Anatomy of tunnel congestion: causes and implications for tunnel traffic management

Athina Tympakianaki\textsuperscript{a*}, Haris N. Koutsopoulos\textsuperscript{ab}, Erik Jenelius\textsuperscript{a}

\textsuperscript{a}Department of Transport Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

\textsuperscript{b}Department of Civil and Environmental Engineering, Northeastern University Boston, MA 02115

* Corresponding author. Tel.: +46 8 7908891.

Email addresses: athinat@kth.se (A. Tympakianaki), h.koutsopoulos@neu.edu (H. N. Koutsopoulos), jenelius@kth.se (E. Jenelius)

Abstract

Purpose

Tunnel congestion is an important problem often dealt with disruptive traffic management strategies, such as closures. The paper examines the underlying causes of tunnel congestion and their impact on the most appropriate strategy to be implemented. It tests empirically the hypothesis that the cause of congestion may be different from day to day, even for the same tunnel. Furthermore, the effectiveness of tunnel management strategies may differ, depending on the cause.

Method

A methodology is proposed to test those hypotheses using empirical data and simulation, consisting of: (i) Cluster analysis of historical traffic data to identify distinct congestion patterns; (ii) in-depth analysis of the underlying demand patterns and associated bottlenecks; (iii) simulation analysis to identify effective strategies for each demand pattern; (iv) classification analysis which assigns a particular day to one of the patterns based on the available data.

Results

The methodology is demonstrated for a congested tunnel in Stockholm, Sweden, which is subject to frequent temporary closures for safety reasons. The cluster analysis indicates two different congestion patterns, with different spatiotemporal characteristics. Multiple strategies based on tunnel closures and metering are implemented and evaluated for both traffic patterns. Appropriate closure timings can improve the traffic conditions in the tunnel, while metering is found to be the most promising strategy overall. Further, classification of a new day is shown to help predict tunnel congestion early and may inform the strategy recommendation.

Conclusions

The analysis emphasizes the identification of distinct traffic patterns leading to tunnel congestion and recognizes that the recommendation of the most effective tunnel management strategies on a given day depends on the underlying causes of congestion.
Keywords: Tunnel traffic management; data-driven analysis; clustering; simulation.

Acknowledgements

This research has been co-funded by Google through the Better Cities program, the Swedish National Transport Administration, and the TRENOp Strategic Research Area. The authors would like to thank Trafik Stockholm for providing the data and for their valuable input.

1. Introduction

Freeway traffic management is becoming increasingly important as travel demand grows in many urban areas. Tunnel traffic management, in particular, is a challenging problem due to the need for safe conditions (e.g. smooth evacuation in case of incidents) in addition to high throughput. Hence, the criteria for efficient tunnel traffic management strategies may be more rigorous compared to freeway traffic management in general [1].

There are several causes of tunnel traffic congestion, related to demand or supply factors or a combination of both. For example, an increase in demand during peak hours exceeding the tunnel capacity will eventually result in congestion. Moreover, the existence of weaving areas inside the tunnel may generate bottlenecks. In more complex networks, traffic conditions affecting the tunnel may develop downstream or upstream the tunnel. Tunnel traffic conditions are also particularly sensitive to drivers’ behavior [2]. Consequently, adequate tunnel traffic management policies should be developed, which are robust with respect to the multiple traffic patterns that may develop into congestion.

A strategy often deployed in practice to deal with recurrent tunnel traffic congestion, for example by the State Road Traffic Authority (VLB) in Berlin [3], is entrance closure. The TOB (Tunnel Ortskern Ortsteil Britz) tunnel is part of the Berlin urban freeway ring and it is often congested during morning peak hours due to high demand that exceeds the tunnel capacity. According to the German guidelines for tunnel operations [4] the tunnel should be protected from periodic congestion by closing the upstream entrance, hence controlling the inflow. The closure times and durations are based on empirical analysis of flow and speed data from typical congested days. Another example is the 5 km long Södra Länken tunnel in Stockholm, Sweden. Due to high demand exceeding its capacity, congestion often occurs in the tunnel. Furthermore, queues that occur at a highly congested section of the freeway downstream the tunnel tend to spill back into the tunnel. In order to prevent traffic jams from forming inside the tunnel and ensure safety, the local traffic authorities have adopted a preventive strategy of temporarily closing some of the tunnel entrances in order to allow queues to dissipate.

There are a number of studies in the literature focusing on tunnel traffic management. In [5] an operational study of traffic flow in the Holland Tunnel between New York and New Jersey is presented. A bottleneck inside the tunnel caused congestion and capacity drops with shockwaves propagating backwards to the tunnel entrance. In order to maintain flow at high levels, a vehicle platooning approach was used. Specifically, shockwave speeds caused by the bottleneck were estimated from flow and density measurements and the average vehicle speeds at the bottleneck during peak and off-peak hours were determined. The information was then used to inform traffic operators about the optimum flow level that should be reached. Consequently, the target of the control strategy was to limit the number of entering vehicles such that the speed at the bottleneck was maintained above the critical threshold below which shockwaves developed. In order to achieve that, gaps were periodically
introduced into the traffic stream at the tunnel entrance; hence the headway of vehicles at the bottleneck was controlled. In [6] an experimental control system was deployed for the Lincoln tunnel to limit the inflow and prevent congestion by metering the traffic entering the tunnel. It was shown that density is an essential measure for the operator in order to increase throughput and maintain high speeds. An integrated dynamic traffic management system is evaluated in [7] for the Central Artery/Tunnel network in Boston. The main performance measures considered for the control system were the level of service, and the exposure to Carbon Monoxide (CO). Different control strategies were evaluated in the context of an incident scenario that caused congestion. The control system strategies included lane control signals, variable speed limit signs, ramp-metering and route guidance. One of the main findings was the usefulness of microscopic simulation in order to explicitly capture the interactions between drivers’ responses and traffic control. The study concluded that the integration of different strategies has to be optimized and that route guidance should be based on predicted, not instantaneous, traffic conditions. In [8] the authors used a simulation approach to evaluate different traffic management strategies for the congested Hsueh-Shan Tunnel in Taiwan. An incident was simulated and the traffic management aimed at preventing queue formation at the tunnel entrance. The strategies included combinations of ramp closure, ramp metering, route guidance, and lane control. Mesoscopic simulation was used for the evaluation. The results showed that ramp metering provided the best performance compared to the other strategies. In summary, a number of existing studies evaluate different strategies (e.g. ramp closure, metering, route guidance, etc.) to deal with tunnel congestion. Some of the studies focus on the selection of the most appropriate measurement variables and threshold values in order to achieve the best performance. However, very little work has been reported that looks at the causes of congestion inside the tunnel and how they may impact the strategy to be implemented on a given day. Depending on the cause of congestion, which may vary between days, different traffic management strategies may be most adequate. For this purpose, sufficient information including traffic patterns, geometry factors, etc., should be collected in order to observe variations in the demand patterns and potential bottleneck locations. The aim of this paper is to test the hypothesis that recurrent tunnel congestion may be generated by different sources and that tunnel management strategies deployed to mitigate the impacts depend on the underlying cause. Advances in sensor technologies enable the collection of data that can be used for the analysis of complex network dynamics and also the deployment of advanced tunnel management strategies. The proposed approach involves empirical analysis of the network dynamics and congestion patterns that lead to the need for intervention, and subsequent development and evaluation of adequate tunnel management strategies under recurrent traffic conditions that are responsive to the traffic conditions on a given day. A data-driven cluster analysis, utilizing extensive historical information of the network traffic states, is used to identify different congestion patterns and the underlying demand patterns that cause them. Subsequently, simulation analysis is deployed to evaluate alternative traffic management strategies for each of the different demand patterns identified in the first phase. The main advantage of simulation is that a range of factors related to the recommended...
strategies can be incorporated and tested thoroughly. Finally, on-line traffic measurements are used to classify a new day into one of the identified congestion patterns early, which allows the corresponding recommended traffic management strategy to be deployed in a timely manner.

Two main hypotheses are examined: a) different congestion patterns exist that generate the need for intervention; and, b) depending on the causes of congestion, different types of strategies are required.

Two main causes of tunnel congestion are investigated: a) congestion originates at bottlenecks downstream the tunnel and propagates upstream; and, b) congestion is triggered inside the tunnel due to weaving sections. Depending on the causes of congestion, the implications in the tunnel traffic conditions are expected to vary. Consequently, adequate tunnel management strategies are required that are adjusted to the different responses to deal with congestion, e.g. location and time of intervention. Hence, the recognition of the congestion patterns is considered important to increase the effectiveness of the different tunnel management strategies. This relationship has previously not been explicitly recognized in the literature.

The main questions addressed in the paper are thus:

- Are there multiple causes of tunnel congestion and how to identify them?
- To what degree do the underlying causes of congestion impact the strategy to be implemented in order to mitigate or even avoid congestion inside tunnels?
- Is it possible to identify early the congestion pattern and be more proactive in terms of the strategy to be implemented?

The paper is organized as follows. Section 2 introduces the proposed methodology identifying distinct tunnel traffic evolution patterns, the evaluation of different traffic management strategies through traffic simulation. An approach is also suggested that can be used on-line to classify the emerging traffic to one of the identified patterns. Section 3 presents a case study involving a congested tunnel in Stockholm, Sweden and discusses the scope of the application and the data available for the analysis. Section 4 presents the results and discusses the recommended tunnel traffic management strategies. Section 5 concludes the paper.

2. Methodology

Assuming adequate data on traffic conditions the approach to address the main research hypotheses includes the following steps:

1. Clustering of historical traffic data to identify distinct congestion evolution patterns.
2. Characterization of the underlying demand patterns and bottlenecks related to each observed congestion pattern.
3. Simulation-based evaluation to identify effective traffic management strategies for each identified congestion pattern scenario.
4. On-line classification using data as they become available to assign the evolving traffic conditions to one of the identified congestion patterns for timely recommendation of the appropriate traffic management strategy to be implemented.
Fig. 1 summarizes the main components of the approach. Each step is described in more detail in the following sections.

**Fig. 1** Analysis of tunnel congestion causes and evaluation of tunnel management strategies.

### 2.1 Data-driven congestion characterization

Variations in traffic congestion patterns may reveal different conditions that trigger tunnel congestion, and different control actions may be appropriate depending on the cause. The proposed data-driven analysis uses cluster analysis to categorize distinct congestion patterns from archived data based on their spatiotemporal characteristics. A more in-depth analysis then follows in order to distinguish the conditions that generate the congestion patterns (e.g. different locations and times of bottlenecks) based on two different hypotheses. The first is that congestion originates downstream the tunnel and propagates upstream and, the second is that the traffic conditions inside the tunnel generate congestion.
Cluster analysis

With the proliferation of large data sets, clustering is increasingly used in transportation studies [9]. Clustering requires a number of features that are characteristic of the traffic patterns. Possible traffic characteristics to use as features include speed, density, speed propagation measurements, etc.

Clustering methods belong in mainly two groups: hierarchical and partitional clustering. In hierarchical clustering, successive agglomerative or divisive procedures are followed, i.e. by merging smaller clusters or splitting larger ones. Appropriate clusters can be determined by visually analyzing a dendrogram that shows the clustering stages and the variation within the clusters. In partitional clustering (e.g. the k-means algorithm), the number of clusters is specified in advance, hence the clustering results in a unique partitioning for a given data set. The selection of the number of clusters is then based on the contribution of each additional cluster to some measure of performance (e.g. variability). For the purpose of this paper, Ward’s hierarchical clustering method ([9], [10]) is chosen for the categorization of the examined days. However, partitioning methods may also be applied, especially for very large data sets.

Each day of measurements is initially assigned to its own cluster and pairs of clusters are combined step by step until one cluster remains. The merging of clusters is done in such a way that the within-cluster variability is reduced. The adequate number of clusters can be decided using a dendrogram, which visualizes the variation within the clusters for different steps of the clustering procedure. In the first step of the clustering procedure, two days are combined into one cluster.

Let $y_{dlt}$ be the traffic measurement at location $l$ in time interval $t$ on day $d$, $L$ the total number of locations and $T$ the total number of time intervals used in the analysis. The cluster membership is assessed by calculating the within-cluster variability for all potential clusters:

$$\sum_{d \in C_k} \sum_{l=1}^L \sum_{t=1}^T (y_{dlt} - \bar{y}_{kt})^2,$$

where $C_k$ is the set of days in cluster $k$, and $\bar{y}_{kt}$ is the mean value across all days in cluster $k$.

The criterion for merging different clusters is the minimization of the increase in the total within-cluster variability after two clusters are joined, i.e. the sum of squared distances.

Given the hypotheses regarding the causes of tunnel congestion, the features for the cluster analysis should provide information about the traffic conditions downstream the tunnel and the tunnel itself. Hence, the features that are considered in this analysis consist of traffic measurements at critical locations downstream the tunnel at potential bottlenecks and inside the tunnel between merging and diverging segments.

Characterization of congestion patterns and bottleneck analysis

Following the cluster analysis, the distinct characteristics of the congestion patterns in each cluster are analyzed and explained in terms of traffic flow characteristics. The causes of tunnel congestion may be related to e.g. variations in the demand, the onset and the duration of congestion, different bottlenecks, etc.. The detailed analysis considers the spatiotemporal traffic patterns based on representative historical traffic information for each distinct congestion pattern identified by the cluster analysis. Kerner’s three-phase traffic theory [11] provides means to support this analysis. According to this theory, traffic patterns are
categorized in three phases: (a) free flow \((F)\), (b) synchronized flow \((S)\) and, (c) wide moving jam \((J)\). The criteria to distinguish between the synchronized flow and wide moving jam are based on qualitative empirical spatiotemporal features. In particular, wide moving jams propagate through any bottleneck and other traffic states keeping the mean downstream front velocity. On the other hand, the synchronized flow phase is usually fixed at the bottleneck. Transitions \(F \rightarrow S\) are expected to occur at the bottlenecks, while \(S \rightarrow J\) may occur later and further upstream the bottleneck location. In [12] the common features of Kerner’s three-phase theory are identified for traffic congestion patterns on freeways in USA, UK, and Germany. The theory is used to examine the two causes of tunnel congestion that are investigated as mentioned in Section 1.2. In particular, if the cause of congestion originates downstream the tunnel and propagates upstream, wide moving jams are expected to be observed. On the other hand, if congestion is triggered inside the tunnel due to weaving sections, the spatiotemporal features should indicate synchronized flow.

### 2.2 Simulation-based evaluation of tunnel traffic management strategies

The aim of the simulation analysis is to replicate the different traffic patterns leading to tunnel congestion, and to evaluate the effectiveness of various tunnel traffic management strategies under each demand pattern.

The main control strategies used in tunnel traffic management include entrance closure, metering, and combinations of metering and closure. The temporary closure of some tunnel entrances is a drastic traffic management strategy often used when congestion occurs inside a tunnel in order to reduce the inflow. The main criterion for the timing of the closure is the evolution of congestion based on traffic measurements (e.g. speeds and densities) at critical locations (e.g. heads of bottlenecks) along the tunnel. Different combinations of entrances to be closed may be investigated, based on the inflow. An obvious disadvantage associated with this strategy is the impact on the network due to the forced detouring of vehicles.

Metering has been used extensively for freeway traffic control ([13], [14]). Metering is typically applied on ramps, depending on the location of the bottlenecks and ramps’ storage capacity. Moreover, metering strategies can use either real-time traffic measurements or historical traffic information. This paper explores the possibility of using ramp metering as an alternative to the drastic action of temporal tunnel entrance closures, as preliminary evidence from prior studies suggests that it can be effective.

Metering can be implemented at different tunnel entrances that are expected to have an effect on the traffic conditions inside the tunnel, while the time and duration of the metering may be chosen such that the evolution of congestion inside the tunnel is avoided.

The following design aspects are considered important for the effectiveness of the different strategies, and adequate values or parameters associated may be obtained from the data-driven analysis:

- Timing (closure, metering)
- Duration (closure, metering)
- Location (i.e. tunnel entrances)
- Metering rate
2.3 On-line pattern classification for strategy recommendation

Proactive recommendation of the strategy to be implemented on a given day, before tunnel congestion has fully developed, can benefit from on the findings from the cluster analysis. The similarity of a new day with the identified types of days can be measured using a classification method. Early identification of the type of congestion pattern is critical so as to be proactive and able to recommend the most adequate strategy.

The classification starts from the same time interval as the historical cluster analysis, but only uses available measurements up to the most recent time interval $T_0 \leq T$, where $T$ is the last time interval used in the cluster analysis. Therefore, classification has to be based on partial information early on, and updated as more information is available. The same features are used as in the cluster analysis, i.e. same traffic metrics and measurement locations. A variant of the minimum-distance classifier (MDC) is used. The MDC method uses the distances between the features of the new day and the mean values of the features across all days in each cluster as the classification criterion. The cluster with the shortest distance is chosen for the classification. A more detailed description of MDC can be found in [15]. The Euclidean distance is chosen as the distance metric. Specifically, for each cluster, the average distance between the features of the new day and each day $d$ in the cluster is calculated as:

$$D_k = \frac{1}{N_k} \sum_{d \in C_k} \sqrt{\sum_{l=1}^{L} \sum_{t=1}^{T_0} (y_{lt}^d - y_{dt})^2},$$

where $y_{lt}^d$ is the measurement from the current day, at location $l$ in time interval $t$ on day $d$, $L$ is the total number of locations and $N_k$ is the number of days in cluster $k$. The new day is assigned to the cluster that corresponds to the lowest average distance.

The classification is updated, for example every 15 minutes, as new traffic measurements become available. Once the congestion pattern is reliably classified into a specific cluster, the most adequate strategy can be recommended. However, critical in this approach is the stability of the classification, i.e. the ability to make the current classification detection early on in the day.

3. Case study

The analyses steps presented in section 2 are applied to a tunnel on the Södra Länken (“The southern link”) freeway section in Stockholm, Sweden (see Fig. 2). The Södra Länken freeway is 6 km long, of which 4.7 km is in tunnels. It connects Essingeleden in Stockholm with Värmöleden (county road 222) in the Nacka municipality (see Fig. 2). In the westbound direction, Södra Länken merges with flow coming from the south (E4). The tunnel has in general two lanes in each direction and a speed limit of 70 km/h. There are six entry points and six exit points (including the mainline entrance and exit). Södra Länken was designed to serve 60,000-70,000 vehicles per day when it opened in 2004. Today, however, the daily traffic demand is 90,000-100,000 vehicles per day. The network of interest is equipped with radar detectors every 500 meters as part of the Motorway Control System (MCS), which report counts and speeds per minute and per lane (indicated by dots in Fig. 2).
Due to the high demand, in the tunnel and the downstream freeway, entrance closures are often used as the main traffic control strategy. Fig. 3a depicts the number of tunnel closures per year between 2005 and 2015, due to recurrent congestion only; other causes of closures, e.g. incidents, are excluded. As can be seen, there is an increasing trend in the number of closures over time. As Fig. 3b illustrates, the average closure time between 2005 and 2015 is about 45 minutes. The closures occur most commonly during the morning peak between 6 am and 8 am, and occasionally in the evening peak.
In order to prevent traffic jams from forming inside the tunnel and ensure safety, the local traffic authorities have adopted a preventive strategy of temporarily closing some of the tunnel entrances to allow the queues to dissipate. Before any tunnel closure takes place, warning messages are broadcast via radio, displayed on variable message signs (VMS) (Fig. 4) and posted on Stockholm’s traffic information website. The messages inform drivers about the congestion in the tunnel, but provide no indication about which alternative routes drivers should take.

![Map with VMS locations](image)

**Fig. 4** Variable message sign (VMS) locations.

The tunnel closure decision-making process is based on a combination of traffic measurements at predetermined locations. Real-time average speed measurements per minute are observed at the sensor locations, depicted as red rectangles on Fig. 5. The on- and off-ramps are depicted with arrows and the circles represent roundabouts. The most critical tunnel entrances and exits are also indicated. Based on certain conditions, including speed thresholds and a time window of observing low speed values, a warning for queues inside the tunnel is first given; if low speeds persist, a closure is triggered. The speed thresholds are 25 km/h for locations 1 and 4, and 20 km/h for locations 2 and 3 (see Fig. 5). A necessary condition for the tunnel to be closed is that speeds at location 4 drop below the corresponding threshold for a predefined time window (5 minutes).
The tunnel warnings and closures have significant impacts for travelers. During closures travelers experience long delays, either due to congestion or their attempts to find alternative routes through the surrounding urban network, which is not designed to handle large volumes of rerouted traffic. Thus, it is important to identify and analyze the reasons leading to tunnel congestion to be able to implement more effective and proactive tunnel traffic management.

4. Results and Analysis

The approach proposed in Section 2 is used to investigate the traffic patterns leading to tunnel congestion, evaluate the current tunnel closure policy and recommend alternative strategies to mitigate the negative impacts of the closures, and assess the possibility of early strategy recommendation on-line.

Historical traffic measurements from the MCS sensors at different locations inside and downstream the tunnel are used for the empirical analysis. More specifically, the westbound direction and the morning period are selected for the analysis. A set of 25 weekdays in April and October 2014, including days with and without tunnel closures, is selected. Days with incidents or sensor failure are excluded.

4.1 Data-driven congestion characterization

Cluster analysis

Clustering is used to distinguish the temporal and spatial variations among days in the data set and to potentially identify different causes of tunnel congestion. Speed measurements at the MCS sensor locations by time of day are used as features in the analysis. In particular, during the morning period measurements from $L = 25$ MCS sensor locations inside and downstream the tunnel are aggregated over 5-minute intervals for a time period starting at 6 am up to the earliest tunnel closure (7:15 am) that is observed among the days in the data set, in total $T = 15$ time intervals. This is done in order to avoid misclassification of days due to the traffic pattern changes caused by the tunnel closure.
Fig. 6 shows the dendrogram from the application of the hierarchical clustering method. The different days in the data set are presented as leaf nodes and are assigned to specific clusters indicated by clades. Three clusters are identified based on the similarity of the different days indicated by the distance between clusters reported on the y-axis. The selected clusters are numbered and marked by red rectangles in Fig. 6. Cluster 1 includes days without congestion inside the tunnel that require no intervention, while clusters 2 and 3 include days with congestion. On several occasions the congestion triggers warnings and/or closures. The main difference between clusters 2 and 3 is with respect to the time and space evolution of congestion. The differences are highlighted by the speed contour diagrams along the tunnel for select days from clusters 2 and 3 shown in Fig. 7. The boundaries of the space-time plots are the tunnel entrance from the east (entrance 1) and the end of the Södra Länken before the merging with E4 northbound (see Fig. 5).
Days in cluster 2 are characterized by congestion that is generated downstream the tunnel exit and spills back inside the tunnel. Multiple bottleneck locations can be identified inside the tunnel (see Fig. 7), probably generated by several weaving segments with consecutive off- and on-ramps. Between the weaving segments, vehicles may find space to accelerate; hence there are some regions with higher speeds inside the tunnel. The blank areas on the contour plots represent periods during which the tunnel was closed.

**Fig. 7** Speed contour plots for selected days in clusters 2 and 3.
The contour plots for days in cluster 3 reveal more severe congestion, which originates downstream the tunnel and spills back faster inside the tunnel compared to the patterns observed in cluster 2. Another difference is that congestion starts earlier compared to the days in cluster 2.

Characterization of congestion patterns and bottleneck analysis

In the case of Södra Länken, the traffic phases of Kerner’s theory can be identified in each of the two clusters. More specifically, in the speed contour plots in cluster 2, patterns with mainly the synchronized flow phase and some wide moving jams are identified, while wide moving jams are mainly observed in cluster 3. The synchronized flow phase is fixed at the different bottlenecks along the section, while moving jams are generated mainly downstream the tunnel exit and propagate upstream through other bottlenecks.

One day from each cluster is selected for further analysis. Fig. 8 shows the contour plots for the 9th of April 2014 (cluster 3) and 1st of April 2014 (cluster 2) based on speed observations in one minute intervals. It can be observed that on the 9th of April congestion starts at some location downstream the tunnel and propagates rapidly inside the tunnel. Moving jams are formed (marked on Fig. 8b) inside the tunnel at approximately the same mean speed as the downstream front. On April 1st some moving jams can be identified; however they are not as extended in space and time as on April 9th. Congestion inside the tunnel seems to be triggered by the downstream bottleneck, but develops mainly due to the several on- and off-ramps and associated weaving sections along the tunnel.

Fig. 8 Speed contours per minute interval for: (a) 9th and (b) 1st of April 2014.

Fig. 9 depicts three fixed bottlenecks generated inside and downstream the tunnel, based on the speed contours analysis. In particular, “Bottleneck 1” is triggered by the downstream conditions, on Essingeleden (E4 NB), in conjunction with the merging of vehicles coming from E4 northbound. The second cause of congestion inside the tunnel is associated with “Bottleneck 2” and “Bottleneck 3”. The heads of those two bottlenecks are generated at critical segments where weaving phenomena caused by on- and off-ramps are observed.
Congestion at those weaving segments may occur independently of the conditions downstream the tunnel exit. The most critical measurement locations that belong to these bottlenecks and are used by the current tunnel traffic management strategy are numbered in Fig. 9.

The results are consistent with the hypothesis that for days when congestion is mainly triggered by a downstream bottleneck (days in cluster 3) congestion inside the tunnel propagates rapidly across the multiple bottlenecks, while for other days (cluster 2) congestion is the result of the weaving segments along the tunnel and evolves independently of the downstream conditions.

![Diagram of Södra Länken freeway and important sensors](image)

**Fig. 9** Main bottlenecks on Södra Länken freeway and the locations of important sensors.

### 4.2 Simulation-based evaluation of alternative strategies

Based on the results of the cluster analysis, typical days for each cluster are simulated under alternative control strategies. The microscopic traffic simulation model TransModeler 4.0 [17] is used. The network consists of the main freeway section, including the tunnel, the downstream E4 freeway and all tunnel on-/off-ramps (Fig. 5); in total 57 links and 77 O-D pairs. The initial demand matrix was extracted from the Swedish national travel demand model SAMPERS [18] and adjusted to fit counts on certain locations for the typical days that are analyzed. The simulation period is 06:00 - 09:00 am.

During a tunnel closure, redistribution of demand is expected at the entrances that are closed. In the current tunnel traffic management approach, mostly entrances 1 and 3 are closed. The average number of vehicles that are diverted during the morning peak (see Fig. 5) is about 2000 veh/h for entrance 1 and 600 veh/h for entrance 3. In a parallel study [16] a detailed analysis of demand redistribution related to tunnel closures was conducted. The key findings
are that during closures vehicles take alternative paths to enter the tunnel from other entrances or take other routes to reach their destination. Part of the flow reduction at entrance 1 appears at entrance 2, where the average flow during closures increases by 40%. These redistribution patterns are used to develop the demand flows during closures.

The performance measure used for the evaluation of the different tunnel traffic management strategies is travel times for critical O-D pairs. For the purpose of this study the O-D pair originating from entrance 2 and ending downstream of entrance 6 (Fig. 5) is of interest since it captures the main conditions inside the tunnel. The distance of the O-D pair is 5.5 km.

Two closure days (1st and 9th of April 2014), representing the two identified types of congestion patterns, are simulated. The corresponding O-D matrices are generated based on observed vehicle counts from the two days. The two demand pattern scenarios are considered as the base scenarios used for the evaluation of the different tunnel management strategies. In the base scenarios no tunnel closure or any other strategy is utilized. Fig. 10 displays the simulated speed profiles for each demand scenario at the key locations (see Fig. 9) that are used by the existing logic to trigger warnings and closures.
Fig. 10 Simulated speeds inside the tunnel for April 1st (top) and April 9th (bottom).

For each day, the warning and closure times according to the current practice are identified. One of the conditions for closure is that congestion has reached section 4 (Fig. 5), where sensor 24 is located. As marked in Fig. 10, the warning period for April 1st is triggered between 7:03 - 7:09 am, while the conditions for closure are first triggered around 7:35 - 7:40 am. For April 9th the warning period is triggered between 6:54 - 6:59 am, and the closure between 7:07 – 7:15 am. Based on the speed conditions of the base scenarios, different simulation scenarios are created representing alternative implementations of control strategies with respect to the main design factors identified in Section 2.2.

Closure strategies

Three main entrances are included in the experimental design as candidates for closure. Specifically, closures of tunnel entrances 1, 3 (also used in the current closure logic), and 4
are considered. Closing either entrance 1 or entrance 4 is expected to significantly reduce the mainstream tunnel flow due to their high inflow. Different scenarios with respect to time and duration of closure are examined and the best performing ones are discussed in more detail.

Different closure start times are evaluated based on the simulated traffic conditions inside the tunnel, including the start time according to the current practice. According to current practice, tunnel closure should be triggered around 7:30 am for April 1st and around 7:00 am for April 9th (see Fig. 10). The selection of other closure times is based on the conditions downstream and inside the tunnel, specifically the times at which congestion reaches sensors 28 and 34 inside the tunnel, and sensors 42 and 43 downstream the tunnel. Different closure durations are also investigated; the results here focus on the scenarios with 30 minutes duration as they indicate the best performance. Table 1 summarizes the design of the evaluated closure strategies. The closure times that correspond to current practice are presented as the ‘current tool’ scenario.

**Table 1** Simulated closure strategies.

<table>
<thead>
<tr>
<th>Closure strategy</th>
<th>April 1st</th>
<th>April 9th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time (hh:mm)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>06:30</td>
<td>07:00</td>
</tr>
<tr>
<td></td>
<td>07:30 (current tool)</td>
<td>06:30</td>
</tr>
<tr>
<td>Entrances 1&amp;3</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Entrances 1&amp;4</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

For each closure strategy, the O-D matrix is adjusted according to the observations discussed in Section 3.3.

- When entrance 1 is closed, most vehicles upstream the entrance follow the alternative path to the city (Fig. 9). However, some vehicles use entrance 2 to enter the tunnel, increasing its inflow by 40%.
- When entrance 4 is closed, 40% of the affected flow is redistributed to entrances 6 and 7.

Fig. 11 illustrates the performance in terms of travel times of the different closure strategies, including the base case of no action.
The closure strategies show significant reduction in travel times during and after the tunnel closure compared to the base scenario without closure. In particular, significant improvements are achieved if one of the two main entrances (1 or 4) is closed. For April 1st, the best performance is observed when the closure begins at the onset of congestion inside the tunnel, i.e. at 6:30 am. Notable reductions in travel times are observed which are maintained during the whole simulation period, in contrast to strategies involving later closures such as the current practice.

For the second demand pattern (April 9th), early closures at the onset of congestion downstream do not provide the same improvement for the whole simulation period. As observed in Fig. 11b, for the strategies with closures at 6:30 am and 6:45 am, travel times start increasing again after a while. The optimum closure time is 7:00 am, which coincides with the time that the warning is triggered in the current practice.
The observed differences between the best tunnel closure strategies for the two demand scenarios may be explained by the distinct traffic patterns they represent. In particular, for April 1st, a tunnel closure before the weavingle bottlenecks are generated alleviates or even prevents the evolution of congestion locally. For April 9th, where the main sources of congestion inside the tunnel are the downstream bottlenecks, the temporary tunnel closure at the onset of downstream bottlenecks does not give the expected improvement. One may expect that the propagation of the downstream bottleneck is slower as less flow is approaching the bottleneck, but the strategy does not control the congestion at the head of the downstream bottleneck. Hence, after the tunnel re-opens congestion starts propagating backwards. Therefore, a closure at the time when congestion has reached the tunnel provides overall better performance. Nevertheless, in all cases, early closures at the onset of congestion lead to faster congestion recovery, sustained and for longer periods, compared to implementing later control actions.

**Metering strategies**

The entrances to implement metering were selected based on their inflow and location. One strategy involves the metering at the main tunnel entrance (entrance 1) in combination with metering at an on-ramp after the tunnel exit (entrance 6), which belongs to bottleneck 1 (see Fig. 9) and has on average 1200 veh/h inflow during peak hours. The metering of entrance 4 that belongs to bottleneck 3 is also examined in combination with two on-ramps (2 and 6) that have the lowest inflow compared to the main entrance. Another scenario involves metering of only those on-ramps.

The selection of metering timing is based on the time when congestion occurs on each day. It is expected that if metering is implemented early in the morning, right before the onset of congestion downstream or inside the tunnel, the propagation of congestion farther inside the tunnel is prevented. At most of the entrances that are controlled, metering starts at 6:30. Different metering durations are also investigated. Metering on entrance 1 is implemented for 30 or 45 minutes. Secondary entrances (2, 4 and 6) are metered for longer periods, between 75 and 120 minutes. Investigation of the optimal design of the metering strategy is out of the scope of this paper. Hence, simple metering implementations using fixed rates are used. The maximum allowable flow during metering is 1200 veh/h for entrance 1 and 900 veh/h for entrances 2, 4 and 6. Table 2 summarizes the different metering scenarios indicating the regulated entrances and the corresponding time period when metering is active.

<table>
<thead>
<tr>
<th>Table 2 Simulated metering strategies.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td><strong>April 1st</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td><strong>April 9th</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

The OD travel times for both days and metering strategies are shown in Fig. 12. Traffic conditions are improved for the strategies with metering at the main entrances 1 or 4,
combined with metering at downstream entrances, for both simulated days. The hypothesis that early metering will prevent congestion inside the tunnel is confirmed by the results. Regarding the metering strategy for the two secondary entrances (2 and 6), the OD travel times for both days show some improvement compared to the base scenario, although not very significant. Controlling the on-ramps with the highest inflow provides greater benefits than metering secondary entrances.

Compared to the results for the closure strategies, the obvious benefit of the metering is that vehicles are allowed inside the tunnel; the flow is smoother and the propagation of congestion is avoided for the whole simulation period. Metering seems to successfully decrease or prevent congestion, independently of the congestion patterns, i.e. whether originating from a downstream bottleneck or weaving segments inside the tunnel. However, slightly better performance is observed for April 9th since metering of the downstream entrance (entrance 6) contributes to the improvement of the traffic conditions near the downstream bottleneck. In combination with metering of the upstream entrances, congestion is prevented inside the tunnel.

**Fig. 12** Travel times for best metering strategies for: (a) April 1st and (b) April 9th.
Combined closure and metering strategies

The experiments with combined closure and metering strategies involve the closure and metering of secondary entrances only. It is investigated whether closure of the main entrances can be avoided by closing only one entrance with low inflow and metering another. Hence, closure is examined at entrance 3 at the head of bottleneck 3 and metering is applied to entrance 6. Different closure times are examined; however, the best time for each simulated day corresponds to the time when congestion starts propagating at bottleneck 3. If the closure occurs earlier the effects of the closure will only be temporary and congestion will form again. The closure duration is varied between 15 and 30 minutes. On-ramp metering starts at 6:30 am for April 9th and at 6:45 am for April 1st. The duration of metering is varied between 90 and 120 minutes. Table 3 summarizes the combined closure and metering strategies.

Table 3: Simulated combined strategies.

<table>
<thead>
<tr>
<th>April 1&lt;sup&gt;st&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>April 9&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 13 shows the OD travel times for the two demand scenarios. For both days, travel times are reduced compared to the base scenario, independently of the closure duration. Slightly larger impacts are observed for April 9<sup>th</sup> where metering of entrance 6 provides significant improvement. The combined closure and metering strategy shows promising performance as significant travel time reductions can be achieved without closing entrances with high inflow (e.g. entrance 1), but rather secondary entrances with relatively low flow.
Fig. 13 Travel times for best metering and closure scenarios (a) April 1st, and (b) April 9th.

Discussion

Among the three types of strategies investigated, the most promising category overall, for both demand patterns, is metering. First, the tunnel throughput capacity can be more effectively controlled compared to the temporary closure policy. Most importantly, the impacts of the drastic closure policy on the network (e.g. detouring of vehicles) can be mitigated or even avoided. Moreover, the flexible metering scheme can be deployed for different traffic conditions and needs (e.g. based time-of-day, demand levels, duration, etc.).

The closure strategy can improve traffic conditions inside the tunnel; however, the effectiveness of the strategy depends on the selection of appropriate time and duration of the closure, as well as on the congestion pattern. The impact of tunnel closures is found higher for the demand pattern where congestion is generated due to bottlenecks inside the tunnel (April 1st) compared to the pattern where congestion is caused by downstream bottlenecks (April 9th).

In the combined closure and metering strategy, overall better traffic conditions are achieved for April 9th, since the entrances to which metering and closure were implemented were closer to the downstream bottlenecks. The combined strategy can be an alternative to the closure of two entrances where fewer vehicles are affected by the closure.
The measures of effectiveness (MOE) of the best design for each of the strategies are summarized in Table 4. The OD travel times are presented for the base scenario and the percentage travel time reduction resulting from each strategy is reported for comparison.

**Table 4** MOE comparison of the best strategies.

<table>
<thead>
<tr>
<th>MOE</th>
<th>Base Travel times (min)</th>
<th>Best closure % reduction</th>
<th>Best metering % reduction</th>
<th>Best combined % reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>April 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>April 9&lt;sup&gt;th&lt;/sup&gt;</td>
<td>April 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>April 9&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>07:00</td>
<td>12</td>
<td>13</td>
<td>50</td>
<td>46</td>
</tr>
<tr>
<td>07:15</td>
<td>13</td>
<td>15</td>
<td>46</td>
<td>53</td>
</tr>
<tr>
<td>07:30</td>
<td>17</td>
<td>20</td>
<td>59</td>
<td>50</td>
</tr>
<tr>
<td>07:45</td>
<td>17</td>
<td>22</td>
<td>53</td>
<td>45</td>
</tr>
<tr>
<td>08:00</td>
<td>19</td>
<td>19</td>
<td>63</td>
<td>47</td>
</tr>
<tr>
<td>08:15</td>
<td>16</td>
<td>18</td>
<td>63</td>
<td>44</td>
</tr>
<tr>
<td>08:30</td>
<td>13</td>
<td>16</td>
<td>54</td>
<td>44</td>
</tr>
</tbody>
</table>

It is noticeable that for both days the highest overall improvement is obtained by the metering strategy. The closure strategy provides the second highest improvement for April 1<sup>st</sup>; however, for April 9<sup>th</sup> the combined strategy is overall more effective.

The effectiveness of the different strategies can also be observed in the speed contours along the tunnel. Fig. 14 presents the plots for the two days and for the base, best closure and best metering scenarios. Compared to the base case, the benefits of the metering strategy distinguish compared to the closure strategy. In particular, for both days, the traffic congestion inside the tunnel is alleviated with metering.
4.3 On-line classification for strategy recommendation

The identification of different traffic congestion patterns from historical data using cluster analysis enables the on-line classification of the emerging congestion patterns. For the classification, the same traffic data (variables, locations and time interval lengths) as in the cluster analysis are used. The classification uses data starting at 6 am and ending at $t^0$, where $t^0$ is the current time, increased in 15-minute increments from 6:15 up to 7:15.
The minimum distance based classifier discussed in Section 2.3 is used to identify the emerging traffic patterns on-line and hence, be more proactive in terms of the tunnel traffic management strategies to be used. In order to identify how early the unknown traffic pattern can be predicted, classification was applied using data as they would arrive at the traffic control center. The classification was assumed to take place every 15 minutes, each time using only the data available up to that point. The results are summarized in Table 5.

Table 5 Classification results.

<table>
<thead>
<tr>
<th>Observation intervals</th>
<th>April 4th (closure)</th>
<th>April 8th (closure)</th>
<th>April 29th (closure)</th>
<th>May 8th (closure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:00 – 6:15</td>
<td>Cluster 1</td>
<td>Cluster 1</td>
<td>Cluster 3</td>
<td>Cluster 1</td>
</tr>
<tr>
<td>6:00 – 6:30</td>
<td>Cluster 1</td>
<td>Cluster 1</td>
<td>Cluster 3</td>
<td>Cluster 1</td>
</tr>
<tr>
<td>6:00 – 6:45</td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>6:00 – 7:00</td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>6:00 – 7:15</td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 2</td>
<td>Cluster 2</td>
</tr>
</tbody>
</table>

On April 4th no congestion developed inside the tunnel. The day is correctly classified as an uncongested type of day (cluster 1) from the earliest up to the latest measurement periods. April 29th is classified early as the type of day where congestion originates at downstream bottlenecks (cluster 3). This classification remains consistent at all subsequent times. Three tunnel entrances, instead of two were closed on that day, between 7:00 – 7:50 am, indicating more severe congestion compared to other days.

April 8th and May 8th are two days with closures starting at 7:50 and 8:20, respectively. Initially, both days are assigned to cluster 1, indicating very low or no congestion. However, when more observations become available (e.g. after 45 minutes) the days are assigned to cluster 2 representing days with congestion mainly triggered by bottlenecks inside the tunnel.

For days with tunnel congestion caused by traffic bottlenecks starting downstream the tunnel, the classification is able to predict early the congestion pattern that develops, before congestion has propagated far upstream inside the tunnel. For the distinction between days with mainly upstream congestion (cluster 2) and other days with lower or no congestion (cluster 1), larger periods of observation are needed (i.e. 45 min) to accurately predict how the day will behave. Finally, in some cases the classification results may indicate that the new day is similar to more than one traffic patterns.

5. Conclusion

The paper presented an empirical analysis to observe and explain the causes of tunnel traffic congestion using a data-driven approach and implications for tunnel management. Simulation-based analysis is used to evaluate different tunnel management strategies under different congestion patterns. Cluster analysis is used to identify distinct traffic evolution patterns leading to tunnel congestion. Subsequently, simulation analysis is deployed to design effective traffic management strategies for the identified traffic patterns. On-line early classification, of a new day into one of the identified traffic patterns, which allows the corresponding recommended traffic management strategy to be proactively deployed, is very important. An important contribution of the paper is the recognition that the most effective control strategy depends on the underlying causes of congestion inside the tunnel. This relationship has previously not been explicitly recognized in the literature.
The approach was used for the study of the Södra Länken tunnel in Stockholm, Sweden. The clustering analysis indicates at least two distinct traffic congestion patterns, which differ with respect to the onset and the propagation of congestion. On some days, congestion is triggered by downstream bottlenecks and propagates rapidly inside the tunnel generating wide moving jams. On other days, the downstream congestion spills back slower, while congestion inside the tunnel mainly develops due to the weaving segments inside the tunnel. The simulation results indicate that closures should start at the onset of congestion, not later, in order to have the greatest effect. Moreover, in order to prevent congestion in the tunnel, it is necessary to give higher importance to the measurement locations at the head of identified bottlenecks inside the tunnel compared to waiting for congestion to reach the far upstream locations, as the current logic does. Among the strategies that were assessed, metering was found to be the most promising strategy. The main advantage of metering is that impacts on the surrounding network due to vehicles detouring can be avoided.

The case study further examined how early the evolution of congestion on a new day can be predicted on-line, so as to be more proactive in terms of the intervention that is needed. The classification method used for this purpose proved to be quite robust for the days examined. The results indicate that days with congestion originating downstream the tunnel can be predicted early at the onset of congestion, whereas the identification of days with congestion mainly triggered by bottlenecks inside the tunnel requires observations over a longer period.

The proposed methodology is generally applicable to freeway tunnels or networks, assuming that sufficient traffic information is available. In particular, in order to understand the traffic dynamics in the network and identify the traffic congestion patterns through data-driven analysis, historical traffic information should be available at critical locations and for a representative set of days. Of course, the need for intervention is network specific. Moreover, the requirements for the implementation of the investigated traffic management strategies vary depending on the strategy. For the entrance closure strategy it is necessary that alternative routes exist to divert the traffic. With respect to the metering strategy, the storage capacity of the on- or off-ramps that are metered should be sufficient in order to avoid queue spill back on the freeway mainline or on other roads. Further, more advanced metering methods may be used that take into account and adapt to on-line information of the traffic conditions (e.g. densities).

References


