ESTIMATION OF STOCHASTIC TRAFFIC CAPACITY USING EXTREME VALUE THEORY AND CENSORING: A CASE STUDY IN SALEM, NEW HAMPSHIRE

Eric M. LAFLAMME[1], Paul J. OSSENBRUGGEN[2]

[1] *Plymouth State University, Department of Mathematics, Plymouth, NH, USA*

[2] *University of New Hampshire, Department of Mathematics and Statistics, Durham, NH, USA*

Contact:
1) emlaflamme@plymouth.edu, 2) paul.ossenbruggen@unh.edu

Abstract:

In this work, we introduce a method of estimating stochastic freeway capacity using elements of both extreme value theory and survival analysis. First, we define capacity data, or estimates of the capacity of the roadway, as the daily maximum flow values. Then, under a survival analysis premise, we introduce censoring into our definition. That is, on days when flows are sufficiently high and congestion occurs, corresponding flow maxima are considered true estimates of capacity; otherwise, for those days that do not observe high flows or congestion, flow maxima are deemed censored observations and capacities must be higher than the observations. By extreme value theory, the collection of flow maxima (block maxima) can be appropriately approximated with a generalized extreme value (GEV) distribution. Because of small sample sizes and the presence of censoring, a Bayesian framework is pursued for model fitting and parameter estimation. To lend credence to our proposed methodology, the procedure is applied to real-world traffic stream data collected by the New Hampshire Department of Transportation (NHDOT) at a busy location on Interstate I-93 near Salem, New Hampshire. Data were collected over a period of 11 months and raw data were aggregated into 15-minute intervals. To assess our procedure, and to provide proof of concept, several validation procedures are presented. First, using distinct training and validation subsets of our data, the procedure yields accurate predictions of highway capacity. Next, our procedure is applied to a training set to yield random capacities which are then used to predict breakdown in the validation set. The frequency of these predicted breakdowns is found to be statistically similar to observed breakdowns observed in our validation set. Lastly, after comparing our methodology to other methods of stochastic capacity estimation, we find our procedure to be highly successful.

To cite this article:

1. Introduction

When studying traffic flow and volume, a prominent concept is that of roadway 'capacity' (Hyde and Wright, 1986). The Highway Capacity Manual (HCM) defines capacity as 'the maximum hourly rate at which persons or vehicles can be reasonably expected to traverse a point or a uniform section of a lane or roadway during a given time period, under prevailing roadway, traffic and control conditions' (Highway Capacity Manual, 2010). As this definition includes the term 'expected,' the capacity of a freeway facility is likely not a constant value.

It has been shown that capacity varies widely on a daily basis for the same facility and under the same geometric and traffic conditions (Lorenz and Elefteriadou, 2001; Elefteriadou et al., 1995; Persaud et al., 1998; Persaud et al., 2001; Brilon et al., 2005; Cassidy and Bertini, 1999; Kuhne et al., 2006; Li and Laurence, 2015). Moreover, breakdown does not necessarily occur at the same demand levels, but can occur when flows are lower or higher than the numerical value traditionally accepted as capacity (Elefteriadou et al., 1995; Li and Laurence, 2015). Thus, a single value of the capacity value for a freeway facility does not reflect real-world observations and capacity should be considered a random variable that is *stochastic* in nature (Brilon et al., 2005; Brilon and Geistefeldt, 2009; Geistefeldt, 2008; Dong et al., 2017). By considering capacity in this way one is left to identify/estimate capacity as a probability distribution across a range of values (Brilon et al., 2005; Lorenz and Elefteriadou, 2001). The corresponding cumulative distribution of capacity values has become known as the 'capacity distribution function,' F_C (Brilon et al., 2005), and is considered a valuable tool for evaluating roadway performance and efficiency.

In this analysis, we pursue the use of *daily flow maxima* as estimates of capacity. After outlining a theo-

retical justification for using such values, the generalized extreme value (GEV) distribution will be used for their approximation. Based on data collected from a location along Interstate 93 in New Hampshire, we will use the GEV model form to define a capacity distribution function, a measure of stochastic capacity, for the freeway segment. In addition to extreme value theory, this work will apply aspects of lifetime analysis (censoring) and computational Bayesian model-fitting based on Markov Chain Monte Carlo (MCMC) methods. This analysis, as it is presented herein, is a refinement of our previous work presented in Laflamme (2013). This work makes several improvements to that analysis, extends the work analysis to include a comparison to several other methods of capacity estimation, and generally presents a more concise approach to estimating capacity.

2. Methods

2.1. Data and preprocessing

To calibrate our models, we use real-world traffic stream data collected by the New Hampshire (NH) Department of Transportation at a single collection site in Salem, NH, along the northbound lane of I-93 just north of an off-ramp, exit 1, and just south of an on-ramp. Immediately north of this location (downstream), I-93 is physically constricted from three to two lanes (See Figure 1). As stated by Brilon et al. (2005), such sites (locations immediately upstream of a bottleneck) are ideal for the collection of capacity data.

Initially, it was suspected that the downstream on-ramp (See Figure 1) was the source of bottleneck, but data collected near (just upstream) the on-ramp (just downstream from our location) did not support this as very few incidents of congestion were observed there (around 30 breakdowns over 244 days).
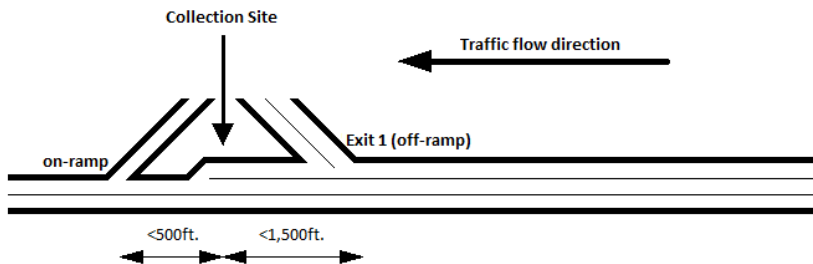

Fig. 1. Illustration of collection site structure along northbound lanes of I-93 in Salem, NH

However, our analysis of the traffic stream data across the highway segment supports a bottleneck located at the lane drop (at the lane drop, at the indicated collection site in Figure 1). In addition to this physical bottleneck, traffic volumes at this site exceed 100,000 vehicles per day (VPD) which far surpass the 60,000 to 70,000 VPD that the roadway was designed to accommodate (http://www.fhwa.dot.gov/construction/accelerated/wsnh0602.cfm). This heavy, daily flow at the site, in conjunction with frequent lane-changing, merging, weaving, etc. to accommodate the off-ramp, results in daily, recurrent congestion. Lastly, it is the opinion of the NH DOT that our collection site, because of the poor geometry and high volumes, is the bottleneck and source of congestion.

Data were collected between April 1 and November 30, 2010. During this time, side-fire radar devices intermittently measure traffic at irregular but frequent time periods about 1 minute apart. Data observations (raw data) consist of the following measurements: vehicle counts, average speed, occupancy, and speed (spot speed) of individual vehicles observed over the interval. 69% of days during the collection period experienced breakdown (breakdown criteria described in detail in the following section) for a total of 228 observed breakdowns (some days observed multiple, distinct breakdowns). Also, during the collection period, there were several incidents of missing values due to scheduled maintenance (scheduled shut-downs of the device) as well as unscheduled 'gaps' where the radar devices stopped collecting data (some lasted for several days).

Next, because radar data are collected over very short, irregular time intervals, these measurements were aggregated into uniform intervals of 15 minutes. The choice of interval length depends on the aim of the study, and estimates based on these intervals can vary substantially depending on the length used. For control studies, analyzing ramp metering systems, for example, very short intervals, sometimes as short as 30 seconds or 1 minute, are required. These short intervals have the ability to capture 'instantaneous' traffic behavior, but these are not sustained trends. Capacity studies, on the other hand, tend to use longer intervals, between 5 and 15 minutes, generally, so as to capture the underlying trend in flow and reduce (filter) noise. That is, 15-minute intervals are recommended to ensure 'stable' flow rates (Smith and Ulmer, 2003; Highway Capacity Manual, 2010) that are especially suitable for macroscopic/speed-flow analyses (HCM, 2010). As this work aims to identify the capacity of the roadway, the maximum sustainable rate, the longer 15-minute interval was used.

Finally, harmonic averages (see Daganzo 1997, for example) were calculated from flow and spot speed observations within each 15-minute interval to produce an aggregated flow rate (q) and aggregated speed (u) in units of vehicles/hour/lane (vph) and miles per hour (mph), respectively. Thus, for each day, aggregation yields u_t and q_t (speed and flow, respectively) series where t represents time-of-day with $t = 1, ..., 96$. Figure 2 illustrates the speed and flow aggregates produced for one week in April. Also, this figure illustrates the stochastic nature of congestion, how high flows typically, but do not necessarily, result in sustained slow speeds.
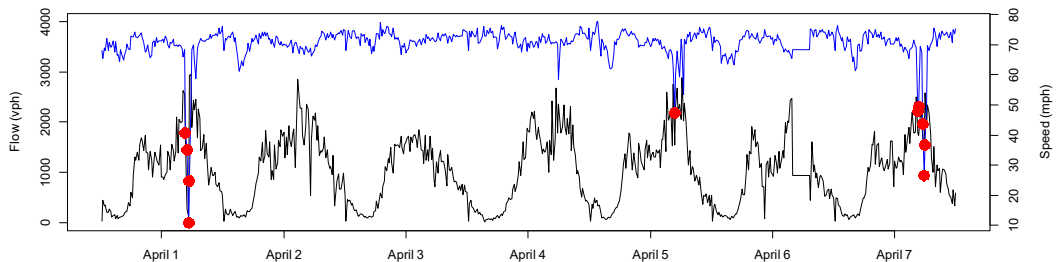

Fig. 2. Speed (blue) and flow (black) 15-minute aggregates for one week (April 1, 2010 – April 7, 2010). Red dots indicate speeds less than 50 mph. Notice that slower speeds correspond to sustained, high traffic flows on the first, fifth, and seventh days. Also, notice that sustained slow speeds do not occur on the third and fourth days, Saturday and Sunday

2.2. Capacity of the roadway

Researchers have come to acknowledge capacity as a stochastic process, and thus, instead of identifying capacity as a single, fixed value, have defined capacity as an entire probability distribution (for example, Elefteriadou et al., 1995). Such distributions are typically functions of traffic flow (Elefteriadou et al., 1995; Brilon et al., 2005), which means that every value across a continuous interval of traffic flows has a corresponding probability of breakdown (a congestion event resulting from traffic flow exceeding the capacity of the roadway). In order to calibrate such distributions of capacity, one must first define 'capacity data,' or estimates of the capacity of the roadway, then find an appropriate distribution, a capacity distribution denoted $F_C(c)$, to summarize these capacity data.

2.2.1. Breakdown/congestion identification

Because 'capacity data' is often dependent on identification of a **_'breakdown'_** or **_'congestion event,'_** we must precisely define these concepts. So, how do we define breakdown or congestion? In several studies, a fixed speed threshold, denoted u^*, is chosen to distinguished between freely flowing and congested states (Banks, 2009; Brilon et al., 2005; Geistefeldt and Brilon, 2009; Habbib-Mattar et al., 2009; Lorenz and Elefteriadou, 2001; Yeon et al., 2009, Li and Laurence, 2015). Using this definition and denoting u_t as the speed for a specified time t, when $u_t > u^*$ and $u_{t+1} < u^*$, a breakdown is identified at time t (See, for example, Lorenz and Elefteriadou, 2001). If, on the other hand, $u_t > u^*$ and $u_{t+1} > u^*$, no breakdown occurs at time t.

In our case, because the timing of the breakdown is not critical to our procedure (see below), we employ the procedure to identify if a particular day observes a breakdown. Specifically, a daily breakdown is identified whenever any of the 96 (15-minute) speed aggregates drop below a predetermined speed threshold, u^*. For threshold-based breakdown identification, no standard approach exists for identifying u^*, but based on visual inspection of our speed aggregates, a value of $u^* = 50$ mph was chosen. This value is similar to thresholds used by Brilon et al. (2005), Geistefeldt and Brilon (2009), Lorenz and Elefteriadou (2001), Yeon et al. (2009), who used fixed values of 47 mph, 50 mph, 43 mph, and 56 mph, respectively. In a related study, using the same I-93 raw data, Laflamme and Ossenbruggen (2017)

used the same definition of breakdown in their study of time-of-day and day-of-the-week on congestion duration and occurrence.

Note: since our choice of u^* is chosen with some degree of arbitrariness, we investigated the effect of using a different threshold of $u^* = 45$ mph, a value closer to the smaller thresholds used in similar studies (we feel larger thresholds, speeds bigger than 50 mph, are approaching freeflow speeds and cannot justifiably be used to identify congested traffic). After performing the analysis presented in the following sections for both $u^* = 45$ and $u^* = 50$, we found no evidence to suggest the resulting distribution of capacity was significantly affected by our choice of threshold. Thus, going forward, we use $u^* = 50$ as originally defined and assume the procedure is not particularly sensitive to thresholds within a realistic range.

An advantage of using speed aggregates to identify breakdown is that, because aggregates correspond to averages based on multiple vehicles, drops below u^* likely correspond to sustained drops in speed and 'true' breakdowns. For the remainder of this work, 'breakdown' or 'congestion' (used synonymously) refers to a transition from sustained speeds above u^* to sustained speeds below u^*. While we are confident that our definition captures true episodes of congestion, it is worth noting that no universal definition of congestion exists. As stated by Zochowska (2014), congestion should be treated as a relative phenomenon where expectations of the road system play a role in the perception of congestion.

2.2.2. Daily flow maxima as 'capacity data'

Based on the current literature related to stochastic capacity, while no one, universally-accepted measure of capacity has been established, **_'breakdown flows,'_** those flows measured immediately before the onset of congestion, have been widely adopted as capacity data, or good estimates of capacity (Elefteriadou and Lertworawanich, 2002; Brilon et al., 2005; Lorenz and Elefteriadou, 2001; Minderhoud et al., 1997). Using our definition and notation above, when breakdown occurs at time t, or when $u_t > u^*$ and $u_{t+1} < u^*$, q_t, the flow measured immediately before breakdown, is identified as the breakdown flow.

One concern when using such breakdown flows to measure capacity, however, is that higher flows often precede the breakdown flows. Often, prior to a freeway section entering a congested state (breaking down), traffic flow increases, peaks, and then begins to drop (Banks, 2009). In such cases, traffic flow is higher before the breakdown flow, and breakdown flow likely underestimates capacity. In this work, rather than using breakdown flows as estimates of capacity, we instead propose the use of *daily flow maxima*, the collection of maximum daily flows observed from the traffic stream data. Daily flow maxima, for these particular cases, would capture the higher flows prior to breakdown and ultimately obtain more representative measures of capacity, the maximum flow that a roadway can sustain. In our case, on days when breakdowns occur, the daily flow maximum is, on average, 484 vph greater than the corresponding breakdown flow. This is suggests that, generally, flow maxima are capturing the higher flows prior to breakdown as illustrated in Figure 3.

Another concern when using breakdown flows is that they are highly dependent on subjective, and often arbitrary, breakdown identification criteria. For example, breakdowns are typically defined by sustained speeds below a fixed speed threshold (50 mph, say). While it is simple to extract a breakdown flow when a breakdown occurs, changing the speed threshold can change the exact moment of breakdown and ultimately change the corresponding breakdown flow. Daily flow maxima, on the other hand, have the advantage of being independent of the methods used to identify breakdowns.

The use of daily flow maxima as estimates of capacity is indirectly supported in literature. That is, Hall and Agyemang-Duah (1991) and Hall et al. (1992) use *maximum pre-breakdown flow*, or the maximum sustained flow measured in some predetermined time window prior to a breakdown, to estimate capacity. Because daily flow maxima themselves typically occur prior to breakdowns (in fact, more than 80% of flow maxima occur within an hour of congestion) , these measures closely resemble maximum pre-breakdown flows, although the two are conceptually different (See Figure 3 for an illustration of breakdown flows, maximum pre-breakdown flows, and daily flow maxima). Minderhoud et al. (1997) discuss the use of 'extreme' values and state that observed maximum volumes (collected over days, for example) and corresponding extreme value statistics can be used to estimate capacity distributions. Next, applicable to our study, is the work of Hyde and Wright (1986) who use a variety of flow maxima to calibrate a capacity distributions based on direct probability methods and asymptotic theory. Lastly, among other methods to identify capacity distributions, Li and Laurence (2015) fit a variety of distributions (Normal, lognormal, Weibull, uniform) to the largest daily, five-minute flow rates (maxima) to data collected from San Diego (California) and Shanghai roadways. In this work, the authors identified little variation (standard deviation of veh/h/ln) in maxima-based capacity estimates over time (Li and Laurence, 2015).
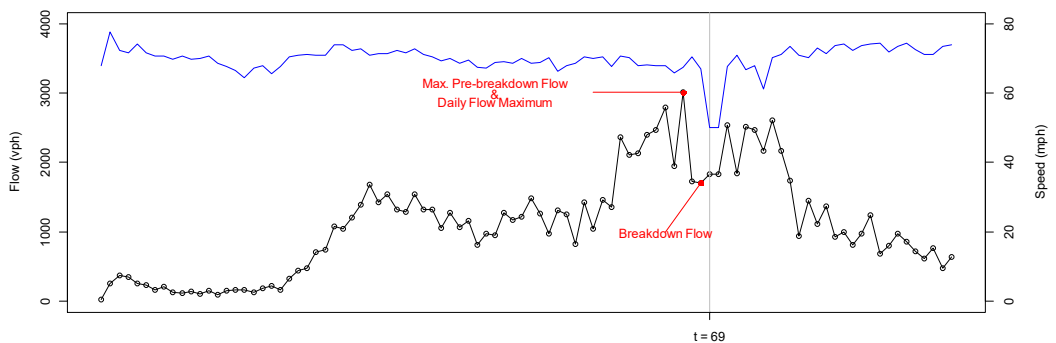

Fig. 3. Illustration of breakdown flow, maximum pre-breakdown flow, and daily flow maxima. Here, breakdown occurs at time point $t = 69$ (grey line). Note that on this day, the maximum pre-breakdown flow and daily flow maxima are identical values (3,008 vph). Also, note that the breakdown flow (1,704 vph) is surely an underestimate of the capacity of the roadway

2.2.3. Censoring

One issue that arises when considering daily flow maxima as capacity data is that high traffic flows are not observed on every day. Even our section of I-93, which typically observes high demand on a daily basis, occasionally experiences lower-than-usual traffic flows without congestion. Intuitively, daily flow maxima extracted from these few days would not be representative of roadway capacity. As stated by the HCM, the capacity for a given facility is the flow rate that can be achieved repeatedly for peak periods of **sufficient** demand (HCM, 2010). To remedy this situation, we introduce censoring into our definition of capacity data. For days when breakdown occurs (as defined, when speed aggregates drop below u^*), the demand is considered 'sufficiently' high, and the associated daily maximum flow is considered a 'true' estimate of capacity. However, on days where demand is insufficient, when breakdowns are not observed (when speed aggregates never drop below u^*), daily conditions are not adequate (sufficiently extreme) to assess the true capacity of the roadway. In these cases, under a survival analysis premise, the corresponding maxima are deemed censored (right-censored) capacity values as the roadway can surely service higher demands. That is, breakdowns would occur at some higher flow rates, and the resulting capacity, the maximum daily flows, would necessarily be larger than the observed value.

Despite the incompleteness associated with censored values, they still contain valuable information and will therefore be considered in the calibration of our capacity distribution (Geistefeldt, 2010). Non-parametrically, the capacity distribution function, $F_C(c)$, has been estimated using the Kaplan-Meier (Kaplan and Meier, 1958)/product limit method (PLM) based on samples that include both censored and uncensored values, a survival analysis approach (Brilon et al., 2005). It has been shown that, under this survival analysis framework, the Weibull model is well-suited and accurately approximates capacity data based on breakdown flows (Brilon and Zurlinden, 2003; Brilon et al., 2005). In a comparison of capacity distributions approaches, Geistefeldt and Brilon (2009) found that using censored data achieves significantly more precise estimates, especially at higher quantiles.

2.2.4. Notation

Let the random variable C denote the capacity of the roadway. Furthermore, let $M_j = \max(q_{j,t=1}, q_{j,t=2}, \ldots, q_{j,t=96})$ be the daily maximum flow for some day j in the collection period where $j = 1, \ldots, 224$. If $u_{j,t} < u^*$ for some time t, or if a breakdown occurs at some time t on day j and flows are deemed sufficient to estimate capacity, then M_j is considered a capacity datum and $M_j = c_J$. Otherwise, if $u_{j,t} \geq u^*$ for all t, or if a breakdown does not occur on day j and flows are deemed insufficient to estimate capacity, then M_j is considered a censored value. In these cases, M_j is less than the capacity of the roadway, or $M_j < c_J$.

2.3. Generalized extreme value distribution for maxima

By extreme value theory, the collection of maxima observed over fixed time units, or blocks, can be approximated by a superclass of distributions called the Generalized Extreme Value distribution, or the GEV. The GEV is given by the following form:

$$G(x) = exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \qquad (1)$$

defined on $\{x: 1 + \frac{\xi(x-\mu)}{\sigma} > 0\}$ with μ and σ the respective location and scale parameters. The shape parameter of the GEV, ξ, characterizes the rate of tail decay, where $\xi > 0$, $\xi = 0$, and $\xi < 0$ correspond to data with heavy tails, light tails, and short tails, respectively. For details related to extreme value theory, the reader is directed to Coles (2001), for example.

In the typical application of extreme value theory to environmental data, data are measured very sparsely (daily rainfall, for example) and blocks are set to long periods (a year, say) to limit bias in estimation. In our case, however, because we have densely measured data, blocks are defined as days. The resulting 'block maxima' are the daily flow maxima extracted from the traffic stream data, which we consider capacity data. Thus, in (1), we replace the random variable X with capacity data C, both censored and uncensored (observed) daily flow maxima, and $G(x)$ becomes a capacity distribution $F_C(c)$.

Although the GEV model assumes that block maxima are extracted from a series of independent observations, the GEV is still a reasonable distribution

form for block maxima extracted from dependent series such as we have (Coles, 2001). Additionally, the GEV approach assumes the collection of daily maxima are identically distributed, which, from observing the 8 months of data, appears to be the case. That is, we observe no true seasonality ('heavy' season), no trend, no oscillations, nor any systematic patterns in maxima.

2.4. Computational Bayesian approach

Daily maxima (capacity data) have two dominant characteristics: small sample sizes and censored values. Because of this, and because capacity model-fitting need not be done in real-time, a computational Bayesian approach was employed. Ozguven and Ozbay (2008) concluded that Bayesian estimation is superior for survival analyses with small samples and censoring. The OpenBUGS statistically software (Lunn et al., 2000) was used for all computation Bayesian estimation, and the analysis and manipulation of all OpenBUGS output, the data containing the Bayesian samples of the parameters, was then performed with the R statistical software (R Core Team, 2015). For a more detailed treatment of computational Bayesian procedures, the reader is directed to any number of sources including Carlin and Louis (2008) and Gelman et al. (2003).

Under the Bayesian framework, identifying the capacity distribution function ($F_C(c)$) is simply identifying the cumulative distribution that most accurately estimates the capacity data. By assuming a model form for this distribution function, the GEV, the objective of the analysis is then simply the estimation of the model parameters.

3. Results

Using asymptotic distributions, Hyde and Wright (1986) found flow maxima are approximated most accurately by short-tailed distributions. This makes intuitive sense as there is an absolute limit to the number of vehicles a road may carry. Thus, finite upper bounds were assumed for the capacity data, and, consequently, GEV shape parameters were assumed to be negative. Somewhat diffuse (semi-informative) prior distributions for the GEV scale and location parameters were assumed, but within a realistic range based on previous model-fitting. Capacities, C, are assumed to be generalized extreme value distributions (GEV), $C \sim GEV(\mu, \sigma, \xi)$, with lo-

cation, scale, and shape parameters μ, σ, and ξ, respectively. Collected data, M_j, are maximum daily traffic flows for day $j = 1, \dots, 244$, each of which is classified as a capacity datum or censored capacity. For the GEV shape (ξ), scale (σ), and location (μ) parameters, semi-informative, uniform prior distributions on (-.75, 0), (0, 10), and (2000, 5000), respectively, are used.

The following results, the posterior analysis, are based on the output of 5,000 MCMC iterations, the first 2,000 discarded as a 'burn-in' period. Convergence and independence from the starting values were checked using the 'coda' package in R (Plummer et al., 2006), the standard tools in such cases. Figure 4 gives fitted distributions (densities and CDFs) for several hundred sets of parameter estimates, and we have included the median fitted distribution as well as upper and lower 5% fitted distributions. For comparison purposes only, to illustrate the effect of including censored values in our model, we have also included GEV distributions fitted to only uncensored values.
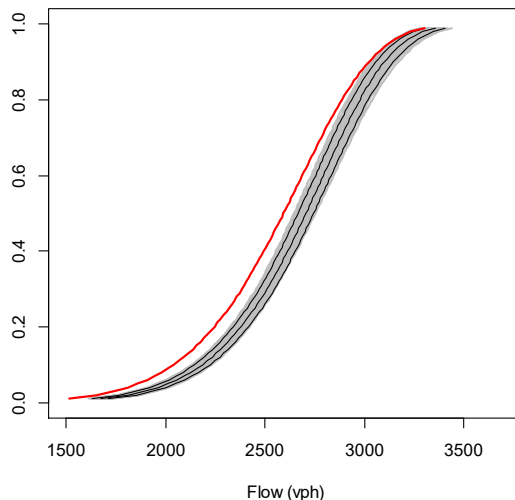

Fig. 4. Comparison of fitted cumulative distributions (right) with and without censored values. Grey lines represent individual fitted curves considering censored values; black lines represent median, upper 5%, and lower 5% curves. Red line represents fitted model to only uncensored values.

From Figure 4, we observe the effect of the inclusion of censored values on the fitted distribution, the capacity distribution function, F_C. Compared to the distribution fitted to uncensored values only (red curve), inclusion of censored values in model-fitting results in a right shift of the distribution (grey and black curves). To objectively assess this shift, and to compare a GEV model using uncensored values to the Bayesian-based GEV using both censored and uncensored values, we performed a two-sample Kolmogorov–Smirnov Test (KS Test). Briefly, the two-sample KS test is a non-parametric test used to determine if two probability distributions differ. Specifically, for a sample of size n with CDF $F(x)$ and a second sample of size m with CDF $G(x)$, the KS test tests $F = G$ (null hypothesis) against $F \neq G$ (alternative hypothesis). If $F_n(x)$ and $G_m(x)$ are the corresponding empirical CDFs of the samples, then the null hypothesis is rejected if the statistic $D = \sup_x |F_n(x) - G_m(x)|$ is larger than some associated threshold. In our case, the KS test of the two capacity distribution functions gave ample evidence to reject the claim that the two distributions are statistically similar ($D = 0.2323$; p-value = 0.0094; sample sizes $n, m = 99$) and underscores the important effect of censoring.

3.1. Validation

In order to assess our methodology, a variety of validation procedures were implemented. First,

a validation procedure was used to evaluate the plausibility of our GEV model choice in predicting the distribution of capacity data not used in the fitting process, to assess the model's probabilistic predictive ability. The cross-validation (CV) technique used here is a random sub-sampling procedure where the data is tested against itself. Under the premise of exchangeability, a training set is first created by randomly selecting 70% of both the censored and uncensored values. Of the remaining 30% of the data, only the uncensored capacity values were designated as a validation set. Since the training and validation sets are non-overlapping, the validation set may be considered truly unobserved data suitable for evaluating the procedure.

Using the Bayesian approach described previously, a GEV model is fit to the training set and then compared to the validation set. This process is repeated numerous times to ensure consistency, and result from one replication is presented in Figure 5. From this result, the histogram represent the distribution of capacity values in the validation set, those capacity values not used in model fitting (the 30% withheld from model fitting). The grey and black lines represent the GEV distribution (density) fitted via Bayesian methods to the training data. We observe the fitted GEV model visually captures the shape of the validation data, and therefore appears to successfully predict the distribution of capacity data.
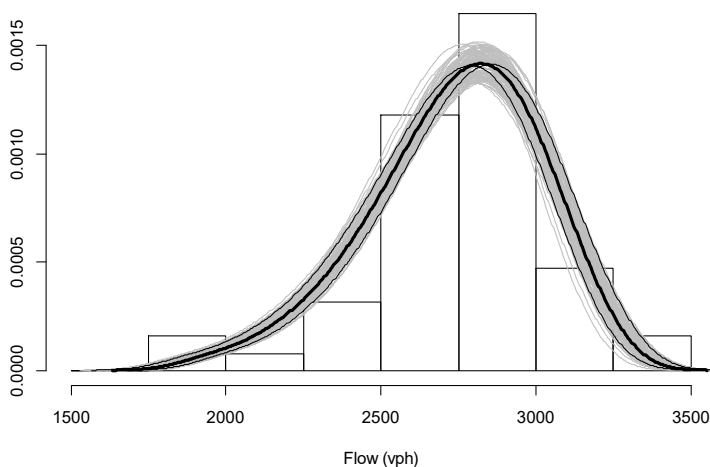

Fig. 5. Histogram of validation data, overlaid densities fitted to training data (grey lines), median training density (thick black line), and upper/lower 5% density curves (thin black lines)

Next, quantiles of the GEV distribution fitted to the training set were compared to the observed quantiles of the validation set via a quantile-quantile plot (see Figure 6). Results show excellent correspondence, and we have visual confirmation that the GEV model class is a successful predictor of capacity data, even at the upper tail (extreme values). Also, as above, this supports the use of censored values in model-fitting as their inclusion yields accurate probabilistic predictions. Lastly, a two-sample KS test was used to compare the distribution of the validation data to the fitted distribution from the training data. The KS test provided no evidence to reject the claim that the two distributions are statistically similar ($D = 0.099822$; p-value = 0.8906; sample sizes $n = 99$, $m = 51$).

We further validate our procedure by assessing the ability of daily flow maxima to approximate freeway capacity and predict breakdown. To do this, a validation procedure was performed in which observed traffic flows were compared to predicted capacity values. To do this, 70% of days in the collection period were randomly designated as a training set, and the remaining 30% of days in the collection period were designated as a validation set. Then, our model-fitting procedure was performed on the training data: daily maxima were extracted, maxima were identified as either capacity or censored values, and a GEV model was fitted using a computational

Bayesian procedure. From the fitted results, the mean parameter estimates, random capacities were generated. Flows from the validation set were then compared to the randomly generated capacity values. Figure 7 shows three days of validation data (flow values withheld from model-fitting) and the corresponding capacity estimates. In this particular case, since observed flows exceed random capacities, breakdowns are predicted on two of the three days.

Since we intend this procedure to be used for probabilistic prediction, we observe how well the procedure is able to predict the number of days when breakdown occurs. To do this, we simply observe the number of days (in the validation set) when breakdown is predicted, and then compare this to the number of days when breakdown actually occurs, when sustained speeds below 50 mph were observed (which we can determine based on speed records for these days). Zurlinden (2003), Brilon et al. (2007), and Geistefeldt and Brilon (2009) performed similar procedures to test the consistency of their capacity distributions. In the end, our procedure predicted that 44 of the 67 days in the validation set would observe breakdown, a rate of about 66%. We know, however, based on the known speeds for the days in the validation set, that 47 of the 67 days in the validation set (or 70%) observed true breakdowns.
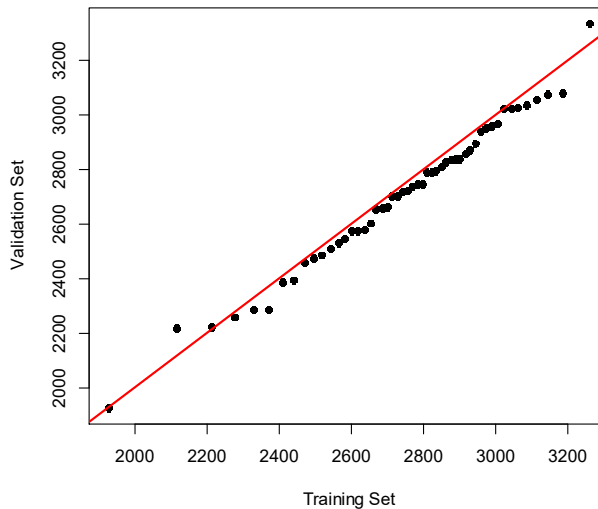

Fig. 6. Quantile-quantile plot comparing quantiles of the GEV model fit to the training set (x-axis) and quantiles of the observed capacities observed in the validation set (y-axis).
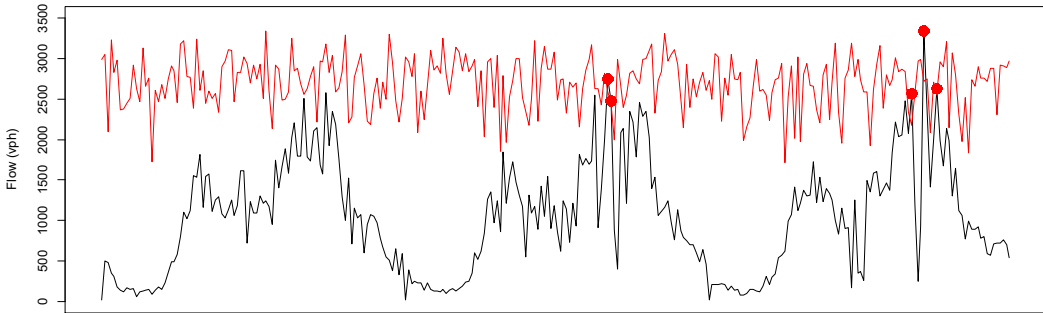
Fig. 7. Flow values (black) for a sample of three days from our validation set (data withheld from model-fitting) and randomly generated capacity values (red) from our training set. Red dots indicate occurrences when observed flow exceeds the capacity of the roadway, breakdown. Note that two of three days observe breakdown

To objectively compare these rates (66% versus 70%), we performed a uniformly most powerful (UMP) unbiased test, the two-sample rate ratio test. Specifically, the test assumes the rates are equal, or that the ratio of these rates is equal to 1. If the rate ratio is very different from 1, then the assumption is rejected and we conclude the rates are different. In our case, the ratio of estimated rates is found to be 0.9361 (the 95% confidence interval for the rate ratio (0.6062, 1.4432)), which is not significantly different from 1 (estimated ratio compared to a binomial distribution; p-value = 0.8341). Thus, we have no evidence to reject the claim that the rates of predicted and observed breakdowns are different. This agreement between predicted and observed proportion of breakdowns supports the use of daily flow maxima as an estimate of capacity, provides evidence that our model is a suitable probabilistic predictor of breakdown, and generally gives credence to our procedure.

3.2. Comparison to other methods of estimating capacity

Since we propose that our procedure for estimating capacity is advantageous over current methods of estimating capacity, a validation procedure is implemented in which our method is compared to other, common methods of capacity estimation. Specifically, we compare our method to the approaches discussed in Brilon et al. (2005), the non-parametric product limit method (PLM) approach (Kaplan and Meier, 1958) and the survival analysis approach using a Weibull distribution (Brilon et al., 2005), as

well as survival-based approaches using both Gaussian (Normal) and logistic distributions. These same methods of capacity estimation were discussed in Kim et al. (2010), who investigated the effect of weather on capacity.

Similar to our validation of our procedure using maximum flow to estimate capacity, we now test the PLM, Weibull, Gaussian, and logistic models. As above, 70% of the days in the collection period were randomly designated as a training set, and the remaining 30% were designated at a validation set. Then, from the training set, censored and uncensored values were identified and each of the four models was used to estimate distributions of capacity. Finally, these fitted distributions were compared to the capacity values identified in the validation set (the values not used in the fitting process). Figure 8 shows the CDFs of the four methods (the resulting capacity functions) as well as the empirical CDF of the observed capacities in the validation set. While the brown line, corresponding to the capacity function based on a logistic model, is closest to the distribution of validation capacities (based on a two-sample KS test with $D = 0.28758$; p-value = 0.007639; sample sizes $n = 99$, $m = 51$) none of the distributions are deemed statistically similar to the observed capacities (based on KS tests). We remind the reader that in the validation procedure for our modeling approach (see Validation), we found the resulting capacity function was statistically similar to distribution of observed capacities (based on KS test).

Note: it is not possible to put our previous result (GEV procedure) and this validation result in the same figure as our procedure and the results below are based on different estimates of capacity (our procedure uses maxima and below we use breakdown flow). However, a direct comparison of our procedure to these approaches will be discussed later.

As a last validation procedure, our method of estimating capacity (GEV, maximum flow) is compared to the four methods above (PLM, Weibull, Gaussian, and logistic models) in terms of actually predicting breakdowns. As previously done for our procedure (see above), we use the fitted distributions (fit to the training datasets) to generate random distributions of capacities. Flows from the validation set (observed flows) are then compared to the randomly generated capacity values for the four different models. Then, we simply observe the number of days (in the validation set) when breakdown is predicted, and then compare this to the number of days when breakdown actually occurs, when sustained speeds below 50 mph were observed (which we can determine based on speed records for these days). As done previously, we compare the number of predicted days experiencing traffic breakdown to the observed number of day experiencing breakdown using a rate ratio test. Recall that 47 of the 67 days in the validation period observed sustained breakdowns. Based on the comparison of observed flows and random capacities generated from the fitted PLM, Weibull, Gaussian, and logistic models, we have predicted breakdowns in 57, 41, 38, and 40 days, respectively. These values agree with the CDF plot (Figure 8) as the Gaussian curve, among those four distributions, has the highest proportion of large capacity values and we would thus expect this distribution to yield the lowest number of breakdowns. Based on the p-values associated with the rate ratio tests, we find that none of the predicted rates are statistically different from the observed rate (47 of 67 days). That said, the predicted number of days with breakdown from our procedure, 44 (see last paragraph in section 3.1 above), is the closest to the observed number of days with breakdown.

While the four models and our procedure all do well to predict the number of days with breakdowns, our procedure is especially accurate in terms of total number of breakdown predicted in the validation period. During the validation period, as mentioned above, there were 47 of the 67 days that observe traffic breakdown. But, during some of these 47 days, multiple, distinct breakdowns were observed (when breakdown occurs, then has a sustained period of freely flowing traffic before another breakdown). In fact, we observe 65 distinct breakdowns in this validation period. Using our GEV procedure and maximum flows, after fitting a distribution to the training data, generating random capacities, and comparing these random capacities to observed flows from the validation period, we find 74 predicted breakdowns (compared to the 65 true breakdowns). To statistically test if these counts are different, we use a Chi-square test that assumes the counts come from a common Poisson distribution (this is equivalent to a Z-test of the difference of two Poisson rates). The corresponding test statistic is the accumulation of squared differences between observed and expected counts under the assumption of a common Poisson rate parameter. This statistic, X^2, follows a Chi-square distribution with one degree of freedom (d.f. = 1). In our case, comparing 65 observed breakdowns to our predicted 74 breakdowns, we find no evidence to reject the assumption that these are realizations from a common distribution ($X^2 = 0.583$, degrees of freedom (d.f.) = 1, p-value = 0.4452). Similar tests of the other four models (PLM, Weibull, Normal, and logistic models) found that only the PLM predicted a statistically dissimilar number of breakdowns (125) than the observed count ($X^2 = 18.947$, d.f. = 1, p-value = 0.00001). It is worth noting, however, that our procedure based on maximum flows gave a predicted value closest to the observed number of breakdowns.

For comparison purposes, a purely deterministic measure of capacity was tested using the same procedure. Using a fixed capacity value of 2,656 vph, the mean of breakdown flows in the training period, and using the procedure outlined above, we have predicted breakdowns on only 27 of the 67 days. Using a rate ratio test, this is ratio is found to be statistically dissimilar (estimated ratio of 0.5744, p-value = 0.0265) to the observed rate of 47 out of 67. Furthermore, a deterministic capacity measure predicts a total of 50 breakdowns during the validation period, a count found to be nearly statistically dissimilar ($X^2 = 1.9565$, d.f. = 1, p-value = 0.1619) to the observed number of breakdowns (65).
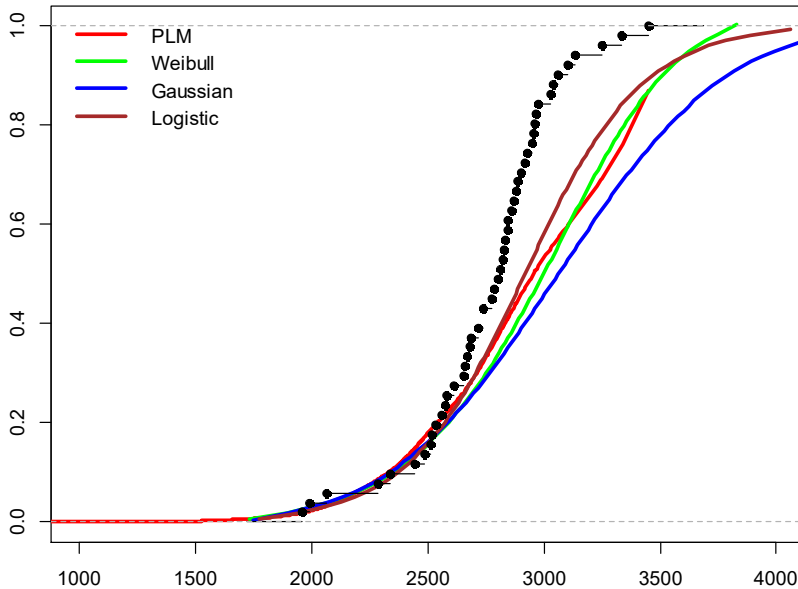
Fig. 8. Capacity distribution functions based PLM, Weibull, Gaussian, and logistic models (red, green, blue, and brown curves, respectively). These models were fit to training data (both capacity and censored values) and compared to the distribution of capacity values observed in the validation set (black dots)

Note: for all of the validation procedures, we focus on *probabilistic* predictions and, as such, focus on predicting distributions and frequency of predicted breakdowns. We cannot reasonable expect any of these procedures (either our procedure based on maximum flows or other procedures using breakdown flows) to predict the exact timing of a breakdown.

4. Discussion and Conclusion

In this work, a number of techniques were examined to offer a new perspective on the estimation of stochastic capacity. First, we have made a case for considering daily flow maxima as estimates of roadway capacity. When daily maxima correspond to breakdowns, daily maximal flows are considered capacity data; otherwise, these maxima are considered censored (right-censored) estimates. Extreme value theory suggests such estimates of capacity, maximum flows, may then be suitably approximated by the generalized extreme value (GEV) distribution for block maxima. To introduce the censored values, and to address small sample size, a Bayesian framework was implemented using semi-informed priors.

Validation of the procedure, using real-world data, provides evidence that the combined application of methodologies (extreme value analysis, censoring) can yield accurate distributions of capacity. Comparisons to other methods of capacity estimation show illustrate the validity of the approach. We note that, since the procedure was applied to just one location (just one set of data), we make no claim regarding transferability. Rather, we offer the analysis as a 'proof of concept' of our procedure. It is not clear that process can be suitably applied to locations dissimilar to that used in this analysis, and future work will certainly explore the application of the technique to more diverse road sections. With additional locations, a hierarchical model may possibly be pursued.

Going forward, there are numerous other areas into which our work can be extended. First, this analysis considered the roadway as a homogenous unit and made no distinction between lanes (left, right, median, etc.). Identifying capacity distributions for individual lanes is a logical extension of this work as evidence suggests that traffic behavior between lanes may be quite dissimilar (Cassidy and Bertini,

1999). Also, the effect of driver behavior and traffic composition should be investigated. For example, in separate analyses, Hall (1995) and Bharadwaj (2016) identified that lane changes (weaving) and heterogeneity of traffic, respectively, have appreciable effects on the capacity of the roadway. Next, following Brilon et al. (2005), we may investigate the effect of external conditions such as rain/ inclement weather into our capacity estimates. Along these lines, Ponzlet (1996) demonstrated that capacities vary according to external conditions such as dry/wet road surfaces or daylight/darkness. So long as detailed, local records are available, identifying the effects of weather variables on capacity distributions is feasible. Lastly, and ultimately, our capacity models and corresponding predictions of congestion/breakdown could be used to identify the 'level of impact' on the system. This is, of course, a complicated topic that incorporates a variety of factors (among which are traffic flows), and a topic pursued by Zochowska (2014) in an urban network setting. Ultimately, as concluded by Zochowska (2014) for city systems, estimates of level of impact (probability of congestion, say) could be conveyed to the user to improve the functioning of the transportation system.

References

[1] AGYEMANG-DUAH, K. & HALL, F.L., 1991. Some Issues Regarding the Numerical Value of Freeway Capacity, Highway Capacity, and Level of Service. *International Symposium on Highway Capacity*. Karlsruhe, Netherlands, 1991. Rotterdam: AA Balkema, pp. 1-15.

[2] BANKS, J.H., 2009. Flow Breakdown at Freeway Bottlenecks: Evidence from Automated Flow Analysis. *Transportation Research Record: Journal of the Transportation Research Board*. **2099**(2009), pp. 14-21.

[3] BHARADWAJ, N., MATHEW, S., PANI, A., ARKATKAR, S., JOSHI, G., & RAVINDER, K., 2016. Effect of Traffic Composition and Emergency Lane on Capacity: A Case Study of Intercity Expressway in India. *Transportation Letters*. DOI: 10.1080/19427867.2016.1265237

[4] BRILON, W. & GEISTEFELDT, J., 2009. Implications of the Random Capacity Concept for Freeways. *2nd International Symposium on Freeway and Tollway Operations*. Honolulu, HI, 2009.

[5] BRILON, W., GEISTEFELDT, J., & REGLER, M., 2005. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proceedings of the 16th International Symposium on Transportation and Traffic Theory.* College Park, Maryland, 2005, pp. 125-144.

[6] BRILON, W., GEISTEFELDT, J. & ZURLINDEN, H., 2007. Implementing the Concept of Reliability for Highway Capacity Analysis. *Transportation Research Record: Journal of the Transportation Research Board*. **2027**(2007), pp. 1-8.

[7] BRILON, W. & ZURLINDEN, H., 2003. Overload probabilities and traffic activity as design criteria for road traffic systems. *Research in Road Construction and Road-Traffic Technique* (series). **870**(2003).

[8] CARLIN, B. P., & LOUIS, T.A., 2008. *Bayesian Methods for Data Analysis*. 3rd edition. Boca Raton, FL: Chapman and Hall/CRC.

[9] CASSIDY, M.J. & BERTINI, R.L., 1999. Some Traffic Features at Freeway Bottlenecks. *Transportation Research, Part B: Methodological.* **33**(1), pp. 25-42.

[10] COLES, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

[11] DONG, S., MOSTAFIZI, A., WANG, H., & LI, J., 2017. A Stochastic Analysis of Highway Capacity: Empirical Evidence and Implications. *Journal of Intelligent Transportation Systems*. DOI: 10.1080/15472450.2017.1396898

[12] ELEFTERIADOU, L. & LERTWORAWANICH, P., 2002. A Methodology for Estimating Capacity at Ramp Weaves Based on Gap Acceptance and Linear Optimization. *Transportation Research, Part B: Methodological.* **37B**(5), pp. 459-483.

[13] ELEFTERIADOU, L., ROESS, R.P., & MCSHANE, W.R., 1995. Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record: Journal of the Transportation Research Board.* **1484**(1995), pp. 80-89.

[14] GELMAN, A., CARLIN, J.B., STERN, H.S. & RUBIN, D.B., 2003. *Bayesian Data Analysis*. 2nd edition. Boca Raton, FL: Chapman and Hall/CRC.

[15] GEISTEFELDT, J., 2008. Empirical Relation between Stochastic Capacities and Capacities Obtained from the Speed-Flow Diagram. *Proceedings of the Symposium of the Fundamental Diagram: 75 years*. Woods Hole, MA, 2008.

[16] GEISTEFELDT, J., 2010. Consistency of Stochastic Capacity Estimations. *Transportation Research Record: Journal of the Transportation Research Board.* **2173**(2010), pp. 89-95.

[17] GEISTEFELDT, J. & BRILON, W., 2009. A Comparative Assessment of Stochastic Capacity Estimation Methods. *18th International Symposium on Transportation and Traffic Theory*. Hong Kong, 2009.

[18] HALL, F.L. & AGYEMANG-DUAH, K., 1991. Freeway Capacity Drop and the Definition of Capacity. *Transportation Research Record: Journal of the Transportation Research Board.* **1320**(1991), pp. 91-98.

[19] HALL, F.L., HURDLE, V., & BANKS, J., 1992. Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways. *Transportation Research Record: Journal of the Transportation Research Board.* **1365**(1992), pp. 12-17.

[20] HALL, R.W., 1995. Longitudinal and Lateral Throughput on an Idealized Highway. *Transportation Science*. 29(2), pp. 118-127.

[21] HYDE, T. & WRIGHT, C.C., 1986. Extreme Value Methods for Estimating Road Traffic Capacity. *Transportation Research, Part B: Methodological.* **20B**(2), pp. 125-138.

[22] KAPLAN, E.L. & MEIER, P., 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association.* **53**, pp. 457-481.

[23] KIM, J., MAHMASSANI, H. & DONG, J., 2010. Likelihood and duration of flow breakdown: modeling the effect of weather. *Transportation Research Record: Journal of the Transportation Research Board.* **2188**(2010), pp. 19-28.

[24] KUHNE, R.D., MAHNKE, R., & HINKEL, J., 2006. Modeling the Effects of Corridor Control Systems on Road Capacity. *5th International Symposium on Highway Capacity and Quality of Service*. Yokohama, Japan, 2006, pp. 289-298.

[25] LAFLAMME, E.M., 2013. Extreme value theory: Applications to estimation of stochastic traffic capacity and statistical downscaling of precipitation extremes. PhD. Thesis, University of New Hampshire (Durham, NH, USA).

[26] LAFLAMME, E.M. & OSSENBRUGGEN, P.J., 2017. Effect of time-of-day and day-of-the-week on congestion duration and breakdown: A case study at a bottleneck in Salem, NH. *Journal of Traffic and Transportation Engineering (English Edition).* **4**(1), pp. 31-40.

[27] LI, Z. & LAURENCE, R., 2015. An Analysis of Four Methodologies for Estimating Highway Capacity from ITS Data. *Journal of Modern Transportation*. 23(2), pp. 107-118.

[28] LORENZ, M. & ELEFTERIADOU, L., 2001. Defining Freeway Capacity as Function of Breakdown Probability. *Transportation Research Record: Journal of the Transportation Research Board.* **1776**(2001), pp. 43-51.

[29] LUNN, D.J., THOMAS, A., BEST, N., & SPIEGELHALTER, D., 2000. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing.* **10**(2000), pp. 325–337.

[30] MINDERHOUD, M.M., BOTMA, H., & BOVY, P.H.L., 1997. Assessment of Roadway Capacity Estimation Methods. *Transportation Research Record: Journal of the Transportation Research Board.* **1572**(1997), pp. 59-67.

[31] OZGUVEN, E. E. & OZBAY, K., 2008. Nonparametric Bayesian Estimation of Freeway Capacity Distribution from Censored Observations. *Transportation Research Record: Journal of the Transportation Research Board.* **2061**(2008), pp. 20-29.

[32] PERSAUD, B., YAGAR, S., & BROWNLEE, R., 1998. Exploration of the Breakdown Phenomenon in Freeway Traffic. *Transportation Research Record: Journal of*

the Transportation Research Board.
1634(1998), pp. 64-69.

[33] PERSAUD, B., YAGAR, S., TSUI, D., & LOOK, H., 2001. Study of Breakdown-Related Capacity for a Freeway with Ramp Metering. *Transportation Research Record: Journal of the Transportation Research Board.* **1748**(2001), pp. 110-115.

[34] PLUMMER, M., BEST, N., COWLES, K., & VINES, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News.* **6**(2006), pp. 7-11.

[35] PONZLET, M., 1996. Dynamik der Leistungsfähigkeiten von Autobahnen (Dynamics of Freeway Capacity*).* PhD. Dissertation, Institute for Transportation and Traffic Engineering, Ruhr-University Bochum.

[36] R CORE TEAM, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: http://www.R-project.org/

[37] TRANSPORTATION RESEARCH BOARD, 2010. Highway Capacity Manual (HCM). Washington, D.C.

[38] U.S. DEPARTMENT OF TRANSPORTATION, FEDERAL HIGHWAY ADMINISTRATION, 2012. *Localized Bottleneck Reduction Program*. U.S. Dept. of Transportation, Federal Highway Administration. [Viewed 2012]. Available from: http://www.ops.fhwa.dot.gov/bn/lbr.htm

[39] YANG, X. & ZHANG, N., 2005. The Marginal Decrease of Lane Capacity with the Number of Lanes on Highway. *Proceedings of the Eastern Asia Society for Transportation Studies.* **5**(2005), pp. 739–749.

[40] ZOCHOWSKA, R., 2014. Selected Issues in Modelling of Traffic Flows in Congested Urban Networks. *Archives of Transport*. 29(1), pp. 77-89.

[41] ZURLINDEN, H., 2003. Ganzjahresanalyse des Verkehrsflusses auf Strassen (Whole year analysis of traffic flow on highways). *Schriftenreihe des Lehrstuhls fuer Verkehrswesen der Ruhr-Universitaet Bochum.* **26**.

Declarations