

AN ESTIMATOR FOR TRAFFIC BREAKDOWN PROBABILITY BASED ON CLASSIFICATION OF TRANSITIONAL BREAKDOWN EVENTS

Petter Arnesen Ph.D. (corresponding author)

Department of Transport Research
SINTEF Technology and Society
Telephone: (0047) 98083886; E-mail: petter.arnesen@sintef.no

Odd A. Hjelkrem Ph.D.

Department of Transport Research
SINTEF Technology and Society
Telephone: (0047) 98482278; E-mail: oddandre.hjelkrem@sintef.no

26.04.17

ABSTRACT

In this paper we propose a new estimator for calculating the probability of traffic breakdown as a function of traffic demand. Traffic breakdown is a well studied phenomena within previous literature, and is of great importance to traffic planners and controllers. The proposed estimator has an appealing intuition and is able to overcome several of the problems associated with previously proposed methodology. The input to the estimator is a set of aggregated (typically 5 minutes) traffic observations classified to either breakdown or non-breakdown state, and a customized and fast algorithm for this purpose is proposed. Lastly, we apply the classification algorithm and breakdown probability estimator to a large data set consisting of several observation sites on the Norwegian road network, and compare our estimator to a previously defined estimator.

Keywords: Congestion; Probabilistic capacity; Traffic breakdown; Traffic flow

1 INTRODUCTION

Congestion in traffic flow is responsible for significant delays for travelers. Congestion occurs when a traffic volume increases beyond a critical value, causing a breakdown shift from an uncongested flow to congested state. The mechanics of traffic breakdown has been extensively researched in the past, see for instance Elefteriadou et al. (1995), Persaud et al. (1998), Brilon et al. (2005), Kerner (2009), Elefteriadou et al. (2011) and the references therein. A common perception of the relationship between the three important variables flow, speed and density is represented by the fundamental diagram of traffic flow (May, 1990). According to the fundamental diagram, each set of prevailing conditions has a specific capacity, which is the maximum value of traffic volume. An increase in density while at the capacity will result in a decreased speed and flow, also known as a breakdown. Empirically, the transition from uncongested to congested flow results in a large drop in speed (Sugiyama et al., 2008). If this phase transition could have been avoided, delays would decrease, resulting in an increased benefit for travelers. With the traffic jams worldwide in mind, the potential for total increase in benefit is substantial.

Estimating the probabilistic nature of phase transitions in traffic

Some of the first to document the probabilistic behavior of traffic breakdowns were Elefteriadou et al. (1995). They concluded from empirical observations that the capacity was a probabilistic value rather than deterministic, and developed a model which takes into account vehicle clusters on ramps and freeway flow. Before this, the capacity was thought of as a fixed value for each location, and that a traffic flow beyond the capacity would induce a breakdown. Several other studies later confirmed that breakdowns are probabilistic, which has resulted in a widely accepted view on breakdowns as a probabilistic event in recent literature (for instance Brilon et al. (2005), Sugiyama et al. (2008), Elefteriadou et al. (2011), Xu et al. (2013) and Chen et al. (2014)).

The probability of a breakdown is usually represented as a function of traffic flow. At low traffic volumes, the probability of a breakdown is zero. At very high volumes, it is almost certain that a breakdown will occur. The phase transition between free flow and breakdown is usually represented with an "S"-curve, as shown in Figure 1.

Several studies have calculated the probability function from empirical data, e.g. Elefteriadou et al. (1995), Persaud et al. (1998) and Lorenz and Elefteriadou (2000). The main features of the approach was to group observations of flow in multiples of 100 vehicles per hour, and find the frequency of observations where a breakdown has occurred as N_B . The probability P_B for a breakdown at a certain input flow was then calculated as

$$P_B = N_B/N_T, \quad (1)$$

where N_T is the frequency of all observations in the specific group of flow values. A calculation of P_B for each group of flows should lead to a function as proposed in Figure 1.

As empirical data of sufficient size and quality has been difficult to obtain, several researches have used simulations to produce probability functions for traffic breakdown, where Markov chain techniques have been the most common approach. Most common of these simulations have used Markov Chains. Wang et al. (2010) used a discrete time Markov chain (DTMC) to model the traffic state transition, and were able to calculate the breakdown probability as a function of density. Shiomi et al. (2011) developed a traffic flow model for bottlenecks for estimating breakdown probability. The model consisted of platoon formation and a speed transition

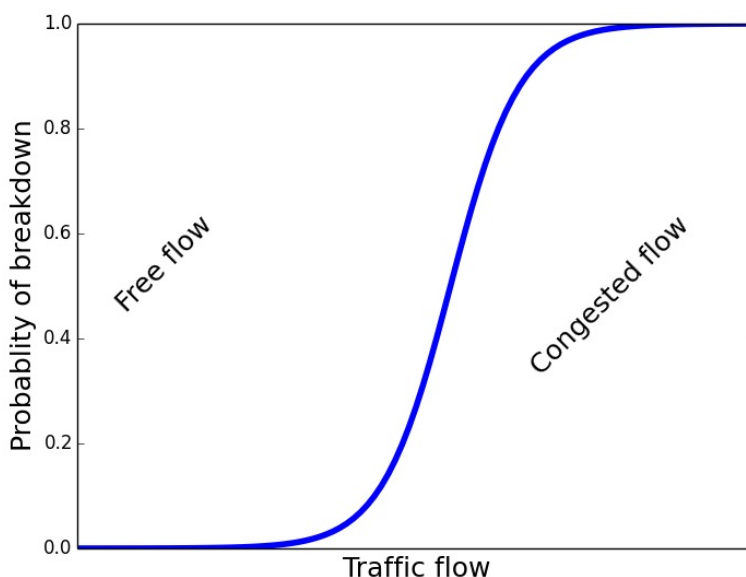


FIGURE 1 A hypothetical representation of the probability of a traffic breakdown as a function of traffic flow.

model based on an absorbing Markov chain model. Xu et al. (2013) used a Monte Carlo model for simulating freeway traffic, and calculated the breakdown probability as a function of input flow. Chen et al. (2014) proposed a queuing model to describe traffic breakdown phenomena caused by perturbations of on-ramp merging vehicles. Kerner et al. (2002) calculated the breakdown probability from free to synchronized flow from simulations. Davis (2006) developed a methodology for preventing transitions to synchronous flow at bottlenecks with ramps.

A common feature for studies where traffic flow has been simulated, is that the definition of probability is defined by P_B . However for empirical data, Elefteriadou et al. (2011) argues that this is not correct, due to the fact that the chance of observing a certain volume is not taken into account. Therefore, Elefteriadou et al. (2011), as well as Brilon et al. (2005), instead used the Product Limit Method (PLM) to estimate the breakdown probability. A consequence of this method is that the probability function is rarely calculated for high traffic volumes.

A majority of the presented studies on this topic use advanced statistical methods, e.g. Markov chains and queuing theory, to produce simulations and P_B to calculate the breakdown probability, or use the PLM-estimator on empirical data. Both of these approaches has limitations, and we therefore propose to use a more intuitive approach to calculate the breakdown probability based on empirical data. Compared to the simulation techniques our approach is computer effective and model based simulation is not necessary. Moreover our estimator is able to calculate probabilities also for higher volumes, which the PLM-estimator typically fail to do. In addition, as pointed out in Xu et al. (2013), the PLM-estimator seems to overestimate the capacity of the road, which does not seem to be the case using our method.

The driver behavior is dependent on surroundings, and the dependence at one detection point is not necessarily the same as for another point Ranney (1999). Several previous studies has

collected data from a limited number of sites (for instance Wang et al. (2010) and Elefteriadou et al. (2011)), whereas we consider multiple observation sites.

In the next section, we present our proposed methodology for classification of observations as either (i) free flow or (ii) congested. Then, we present the estimator based on phase transitions, and how the probability distribution function is estimated. Finally, we illustrate how this can be used to avoid breakdowns.

2 A METHOD FOR ESTIMATING THE PROBABILITY OF TRAFFIC BREAKDOWN

Our proposed methodology assumes that single vehicle registrations of speed is available from at least one where traffic breakdown is observed. An example of such a data set is given in the data section below along with a discussion of the uncertainty of such registrations. In line with previous literature (Shiomi et al., 2011; Elefteriadou et al., 2011; Xu et al., 2013), the data is aggregated to five minute intervals. That is, for each five minute interval we count the number of passing vehicles and calculate their mean speed. The shape of the dependency structure between these two parameters is well known shown in Figure 2 for two of the observation sites available in the data set presented in Section 3. From these flow-speed plots one can clearly observe a decreasing speed as

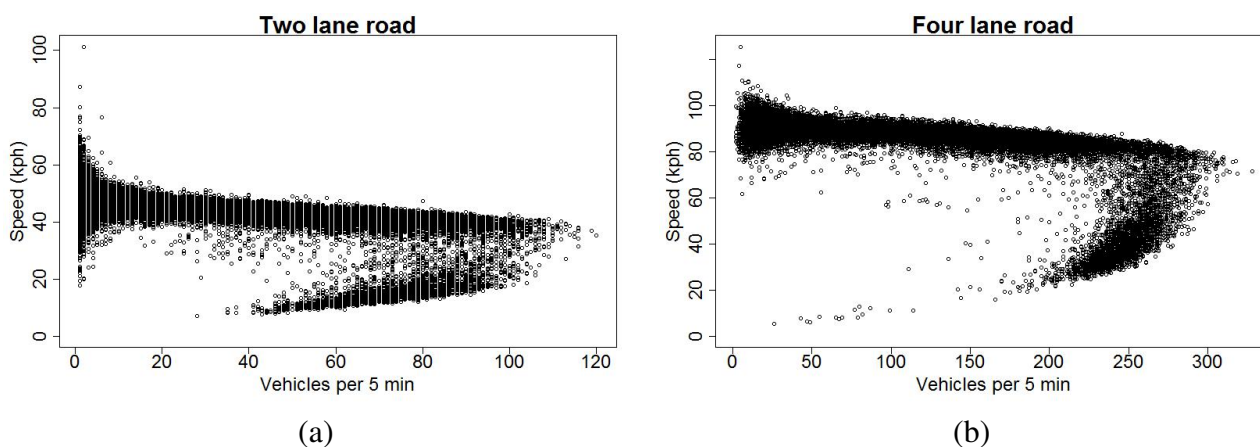


FIGURE 2 Observations of the number of vehicles and their mean speed within five minute intervals on a two lane road (a) and on a four lane road (b).

a function of increasing traffic volume, and moreover a breakdown pattern when the traffic volume becomes too large. In a state of breakdown traffic slows, and the number of vehicles passing the observation point decreases, resulting in the characteristic concave form (relative to the y-axis) in these plots. It is essential to understand the stochastic nature of this process. Firstly, the variability of the registered mean speed for a given traffic volume. Note that as each point represents the mean of the speed of passing cars the increase of variability as the traffic volume decreases towards zero is natural due to the decreasing number of cars for which the mean is calculated. And secondly, more importantly, the variability of the traffic volume that result in a breakdown. The stochastic nature of the road capacity (i.e. the maximum traffic volume that can pass a certain point on the road in a certain time frame) is often considered in the literature (Elefteriadou et al., 1995; Shiomi et al., 2011; Elefteriadou et al., 2011). However, we focus on the stochastic properties of the

amount of traffic that results in a breakdown. This motivates the estimation of a function returning the probability of a breakdown given a traffic volume.

Before presenting our proposed algorithm and probability estimator we note that there are several informative ways of representing the data shown in Figure 2. For instance the well-known relationship

$$d = V/f, \quad (2)$$

where d is the density of the traffic, V is the volume and f is the speed, can be used to consider speed as a function of density instead of volume (See Figure 3). Another example of representation

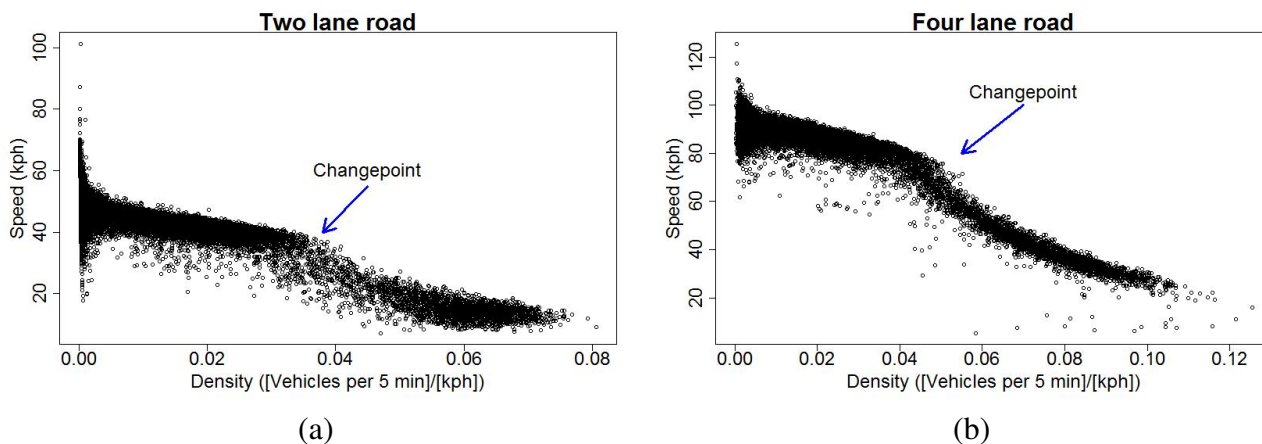


FIGURE 3 Observations of the density of vehicles and their mean speed within five minute intervals on a two lane road (a) and on a four lane road (b).

is to consider delay as a function of density, where we define delay to be the time difference between driving 1 m with the calculated mean speed and the road's speed limit. In the method presented below we use these two relationships to identify observations in a breakdown state.

Our proposed method for estimating the breakdown probability as a function of traffic volume consist of three parts

1. Classify each observation to a breakdown or non-breakdown state.
2. Identify all successive observations where the former is in non-breakdown state and the latter is in a breakdown state.
3. Estimate the breakdown probability using a defined estimator and point 1. and 2. above.

The main focus in this paper is on the estimator used in the third point in the list above. It is important to note that our estimator for the breakdown probability can be used regardless of which method classifying the observations to breakdown or non-breakdown states. Below we propose a classifying algorithm, but we stress the point that *any* algorithm that successfully divide the observations into breakdown or non-breakdown states can be used. For instance, a manually coding of each 5 minute interval will be sufficient.

Classification

In this section we propose an algorithm for classifying each observation to either be in a breakdown or a non-breakdown state. The algorithm consist of two parts. Firstly, we do a rough classification using change point analysis available through the *changepoint* -package (Killick et al., 2016) within the statistical software R (R Core Team, 2015). Secondly, we use linear regression to iteratively reclassify breakdown observations that are close to the regression line to be in a non-breakdown state. In the following these two parts will be explained in detail.

We start by performing a rough classification of each observation. This is done by applying the *changepoint*-package in R in two steps. First, we order the observed mean speeds by the observed densities and apply the *changepoint*-package in R to find the one point of density where the mean and the variance of the speed changes the most. Second, we order the observed delays by the observed densities and apply the *changepoint*-package to find the density where the mean and the variance of the delay changes the most. Recall that delay is defined to be the time difference between driving 1 m with the calculated mean speed and the road's speed limit. These two independent runs of the *changepoint*-package identifies two different points of density where there is a change in the dynamic of the traffic flow, and as a rough classification we choose the maximum of these two points to separate breakdown observations for non-breakdown observations. I.e. if an observation has a density lower than what is obtained from the changepoint analysis we classify this observation to be a non-breakdown points, and if an observation has a density that is higher we classify this observation to be a breakdown point. We stress that this is just a rough classification method that has proven helpful for us as an input to the finer classification method presented next, and a result from such a rough classification is shown in plot (a) and (b) in Figure 4. As shown within the blue circles in these two figures, multiple observations seems to be misclassified to breakdown state (red points) for high volumes. To get a better classification we perform the following iterative procedure. Using the black observations, i.e. in non-breakdown state, we estimate a regression line on the form

$$f = \beta_0 + \beta_1 V + \beta_2 V^2 + \epsilon, \quad (3)$$

where β_0 , β_1 , and β_2 are parameters to be estimated, and $\epsilon \sim N(0, \sigma^2)$ is normally distributed noise with zero mean and variance σ^2 . Using this estimated line we identify all observations classified to the breakdown state that are closer than $3\hat{\sigma}^2$, and reclassify these observations to the non-breakdown state. This last step is repeated until no change in the sets of breakdown and non-breakdown observations are observed. In our experience two to four iterations are enough for this procedure to converge. For the rough classification shown in plot (a) and (b) in Figure 4, the result from this last iterative procedure is respectively shown in plot (c) and (d) in the same figure. Now we see that the observations within the blue circles in these plots have a more correct classification, that is fewer observations with high volume and high speed are classified to breakdown state.

The classification method presented in this section is just an example of how such observations can be divided into breakdown and non-breakdown states. There exists more sophisticated and generic classification (or cluster) algorithms than the one presented here (Friedman et al. (2001)), however our procedure performs well enough for these type of data. In addition, the method is very easy to implement, and very computer effective. For instance, some of the observation sites given in the data set in Section 3 have more than 100 000 observations, and with the method presented here the classification of even these locations are completed within seconds.

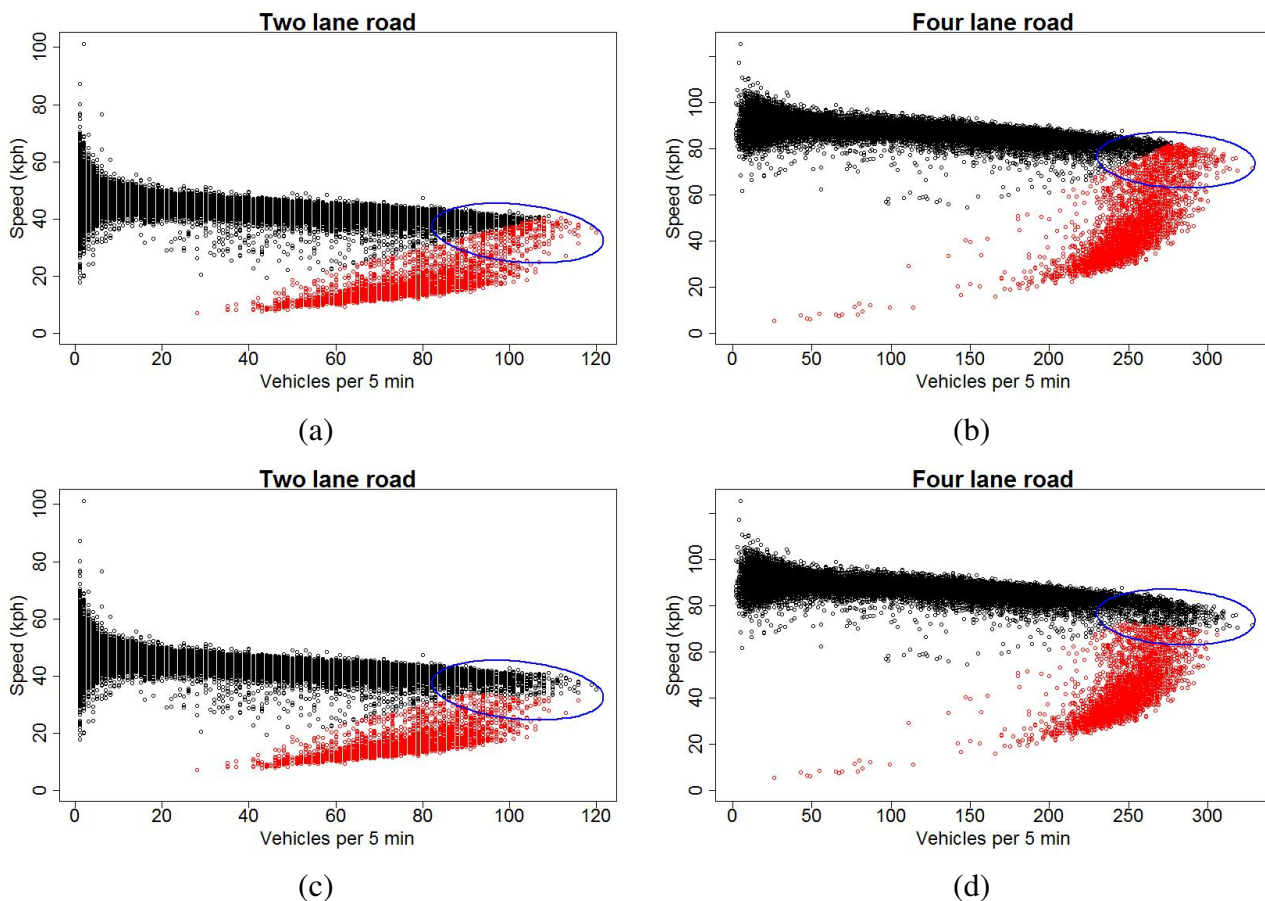


FIGURE 4 Rough ((a) and (b)) and finer ((c) and (d)) classification of observations to breakdown (red) or non-breakdown (black) states on a two lane road ((a) and (c)) and on a four lane road ((b) and (d)).

Probability estimation

When a proper classification of the observations is found we define a transition *point* to be:

Definition 1 For observations $x_t = (V_t, f_t, C_t)$, where $t = 1, \dots, T$ is an index for the successive five minute intervals and C_t is the classification of each observation to either breakdown ($C_t = 1$) or non-breakdown state ($C_t = 0$), a transition point is a traffic volume V_t where $C_t = 0$ and $C_{t+1} = 1$.

Following this definition a transition point is the traffic volume in a five minute interval classified to the non-breakdown state immediately followed by a five-minute interval classified to the breakdown state, see Figure 5 for an illustration.

The traffic volume for an observation in a breakdown state may change significantly after the transition from a non-breakdown state, as demonstrated in Figure 5. Therefore, we use the non-breakdown observation as an estimate for the traffic demand. We assume by this that the flow does not change dramatically within each five-minute observation, an assumption also applied by others in the literature (Shiomi et al., 2011; Elefteriadou et al., 2011).

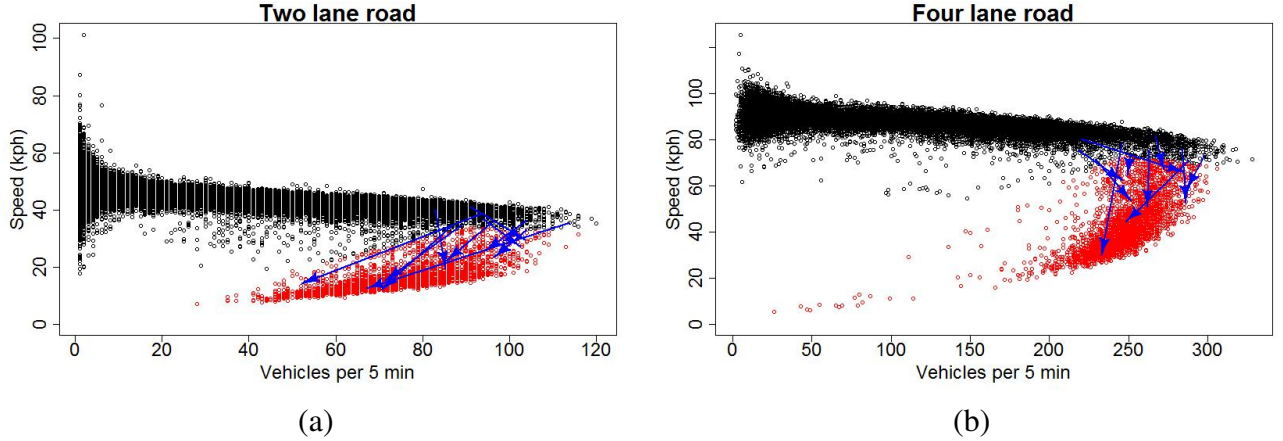


FIGURE 5 Arrows pointing from ten transition points on a two lane road (a) and on a four lane road (b).

To estimate the probability of a breakdown given a traffic volume, V , we firstly define the following functions

$$Q(V) = \sum_{t=1}^T I(V_t \leq V) I(C_t = 0) I(C_{t+1} = 1), \quad (4)$$

which simply counts the number of transition points with a traffic volume smaller than or equal to V , and

$$R(V) = \sum_{t=1}^T I(V_t \geq V) I(C_t = 0) I(C_{t+1} = 0), \quad (5)$$

which counts the number of observations classified to the non-breakdown state that is not a transition point and with a traffic volume larger than or equal to V . Lastly, we use the following estimator to calculate the breakdown probability, $p(V)$, given a traffic volume V

$$p(V) = \frac{Q(V)}{Q(V) + R(V)}. \quad (6)$$

There are several advantages using this estimator. As pointed out by Shiomi et al. (2011) there are (at least) two reasons why it is fewer observations at higher volumes. Firstly, the rush hour is limited to only a small portion of the day, resulting in more observations with lower traffic volume. Secondly, the transition point is a stochastic variable resulting in breakdowns also before the highest possible volumes are obtained. The idea of the estimator is not only to use the observations at a specific volume V , but also to use both the observations that resulted in transition points at volumes lower than V , and observations that did *not* result in transition points at volumes smaller than V . In addition, our estimator $p(V)$ is guaranteed to be zero at very small volumes, is one at very high volumes, and is monotonically increasing in between these limits. This estimator is also very fast to evaluate, avoiding (possible) computer intensive methods such as Markov chains and queuing theory. This can be very convenient in particular when handling large data sets (see Section 3 for an example). In the next section we apply our estimator to a simple example, to give an intuitive and physical interpretation of (6). This example shows that our estimator simply is the classical

statistical estimator for probability, namely the number of observations of interest divided by the total number of observation, and that our contribution is simply how we view an observation.

Simple and hypothetical example to explain estimator

Here we provide a small and hypothetical example to give an intuitive and physical interpretation of our estimator in (6). In this experiment the traffic is observed for 19 non-congested 5 minute intervals. For each 5 minute interval the number of passing vehicles are counted (the traffic volume), and it is registered whether or not the traffic goes into breakdown state in the following 5 minute interval. In other words, this corresponds to a manual classification of observations to breakdown or non-breakdown state. Let us assume that the result from such a study is the observed volumes $V_{obs} = (15, 22, 60, 50, 90, 40, 75, 35, 70, 70, 70, 10, 50, 90, 60, 10, 20, 45, 45)$. In Figure 6 these observations are plotted with \times , and they are colored green if the traffic did not go into a breakdown state and red if the traffic did go into a breakdown state. In the same figure we draw lines from the observations, indicating the following two reasonable assumptions; (I) For a traffic volume V that did not result in a breakdown state, hypothetically removing vehicles from the exact same traffic flow would not have resulted in a breakdown either, and (II) for a traffic volume V that did result in a breakdown state, hypothetically adding vehicles to the exact same traffic flow would also have resulted in traffic breakdown. We emphasize here that by exact same traffic flow we mean the exact same conditions, drivers, vehicles and in all aspects the exact same situation only with one less or one more vehicle, respectively. Thereby we add green lines going towards lower traffic volumes for non-breakdown events and red lines going towards higher volumes for breakdown events, to indicate that the observations is considered to be valid also for lower and higher volumes, respectively. To calculate the probability of breakdown for instance at the traffic volume $V_2 = 50$ (see again Figure 6), we simply count the number of horizontal red lines intercepting the vertical dotted blue line going up from V_2 and divide by the total number of horizontal lines (both green and red) intercepting the same line. That is, the probability of a breakdown at V_2 would be $P(V_2) = 3/(3 + 5) = 3/8$. Likewise for the volume V_1 and V_3 in Figure 6 the probability of a breakdown will be $P(V_1) = 0/(0 + 9) = 0$ and $P(V_3) = 6/(6 + 1) = 6/7$, respectively.

Using observations for higher volumes (for observations not resulting in breakdown state) and lower volumes (for observations resulting in breakdown state), has parallels to censoring techniques used in survival analysis, where the observations surviving the end of the experiment still are used in the resulting estimations.

3 DATA EXAMPLE: OBSERVATIONS ON THE NORWEGIAN ROAD NETWORK

In this section we present an example data set that our proposed method is applied to. Single vehicle registrations from observation sites with regular breakdowns were made available from the Norwegian Public Roads Administration (NPRA). The observation sites are located all across Norway (see Figure 7). Data is collected from roads with different speed limits, for roads with either one lane in each driving direction (26 lanes in total), or for roads with two lanes in each direction (23 in total).

Single vehicle detections are associated with large uncertainty. However, the data sets collected from such registrations are often very large, which means that the law of large number can be applied to argue that the resulting estimates are accurate. For instance, after aggregating the data example set given above to five minutes intervals, there are a total of 1 357 135 observations

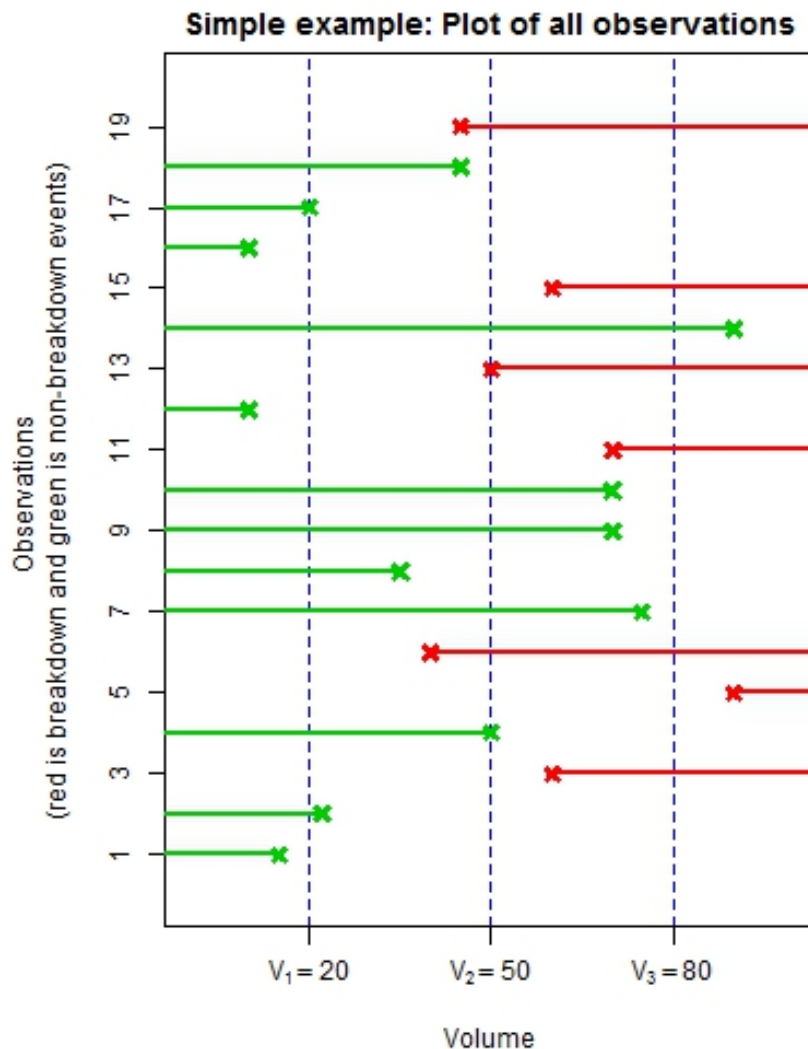


FIGURE 6 Observations from our simple example where 5 minute intervals resulting in traffic breakdown are shown in red and green otherwise.

in the two lane road case and 673 396 observations in the four lane road case.

The objective of this study is to estimate one breakdown probability curve for the two lane road case and one curve for four lane road case. With sufficient amount of data, one can also use the above method to estimate one curve for each of the road. However, note that we use the transition points in our estimator and not all the observations in a breakdown state, so the data collection from such a single observation site should be large. Also, the purpose in our work with this data set was to estimate breakdown curves to use in a transportation model for the Norwegian road network. Therefore, we need representative probability curves for all roads on the network, and thus we chose to mix several road types, speed limits etc. together.

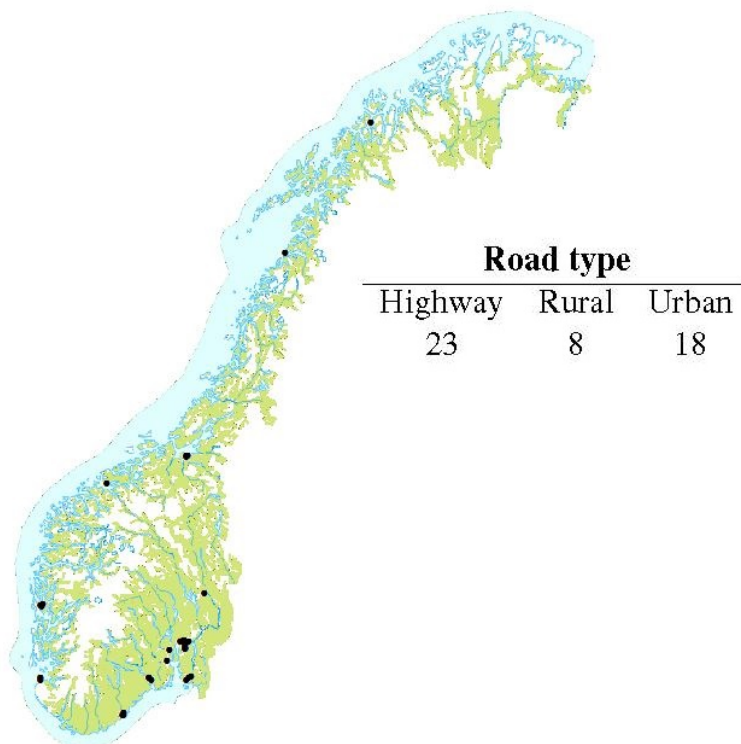


FIGURE 7 All observation sites for single vehicle detection on the Norwegian road network where breakdowns are regularly observed for the two lane and the four lane case. A table with counts of the observation sites for different road types are also shown.

4 ESTIMATED PROBABILITY CURVES FOR THE NORWEGIAN ROAD NETWORK

In this section we present our estimated breakdown probability curves for the data set presented in Section 3. We run our classification procedure for each observation site independently and use our estimator in Equation (6) jointly on all the two lanes and jointly on all the four lanes road, see plot (a) and (b) in Figure 8. The estimated breakdown probabilities form an S-shape, which is a well known shape for these types of curves. The breakdown probability increases rather quickly after 80-90 vehicles per five minutes for the two lane road case, and after 230-240 vehicles per five minutes for the four lane road case. These results seem reasonable compared to for instance Shiomi et al. (2011) and Wang et al. (2010), however less comparable to for instance Brilon et al. (2005) and Elefteriadou et al. (2011). These two latter papers evaluate both the well known non-parametric PLM estimator and a parametric approach using maximum likelihood estimation with censoring, which, similar to our estimator, both use identified transition points and non-breakdown points to estimate the probability breakdown curve. We have implemented both these two estimators where the Weibull distribution is chosen in the parametric case, which is the standard choice in the literature, see for instance Brilon et al. (2005) or Shiomi et al. (2011). The results from these estimators is shown with a grey line and a dotted grey line in plot (a) and (b) in Figure 8. Both the estimated probability curves using PLM and Weibull is located to the right of our curve. In addition, non-parametric PLM fails to estimate the breakdown probabilities higher than the last

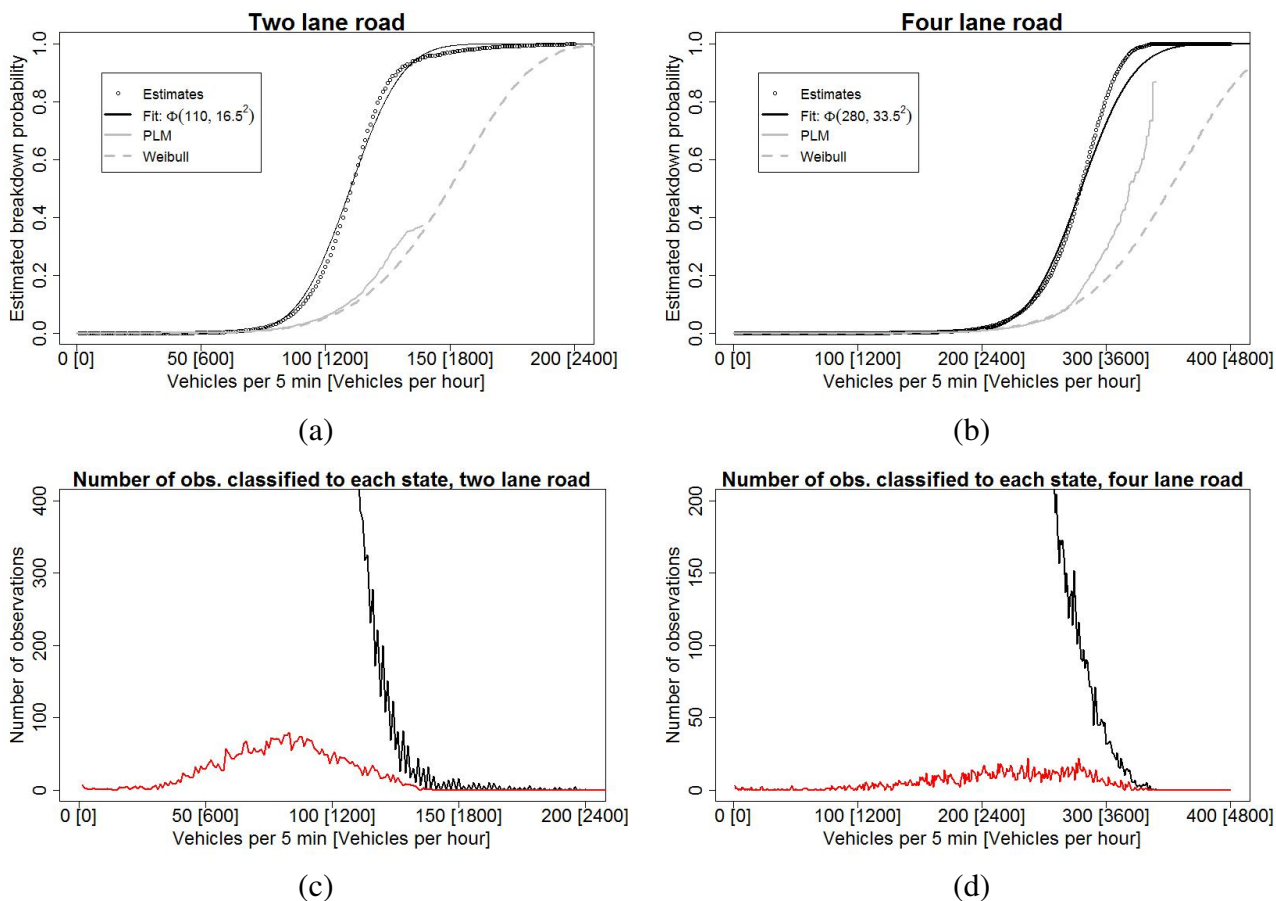


FIGURE 8 Top: Estimated breakdown probabilities (dots) for the two lane road case (left) and the four lane road case (right). Black lines estimate the breakdown probabilities, while grey lines and grey dotted lines represent the probability curve using PLM and parametric Weibull estimation, respectively. Bottom: The number of observed non-breakdown points (black line) and transition points (red) for each traffic volume in the two lane road case (left) and the four lane road case (right). Corresponding hourly flows are showed in brackets on the x-axis to make comparison to previous literature easier.

observed transition point. To investigate the difference between the three estimators we plot the number of observed data points and the number of observed transition points as a function of the observed traffic volume, see plot (c) and (d) in Figure 8. It is interesting that for the two lane road case the PLM and Weibull both estimates a breakdown probability of only 0.4 when the traffic volume is approximately 140 vehicles per five minute, although all transition points are observed before this volume. However, since some non-breakdown observations are above this volume the estimated probabilities for PLM and Weibull can only reach this level in this case. We would argue, based on the plots in Figure 8, that the breakdown probability is underestimated using the PLM and Weibull for this data set. This is an observation done by other in the literature as well, see for instance Shiomi et al. (2011). We would argue that our estimator, or really the way we interpret the data, is able to capture and model the physical phenomenon of traffic breakdown probability better than previously defined estimators. This is because we are able to include in our estimator

the assumption that if no observations above a certain traffic volume is observed, it is highly likely that the probability of breakdown at this traffic flow would be close to one. Assuming of course that the traffic flow is observed for a significant amount of time and at a point on the road network where traffic breakdown regularly occurs. See again Figure 6, for an illustration of how we extract more information from each observations by two simple and reasonable assumptions.

The estimator gives one probability estimate for each discrete value of V . To apply these estimates (for instance in a transport model), one can either use these probabilities directly, or one can fit a continuous function to the calculated probabilities. The lines shown in Figure 8 represents such a fit. In particular we have estimated a cumulative Gaussian function $\Phi(V; \mu, \sigma^2)$ to the estimated breakdown probabilities, which of course results in a completely symmetric curve around μ . In fact, the estimate for μ represents the volume $V = \mu$ where it is a 0.5 probability for the traffic going into a breakdown state. As we can see from Figure 8 the fitted curves follows the estimated breakdown probabilities closely, especially for lower values of V . The estimated probabilities seems however to increase more steeply close to one than close to zero, especially in the four lane road case. This suggests that a completely symmetric function might not be totally appropriate, however other functions can easily be used to obtain a better fit.

5 CONCLUDING REMARKS

In this paper we propose an estimator for calculating the probability for breakdown in traffic given a traffic demand. The estimator is based on classification of aggregated single vehicle registrations to either breakdown or non-breakdown state. This methodology has an appealing interpretation and is a computer effective alternative to previously suggested methods. We apply our method to a large data set consisting of more than 100 000 000 single vehicle registrations from multiple observation sites all over Norway, and are able to estimate breakdown probability curves as a function of traffic volume.

Several actions might be taken by the authorities to avoid traffic congestion, redirecting parts of the traffic flow to alternative routes, and warning travelers of the possibility of congestion. For any of these alternatives to be effective, which includes avoid taking action whenever traffic congestion is less likely, one needs to be able to predict with some certainty the probability of a traffic breakdown. The method presented here contributes in this direction as it provides an intuitive estimate for a traffic breakdown given a traffic demand.

Avoiding traffic congestion can not be achieved by limiting the traffic flow alone, but also by somehow moving the probability curve to higher volumes. For instance a traffic flow with connected and/or automated vehicles should get to higher volumes before the probability of a breakdown becomes significantly different from zero. The effect of connected or automated vehicle on traffic flow is to some degree studied through simulations in the literature (Dresner and Stone, 2004; van Arem et al., 2006; Lee and Park, 2012; Litman, 2015), and all studies suggests a more stable traffic flow. Empirical data is non-existing for such problems, but if data with information of the amount of connected or automated vehicles within the traffic flow is collected, this methodology can be used to study the effect of such intelligent transport systems (ITS) on the breakdown probability curve.

In addition to what is mentioned above, several options for further work can be pursued from these results. This methodology is a first step in the direction of predicting breakdown events. The development of an ITS service predicting traffic congestion on a day-to-day basis could apply

the results presented in this paper. However issues like what is an acceptable breakdown probability and what is the best actions to take in a state of high breakdown probability needs to be properly addressed for such an ITS service. When traffic flow increases the influence of other traffic on a single vehicle increases (Ranney, 1999). This theory leads to the hypothesis that for instance the amount of heavy vehicles might change the probability curve for traffic breakdown. As oppose to the case with connected or automated vehicles, data are available for the amount of heavy vehicles within the traffic flow so this hypothesis could be tested. This is currently work in progress.

6 ACKNOWLEDGMENTS

We thank the NPRA for funding the project for this paper, and for providing the data sets.

REFERENCES

- Brilon, W., Geistefeldt, J., and Regler, M. (2005). Reliability of freeway traffic flow: a stochastic concept of capacity. In *Proceedings of the 16th International symposium on transportation and traffic theory*, volume 125143. College Park Maryland.
- Chen, X. M., Li, Z., Li, L., and Shi, Q. (2014). A traffic breakdown model based on queueing theory. *Networks and Spatial Economics*, 14(3-4):485–504.
- Davis, L. (2006). Controlling traffic flow near the transition to the synchronous flow phase. *Physica A: Statistical Mechanics and its Applications*, 368(2):541–550.
- Dresner, K. and Stone, P. (2004). Multiagent traffic management: A reservation-based intersection control mechanism. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 530–537. IEEE Computer Society.
- Elefteriadou, L., Kondyli, A., Washburn, S., Brilon, W., Lohoff, J., Jacobson, L., Hall, F., and Persaud, B. (2011). Proactive ramp management under the threat of freeway-flow breakdown. *Procedia-Social and Behavioral Sciences*, 16:4–14.
- Elefteriadou, L., Roess, R. P., and McShane, W. R. (1995). Probabilistic nature of breakdown at freeway merge junctions. *Transportation Research Record*, (1484):80–89.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.
- Kerner, B. S. (2009). *Introduction to modern traffic flow theory and control: the long road to three-phase traffic theory*. Springer Science & Business Media.
- Kerner, B. S., Klenov, S. L., and Wolf, D. E. (2002). Cellular automata approach to three-phase traffic theory. *Journal of Physics A: Mathematical and General*, 35(47):9971.
- Killick, R., Haynes, K., and Eckley, I. A. (2016). *changeoint: An R package for changepoint analysis*. R package version 2.2.1.
- Lee, J. and Park, B. (2012). Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment. *IEEE Transactions on Intelligent Transportation Systems*, 13(1):81–90.
- Litman, T. (2015). Autonomous vehicle implementation predictions: Implications for transport planning. In *Transportation Research Board 94th Annual Meeting*, number 15-3326.
- Lorenz, M. and Elefteriadou, L. (2000). A probabilistic approach to defining freeway capacity and breakdown. In *Proceedings of the 4th International Symposium on Highway Capacity*, pages 84–95. Transportation Research Board Washington, DC, USA.
- May, A. (1990). *Traffic flow fundamentals*. Prentice Hall.

- Persaud, B., Yagar, S., and Brownlee, R. (1998). Exploration of the breakdown phenomenon in freeway traffic. *Transportation Research Record: Journal of the Transportation Research Board*, (1634):64–69.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ranney, T. A. (1999). Psychological factors that influence car-following and car-following model development. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4):213 – 219.
- Shiomi, Y., Yoshii, T., and Kitamura, R. (2011). Platoon-based traffic flow model for estimating breakdown probability at single-lane expressway bottlenecks. *Transportation Research Part B: Methodological*, 45(9):1314–1330.
- Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., ichi Tadaki, S., and Yukawa, S. (2008). Traffic jams without bottlenecks? experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics*, 10(3):033001.
- van Arem, B., van Driel, C. J. G., and Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics.
- Wang, H., Rudy, K., Li, J., and Ni, D. (2010). Calculation of traffic flow breakdown probability to optimize link throughput. *Applied Mathematical Modelling*, 34(11):3376–3389.
- Xu, T.-d., Hao, Y., Peng, Z.-r., and Sun, L.-j. (2013). Modeling probabilistic traffic breakdown on congested freeway flow. *Canadian Journal of Civil Engineering*, 40(10):999–1008.