Estimating the Safety Effects of Congestion Warning Systems using Carriageway Aggregate Data

van Lint, Hans; Nguyen, Tin Thien; Krishnakumari, Panchamy; Calvert, Simeon C.; Schuurman, Henk; Schreuder, Marco

DOI
10.1177/0361198120945319

Publication date
2020

Document Version
Final published version

Published in
Transportation Research Record

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
2020, Vol. 2674(11) 278–288

DOI: 10.1177/0361198120945319
journals.sagepub.com/home/trr
SAGE

Estimating the Safety Effects of Congestion Warning Systems using Carriageway Aggregate Data

Hans van Lint[1], Tin Thien Nguyen[1], Panchamy Krishnakumari[1], Simeon. C. Calvert[1], Henk Schuurman[2], and Marco Schreuder[2]

Abstract

Is it possible to use *just* aggregate carriageway data for the evaluation of congestion warning systems (CWS) in large networks—or any system affecting traffic safety for that matter? In this paper, two hypotheses related to this question are tested. The first hypothesis is that it can be done by comparing large-scale congestion patterns on road stretches with and without CWS. The underlying rationale is that heterogeneous congestion patterns with many disturbances, frequent wide moving jams, and large speed differences result in more potentially unsafe traffic conditions than more homogeneous congestion patterns. The second hypothesis is that it is possible to compare differences in average (maximum) deceleration distributions into congestion waves between road stretches with and without CWS. Both hypotheses have been tested for similar bottlenecks with similar demand patterns and the results suggest the first hypothesis must be rejected. Although the idea seems plausible (CWS result in more homogeneous congestion patterns) there were too many confounding factors in the data to make the case. However, persuasive evidence was found for the second hypothesis. Statistically significant differences were found between (maximum) deceleration distributions on road stretches with and without CWS that suggest CWS do—as expected—contribute positively to traffic safety. It thus seems to be possible to monitor safety effects using just average speeds. However, the method is limited to providing relative comparisons. Furthermore, to fully rule out the effects of unobserved factors, more evidence and validation with microscopic data are needed.

Congestion warning systems (CWS) belong to the oldest, and still successful, intelligent transportation systems (ITS) applied on freeway networks. For example, in the Netherlands, the so-called MTM (motorway management and signaling) system was installed on large parts of the Dutch freeway network in the mid-80s and is still considered one of the important contributors to the (dramatic) decrease in rear-end collisions on freeways since then—positive results that are also reported elsewhere (*1*). The principle idea is simple: use data from sensors (e.g., induction loops) to detect congestion downstream, and, if this is the case, display a warning pictogram or reduced advisory speed limit on variable message signs on gantries, to warn upstream traffic of the (backward moving) congestion wave ahead. Over time, these systems have been improved with more accurate and lane-specific loop detectors; and more diversified lane-specific messages, but the essence has not changed. What has changed is that in some cases these CWS (both sensors and variable message signs) are also used for different purposes, for example, for speed homogenization or proactive congestion management with dynamic speed limits (*2*, *3*). With the arrival of alternative sensing systems (other than loops), and the possibility of using the vehicle itself as both sensor and actuator, one of the key questions for road operators today is whether continuous investments in infrastructure-based CWS is a cost-effective strategy for the near-term (and longer term) future. This policy question forms the background of this study.

Cost-effectivity in the context of this paper relates to how well current infrastructure-based CWS perform in relation to preserving/increasing traffic safety in comparison with possible alternative CWS that make use of alternative sensing and actuation, particularly in-vehicle.

[1]Faculty of Civil Engineering and Geosciences, Department of Transport & Planning, Delft University of Technology, Netherlands
[2]Rijkswaterstaat Water, Verkeer en Leefomgeving, (Executive arm Ministry Infrastructure & Water Management), Netherlands

Corresponding Author:
Hans van Lint, J.W.C.vanLint@tudelft.nl

The prospect of increasing percentages of connected (and automated) vehicle automation opens up many new possibilities for traffic management in general. Whereas much research attention around that theme goes to understanding and evaluating the effects of dynamic speed limits on throughput and congestion, much less attention goes to evaluating the effects of these systems on traffic safety under such changing conditions (*3–6*). For example, what is the optimal mix of in-car and infrastructure-based sensing and actuation for congestion warning during a transition to (eventually) 100% connected (and possibly automated) vehicles in relation to potentially prevented accidents? The key methodological difficulty in answering this question is that, despite the fact that traffic is the number eight cause of death globally, and even number one for people under 30, traffic accidents are extremely rare events relative to the total amount of traffic interactions (*7*). For example, the accident rate in the Netherlands is in the order of 1 in 10^6 km driven, and the fatality rate is about 4 in 10^9 km driven (*8*). Since CWS support drivers in their perception of congestion downstream, the key of successful CWS lies in detecting this congestion. The requirements for CWS thus boil down to minimizing type I errors (detecting congestion when there is none) and time-to-detection (congestion); under the constraint of *zero* type II errors (not detecting congestion when it is present). Algorithms with densely placed loop detectors are fairly good at doing this job, so for alternative CWS systems typical research questions include: Which factors govern the robustness of these alternative solutions? How sensitive are these new combinations of (vehicle- and infrastructure-based) sensing, algorithms, and actuation for CWS to errors in the data, in the communication (protocols)? What role is played by errors in human perception and response when bringing CWS (partially) in-vehicle?

These "what if" type research questions relate to the ex-ante assessment of CWS. There are also ex-post research questions, which are at least as relevant as ex-ante questions and suffer from similar methodological problems. Since accidents are rare, accident statistics alone may not give an unbiased assessment of the efficacy of CWS, particularly without contextual data that depict causal mechanisms. For example, it is difficult to find strong empirical (behavioral) evidence in field data that a head-tail collision would have been preventable with a CWS, or vice versa, that an accident on a road with CWS would have also happened without CWS. There are many contextual factors that come into play here. For example, maybe the driver did not see the braking lights downstream because they were distracted by something inside or outside the vehicle regardless of CWS. Maybe they took unacceptable risks. Or maybe other factors (geometry, traffic mix, etc.) played a role in a particular case. Accident records rarely contain sufficiently reliable information to warrant such inference, let alone allow conclusions to be drawn from it.

The alternative for accident statistics are surrogate safety measures as indicators (proxies) for traffic (un)safety and thus of the efficacy of CWS (*9–12*). However, to compute most surrogate measures, individual speeds and headways are required from either individual vehicle passages at a local detector (e.g., a loop) or vehicle trajectories. This makes ex-post evaluation of CWS over entire networks expensive and impractical at the very least. An ideal scenario would be to derive alternative surrogate safety measures from data that are readily available on the entire road network, for example, flows and speeds from loops or other infra-based sensors, or even average speeds fused from a multitude of sources, including probe vehicle data. This would allow network-wide monitoring and assessment of CWS, and this is the objective of this paper. It discusses two such methods and is outlined as follows.

First, in the next section, the two hypotheses underneath these methods and the associated assessment frameworks to test these are formulated and discussed. Next, the data used and some of the issues related to filtering and processing are briefly overviewed. Then, the results from applying the assessment frameworks are presented in two separate sections (one for each hypothesis), in which relevant methodological detail (e.g., clustering and statistical testing), and, of course, the results are discussed. The paper closes with a synthesis and critical discussion of the findings and conclusions.

Hypotheses and Assessment Frameworks

The a priori reason why deriving surrogate safety measures from carriageway aggregate data for the assessment of CWS seems feasible is that CWS inform drivers about (unsafe) downstream drops in speed. CWS thus specifically give information about, and affect, *longitudinal* movement (speed choice and acceleration). These longitudinal speed dynamics are captured well by carriageway average speeds. Particularly around capacity, drivers are largely constrained in choosing their desired speed because a significant fraction of drivers are car following. The assumption is therefore that carriageway average speed dynamics are informative of the potential unsafe traffic conditions individual drivers may experience. More precisely, it is assumed that average space-time contour plots of speeds and the derived acceleration patterns from these contour plots are informative of potential unsafe traffic conditions.

Before discussing which two hypotheses will be tested in this paper, two remarks are made. First, both hypotheses are tested against the null hypothesis (H0) that

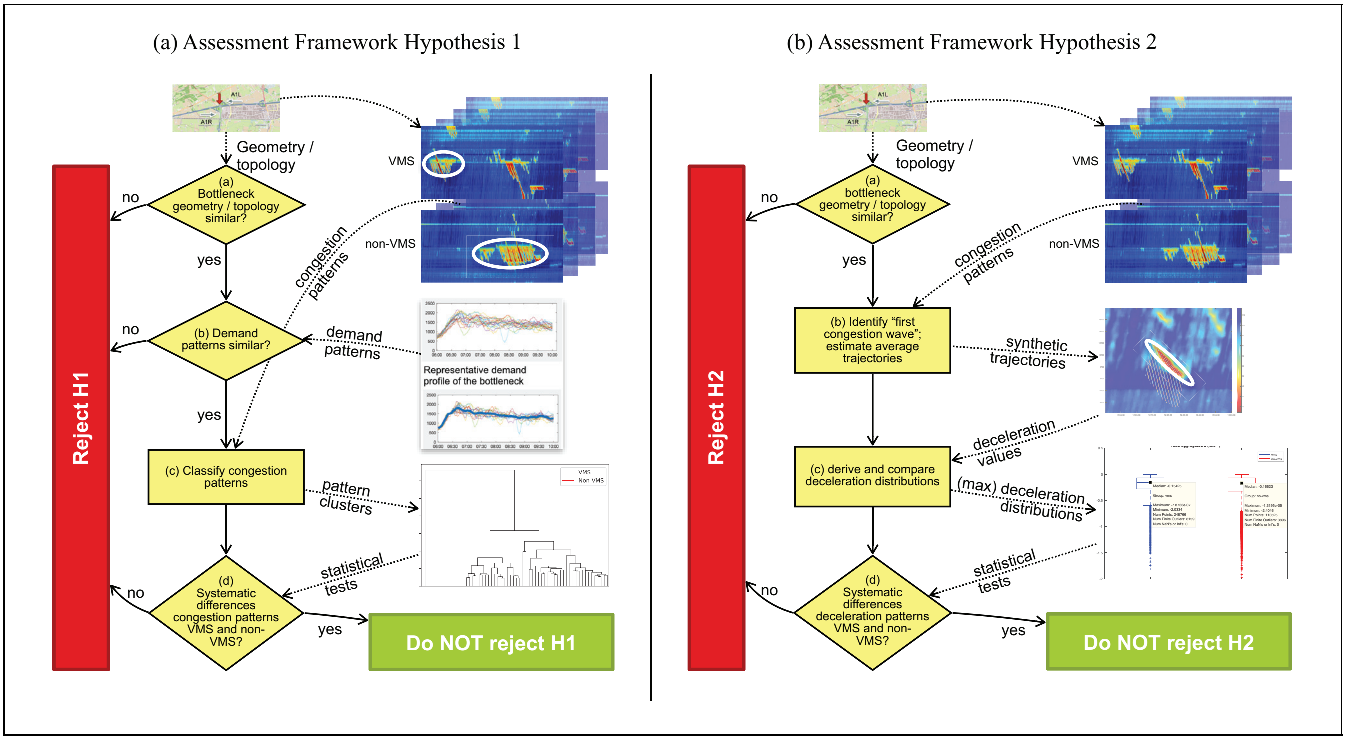

Figure 1. Assessment frameworks for: (*a*) Hypothesis 1 (Safety effects of congestion warning systems [CWS] can be assessed by comparing large-scale congestion patterns at comparable regular bottlenecks) and (*b*) Hypothesis 2 (It is possible to assess safety effects of CWS by comparing distributions of [maximum] average decelerations at regular bottlenecks).
Note: VMS = variable message signs.

differences between either congestion patterns or deceleration distributions are because of other (non-observed) factors than the presence of a CWS. However, claiming that this null hypothesis can be rejected completely will be refrained from. In the authors' view the best that can be done is making it plausible that the presence of CWS is an important contributing factor, that is, NOT rejecting the alternative hypotheses. Second, since the CWS considered in this paper convey information using variable message signs (VMS), the terms VMS and CWS are used interchangeably.

Hypothesis 1 (H1): It is possible to assess safety effects of CWS by comparing large-scale congestion patterns at comparable regular bottlenecks under comparable demand patterns.

The underlying rationale of this hypothesis is that, in heterogeneous congestion patterns with many disturbances—frequent wide moving jams and, thus, large speed differences—more potentially unsafe vehicle interactions take place than in homogeneous congestion patterns with very few or no disturbances and thus smaller speed differences. In case it is found that congestion patterns on non-CWS road stretches are systematically more heterogeneous than patterns on CWS stretches, this provides indirect (corroborating) evidence that CWS may result in safer traffic conditions.

Figure 1*a* sketches the assessment framework to test this hypothesis. The key of this framework is to derive and classify (label) congestion patterns such that it is possible to statistically test systematic differences between the labels (i.e., type of congestion patterns) occurring at stretches with and without CWS. So, given two road stretches (one equipped with CWS and one without), the logical flow is as follows:

(a) First assess whether the geometry and topology of both regular bottlenecks (from which the congestion patterns emanate) are sufficiently similar. In this case, this means (i) the same number of lanes up- and down-stream, (ii) the same bottleneck type (e.g., on ramp, off ramp, or weaving section), with (iii) by and large similar dimensions (length merging sections, etc.)
(b) Then construct congestion patterns using the adaptive smoothing method (ASM) (see below) and group these according to similarity in demand patterns (*13*, *14*). The idea is that congestion patterns can only be compared if the underlying demand pattern is similar; otherwise, differences

in congestion patterns can be attributed just as well to differences in demand patterns.

(c) If sufficient pairs of congestion patterns are found on similar bottlenecks with similar demand patterns, these congestion patterns are now classified using the method proposed by Nguyen et al. (*15*). This method automatically derives highly compact feature vectors from congestion patterns that include amongst other things the number of wide-moving jams, and spatiotemporal extent. Using these small but informative feature vectors it is possible to build a cluster-tree of the patterns.

(d) With each pattern clustered it is possible to use statistical tests to determine whether the feature vectors that represent the congestion patterns for the CWS and non-CWS patterns systematically differ.

Hypothesis 2 (H2): It is possible to assess safety effects of CWS by comparing distributions of (maximum) average decelerations at regular bottlenecks.

The underlying rationale of this hypothesis is similar as for hypothesis 1: higher average (maximum) decelerations are indicative of potentially unsafe vehicle interactions. If systematic differences in (maximum) deceleration distributions are found between CWS and non-CWS road stretches in this direction (i.e., larger [maximum] decelerations), this provides indirect (corroborating) evidence that CWS may result in safer traffic conditions.

Figure 1*b* sketches the assessment framework to test this hypothesis. The key of this framework is to derive these deceleration distributions such that systematic differences can be statistically tested. Given two road stretches, one equipped with CWS and one without, the logical flow is as follows:

(a) Again, first assessed is whether the geometry and topology of both bottlenecks from which congestion emanates are sufficiently similar (see above).

(b) Then, instead of looking at feature vectors of entire congestion patterns, deceleration patterns into the tails of congestion waves emanating from the bottlenecks are looked at. The criterion for selecting such waves is that there is a sufficient amount of free-flowing traffic upstream of it, therefore the term "first" congestion waves in Figure 1*b*. Average (longitudinal) vehicle trajectories $s_i(x,t)$ are then estimated using the so-called Filtered-Speed-Based (FSB) travel time estimation method into this congestion wave (*16*). This reduces to drawing (synthetic) trajectories through area in the speed contour plot upstream of the wave. The slope of this trajectory ds_i/dt at point (t,x) equals the speed $v(t,x)$ in the contour plot at point $\{t,x\}$. From these trajectories, only those parts are then selected in which each synthetic "vehicle"i decelerates $(a_i(t,x) \leq 0)$, and partition this sub trajectory into segments s_{ij} with fixed length Δx of around 10 m.

(c) For each resulting trajectory segment s_{ij} an average deceleration value $d_{ij} = |a_{ij}|$ is computed. For each trajectory i a maximum deceleration $d_i^{\max} = \max_j(d_{ij})$ is also identified. These are then used to construct two empirical distributions, constituted by all decelerations, and the maximum decelerations, respectively.

(d) It is now possible to use statistical tests to assess whether systematic differences between distributions from CWS and non-CWS road stretches are observed. It is important to note that, since smoothed speeds are used (ASM computes a weighted average speed on each point (t,x)), distributions of smoothed decelerations are compared. Discussed below is whether this has consequences for the statistical tests (the hypothesis is that it does not).

Data and Processing

The data for this study come from the national data warehouse of traffic information (NDW) (www.ndw.nu), which are continuously collected and processed in the DiTTLAB cloud environment. In this environment there are over 7 years of speeds and flows for the entire Dutch freeway network and the main motorway network (about 1 TB raw data per year), including a detailed graph model to combine data from different sources, and a range of processing and filtering services for these data.

Data

For this research, 3 months of data in 2017 were selected. A challenge is that road stretches without VMS signaling (so without CWS), have a far less densely spaced set of loop detectors. The distance between loop detectors on those stretches is between 1000 and 2500 m in contrast to the 400 to 600 m loop spacing on most signaled road stretches. In some cases, no loop detectors are available at all. Figure 2 (*right*) shows a map with the freeway network in the Netherlands, in which thin lines indicate road stretches without signaling; thick blue lines road stretches with signaling; and thick semi-transparent lines those road stretches selected for this project—Figure 2 (*left*) zooms in on one of these.

To remedy this problem (fewer loop detectors on non-CWS road stretches), two data sources are used:

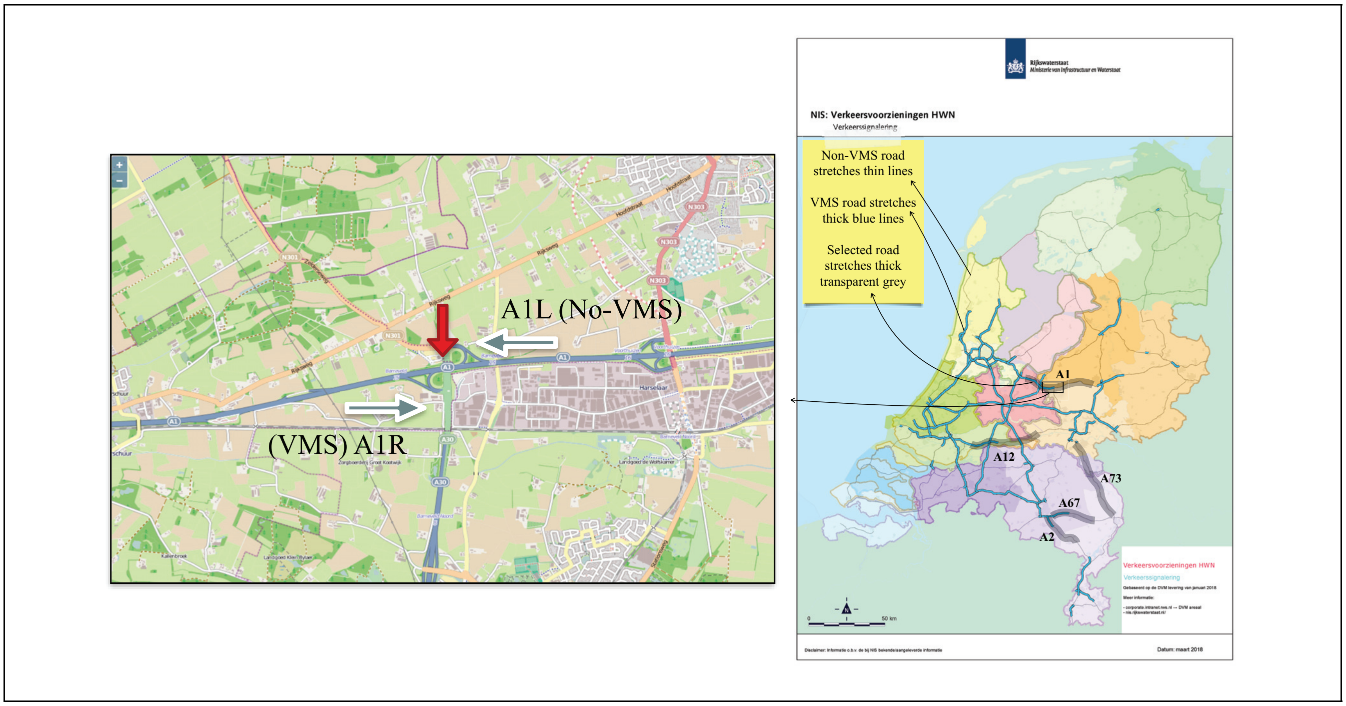

Figure 2. Example road stretches with and without congestion warning systems (CWS) (*left*), and map of CWS on Dutch freeways (*right*) annotated with all road stretches selected for this research.
Note: VMS = variable message signs.

- Loop detectors measuring minute averaged speeds and flows solely to estimate demand patterns into the bottleneck, which are used to test hypothesis 1.
- Floating car data (FCD) in the form of average speeds over 50-m segments to construct the speed contour maps needed to test both hypotheses. The underlying data are sampled vehicle trajectories (i.e., records of vehicle location, time, and speed) sourced from a popular Dutch route information and navigation app (FlitsMeister), which are (in this aggregated form) made available by NDW for the entire Dutch road network.

A potential complication for this study is that the investment decision for (or against) installing CWS on particular road stretches over the years is/was closely related to the frequency with which congestion occurred on these road stretches. Although in the North, East, and South-West part of the network congestion *does* occur, the duration and severity are indeed usually less than in the West, central, and mid-South part. As illustrated in Figure 2 (*right*), most CWS are located in the more congested Western part of the Netherlands. This may bias the outcomes of testing hypothesis 1. To make sure that in comparing congestion patterns (H1) and deceleration patterns (H2) it is possible to attribute the differences to whether or not a CWS is installed, the following rationale is used:

For hypothesis 1, entire congestion patterns are compared. Roughly speaking, two categories of factors determine the shape and characteristics of these congestion patterns: (1) the demand pattern over time into the bottleneck; and (2) the supply characteristics of the bottleneck, which are assumed to be largely determined by the bottleneck topology/geometry. A fair comparison between congestion patterns on CWS and non-CWS road stretches thus requires that both demand and supply characteristics on those road stretches are similar. For hypothesis 2, deceleration patterns are compared into a single congested wave. In this case, it is argued that, for a fair comparison, the larger scale demand pattern over time does not matter much, because only very short time periods are looked at, during which platoons of vehicles drive into the congestion wave. All drivers upstream MUST decelerate, and the hypothesis is that the presence of CWS affects how they do this. The assumption is that sufficient similarity in (the severity of) these waves may be expected in case the topologies of bottlenecks are similar.

In summary, the CWS and non-CWS cases are compared, such that if differences are found in congestion patterns (H1) and deceleration patterns (H2), respectively, these differences can be attributed to the presence of a CWS on one of the two road stretches. To this end, topological similarity (both hypotheses) and similarity in demand patterns (H1 only) are required.

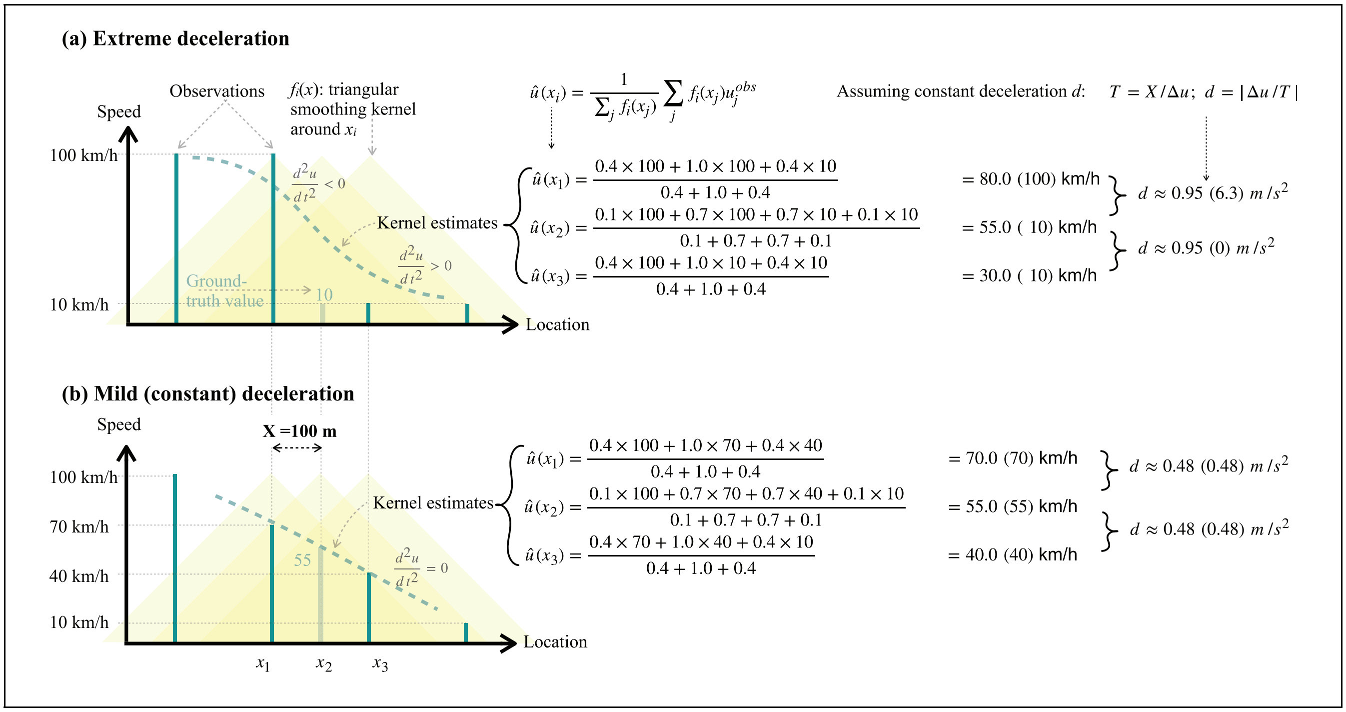

Figure 3. Two cases of decelerations that demonstrate a systematic estimation bias by kernel smoothers.

Data Smoothing and Consequences for Assessment Methods

All selected data are smoothed using the ASM (*13*, *14*). The ASM computes average speed $u(t,x)$ as a weighted average of speed observations $u^{\text{obs}}(t_k, x_d)$ around the point $\{t,x\}$. In this averaging process, two types of weight are used. The first weight, which can be computed by any reasonable kernel centered around $\{t,x\}$, is inversely proportional to the distance in space $|x_d - x|$ and time $|t_k - t|$ of each observation to the estimation point. By skewing the axes along the two dominant wave speeds, that is, at around –20 km/h in congestion, and slightly below average vehicle speed otherwise, the ASM preserves wave patterns in the traffic data, which would be destroyed by averaging orthogonally, that is, between locations or time instants. This process results in two weighted averages: one for congestion and one for free-flowing traffic, respectively—the difference between the two relates to which data points are weighed more heavily. To obtain one final weighted average, the two estimates are combined using a *second* (non-linearly computed) weight based on a principle that informally can be understood as "congestion wins." This principle follows the intuition that—at least on freeways—it is more likely that congestion waves that *seem* connected, are indeed connected (propagate backwards), rather than that they contain "holes," that is, small road stretches with freely flowing traffic over very short time intervals. The traffic dynamics of the latter case are not impossible, but in most cases much less plausible. This "congestion wins" principle is implemented in the ASM by taking the minimum of the two speed estimates to determine this second non-linear weight. The ASM is powerful in that it accurately reconstructs speed contour maps even if up to 50% of the observations is missing (*13*). However, the ASM is a low pass filter. This means that extreme changes in speeds (accelerations and decelerations) are smoothed out. These extremes, however, are precisely needed for the assessment based on hypothesis 2.

The example case in Figure 3 supports the intuition that, nonetheless, the assessment in hypothesis 2 is valid because the "error" because of smoothing is a *systematic* bias—it always has the same (predictable) consequences for the resulting statistics. For illustration purposes, in Figure 3 a single linear kernel is used to smooth the speed along the trajectory of a vehicle, so that some back-of-the-envelope computation is possible. The rationale is, however, the same as with reconstructing vehicle trajectories on a speed contour map with just the free-flow kernel of the ASM. This point is returned to below. For clarity in this section, deceleration is considered as a positive number ($d(t) = |a(t)|, a(t) \leqslant 0$), since for hypothesis 2 trajectory segments are considered in which vehicles decelerate *only*. First observe that smoothing speeds $u(t)$ result in both over- and under-estimations of the real decelerations $d(t) = du/dt$. Figure 3*a* illustrates an extreme deceleration (emergency stop) and shows that over- and under-estimation of deceleration directly relates to the sign of the rate of change in deceleration (jerk $j(t) = d^2u/dt^2$) along the vehicle trajectory. The error induced by smoothing results in *underestimation* of deceleration for $j(t) < 0$; and in *overestimation* of decelerations for $j(t) > 0$. In case $j(t) = 0$, as in Figure 3*b*, the reconstruction is exact. Since the smoothing kernel computes a weighted average, it effectively distributes the deceleration pulse over a larger part of the vehicle trajectory resulting in approximately symmetrical positive and negative "errors" around this pulse. As long as the differences in magnitude of deceleration pulses (congestion waves) are still observable in the smoothed data, *which for normal parameter settings of the ASM is always the case*, the statistics of smoothed decelerations (mean and variance) will be proportional to the statistics of the underlying raw decelerations.

Note that the ASM is much better suited for smoothing vehicle trajectories and preserving their characteristics than the linear filter in Figure 3. In the ASM, as speeds fall below the threshold between free-flowing and congested traffic, most weight will be assigned to the congested kernel along the dominant congested wave speed, which means that *in congestion most of the smoothing occurs perpendicular to* (*and not along*) *vehicle trajectories*. In the deceleration zones (into shockwaves) the smoothing is adaptive (hence the name ASM), in that it increasingly takes place along this dominant congested wave speed.

In summary, if statistical differences are found in the smoothed data, it can be guaranteed that these differences coincide with larger differences *of the same sign* in the underlying (not observable) real deceleration values. In Figure 3*a*, the computed decelerations (on the right) over the deceleration pulse equal $0.95\,\mathrm{m/s^2}$, which is considerably smaller than during the underlying emergency deceleration ($6.3\,\mathrm{m/s^2}$), but still nearly twice as large as the smoothed values in the "mild deceleration case" in Figure 3*b*, where the smoothed decelerations are in fact equal to the underlying value of $0.48\,\mathrm{m/s^2}$. In the ASM expect smaller differences can be expected between filtered and (underlying) real data—the results below will confirm this.

Clustering and Comparing Congestion Patterns

In this section, the steps in the assessment framework in Figure 1*a* are followed. In the first section, the first two (finding comparable bottlenecks and demand patterns) are discussed, and in the second section, the last two steps (comparing congestion patterns and statistical testing) are discussed.

Finding Comparable Bottlenecks with Comparable Demand Patterns

The criteria for comparable bottlenecks with and without CWS/VMS are *qualitative*: geometries are looked for with similar discontinuities (on ramp, off ramp, lane drop, weaving section) with approximately similar dimensions; traffic rules (e.g., maximum speeds); and the same number of lanes up- and down-stream of the bottleneck. As shown in Figure 2 (*right*), a limited number of road stretches have been selected with multiple bottlenecks that could be compared by hand. The result is eight sets of two comparable bottlenecks, one with and one without CWS. Note that two non-CWS bottlenecks (A73L and R) were also considered, because the Dutch highway authority (Rijkswaterstaat, RWS) was interested in the deceleration statistics there too—these are, however, not considered in the statistical tests since there is no comparison to make. Qualitative similarities do not guarantee that different congestion patterns may arise just because one has CWS and the other does not. Clearly, there are many other factors that may explain such differences. A priori the most important factors seem to be the demand pattern into the bottleneck, and the capacity distribution—emerging from collective driving behavior—at the bottleneck. Because of the limited project duration, it was necessary to make the (debatable) assumption that the capacity distributions of bottlenecks are sufficiently "covered" by similarities in bottlenecks. The assumption is debatable, because also differences in traffic mix may lead to systematic differences in driving behavior and thus influence differences in congestion patterns. This point is returned to further below.

The criterion for comparable demand patterns is *quantitative*. For the eight sets of comparable bottlenecks, for each congestion pattern the first loop detector far enough upstream of the bottleneck was selected so that the congestion waves do not spill over this detector location (see Figure 4*c*). From these loop detectors the inflow pattern (demand pattern into this bottleneck) was collected for every congestion pattern. For each bottleneck average daily patterns were constructed (see Figure 4*a* for an example). Average patterns were constructed, rather than individual patterns, at the bottlenecks to allow for some margin of dissimilarity between individual days—without this step no clear patterns were discernable. Finally, clustering was used to construct the dendrogram, which is a tree that visualizes the numerical distance (pat_1, pat_2), where pat_X is the time series of average (weekday) demands (see Figure 4*b*). In Figure 4*b* an example of three patterns A, B, and C is highlighted, in which A and C are more dissimilar than A and B. Two distance metrics D were tested to construct this

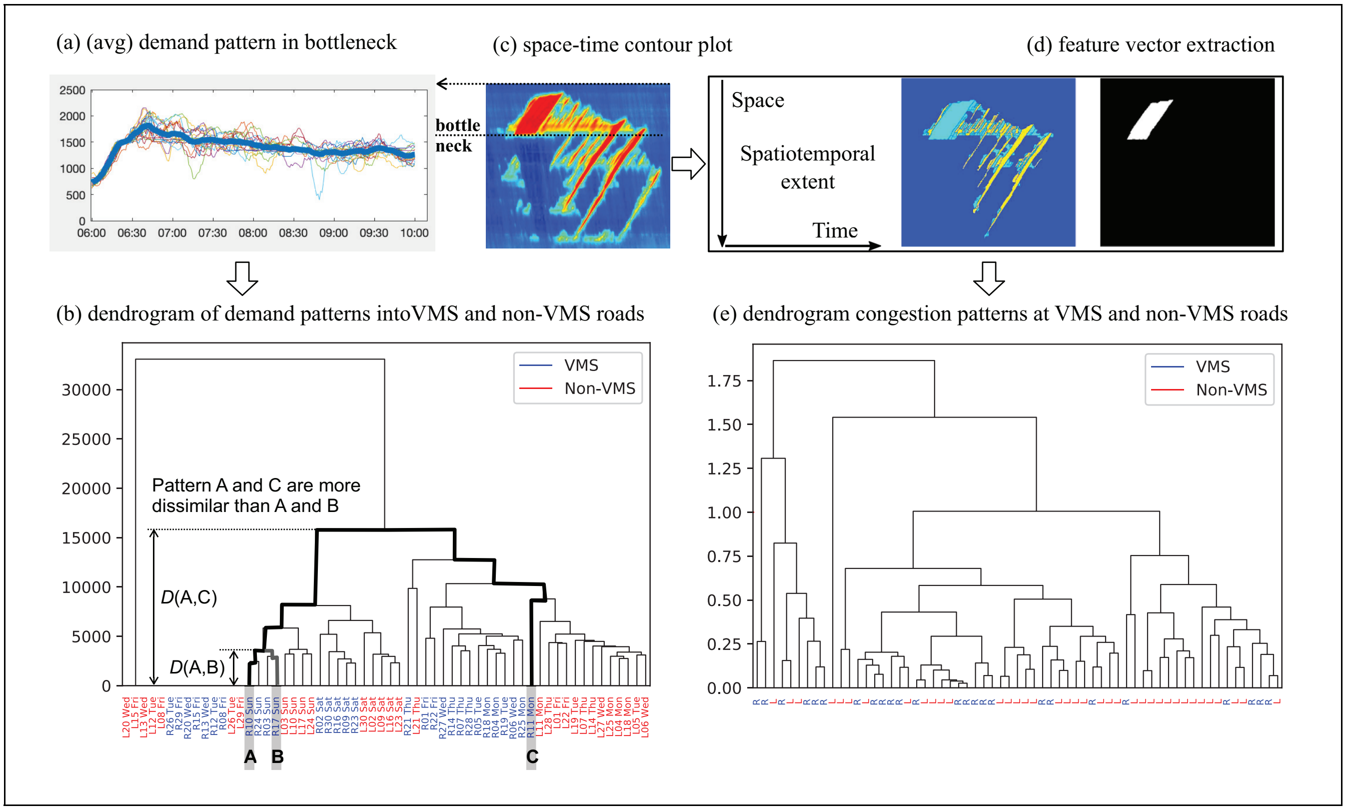

Figure 4. Methods and clustering results for hypothesis 1. (For an explanation of the subfigures see running text.)
Note: VMS = variable message signs.

dendrogram: the Euclidean distance between the average demand time series, and the Euclidean distance between the same time series aligned with a dynamic time warping (DTW) algorithm. Both gave very similar results, so the simplest one (without time-warping) was used.

Using the demand dendrogram Figure 4*b*, pairs of similar bottlenecks that have the closest demand patterns are selected and the feature extraction method in Nguyen et al. is applied to the associated congestion patterns of each pair (*15*). This method automatically derives highly compact feature vectors from congestion patterns (speed contour plots) that include amongst other things the number of wide-moving jams, the amount of homogeneous congestion, and the total spatiotemporal extent of the congestion pattern. This results in small, equal-sized and highly informative feature vectors with which in turn a dendrogram of congestion patterns can be constructed as in Figure 4*d*. As with the demand patterns in Figure 4*b*, the Euclidean distance between two feature vectors is used as distance measure.

Comparing Congestion Patterns?

The question now is what can be concluded from these two dendrograms. Unfortunately, when examining the cluster-trees (dendrograms) of demand patterns and congestion patterns, in Figure 4*b* and *d* respectively, *both seem to suggest systematic differences between road stretches with and without CWS*. This calls into question the plausibility of hypothesis 1. Consider, for example, the pattern with label C in Figure 4*b*. To the right of this pattern there is a cluster of "blue" road stretches (with CWS) and to the left there is a cluster of "red" road stretches (without CWS). All demand patterns on the "blue" (with CWS) road stretches are more similar to each other than to demand patterns on the "red" road stretches (without CWS). This means that it is (just as) likely that the congestion patterns on these "blue" road stretches differ from their "red" counter parts because the demand patterns are different, than whether or not a CWS is installed.

In general terms, a statistical comparison between congestion patterns that occur at two bottlenecks with average demand patterns $dpat_1$ and $dpat_2$ (with and without CWS respectively) is meaningful only if ($dpat_1$, $dpat_2$) (the vertical axis of the dendrogram in Figure 4*b*) is relatively small, that is, smaller than, for example, $\mathcal{D}(dpat_1, dpat_3)$, where $dpat_3$ is the average demand pattern of a third bottleneck *with* CWS. If two demand patterns at similar bottlenecks with CWS look more alike than two

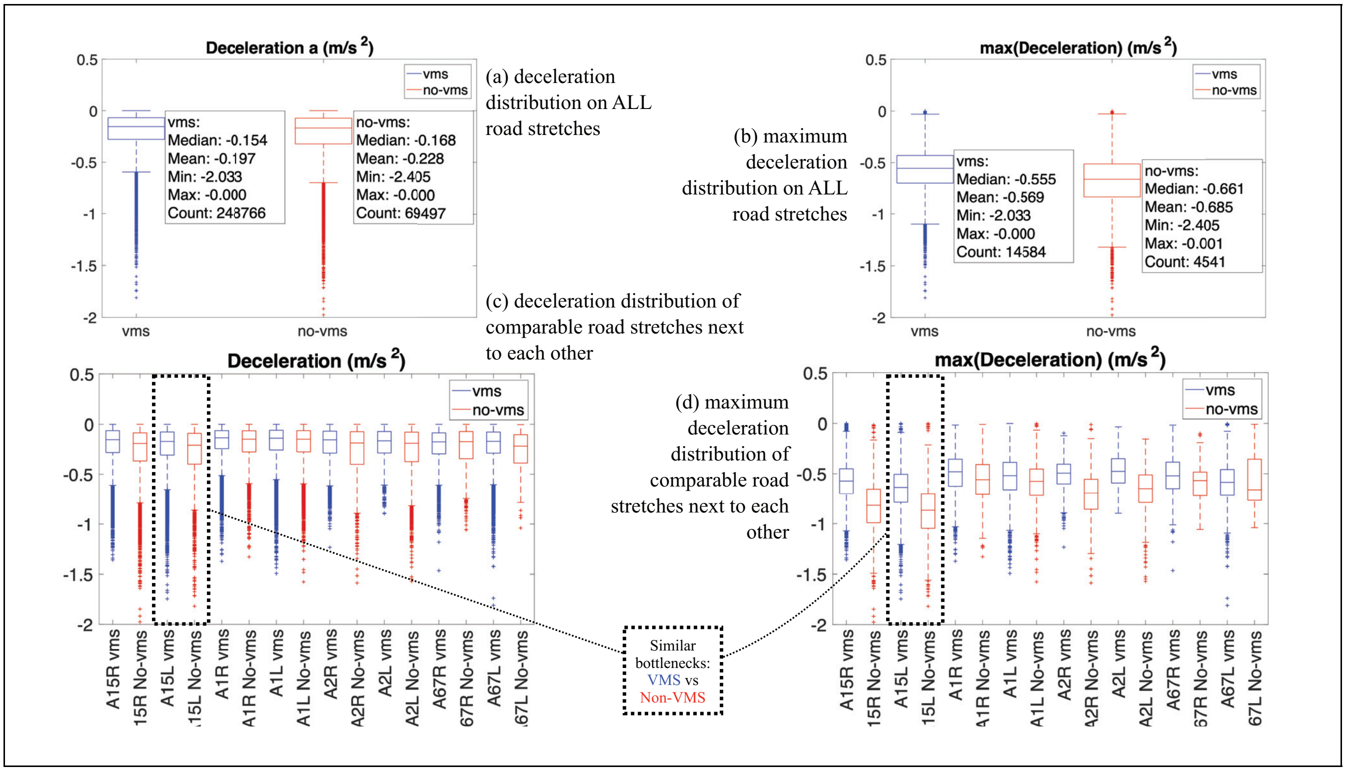

Figure 5. Results comparison (maximum) deceleration distributions for hypothesis 2.
Note: VMS = variable message signs.

Table 1. Results Statistical Tests for Hypothesis 2

	Kolmogorov-Smirnoff test				*t*-test	
	All decelerations		Maximum decelerations		Maximum decelerations	
	p-value	Samples	*p*-value	Samples	*p*-value	Samples
Overall	1,68E-183	248766	6,57E-120	69497	5,72E-184	4541
A1R	1,02E-15	44576	1,37E-10	6530	3,58E-08	388
A1L	2,34E-17	44741	2,15E-12	28149	1,76E-15	1345
A2R	1,09E-40	11771	1,73E-54	3247	3,14E-72	365
A2L	4,40E-23	1947	6,87E-23	9609	3,51E-31	1100
A15R	5,60E-106	49523	1,67E-108	13812	1,33E-162	836
A15L	6,33E-55	58165	1,89E-41	5686	1,85E-67	354
A67R	1,05E-08	6917	4,26E-04	2178	7,52E-04	124
A67L	3,42E-05	31126	1,47E-01*	286	5,15E-01*	29

*Reject hypothesis 2; Otherwise: DO NOT reject hypothesis 2.

demand patterns at bottlenecks with and without CWS, respectively, this is a tell-tale sign that differences in congestion patterns have more to do with differences in demand patterns than that one of the two has CWS. The easiest way to deduce this from Figure 4*b* is to look at the horizontal axis: here it is possible to see very clear clusters of blue (with CWS) and red (without CWS). There are also clusters visible at the dendrogram of congestion patterns Figure 4*d*, although they are less pronounced. The conclusion has to be that hypothesis 1 must be rejected—even without formal statistical testing.

Comparing (Maximum) Deceleration Distributions

Table 1 and Figure 5 show the results of applying the assessment framework in Figure 1*b* to the data. In this case, the results are very different than those for hypothesis 1. The distribution of deceleration values in the non-VMS cases differ systematically from the VMS cases. This holds for the total distribution, that is, all deceleration values for all bottlenecks as illustrated by the box-plots in Figure 5*a*, and the distribution of all maximum decelerations for which Figure 5*b* shows the box-plots.

To compare the distributions quantitatively, the Kolmogorov-Smirnoff (KS) test and a two-sample *t*-test are used, which test the (maximum) distances between the empirical distributions, and the differences between the means of these distributions, respectively. In the latter case, the maximum deceleration distributions *only* were considered; the distributions of all decelerations are too heavily skewed for a *t*-test. Both tests give statistically significant results for the total distributions of (maximum) decelerations (Table 1 top row). This also holds when comparing the distributions of all (maximum) decelerations at individual (comparable) bottlenecks in Figure 5, *c* and *d*, respectively. There is one exception in Table 1 (bottom row) (the A67L), in which the sample is simply too small to reject the null hypothesis. Having said that, the number of observations in all other cases is very large, which usually means a KS test will dismiss the possibility of differences because of chance even if the distributions are close. This implies it is not possible to dismiss the probability that other factors than CWS may be of influence.

However, the systematic results do provide persuasive evidence that the differences found can at least to a significant degree be attributed to the presence of congestion warning using VMS. This is illustrated in Figure 6 which shows cumulative deceleration distribution for one of the bottleneck pairs on the A15R. Clearly, higher decelerations cover a much larger portion of the distribution on the segment without VMS than on the road segment with VMS.

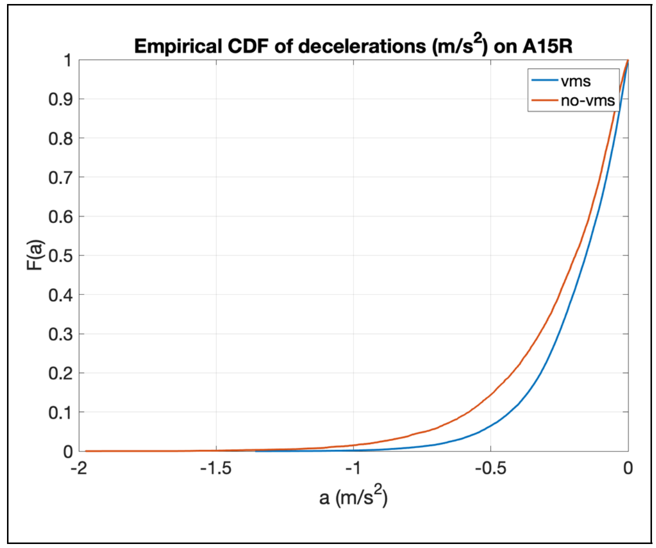

Figure 6. Empirical cumulative distribution functions (CDF) for a pair of congestion warning systems (CWS) and non-CWS road segments (on the A15R).
Note: VMS = variable message signs.

Conclusion and Discussion

In this paper, two hypotheses are tested related to assessing safety effects of measures on road networks using carriageway aggregate data *only*. The first (H1) is that differences in large-scale congestion patterns can be used as proxies for (un)safety. The underlying rationale is that heterogeneous patterns with many disturbances, frequent wide moving jams, and large speed differences result in more potentially unsafe traffic conditions than more homogeneous congestion patterns. The second hypothesis (H2) is that differences in average (maximum) deceleration distributions into congestion waves can be used as a proxy for (un)safety.

It turns out that sufficient evidence was not found to support H1. It was not possible to attribute the differences between large-scale congestion patterns to the presence of CWS. However, because the method in itself is interesting and because reporting negative results is good science, it was decided to include these results nonetheless. *Conclusive evidence was found for H2.*

From the results, it can be concluded that distributions of smoothed carriageway average (maximum) decelerations differ systematically between road segments equipped with CWS and those without CWS. By extension, it can be concluded that these distributions will also differ for the actual underlying decelerations of vehicles. This is positive, because it potentially offers road operators such as Rijkswaterstaat a surrogate safety measure based on average carriageway data, with which they can economically and continuously monitor the effects of CWS (or any other ITS) on traffic safety.

The term "potentially" is used, because there are likely confounding factors that may—in different degrees for the bottlenecks investigated in this study—explain some of the systematic differences observed. These probably include traffic mix (e.g., % trucks), and, for example, local factors like road inclination, visibility lines, and distractions. It was not possible to account for these factors on the basis of the information available. A more in-depth assessment that involves checking for these factors, and validating the method with microscopic data is needed to confirm the hypothesis more rigorously.

However, given the systematic differences found, the evidence is persuasive enough to attribute at least a portion of the differences to the presence of CWS. This conclusion is all the more justified because it is highly plausible that warning drivers about a dramatic speed change (a deceleration) ahead will result in smoother and (in the absolute sense) smaller decelerations.

Acknowledgment

The work presented in this article is part of a collaboration between the Data Analytics and Traffic Simulation Laboratory (DiTTLAB.tudelft.nl) and Rijkswaterstaat (RWS), the executive arm of the Dutch Ministry of Infrastructure and the Environment. RWS is responsible for, amongst other things, the highway and motorway networks in the Netherlands.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Van Lint, Nguyen, Krishnakumari, Calvert, Schreuder, Schuurman; data collection: Nguyen, Krishnakumari; analysis and interpretation of results: Van Lint, Nguyen, Krishnakumari; draft manuscript preparation: Van Lint. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is sponsored by Rijkswaterstaat (RWS) under grant nr 31145099, project “Samenwerking DITT LAB irt bepalen meerwaarde MTM” (collaboration DiTTLAB-RWS in relation to determining added value of the Motorway Management System).

References

1. De Pauw, E., S. Daniels, L. Franckx, and I. Mayeres. Safety Effects of Dynamic Speed Limits on Motorways. *Accident Analysis and Prevention*, Vol. 114, 2018, pp. 83–89.
2. van Nes, N., S. Brandenburg, and D. Twisk. Improving Homogeneity by Dynamic Speed Limit Systems. *Accident Analysis and Prevention*, Vol. 42, No. 3, pp. 944–952.
3. Mittal, A., E. Kim, H. S. Mahmassani, and Z. Hong. Predictive Dynamic Speed Limit in a Connected Environment for a Weather Affected Traffic Network: A Case Study of Chicago. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 13–24.
4. Karafyllis, I., and M. Papageorgiou. Feedback Control of Scalar Conservation Laws with Application to Density Control in Freeways by Means of Variable Speed Limits. *Automatica*, Vol. 105, 2019, pp. 228–236.
5. Han, Y., A. Hegyi, Y. Yuan, S. Hoogendoorn, M. Papageorgiou, and C. Roncoli. Resolving Freeway Jam Waves by Discrete First-Order Model-Based Predictive Control of Variable Speed Limits. *Transportation Research Part C: Emerging Technologies*, Vol. 77, 2017, pp. 405–420.
6. Hegyi, A., S. P. Hoogendoorn, M. Schreuder, and H. Stoelhorst. *The Expected Effectivity of the Dynamic Speed Limit Algorithm Specialist - A Field Data Evaluation Method. Proc., 2009 European Control Conference, ECC 2009*, Budapest, Hungary, 2014.
7. World Health Organization. Top 10 Causes of Death. 2018. http://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/. Accessed November 2018.
8. SWOV. Verkeersveiligheidscijfers: verkeersongevallen (Traffic Safety Numbers: Traffic Accidents). https://www.swov.nl/feitenencijfers/verkeersveiligheidscijfers-verkeersongevallen. Accessed February 14, 2019.
9. Kachroo, P., and A. Sharma. Theory of Safety Surrogates using Vehicle Trajectories in Macroscopic and Microscopic Settings: Application to Dynamic Message Signs Controlled Traffic at Work Zones. *Transportation Research Part C: Emerging Technologies*, Vol. 91, 2018, pp. 62–76.
10. Tak, S., S. Kim, and H. Yeo. Development of a Deceleration-Based Surrogate Safety Measure for Rear-End Collision Risk. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 5, 2015, pp. 2435–2445.
11. Behbahani, H., N. Nadimi, and S. S. Naseralavi. New Time-Based Surrogate Safety Measure to Assess Crash Risk in Car-Following Scenarios. *Transportation Letters*, Vol. 7, No. 4, 2015, pp. 229–238.
12. Li, Z., S. Ahn, K. Chung, D. R. Ragland, W. Wang, and J. W. Yu. Surrogate Safety Measure for Evaluating Rear-End Collision Risk Related to Kinematic Waves Near Freeway Recurrent Bottlenecks. *Accident Analysis and Prevention*, Vol. 64, 2014, pp. 52–61.
13. van Lint, J. W. C., and S. P. Hoogendoorn. A Robust and Efficient Method for Fusing Heterogeneous Data from Traffic Sensors on Freeways. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 25, No. 8, 2010, pp. 596–612.
14. Treiber, M., and D. Helbing. Reconstructing the Spatio-Temporal Traffic Dynamics from Stationary Detector Data. *Cooperative Transportation Dynamics*, Vol. 1, 2002, pp. 3.1–3.24.
15. Nguyen, T. T., P. Krishnakumari, S. C. Calvert, H. L. Vu, and H. van Lint. Feature Extraction and Clustering Analysis of Highway Congestion. *Transportation Research Part C: Emerging Technologies*, Vol. 100, 2019, pp. 238–258.
16. Van Lint, J. W. C. Empirical Evaluation of New Robust Travel Time Estimation Algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2160: 50–59.