The Expected Effectivity of the Dynamic Speed Limit Algorithm
SPECIALIST – a Field Data Evaluation Method

A. Hegyi, S.P. Hoogendoorn, M. Schreuder, H. Stoelhorst

Abstract—Shock waves on freeways can be resolved by dynamic speed limits. In an earlier publication we presented an approach to solve shock waves based on shock wave theory, the so-called SPECIALIST algorithm, and the potential of the algorithm was established by simulation. In this paper a method is presented to evaluate the expected effectivity of the SPECIALIST algorithm based on field data.

The effectivity depends on several factors that determine together whether a certain shock wave is solvable or not. The most important factors are (1) the traffic state in and around the shock wave, (2) the value of the displayed speed limits in combination with the compliance of the drivers with the displayed speed limits, and (3) the delay between the traffic state measurements and the actuation of the speed limits.

To evaluate the expected effectivity of SPECIALIST an algorithm is developed that identifies the shock waves and their relevant properties in the data and evaluates the solvability per shock wave.

The presented methodology is applied to traffic data from the A12 freeway in the Netherlands for 87 morning peaks. The frequency of the occurrence of solvable shock waves is determined as a function of the effective speed limit and the delays in the system.

1. INTRODUCTION

On freeways basically two types of traffic jams can occur: jams with the head fixed at a bottleneck location and jams that have an upstream moving head and tail. Here we focus on the second type, which are often called shock waves [1] or wide moving jams [2], and we use the term shock wave for these jams. These jams are typically short jams (say, 1–2 km) that propagate upstream, due to the incoming vehicles at their tail and the leaving vehicles at their head. They can remain existent for a long time and distance [2]. Consequently, every vehicle that enters the freeway upstream of the jammed area will have to pass through the jammed area, which increases travel times, creates potentially unsafe situations, and increases noise and air pollution by braking and accelerating vehicles. Shock waves typically have a significantly lower outflow than the capacity of the freeway, which motivates the idea that traffic flow can be improved by resolving shock waves. The difference between the free flow capacity and the queue discharge rate is around 30% [2].

In the literature two main approaches can be found to dynamic speed limit control aiming at flow improvement. The first emphasizes the homogenization effect [3]–[6], whereas the second is focused on preventing traffic breakdown or resolving existing jams by reducing the flow by means of speed limits [1], [7], [8]. The basic idea of homogenization is that speed limits can reduce the speed (and/or density) differences, by which a more stable (and safer) flow can be achieved. The homogenization approach typically uses speed limits that are above the critical speed (i.e., the speed that corresponds to the maximal flow). So, these speed limits do not limit the traffic flow, but only slightly reduce the average speed (and slightly increase the density). In theory this approach can increase the time to breakdown slightly [3], but it cannot suppress or resolve shock waves.

The flow reduction approach focuses more on preventing or resolving too high densities, and also allows speed limits that are lower than the critical speed in order to limit the inflow to these areas. The flow reduction by the speed limits is due to two mechanisms: first, at the moment when the speed limit is reduced the traffic will travel with the same density but with a lower speed, which leads to a lower flow, and second, the maximum flow that corresponds to speed limits below the critical speed is less than the capacity. The goal of the flow reduction is to resolve jams by limiting the inflow to them. By resolving the jams (the bottlenecks) higher flows can be achieved in contrast to the homogenization approach as demonstrated in [1], [7], [8].

We presented in an earlier publication an approach to dynamic speed limit control to eliminate shock waves on freeways that is based on shock wave theory, the so-called SPECIALIST algorithm (SPEed ControllIng ALgorIthm using Shock wave Theory) [9]. It was shown by simulation that the algorithm is capable of resolving shock waves, and that the improvement of the total time that the vehicles spend on the freeway is in the range of 10–19%, which is comparable with approaches using other control techniques [1], [8]. The SPECIALIST algorithm has features that are attractive for real-world application: it is based on a simple principle, it has a very low computational demand, and its tuning parameters have a clear physical interpretation.

For a real-world application it is desired to assess its effectivity under real traffic conditions. Therefore, we investigate the relation between the most important conditions and the resulting performance. These conditions include:

- The traffic conditions in the shock wave and upstream and downstream of it. A shock wave may not always be solvable depending on the traffic conditions. We evaluate the frequency of the occurrence of solvable shock waves based on real traffic data.
• The effective speed limits. Based on the theory of SPECIALIST it is known that the required length of the speed-limited stretch increases with increasing speed limit value. If the length of the stretch is limited, then this may imply that certain shock waves are not solvable. Furthermore, the compliance of the traffic with the displayed speed limits may depend on the type of enforcement, and may therefore result in different effective speeds for different enforcement levels. We investigate the relation between the effective speed limit and the solvability of the shock waves.

• The measurement and actuation delay. Given the real traffic data collection systems and the dynamic speed limit displays there may be delays between the measurement and actuation in the range of 30s–2 min. In the time between the measurement and actuation the shock waves propagate upstream, and in some cases shock waves grow, and consequently a longer delay may mean that the shock wave is not solvable anymore. We investigate the relation between the delays and the solvability of the shock waves.

The theory of the algorithm will be discussed in Section II. The theory is translated into an algorithm in Section III, and the approach to the data analysis will be discussed in Section IV. This approach is applied to real traffic data from the A12 freeway in the Netherlands in Section V. In Section VI we summarize the main conclusions.

II. THEORY OF SPECIALIST

The theory of resolving shock waves by dynamic speed limits is based on the shock wave theory as presented by Lighthill and Whitham in their famous paper [10]. Before explaining the approach for shock wave resolution we explain a fundamental relationships in shock wave theory.

A. Shock wave theory

Although shock wave theory goes further than what is presented here, we only present one fundamental aspect, which is necessary to understand the remainder of the paper.

One of the most basic relationships in shock wave theory is the relationship between the time-space graph of the traffic states (as shown on the left in Fig. 1) and the density-flow graph (as shown on the right in Fig. 1). The time-space graph shows the traffic states on a road stretch (along the vertical axis) and their propagation over time (in the horizontal direction). In the figure a short traffic jam is shown that propagates upstream (area 2) and which is surrounded by traffic in free-flow (areas 1). The density-flow diagram shows the corresponding density and flow values for these states. Shock wave theory states that the front (boundary) between two states in the left figure has the same slope as the slope of the line that connects the two states in the right figure. Note that the slopes in both figures have the unit of km/h. The orange lines (light gray in black and white) indicate the fundamental diagram (as a reference).

The importance of this relationship is that if the different traffic states on a freeway stretch are known, then their future evolution can be predicted by describing the fronts between them. This basic relationship will be used in the theory for resolving shock waves.

B. Resolving shock waves

The approach to resolve shock waves consists of different phases and starts with a shock wave similar to the example above.

Phase 1. Assume a shock wave is detected on the freeway as shown in Fig. 2. (How the shock wave is detected will be explained in Section III.) We assume that the traffic state upstream (state 6) and downstream (state 1) of the shock wave is in free flow which is generally the case in real traffic. In Fig. 2 the phases are indicated at the top of the left subfigure. For the sake of readability of the figures we assume that state 1 and state 6 are equal, but the theory also holds for the case when they are unequal.

Phase 2. As soon as the shock wave is detected the speed limits upstream of the shock wave are switched on. This leads to a state change in the speed-controlled area from state 6 to state 3 (in Fig. 2 approximately from 4-8 km), and to the boundary between areas 6 and 3. State 3 has the same density as state 6, as the density does not change when the speed limits are lowered on a longer stretch: no vehicles can suddenly appear or disappear. However, the flow of state 3...

Fig. 1. According to the shock wave theory the propagation of the front between two traffic states (left figure) has the same slope as the line connecting the two states in the density-flow diagram (right figure). The arrow indicates the travel direction.

Fig. 2. The four phases of the SPECIALIST algorithm. Phase 1: The shock wave is detected. Phase 2: Speed limits are turned on in areas 2, 3, and 4. The shock wave dissolves. Phase 3: The speed-limited area (area 4) resolves and flows out efficiently. Phase 4: The remaining area 5 is a forward propagating high-speed high-flow wave.
is lower than that of state 6 due to the combination of the same density with a lower speed.

As shown by the density-flow graph, the front between states 2 and 3 will propagate backwards with a lower speed than the front between states 1 and 2, which resolves the shock wave after some time. The required length of the speed-limited stretch depends on the density and flow associated with state 2 and the physical length of the detected jam. We choose the length of the speed-limited stretch such that the creation of state 3 exactly resolves the shock wave.

At the upstream end of the speed-limited area traffic will flow into this area with the speed equaling the speed limit and with a density that is ‘in accordance’ with the speed – typically significantly higher than the density of state 3 (which was the density corresponding to free-flow). This state is called state 4, and the front between states 6 and 4 will propagate upstream.

**Phase 3.** When the shock wave (area 2) is resolved there remains an area with the speed limits active (state 4) with a moderate density (higher than in free-flow, but lower than in a shock wave). A basic assumption in this theory is that the traffic from such an area can flow out more efficiently than a queue discharging from full congestion as in the shock wave. So, the traffic leaving area 4 will have a higher flow and a higher speed than state 4, represented by state 5. This leads to a backward propagating front between states 4 and 5, which resolves state 4 as shown in Fig. 2.

**Phase 4.** What remains is state 5, and state 6 upstream and state 1 downstream of it. The fronts between states 1 and 5, and between states 6 and 5 propagate downstream, which means that eventually the backward propagating shock wave is converted into a forward propagating wave leading to a higher outflow of the link as shown in Fig. 2.

Obviously, not all traffic situations are suitable to construct the above control scheme. The exact requirements for such a scheme will be discussed in Section III-B when the solvability assessment is discussed.

### III. ALGORITHM DEVELOPMENT

Based on the theory in Section II an algorithm is developed that is suited for real-world implementation, by taking into account the following typical properties of a real-world traffic control system:

- The measurements of traffic monitoring systems are discrete in time and space, with a typical sampling time of 30s, or 1, 2, or 5 minutes, and a typical detector spacing of 0.5, 1, 2, or more kilometers.
- Traffic density is typically not measured, but speeds and flows often are.
- The speed limit actuation may occur with a fixed time step or in an event-based manner depending on the technical properties of the speed limit actuation system.

Here we present only the global steps of the algorithm with the focus on the part that is relevant for the data analysis. We refer the interested reader to [9] for the exact details.

The steps of the algorithm can be described as follows:

1. **Shock wave detection.** When new flow and speed measurements arrive, test whether there is a shock wave on the considered stretch. If there is no shock wave present then wait for the next measurement and go to step 1, otherwise continue with step 2.

2. **Control scheme generation.** Based on the traffic states, calculate the control scheme according to the theory of Section II (i.e., the location of the fronts and the active speed limits is determined).

3. **Solvability assessment.** Determine the solvability of the detected shock wave based on the measurements in the shock wave and upstream and downstream of it (solvability criteria are discussed below). If it is not solvable then wait for the next measurement and go to step 1, otherwise continue with step 4.

4. **Control scheme application.** Determine the applicable speed limits for the current moment based on the control scheme. The speed limits that are in areas 2, 3 and 4 are activated. Repeat this step regularly (e.g., every second) and ignore new measurements until no speed limits need to be set anymore according to the control scheme and thus this control scheme is finished (i.e., area 4 has been resolved). Wait for the next measurement and go to step 1.

In the data analysis part of this paper the shock wave detection and the solvability assessment play an important role. Therefore we discuss these steps in more detail here.

### A. Shock wave detection

In the shock wave detection step the shock wave is detected by using thresholds \( v_{\text{max}} \) (km/h) for the speed and \( q_{\text{max}} \) (veh/h) for the flow measurements, by assuming that in segment \( i \), a shock wave is present if \( q_i \leq q_{\text{max}} \) and \( v_i \leq v_{\text{max}} \). When there are no other types of jams on the considered stretch, this identifies the location of the shock wave.

### B. Solvability assessment

For the assessment of the solvability first all the traffic states in the control scheme are determined. The traffic states are denoted \( v_{[j]} \), \( q_{[j]} \), \( \rho_{[j]} \), \( j \in \{1, \ldots, 6\} \) for the six states. Since \( q_{[j]} = \rho_{[j]} v_{[j]} \), it is sufficient to determine two of the three values. The areas directly upstream and downstream of the shock wave that are not falling below the thresholds are classified as free-flow, and the measurements from these areas are determined to calculate the average states upstream and downstream of the shock wave.

The average flow \( \bar{q} \) (veh/h/lane) and average density \( \bar{\rho} \) (veh/km/lane) upstream (downstream) are approximated by

\[
\bar{q} = \frac{1}{N} \sum_{i \in I} q_i, \quad \text{and} \quad \bar{\rho} = \frac{1}{N} \sum_{i \in I} (q_i/v_i),
\]

where \( q_i \) (veh/h/lane) and \( v_i \) (km/h) are respectively the flow and speed measurements of segment \( i \), and \( I \) is the set of appropriately chosen segment indices and \( N \) the number of segments considered. This determines states 1 and 6. As a very low speed is associated with state 2 and induction loop detectors are known to be inaccurate for low speeds, the density of state 2 is determined by \( \rho_{[2]} = \rho_{[1]} + (q_{[2]} - q_{[1]})/v_{[1,2]} \), where \( v_{[1,2]} \) denotes the propagation speed of the
head of the shock wave. The propagation speed of the head of the shock wave can be taken from off-line data, since it is well-known to be fairly constant (around -18 km/h). State 3 directly follows from the density of state 6 and the used speed reduction. The speed of state 4 equals the speed of the speed limits. However, the density of state 4, and the density and flow of state 5 do not follow from the shock wave theory and can be considered as a design variables for which heuristic tuning rules can be given. For now, it is sufficient to assume that they have a fixed value.

When all six states are determined, the control scheme can be constructed. This involves the determination of the various fronts by solving straightforward linear equations based on the traffic states and shock wave theory. Due to space limitations we do not present the equations here.

In the construction of the control scheme two conditions play an important role. First, the delay between the moment that the traffic is physically measured and that the according speed limits are displayed may be in the order of several minutes. The effect of the total delay on the control scheme is that the speed limits will be scheduled in the future to be actuated, and thus the state change from state 6 to state 3 is the expected moment when the displays will physically display the speed limit drop.

The second condition that plays an important role in the control scheme construction is the effective speed that the traffic will assume when the speed limits has been lowered. This depends on both the displayed speed limit value and the compliance with the speed limits. In the context of the SPECIALIST algorithm the relevant quantity is the speed at which the traffic eventually will drive for which we will use the term effective speed limit. In the control scheme generation the effective speed limit is taken into account by assigning this speed to the states 3 and 4. The effective speed limit can be estimated based on traffic measurements.

After the construction of the control scheme the solvability is assessed. A shock wave is classified as solvable if:

- the control scheme can be constructed according to the theory given the traffic states 1–6. The necessary and sufficient conditions for the constructions of the control scheme are:
  - the head and tail of area 2 should converge, otherwise the shock wave will not be resolved,
  - the same applies for area 4,
  - state 5 should have a higher flow and a higher density than state 1, otherwise there is no forward propagating front between these two states,
- the speed in area 6 should be higher than the speed limit, otherwise the speed limits do not have any effect.
- the necessary length of the speed-limited trajectory should be smaller than the upstream free-flow area,
- if the speed-controlled area should be limited by the physical availability of speed limit signs the controlled area following from the scheme should fall inside the physically available speed limits.

If these conditions are satisfied, the shock wave is classified as solvable.

C. Tuning

In the algorithm there are several parameters that can be selected by the designer of the control system: $v_{\text{max}}$, $q_{\text{max}}$, $\rho_{[4]}$, $\rho_{[5]}$, and $\theta_{[5]}$. For these variables heuristic tuning rules can be given, partially based on offline traffic data and partially based on the online (closed-loop) behaviour of the algorithm. Due to space limitations we refer the interested reader to [9].

IV. DATA ANALYSIS APPROACH

The goal of the data analysis is to make an assessment of the real-world applicability of the SPECIALIST algorithm. In this context the major questions are:

- How frequently do shock waves occur?
- How many of them are solvable according to the conditions in Section III-B?
- What is the relation between the number of solvable shock waves and the delays in the system?
- What is the relation between the number of solvable shock waves and the value of the effective speed limit?

Similarly to the algorithm the data analysis method is using speed and flow measurements with a discrete time step and discrete spacing with typical values as mentioned in Section III. The approach consists of two major steps, the identification of the shock waves in the data and the evaluation of the solvability of each shock wave.

A. Shock wave identification

The shock wave identification algorithm is based on the typical characteristics of shock waves: the speed and flow are very low, and the propagation speed of the head is fairly constant (around -18 km/h). The general idea of the algorithm is to use the property of the fairly constant propagation speed of the head to distinguish a shock wave from other types of jams (such as jams at a fixed bottleneck), and to detect the head of the shock wave at each time instance $k = kT$, where $T$ is the sampling time of the measurements.

When the head locations of the shock wave are determined for each time instance then the rest of the shock wave is identified by using the thresholds $v_{\text{max}}$ and $q_{\text{max}}$ for the speed and flow respectively.

Due to the thresholding and the noise in the traffic process and the measurements it may occur that the speed or flow of a shock wave does not fall below the thresholds for a given time instance and detector location, while they do fall below the thresholds for a subsequent (and previous) time instance and detector location. Since such a pattern is still related to the same shock wave we allow short interrupts in time and space of length $t_{\text{max INTERRUPT}}$ [h] and $x_{\text{max INTERRUPT}}$ [km].

Summarizing the shock wave detection algorithm, the following steps are taken:

- For $k = 1, \ldots$ the first occurrence of the head of a shock wave is detected by taking the most downstream location for which the speed and flow fall below the thresholds $v_{\text{max}}$ and $q_{\text{max}}$ respectively.
A. Hegyi et al.: The Expected Effectivity of the Dynamic Speed Limit Algorithm SPECIALIST – a Field Data Evaluation Method

TuA2.4

• If such a head is found, the head location at the subsequent time instances is detected by selecting the location that satisfies the following requirements:
  - it is a location with a speed and flow that satisfies the thresholds \( v_{\text{max}} \) and \( q_{\text{max}} \).
  - it is a location that is consistent with the propagation speed of the shock wave head. Since the space discretization step is rather coarse (around 500-600m) compared to the time discretization step (1 min) there may be relatively large differences between the location where shock wave head is detected (the detector location) and where it physically is. Therefore, the exact formulation regarding the head propagation speed is that for a given shock wave the current location of the head \( x_c(k_c) \) should be in the range \[ x_i(k_i) - \frac{x_{\text{error}}}{T(k_c - k_i)} + v_{\text{max,head}} * T(k_c - k_i), x_i(k_i) + \frac{x_{\text{error}}}{T(k_c - k_i)} + v_{\text{min,head}} * T(k_c - k_i) \], where \( x_i \) is for one given shock wave the location of its head in the \( i \)-th time step, and the term \( \frac{x_{\text{error}}}{T(k_c - k_i)} \) is the allowed error margin, which is decreasing over time, and \( v_{\text{min,head}} \) and \( v_{\text{max,head}} \) are the minimum and maximum allowed propagation speeds of the jam head.
  - if no such head location is detected then an interruption of at most \( t_{\text{max,interrupt}} \) is allowed. If there is no such head for a longer time, then the pattern is classified as two shock waves.

• When all shock wave heads are detected the complete jam pattern is identified by classifying all locations as a part of the jam that fulfill the following criteria:
  - The locations are upstream of the head and the speed and flow satisfies the thresholds \( v_{\text{max}} \) and \( q_{\text{max}} \).
  - The locations (including the head) are contiguous with interruptions of no more then \( x_{\text{max,interrupt}} \).

• When all shock waves are identified, the average traffic states upstream and downstream of them are determined, and are used for the solvability evaluation.

B. Solvability evaluation

The evaluation of the traffic data for several days containing several shock waves is performed for each time instance during each shock wave in each day separately. The solvability of a shock wave is determined according to the approach discussed in Section III-B, except for the front speed of the shock wave, which is not taken to be constant (as in the algorithm), but is determined based on the position of the head at the first and the last moment of detection in the data.

The dependency of the solvability on the effective speed limit is evaluated by selecting state 3 in SPECIALIST according to the effective speed limit for various speed limit values. The dependency of the solvability on the delay is evaluated by evaluating the solvability of each shock wave for each time step during its existence and taking taking the first number of steps (1, 2, \ldots) for which the shock wave is solvable. For example if the shock wave is solvable for the first ten minutes then the maximal delay for which is shock wave is expected to be solvable is 9 minutes.

V. DATA ANALYSIS PRACTICAL APPLICATION

The freeway stretch that we consider is a part of the Dutch A12 freeway and has three lanes and a length of approximately 14 km going from the connection with the N11 at Bodegraven up to Harmelen as shown in Fig. 3. The stretch includes a few on-ramps, however the on-ramp volumes do typically not create jams on the freeway. The shock waves are often created around km 48 (in the figure behind the red A12 sign).

The stretch is equipped with double loop detectors with a typical spacing of 500 to 600m, measuring the average speed and flow every minute. Above each detector there is a VMS panel that displays the speed limit.

For the data analysis 92 morning peaks (6:00-11.00 am) were considered in the period January-April 2006. During the pre-selection of the data several days were discarded due to a significant amount of missing data. In the remaining 92 days, data was missing only occasionally for some combinations of time and location.

The parameters used in the evaluation were: \( v_{\text{max}} = 80 \) km/h, \( q_{\text{max}} = 3600 \) veh/h, the speed and flow of state 5 \( v_5 = 75 \) km/h, \( q_5 = 6600 \) veh/h, the density of state 4 was selected such that the speed is according to the effective speed limit and the state is on the line that connects state 1 and state 2 in the density-flow graph, the front propagation speed parameters \( v_{\text{min,head}} = -15 \) km/h, \( v_{\text{max,head}} = -23 \) km/h, \( t_{\text{max,interrupt}} = 5 \) [min] and \( x_{\text{max,interrupt}} = 1.5 \) [km].

Furthermore, shock waves of duration of 1 minute were discarded, as they were apparently quickly resolved without intervention, and if traffic data was missing in or around the shock wave then it was classified as not solvable.

The expected measurement delay in the Dutch case is around 1–2 min. The actuation delay is in the order of seconds and is considered to be negligible. Therefore, delays in the range of 0–3 min are investigated. Furthermore the effective speed limits in the range 50-100 km/h is investigated.

A. Results

In the investigated 92 morning peaks there were five morning peaks for which more than 20 shock waves were detected. These shock waves were so densely spaced that there was virtually one large traffic jam with some minor free-flow area’s in between. As these are no real shock waves, these days were also discarded and the rest of the analysis was performed for the remaining 87 days.

In the remaining 87 days 407 shock waves were found for which the solvability for various delays and effective speed limits is determined. In Figure 4 the relation between the effective speed limit, the measurement delay and the number of the solvable shock waves is shown. It is clear that both increasing measurement delay and an increasing effective
shock waves out of 407 (approximately 6%). This leads to between 21 and 33 resolvable shock waves, depending on the effective speed limit 70 km/h (if 60 km/h is displayed with 1.5 min and for the system delay. In the case of a delay of 1.5 minute and an effective speed limit of 70 km/h approximately 6% of the shock waves are expected to be solvable.

A note should be added here about the interpretation of the results. In a real-world application, the parameters of the algorithm need to be tuned online, as mentioned in Section III-C. The “shape” of the control scheme, and consequently, the expected effectivity, may depend on these parameters. As online tuning is not possible with offline data, a robustness/sensitivity analysis would be needed to investigate the dependency of the expected performance on these parameters, which is a topic for future research.

ACKNOWLEDGMENTS

Supported by the DYNAMAX project of the Dutch Ministry of Transport, Public Works and Water Management.

REFERENCES


