality be attained? *Statistics & Probability Letters* 91, 1–5.
- Omer, M., Mostashari, A., Nilchiani, R., 2013. Assessing resilience in a regional road-based transportation network. *International Journal of Industrial and Systems Engineering* 13 (4), 389–408.
- Ouyang, M., Dueñas-Osorio, L., Min, X., 2012. A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety* 36–37, 23–31.

- Qi, G., Li, X., Li, S., Pan, G., Wang, Z., Zhang, D., 2011. Measuring social functions of city regions from large-scale taxi behaviors. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011. pp. 384–388.
- Reed, D. A., Kapur, K. C., Christie, R. D., 2009. Methodology for assessing the resilience of networked infrastructure. *IEEE Systems Journal* 3 (2), 174–180.
- Saberi, M., Mahmassani, H., Hou, T., Zockaie, A., 2014. Estimating network fundamental diagram using three-dimensional vehicle trajectories: Extending Edie’s definitions of traffic flow variables to networks. *Transportation Research Record: Journal of the Transportation Research Board* (2422), 12–20.
- Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., Ratti, C., 2014. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences* 111 (37), 13290–13294.
- Warren, R., Smith, R. F., Cybenko, A. K., 2011. Use of mahalanobis distance for detecting outliers and outlier clusters in markedly non-normal data: A vehicular traffic example. Tech. rep., DTIC Document.
- Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G., Huang, Y., 2010. T-drive: driving directions based on taxi trajectories. In: Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems. ACM, pp. 99–108.

- Zhan, X., Hasan, S., Ukkusuri, S. V., Kamga, C., 2013. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies* 33, 37–49.
- Zhan, X., Qian, X., Ukkusuri, S. V., Sept 2016a. A graph-based approach to measuring the efficiency of an urban taxi service system. *IEEE Transactions on Intelligent Transportation Systems* 17 (9), 2479–2489.
- Zhan, X., Ukkusuri, S. V., Yang, C., 2016b. A probabilistic urban link travel time estimation using large-scale taxi data automation in construction. *Invited Paper for Special Issue on Big Data in Civil Engineering* 72, 237–246.
- Zhang, D., Zhao, J., Zhang, F., He, T., 2015. UrbanCPS: a cyber-physical system based on multi-source big infrastructure data for heterogeneous model integration. In: *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*. ACM, pp. 238–247.
- Zheng, Y., Liu, Y., Yuan, J., Xie, X., 2011. Urban computing with taxicabs. In: *Proceedings of the 13th International Conference on Ubiquitous Computing*. ACM, pp. 89–98.
- Zhu, S., Levinson, D., Liu, H. X., Harder, K., 2010. The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. *Transportation research part A: policy and practice* 44 (10), 771–784.