Examination of the Effects of Dynamic Speed Limit on Shock Waves with a Simulation Technique

Viktor Nagy\textsuperscript{1*}, Gergely Kruppa\textsuperscript{1}, Balázs Horváth\textsuperscript{1}

\textsuperscript{1} Department of Transport, Széchenyi István University, Egyetem tér 1., H-9026 Győr, Hungary
\textsuperscript{*} Corresponding author, e-mail: nagy.viktor@sze.hu

Received: 21 March 2023, Accepted: 16 October 2023, Published online: 13 November 2023

Abstract
The aim of motorway networks is to ensure smooth, high volume road traffic. The problem of anomalies in the system has been a long-standing concern for researchers, and with the widespread use of motorway networks a series of studies have been published on the subject. Speed limitation is one of the most important tools of mitigating disturbances. This study is looking for a solution to the problem known as the "shock wave effect". The research is based on a simulation method in a PTV VISSIM environment. The software can be used to examine certain traffic conditions, and then apply dynamic speed limitation based on these conditions. The Built Environment Information Platform (BENIP) is based on the idea that the built environment, the traffic, and the flow of information between them are closely related. The relevance of the study lies in the fact that a point-based notification of speed limits can be used for vehicles – a technology that is available in most cases at the infrastructure level on motorways – thus, improving road capacity and traffic safety.

Keywords
shock wave effect, phantom jams, dynamic variable speed limit, variable message signs, VMS, VISSIM

1 Introduction
The design of the motorway networks aims to ensure a smooth flow of traffic. There are no at-grade crossings, no traffic lights, no oncoming traffic. In default cases, reducing the speed to zero or drastically lowering the maximum permitted speed is said to be unjustified and it never occurs in normal conditions. Some undesirable disturbance appearing in the system is referred to as congestion. We often see such incidents, for example in the case of accidents or congestion caused by road work. They are known as bottlenecks, where the number of downstream lanes is fewer than the number of upstream lanes. To mitigate this type of congestion we can find several proposals, such as ramp-metering. Based on the cited references, ramp metering seems to be one of the best solutions to lowering the volume of incoming traffic, there are isolated (Hegyi et al., 2002) results, but also whole research groups involving in the ramp metering theory with several future ready solution like the study of Gáspár and Németh (2019), or a whole theoretical and practical approach by Luspay et al. (2012). At the ramp metering problem there are also multicriteria solution like described by Csikós and Varga (2012). Another solution is the optimising lane changes before congestion (Yuan et al., 2021), or a combination of both (Zhang and Ioannou, 2017), (Vrbanić et al., 2021). However, several other phenomena can also cause congestion. This type of congestion is caused by traffic intensity, like described its connection with capacity by Sarwar et al. (2021). Some researchers did analytical approaches in the past (Daganzo and Geroliminis, 2008), but we found the state-of-the-art simulation much appropriate method to discover the connections between the elements and actors (elements and actors mean here users of the examined transport system, like vehicles and drivers). Although this kind of research is just small building block of the whole, we have to see that finally our efforts aiming to reach a more efficient and effective transport system, regarding resource usage, like described by Szabó et al. (2021).

In our research, we investigated how to solve the so-called "phantom congestion" defined by Sugiyama et al. (2008) using simple solutions that do not require continuous vehicle to infrastructure (V2I) communication. The other way round (e.g., continuous communication) is already researched (Ishikawa and Arai, 2016; Won et al., 2017). We anticipate that speed limitation as a tool would help
to optimise the use of the infrastructure and the dynamically varying speed limitation would produce better results than the simple static case. The variable message signs (VMS) can effectively influence driver behaviour (Ghosh et al., 2018; Yan and Wu, 2014) therefore VMS could have an effect on traffic safety (Lee and Cho, 2015; Yu and Abdel-Aty, 2014) or on emission in urban area (Kölbl et al., 2015). Thus, complexity of the transportation system (infrastructure, driving behaviour, IT solutions, communication) demonstrating the importance of the interrelation between the built environment, information flow and traffic, which is the basis of the Built Environment Information Platform (BENIP) idea (Horváth et al., 2023).

2 Problem description

Under low traffic, free-flow conditions, the idiosyncratic driving behaviour of some participants is not a problem, and the reaction time and habits of drivers do not play a significant role. However, as soon as a critical amount of traffic volume (Kerner's minimum and maximum capacity) is reached (through increasing traffic density), the flow changes from a stable to a so-called metastable state, as shown in Fig. 1 based on the theory of Kerner (2004).

Traffic density is defined as the number of vehicles on a road section of a given length. For high intensity traffic, it obviously takes a higher value, which is only possible, if headway are reduced. When headway are reduced, we must react to the surrounding traffic. For these reasons, a sudden deceleration forces the following vehicle (behind the braking one) also to slow down. Practice shows that the following vehicle reduces its speed more intensely than the first braking action. This is mainly due to human reaction time and the stochastic driving style of humans (Orosz et al., 2006).

According to Horn and Wang (2018), the distance between two cars acts as a type of shock absorber that can absorb the disturbance caused by such speed differences.

To maintain sufficient headway between two vehicles after a braking event, the following vehicle must brake harder (due to the additional distance travelled during the human reaction time). This is true for any other following vehicle that is within a critical distance. Unless there is a larger section with lower traffic density where this process is mitigated, traffic may stop or slow down considerably for several hundred metres upstream. Some researchers refer to this shock-wave-like congestion as "phantom congestion" (Suijs et al., 2015), which results only from traffic instability.

3 Examination method

In the research we used simulation techniques to observe the behaviour of highway traffic. Fig. 2 describes the concept of our examination.

3.1 Simulation environment

Our research is based on a microscopic traffic simulation environment created by PTV VISSIM software. In the software, the behaviour of each vehicle can be defined, such as the timing of lane changes, the rate of braking, and the rate of accelerations (Zhang and Ioannou, 2017).

The "Attribute Modification" function available in the software was used in the research. It allows for a dynamic modification of any parameter of any object in the network, continuously during the simulation run. By having designed a suitable algorithm, we could examine the traffic flow under varying speed limit values.

In the initial stages of the research, we investigated whether the phenomenon occurs under simulation conditions. After having been convinced of this, subsequently we generated congestions under controlled conditions. Two types of simulations were evaluated. First,
we allowed vehicles to pass without any intervention and conducted our measurements, while in the second, a pre-defined algorithm controlled the traffic speed. To ensure the validity of the results, both cases were run with three different "random seed" values. All other initial parameters were the same in all six cases.

In the model, we collected our measurement data pointwise, and the communication with the vehicles assumed also pointwise road signs with fixed locations. Therefore, we did not assume any continuous communication channel between the infrastructure and the vehicles; thus, we used real-life solutions that are currently in use.

The geometrical requirements for the track were limited to the requirement that it should not affect the flow of traffic. A straight section seems obvious, but finally a spiral track with an intro section was chosen. This is due to the limitations of our software licence, which is limited to an area of 1500 m × 1500 m. The track, thus, fitted entirely onto the display screen, so we could see the entire process.

The network included a 1200 m long buffer section where traffic was loaded onto the network. Traffic events on this section were not analysed and not included in the results, leaving space and time for vehicles to stabilize their headways and desired speeds.

Then, the test section was divided into 15 parts, each 1000 m long. Finally, a 5000 m long run-out section was added. The lengths of the sections were chosen so that they were not too short, which could make the deployment unnecessarily expensive or too sensitive to disturbances and not too long, which would reduce the impact of the algorithm on the network. The speed limit points were at the connection of every section, and the data collection points were on the same places. This is important because we did not take data of the network that assumed any communication between the infrastructure and the vehicles, which may not be impossible today, but not too widespread. This was to ensure that the model could work even in real life. The road had two traffic lanes along its entire length, each of the lanes being 3.5 meters wide.

3.2 Traffic inputs
The traffic load parameters were the same for all examination variants to keep the results comparable. The vehicle composition was homogeneous, so only vehicles with the same parameters were loaded to the network. They were modern passenger vehicles, suitable for motorway speeds in all cases. The main reason for this was that the speed difference between vehicles with significantly different main parameters (a trailer capable of 90 km/h and a passenger car capable of 130 km/h) could be so large that it may have caused disturbance in the system. The homogeneous composition was used to keep the effect of all other factors as low as possible. However, the actual speed of each vehicle followed a predefined distribution.

It should be noted that the "desired speed distribution" means the free-flow speed of each vehicle, i.e., the speed the vehicle wishes to use when there are no forcing traffic situations. This is the "free-flow" state of traffic, indicated by the letter "F" as shown previously. The default distributions of the software were used for this study. Fig. 3 shows the speed distribution associated with 130 km/h, but all speed values are similar.

In determining the traffic volume, we wanted to identify critical values at which traffic could enter the Kerner's metastable traffic condition (Kerner, 2004). In this case, not only the traffic volumes are so high that vehicles are forced to react to one another (no free flow), but the traveling of the vehicles is also not synchronized; in other words, a slower but steady flow of traffic has not been established.

In these conditions, congestion spreads like a shock wave, there are major differences in speed and headway among participants; therefore, the traffic is unstable. We wish to note that this was our goal in this research, i.e., to shift from an unstable state to a slower but loss-minimising stable, synchronised state, smoothing out the speed and headway differences mentioned above. The number of vehicles is based on the research of Yuan et al. (2021), which stated that the throughput of the road begins to decrease between 3000 and 4000 vehicles/hour, at which point it deviates negatively from the line indicating the absence of disturbance. Based on the results, it would be a workable solution to choose the value of 4000 vehicles/hour and work consistently with it.

![Fig. 3 Speed distribution of vehicles](image)
However, previous studies have shown that over a certain traffic volume, the intervention becomes useless because of the shock absorption phenomenon. Thus, congestion can extend for kilometers if there is no space for expansion in the system. In order to handle the transitions and changes as well as to give enough expansion space, we used traffic fluctuation between 3800–4200 vehicles/hour.

3.3 Simulation parameters

The length of the simulation was set for one hour. The first 10 minutes were the recharge phase, therefore, no data were collected from this interval. After the first 10 minutes, the sampling interval was set for 1 minute, which was collected until the end of the simulation. The sampling value was adapted to the algorithm run interval, which also controlled the traffic in every minute. The speed limit points were placed at every 1000 m. The speed limits changed between an interval from 60 to 130 km/h with a difference of 10 km/h. To ensure a controlled simulation, sudden speed reduction points were set up at 3 separate locations, which were activated at different times, and they forced vehicles to brake intensively. The braking points and their activation times were as follows:

- at 4 km at the 10th minute,
- at 8 km at the 15th minute,
- at 12 km at 20 minutes.

All of the braking points activated in every single simulation, so the three simulation runs contain the same three braking event.

4 Algorithm description

At the end of each minute, the algorithm performed the necessary interventions on every examined section. The method was based on the speed factor (which is practically the utilization of the current speed limit) that can be calculated using the following formula:

\[
SF_{i,j} = \frac{AS_{i,j}}{S_{i,j}},
\]

where:

- \(SF_{i,j}\) is the speed factor on the \(i\)-th section in the \(j\)-th minute,
- \(AS_{i,j}\) is the average speed on the \(i\)-th section in the \(j\)-th minute,
- \(S_{i,j}\) is the current speed limit on the \(i\)-th section in the \(j\)-th minute.

The calculated speed factor affects the "modified speed" value for the current section in the next minute. We used the VISSIM software Attribute Modification function (PTV Group), where the program counts and refresh the speed limits in every single minutes as follows:

if \(SF_{i,j} > 0.8\)
then \(MS_{i,j+1} = S_{i,j} + 20\)

if \(SF_{i,j} > 0.7\) and \(SF_{i,j} \leq 0.8\)
then \(MS_{i,j+1} = S_{i,j}\)

if \(SF_{i,j} \geq 0.6\) and \(SF_{i,j} < 0.7\)
then \(MS_{i,j+1} = S_{i,j} - 20\)

if \(SF_{i,j} \geq 0.5\) and \(SF_{i,j} < 0.6\)
then \(MS_{i,j+1} = S_{i,j} - 30\)

if \(SF_{i,j} \geq 0.4\) and \(SF_{i,j} < 0.5\)
then \(MS_{i,j+1} = S_{i,j} - 40\)

if \(SF_{i,j} < 0.4\)
then \(MS_{i,j+1} = S_{i,j} - 50\)

where:

- \(MS_{i,j+1}\) is the modified speed on the \(i\)-th section in the \(j + 1\)-th minute.

The speed limits may vary between 60 km/h and 130 km/h, with no speeds above or below these limits. In addition, the algorithm monitors the next 3 km long sections and the newly calculated speeds on them. A given limitation will therefore also influence 3 sections upstream, as follows:

\[
MS_{i,j+1} = MS_{i+1,j+1} + 10,
\]

\[
MS_{i,j+1} = MS_{i+2,j+1} + 20,
\]

\[
MS_{i,j+1} = MS_{i+3,j+1} + 30.
\]

Thus, four different modified speed values have been developed for each minute for each section, which are:

- modified speed based on the speed factor,
- modified speed based on the speed of the first next section,
- modified speed based on the speed of the second next section,
• modified speed based on the speed of the third next section.

The algorithm examined the four different speed values and selected the lowest one with a minimum of 60 km/h and a maximum of 130 km/h. Then, it will be the speed limit on the current section for the following minute.

To evaluate the results of the simulation, we divided the speed ranges into three categories:
• above 80 km/h: normal,
• 60–80 km/h: acceptable,
• 0–60 km/h: dangerous.

A speed limit of 80 km/h is in use for some vehicle types in a motorway environment in Hungary, so this has not yet been considered as a dangerous speed drop, while the Highway Code states that a vehicle what unable to travel with 60 km/h is prohibited from using the motorway, so this has been set as the limit of the danger zone.

5 Results
5.1 Average travel time
The average travel time is a particularly sensitive area due to the fact that by decreasing the speed limit, we will ideally ensure that there will be no significantly faster participants than the current speed limit. On the contrary, if the average traffic speed naturally decreases, e.g., due to congestion, or significantly increased traffic density, there are spots where the limiting factor may be reduced and traffic could speed up. This property implies that the speed would remain unnecessary low if the speed limiting algorithm cannot respond to these spots. Thus, the travel time would be increasing compared to an uncontrolled case, which is unfavourable like in the Fig. 4.

We can see that the algorithm can deal with the problems since the average travel times did not increase significantly. During the whole simulation, the average travel time based on the three simulations were 856 seconds in the uncontrolled case and 859 seconds in the controlled case.

5.2 Average speed
Average speed values were measured in every minute for each test section (per kilometers). It has an advantage over the average speed over the whole network because it also allowed us to see the spatial effects. The results of the three independent pairs of measurements were averaged. The primary reason for this being our usage of artificial barriers with the same parameters in all cases so that the measurements gave similar results. Fig. 5 shows the uncontrolled case while Fig. 6 shows the controlled case.

We can see how the shockwave propagates backwards. We can see that, if the algorithm detects a drop in speed on a given section, it starts to use a staggered speed limit on the earlier sections too, thus slightly damping the arrival rate of vehicles that could strengthen further the congestion. Overall, the area of the critical zones, their length in time and space has not been drastically reduced. The reason for that could be that the artificial congestion generation was too strong and for this there was no answer, which could resolve this. And also, maybe the graph is not sensitive enough to subtle differences.

It is also important to note the phenomenon that starts at the section 15 and continues until section 12 and would certainly continue if the simulation had allowed time for this
to happen. Here, a spontaneous (not artificially generated) congestion starts, simply because of the metastable traffic condition. The minimum average speed measured also falls between 40 km/h and 60 km/h. However, this was resolved by the algorithm with a minor intervention within a few minutes before a significant congestion. You can see how the algorithm works by slightly smoothing the amplitude of the speed variations, thus stabilising the traffic. Table 1 shows how the average speeds changed on the sections.

For the whole network, in the uncontrolled case, there were 371 minutes in total over the 3 simulation runs where the average speed on a section was below 60 km/h, while in the controlled case this was reduced to 301 minutes. On the other hand, the times spent in the normal speed range also decreased and slipped into the acceptable range.

5.3 Stops
An important result was achieved in terms of stops. The number of stops is shown in Fig. 7 over time. Significantly higher values were obtained for uncontrolled cases. Braking to a stationary state is a serious road safety risk. The number of stops reached a three times lower value than in the most extreme case (measurement pair 3–28.8%) and decreased to the half in the other two cases (45.5% and 53.9%).

In addition to the number of stops, the average standstill time also decreased, on average by 41% in the three cases, from 6.07 seconds to 3.58 seconds. The total standstill times required for stops for all measurements is shown in Fig. 8. The average of the uncontrolled cases is 6006 minutes. The same value for the controlled case is 3494 minutes.

Looking at these results it is also noticeable, that despite the average speeds have decreased slightly over the whole network, the deviation of speeds (what cause dangerous situations) has also decreased, and the number and time of extreme cases in the dangerous range (0–60 km/h) has been reduced.

6 Conclusion
The shockwave effect and spontaneous congestions in high traffic intensity cases are real problems on our motorway networks. We know from previous research that dynamic speed limit can be remarkably effective in mitigating these problems, and we have also seen that an algorithm can be created in PTV VISSIM environment that responds to congestion in an adaptive and efficient way. In the research, we considered that all network elements should be easily available, thus avoiding the need for new technologies that will certainly be available in the medium-term future but are not yet widespread today. Therefore, we used only point-based calculated data and communicated only point-based instructions to users.

As a result, the measurement shows that the number of stops is drastically reducible and the standstill times can also improved. Although we did not increase the speed of traffic flow, we managed to shift it from an unstable state to a stable one, thus significantly improving traffic safety.

As a further direction of research, we would like to investigate how the algorithm reacts in the case of mixed traffic with different vehicle types. To see what the optimal distance is where the traffic manipulation works well, but where the spacing of the measurement points is not too dense. It is also worthwhile to develop a computational method to calibrate the speed factors as speed modification parameters. Further simulation runs would investigate the impact of the algorithm on spontaneously occurring, less drastic congestion, which seems promising to manage due to its fast response, and based on the above-mentioned ideas, we would like to investigate how much the travel time can be reduced. In addition, we would like to examine the impact on emission.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Travel time spent in individual speed categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed [km/h]</td>
<td>Uncontrolled [pc × min]</td>
</tr>
<tr>
<td>80 –</td>
<td>1773</td>
</tr>
<tr>
<td>60 – 80</td>
<td>106</td>
</tr>
<tr>
<td>00 – 60</td>
<td>371</td>
</tr>
</tbody>
</table>
References


