Journal of Big Data Analytics in Transportation (2020) 2:61–74
https://doi.org/10.1007/s42421-020-00017-w

ORIGINAL PAPER

Improving Probe-Based Congestion Performance Metrics Accuracy by Using Change Point Detection

Vesal Ahsani[1] · Anuj Sharma[1] · Chinmay Hegde[2] · Skylar Knickerbocker[3] · Neal Hawkins[3]

Received: 19 March 2020 / Revised: 28 March 2020 / Accepted: 3 April 2020 / Published online: 15 April 2020

Abstract
Probe-based speed data provide great value to agencies; especially, in areas which are not feasibly covered by traffic sensors. However, as with sensors, probe data are not without nuance and issues like latency prevent alignment between calculated metrics by data source. In recent years, there has been a strong impetus on using data-driven decision making. Data-driven insights have become critical for smart mobility. To support data-driven decision making, Federal Highway Administration has procured probe data feeds and provides free access to state and local agencies as National Performance Measures Research dataset (NPRMDS). In addition to the NPRMDS, several state agencies subscribe to a paid probe data provider for obtaining real-time streams of high-resolution probe data. These datasets are used to generate nationwide urban mobility reports as well as reports focusing on certain jurisdiction. These mobility reports are integrated in several Transportation System Management and Operations (TSMO) plans which are often used to drive several resource allocation projects. This paper examines accuracy of methodology used to derive two frequently used performance measures in the mobility reports; namely, number of congested hours and number of congestion events. An improve methodology is then proposed to find accurate estimates for number of congested hours and number of congested incidents.

Keywords Probe data · Sensor data · Congestion detection · Number of congested events · Congested hour · Fixed-speed threshold · Change point detection

Introduction

In recent years, there has been a strong impetus on using data-driven decision making. Data-driven insights have become critical for smart mobility. Map-21, the Moving Ahead for Progress in the 21st Century Act (P.L. 112–141), asks state departments of transportation (DOTs) and agencies to monitor and report mobility performance measures. To support data-driven decision making, in 2013, Federal Highway Administration has procured probe data feeds and provides free access to state and local agencies as National Performance Measures Research dataset (NPRMDS). INRIX is the current provider of NPMRDS data records. The database contains billions of records that fully cover the whole National Highway System (NHS) which includes all US interstates and highways. The hope is that all project decisions might be improved through use of probe-based data as opposed to only relying on infrastructure mounted sensors.

However, there are some critical points that state DOTs and transportation agencies should consider when using a probe data stream, such as INRIX. Some advantages and limitations of INRIX are as follows:

- In terms of geographic coverage, INRIX has been evaluated for interstates and noninterstates and has been shown to be reliable for almost all times of day on interstates.
- INRIX is more reliable during the day than at night, especially during peak hours.
- Regarding incident detection, INRIX is reliable for detecting merely congestion, especially recurring congestion. Generally, there are two types of congestion, recurring and nonrecurring. Recurring congestion is regarded as the congestion caused by the routine traffic in a nor-

✉ Vesal Ahsani
vesalahsani@gmail.com

1 Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, IA, USA

2 Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA

3 Institute for Transportation, Iowa State University, Ames, IA, USA

mal environment which is somehow expected, whereas nonrecurring congestion is unexpected and is most likely caused by an incident. Nonrecurring congestion may emerge as a result of a variety of factors like lane blocking crashes or disabled vehicles, work zone, lane closures, adverse weather conditions, etc. When INRIX detects congestion, it gets all the information related to the congestion, such as the duration of the congestion.
- There is almost always a time delay (latency) for INRIX congestion detection.

The Regional Integrated Transportation Information System (RITIS), provides use of NPMRDS data including travel time per mile (reliability), delay, duration of congestion, number of congested events, congested hour, congested mile, congestion intensity, speed drop, etc. Therefore, further reliance on this emerging source of data requires assurance that the data represent what is actually being experienced on the roadway. The goal of this paper is to compare the precision of performance measure methodologies between probe and sensor-based sources.

The number of congested events is a performance measure, which explains how reliable probe-sourced data are in detecting congestion (recurring and non-recurring). This performance measure is actually the number of congested events reflected within the probe data. On the other hand, the congested hours of a segment reflect the summation of hours vehicle speeds are below a defined speed threshold. A state-wide analysis reveals both the location and magnitude of congestion with the information aggregated by year, month, day of week (DOW) and time of day (TOD). Congested hour calculations require comparing each minute of measured speed data, for all state-wide segments, to the congestion threshold. If the probe-sourced speed data are both "real time", as opposed to historical and below the fixed-congestion threshold, that 1 min of time is considered as "congested". Summation of these congested minutes (reported in units of hours) is defined as the total number of congested hours.

Literature Review

A literature review shows that congested hour and number of congested events are two essential performance metrics which allow transportation planners and policy makers to more effectively allocate resources to address and improve network. Regarding costs of congestion, some organizations and agencies only consider costs of recurring congestion, while others include non-recurring costs in addition to recurring.

Recurring congestion is very common in the U.S. with travellers expecting and planning for some delay; particularly, during peak hours. Many commuters modify their schedules or assign additional time to allow for these typical traffic delays. In contrast, non-recurring, unexpected delays, can have severe impacts on motorist's safety and mobility. Motorists want to be confident that a trip that takes 30 min today will also take 30 min tomorrow and so travel time reliability calculates the extent of this unexpected delay. Reliability is formally defined as the consistency or dependability in travel times, as measured from day to day or across different time periods of the day.

Delay is also important to many users of transportation systems, from passenger vehicle and truck drivers, transit riders, freight shippers, pedestrians, etc. Reliability is valuable for personal and business travellers as it allows them to use their time better. Additionally, it is a priceless service that can be afforded on privately-financed or privately-operated routes. That is why transportation planners should consider delay, congested hour, number of congested events and travel time reliability as essential performance measures because they are so vital for transportation system users.

Transportation agencies and regional planning organizations increasingly utilize travel time performance as reliability and variability measure (Nam et al. 2005). In 2003, Bell and Lida defined travel time reliability as the probability of on-time arrival (Bell and Iida 2003). Lei Lin and his colleagues compared travel times calculated from connected vehicles to those from road sensors (Lin et al. 2018, 2019). In addition, Lomax et al. in the same year introduced travel time reliability as a variability of travel time that commuters experience and as a consistency of a specific mode during a certain period of time (Lomax et al. 2003a, b). Additionally, Lomax recommended a focus on duration, extent and intensity of a congestion and reliability measures in addition to travel time alone in order to assess transportation system performance. Adapting different methods to measure traffic congestion intensity helps to rank and prioritize congested segments and in providing a more comprehensive spatial and temporal understanding of congestion duration, extent and severity.

In this study, we attempt to evaluate the reliability of probe-sourced data (INRIX) using two performance measures; congested hour and the number of congested events. The study will introduce a new method for detecting traffic congestion and reductions in speed.

Wide Area Probe Data

State Departments of Transportation (DOT) and many transportation agencies use infrastructure sensors to collect comparatively accurate real-time traffic-related information such as occupancy, traffic speed for each lane and vehicle class. The cost to deploy and maintain these infrastructure-based sensors can be high. The other major limitation for fixed sensors is their geographical scalability; many units must be installed on the roadsides to adequately determine and measure the traffic situation in any particular area (Young 2007). The access to power and communications leans towards major freeways, interstates and critical urban areas rather than an even distribution a state. The lack of sufficient coverage on highways and arterials generate the desire for DOTs to consider augmenting existing traffic data collection with probe-based services for wider coverage under limited budgets.

Advancements in telecommunication and wireless technologies have facilitated new ways to collect traffic data, process the information and analyse the data. Probe-based technologies can be used to collect traffic-related information from millions of mobile devices, GPS-enabled vehicles and other sources. Probe-based methods of measuring travel time and speed data can easily be scaled across large networks without need for deploying additional infrastructure (Young 2007). This can allow state agencies to cost-effectively use a single uniform source of data for monitoring traffic across most roadways (FHWA 2013). INRIX, HERE and TomTom are some of the established third-party providers of such probes.

Various studies have been carried out to compare the reliability and accuracy of probe data with sensor data from radar sensors, loop detectors, etc. (Adu-Gyamfi et al. 2017; Coifman 2002; FDOT 2012; Feng et al. 2010; Haghani et al. 2009; Kim and Coifman 2014; Lindveld et al. 2000). Many of these studies evaluated probe performance using travel time reliability measures, such as the 90th or 95th percentile of travel time, standard deviation, percentage of variation, buffer time index (BTI), planning time index (PTI), travel time index (TTI), frequency of congestion, failure rate (with respect to average), on-time arrival, misery index, congestion detection latency, count of congestion, congestion duration, reliability curve, hourly traffic volume, congested hours, etc. (Aliari and Haghani 2012; Araghi et al. 2015; Belzowski et al. 2014; Cookson and Pishue 2016; Day et al. 2015; FHWA 2017; Gong and Fan 2017; Lomax et al. 2003a, b; Mcleod et al. 2012; MoDOT 2017; Peniati 2004; Pu 2012; Remias et al. 2013; Schrank et al. 2012, 2015; Sekuła et al. 2017; Turner 2013; Venkatanarayana 2017; WSDOT, 2013, 2014; Zheng et al. 2018). An overview of these studies and the performance measures used to evaluate the reliability of probe data is provided in Table 1.

Fixed Threshold

Traffic congestion has become one of the most expensive problems in the world, especially in large cities and metropolitan areas. Effectively addressing congestion requires the ability to use real-time traffic data towards improved timely decision making. Traffic flow parameter-based detection methods have been widely accepted since they can be implemented automatically and are not affected by weather conditions. Many congestion detection methods based on the traffic flow parameters have been studied. Dudek, Messer and Nuckles developed the California method in 1974, which has been widely accepted and applied in traffic congestion and incident detection. The California algorithm is mostly used as a basis of comparison between congestion and incident detection methods (Dudek et al. 1974). McMaster's incident detection method was developed by Persaud in 1990. Many other methods were also developed in the following years and all are convenient to be used in practice. The difficult task is to define the threshold values of these methods which are often subjective according to experience (Persaud et al. 1990).

When these same subjective threshold values are used in performance measures, such as the number of congested events and congested hours, this can lead to erroneous decisions. The number of congested events is a performance measure that explains how reliable probe data are in detecting congestion (recurring and non-recurring) compared to a benchmark dataset (sensors). Average number of hours when the vehicle speeds are less than 90 percent of free-flow speed (FFS) is considered as congested hour. For instance, when the FFS is 60 mph, congested hour is computed as the average number of hours when vehicle speeds are less than 54 mph. This performance measure is typically computed only for weekdays from 6 am to 10 pm.

Data

The sources of data utilized in this work are explained in this section.

Probe-Sourced Data

With the help of today's technologies, e.g., connected vehicles and smartphones, probe data can leverage both historic and real-time data to report on transportation network operations. This study used both historical and real-time traffic data collected through the INRIX TMC monitoring platform. For each of TMC segment, speed, average length of segments and corresponding date and time of traverse, are provided each minute.

Infrastructure-Mounted Sensors

The benchmark data utilized in this study were provided by Wavetronix sensors which uses radar technologies for collecting traffic-related data. Although admittedly sensors might have some inherent errors, Wavetronix Smart Sensors have been commonly used for comparison purposes in

Table 1 Overview of the studies and the performance measures used to evaluate reliability of probe-sourced data

Study	Source of probe data used	Performance measures
Pu (2012)	Not mentioned	95th percentile travel time, standard deviation, coefficient of variation, percent variation, skew statistic buffer index (w.r.t. average), buffer index (w.r.t. median), planning time index, frequency of congestion, failure rate (w.r.t. average), failure rate (w.r.t. median), travel time index
Lomax et al. (2003a, b)	Not mentioned	Travel time window, percent variation, variability index, displaying variation, buffer time, buffer time index, planning time index, travel rate envelope, on-time arrival, misery index
Turner (2013)	INRIX	Annual hours of delay per mile, hours of target delay per mile, Travel Time Index, Planning Time Index, top N congested segments
Uno et al. (2009)	Not mentioned	Average travel time, covariance of travel time, level of service (LOS)
Rakha et al. (2010)	Not mentioned	Travel time coefficient of variation
Day et al. (2015), Remias et al. (2013)	INRIX	Congestion hours, distance-weighted congestion hours, congestion index, speed profile, speed deficit, travel time deficit, congestion cost, top N bottlenecks
MoDOT (2017)	Not mentioned	Average travel time per 10 miles, additional travel time needed for on-time arrival (80% of time), annual congestion costs
FHWA (2017)	NPMRDS	Congested hours, planning time index, travel time index
Schrank et al. (2012, 2015)	INRIX	Travel speed, travel delay, annual person delay, annual delay per auto commuter, total peak period travel time, travel time index, planning time index, number of rush hours, percent of daily and peak travel in congested conditions, percent of congested travel
WSDOT (2013, 2014)	Not mentioned	Lane-miles congested, total and cost of delay, travel time index
Sharma et al. (2017)	INRIX	Congestion detection latency, count of congestion, congestion durations, buffer time index, reliability curve
Hu et al. (2015)	INRIX	Delay saving, buffer index, 95th percentile travel time
Cookson and Pishue (2016)	INRIX	INRIX travel time index, wasted time in congestion
Aliari and Haghani (2012)	INRIX	Travel time, average speed
Gong and Fan (2017)	INRIX	Travel time reliability, planning time index, frequency of congestion
Sekuła et al. (2017)	INRIX	Hourly traffic volume
Venkatanarayana (2017)	INRIX, NPMRDS	Traffic delay, planning time index, travel time index, AASHTO reliability indexes (RI80, for all days and weekdays), congested hours and congested miles

various studies (Chakraborty et al. 2018; Lu et al. 2014; Poddar et al. 2018; Sharifi et al. 2011). Each Wavetronix sensor unit consists of a side-fire radar and hard-wired power for real-time processing of traffic data, such as speed, volume, etc. Wavetronix sensors provide high-resolution traffic data every 20 s.

In this paper, speed is the only traffic parameter which is utilized from Wavetronix sensors and INRIX segments. Table 2 indicates the statistics of the probe and sensor data that were utilized in this paper. Also, Fig. 1 shows speed distribution for INRIX and Wavetronix over 5 routes across Iowa in 2017.

Five different routes with 64 sensor–segment pairs were chosen in the state of Iowa over the year 2017. Figure 2 shows the segments and sensors for all five routes across Iowa. Moreover, Table 3 shows the number of segment–sensor pairs in each considered route. The speed limit of each route varies from 45 to 70 mph and it is considered in our analysis.

Table 2 Descriptive statistics of probe and sensor data used in this study

Parameters	Min	Max	Mean	Standard deviation
INRIX speed (mph)	2	92.6	61.3	5.8111
INRIX speed < 45 mph (mph)	2	44.9	30.7191	9.0671
INRIX segment length (mile)	0.0127	4.3673	1.5707	0.8636
Wavetronix speed (mph)	1	98.2	65.8	9.2543
Wavetronix speed < 45 mph (mph)	1	44.9	25.0162	11.1052

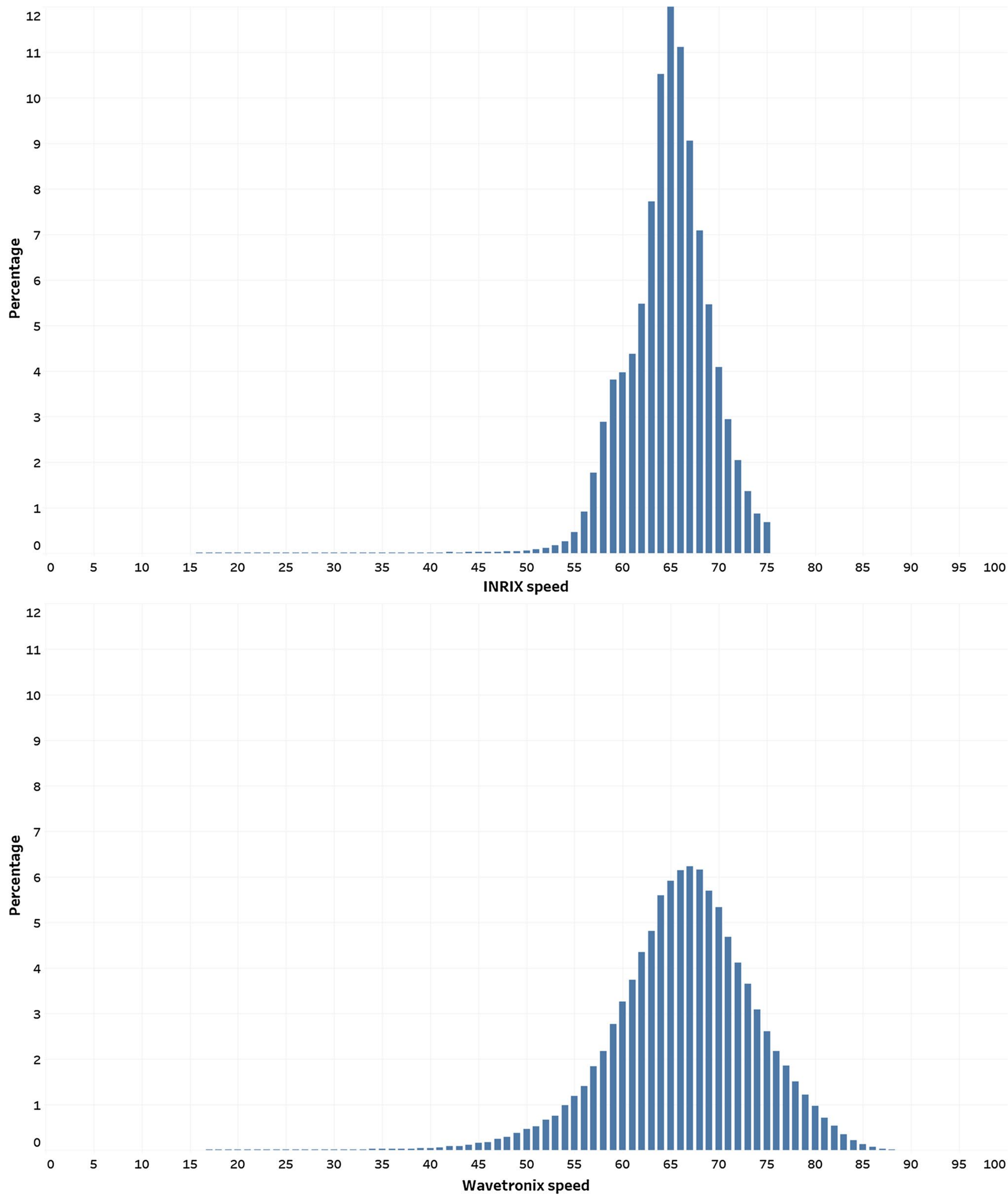

Fig. 1 Distribution of speed for INRIX and Wavetronix over 5 routes across Iowa in 2017

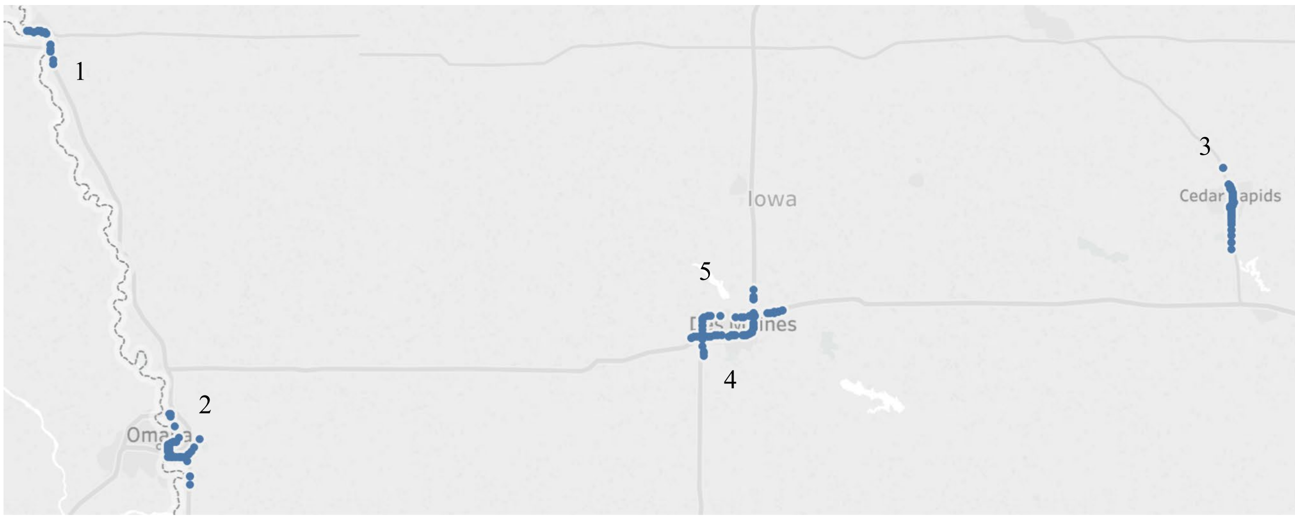

Fig. 2 Location of sensors and segments on 5 different routes in Iowa

Table 3 Number of sensors and segments on 5 different routes in Iowa

Route	Corridor	Number of segment–sensor pairs
1	I-29	8
2	I-29/80	12
3	I-380	16
4	I-235	15
5	I-35/80	13

Data Stream and Pre-processing

In real-world scenarios, since most raw datasets are incomplete, highly susceptible to noise and inconsistent due to sensor failures, measurement technique errors or data volume, data pre-processing plays a key role in detecting and correcting corrupt and erroneous traffic-related data. Since storing and analyzing huge amounts of INRIX and Wavetronix data need proper infrastructure and computational power to manage the large volume of data, a high-performance computing cluster should be utilized for data processing. A Hadoop Distributed File System (HDFS) was utilized for data storage and the map reducing was utilized for processing.

Preliminary Analysis

In the following section, a very traditional and common method of congestion detection is examined which utilizes a fixed-threshold speed and demonstrates how unreliable and erroneous the process can be. After that, an improved traffic congestion identification method is proposed and the number of congested events and congested hours are computed as performance measures.

In the preliminary stage, an analysis of a specific number of sensor–segment pairs in the state of Iowa was conducted and the results compared for different scenarios. For this purpose, ten sensor–segment pairs were chosen in the Des Moines metropolitan area. Performance measures were calculated for the entire year of 2016. The data were limited to the period of 5 am to 10 pm because the reliability of the Wavetronix sensors (benchmark data) is lower

Table 4 Reliability of probe data in detecting congestion events using fixed threshold method

		INRIX			
		Detect congestion		No congestion detected	
Wavetronix	Detect congestion	True positive: 343		False negative: 202	
		Recurring	Non-recurring	Recurring	Non-recurring
		304	39	190	12
	No congestion detected	False positive: 81		True negative: –	
		Recurring	Non-recurring	Recurring	Non-recurring
		70	11		

during the low volume late night hours. Also, the minimum duration for congestion was set to be greater than or equal to 15 min. Table 4 shows the reliability of probe data in detecting congested events. The fixed-threshold congestion detection method which utilized in this study is thoroughly explained in our previous paper (Ahsani et al. 2019). To have a brief explanation, the threshold speed in traditional congestion detection method is computed by subtracting twice the interquartile range from median speed for each 15-min period for each weekday from 8-week history. All the speeds below this threshold and 45 mph simultaneously are considered as non-recurring congestion while speeds above the computed threshold, but below 45 mph are considered as recurring congestion.

The first performance measure computed for this analysis is the number of congested events, as shown in Table 4. True positive (TP) represents a similar congested event which is detected by both Wavetronix sensor and INRIX segment; false negative (FN) means a congested event was detected by Wavetronix, but not INRIX; false positive (FP) denotes a congested event was detected by INRIX but not Wavetronix; and true negative (TN) represents a congested event which is detected by neither Wavetronix nor INRIX. It should be noted that there is no value for TN in the table below since we do not know the true number of congested events (not detected by either) which is why Wavetronix sensors are considered the benchmark for this analysis.

Table 4 shows that the number of congested events detected by both datasets was 343. There were 202 events not detected by INRIX or 37% of events missed. FP are the other way around an additional false alert that an operator would spend time on that did not actually occur. R and NR represent recurring and non-recurring congestion, respectively.

Table 4 shows a large discrepancy in the number of congested events detected by both the Wavetronix and INRIX. The measures do not imply there is a problem in the structure of the congestion detection algorithm but instead represent errors in the congestion detection method. Thus, it is imperative to come up with a solution to this issue which will be discussed further in this paper.

The second performance measure computed for this analysis is congested hours which is calculated using a fixed-threshold speed of 45 mph. Figure 3 compares congested hours for INRIX against Wavetronix sensors (benchmarked dataset) for two different routes in Iowa under three different scenarios; daily, weekly and monthly. As shown in the figure, no pattern can be recognized in the diagrams for either routes 1 or 2. In an ideal diagram, all points would be plotted close to a 45 degree line, which is not the case and suggests a less than ideal agreement.

Methodological Flaws

Figure 4 shows a sample daily speed profile at the same location for INRIX and Wavetronix data. Point A shows a drop in speed which is detected by both INRIX (blue) and Wavetronix (orange). The latency (delay) between INRIX and Wavetronix can also be seen at point A with the INRIX data detecting the slowdown after the Wavetronix. A major problem contributing to the discrepancy in the number of congested events is latency.

Point B shows a speed drop in both time series, but they occurred above the 45 mph threshold line. In other words, both datasets detect a considerable speed drop, but are not identified as "congested" since they are above the predefined threshold (45 mph). At point C, part of the INRIX time series goes above the 45 mph threshold line indicating that it was uncongested for that period. Similarly, point D indicates a small drop (still greater than 15 min) labelled as congested for Wavetronix, but not for INRIX. These example contradictions compelled us to consider the detection algorithm and identify alternatives to use a fixed-speed threshold for performance calculations.

According to the research conducted by Adu-Gyamfi et al. (2017), it is recommended to consider 12 min as the maximum allowable latency (delay) time between sensor and segment reported traffic speeds. In our analysis, we examined the distribution of both detection and recovery latencies. Figure 5 shows that expanding the maximum allowable latency to 16 min yields a much higher agreement between Wavetronix and INRIX datasets. The detection latency is defined by subtracting Wavetronix detection time from INRIX detection time. For instance, if a congestion is detected at 4:00 PM and 4:06 PM by Wavetronix and INRIX, respectively, the detection latency would be + 6 min implying 6 min of delay in congestion detection using INRIX. The same happens for recovery latency. It should be noted that negative latency means INRIX is detected or recovered a congestion earlier than Wavetronix which occurs very rare.

Based on the methodological concerns shown, it was concluded that an alternative method to capture big changes in speed profile (slope) should be considered and that this should be free of any fixed threshold. To capture the maximum number of speed drops, a change point detection algorithm was considered.

Methodology

Change Point Detection Algorithm

Time-series analysis is used widely in fields, such as medicine, aerospace, finance, business, entertainment and transportation. Time-series data are sequences of temporal

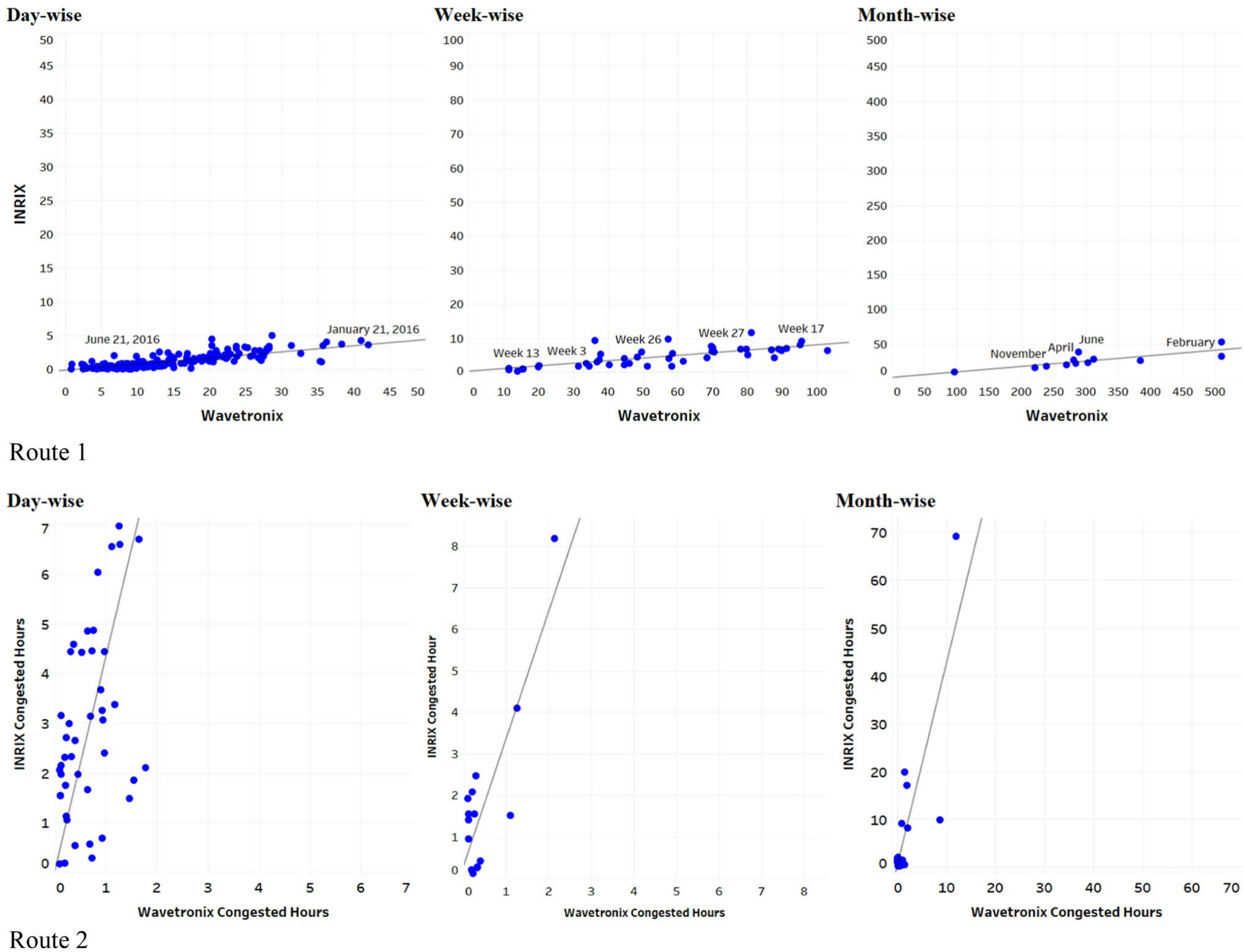

Fig. 3 Day-wise, week-wise and month-wise congested hours for INRIX vs Wavetronix computed using a fixed-threshold method for Route 1 (upper) and Route 2 (lower) in Iowa

measurements that describe the behaviour of systems. These behaviours can vary over time due to external circumstances and/or internal systematic changes (Montanez et al. 2015). Change point detection (CPD) is a method of finding sudden changes in the data when a property of the time series changes (Kawahara et al. 2012). Change point detection is similar in concept to segmentation, edge detection, event detection and anomaly detection all of which are commonly used in industry. Change point detection is also used to model and predict events like medical condition, climate change, speech recognition, image analysis and human activity and preferences. Generally, a change point detection algorithm has two parts which are the search method and cost function. The search method solves the change point detection problem with a known or unknown number of segments. The cost function measures the goodness-of-fit for the sub-signal to a specific model. In this analysis, a bottom-up segmentation method performed better than other search methods including dynamic programming, pruned exact linear time (PELT), binary segmentation and window-based change point detection. For the cost function, a kernelized mean change outperformed other functions including least absolute deviation, least squared deviation, Gaussian process change, linear model change, autoregressive model change and Mahalanobis-type metric.

Bottom-up Change Point Detection

A bottom-up change point detection is a sequential approach used to perform fast signal segmentation. It is a generous procedure contrary to binary segmentation. It starts with many change points and successively deletes the less significant ones. As the first step, the signal is divided in numerous segments along a regular grid. Then, adjacent segments are successively merged according to their similarities. The benefits of a bottom-up segmentation includes low complexity (of the order of $O(n\log n)$, where n is the number of samples), the ability to extend any single change point detection

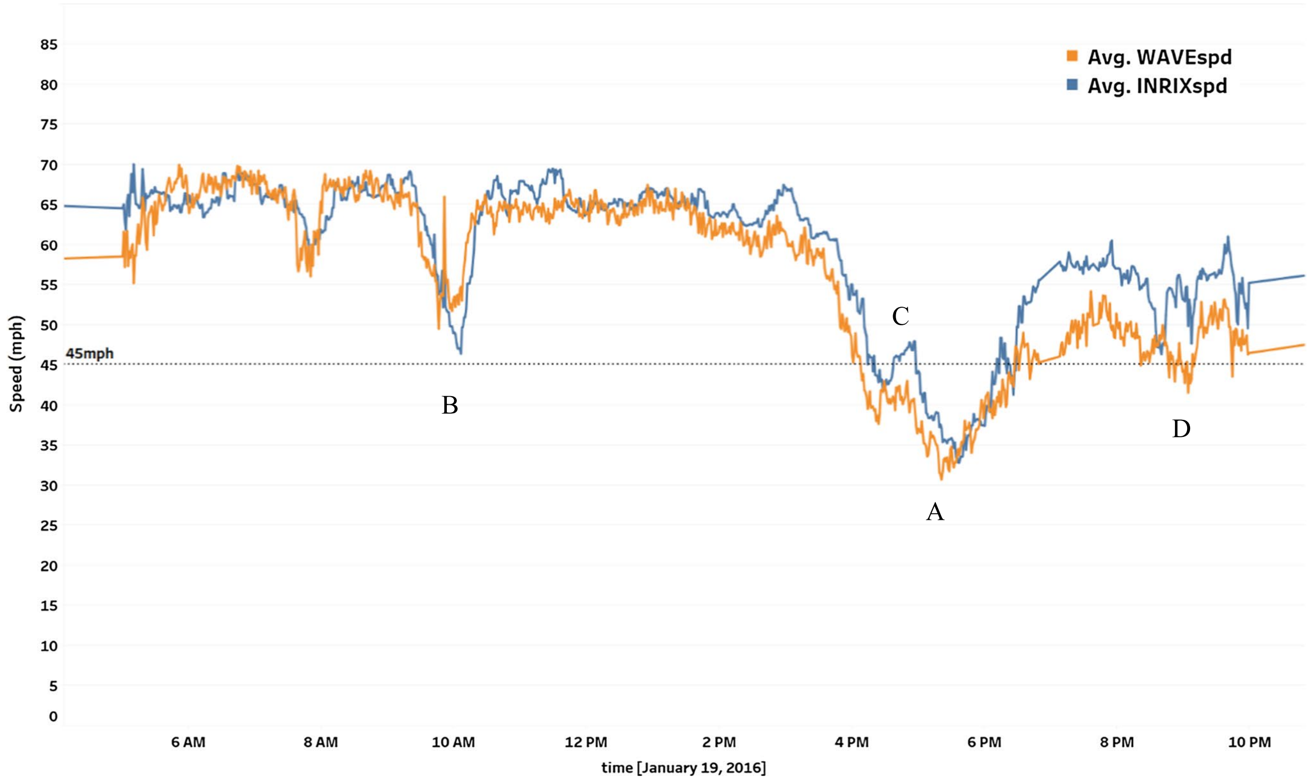

Fig. 4 Speed time series of INRIX (blue) and Wavetronix (orange)

method to multiple change points and; finally, the ability to perform in any number of regimes whether already known or not.

Kernelized Mean Change

In this method, we assumed a positive semi-definite kernel $k(\cdot,\cdot) : R^d \times R^d \mapsto R$ and its associated feature map $\Phi : R^d \mapsto H$ (where H is an appropriate Hilbert space), this cost function is able to detect changes in the mean of the embedded signal $\{\Phi(y_t)\}_t$ (Arlot et al. 2012; Arthur et al. 2020). Formally, for a signal $\{y_t\}_t$ on an interval I,

$$c(y_I) = \sum_{t\in I} ||\Phi(y_t) - \bar{\mu}||_H^2 \qquad (1)$$

where $\bar{\mu}$ is the empirical mean of the embedded segment $\{\Phi(y_t)\}_t \in I$. Also, kernel is the radial basis function (rbf):

$$k(x, y) = \exp\left(-\gamma||x - y||^2\right). \qquad (2)$$

where $\|\cdot\|$ is the Euclidean norm and $\gamma > 0$ is the so-called bandwidth parameter. It is determined based on the median heuristics. In other words, it is equal to the inverse of median of all pairwise distances.

Figure 6 shows a sample time series with two drops; one is sharp and the other one is smooth. The proposed method is applied on the time series and successfully detected both drops.

Final Results

Based on the analysis of the same number of segment–sensor pairs over the same period of time, Table 5 indicates significant improvements in calculating the number of congested events. It clearly shows that almost all of congestion are detected by both datasets and the number of missing congestion and false detected congestion by INRIX has decreased significantly. Additionally, Fig. 7 shows updated daily congested hour computations by the change point detection method. As can be seen, it has significantly improved and is very close to the ideal situation which is a 45∘ line. Using the proposed method, calculated congested hour by INRIX is very close to the benchmark congested hour calculated by Wavetronix dataset.

The change point detection algorithm delivered a higher accuracy and significantly improved congestion detection compared to the traditional fixed-threshold method. As it only applied to a limited number of sensor–segment pairs on one specific route, it was decided to develop and test this method on different routes with an increased number of locations.

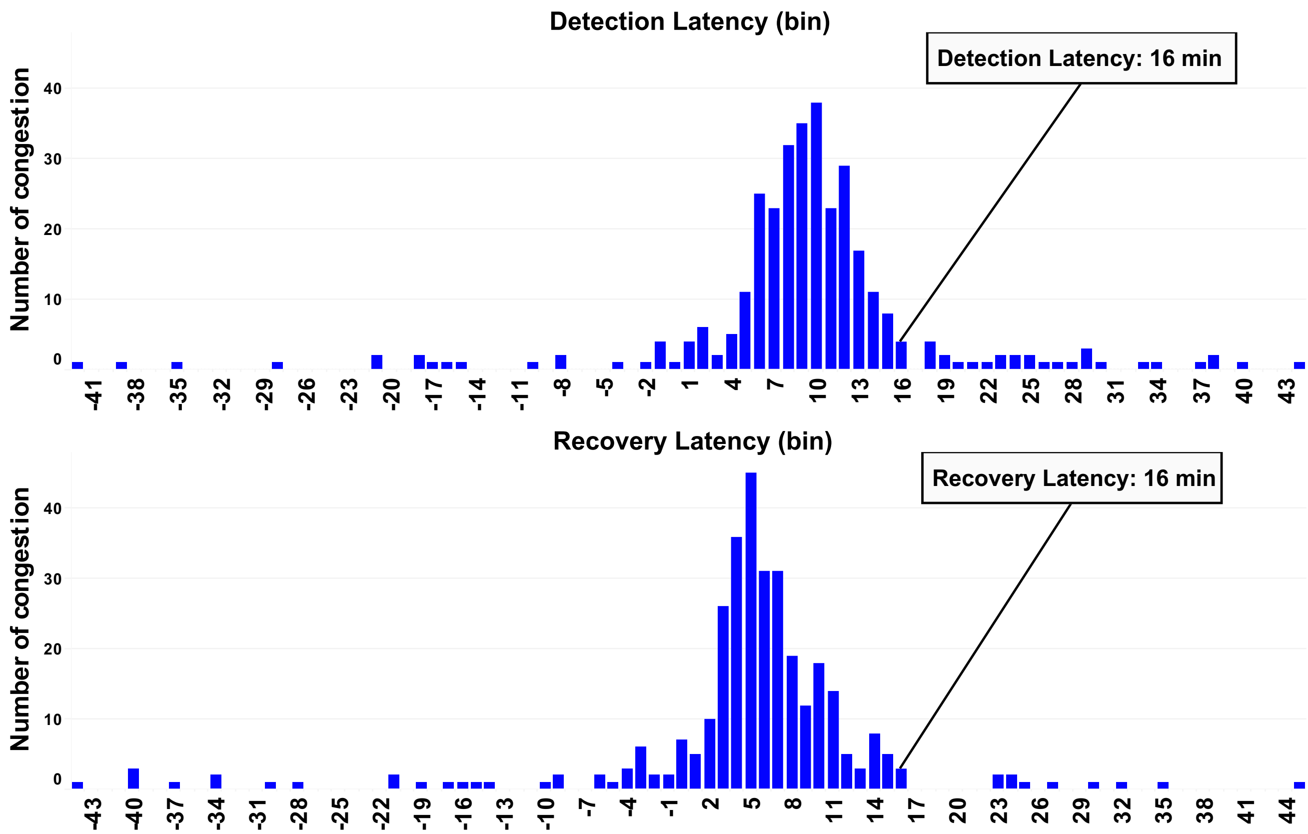

Fig. 5 Distribution of **a** detection latency and **b** recovery latency

Fig. 6 Change point detection method with bottom-up segmentation as search method and kernelized mean change as cost function. Two speed drops are detected in red

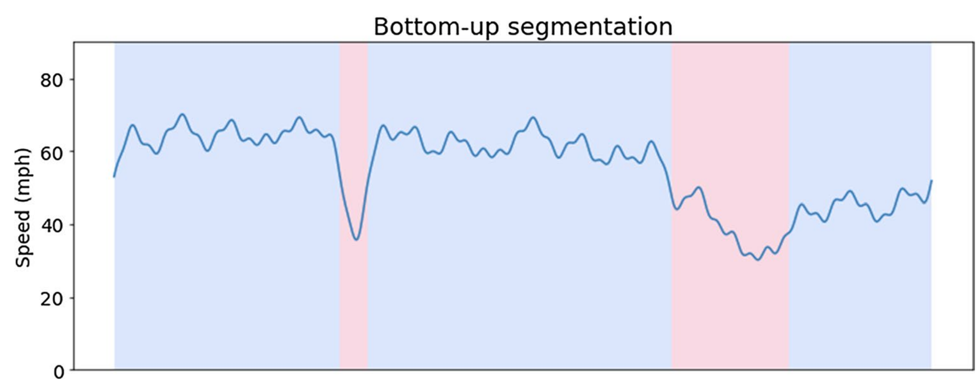

Table 5 Reliability of probe data in detecting congestion events using change point detection method

		INRIX			
		Detect congestion		No congestion detected	
Wavetronix	Detect congestion	True positive: 794		False negative: 19	
		Recurring	Non-recurring	Recurring	Non-recurring
		732	62	17	2
	No congestion detected	False positive: 16		True negative: –	
		Recurring	Non-recurring	Recurring	Non-recurring
		15	1		

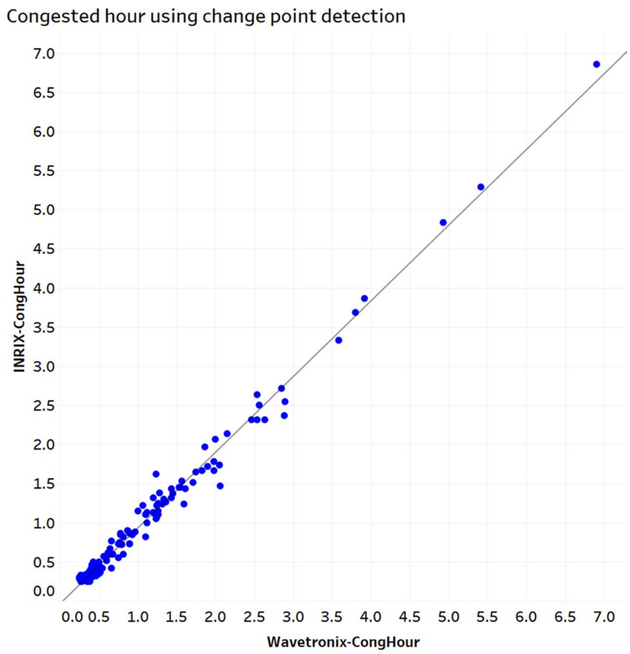

Fig. 7 Congested hour of INRIX vs Wavetronix computed by change point detection method

Table 6 Reliability of probe data in detecting congestion events for 5 major routes using change point detection method

Route	True positive (total/R/NR)	False negative (total/R/NR)	False positive (total/R/NR)
I-29	274/226/48	3/1/2	5/3/2
I-29/80	151/114/37	4/1/3	2/2/0
I-380	96/30/66	0/0/0	1/1/0
I-235	782/559/223	5/3/2	13/11/2
I-35/80	515/395/120	3/1/2	18/7/11

R recurring congestion, *NR* non-recurring congestion

Table 6 demonstrates the high accuracy using the change point detection method in calculating the number of congested events and speed drops using probe-sourced and sensor-based datasets. Table 6 contains five different routes with 64 sensor–segment pairs which were chosen in the state of Iowa over the year 2017.

Regarding congested hour as a second performance measure, Fig. 8 shows the significant improvement in calculation using the change point detection algorithm. Similar to Fig. 7, Fig. 8 confirms the robustness of the proposed method in calculating congested hour for different routes in Iowa.

Conclusion and Recommendation

This research evaluated probe-sourced streaming data from INRIX, to study its characteristics as a data source for calculating traffic performance measures. For this purpose, Wavetronix, a commonly used infrastructure sensor data source, was selected as the benchmark. The agreement between data sources was evaluated by two different measures; number of congested events and congested hours.

For both performance measures, a traditional fixed-threshold congestion detection method was initially used. The lack of efficiency and high number of errors in congestion detection by probe data and the lack of overlaps between probe congested hour data and Wavetronix inspired the development of a solution for congestion detection. Consequently, a change point detection method was utilized and its accuracy was proven by applying the new method to five different routes in Iowa. Finally, the concurrence of the two sensor systems in terms of congestion or no congestion is very promising in the context of NPMRDS. These results are important because the error mechanisms are very different in the two systems; hence, it is unlikely to get false concurrence. Also, the discrepancies between the two systems should not be taken as errors of the INRIX, certainly an INRIX error could trigger one of these, but so too a radar error, but also the simple difference between link speed and point speed can cause a discrepancy when neither sensor is in error.

The following recommendations are offered for transportation agencies who are augmenting traditional traffic data with probe-based services for wider coverage under restricted budgets:

Probe-based speed data provide great value to agencies especially in areas not covered by sensors. However, as with sensors, probe data are not without error. Therefore, it is critical that agencies understand these issues and continue to examine and consider alternative methods to remove error prior to calculating and reporting performance metrics to the public.

Change point detection appears to address errors observed when calculating traffic performance measures using a fixed-speed congestion threshold. Agencies should consider this method when using probe data to calculate performance measures.

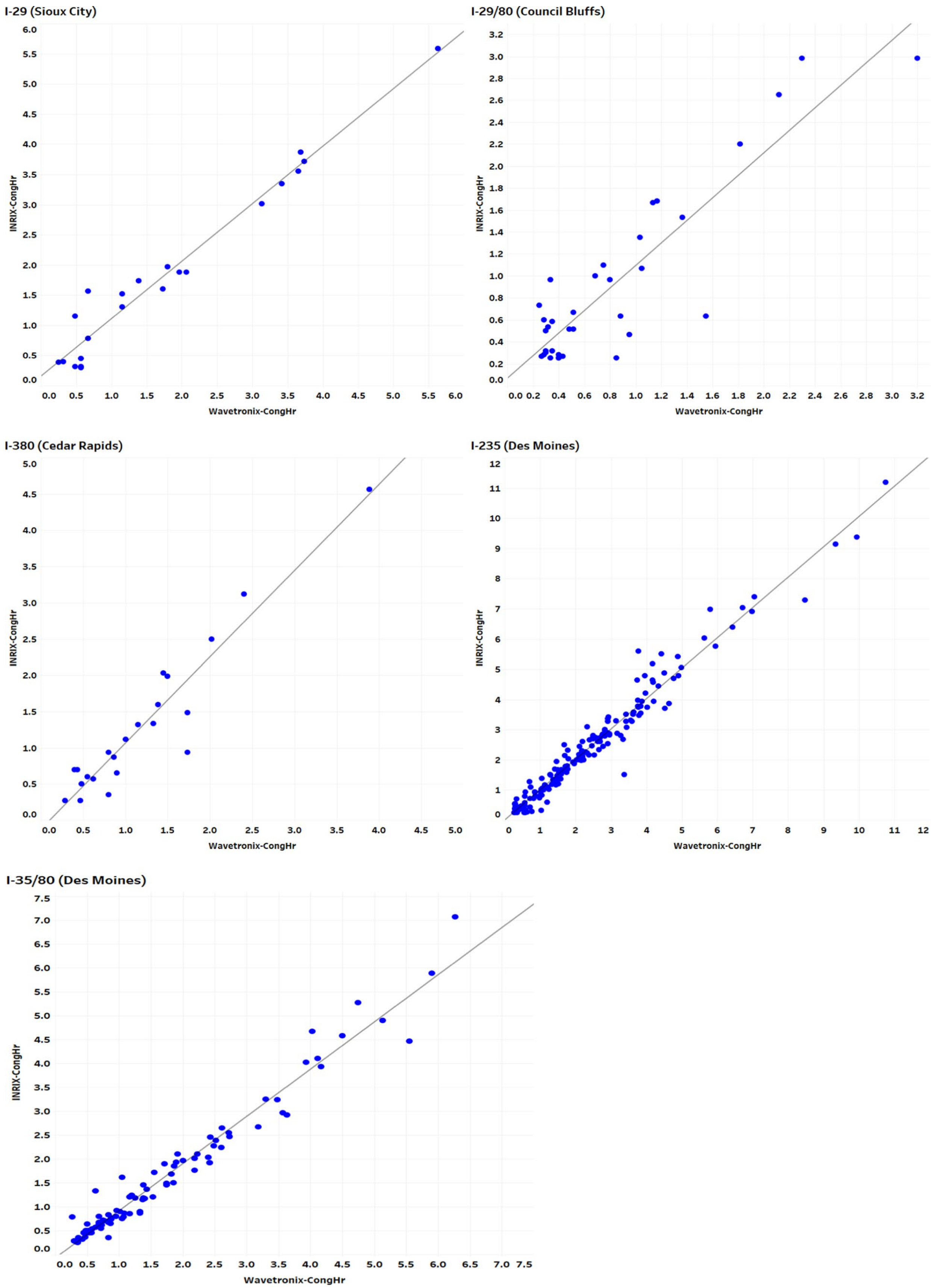

Fig. 8 Congested hour of INRIX vs Wavetronix for 5 major routes computed by change point detection method

Future Work

The great potential of probe data encourages deeper exploration into the characteristics of this data source, to build models that encourage traffic experts to trust probe-based reports without need for cross-checking or further validation. Also, the methodology will be extended to differentiate between recurring and non-recurring in future papers from lab. Also, the reliability measures need to be improved.

The authors plan to compute other important performance measures including delay and travel time per mile (reliability) and check their efficiency and accuracy using proposed change point detection method against traditional method. Moreover, authors attempt to develop models for the potential use in automatically correcting latency measurements from probe data.

Acknowledgements Our research results are based on the work supported by the Iowa DOT Office of Traffic Operations Support Grant. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Iowa DOT.

References

Adu-Gyamfi YO, Sharma A, Knickerbocker S, Hawkins N, Jackson M (2017) Framework for evaluating the reliability of wide-area probe data. Transport Res Rec J Transport Res Board 2643:93–104. https://doi.org/10.3141/2643-11

Ahsani V, Amin-Naseri M, Knickerbocker S, Sharma A (2019) Quantitative analysis of probe data characteristics: coverage, speed bias and congestion detection precision. J Intell Transport Syst 23(2):103–119

Aliari Y, Haghani A (2012) Bluetooth sensor data and ground truth testing of reported travel times. Transport Res Rec J Transport Res Board 2308:167–172. https://doi.org/10.3141/2308-18

Araghi BN, Hammershøj Olesen J, Krishnan R, Tørholm Christensen L, Lahrmann H (2015) Reliability of bluetooth technology for travel time estimation. J Intell Transport Syst 19(3):240–255. https://doi.org/10.1080/15472450.2013.856727

Arlot S, Celisse A (2012) Kernel change-point detection. Citeseer. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.751.542&rep=rep1&type=pdf. Preprint arXiv:1202.3878

Arthur G, Karsten B, Malte R et al (2020) A kernel two-sample test. Jmlr.Org. https://www.jmlr.org/papers/v13/gretton12a.html

Bell MGH, Iida Y (2003) The network reliability of transport . In: Proceedings of the 1st international symposium on transportation network reliability (INSTR), Pergamon

Belzowski BM, Ekstrom A (2014). Stuck in traffic: analyzing real time traffic capabilities of personal navigation devices and traffic phone applications. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/102509/102984.pdf?sequence=1&isAllowed=y

Chakraborty P, Adu-Gyamfi YO, Poddar S, Ahsani V, Sharma A, Sarkar S (2018) Traffic congestion detection from camera images using deep convolution neural networks. Transp Res Rec 2672(45):222–231. https://doi.org/10.1177/0361198118777631

Coifman B (2002) Estimating travel times and vehicle trajectories on freeways using dual loop detectors. Transport Res Part A Policy Pract 36(4):351–364. https://doi.org/10.1016/S0965-8564(01)00007-6

Cookson G, Pishue B (2016). INRIX global traffic scorecard. Inrix Global Traffic Scorecard, 44. https://media.bizj.us/view/img/10360454/inrix2016trafficscorecarden.pdf

Day C, Remias S, Li H, Mekker M, Mcnamara M, Cox E, Wasson J (2015) 2013–2014 Indiana mobility report. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1006&context=imr

Dudek C, Messer C, Record NN-TR (1974) Incident detection on urban freeways. https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_483374_38

FDOT (2012) Probe data analysis evaluation of NAVTEQ, TrafficCast, and INRIX® travel time system data in the Tallahassee region evaluation of NAVTEQ, TrafficCast, and INRIX® travel time system data. https://www.fdot.gov/traffic/ITS/Projects_Deploy/2012-03-26_Probe_Data_Analysis_v2-0.pdf

Feng W, Bigazzi A, Kothuri S, Bertini R (2010) Freeway sensor spacing and probe vehicle penetration: impacts on travel time prediction and estimation accuracy. Transport Res Rec J Transport Res Board 2178:67–78. https://doi.org/10.3141/2178-08

FHWA (2013) Work zone performance measurement using probe data. https://ops.fhwa.dot.gov/wz/resources/publications/fhwahop13043/fhwahop13043.pdf

FHWA (2017) 2016 urban congestion trends. https://ops.fhwa.dot.gov/publications/fhwahop17010/fhwahop17010.pdf

Gong L, Fan W (2017) Applying travel-time reliability measures in identifying and ranking recurrent freeway bottlenecks at the network level. https://doi.org/10.1061/JTEPBS.0000072

Haghani A, Hamedi M, Sadabadi KF (2009) I-95 Corridor Coalition Vehicle Probe Project: Validation of INRIX data July–September 2008. https://www.i95coalition.org/wp-content/uploads/2015/02/I-95-CC-Final-Report-Jan-28-2009.pdf

Hu J, Fontaine MD, Park BB (2015) Field evaluations of an adaptive traffic signal—using private-sector probe data. J Transport Eng 142(1):1–9. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000806

Kawahara Y et al (2012) Sequential change-point detection based on direct density-ratio estimation. Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.10124

Kim S, Coifman B (2014) Comparing INRIX speed data against concurrent loop detector stations over several months. Transport Res Part C Emerg Technol 49:59–72. https://doi.org/10.1016/j.trc.2014.10.002

Lin L, Li W, Peeta S (2019) Efficient data collection and accurate travel time estimation in a connected vehicle environment via real-time compressive sensing. J Big Data Anal Transport 1(2):95–107

Lin L, Pccta S, Wang J (2018) Efficient collection of connected vehicle data based on compressive sensing. In 2018 21st international conference on intelligent transportation systems (ITSC), pp 3427–3432. IEEE, New York

Lindveld CDR, Thijs R, Bovy PHL, der Zijpp NJ (2000) Evaluation of online travel time estimators and predictors. Transport Res Rec 1719:45–53

Lomax TJ, Texas Transportation Institute, & National Research Council (U.S.), Transportation Research Board (1997) Quantifying congestion. National Academy Press. https://books.google.co.uk/books?hl=en&lr=&id=q6NSRBQ3uGIC&oi=fnd&pg=PA12&dq=Lomax,+T.%3B+Turner,+S.%3B+Shunk,+G.+NCHRP,+R.+398:+Quantifying+Congestion%3B+Transportation+Research+Board:+Washington,+DC,+USA,+1997.&ots=NQiYk5spRv&sig=0FdQfJEDPXiAI-ijLJMbrFGpwFw#v=onepage&q&f=false

Lomax T, Schrank D, Turner S, Margiotta R (2003) Selecting travel reliability measures. Texas Transportation Institute, Cambridge Systematics

Lomax T, Schrank D, Turner S, Margiotta R (2003) Selecting travel reliability measures. https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2003-3.pdf

Lu X, Lee J, Chen D, Bared J, Dailey D, Shladover SE (2014) Freeway micro-simulation calibration: case study using Aimsun and VISSIM with detailed field data. In: Transportation Research Board 93rd annual meeting, Washington, D.C, 12–16 January, pp 1–17

Mcleod DS, Morgan G, Mcleod M. (2012) Florida's mobility performance measures and experience. In: Transportation Research Board, 91st annual meeting

MoDOT (2017) Tracker: measures of departmental performance. https://www.modot.org/about/documents/Tracker_July17/July2017FinalTracker.pdf

Montanez G, Amizadeh S, AAAI NL (2015) Inertial hidden Markov models: modelling change in multivariate time series. Aaai.Org. https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/viewFile/9475/9470

Nam D, Park D, Khamkongkhun A (2005) Estimation of value of travel time reliability. J Adv Transport 39(1):39–61. https://doi.org/10.1002/atr.5670390105

National Academies of Sciences, Engineering (2008) Cost-effective performance measures for travel time delay, variation, and reliability. National Academies Press, Washington, D.C. https://doi.org/10.17226/14167

Peniati J (2004) Operational solutions to traffic congestion

Persaud B, Hall F, Record L H-TR (1990) Congestion identification aspects of the McMaster incident detection algorithm. Trid.Trb.Org. https://trid.trb.org/view/352877

Poddar S, Ozcan K, Chakraborty P, Ahsani V, Sharma A, Sarkar S (2018) Comparison of machine learning algorithms to determine traffic congestion from camera images. In: Transportation Research Board 97th annual meeting, Washington, DC, 7–11 Jan 2018. https://trid.trb.org/view/1496348

Pu W (2012) Analytic relationships between travel time reliability measures. Transport Res Rec J Transport Res Board 2254(1):122–130. https://doi.org/10.3141/2254-13

Rakha H, El-Shawarby I, Arafeh M (2010) Trip travel-time reliability: Issues and proposed solutions. J Intell Transport Syst Technol Plan Oper 14(4):232–250. https://doi.org/10.1080/15472450.2010.517477

Remias S, Brennan T, Day C, Summers H, Cox E, Horton D, Bullock D (2013) 2012 Indiana mobility report. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1004&context=imr

Schrank D, Eisele, B, Lomax T, Bak J (2015) 2015 Urban mobility scorecard, vol 39. Texas A&M Transportation Institute

Schrank D, Eisele B, Lomax T (2012) TTI's 2012 urban mobility report. Texas A&M Transportation Institute. https://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-report-2012.pdf

Sekuła P, Marković N, Laan ZV, & Sadabadi KF (2017). Estimating historical hourly traffic volumes via machine learning and vehicle probe data: a Maryland Case Study. https://arxiv.org/abs/1711.00721

Sharifi E, Hamedi M, Haghani A, Sadrsadat H (2011) Analysis of vehicle detection rate for bluetooth traffic sensors: a case study in Maryland and Delaware. In: 18th ITS world congress, Orlando, FL, 16–20 Oct 2011

Sharma A, Ahsani V, Rawat S (2017). Evaluation of opportunities and challenges of using INRIX data for real-time performance monitoring and historical trend assessment

Subrat M, Matthew Wolniak EB, Sadabadi KF (2015) 2015 Maryland state highway mobility report

Sun Z, Gu W, Feng J, International X Z-P (2011) Threshold value based traffic congestion identification method. Researchgate.Net. https://www.researchgate.net/profile/Zhanquan_Sun/publication/267244613_Threshold_Value_Based_Traffic_Congestion_Identification_Method/links/55aac7eb08aea9946724163f.pdf

Turner, S. (2013). Developing twin cities arterial mobility performance measures using GPS speed data. https://www.lrrb.org/pdf/201314.pdf

Uno N, Kurauchi F, Tamura H, Iida Y (2009) Using bus probe data for analysis of travel time variability. J Intell Transport Syst 13(1): 2–15. https://doi.org/10.1080/15472450802644439

Venkatanarayana R (2017) Considerations for calculating arterial system performance measures in Virginia. https://www.virginiadot.org/vtrc/main/online_reports/pdf/17-r2.pdf

WSDOT (2013) The 2013 Corridor capacity summary. https://wsdot.wa.gov/publications/fulltext/graynotebook/CCS13.pdf

WSDOT (2014) Gray notebook 52—for the quarter ending December 31, 2013. https://wsdot.wa.gov/publications/fulltext/graynotebook/Dec13.pdf

Young S (2007) Real-time traffic operations data using vehicle probe technology. In: 2007 mid-continent transportation research symposium. https://www.ctre.iastate.edu/pubs/midcon2007/YoungVehicleProbe.pdf

Zheng F, Li J, van Zuylen H, Liu X, Yang H (2018) Urban travel time reliability at different traffic conditions. J Intell Transport Syst 22(2):106–120. https://doi.org/10.1080/15472450.2017.1412829