Natural Hazards
https://doi.org/10.1007/s11069-020-04427-5

ORIGINAL PAPER

# Estimating impacts of recurring flooding on roadway networks: a Norfolk, Virginia case study

**Shraddha Praharaj[1] · T. Donna Chen[1] · Faria T. Zahura[1] · Madhur Behl[2] · Jonathan L. Goodall[1]**

Received: 10 June 2020 / Accepted: 11 November 2020


**Abstract**

Climate change and sea level rise have increased the frequency and severity of flooding events in coastal communities. This study quantifies transportation impacts of recurring flooding using crowdsourced traffic and flood incident data. Agency-provided continuous count station traffic volume data at 12 locations is supplemented by crowd-sourced traffic data from location-based apps in Norfolk, Virginia, to assess the impacts of recurrent flooding on traffic flow. A random forest data predictive model utilizing roadway features, traffic flow characteristics, and hydrological data as inputs scales the spatial extent of traffic volume data from 12 to 7736 roadway segments. Modeling results suggest that between January 2017 and August 2018, City of Norfolk reported flood events reduced 24 h citywide vehicle-hours of travel (VHT) by 3%, on average. To examine the temporal and spatial variation of impacts, crowdsourced flood incident reports collected by navigation app Waze between August 2017 and August 2018 were also analyzed. Modeling results at the local scale show that on weekday afternoon and evening periods, flood-impacted areas experience a statistically significant 7% reduction in VHT and 12% reduction in vehicle-miles traveled, on average. These impacts vary across roadway types, with substantial decline in traffic volumes on freeways, while principal arterials experience increased traffic volumes during flood periods. Results suggest that analyzing recurring flooding at the local scale is more prudent as the impact is temporally and spatially heterogeneous. Furthermore, countermeasures to mitigate impacts require a dynamic strategy that can adapt to conditions across various time periods and at specific locations.

**Keywords** Recurring flooding · Crowd-sourced data · Data predictive model · Impact analysis

✉ T. Donna Chen
tdchen@virginia.edu

Extended author information available on the last page of the article

Published online: 04 January 2021

## 1 Introduction

Recurring flooding is a type of disruption commonly observed in coastal cities due to heavy rainfall, high tides, or both. Historically, recurring flooding has been a low-frequency, low spatial-, and temporal-scale disruption to the transportation system, assumed to have minor impacts. However, in recent years, rising sea levels and coastal flooding are increasingly affecting coastal communities across the USA, with almost 30 coastal cities witnessing more than double the number of annual flood days in the 2010s as compared to the 1950s (US EPA 2016, p. 36). NOAA (Sweet et al. 2016, p. 17) projects tide-related flooding in east coast cities in the USA to increase threefold by 2030 and tenfold by 2050, relative to 2019 estimates. Tidal flooding, combined with rainfall-induced flooding, is expected to increase the number of flood events in these cities even further. With continued relative sea level rise, recurring flooding is expected to occur more frequently, and propagate to more inland locations.

As recurrent flooding frequency and intensity increases, there is a growing need to understand the subsequent impacts of these flood events on people and civil infrastructure (traveler response, frequency and duration of roadway closures, reduction of infrastructure life, impact on stormwater drainage capacity, etc.). Coastal recurring flooding is considered a minor disruption compared to consequences of catastrophic storms. However, inundated areas in coastal cities greatly deteriorate the mobility of road users, by increasing travel delay and by disrupting their ability to complete trips. The existing literature on transportation disruptions due to flooding are largely focused on major storms, with much of the research oriented towards evacuation and rehabilitation efforts, and not the recovery of daily transportation activities.

Only a few studies have examined the transportation impacts of recurring flooding through projected data and scenario analysis (e.g., Suarez et al 2005; Chang et al 2010). Due to lack of availability of real-time crowdsourced datasets, none of the previous studies have used empirical data to observe the impacts of recurrent flooding on the roadway network. This study is the first to utilize empirical data to examine the impact of recurring flooding on roadway users, using a combination of agency-provided and crowdsourced datasets in Norfolk, Virginia. The analysis in this study is twofold: first, analyzing the daily (24 h) impacts on a citywide scale using agency-provided flood report data, then analyzing the time-of-day (TOD) impacts on a localized scale using crowdsourced flood report data.

## 2 Literature review

There are few studies focused on recurrent flooding in the transportation infrastructure resilience literature due to its historical categorization as low severity and low frequency. Among studies examining recurring flooding events, most use projected transportation and hydrological data to create disruption scenarios for predicting roadway impacts. Lu and Peng (2011) developed an accessibility-based analysis to quantify roadway network vulnerability to sea level rise (SLR). They considered land use and population variables in defining an accessibility index in the Miami, Florida network. Their model assessed portions of the roadway network and traffic analysis zones that would be inundated at different SLR scenarios. Jacobs et al. (2018) combined flood projection maps with annual average daily traffic data (AADT) from the US Federal Highway

Administration's Highway Performance Monitoring System along east coast highways. They estimated the current total vehicle hours of delay due to recurring flooding at over 100 million hours annually, and projected this delay will increase to 160 million vehicle-hours by 2020 and 1.2 billion vehicle-hours by 2060. On a citywide scale, Suarez et al. (2005) estimated the indirect costs of increased flooding in Boston by examining the effects of coastal flooding due to SLR and riverine flooding due to heavy rainfall events. The study simulated these effects in an urban transportation model and projected an increase in delay and lost trips of around 80% in 2100 compared to 2000, with an assumed SLR of 0.3 cm per year and an increase in intensity of heavy rainfall events of 0.31% per year. Sadler, et al. (2018) estimated the impact of SLR on flooding of roadways, by running different high tide scenarios for the cities of Norfolk and Virginia Beach. Critical roadways vulnerable to flooding were identified based on the annual average weekday daily traffic, elevation of roadways, and different high tide and storm surge scenarios. The study yielded an annual generalized estimate that nearly 10% of major roadways would be affected for every high tide event by the year 2100. As a part of a larger study in Portland, Oregon, Chang et al. (2010) used predicted flooding frequency and locations based on hydrological models to determine the impacts of coastal flooding on the roadway network in 2035, using the four-step regional travel demand model. The study found an inconsistent relationship between precipitation and travel disruption impacts and estimated a negligible change in vehicle-miles traveled (VMT). However, vehicle-hours of delay increased by up to 10% in one of the sub-areas analyzed.

None of the studies discussed thus far use empirical data for analysis. Only a few studies have characterized the impact of flood events on transportation systems using empirical data, and these studies focus on large-scale disruptions. For example, New York City taxi and subway ridership datasets were made publicly available for 2010 through 2013, during which hurricanes Irene and Sandy significantly disrupted the transportation and power networks in the area. Zhu et al. (2016) and Donovan and Work (2017) used these datasets to propose new methodologies to quantify city-scale transportation system resilience to extreme events. Zhu et al. presented resilience curves, which showed that Hurricane Sandy had a slower transportation recovery rate than Hurricane Irene. Resilience of the roadway network was found to be higher in both disruptions compared to the subway network. In the post-disruption period of Hurricane Sandy, Donovan and Work found an increase in delay of over two minutes per mile about two days after the hurricane had struck, although a faster traffic flow was observed during most of the post-disruption period.

A significant challenge to using real-time data for estimating the impacts of disruption incidents is simply the lack of availability of such data through traditional agency sources. Installing sensors on the roadway network to obtain comprehensive real-time information is cost-prohibitive, which makes passively generated crowdsourced data an attractive source for transportation analysis. Crowdsourced data is not regulated and may contain erroneous reporting due to misunderstanding, confusion, carelessness, incompetence, or even intent to deceive (Ouyang et al 2016). This data, however, may still contain useful information to improve the understanding of any situation. Various studies (e.g., Amin-Naseri et al. 2018; Lenkei 2018; Goodall and Lee 2019) have conducted analyses to quantify traffic incidents through the crowdsourced navigation app, Waze (owned by Google). This study is the first to apply crowdsourced real-time data to assess the impacts of recurring flooding, using citizen-reported flood incident data from Waze and crowdsourced traffic data from location-based service (LBS) app data aggregator Streetlight (founded, 2011). The goal of this study consists of two parts to contrast the extent of analysis possible with traditional vs. crowdsourced data: analyzing citywide impacts of recurring flooding using agency-obtained flood incident data, and localized impacts using crowdsourced flooding incident data.

## 3 Data sources

For this study, a combination of agency-provided traffic volume data in limited locations and crowdsourced LBS data is used to build a predictive model which estimates the traffic volumes across the entire Norfolk roadway network. Transportation datasets include agency-provided roadway geometry data along with agency-provided and crowdsourced traffic volumes. Hydrology datasets include agency-provided tide and rainfall data, along with agency-provided and crowdsourced flood incident data. The following subsections discuss each dataset in detail, and Table 1 shows basic summary statistics for the various datasets.

### 3.1 Roadway characteristics data

The roadway characteristics considered in the data predictive model consist of geometric features obtained from Hampton Roads Regional Travel Demand Model (HRRTDM), provided by Virginia Department of Transportation (VDOT), and include number of lanes, posted speed limit, and per lane capacity for each of 7737 unique links in the city of Norfolk. Thus, the roadway network analyzed in this study was limited to the links in HRRTDM (shown in blue in Fig. 1), which includes interstates, freeways, arterials, and collectors in Norfolk. Minor streets (shown in gray in Fig. 1) are represented by aggregate centroid connectors (blue links in Fig. 1 that end at a cluster of gray links), but are not individually analyzed.

### 3.2 Traffic volume data

#### 3.2.1 Agency-provided traffic volume data

VDOT collects traffic volume data at 12 permanent continuous count stations (CCS) on freeways and arterials within the city of Norfolk (shown in orange in Fig. 1). This data is collected at 15-min intervals throughout the year and, for this study, was obtained at all 12 count stations for 2017 and 2018. These volumes were used as ground truth traffic volumes for model development in the data predictive model framework for volume estimation, explained in the Methods section.

#### 3.2.2 Crowdsourced traffic volume and speed data

Streetlight Data (http://www.streetlightdata.com/founded 2011) is a commercial platform that provides road segment volume data, origin–destination (OD) analysis data, and zonal activity data. In this platform, Streetlight (StL) trip indices (estimated link volumes) and travel speeds are projected from signals or pings (called StL trip counts) generated from applications using LBS on mobile phones, tablets, connected cars, and other electronic devices. LBS data-enabled devices are reported to have an

**Table 1** Summary statistics for datasets utilized in this study

| Dataset (data source) | # locations | Frequency of data collection | Minimum value | Maximum value |
|---|---|---|---|---|
| Traffic volumes (VDOT CCS) | 12 | 15 min | 0 vehicles | 2080 vehicles |
| Traffic volumes (trip counts from Streetlight Data) | 7737 | Time period (3 or 6 h) | 0 pings | 863 pings (during 9a-3p period) |
| Flood incident data (City of Norfolk) | Theoretically all | Daily | 1 location in a single day | 40 locations in a single day |
| Flood incident data (WAZE) | Theoretically all | 1 min | 1 location in a single day | 107 locations in a single day |
| Rainfall data (HRSD) | 7 | 1 h | 0 inches | 2.26 inches |
| Tide level data (NOAA) | 1 | 6 min | −10.59 m | 13.61 m |

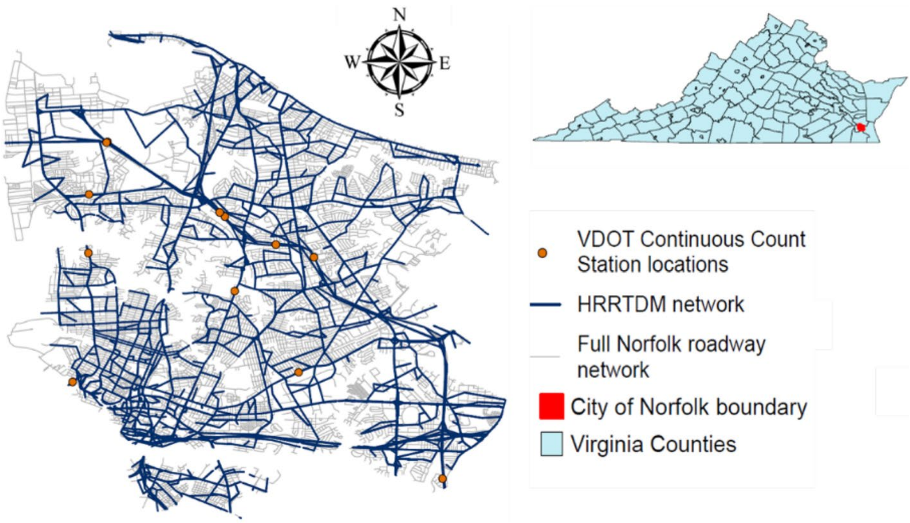



**Fig. 1** City of Norfolk, Virginia roadway network coverage with data sources

approximately 23% penetration rate among all traffic (Streetlight Insight 2018). Thus, with sufficient samples of StL trip counts across the Norfolk roadway network, Streetlight Data is able to provide significantly greater spatial coverage of traffic volume estimates (StL trip indices) compared to the VDOT CCS data. The network links coded in blue in Fig. 1 show the extent of the spatial coverage requested from Streetlight Data for this study. Most of the links from the HRRTDM model coincide with the OpenStreetMap (OSM) layer used in Streetlight Data. The links in the roadway network that do not align with OSM layer either use the volumes of links that they intersect, or are marked as not-applicable.

For this study, data was requested from Streetlight Data's online platform for five TOD periods (12a-6a, 6a-9a, 9a-3p, 3p-6p, and 6p-12a) for each day in the study period (January 2017 to August 2018). The time period classification (within each day) matches the peak and off-peak period definitions in the HRRTDM. The platform then processed and filtered the data which matched the input requirements and then gave an output file. However, links without any StL trip data during a time period are omitted from the Streetlight output file (and thus omitted in this study). For the citywide flood impact analysis, 54% of roadway links have time period matches with StL trip data, while in the localized flood impact analysis, 67% of the links have time period matches with StL trip data.

The calibration process of the StL trip indices on these links and in zonal analyses is internal to Streetlight Data as a part of their data cleaning and imputation process, but is based on AADT metrics from VDOT roadways in Bristol, VA (Streetlight Insight 2018). Thus, direct application of the StL trip indices as link volume estimates is inappropriate for other parts of Virginia, including Norfolk. In examining the ratio of VDOT CCS volumes to StL trip indices for 35 randomly sampled days in 2017, these ratios were closer to 1 during peak periods of travel, but ranged from 0.2 to 26 at other time periods in the day (with a median value of 1.57 across off-peak periods). Thus, in this study, raw StL trip counts are used as an input into a data predictive model which then estimates traffic volumes, described in the Methods section.

To verify the speed data, StL link speed estimates were compared with INRIX data provided by RITIS (a relatively more established and commercially available data source which estimates travel speeds and travel times based on location information emitted by GPS-based mobile devices) for two weekdays in March 2017 across all CCS locations for all time periods. There were no statistically significant differences observed when comparing both speed datasets. Thus, this study utilizes the Streetlight link travel time data directly in the analysis.

## 3.3 Flood incident data

### 3.3.1 Agency-provided flood incident data

Flood incident data from the City of Norfolk was collected from city employees' reported flood locations in a mobile phone application (System to Track, Organize, Record, and Map (STORM)). Due to the lack of a timestamp associated with the flood reports (only dates were included), flood report data is coded as a binary variable, with any day with one or more flood reports considered a flood day (FD) and any day without flood reports considered as a non-flood day (NFD). The spatial distribution of flood locations could not be considered while using the city's flood incident dataset due to a lack of citywide spatial representation of the small sample of reports for each FD. The incident data collected spanned from January 2017 to August 2018 with floods reported on 10 unique days. This data is used to analyze the city-wide flood impacts.

### 3.3.2 Crowdsourced flood incident data

The mobile navigation application Waze also collects flood incident reports (alongside other incident data like road closures and congestion) via their real-time information reporting tool. The application provides aggregated user-reported incident data via its Waze for Cities data sharing program, open to public entities worldwide. In this study, Waze (2017–2018) timestamped and location-specific incident reports related to flooding in Norfolk (106 unique days with flooding between August 2017 and August 2018) are analyzed. While the Waze flood report data is not comprehensive of all instances of roadway flooding in Norfolk, its spatial coverage is significantly greater than the agency data available through City of Norfolk.

## 3.4 Hydrological data

The hydrological characteristics considered in this study include rain and tidal gauge data. The rainfall data, collected at 15 min intervals, is from the Hampton Roads Sanitation Department (HRSD), which has seven rain gauge stations in the city. Tide level data is available through the sole tidal gauge in the city at Sewell's Point, and data collected every six minutes is archived and obtained by NOAA Tides and Currents. These two datasets were aggregated to match the time periods specified in the traffic volumes data description.

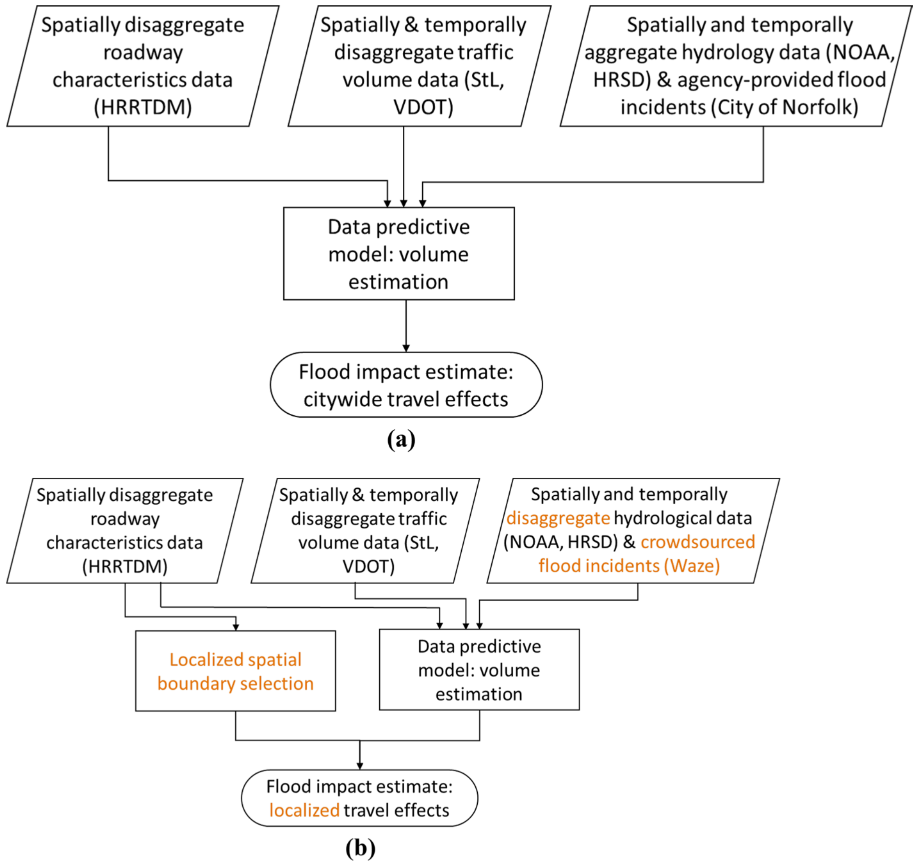



**Fig. 2** **a** Citywide flood impact analysis framework. **b** Localized flood impact analysis framework

## 4 Methods

The overall framework to estimate citywide impacts of recurring flooding is shown in Fig. 2a. For localized impact analysis (Fig. 2b), two components of the framework are changed. First, a localized boundary around flood incident reports is selected based on the roadway network structure and AADT of adjacent roadway links. Second, Waze flood incident report data is used in place of the City of Norfolk data.

### 4.1 Data predictive model: volume estimation

For the volume estimation step of the flood impact analysis, different roadway, traffic flow, and hydrological variables were used to create a data predictive model, which uses a set of input variables to provide traffic volume estimates on each roadway link for each time period. For the citywide flood impact analysis, the number of lanes, speed limit, and capacity per lane data were collected from the HRRTDM; trip counts and speeds were collected

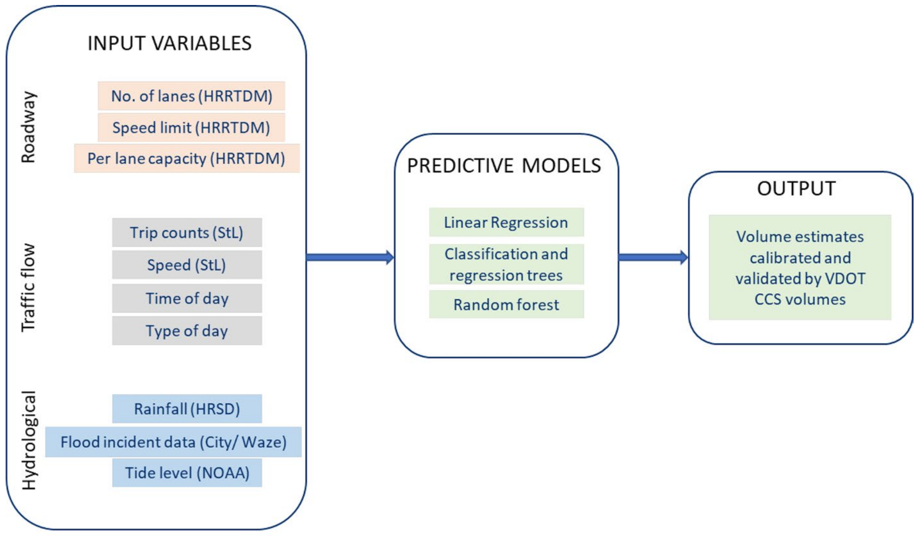



**Fig. 3** Traffic volume prediction process

from Streetlight Data; tide levels from NOAA; rainfall values from HRSD (averaged over the city); and flood incident reports from City of Norfolk. In the localized impacts study, the same datasets are used, except the City of Norfolk flood incident report data is replaced by Waze flood incident reports, and the rainfall values are interpolated at the point of the flood incident report (from 7 rain gauges across the city, using the inverse distance weighting [IDW] technique in ArcGIS). As seen in Fig. 3, several predictive models were tested in this study to predict the link volumes in order to determine a preferred model with the best prediction accuracy without overfitting the data.

A linear regression model was first developed as a baseline model for comparison. Classification and regression trees (CRT) and Random forest (RF) models, which group data points with similar dependent variable values together based on their independent variables, were also developed.

In CRT models, a parent node in the CRT is divided based on any independent variable into two child nodes, such that each child node is more homogenous (or less impure) than the parent node. Homogeneity is measured by the least squared deviation measure of impurity (within-node variance). The process continues until constraints, such as a minimum number of cases per node, maximum tree depth, node homogeneity, or a minimum change in improvement, are satisfied. In this study, 70% of the observations were reserved for training the dataset, and 30% were reserved for validation. Through trial and error, a 50–20 split of data in parent and child nodes was used (a minimum of 50 observations from the dataset in the parent node, and a minimum of 20 observations in the child node), which was pruned to avoid overfitting. Pruning reduces the size of decision trees in an attempt to prevent the nodes from being too specific (thereby keeping the model more generalized).

In random forests, similar to the CRT models, a 70–30 split of observations are used for training and testing the dataset, respectively. Random sampling of data subsets is performed on the training dataset to fit the samples into a model prediction, while reducing the total error in the model. The response variables are divided into groups until the resulting predictions reach a minimum amount of node impurity (sum of the squared deviations between the predicted and actual value, a measure of error). Random forests are a strong modeling technique and much more robust than a single decision tree. They aggregate many decision trees to limit overfitting as well as error due to bias and therefore yield useful results. CRT models are also prone to overfitting the data, with random forest addressing the issue by creating various groups of randomly selected regression trees while running the model.

Once the model is developed, errors are calculated for training and testing the data, which are used as criteria for selecting the appropriate model. Errors calculated for these models are the root mean squared error (RMSE) and normalized root mean squared error (NRMSE), given by Eqs. 1 and 2.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{I}\left(v_{\mathrm{obs},i} - v_{\mathrm{model},i}\right)^2}{n}} \tag{1}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{v_{\mathrm{obs,max}} - v_{\mathrm{obs,min}}} \tag{2}$$

where $i$ = roadway link, $v_{\mathrm{obs},i}$ = observed VDOT CCS volumes, $v_{\mathrm{model},i}$ = predictive model's estimated volumes.

### 4.2 Flood impact estimation (citywide and localized analysis)

The citywide impacts of roadway flooding borne by travelers on the Norfolk roadway network are accounted for by comparing the 24 h VHT across the entire city on a day with a recorded flood incident versus days without a flood incident. To assess the citywide VHT on a FD, the products of the estimated link volume and the average travel time for each link are aggregated across all TOD periods (Eq. 3). FD traffic volumes and VHT were compared with an average NFD to estimate the network-wide impacts of recurring flooding. For each FD, four NFDs were selected to obtain an average NFD (and its associated link volume and travel speeds) (Eq. 4). NFDs are selected from comparable days (workdays measured against other workdays, non-work days—weekends and holidays—measured against other non-work days) within 3 weeks prior to and after the FD to minimize effects of seasonal and weekday versus weekend traffic variation. However, the 3 days immediately before and after the FD were excluded to minimize potential anticipatory and residual traffic effects of the flood incident. Conceptually, this is similar to the approach taken by Zhu et al (2016, p. 2599), where the data was compared to the same day in the prior year to observe differences in traffic flow while accounting for seasonal traffic variation.

In the localized analysis, Waze flood incident reports are assigned to a TOD flood period (FP) based on the timestamp of the report, and the comparable non-flood periods (NFPs) are defined as the same TOD periods during the three weeks (of the same day type, e.g., workday or non-work day) prior to and after the FP, when no flood incident was reported within a one mile radius of the location of the flood incident report. The incidents reported in Waze have an associated time duration (time between when a flood incident is first reported to 30 min after the last "thumbs up," indicating that report is still true; or until someone reports a "thumbs down," indicator that the report no longer holds true). The maximum duration of an incident reported in Waze is under 3 h, which is shorter than any of the TOD periods considered for classification. Thus, the analysis here only accounts for impacts within the time period that contains the flood report (and not subsequent time periods). For each FP, all candidate NFPs within three weeks prior to and after the FP are considered to obtain an average NFP for link volume and travel speed comparisons.

Previous studies have shown that precipitation affects travel decisions and choices differently for peak and off-peak periods, weekday and weekend traffic, and in different seasons (Böcker et al. 2013, pp. 79–80). Thus, this study distinguishes the FDs (or FPs) with rainfall from those without, when considering candidate NFDs (or NFPs). Flood events in the study area are assumed to occur due to two environmental conditions: high tide, rainfall, or both. The candidate NFDs (or NFPs) for a high tide only FD (or FP) were picked from days (or periods) with no rainfall within the comparison window. For FDs (or FPs) with rainfall, the NFDs (or NFPs) were chosen from days (or periods) that experienced rainfall, but recorded no flooding. Traffic impacts due to flooding are then evaluated using Eq. 5, where change in VHT is assessed across links on a FD (or FP) compared to a NFD (or NFP).

$$\mathrm{VHT}_{iF} = \sum_{j} (v_{i,j} * \mathrm{tt}_{i,j})_F \tag{3}$$

$$\mathrm{VHT}_{i\mathrm{NF}} = 1/K\left(\sum_{k=1}^{K} \sum_{j} (v_{i,j} * \mathrm{tt}_{i,j})_{\mathrm{NF}k}\right) \tag{4}$$

$$\Delta\mathrm{Travel} = \sum_{i} [\left(\mathrm{VHT}_i\right)_F - \left(\mathrm{VHT}_i\right)_{\mathrm{NF}}] \tag{5}$$

where $i$ = roadway link, $j$ = time-of-day (TOD) period, $K$ = maximum number of comparable days/periods considered for average NFD/ NFP calculations (4 for citywide impacts and all possible NFPs during 3 weeks prior to and after flood periods for localized impacts), $\Delta$ Travel = change in VHT due to flooding, measured in veh-hrs, $\mathrm{tt}_{i,j}$ = travel time on segment $i$ during TOD $j$ on a FD/FP or NFD/NFP, $v_{i,j}$ = traffic volume on segment $i$ during TOD $j$ on a FD/FP or NFD/NFP, $(\mathrm{VHT}_{i,j})_F$ = vehicle-hours of travel on segment $i$ during TOD $j$ on a FD/ FP, $(\mathrm{VHT}_{i,j})_{\mathrm{NF}}$ = vehicle-hours of travel on segment $i$ during TOD $j$ on a NFD/FP.

### 4.3 Localized spatial boundary selection (localized analysis only)

Since crowdsourced flood incident reports from Waze are spatially and temporally disaggregate, it provides an opportunity to analyze flood impacts on a more localized scale, as flooding on one roadway link is unlikely to impact traffic throughout City of Norfolk in a homogeneous way. Thus, in the localized flood analysis, a spatial boundary can be used to define the affected area around the location of the flood incident report. The roadway links within the localized boundary were selected as those most likely to be affected by the flood incident, based on the network theory measure hub dependence (Rodrigue et al. 2017). Hub dependence, or *Hvalue,* is a measure of a node's vulnerability and represents the share of traffic borne by the highest volume traffic link among all links connected to a node, and is calculated as:

$$Hvalue = \frac{\left(\mathrm{AADT}_{ij}\right)_{\max}}{\sum_{j=1}^{J}\left(\mathrm{AADT}_{ij}\right)} \tag{6}$$

where $i$ = current node, $j$ = adjacent node, $J$ = maximum number of nodes that are connected to node $j$, $\mathrm{AADT}_{ij}$ = annual average daily traffic of the link between node $i$ and node $j$.

Weak nodes, exhibiting higher hub dependence values, are heavily dependent on the conditions of the connected links for movement of traffic, and disruptions on any link connecting to the weak node would greatly affect operations at the node. On the other hand, stronger nodes with lower hub dependence values have a more even distribution of traffic among the links they are connected to. Thus, their operations are less likely to be affected by disruptions of travel on any single connecting link. In a sense, nodes with lower hub dependence values may be more resilient when facing incidents and disruptions (Ducruet 2008).

Hub dependence values for all nodes in the network based on 2009 AADT (collected from the HRRTDM) were calculated. To define the spatial boundary of roadway links impacted by a flood incident report, the node nearest to the flood report observation is assigned as the affected node and all the links connected to that node as affected links, by default. Then, the *Hvalue* for nodes connecting to these affected links are compared. If the adjacent node has a lower *Hvalue* than the affected node, it implies that the affected node is relatively stronger, and does not rely as much on the adjacent node for movement of traffic. In this case, the spatial boundary was cut off at the link leading to the adjacent weaker node. On the other hand, if the adjacent node has a higher hub dependence value than the affected node, it implies that the adjacent node's operations are highly affected by operations of the affected node. In this case, the adjacent weak node becomes an affected node, and the spatial boundary is expanded to include the set of links connected to the new affected node(s). A 10% threshold (see Fig. 4) was used to ensure a sufficient difference in hub dependence values of adjacent nodes is observed before a boundary is set, and to prevent nodes with similar hub dependence values from being excluded from the boundary. The process of increasing the spatial boundary for each Waze-reported flood incident is repeated until there is a node that has a minimum 10% smaller hub dependence value than its previous adjacent node, or until the boundary of the study network (City of Norfolk border) was reached. This process of spatial boundary selection by hub dependence metric is illustrated in Fig. 4.

Figure 5 shows an example of the spatial boundary selection starting with a flood report assigned to node *A* with $H_{\text{value}} = 0.427$. Following the procedure in Fig. 4, the links within the localized boundary are shown in yellow. The links connected to node *A* are included in the spatial boundary by default. Next, adjacent nodes *B, C, D,* and *E* were evaluated. Since

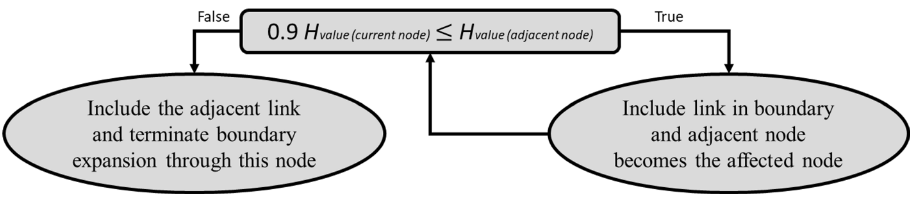



**Fig. 4** Framework for estimating localized spatial boundary of flooding impacts

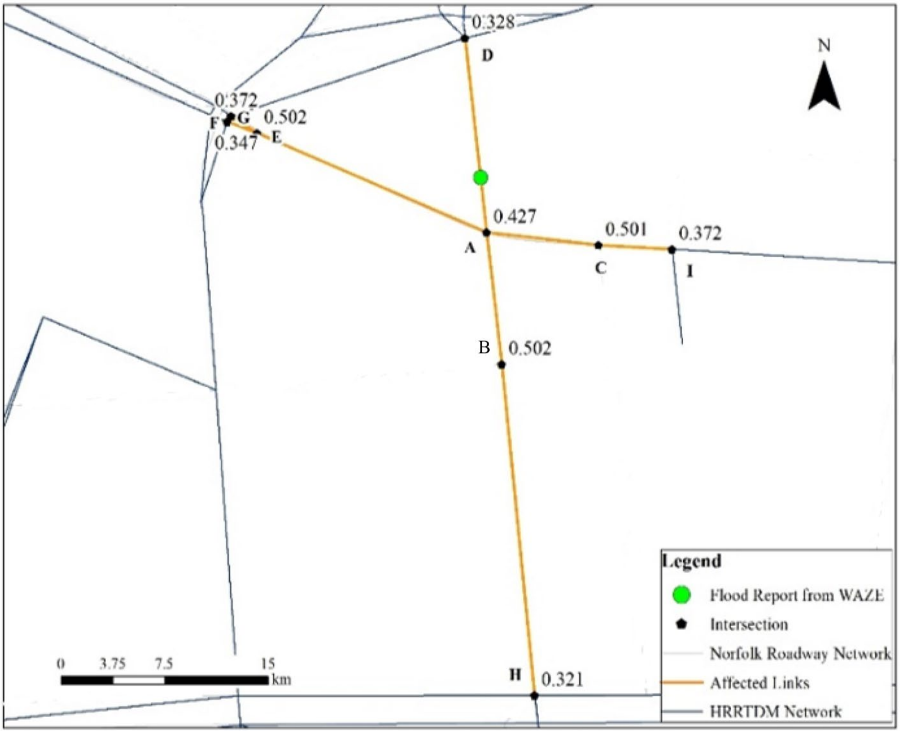



**Fig. 5** Localized spatial boundary assignment with flood incident assigned to Node A

the *Hvalue* of *B* was greater than that of *A*, *B* (and its connecting links) were included in the boundary. Moving on to the next adjacent node, *H*, which has an *Hvalue* less than 90% of node *B*, the boundary is terminated at node *H*. This propagation of spatial boundary takes place through each link that is connected to node *A* and terminates when the hub dependence value falls below 90% of the previous node's *Hvalue*.

## 5 Results and discussion

### 5.1 Flood impact estimation at CCS locations

Variation of traffic across all VDOT CCS locations with available data was first compared to understand the baseline roadway network impacts due to flood incidents. The CCS are strategically placed on major arterials and freeways where there are no historic congestion or bottlenecking issues to ensure accurate volume estimates. Due to the specific location selection criteria and sparse spatial representation of CCS across Norfolk, an accurate estimation of the flooding impacts throughout the network cannot be made, but general trends can be observed, as shown in Table 2. The CCS data was collected on the same FDs and NFDs (January 2017–August 2018) considered in the citywide flood impact analysis, at 15 min intervals throughout the day. The table shows average speeds and volumes at across 15 min intervals.

**Table 2** CCS volume and speed trends, categorized by roadway type

| Count station ID | Facility type | FD average volume (veh/15 min) | NFD average volume (veh/15 min) | Δ volume (%) | Δ volume *p* value | FD average speed (mph) | NFD average speed (mph) | Δ speed (%) | Δ speed *p* value |
|---|---|---|---|---|---|---|---|---|---|
| 150114_1 | Principal arterial | 188.35 | 210.80 | −10.6 | <0.001 | 37.64 | 38.77 | −2.9 | <0.001 |
| 050169_2 | Freeway | 709.50 | 789.70 | −10.2 | <0.001 | 59.94 | 62.10 | −3.5 | <0.001 |
| 050306_2 | Freeway | 594.59 | 655.47 | −9.3 | <0.001 | 60.00 | 62.44 | −3.9 | <0.001 |
| 150065_4 | Freeway | 553.78 | 627.52 | −11.8 | 0.001 | 71.91 | 77.76 | −7.5 | <0.001 |
| 150110_2 | Principal arterial | 88.76 | 106.74 | −16.8 | <0.001 | 39.95 | 46.94 | −14.9 | <0.001 |
| 150110_4 | Principal arterial | 91.77 | 111.75 | −17.9 | <0.001 | 35.52 | 38.63 | −8.1 | <0.001 |
| 150119_2 | Principal arterial | 174.11 | 188.87 | −7.8 | <0.001 | 36.84 | 39.01 | −5.6 | <0.001 |
| 150119_4 | Principal arterial | 144.77 | 160.87 | −10.0 | <0.001 | 36.19 | 37.14 | −2.6 | <0.001 |
| 150120_4 | Principal arterial | 98.32 | 109.86 | −10.5 | <0.001 | 43.84 | 48.55 | −9.7 | <0.001 |
| Total | All | 293.77 | 329.06 | **−10.7** | – | 46.87 | 50.14 | **−6.5** | – |

Bold shows the aggregate value of the table

The CCS volumes and speeds on the 10 FDs (as reported by the City of Norfolk) were compared with their respective NFD counterparts. The NFDs used in this study come from the 3 weeks within (before and after) the flood incident report day, excluding the week of the flood incident. The data was compared at 15 min intervals for the 24 h day and then aggregated over the 10 FDs in the 20 month study period. A two-sample, one-tailed paired Student's *t* test was conducted, and revealed link volumes and speeds at CCS locations were statistically significantly lower on FDs than on NFDs (with all $p$ values $< 0.01$). An average 11% decrease in traffic volumes and 7% decrease in travel speeds were observed across the CCS locations on FDs. This result suggests that FDs consistently experience decreased traffic demand. At the same time traffic volumes are decreased, those who are traveling on FDs also experienced slower speeds, which is indicative of increased travel times.

### 5.2 Traffic volume estimation model training and validation

While general trends of the traffic impacts of recurring flooding can be observed with the spatially limited CCS data, a network-wide impact assessment requires more spatial coverage. Here, the proposed data predictive model (using agency-provided roadway characteristics, hydrology data, and flood reports along with crowdsourced traffic flow data) estimates volumes across all freeway and arterial links in Norfolk. To create the ground truth dataset for model calibration (training) and validation (testing), all the days in the 20 month period were divided into categories based on environmental conditions. The days were categorized as combinations of three levels of rainfall (rainfall = 0 in., 0 < rainfall ≤ 0.5 in., and rainfall > 0.5 in.) and three levels of tide (tide level < 1 ft, 1 ft < tide level ≤ 2 ft, and tide level > 2 ft), thus creating 9 combinations of environmental conditions based on rainfall and tide levels. Twenty percent of the days in each category were randomly selected to create the ground truth dataset, ensuring representation of all combinations of rainfall and tide conditions.

Linear regression, CRT, and random forest models were developed with all the variables previously mentioned in three categories: hydrological, roadway, and traffic flow characteristics. The model fits (measured by RMSE and NRMSE values), along with statistically significant variables, are shown in Table 3 for comparison across models. The two random forest model specifications outperformed linear regression and CRT in terms of model fit. For random forest models, the first model (RF1) used only the roadway and traffic flow characteristics as input variables. In this model, the StL dynamic crowdsourced trip counts had less importance than other static variables such as number of lanes and type of day, which is counterintuitive. When the hydrological variables are introduced into the random forest mode specification (RF2), tide level and rainfall were found to be the least important variables, but the StL trip count became the highest significance variable, which is intuitive. In the RF2 model, other relatively high importance variables described patterns associated with traffic flow in specific environments, such as TOD, per lane capacity, posted speed limit, and link speed. This RF2 model specification also proved to be the best performing (with the lowest RMSE and NRMSE).

### 5.3 Citywide roadway network impacts

The RF2 model was propagated to the HRRTDM roadway network in Norfolk to predict the volumes on each roadway segment across all TODs. The HRRTDM roadway network

**Table 3** Comparison of data predictive models

| Model type | RMSE | NRMSE | Significant/high importance variables* | Insignificant/low importance variables** |
|---|---|---|---|---|
| Linear regression | 2384.43 | 0.085 | Rainfall<br>Tide level<br>Flooding<br>Number of lanes<br>Posted speed limit<br>TOD<br>StL trip counts | Per lane capacity<br>Segment speed<br>Type of day |
| CRT | 2512.22 | 0.157 | StL trip count<br>Posted speed limit<br>TOD | Rainfall<br>Tide level<br>Flooding<br>Number of lanes<br>Per lane capacity Segment speed<br>Type of day |
| RF1 with roadway and traffic characteristics | Train: 1573.97<br>Test: 3399.23 | Train: 0.026<br>Test: 0.058 | TOD<br>Per lane capacity<br>StL trip counts<br>Posted speed limit | Number of lanes<br>Type of day |
| RF2 with roadway, traffic, and hydrologic variables | Train: 1341.98<br>Test: 2865.60 | Train: 0.022<br>Test: 0.048 | StL trip counts<br>TOD<br>Per lane capacity<br>Posted speed limit<br>Segment Speed | Type of day<br>Tide level<br>Rainfall |

Variable importance predictive power of the variables in the random forest model

*High importance variables: normalized variable importance over 0.5

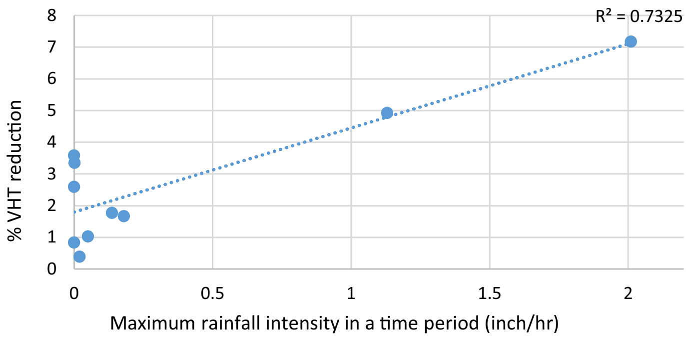
**Low importance variables: normalized variable importance under 0.1

consists of 7736 segments, which were fed into Streetlight Data to retrieve the associated StL trip counts, segment speed, and travel time on each segment. StL trip counts and segment speed, along with other roadway and hydrological variables, were used as inputs into the random forest model (RF2) to obtain volume estimates on FDs and NFDs. Total VHT on FDs and NFDs was calculated per Eq. 2. There were 11 FDs recorded in the 20 month analysis period by City of Norfolk employees. One of the FDs was discarded due to insufficient comparable NFDs within the 6 week window. Table 4 shows the total VHT on each FD compared to the corresponding NFDs.

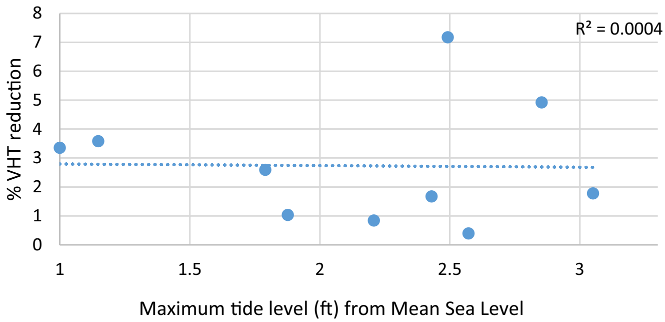
Table 4 shows that, based on the predicted vehicle volumes, network-wide VHT was consistently reduced on FDs compared to NFDs in the citywide analysis, on average by 3%. This decrease is consistent with trends in the CCS analysis, though not substantial. The result may be attributed to two factors: (1) cumulative change in VHT may not be a sufficient metric for quantifying the effect of flooding, and (2) the spatial aggregation at the city-level may be too large for assessing the impacts of local recurring flooding. Since the VDOT CCS data also showed a reduction in travel speeds and personal vehicle volumes on FDs, it is likely that the individual effects of increased travel time and reduced volumes were somewhat nullified when multiplying the two for the cumulative effect measured in VHT. Decreased network VHT may imply higher rates of abandoned trips, which would signify an economic impact of recurring flooding (due to decreased business transactions, work productivity loss, etc.). Considering a net difference in VHT over the roadway

**Table 4** Citywide network impact summary

| Flood report date | Max tide level (ft) | Max rainfall intensity (in/hr) | FD VHT (veh-hrs) | NFD VHT (veh-hrs) | FD-NFD VHT (veh-hrs) | Δ VHT (%) |
|---|---|---|---|---|---|---|
| 3/6/2017 | 1.000 | 0.002 | 6,261,710 | 6,478,423 | −216,713 | −3.35 |
| 4/5/2017 | 2.208 | 0.000 | 6,594,179 | 6,649,965 | −55,786 | −0.84 |
| 4/12/2017 | 1.148 | 0.000 | 6,311,667 | 6,546,158 | −234,491 | −3.58 |
| 8/29/2017 | 2.854 | 1.130 | 6,005,899 | 6,316,644 | −310,745 | −4.92 |
| 9/5/2017 | 1.877 | 0.050 | 6,416,229 | 6,482,857 | −66,627 | −1.03 |
| 3/6/2018 | 3.051 | 0.137 | 6,133,580 | 6,244,388 | −110,808 | −1.77 |
| 5/18/2018 | 2.430 | 0.180 | 6,447,560 | 6,557,150 | −109,590 | −1.67 |
| 5/22/2018 | 1.791 | 0.000 | 6,462,511 | 6,634,472 | −171,960 | −2.59 |
| 8/11/2018 | 2.493 | 2.010 | 4,503,163 | 4,851,056 | −347,898 | −7.17 |
| 8/13/2018 | 2.572 | 0.020 | 6,526,115 | 6,551,927 | −25,813 | −0.39 |




**(a)**




**(b)**

**Fig. 6** Comparison of roadway network impacts by hydrological variables

network of the entire city may also temper the more significant local effects of flooding experienced by specific areas within the city. Thus, a more spatially disaggregate analysis of flooding impacts on the roadway network is necessary to fully understand the effects of recurring flooding.

While the sample of flood days is small ($N = 10$), relationships between hydrological variables and traffic impacts on the roadway network still appear to exist. Figure 6a and b shows the relationship between rainfall intensity, tide level, and VHT reduction. Increasing rainfall intensity had a relatively consistent positive correlation with reduction of VHT (Fig. 6a); however, no relationship appears to exist between tide level and reduction in VHT (Fig. 6b). It is possible that effects of tide-induced flooding are more local than that of rain-induced flooding. In other words, tide-induced flooding impact roadway segments near the shoreline in a spatially and temporally consistent manner. However, this temporal and spatial granularity cannot be analyzed with the City of Norfolk flood report data, which lacks timestamps and representative spatial coverage.

## 5.4 Localized roadway network impacts

Using City of Norfolk flood incident reports which are neither spatially nor temporally disaggregated, only a citywide analysis of recurring flooding impacts is feasible. However, Waze flood incident report data contains both timestamp and location (latitude and longitude) data, allowing for local analysis of recurring flooding impacts. In this section, the analysis of recurring flooding impacts is defined by five TOD periods (consistent with HRRTDM definitions) and a local geographic boundary around the location of the flood incident report (see Sect. 4.3 for methodology for localized spatial boundary selection). StL trip count data is too sparse to be aggregated at an hourly interval, particularly for off-peak travel periods, hence the selection of TOD periods for analysis. Total VHT during a flood period (FP) and an average non-flood period (NFP) is estimated using Eqs. 3 and 4, after obtaining the link volumes from the RF2 model (normalized RMSE for the training data: 0.03, testing data: 0.07). The candidate NFPs used in this study come from 3 weeks before and after the flood incident report, excluding any TOD with another reported flood incident within one mile radius of the original report. After removing the FPs that did not have any candidate NFPs, fell outside the roadway network being analyzed (e.g., on local streets, ramps, or centroid connectors), or had insufficient Streetlight Data trip counts to predict link volumes inside the localized spatial boundary, 340 flood report observations remained (representing 51% of the original Waze flood reported incidents between August 2017 and August 2018).

A link-by-link impact analysis for all roadway links within the localized boundaries was conducted to compare travel during FPs compared to NFPs. The localized network impact summary of the link-by-link analysis (categorized by time period) for workdays is shown in Table 5.

The discussion about localized impacts of flood incidents here is focused on workdays, because the number of Waze flood reports on non-workdays (weekends and holidays) was too small per TOD period for analysis (*N* for each non-workday TOD ranged from 2 to 24). On workdays, across almost all time periods, the majority of affected links experience a decline in both speed and traffic volume during FPs compared to NFPs (example for 3p-6p and 6p-12a time periods shown in Fig. 7), as indicated by the median percent differences in Table 5. However, the average link speed and volume changes during FPs compared to NFPs are predominantly positive, indicating a few links experiencing significantly higher speeds or volumes during FPs. Per Table 5, workday evening (3p-6p) and overnight (6p-12a) periods experience the most statistically significant reductions in travel speed, traffic volumes, and change in vehicle hours and miles of travel during flood periods compared to non-flood periods, and are examined in more detail here. As seen in Fig. 7, for the workday 6p to 12a time period, the affected links observe a maximum 100% decrease in volume, while select links experience volume increases in excess of 300%. The significant change in traffic volumes in the workday evening and night periods (3p-6p and 6p-12a) is also reflected in the VMT measures, which show a reduction in travel during FPs compared to NFPs, with an aggregate VMT decline of 12% across all affected links. There are similar statistically significant reductions in VHT observed in the evening and night periods, suggesting that road users are either avoiding travel, or changing their destinations.

Figure 8 shows the spatial distribution of links experiencing the greatest increase (top tenth percentile, marked in blue) and greatest decrease (bottom tenth percentile, marked in orange) in traffic volumes during 3p-6p (Fig. 8a and b) and 6p-12a (Fig. 8c and d). The links in black represent all the other links that were considered in the localized spatial boundary analysis for the TOD period (middle 80th percentile). A total of 2167 unique links are considered in the localized boundary analysis across all time periods, with the 3p to 6p time period showing the highest number of impacted links (985 unique links, with several links affected multiple times across multiple flood reports between August 2017 and August 2018). As shown in Fig. 8, the Hampton Boulevard corridor is highly impacted during both TOD periods, with some flood incidents causing a large increase in traffic

**Table 5** Localized network impact summary for workdays

| Day type | Period | # of Waze flood reports | Δ Speed | | | Δ Volume | | | Δ VHT | | Δ VMT | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | (FP-NFP) | | | (FP-NFP) | | | (FP-NFP) | | (FP-NFP) | |
| | | | Mean % diff | Medi-an % diff | *p* value | Mean % diff | Medi-an % diff | *p* value | Aggr. % diff | *p* value | Aggr. % diff | *p* value |
| Workday | 12a-6a | 10 | 8 | − 5 | 0.41 | 2 | − 19 | 0.01* | − 25 | 0.04* | − 16 | 0.05* |
| Workday | 6a-9a | 26 | 1 | − 4 | 0.13 | − 3 | − 11 | 0.06 | − 1 | 0.89 | − 9 | 0.00* |
| Workday | 9a-3p | 59 | 19 | 0 | 0.42 | 21 | 0 | 0.00* | 4 | 0.08 | 3 | 0.11 |
| Workday | 3p-6p | 126 | 23 | − 4 | 0.15 | 1 | − 3 | 0.00* | − 9 | 0.00* | − 16 | 0.00* |
| Workday | 6p-12a | 65 | 23 | 0 | 0.00* | 6 | − 3 | 0.02* | − 6 | 0.01* | − 5 | 0.04* |

*Statistically significant at 95% confidence level

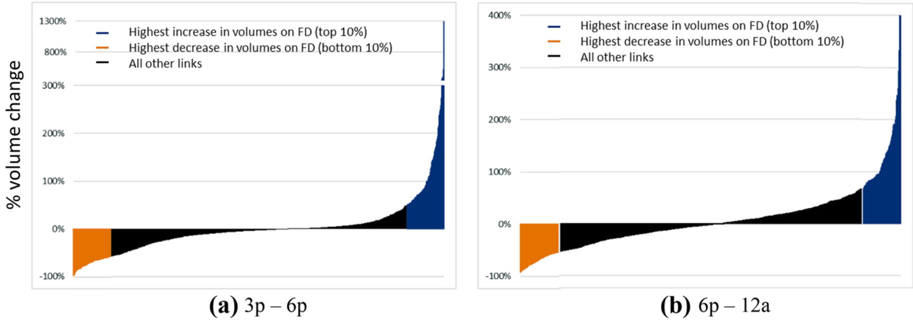



**Fig. 7** Distribution of change in traffic volume between flood periods & non-flood periods

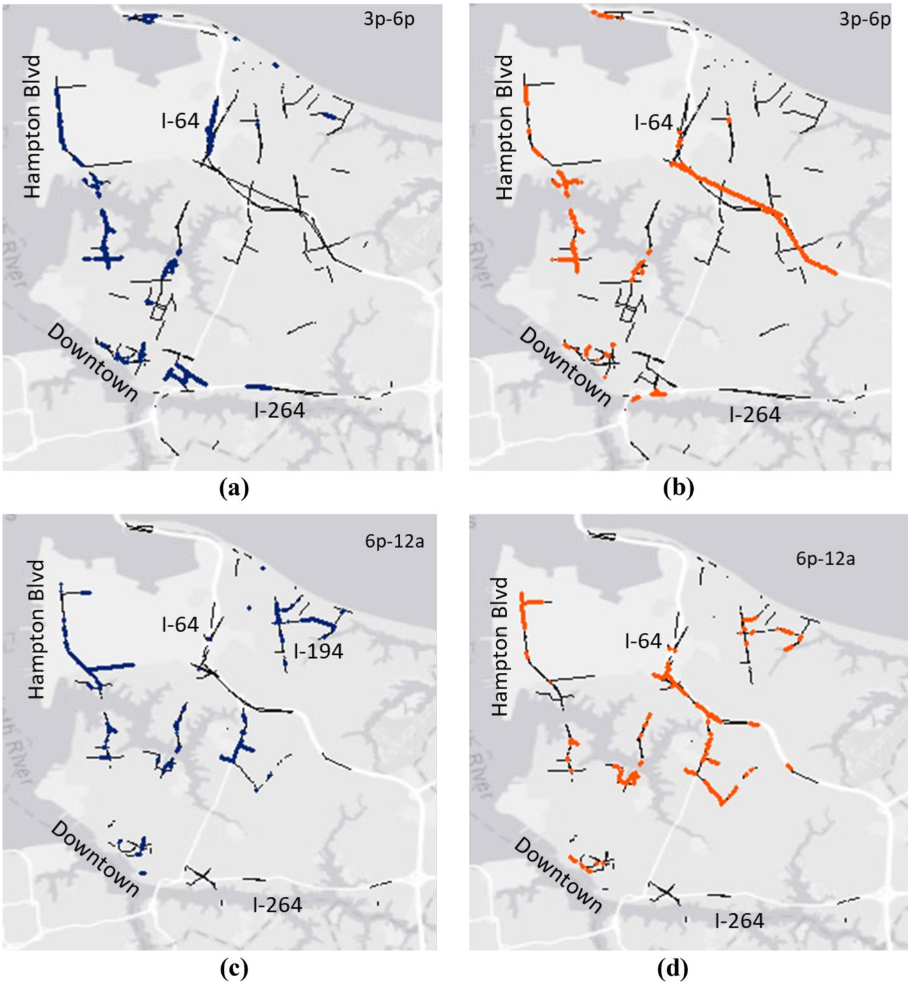



**Fig. 8** Highly affected links during flood periods

volumes, and some flood incidents causing a large decrease in volumes, compared to the NFPs. In the downtown area, the majority of impacted links experience increased traffic volumes. On the other hand, interstate corridors (I-64 and I-264) generally see decreased volumes during flood periods. A few of the highly impacted links (in orange or blue) appear in both maps, implying that these links (25% of highly impacted links during 3p-6p period and 14% of highly impacted links in the 6p-12a period) experience both increasing and decreasing traffic volumes during different flood events.

Impacts of flood incidents on the roadway network are also analyzed by roadway functional classification. The functional classification of affected links in the analysis includes interstates, minor freeways, principal arterials, major arterials, minor arterials, minor collectors, and local collectors. Figure 9 shows the distribution of links experiencing greatest increase and decrease in traffic volume during FPs by roadway functional classification. The most impacted links within the localized boundary analysis were found to be interstates, principal arterials, minor arterials, and minor collectors (other functional classifications had less than 5 links impacted in the top and bottom 10th percentiles).

As seen in Fig. 9, there are relatively few links impacted by flooding in the morning periods (12a-6a and 6a-9a), which is a result of lower traffic volumes and fewer flood incident reports for the overnight and early morning periods. In the remainder three time periods, interstate links within the localized boundary generally observe a decline in traffic volumes when a flood incident is reported. Principal arterials experience a decline in traffic volumes during the midday period (9a-3p) during FPs compared to NFPs, but see traffic volumes rise in the evening and night periods (3p-12a). This implies that during the evening peak and night time periods, when high numbers of flood incidents are reported, more road users opt to use the principal arterials (likely switching routes from interstates). It is possible that travelers perceive the access-controlled nature of interstate corridors as a disadvantage during flood periods, as their ability to switch routes dynamically is limited by access to ramps. Minor arterial links also appear to experience increased traffic volumes during FPs. Minor collector roads, though affected, do not show a consistent trend in direction of volume change by time period.

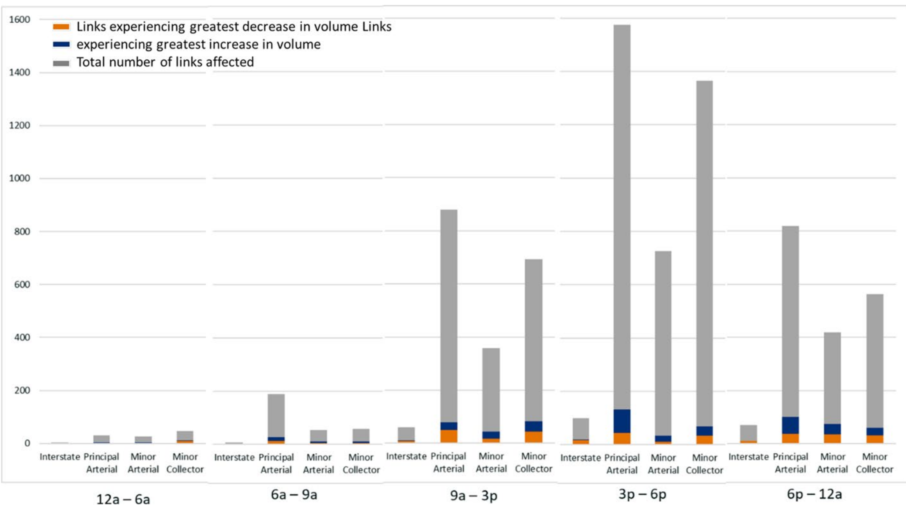



**Fig. 9** Distribution of most impacted links by roadway facility type

## 6 Conclusions and limitations

Prior studies examining recurring flooding and subsequent impacts on the transportation network have used projected or simulated data. This study is the first to use empirical data to assess such impacts, leveraging crowdsourced data to expand the spatial and temporal coverage of agency datasets in understanding the dynamic effects of recurring flood disruptions on roadway users. With recurring flooding becoming an increasing concern for coastal cities, this type of analysis demonstrates a framework for combining limited agency flood incident report data with crowdsourced flood reports, to understand the subsequent impacts on road users (in response to recurring flooding).

The study first estimates the citywide impact of recurring flooding on the Norfolk, Virginia roadway network. Due to lack of comprehensive traffic volume data, a framework to estimate traffic volumes using agency-provided and crowdsourced data was established, expanding the spatial coverage of volume data. These volume estimates show a 3% decline in VHT on FDs (as compared to NFDs) during the 20 month study period. With VHT being a cumulative measure of travel speeds and vehicle volumes, a simultaneous decrease in volumes and increase in travel times would not be sufficiently described by a single measure like VHT, especially when aggregated across the entire city which contains many links unaffected by flood events. Thus, the second part of the study examined the localized impacts of recurring flooding near the location of the crowdsourced flood incident reports. Results suggest that majority of links within impacted areas show a decline in speeds and volumes during flood periods. Volume estimates show a significant change in traffic volumes in the workday evening and night periods during FPs, with an aggregate VMT decline of 12% across all affected links. However, select links experience sizable increases in speeds and volumes. There are similar statistically significant reductions in VHT observed in the evening and night periods, suggesting that travelers are either avoiding travel, or changing their destinations. Particularly in the evening and night periods, localized analysis results point to reductions in travel during FPs, with decreased volumes on interstate corridors and increased volumes on principal and minor arterials, compared to NFPs (suggesting route shifts as a result of flooding).

Results of this study strongly suggest that the impact of recurring flooding events on transportation networks is local; thus, a citywide or regional analysis is not recommended due to the heterogeneous effects of flooding across various links. Analysis across a city or region may underestimate the impact of recurring flooding on travelers, as they abandon trips and shift routes in specific subareas. Spatial and temporal disparities in travel impacts are better explained through the localized impact assessment. Since recurring flooding is dynamic event, the framework provided in this study can serve as a precursor to identify recurring problem areas and periods for agency mitigation. In the case of Norfolk, since evening periods are more impacted than morning periods, mitigation efforts could be concentrated in key areas during those evening periods.

This study has certain limitations. First, like all studies that use crowdsourced data, the fidelity and accuracy of the data is an issue. Waze incident data is reported by Waze users, and there is no ground truth roadway flooding data to enable assertion of trustworthiness measures on the crowdsourced flood incident data. Regardless of trustworthiness, all Waze incident reports are included in the analysis. Second, the ground truth data used to train the data predictive model is obtained from VDOT CCS locations which are located mostly on principal arterials or freeways. Thus, the model's ability to predict traffic volumes on roads of lower functional classification is limited. In this study, only links contained in the regional travel demand model network are included (thereby excluding roads in the lowest functional classifications). Furthermore, there is some spatial mismatch between links in the Streetlight Data OSM layer and the HRRTDM roadway network, and these mismatches occur mostly on minor roads, further exacerbating the accuracy of volume predictions on these links. It is important to note that the results and conclusions in this study are focused on major roads, due to these limitations. Lastly, the Streetlight dataset analyzed here only considers light duty vehicle travel; thus, heavy vehicle movement is not captured. Diversion from freight schedules incurs significant economic costs, which has not been quantified with this study.

Nonetheless, the data predictive framework presented in this study can be generalized to be applied to other crowdsourced datasets, which are highly valuable when and where agency data is limited. This framework is applicable to a wide range of incident analyses such as congestion or accident analysis, post-disruption analysis, etc. Additionally, with the emergence of smarter cities and increasing availability of crowdsourced data capturing real-time traffic flow and hydrological variables, this framework can be a key component in strategic traffic rerouting and dynamic stormwater management to mitigate the impacts of recurring flooding, to ultimately increase resilience of coastal cities against such events.

**Author contributions** The authors confirm contribution to the paper as follows: study conception and design: T. D. C, S. P, M. B, J. G; data collection: S. P; analysis and interpretation of results: S. P, F. Z, T. D. C, J. G; draft manuscript preparation: S. P, T. D. C, F. Z, J. G. All authors reviewed the results and approved the final version of the manuscript.

**Funding** This project is funded by National Science Foundation's Critical Resilient Interdependent Infrastructure Systems and Planning (CRISP) Award: 1735587.

## Compliance with ethical standards

**Conflict of interest** The authors declare that there is no conflict of interest.

## References

Amin-Naseri M, Chakraborty P, Sharma A, Gilbert SB, Hong M (2018) Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze. Transp Res Rec J Transp Res Board 2672(43):34–43

Böcker L, Dijst M, Prillwitz J (2013) Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. Transp Rev 33(1):71–91

Chang H, Lafrenz M, Jung IW, Figliozzi M, Platman D, Pederson C (2010) Potential impacts of climate change on flood-induced travel disruptions: a case study of Portland, Oregon, USA. Ann As Am Geogr 100(4):938–952

Donovan B, Work DB (2017) Empirically quantifying city-scale transportation system resilience to extreme events. Transp Res Part C Emerg Technol 79:333–346

Ducruet C (2008) Hub dependence in constrained economies: the case of North Korea. Marit Policy Manag 35(4):377–394

Frequently Asked Questions about StreetLight Insight (2018) www.streetlightdata.com/support/faq/general. Accessed on Apr 4, 2018

Goodall N, Lee E (2019) Comparison of Waze crash and disabled vehicle records with video ground truth. Transp Res Interdiscip Perspect 1:100019

Jacobs JM, Cattaneo LR, Sweet W, Mansfield T (2018) Recent and future outlooks for recurring flooding impacts on roadways on the US East Coast. Transp Res Rec J Transp Res Board 2672(2):1–10

Lenkei Z (2018) Crowdsourced traffic information in traffic management: evaluation of traffic information from Waze. Dissertation, KTH School of Architecture and Built Environment. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-239178
Lu QC, Peng ZR (2011) Vulnerability analysis of transportation network under scenarios of sea level rise. Transp Res Rec J Transp Res Board 2263:174–181
Ouyang RW, Kaplan LM, Toniolo A, Srivastava M, Norman TJ (2016) Aggregating crowdsourced quantitative claims: additive and multiplicative models. IEEE Trans Knowl Data Eng 28(7):1621–1634
Rodrigue JP, Comtis C, Slack B (2017) The geography of transport systems, 4th edn. Routledge, Abingdon
Sadler JM, Goodall JL, Morsy MM, Spencer K (2018) Modeling urban coastal flooding severity from crowd-sourced flood reports using poisson regression and random forest. J Hydrol 59:43–55
Suarez P, Anderson W, Mahal V, Lakshmanan TR (2005) Impacts of flooding and climate change on urban transportation. Transp Res Part D Transp Environ 10(3):231–244
Sweet W, Park J, Marra J, Zervas C, Gill S (2016) Sea level rise and recurring flood frequency changes around the United States. Publication NOAA Technical Report NOS CO-OPS 073. National Oceanic and Atmospheric Administration, U.S. Department of Commerce
System to Track, Organize, Record, and Map (STORM). http://gisapp1.norfolk.gov/stor-mmap
US EPA (2016) Climate change indicators in the United States: coastal flooding, United States environmental protection agency. Fourth Edition. EPA 430-R-16-004. www.epa.gov/climate-indicators. Accessed May 10 2018
Waze (2017–2018) via the Waze for Cities data-sharing program, https://www.waze.com/wazeforcities. Accessed Aug 2017 to 2018
Zhu Y, Ozbay K, Xie K, Yang H (2016) Using big data to study resilience of taxi and subway trips for hurricanes Sandy and Irene. Transp Res Record J Transp Res Board 2599:70–80



## Affiliations

**Shraddha Praharaj[1] · T. Donna Chen[1] · Faria T. Zahura[1] · Madhur Behl[2] · Jonathan L. Goodall[1]**

Shraddha Praharaj
spraharaj@virginia.edu

Faria T. Zahura
fz7xb@virginia.edu

Madhur Behl
madhur.behl@virginia.edu

Jonathan L. Goodall
goodall@virginia.edu

[1] Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA

[2] Department of Computer Science, University of Virginia, Charlottesville, VA, USA