



ORIGINAL PAPER

Open Access



A machine learning approach for predicting hurricane evacuee destination location using smartphone location data

Prosper K. Anyidoho¹, Xinglong Ju², Rachel A. Davidson^{1*}  and Linda K. Nozick³

Abstract

Evacuation destination choice modeling is an integral aspect of evacuation planning. Outputs from such models are required to estimate the clearance times on which evacuation orders are based. The number of evacuees arriving at each destination also informs allocation of resources and shelter planning. Despite its importance, evacuee destination modeling has not received as much attention as identifying who evacuates and when. In this study, we present a new approach to identify evacuees and determine where they go and when using privacy-enhanced smartphone location data. We demonstrate the method using data from four recent U.S. hurricanes affecting multiple geographies (Florence 2018, Michael 2018, Dorian 2019, and Ida 2021). We then build on those results to develop a new machine learning model that predicts the number of evacuees that move between pairs of metropolitan statistical areas. The machine learning model incorporates hurricane characteristics, which have not been thoroughly exploited by existing methods. The model's predictive power is comprehensively evaluated through a tenfold cross validation, holdout validation using Hurricane Ida (2021), and comparison with the traditional gravity model. Results suggest that the new model substantially outperforms the traditional gravity model across all performance indicators. Analysis of feature importance in the machine learning model indicates that in addition to distance and population, hurricane characteristics are important in evacuee destination choices.

Keywords Hurricane evacuation, Destination choice, Machine learning, Smartphone data, Gravity model

1 Introduction

Developing effective and efficient hurricane evacuation strategies requires understanding and being able to predict population behavior during such events. Key aspects of this behavior include who evacuates, when, and where they go. A great deal of research has focused on the first two questions, in particular identifying

variables associated with the evacuation decision (e.g., Baker, 1991; Dash & Gladwin, 2007; Huang et al., 2016; Thompson et al., 2017), and developing quantitative models to predict it, typically using survey data (e.g., Anyidoho et al., 2022; Fu & Wilmot, 2004; Xu et al., 2016). By contrast, hurricane evacuee destination choice has received little attention (2.1), though evacuee destinations are also required by traffic models that predict evacuee movements and can greatly influence their results (Cheng et al., 2008). Such traffic models, in turn, are important for supporting implementation of contraflow, signal timing, and other network control measures during evacuation. Importantly, they are also used to compute *clearance time*, i.e., the time required for all those threatened by a hurricane to evacuate. In the United States today, official evacuation decisions are

*Correspondence:

Rachel A. Davidson
rdavidso@udel.edu

¹ Department of Civil and Environmental Engineering, University of Delaware, Newark, DE 19711, USA

² Price College of Business Management Information Systems, University of Oklahoma, Norman, OK 73019, USA

³ School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA



largely based on a comparison of this model-estimated clearance time and the estimated arrival time of hazardous conditions (NHP 2022).

Recognizing the destination choice gap in the hurricane evacuation population behavior literature, this paper contributes in two main ways. First, we introduce a method to use smartphone location data to identify evacuees from within the larger population and describe their movements, including geographic evacuation destinations. Results are compared to those based on survey data for three recent hurricanes that made landfall in the United States— Florence (2018), Michael (2018), Dorian (2019). Second, we use those results to develop and evaluate a machine learning model to predict evacuee destination choice. Specifically, we train a Zero-inflated Tweedie Extreme Gradient Boosting model (XGBoost) to predict evacuation flows between metropolitan statistical areas (MSAs). It is trained using data from the same three hurricanes, and its predictive power is comprehensively evaluated and compared with the traditional gravity model through cross validation and holdout validation for Hurricane Ida (2021).

Three novel features of these analyses are worth note: (1) use of smartphone location data, (2) inclusion of hurricane-related variables and data from multiple hurricanes, and (3) model formulation. First, although smartphone location data does not directly provide information about the device (i.e., smartphone) user or reasons for their actions, it offers samples orders of magnitude larger than those collected in the past using surveys (>1 million vs. typically 1000–2000 for surveys; see Table 5 in 5.) and a higher resolution description of the sequence and timing of movements, both helpful in better understanding the complexity of the evacuation phenomenon. Second, since hurricanes and the regions they affect can differ substantially, to improve the model's ability to predict destination choice in *future* hurricanes, we include attributes of hurricanes and their forecasts as predictors, and we combine data from three diverse hurricanes that together represent a broad range of conditions and circumstances. All previous models are based on data from a single hurricane (see Table 5 in 5.), and thus have a limited ability to capture hurricane-related drivers of evacuation destination choice and limited generalizability. Third, we use a Zero-inflated Tweedie Extreme Gradient Boosting model (XGBoost), which allows incorporation of the additional hurricane-related features, addresses the overdispersion created by having many origin–destination pairs with zero evacuees, and is capable of learning the nonlinear relationships that exist in this problem. By contrast, most available evacuee destination choice models are gravity models, which fail to capture the unique circumstances created by the

hurricane, instead focusing primarily on populations at origins and destinations and the distances between them as the driving factors of destination choice decisions, and which are parametric and thus limited in their ability to represent the overdispersion and nonlinearities in the data.

In 2, we summarize the literature with a particular focus on the objectives of this analysis. The data, evacuation behavior description method, and predictive destination choice models are described in 6, 7, and 5., respectively. In Sect. 6, we evaluate and compare the models' predictive power empirically. The paper concludes with discussion of the models' usefulness, limitations, implications, and recommendations for future work.

2 Background

2.1 Hurricane evacuee destination choice

Southworth (1991) notes that while evacuees might follow evacuation plans, some dispersion in destination choices will occur, influenced by locations of friends/relatives and hazard characteristics. Barret et al. (2000) refer to the evacuation destination as a location where evacuees tend to seek safety, or one that is recommended by an evacuation plan. Building on these earlier studies, Lindell and Prater (2007) characterized evacuation destinations, where evacuees stay until they return to their homes, by (1) geographic location (city/county/town), and (2) type of accommodation. Accommodation types are related to friends' or relatives' places, hotels/motels, and public shelters. Later studies conducted by Cheng et al. (2008) and Mesa-Arango et al. (2013) identified the home of friends/relatives as the most preferred destination. Both studies also identified factors affecting destination choices, including evacuation distance, whether the destination is affected by the hurricane, population composition of the destination, whether the destination is in a metropolitan statistical area, transportation convenience, the probability of finding a place to stay at the destination, hurricane position at evacuation time, household geographic location, race, income, preparation time, changes in evacuation plans, previous experiences with major hurricanes, household members working during the evacuation, and official evacuation notices.

In current practice, evacuees are assigned to destinations, or evacuation routes, subjectively (NCDOT 2000, Jha et al. 2004, Radwan et al., 2005). Mei (2002) and Modali (2005) argued that, except for the Oak Ridge Emergency Management software, no other trip distribution package existed before their studies. Wilmot et al. (2008), with specified Traffic Analysis Zones, calibrated gravity and intervention opportunity models using survey data from Hurricane Floyd. Separate models were

developed for each destination accommodation type (shelter and friends/family) with gravity models offering slightly better results. The gravity model considered only travel time in the impedance function, but the intervention opportunity model considered evacuation direction and hurricane path. The predictive power of the models calibrated for Hurricane Floyd was measured on data from Hurricane Andrew. Cheng et al. (2008) used the same Floyd survey to develop two multinomial logit models, one for friends and relatives and one for hotels/motels. They summed individual predicted probabilities to determine the incoming number of evacuees at each destination. These two models were evaluated on the training data, and thus do not provide a comprehensive assessment of predictive power. They also require strong aggregation assumptions to have a manageable number of alternatives in the choice set, and do not consider the destination type choice. Cheng et al. (2011), also using the same Floyd survey data, extended the gravity model by incorporating hurricane path, road conditions, and destination accommodation capacity. The impedance to travel was modeled as a function of distance between origin–destination pairs and the distance between destinations and the coastline using negative exponential function and Rayleigh functions, respectively. However, the transferability of this dynamic model has not been tested (Cheng et al., 2011).

The previously mentioned studies relied on survey data, which is limited in size, and only covers short-term travel patterns. A few more recent studies have leveraged the surge in digital mobility data, such as GPS traces, cell phone records, and geotagged tweets, to study human behavior during disasters. Jiang et al. (2021) used long-term human mobility data retrieved from Twitter during Hurricane Mathew (2016) to develop gravity models with social distance as an additional travel impedance factor. Out-of-sample predictive power of the model was measured through two rounds of leave-one-out cross-validation. The model does not include hurricane attributes, however, and was estimated using data from a single event. Hong et al. (2020) predict county-level evacuation flows using large-scale geotagged tweets gathered in Florida during Hurricane Irma; machine learning regression methods; and socioeconomic, hurricane, and evacuation official order features for both origin and destination counties. Hurricane attributes include average daily wind speed, peak fastest 2-min wind speed, fastest 5-s wind speed, and precipitation during Hurricane Irma's landfall, which can be measured pre-evacuation. They argued that hazards such as storm surge might not be available in the future for prediction until after the storm, however, the National Hurricane Center provides forecast information on hurricane attributes, thus we believe they can

be utilized. The Hong et al. (2020) model was evaluated through a hold-out validation in which the model was fitted with 75% of the data and validated with the remaining 25%. In this study, we develop a machine learning method that combines data from multiple diverse hurricanes using smartphone location data. We utilize important hurricane attributes and comprehensively evaluate predictive performance of our model.

2.2 Smartphone location data and evacuation

Here we focus on the general use of location data—from call detail records (CDRs), Twitter, and smartphones—in disaster-related applications (Yu et al., 2018) (rather than analysis of social media content (e.g., (Beigi et al., 2016; Vieweg et al., 2010)). CDRs have been used to examine population movements after the Haiti (2010), New Zealand (2011), and Nepal (2015) earthquakes (Bengtsson et al., 2011; Wilson et al., 2016). For every mobile phone text or call, the CDR contains the time and location of the nearest cell tower, so the frequency and accuracy of location data points can vary widely. Location data points from Twitter also vary and typically are not as frequent as those in smartphone location data. Nevertheless, Twitter data is relatively easily available. Wang and colleagues (Wang et al. 2014, Wang et al. 2016) used Twitter data from Hurricane Sandy and other events to analyze changes in distributions of trip distances during disasters. Studies conducted by Chae et al. (2014) and Chae et al. (2015), analyzed geotagged Twitter data from Hurricane Sandy, the Moore tornado, and the Boston bombing. They discovered, for example, that many individuals in Manhattan went to a supermarket right after an official evacuation order was issued for Sandy. Han et al. (2019) and Martín et al., 2017 used geotagged Twitter data to describe the evacuation in Hurricane Matthew (2016)—number of evacuees, their origins, and destinations. Jiang et al., (2021) used long-term human mobility data retrieved from Twitter to develop gravity models to predict evacuation flows in Hurricane Matthew (2016). Likewise, Hong et al. (2020), analyzed geotagged twitter data from Irma and developed machine learning models to predict evacuation flows. Less frequent are studies using smartphone location data like that here. Song and colleagues in three different studies (Song et al., 2013, 2014, 2017) used GPS records of 1 + million people over a year to analyze human mobility after the Tōhoku earthquake (2011). They identified important places for each individual (e.g., home, work), and demonstrated the change in time spent in each place type from before to after the event. They also developed a Hidden Markov Model to predict an individual's trajectory during a period of time based on her trajectory during the previous period; and predicted possible evacuation from the area using a

Markov Decision Process to predict the evacuation route. Yabe et al., (2019a, 2019b) used pre-disaster web search behavior to predict evacuation for a 2018 flood in Japan and employed smartphone location data to validate that model.

This literature highlights the promise of big location data in disaster studies and suggests some analysis methods. In this paper, we use smartphone location data, which is more precise and frequently sampled than the more commonly used CDR and Twitter data. Whereas much of the literature only describes observed behavior, the research herein centers on predicting destination choice decisions in future hurricanes. Finally, while most available studies analyze data from single events, the method presented here utilizes data across multiple hurricanes, thus multiple geographies and broader hurricane conditions are captured, increasing the generalizability of the resulting models.

3 Data

3.1 Hurricanes

This study uses data from four hurricanes—Florence (2018), Michael (2018), Dorian (2019), and Ida (2021). They cover multiple geographic areas and represent various storm intensities, timings, and characteristics (Table 1, Fig. 1). Hurricane Michael (2018) and Hurricane Ida (2021) formed in the Gulf of Mexico, whereas Hurricane Florence (2018) and Hurricane Dorian (2019) affected the Atlantic coast and were longer-lasting events (Fig. 1). Hurricane Dorian initially struck the Bahamas,

then followed the coast from Florida to North Carolina; Hurricane Florence came straight onshore at North Carolina. The impacts by Dorian were less severe while Florence, Michael and Ida caused substantial impacts (Table 1).

3.2 Smartphone location data

The data were obtained from the Spectus Data Cleanroom platform for 4 hurricanes—Florence, Michael, Dorian and Ida. The mobility data is de-identified and privacy enhanced for this research. Locations were passively collected through the mobile apps based on an opt-in framework. For each device, we have location data on average about 128 times per day or once every 11 min. iOS (iPhone) collects data when the device detects movement, so the sampling is more frequent in the daytime. Android collects data via a time-based sampling methodology. Each observation consists of a unique, anonymized, and encrypted identification number, time stamp, latitude, and longitude. To preserve the device owners' privacy, home areas are upleveled to the census block group level, sensitive points of interest are removed from the dataset, and device IDs are changed randomly every few months, limiting the time over which a particular phone's movement can be tracked.

Spectus mobility data is collected from smartphone applications where location is at the core of the app's functionality. This includes categories such as maps, navigation, weather, and geo-specific retail. All data is collected with the informed consent of fully anonymized

Table 1 Summary of data by hurricane

		Florence	Michael	Dorian	Ida
Hurricane	Landfall location	Southeastern NC	FL Panhandle	Cape Hatteras, NC	Port Fourchon, LA
	Maximum category	Cat. 4	Cat. 5	Cat. 5	Cat. 4
	Category at landfall	Cat. 1	Cat. 5	Cat. 1	Cat. 4
	Damage in U.S	\$24 billion	\$25 billion	\$1.6 billion	\$75 billion
Survey data	Num. survey observations	1578	486	588	n/a
	Num. evacuees in survey	735	205	105	n/a
Smart-phone description	Study area (sq km)	139,416	375,707	159,475	114,218
	Num. devices in initial data	154,684	447,504	392,057	162,694
	Num. evacuating tracts	424	799	586	293
	Num. devices in evac. tracts	29,057	65,456	49,936	44,580
	Num. evacuating devices	11,868	14,767	9,229	8,629
Smart-phone prediction	Num. evacuating devices scaled	452,893	305,194	264,438	87,903
	Num. evacuating devices in MSAs	446,962	298,860	259,591	85,914
	Num. origin MSAs	33	69	37	17
	Num. destination MSAs	743	775	737	803

Sources. Florence: Stewart and Berg (2019). Michael: Beven et al. (2019). Dorian: NCEI (2020) for damage and fatalities, Ida: Beven et al. (2021), Breslin (2018) for power outages, and NWS (2019) for all else

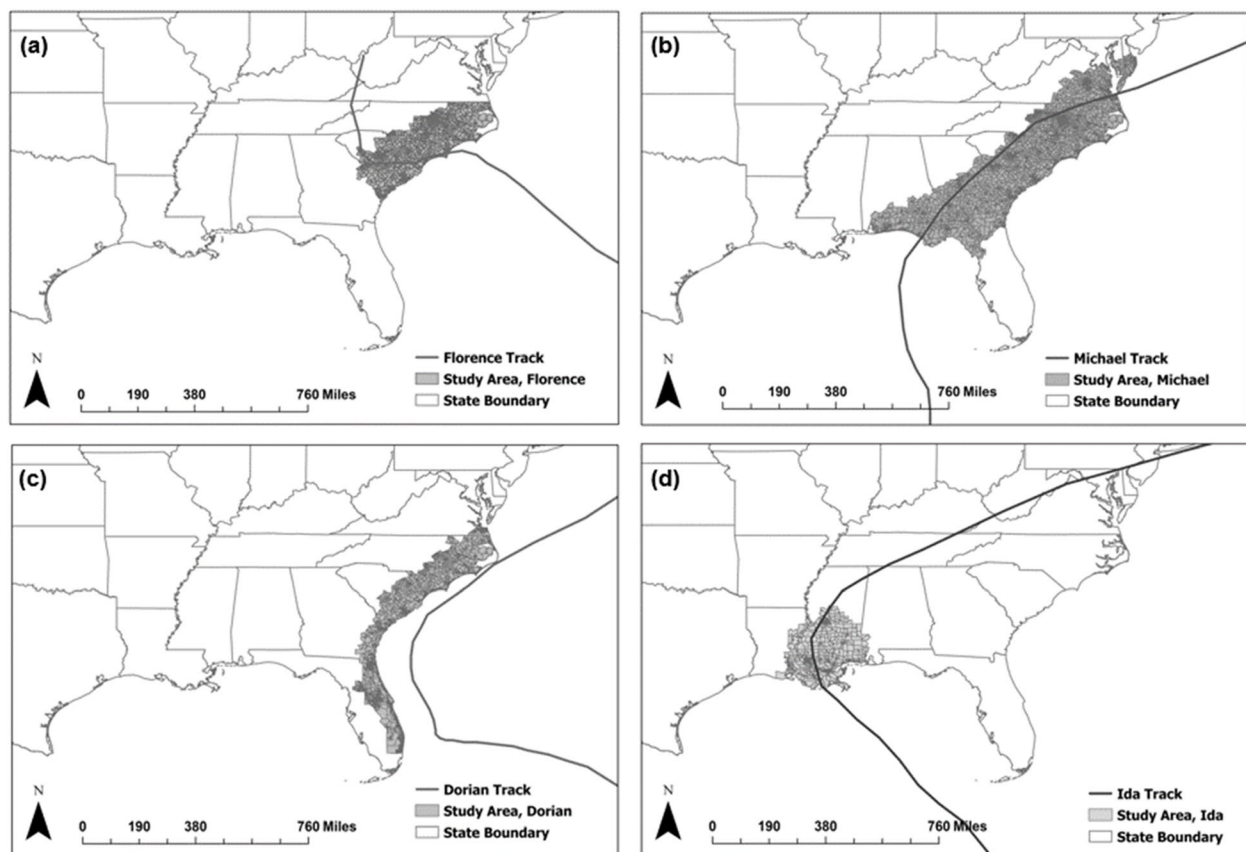


Fig. 1 Hurricane tracks and affected study area for **a** Florence, **b** Michael, **c** Dorian, **d** Ida

users under GDPR and CCPA compliant frameworks. Multiple studies have shown that Spectus data is representative of the general US population at various census area resolutions. For example, a study by Wang et al. (2019) found a 0.91 correlation between smartphone mobility data and census data at the census tract level in the Puget Sound region of Washington state. Another study by Aleta et al. (2020) reported correlation of 0.8 between Spectus data and census data at the census tract level in the Boston area.

3.3 Survey data

We use survey data from three hurricanes (Florence, Michael, and Dorian) (survey data unavailable for Ida) to validate results from the location-based data. The data were gathered through a 5- to 10-min web-based survey which was deployed after each hurricane using the method described in Anyidoho et al. (2022). The survey instrument, which was the same across hurricanes and was also used in Mongold et al. (2020) and Anyidoho et al. (2022), is presented in full in Anyidoho (2021). It includes questions related to (1) respondent location at the time of the event; (2) evacuation behavior, including

if the respondent evacuated, and if so, destination location and type, route, and timing; (3) reasons for evacuation decisions; and (4) individual attributes. Surveys were deployed through hurricane-related groups on social media platforms (e.g., Facebook and Reddit). The samples are reasonably representative of the populations from which they were drawn, across each hurricane, with a few exceptions. With similar distribution across hurricanes, the sample records fewer African Americans/Blacks (1% in sample vs. 17% in the population), has a higher income (\$81,600 sample average vs. \$64,900 population), is less likely to live in mobile homes (6% in sample vs. 13% in population), but equally likely to be employed (81% in sample and population), and have children (39% in sample vs. 40% in population).

4 Smartphone location evacuation description method

The new method to use smartphone location data to identify evacuees and describe their behavior includes eight steps organized in three main parts (Fig. 2): (A) Develop the dataset, (B) identify which smartphones in the dataset represent evacuees, and (C) describe the

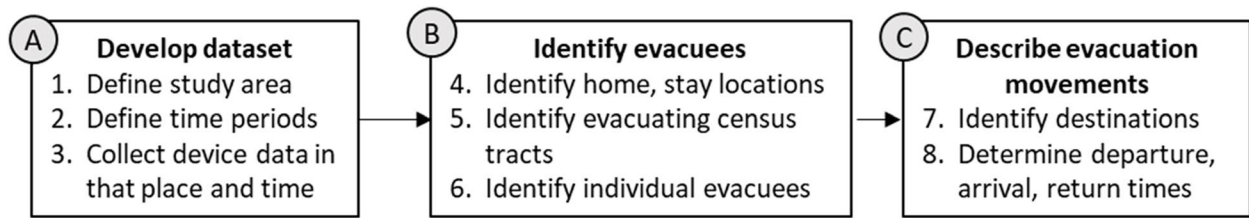


Fig. 2 Method to identify evacuees and describe their behavior using smartphone location data

evacuation movements. Each is described in turn in Sects. 4.1 to 4.3.

4.1 Develop the dataset

The technique adopted for generating the initial dataset aims to focus on areas and times where the hurricane threatened, thus minimizing noise in the data and unnecessary computational time. The approach is also systematic and easily reproducible for new events. In Step 1, for each hurricane, the geographic study area is defined to include all census tracts that were threatened by the storm and thus potentially could have included evacuees, whether or not the area ultimately did experience hazardous conditions. Specifically, to ensure consistency and reproducibility, for each hurricane, the study boundary

includes all census tracts that were forecast to experience wind speed of at least 74 mph. Data on wind hazard was obtained from the National Hurricane Center’s archive on each hurricane (NHC 2021).

In Step 2, we define the study timeline, including the *normal period*, *hurricane period*, and *evacuation period* (Fig. 3). The normal period, used to determine the location functioning as a person’s home at the time the hurricane approaches, is taken to be the three-week period before hurricane formation. While it may be preferable to include additional time windows during the year to establish normal movement patterns, that was not possible due to the limited duration of device IDs. Three weeks prior to hurricane formation was considered sufficient to avoid confusion about home location associated

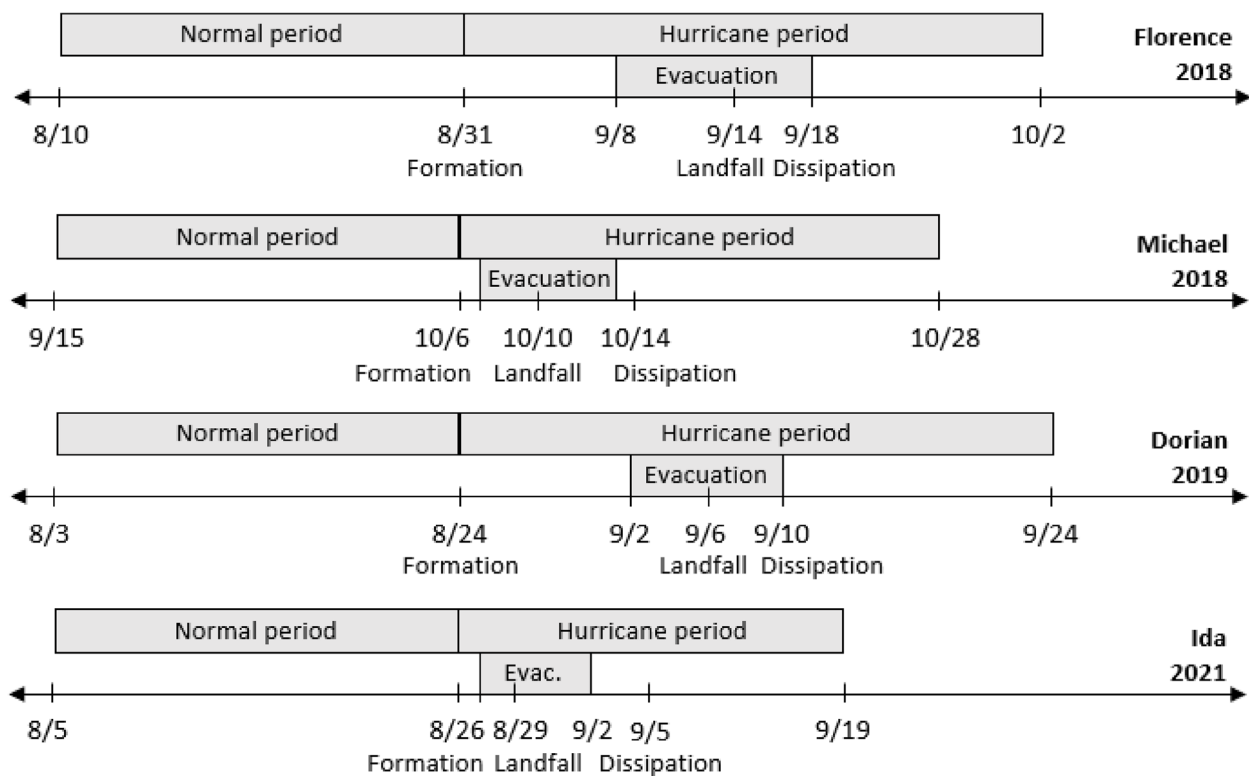


Fig. 3 Timeline of key events for each hurricane

with week-long vacations while not exceeding computational resource constraints. Hong et al. (2020) defined the normal period as two weeks prior to formation. The hurricane period, meant to capture the period of hurricane-related movements, is defined to last from hurricane formation until the time evacuees have returned home, taken to be from formation date to two weeks after dissipation. For each event, the evacuation period, the period during which most evacuees traveled to their destination, is defined as the time within the hurricane period during which travel rates are consistently higher than travel rates in the normal period, excluding holidays (e.g., Labor Day). Specifically, the earliest and latest peak travel date across all census tracts observed after formation mark the start and end of the evacuation period, respectively.

The initial data sample created for analysis is developed in Step 3 by identifying all devices thought to be in the study region during the normal period, and thus representing potential evacuees. For hurricanes making landfall prior to January 2019 (Florence and Michael), we selected devices that appeared in the study area at least 200 times during each week of the normal period. However, for Dorian (2019) and Ida (2021), we sampled devices that have the same home location at the block level for each week within the normal period—Spectus provides weekly home location of devices at the census block level for devices tracked after 2018. The initial dataset includes 154.7, 447.5, 392.0, and 162.7 thousand devices for Hurricanes Florence, Michael, Dorian, and Ida, respectively (Table 1).

4.2 Identify evacuees

Identifying which devices in the dataset represent evacuees and which do not, based only on their movements, is challenging because both evacuations and non-evacuation-related movements occur within the same time and space, and naturally vary over time and space. The approach herein relies on first identifying census tracts in which the aggregate behavior differs substantially from normal behavior, providing evidence of evacuation, and then identifying evacuees within those tracts.

In Step 4, for each device in the dataset, we identify the home location and the stay location for each day of the entire study duration. The home location is assumed to be the centroid of the 100 m by 100 m grid cell in which the individual spends the most time during the normal period. The associated home county, census tract, and state are also recorded. A device's stay location for a single day is defined as the 100 m by 100 m grid cell where the device spends the most time between 6 pm and 9am the following day.

If the stay location on a day is at least one mile from the device's home location, that device is labeled as away

from home that day. Based on that definition, in Step 5, for each day and each census tract, we compute the aggregate travel rate as the number of devices away from home divided by the total number of devices available on that day (i.e., the number for which we could compute the stay location). *Evacuating census tracts* then are those for which the peak rate observed during the hurricane period is at least 15% higher than the peak travel rate in the normal period. In selecting peak rates for the two-time windows, we exclude holidays which may have unusual movement patterns (e.g., Labor Day). To ensure reasonable travel rates, only census tracts with a minimum of 30 devices each day are considered. We assume only individuals with home locations in these evacuating census tracts can be considered evacuees. While people certainly evacuated from other census tracts, in those other tracts there is no evidence that the travel is not normal travel, whereas in the evacuating census tracts there is. Since the evacuee data is being used to develop a geographic destination model, this approach was designed to minimize the number of false positives (i.e., non-evacuees incorrectly labeled as evacuees).

In Step 6, we identify individual evacuees from among those with home locations in the evacuating census tracts. We classify a device as an evacuee if the stay location of the device is at least one mile from home on the *peak evacuation day*. For each census tract, the peak evacuation day is defined as the day in the evacuation time window with the maximum travel rate. Table 1 summarizes the number of evacuating census tracts, number of devices in those evacuating census tracts, and number of evacuating devices (i.e., evacuees) for each hurricane.

4.3 Describe evacuation movements

The destination of each evacuee is defined in Step 7 as the stay location (longitude and latitude) during the hurricane period that is farthest from their home. The destination is also characterized by the associated state, county, census tract, and census block group.

In Step 8, for each evacuee, we determine the home departure time t_d , destination arrival time t_a , and return time t_r , all as date and clock times. The departure time, which denotes the time an evacuee leaves home due to the hurricane, is taken to be the latest time the individual was in the 100 m by 100 m home grid cell prior to the peak evacuation date for the associated census tract.

5 Evacuee destination prediction models

5.1 Data development

For the predictive component of this analysis, we seek to predict the number of evacuees, F_{ij} , that travel from origin i to destination j as a function of feature vectors describing the origin, \vec{x}_i , destination \vec{x}_j , and the

Table 2 Summary statistics of model features

Variable	Florence		Michael		Dorian		All CV ^a		Ida	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Y	19.24	231.70	6.00	82.16	10.15	160.68	10.20	150.92	6.60	107.86
X_{dir}	1011.29	468.85	1025.33	466.52	1097.27	476.71	1040.68	470.95	1083.13	445.98
X_{op}	386,137	565,526	304,872	463,871	533,182	1,029,036	383,218	683,771	271,022	338,268
X_{dp}	322,696	1,092,797	318,219	1,079,968	326,051	1,099,991	321,304	1,088,210	300,151	1,021,771
X_{oh}	204.02	88.90	576.94	346.08	378.21	254.93	437.57	320.61	256.50	133.92
X_{dh}	1080.81	484.08	1110.72	455.78	1105.03	487.32	1102.208	471.02	1235.14	465.54
X_{ol}	530.06	92.06	1016.91	342.60	347.81	150.49	728.73	395.01	1018.04	108.99
X_{dl}	1403.58	509.78	1511.97	489.43	1087.62	464.07	1376.40	518.79	1770.94	524.58
X_{ow}	0.207	0.180	0.132	0.179	0.155	0.290	0.155	0.216	0.124	0.140
X_{dw}	0.010	0.059	0.013	0.066	0.008	0.075	0.011	0.067	0.003	0.028
X_{os}	1.779	2.085	0.343	1.138	1.981	1.646	1.106	1.726	2.424	2.840
X_{ds}	0.108	0.631	0.147	0.812	0.162	0.696	0.142	0.744	0.108	0.742
X_{dir}	248.03	101.48	229.69	119.68	249.02	111.62	239.02	113.93	149.15	134.29
X_{dir}	159.26	141.29	252.72	87.22	260.82	49.23	232.80	104.17	236.31	57.34
X_{lat}	36.22	2.41	35.50	2.65	35.14	2.84	35.58	2.68	34.23	2.45
X_{lon}	-82.52	3.66	-83.95	4.09	-82.77	3.67	-83.31	3.94	-89.42	4.21
X_{jit}	521,213	4,860,089	409,582	4,506,928	623,935	8,38,102	491,462	5,833,094	174,206	1,363,416
X_{jit}	1.61E-07	2.14E-06	9.83E-08	2.57E-06	8.93E-08	2.85E-06	1.11E-07	2.55E-06	2.81E-08	5.20E-07
X_{jit}	0.42	1.48	2.46	18.77	0.92	6.96	1.58	13.85	0.56	3.10
X_{jit}	2.03	9.37	6.24	40.67	0.76	4.14	3.83	29.44	4.37	22.95
X_{jit}	1.30E-05	2.75E-04	7.31E-07	3.25E-05	9.56E-06	1.64E-04	5.90E-06	1.59E-04	9.87E-06	1.77E-04

^a All CV represents data from Florence, Michael and Dorian

interaction between them, \bar{x}_{ij} . The origins and destinations are taken to be MSAs because we observed evacuee origin and destinations are predominantly within MSAs (94% and 93% for origin and destination, respectively, across all hurricanes) and for practical purposes MSAs are sufficiently specific. Modeling origin-destination (OD) pairs as MSAs also helps to reduce the proportion of zeros in the training data. We are most interested in predicting the proportion of evacuees at each destination since other models are available to predict the number of evacuees departing from each origin. Nevertheless, proportions are computationally difficult to model directly, thus in this model, we predict the number of evacuees, F_{ij} .

To derive the required data, we first created an OD matrix of evacuee flows between census tracts, where the origins are census tracts with evidence of evacuation (from Step 5, Sect. 4.2) and the destinations include all census tracts across the United States. Since those flows only reflect people for which we have smartphone location data, we then scale the flows up for each OD pair based on the population at the origin. Specifically, we divide the original number of evacuees by the maximum number of devices observed and multiply by the population at the origin (Table 1). Finally, we aggregated census tract flows to MSAs to obtain a new MSA-level OD matrix of evacuation flows. We define the final set of destinations to include all MSAs within the study area, as well as the approximately 98% MSAs outside the study boundary that are within 2000 km of their respective origins. The final dataset for Hurricanes Florence, Michael, Dorian, and Ida included 1,091,327 evacuees across 118,914 OD pairs (Table 1).

5.2 Features

Based on the literature (2.1), we hypothesize that hurricane evacuation flows are related to: (1) distance between origin and destination, which captures impedance between OD pairs, (2) populations at origins and destinations, which explain the level of production at origins and the attractiveness of destinations, (3) attributes of the hurricane, which could explain the relative hurricane threat at the origins and destinations, (4) features capturing spatial dependency, and (5) joint features which are inspired by the traditional gravity equation, to capture interactions where for each feature, a joint feature is computed by finding a product of the origin and destination values and dividing by square of the distance between the OD pair (Table 2). We also considered socio-demographic features such as the number of hotels/motels and unemployment rate at origins and destinations. However, these features were highly correlated with population

(Pearson correlation coefficient of about 0.98 across hurricanes), thus they were omitted.

Hurricane-specific attributes include, for each origin and each destination: (1) distance from hurricane location (x_{oh}, x_{dh}); (2) distance from projected landfall location along coastline (x_{ol}, x_{dl}); (3) probability of wind speed exceeding 74 mph (x_{ow}, x_{dw}); and (4) potential flood depth in the next 5 days (x_{oss}, x_{dss}) measured a day prior to the peak evacuation date for the corresponding origin. Data on hurricanes and hazards was collected from the National Hurricane Center (NHC 2021). With the availability of dynamic trip distribution models, such as those in Rambha et al. (2021) and Anyidoho et al. (2022), the peak evacuation date for each MSA can be computed in advance; hence, these features can be estimated in a predictive mode.

Spatial dependency was captured by the straight-line travel direction from origin to destination (x_{tdir}) in degrees clockwise from north; travel direction in degrees clockwise from the forward direction of the hurricane at day prior to peak evacuation day (x_{htdir}); and the average latitude and average longitude of the origin and destination (x_{lat}, x_{lon}). The average longitude and latitude values capture positions of each OD pair relative to each other.

5.3 Traditional gravity model

The gravity model is the most common trip distribution model in urban transportation planning. It incorporates the principle that trip patterns are explained by activities at the origin, the relative attractiveness of the destination, and the difficulty of making the trip between a pair of zones. Mathematically, the gravity model is described by the equation:

$$F_{ij} = K \frac{P_i^{\beta_1} P_j^{\beta_2}}{D_{ij}^{\beta_3}} \quad (1)$$

where, in the case of hurricane evacuation, F_{ij} is the flow (number of evacuees) between origin i and destination j ; P_i and P_j are the total populations of the origin i and destination j , respectively; and K is introduced as a scaling parameter analogous to the gravitational constant. In traditional gravity models, D_{ij} represents the resistance to travel between the origin and the destination and could be either travel time or distance. In hurricane evacuation applications, Modali (2005) and Cheng et al. (2008) utilized travel time as the impedance to travel; Jiang and colleagues (2021) used travel distance. In this analysis, D_{ij} is taken to be the shortest straight-line distance between origin and destination. The parameters β_1 , β_2 , and β_3 are estimated for the population at the origin, population at the destination, and travel distance, respectively. There

are numerous approaches to estimating these heuristic parameters. Modali (2005) and Cheng et al. (2008) transformed the problem and estimated the parameters using the popular least squares method. Jiang et al. (2021) used a non-linear optimization technique to estimate these parameters. As in most urban mobility gravity models, we fit the model as a Poisson regression using the *glm* package in R with distance and population as the only features.

Though widely used for other applications, the gravity model has a few key shortcomings for application to hurricane evacuation. First, the destination choices of individuals under a hurricane threat are substantially impacted by the characteristics of the hurricane and its forecast (Southworth, 1991), features not incorporated into traditional gravity models. For example, evacuees are probably less likely to travel to a close destination impacted by the hurricane than a farther destination that is safe. Second, in practice, in hurricane evacuations many possible destinations are not selected because they are too far or are threatened by the hurricane. This results in many OD pairs with zero evacuation flows, making the OD matrix sparse. The associated data are referred to as over-dispersed or zero-inflated with a right skewed distribution. Existing gravity methods, however, ignore this distribution and mostly model their solution using either the least squares approach or a Poisson regression method. Finally, hurricane evacuation is a complex phenomenon, and thus models should be able to capture non-linear relationships. The gravity model is not flexible in capturing such nonlinearities.

5.4 Zero-inflated tweedie XGBoost (XGBoost) model

We propose a machine learning method to predict the evacuation flow between a pair of MSAs. It builds on traditional gravity models and addresses their shortcomings for this application by including hurricane attributes as features; representing overdispersion, zero-inflated data; and capturing nonlinearities.

Specifically, we develop a Zero-Inflated Extreme Gradient Boosting Tweedie Regressor (XGBoost), a meta model of a classifier and a regressor, that predicts the number of evacuees, F_{ij} , that travel from origin i to destination j as a function of a feature set \vec{x} . The feature set \vec{x} is input into the classifier C , which determines whether flow should be zero. If the probability exceeds 0.5, the observation is forwarded to the regressor, R , which outputs a non-negative flow for the given pair. The feature set \vec{x} includes features describing the origin, \vec{x}_i , destination \vec{x}_j , and the joint features between them, \vec{x}_{ij} , i.e., all those listed in Table 2. The

interactive component of the feature space is inspired by the traditional gravity models, such that for a given feature at i and j the number of flows is not only dependent on the features describing the origin and destination, but also the interaction between the origin and the destination. As in Hong et al. (2020), for each feature except the distance between OD pairs, the interactive term is given by the product of the feature values at origin and destination divided by the square of the distance between them.

Both classifier and regressor are gradient tree-boosting algorithms and the regressor is modeled as a Tweedie compound Poisson distribution. The gradient boosting model, a nonparametric machine learning method, combines a series of weak learners (trees) and thus is able to capture complex interactions and produce high accuracies. The gradient boosting method also has hyperparameters which can be tuned to minimize overfitting and obtain desirable outputs. The extreme gradient boosting method, an extension of traditional gradient boosting, was chosen for both classifier and regressors because it converges quickly with fewer iterations, has an inbuilt capability to handle missing values, and enhances model training through parallelization and distributed computing. The objective function for the classifier uses the logistic loss while the regressor is modeled as a Tweedie loss. We chose the Tweedie loss for the regressor since we expect the distribution to be highly skewed to the right. The classifier helps handle overdispersion, the boosting method is capable of fitting a flexible model and capturing complex interactions among predictors, and the Tweedie distribution can capture heavily right-skewed data.

5.5 Hyperparameter tuning

The machine learning model training involves finding the optimal hyperparameters for the given dataset. Both classifier and regressor contain hyperparameters that can be tuned to obtain a suitable model based on the objective function. For both classifier and regressor we considered maximum tree depth, number of estimators, and learning rate as parameters to be tuned. Given that the data is heavily zero-inflated we introduced a custom hyperparameter, k , to control an under-sampling of the zero flows in the training data. Specifically, if there are n_t OD pairs with non-zero flows, we include all n_p and an additional $k*n_t$ randomly OD pairs with zero flows.

When fitting the model for this analysis, using the 105,263 OD pairs from Hurricanes Florence, Michael, and Dorian, for both classifier and regressor, we tuned the hyperparameters by trying all combinations as

follows: maximum tree depth (5, 6, 7, 8, 9, 10); number of estimators (100, 200, 300, 400, 500); learning rate (0 to 0.5 in increments of 0.1); and k (1 to 20 in increments of 1). For each combination of values, the ten-fold cross validation was conducted as described in Sect. 5.6, using the Common Part of Commuters (CPC) score and the coefficient of determination as scoring metrics. The set of parameters with the best model score was taken as the optimal set of hyperparameter values and the corresponding predictions were recorded. The best model reported optimal parameters of 10, 300, 0.2, and 8 for the classifier and 6, 100, 0.1, and 8 for the regressor, for the maximum tree depth, number of estimators, learning rate, and under-sampling parameters, respectively. Similarly, for the XGBoost model with traditional features (XGBoost-trad) the best model reported optimal parameters of 3, 200, 0.1, and 6 for the classifier and, 3, 100, 0.1, and 6 for the regressor, for the maximum tree depth, number of estimators, learning rate, and under-sampling parameter, respectively. There were no hyperparameters for the traditional gravity model.

5.6 Evaluation approach and metrics

To evaluate the out-of-sample predictive power of the evacuee destination models and compare them, we used both cross validation (CV) and holdout validation and computed multiple metrics to capture the various dimensions of interest. Specifically, we (1) trained the gravity and XGBoost models using 105,263 OD pairs across Florence, Michael, and Dorian and

implemented a ten-fold CV for that dataset, and (2) conducted holdout validation to test the models on Hurricane Ida data. CV estimates the expected prediction error over all training sets, rather than for a specific one. That is, it compares the methods rather than specific models with specific coefficient values. Holdout validation, on the other hand, compares specific models but can lead to high variance depending on the particular holdout data used. To minimize the effect of the fold sampling for both models, we repeated the CV ten times, each with a different set of randomly generated folds and averaged the resulting 100 estimates of each error metric.

We evaluate the predictive performance of models using seven metrics along two dimensions of interest (Table 3), the ability to predict the (1) approximate distribution/proportion of evacuees across destinations, and (2) number of evacuees for individual OD pairs. To measure the ability of models to capture evacuation demand distribution, our primary goal, we calculate the mean absolute percentage distribution error (PDE), coefficient of determination (R_j^2) and the root mean squared error ($RMSE_j$). The PDE, a newly proposed metric, compares the proportion of observed and predicted evacuees across destinations. The goodness-of-fit measures (R_d^2 and $RMSE_d$), compare the observed and predicted number of incoming evacuations at destinations.

To measure how well the model estimates the observed OD matrix, we compute Common Part of Commuters (CPC) and the Common Part of Commuters Distance

Table 3 Evaluation metrics

Dimension	Metric	Equation ^a
Incoming evacuees at destinations	Absolute Percentage distribution error	$PDE = \frac{\sum_j P_j - \hat{P}_j }{D}$
	Coefficient of determination for destinations	$R_j^2(F, \hat{F}) = 1 - \frac{\sum_j (\sum_i F_{ij} - \sum_i \hat{F}_{ij})^2}{\sum_j (\sum_i F_{ij} - E)^2}$
	Root Mean Squared Error for destinations	$RMSE_j(F, \hat{F}) = \sqrt{\frac{\sum_j (\sum_i F_{ij} - \sum_i \hat{F}_{ij})^2}{D}}$
Full OD matrix	Common Part of Commuters	$CPC(F, \hat{F}) = \frac{2 \sum_{ij} \min(F_{ij}, \hat{F}_{ij})}{\sum_{ij} F_{ij} + \sum_{ij} \hat{F}_{ij}}$
	Common Part of Commuters Distance Variant	$CPC_d(F, \hat{F}) = \frac{2 \sum_b \min(T_b, \hat{T}_b)}{\sum_b T_b + \sum_b \hat{T}_b}$
	Coefficient of determination	$R^2(F, \hat{F}) = 1 - \frac{\sum_{ij} (F_{ij} - \hat{F}_{ij})^2}{\sum_{ij} (F_{ij} - E)^2}$
	Root Mean Squared Error	$RMSE(F, \hat{F}) = \sqrt{\frac{\sum_{ij} (F_{ij} - \hat{F}_{ij})^2}{N}}$

^a P_j and \hat{P}_j = observed and predicted proportion of evacuees at destination j (0% to 100%); D = total number of destinations; F_{ij} and \hat{F}_{ij} = observed and estimated number of evacuees from origin i to destination j ; E = average number of evacuees across OD pairs; T_b = number of evacuees that traveled within distance b (km); N = total number of OD pairs

Variant (CPC_d), metrics that have mostly been used in urban planning models (Lenormand et al., 2012, Lenormand et al., 2016, Robinson et al. 2017). CPC compares numbers of evacuees between the predicted and observed matrices. It is zero when the two matrices have no entries in common, and one when they are identical. CPC_d measures how well a predicted evacuation matrix recreates trips at the same distances as the observed matrix. In the domain of hurricane evacuations, Modali (2005) and Cheng et al. (2008) captured the idea of this metric by visually comparing the trip length frequency curves for the estimated and the observed data. We also present similar curves for our models in Sect. 6.1. In addition to these metrics, we also include the coefficient of determination (R^2) and the root mean squared error ($RMSE$).

6 Results

6.1 Evacuation behavior description results

After applying the smartphone location evacuation description method (7) to Hurricanes Florence, Michael, Dorian, and Ida, we compare the results to those from the survey data for the first three (6.2.1). Since the smartphone data draws from a larger study area that includes regions far inland (Fig. 1), for a more

direct comparison, we also present *reduced smartphone* results that include data only from the geographic areas that coincide with the survey data area. While both smartphone and survey data have limitations, they are independent approaches to describing evacuation behavior and therefore, agreement between them offers additional confidence in both results.

The smartphone, reduced smartphone, and survey data, respectively, indicate that 41%, 51%, and 57% of people evacuated in Hurricane Florence; 23%, 38%, and 49% in Hurricane Michael, and 20%, 20%, and 16% in Hurricane Dorian. The reduced smartphone and survey data differ by 4 to 11 percentage points, which seems quite reasonable. All three datasets similarly suggest that evacuation rates were highest in Florence and lowest in Dorian. The lower evacuation rates for the full smartphone data make sense given that it includes more inland areas. Potential reasons for the differences in the evacuation rates (and departure curves and distances traveled) estimated by the two datasets include possible sampling bias in the survey data, which was based on a nonrandom sample (Anyidoho et al., 2022), and the smartphone method's prioritization of minimizing false positives (Sect. 4.2), which may result in undercounting evacuees, thus reducing evacuation rates.

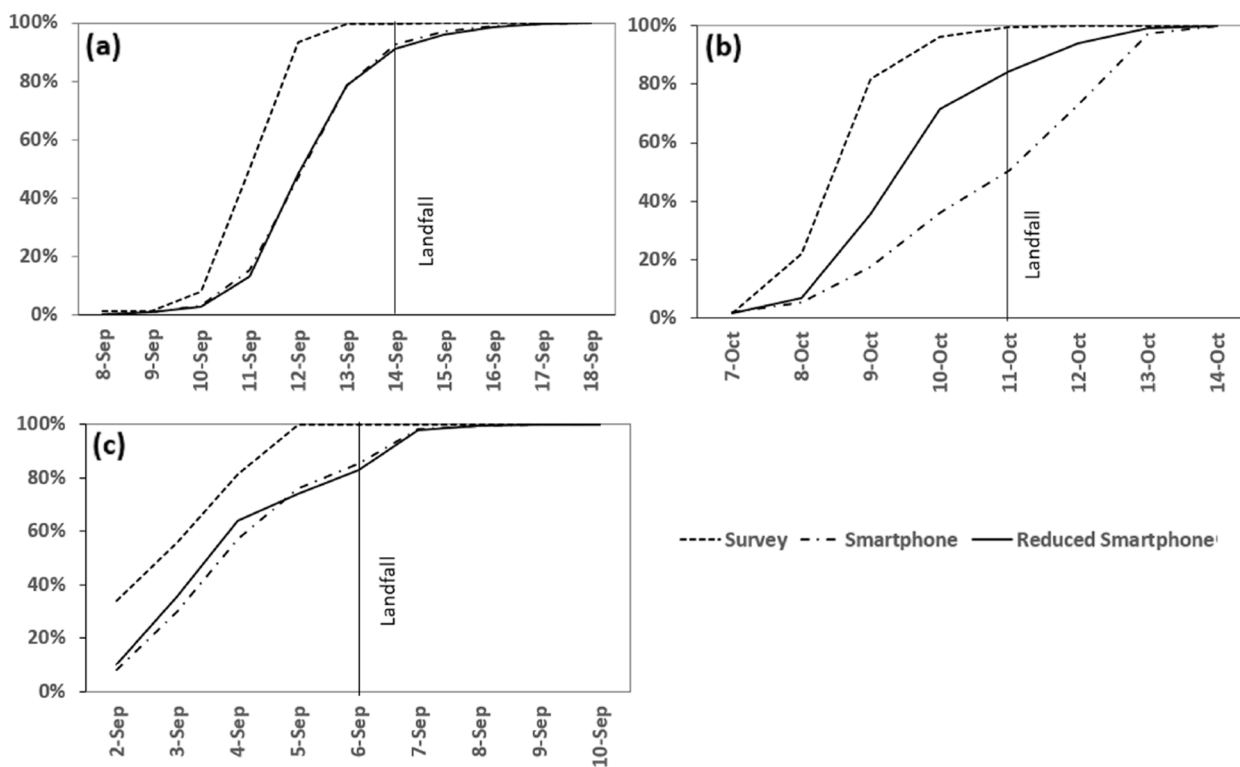


Fig. 4 Comparison of departure curves for survey and smartphone data for a Florence, b Michael, and c Dorian

To further understand evacuation patterns, we compare departure curves across the data types (Fig. 4)—survey, smartphone, and reduced smartphone. Across all hurricanes, evacuations happened more rapidly in the survey data than observed from the smartphone data variations. With landfall as a reference, the survey data suggests at least 98% of departures happened prior to landfall across all hurricanes. The smartphone data reports at least 80% for Florence and Dorian (Fig. 4a and c) and 30% for Michael (Fig. 4b) and the reduced smartphone data indicates similar values, except for Michael with at least 65% evacuating by landfall. The smartphone data covers a wider geographic region including inland areas from which evacuations occur after landfall from flooding. Among the three hurricanes, the study area for

Michael includes the most inland areas (Fig. 1), thus it experiences a significant increase in proportion of evacuees by landfall when these inland areas are removed.

We also compare the distribution of distance traveled by evacuees as estimated using the smartphone and survey data (Fig. 5). It can be observed that, despite the nuances in evacuation rates, across all and individual hurricanes, the travel distance patterns are comparable between the survey and smartphone data. While the range of travel distance for the smartphone data extends beyond that of the survey data, most travels occurred within 2000 km (98%). This aligns with the first assumption laid out by Southworth (1991) that evacuees will travel to the shortest possible destination which is outside the threat zone.

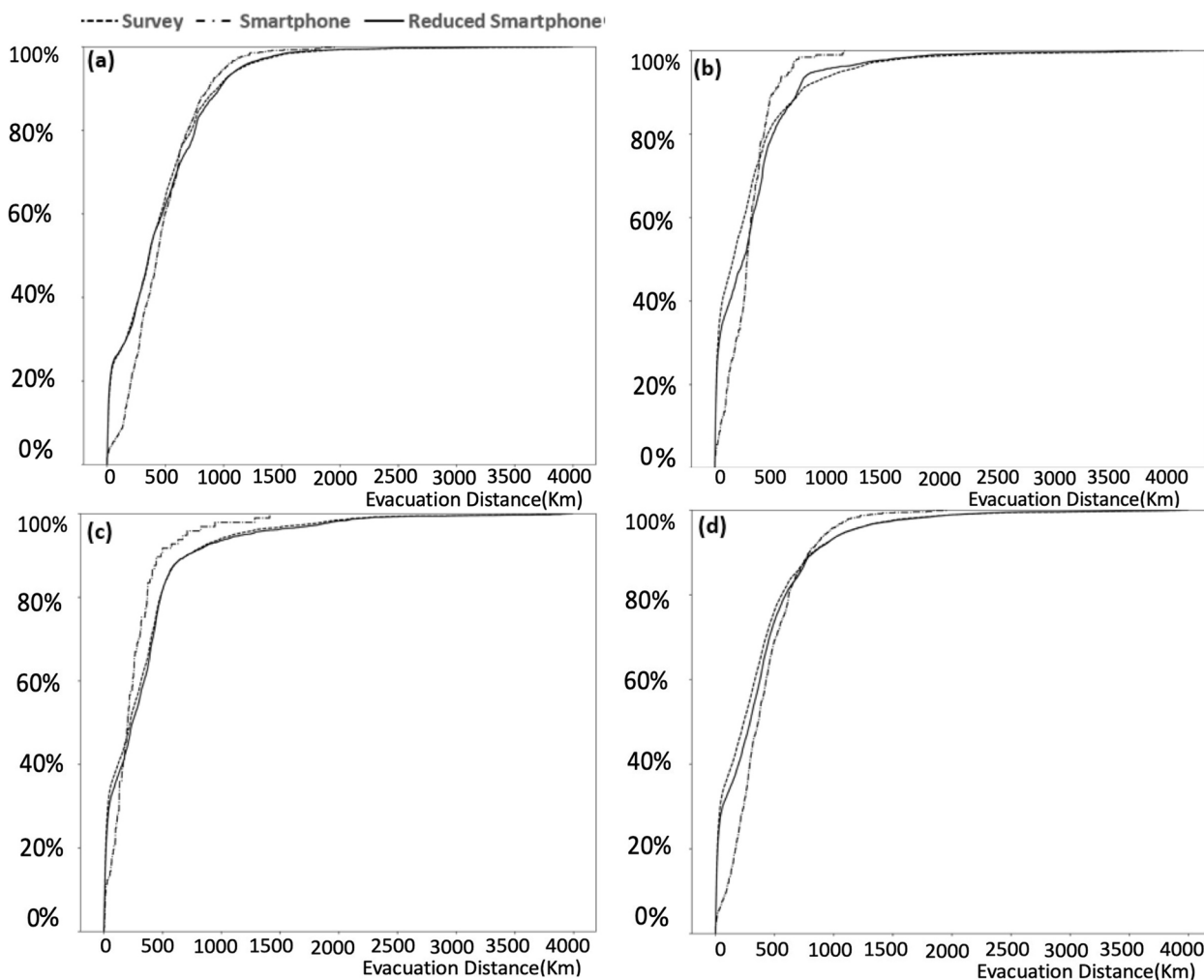


Fig. 5 Comparison of cumulative evacuation distance (limited to trips within 4000 km) for survey and smartphone data for **a** Florence, **b** Michael, and **c** Dorian, **d** All

6.2 Evacuee destination model prediction results

This section compares prediction results for three models—XGBoost, XGBoost-trad, and Gravity—using the evaluation approach and metrics defined in Sect. 5.6. The XGBoost is trained on all features whereas, the XGBoost-trad model is trained on the traditional gravity features (distance and population). For each evaluation metric and each model, metrics are presented for the complete dataset (All) and considering only observations from one hurricane (Florence, Michael, and Dorian). We also present results of model validation for Hurricane Ida (Table 4).

6.2.1 Incoming evacuees at destinations

While the model directly outputs the number of evacuees traveling between individual OD pairs, practical application of the model requires knowing the correct estimate of the proportion/number of evacuees at each destination as this influences planning for shelters and resource allocation. We compare the ability of the models to distribute evacuees at the aggregate level. Results indicate that the machine learning methods (XGBoost and XGBoost-trad) outperform the traditional gravity (Gravity) across metrics for both cross validation results and the holdout results on Ida (Table 4). In terms of the *PDE*, the average percentage of evacuees going to destination j is under- or over-estimated with errors of 0.053, 0.078, and 0.102 percentage

points (see All). Similarly, while errors are larger for the holdout set in Ida, XGBoost and XGBoost-trad report 0.087 and 0.097 percentage points, respectively, versus 0.122 percentage points for the gravity model. The XGBoost outperforming XGBoost-trad confirms the significance of hurricane attributes in destination choice decisions. The spatial distribution of errors shows that all three models generally overpredict for destinations farther from the threat zone and underpredict for destinations closer to the threat zone (Figure 9 in 5.). This may be due to the fact that residents will likely evacuate to the closest but safest destination, thus the model expects evacuees to evacuate to farther destinations not within the threat zone. Also, most of the farther destinations are within larger cities such as Atlanta, which will be more attractive to evacuees. However, the errors are smallest for XGBoost followed by XGBoost-trad and Gravity model. While the PDE uses the proportion of incoming evacuees, the R^2_j and $RMSE_j$ are computed based on the actual number of incoming predicted and observed evacuations (see Figure 8 in 5.).

6.2.2 Individual OD predictions

It is also important to measure how well our models replicate observed OD matrices (Individual OD links), as this is crucial in traffic assignment, implementing contraflow measures and resource allocation. As expected, the metrics suggest that it is

Table 4 Prediction results for models across evaluation metrics

Hurricane	Model	Incoming evacuees at destinations			Individual OD predictions ^a			
		PDE %	R^2_j	RMSE _j	CPC	CPC _d	R^2	RMSE
All	XGBoost	0.053	0.843	1.567	0.544	0.964	0.546	101.635
	XGBoost-trad	0.078	0.598	2.095	0.377	0.952	0.223	132.977
	Gravity	0.102	0.395	2.706	0.272	0.810	0.052	146.923
Florence	XGBoost	0.047	0.852	1.763	0.586	0.931	0.606	145.438
	XGBoost-trad	0.071	0.629	2.456	0.372	0.718	0.253	200.262
	Gravity	0.095	0.447	2.427	0.197	0.680	0.048	226.061
Michael	XGBoost	0.049	0.893	1.506	0.508	0.949	0.485	58.976
	XGBoost-trad	0.067	0.361	1.907	0.360	0.817	0.007	81.850
	Gravity	0.088	0.305	2.899	0.334	0.758	0.093	78.225
Dorian	XGBoost	0.062	0.789	1.409	0.517	0.913	0.464	117.667
	XGBoost-trad	0.097	0.670	1.879	0.404	0.900	0.275	136.787
	Gravity	0.124	0.328	2.762	0.285	0.724	0.035	157.810
Ida	XGBoost	0.087	0.479	1.255	0.334	0.620	0.107	101.952
	XGBoost-trad	0.097	0.506	1.362	0.287	0.665	0.100	102.306
	Gravity	0.122	0.442	2.718	0.275	0.547	0.142	99.935

^a Metrics are defined in Table 3. For *RMSE*, *RMSE_j*, and *PDE*, lower values indicate better performance; for all other metrics, higher values indicate better performance

difficult to accurately predict the exact number of evacuees on each OD pair, however, across all hurricanes, the machine learning models (XGBoost, XGBoost-trad) outperform the traditional gravity model for the four metrics— CPC , CPC_p , $RMSE$, and the R^2 . For all metrics, except for $RMSE$, a value of one suggests the predicted OD matrix perfectly matches the observed OD matrix and zero otherwise. Entries in the predicted OD matrix are identical to 54%, 38% and 27% of entries in the observed OD matrix, for the XGBoost, XGBoost-trad, and Gravity, respectively (see CPC for All).

To our knowledge, the CPC metric has not been used in any of the trip distribution models from the hurricane evacuation literature. Existing studies typically use R^2 and $RMSE$. However, studies by Modali (2005) and Cheng et al. (2008) compared the trip length frequency distribution for observed

and predicted trips, which is similar to the basis on which the CPC_j is computed. We produced similar figures comparing the predicted cumulative trip frequencies to the observed. Most evacuees across individual hurricanes did not travel beyond 300 km (Fig. 6) across all hurricanes. This is reasonable and aligns with the first assumption by Southworth (1991) which states that evacuees will choose destinations that are closer. Generally, across all events, the curve for the XGBoost model closely depicts the observed, followed by the XGBoost-trad as compared to the traditional gravity model. The three models perform well in capturing short distance travels (within 300 km), however, the gravity model and the machine learning model with traditional features (XGBoost-trad) are incapable of capturing long distance travels, even though the XGBoost-trad does a better job. This supports the importance of including

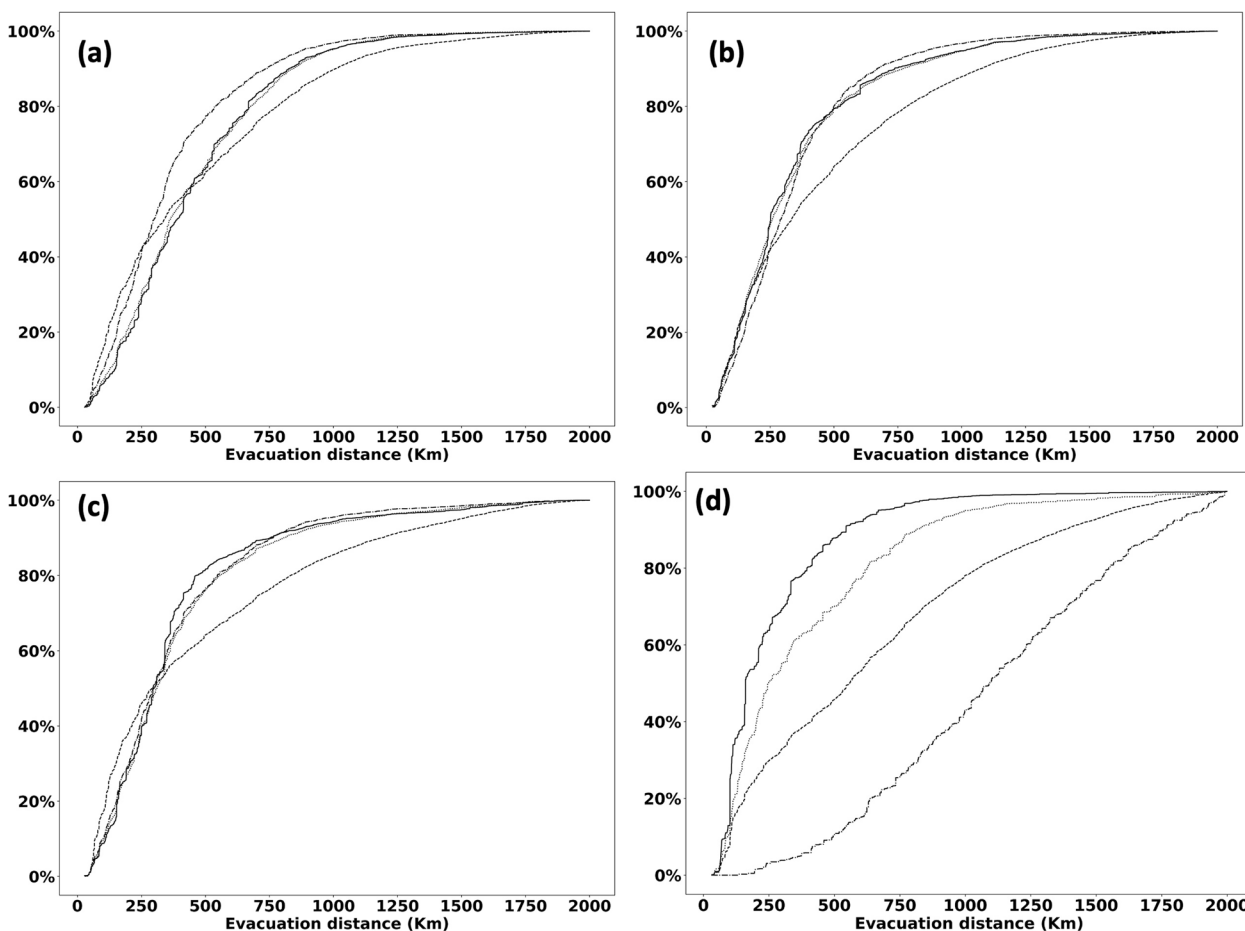


Fig. 6 Comparison of cumulative evacuation distance distribution for observed data and models on **a** Florence **b** Michael, **c** Dorian, **d** Ida

additional features related to the hurricane, as well as utilizing a model that learns non-linear relationships. The Ida validation (Fig. 6d) suggests the models are not as good at predicting travel distance for a hurricane that was not within the training set. They all overestimated the distances evacuees traveled in Hurricane Ida, although XGBoost was much closer than the other models. This suggests the need to include multiple diverse hurricanes when training a model to enhance generalizability.

6.2.3 Feature importance

That XGBoost outperforms XGBoost-trad and both outperform the traditional gravity method suggests that more features than those which are used by the traditional models, are needed to accurately predict trip distributions in hurricane evacuations. To better understand how each feature contributes to the prediction, we rank the features in the XGBoost model

in order of importance based on their SHAP (Shapley Additive exPlanations) values. Lundberg and Lee (2017) introduced the concept of SHAP values for uncovering the most informative features of tree-based machine learning models.

A SHAP value, computed for each observation-feature combination, indicates how much the feature contributes to the evacuation flow prediction. In Fig. 7, the horizontal axis for each feature includes a point for each observation. A red point indicates the observation has a high value for that feature; a blue point indicates it has a low value for that feature. Positive SHAP values (to the right) indicate the feature contributed to *increasing* the predicted evacuation flow; negative SHAP values (to the left) indicate the feature contributed to *decreasing* the predicted evacuation flow. Interpreting the color and SHAP value (i.e., horizontal position) together, Fig. 7 shows for example, that for the joint O-D population feature, x_{pjt} red

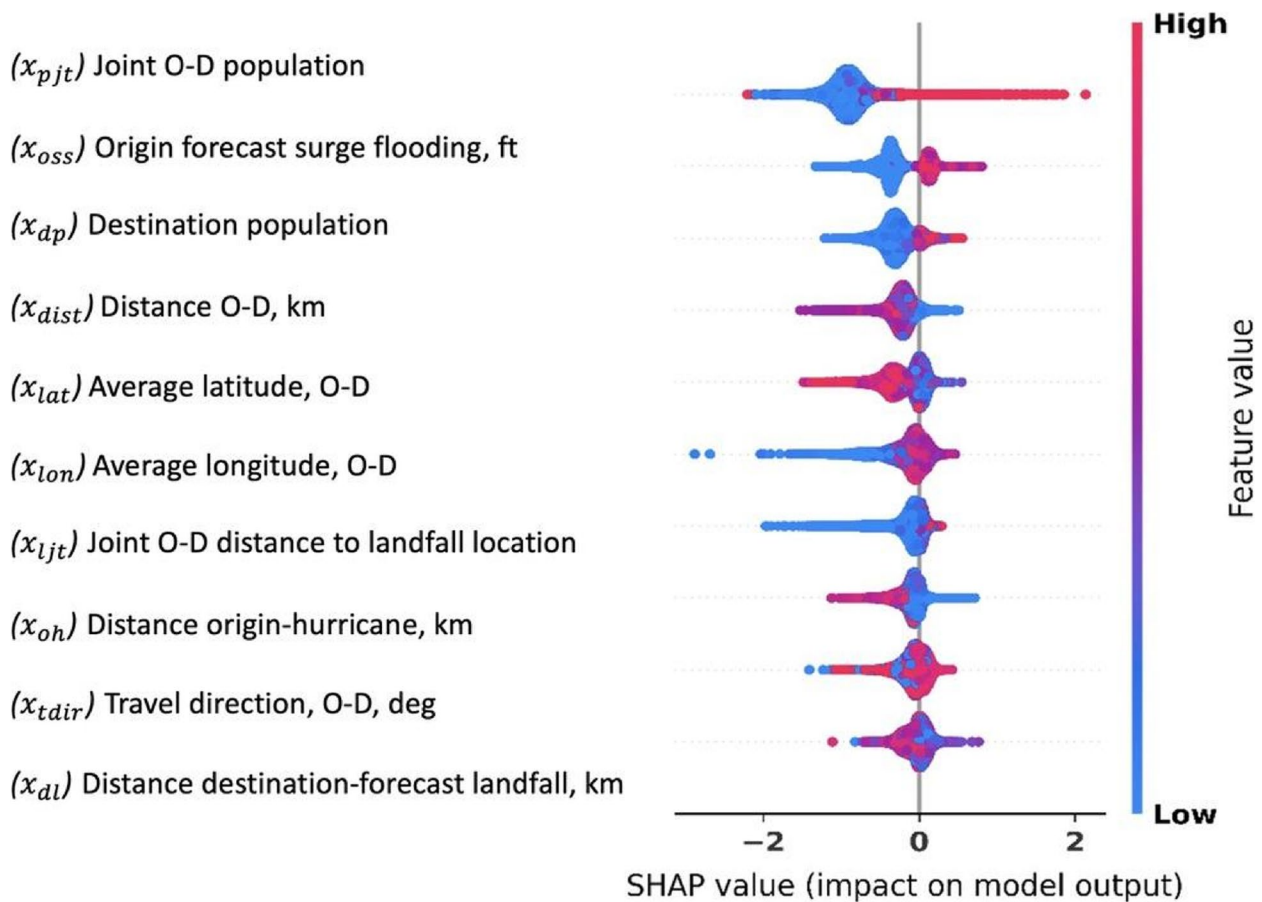


Fig. 7 SHAP feature importance ranking of top 10 features for XGBoost model

points are mostly to the right and blue points to the left. This suggests that high values of joint population (red) have a large positive contribution and low values of joint population (blue) have a large negative contribution. In other words, the joint O-D population has a strong direct relationship with evacuation flow. By contrast, for the distance origin-hurricane feature, x_{oh} , red points are more to the left and blue to the right, suggesting that large distances contribute to reducing evacuation flow predictions, and small distances contribute to increasing flow predictions. That is, the distance from the origin to the hurricane is inversely related to the evacuation flow. While SHAP values are computed for each feature and observation, Fig. 7 provides a global ranking across all observations, with the first being the most important.

We observe that the joint population feature, x_{pjt} , is the most important predictor. On average, the smaller the value of the joint population feature, the smaller the predicted evacuation flow. The SHAP results confirm the importance of population and the distance between OD as represented by the traditional gravity model. Except for population at the origin and destination, travel direction, relative location OD pairs represented by longitude and latitude, the remaining important features are dominated by hurricane-related attributes (origin forecast surge flooding, x_{oss} , joint distance to landfall location, x_{ljt} , distance from origin to hurricane, x_{oh} , and distance from destination to forecast landfall location, x_{dl}). This again indicates that while population and distance are important, the choice of evacuee destinations is also greatly influenced by the characteristics of the event. It is also worth noting that, the origin population does not appear in the top 10 features, which is an indication that the choice of destination is more influenced by the population at the destination. The origin population is more important in estimating evacuation demand.

Another important feature is the relative geographic proximity of the origin and destination, represented by the average of the longitude values at origin and destination. A notable weakness in the traditional gravity model lies in its inability to also capture spatial dependency across OD pairs. For instance, it is expected that two zones that are close will have comparable evacuation rates, and thus will have similar evacuation flows to a particular destination. With the main objective of applying the model to future hurricanes, it is important to integrate spatial awareness and dependency in

the models. This is crucial as a model trained on hurricanes that affected a particular geography might perform poorly when applied to a hurricane that impacted a different geography. This is evident from results from the holdout validation on Hurricane Ida (Table 4).

7 Conclusions and future work

In this study we contribute to the literature on hurricane evacuation population behavior by utilizing smartphone data to understand and predict destination choice decisions at the aggregate level. We offer three main contributions. First, we develop a method to identify evacuees and their final destinations and characterize travel trajectories using smartphone location data. Second, we propose a new machine learning model to predict evacuation flows. Finally, we comprehensively assess the generalizability of the developed method and compare it with the traditional gravity model across multiple evaluation metrics.

The new method to identify evacuees is fast, efficient, and semi-automated, thus it can easily be applied to new events. The machine learning model outperforms the traditional gravity model across all performance indicators. This suggests the need to capture more attributes related to the hurricane itself and to have models that are able to capture the complex interactions in population destination choice decisions.

The approach presented here does have some limitations. First, because the method presented to identify evacuees seeks to minimize false evacuations, there is a possibility of leaving out people that actually evacuated. The new XGBoost model also does not capture the dynamic aspect of evacuations, i.e., when evacuations take place and if destination choices vary with time. Future work should focus on developing algorithms that infer evacuees based on multiple characteristics at the disaggregate level, thus minimizing both false positives and false negatives. Models should also be extended to predict time-dependent destination choice decisions capturing the dynamic hurricane attributes. We also recommend retraining the model with multiple and diverse hurricanes to improve the predictive power of existing models. Finally, since this is the first destination choice model for hurricane evacuations that leveraged smartphone location data, future studies can utilize transfer learning techniques by using model weights from the current models to train new models whenever new hurricane data becomes available.

Appendix

Figures 8, 9, Table 5

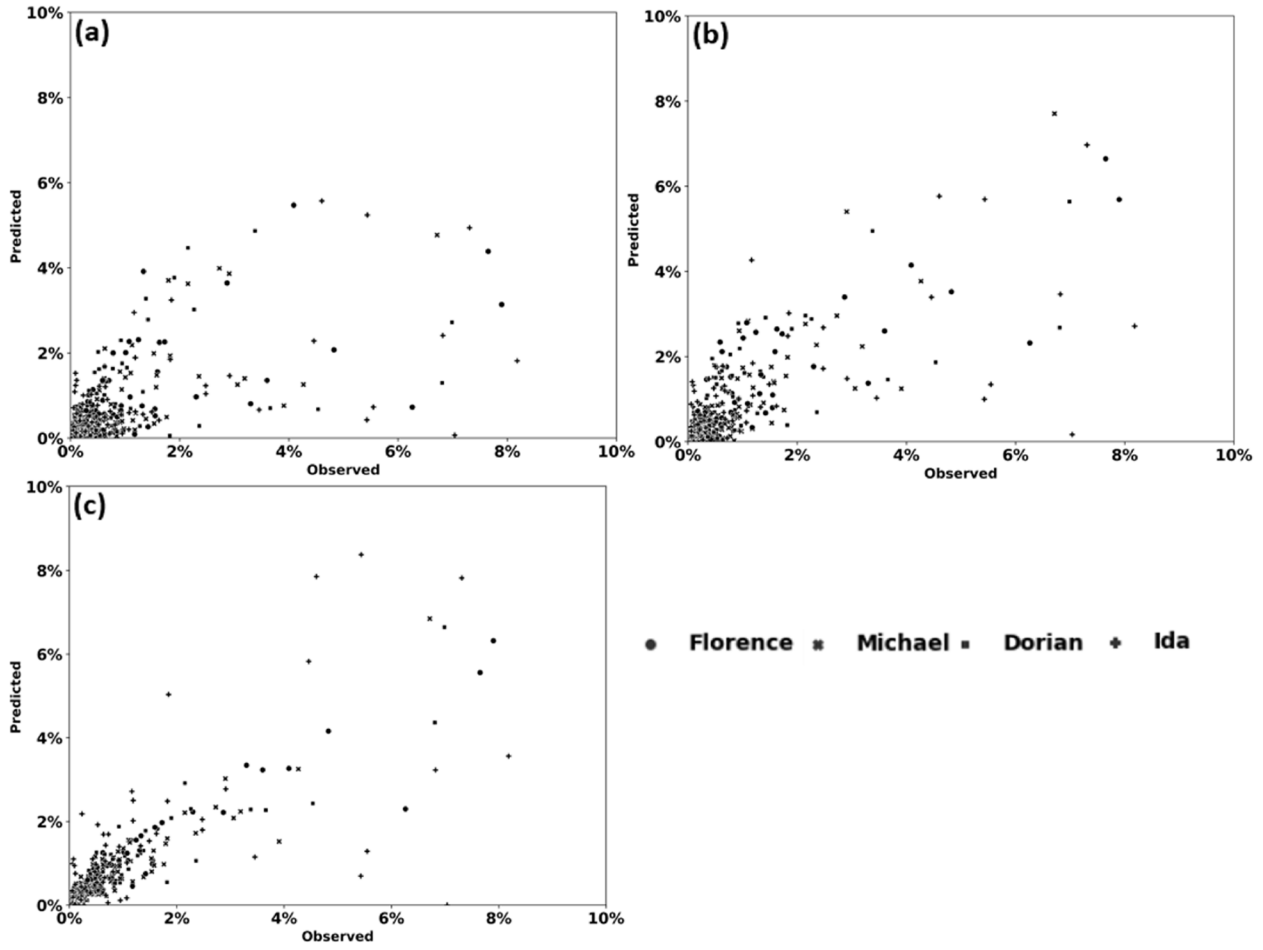


Fig. 8 Scatterplot of proportion of observed vrs predicted incoming evacuees for **a** Gravity, **b** XGBoost-trad, **c** XGBoost

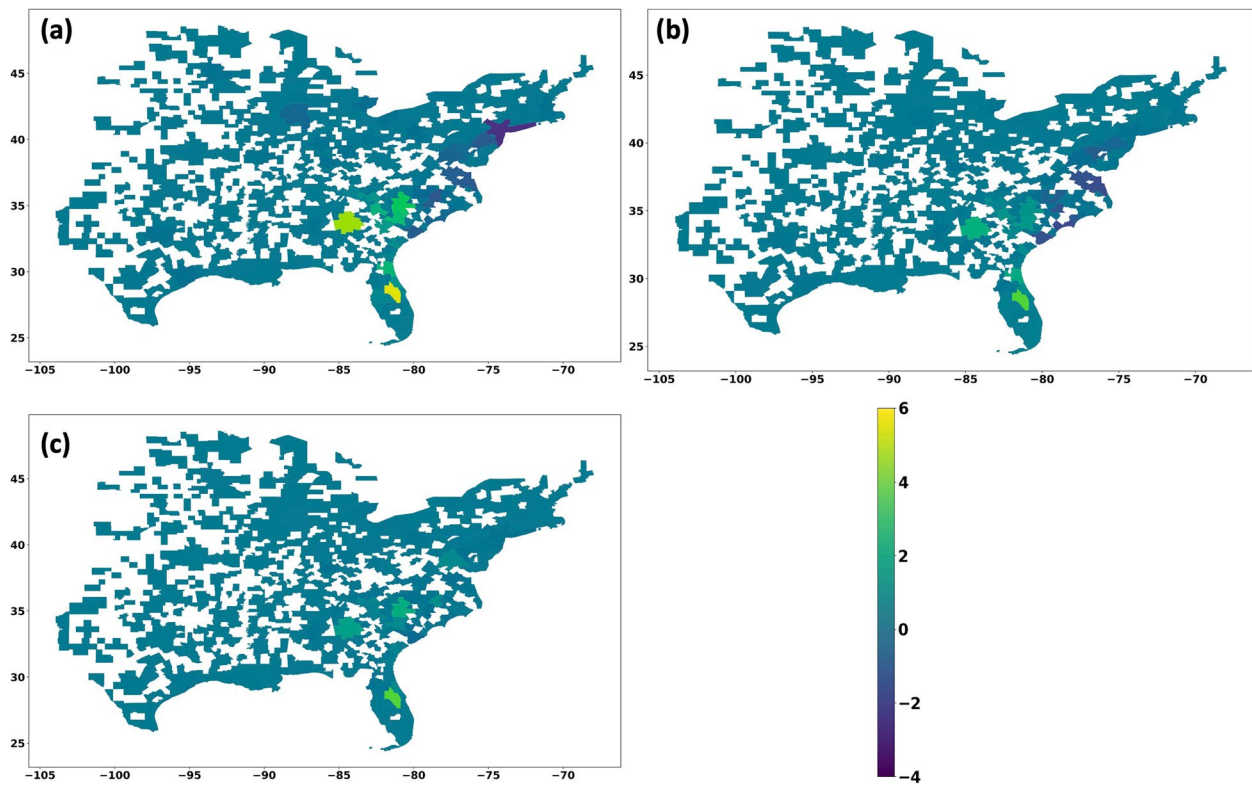


Fig. 9 Map of percentage distribution error on Florence for **a** Gravity, **b** XGBoost-trad, **c** XGBoost

Table 5 Literature on evacuation destination choice prediction

Authors/year	Hurricane	Data Type	Num obs	Model	Evaluation
Chen 2005	Floyd 1999	Telephone survey	1,042	Gravity, intervening opportunity	None
Wilmot et al. 2008	Floyd 1999	Telephone survey	1,042	Gravity, intervening opportunity models	Holdout for gravity on Andrew
Modali 2005	Floyd 1999	Telephone survey	697	Gravity (chi-square optimization)	Holdout on Andrew
Cheng et al., 2008	Floyd 1999	Telephone survey	1,040	Multinomial logit/discrete	None (on train data)
Cheng et al., 2011	Floyd 1999	Telephone survey	1,040	Gravity, time dependent gravity	None
Mesa-Arango et al., 2013	Ivan 2004	Telephone survey	1,419	Nested logit	None
Hong et al. 2020	Irma 2017	Geotagged Twitter	2,020	Machine learning methods	Holdout (75%: 25%)
Jiang et al., 2021	Mathew 2016	Geotagged Twitter	1,286	Gravity	CV

Acknowledgements

We are grateful to Brennan Lake and others at Cuebiq who provided access to the smartphone location data and support in using it. The statements, findings, conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation or Cuebiq.

Code availability

Code available upon request.

Authors' contributions

Prosper K. Anyidoho: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. Xinglong Ju: Conceptualization, Data curation, Formal analysis, Writing – review & editing. Rachel A. Davidson: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. Linda K. Nozick: Conceptualization, Supervision, Funding acquisition.

Funding

This work was supported by the National Science Foundation under Grant CMMI-1331269.

Availability of data and materials

Data available upon request.

Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 18 May 2023 Revised: 17 July 2023 Accepted: 4 August 2023
Published online: 14 September 2023

References

- Aleta, A., Martín-Corral, D., Pastore Y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini, I. M., Jr., Merler, S., Pentland, A., Vespignani, A., Moro, E., & Moreno, Y. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav*, 4(9), 964–971.
- Anyidoho, P. (2021). Prediction of time-dependent population behavior during hurricane evacuations, Master's thesis, University of Delaware, Department of Civil and Environmental Engineering.
- Anyidoho, P. K., Davidson, R. A., Rambha, T., & Nozick, L. K. (2022). Prediction of population behavior in hurricane evacuations. *Transp. Res. Part Policy Pract.*, 159, 200–221. <https://doi.org/10.1016/j.tra.2022.03.001>
- Baker, E. J. (1991). Hurricane evacuation behavior. *Int. J. Mass Emergencies Disasters*, 9(2), 287–310.
- Barrett, B., Ran, B., and Pillai, R. 2000. "Developing a dynamic traffic management modeling framework for hurricane evacuation." Transportation Research Record 1733, Transportation Research Board, Washington, DC., 115–121.
- Beigi, G., Hu, X., Maciejewski, R., and Liu, H. 2016. An overview of sentiment analysis in social media and its applications in disaster relief. In *Sentiment analysis and ontology engineering* (pp. 313–340). Springer, Cham.
- Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & Von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Medicine*, 8(8), e1001083.
- Beven II, J. L., Berg, R., and Hagen, A. (2019). Hurricane Michael. *National Hurricane Center*. Retrieved from https://www.nhc.noaa.gov/data/tcr/AL142018_Michael. Pdf.
- Beven II, J. L., Berg, R., and Hagen, A. (2021). Hurricane Ida. *National Hurricane Center*. Retrieved from https://www.nhc.noaa.gov/data/tcr/AL092021_Ida. pdf.
- Breslin, S. (2018). Hurricane Michael's Staggering Power Loss Documented by NASA. Retrieved from <https://weather.com/news/news/2018-10-17-hurricane-michael-power-outages-nasa>.
- Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers and Graphics*, 38, 51–60.
- Chae, J., Cui, Y., Jang, Y., Wang, G., Malik, A., and Ebert, D. S. (2015, May). Trajectory-based Visual Analytics for Anomalous Human Movement Analysis using Social Media. In *EuroVA@EuroVis* (pp. 43–47).
- Chen, B.(2005). "Modeling destination choice in hurricane evacuation with an intervening opportunity model". LSU Master's Theses. 2595.
- Cheng, G., Wilmot, C. G., & Baker, E. J. (2011). Dynamic gravity model for hurricane evacuation planning. *Transportation Research Record*, 2234(1), 125–134. <https://doi.org/10.3141/2234-14>
- Cheng, G., Wilmot, C. G., and Baker, E. J. (2008). "A destination choice model for hurricane evacuation." Transportation Research Board 87th Annual Meeting Compendium of Papers (DVD), Transportation Research Board of the National Academies, Washington, DC.
- Dash, N., & Gladwin, H. (2007). Evacuation decision making and behavioral responses: individual and individual. *Natural Hazards Review*, 8, 69–77.
- Fu, H., & Wilmot, C. G. (2004). Sequential logit dynamic travel demand model for hurricane evacuation. *Transportation Res Record J Transportation Research Board*, 1882(1), 19–26.
- Han, S. Y., Tsou, M. H., Knaap, E., Rey, S., & Cao, G. (2019). How Do Cities Flow in an Emergency? Tracing Human Mobility Patterns during a Natural Disaster with Big Data and Geospatial Data Science. *Urban Science*, 3(2), 51.
- Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Sci.*, 9, 29. <https://doi.org/10.1140/epjds/s13688-020-00247-6>
- Huang, S.-K., Lindell, M. K., & Prater, C. S. (2016). Who leaves and who stays? a review and statistical meta-analysis of hurricane evacuation studies. *J. Environment Behavior*, 48(8), 991–1029.
- Jha, M., K. Moore, and B. Pashaie. Emergency Evacuation Planning with Microscopic Traffic Simulation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1886, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 40–48.
- Jiang, Y., Li, Z., & Cutter, S. (2021). Social distance integrated gravity model for evacuation destination choice. *International Journal of Digital Earth*, 14(8), 1004–1018. <https://doi.org/10.1080/17538947.2021.1915396>
- Lenormand, M., Huet, S., Gargiulo, F., & Deffuant, G. (2012). A universal model of commuting networks. *PLoS ONE*, 7(10), e45985.
- Lenormand, M., Bassolas, A., & Ramasco, J. J. (2016). Systematic comparison of trip distribution laws and models. *Journal of Transport Geography*, 51, 158–169.
- Lindell, M. K., & Prater, C. S. (2007). Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: Examples from hurricane research and planning. *J. Urban Plann. Dev.*, 133(1), 18–29.
- Lundberg, S. M., & Lee, S. I. (2017, January). A Unified Approach to Interpreting Model Predictions. In *NIPS*.
- Martin, Y., Li, Z., & Cutter, S. L. (2017). Leveraging Twitter to gauge evacuation compliance: spatiotemporal analysis of Hurricane Matthew. *PLoS ONE*, 12(7), e0181701.
- Mei, B. (2002). *Development of trip generation models of hurricane evacuation*, Master's thesis, Louisiana State University, Department of Civil and Environmental Engineering.
- Mesa-Arango, R., Hasan, S., Ukkusuri, S. V., & Murray-Tuite, P. (2013). Household-level model for hurricane evacuation destination type choice using hurricane ivan data. *Natural Hazards Review*, 14(1), 11–20.
- Modali, N. K. (2005). "Modeling destination choice and measuring the transferability of hurricane evacuation patterns." M.Sc. thesis, Louisiana State Univ., Baton Rouge, LA.
- Mongold, E., Davidson, R.A., Trivedi, J., Deyoung, S., Anyidoho, P., 2020. Hurricane evacuation beliefs and behavior of inland vs. coastal populations. *Environ. Hazards* 1–19. <https://doi.org/10.1080/17477891.2020.1829531>.
- National Hurricane Center (NHC). 2021. NHC Data in GIS Formats, Accessed 26 March 2022. <https://www.nhc.noaa.gov/gis/>.
- National Hurricane Program (NHP). 2022. *HURREVAC Key Features*. Available at: <https://www.hurrevac.com>, Accessed 16 Dec 2022.
- National Weather Service (NWS). Hurricane Dorian - September 6, 2019. Retrieved from <https://www.weather.gov/mhx/Dorian2019>.
- North Carolina Department of Transportation. *Hurricane Floyd—Lessons Learned*, Raleigh, 2000.
- Radwan, E., M. Mollaghasemi, S. Mitchell, and G. Yildirim. *Framework for Modeling Emergency Evacuation*. Final report. Florida Department of Transportation, Tallahassee, 2005.
- Rambha, T., Nozick, L. K., & Davidson, R. (2021). Modeling hurricane evacuation behavior using a dynamic discrete choice framework. *Transport Res B-Meth*, 150, 75–100. <https://doi.org/10.1016/j.trb.2021.06.003>
- Robinson, C., & Dilkina, B.N. (2017). A Machine Learning Approach to Modeling Human Migration. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*.
- Song, X., Shibasaki, R., Yuan, N. J., Xie, X., Li, T., & Adachi, R. (2017). DeepMob: learning deep knowledge of human emergency behavior and mobility from big and heterogeneous data. *ACM Transactions on Information Systems (TOIS)*, 35(4), 41.
- Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., and Shibasaki, R. (2013, August). Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1231–1239). ACM.
- Song, X., Zhang, Q., Sekimoto, Y., and Shibasaki, R. (2014, August). Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 5–14). ACM.
- Southworth, F. (1991). Regional evacuation modeling: A state-of-the-artreview. ORL/TM– 11740, Oak Ridge National Laboratory, Oak Ridge, TN.
- Stewart, S. R., and Berg, R. (2019). Hurricane Florence. National Hurricane Center. Retrieved from https://www.nhc.noaa.gov/data/tcr/AL062018_Florence.pdf.
- Thompson, R. R., Garfin, D. R., & Silver, R. C. (2017). Evacuation from natural disasters: a systematic review of the literature. *Risk Analysis*, 37(4), 812–839.
- Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. (2010, April). Microblogging during twonatural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1079–1088). ACM.
- Wang, Q., & Taylor, J. E. (2014). Quantifying human mobility perturbation and resilience in Hurricane Sandy. *PLoS ONE*, 9(11), e112608.

- Wang, Q., & Taylor, J. E. (2016). Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLoS ONE*, *11*(1), e0147299.
- Wang, F., Wang, J., Cao, J., Chen, C., & Ban, X. J. (2019). Extracting trips from multi-sourced data for mobility pattern analysis: an app-based data example. *Transportation research. Part c, Emerging Technologies*, *105*, 183–202.
- Wilmot Chester G. (2008). Modeling Hurricane Evacuation Traffic: Testing the Gravity and Intervening Opportunity Models as Models of Destination Choice in Hurricane Evacuation.
- Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., ... and, Pitonakova, L. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal Earthquake. *PLoS Currents*, *8*.
- Xu, K., Davidson, R. A., Nozick, L. K., Wachtendorf, T., & DeYoung, S. E. (2016). Hurricane evacuation demand models with a focus on use for prediction in future events. *Transportation Research Part A*, *87*, 90–101.
- Yabe, T., Sekimoto, Y., Tsubouchi, K., & Ikemoto, S. (2019b). Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data. *PLoS ONE*, *14*(2), e0211375.
- Yabe, T., Tsubouchi, K., Shimizu, T., Sekimoto, Y., and Ukkusuri, S. V. (2019a). Predicting Evacuation Decisions using Representations of Individuals' Pre-Disaster Web Search Behavior. *arXiv preprint arXiv:1906.07770*.
- Yi, W., Nozick, L., Davidson, R., Blanton, B., & Colle, B. (2017). Optimization of the issuance of evacuation orders under evolving hurricane conditions. *Transportation Research Part b: Methodological*, *95*, 285–304.
- Yu, M., Yang, C., & Li, Y. (2018). Big data in natural disaster management: a review. *Geosciences*, *8*(5), 165.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.