VTT Technical Research Centre of Finland

Estimating the risk of traffic incidents using causal analysis

Kilpi, Jorma; Scholliers, Johan; Koskinen, Sami; Kuusela, Pirkko

Published in:
Proceedings of TRA2020, the 8th Transport Research Arena

Published: 12/03/2020

Document Version
Publisher's final version

Rethinking transport
27–30 April 2020

TRA 2020 HELSINKI

Proceedings of 8th Transport Research Arena TRA 2020, April 27-30, 2020, Helsinki, Finland

Estimating the risk of traffic incidents using causal analysis

Jorma Kilpi[a]*, Johan Scholliers[b], Sami Koskinen[b], Pirkko Kuusela[a]

[a]*VTT Technical Research Centre of Finland Ltd, Tekniikantie 1, Espoo, FI-02044 VTT, Finland*
[b]*VTT Technical Research Centre of Finland, Visiokatu 4, Tampere, 33720, Finland*

Abstract

Traffic management centres aim to keep the traffic flowing regardless of traffic disturbances. This paper presents a new method for quantitative incident analysis via causality and observations. It is applied to estimate the risk of vehicular traffic tunnel closures in the city of Tampere, Finland, where the tunnel on a national main road bypasses the city centre. A tunnel closure rapidly causes traffic jams on alternative routes in the city. Also, traffic incidents near the tunnel may propagate and cause tunnel closures. We restrict our analyses to the westbound direction of the traffic on the main road. We combine various open data sources providing information about traffic and driving conditions. The analysis is based on a probabilistic and statistical framework augmented with causal reasoning. We have identified several event paths from congestion after the tunnel to the tunnel closing, as well as approximate capacity limits near selected critical locations.

Keywords: Traffic Management Centre (TMC); big data; analysis of causality, traffic incident detection, traffic tunnels, congestion.

* Corresponding author. Tel.: +358-040-719-5134;
E-mail address: jorma.kilpi@vtt.fi

1. Introduction

An urban Traffic Management Centre (TMC) aims to keep the traffic in the city fluent. Many different sensors in the city – such as loop detectors at traffic lights, permanent traffic counters in main arterials, floating car data from different fleets, and traffic cameras – provide information that can be used to detect traffic incidents. Traffic incidents at critical locations, e.g. a major crossing or arterial, cause congestions that may spread rapidly throughout the city.

Tampere, with 230 000 inhabitants, is the largest inland city in the Nordic countries. It is located on the isthmus between two large lakes. A 2.4 km long tunnel (Rantatunneli) on the Rantaväylä (Road 12) along the isthmus bypasses the city centre. The national authorities manage the tunnel and the Road 12, whereas the urban TMC manages the traffic in the city. If there is a considerable risk of traffic jams in the tunnel, national authority restricts access to the tunnel, which may cause congestion in alternative routes in the city centre of Tampere. Also, incidents on the part of Road 12 between Vaitinaro and the tunnel's end easily cause the closing of the tunnel in the Westbound direction.

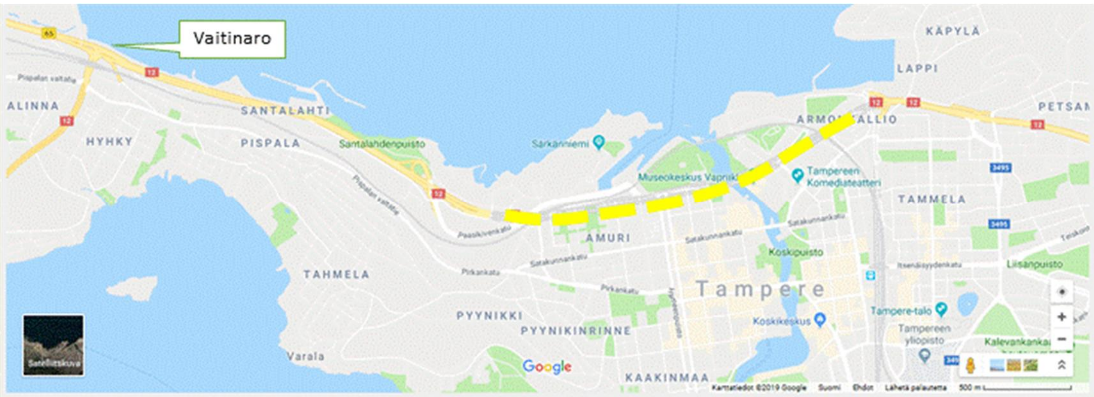

Fig. 1 The tunnel (dashed yellow) and Road 12 (orange) in Tampere

The EU-funded TransformingTransport (TT) project studied how big data and related methods can benefit transport sector. The Integrated Urban Mobility pilot of the project demonstrated the potential of automated processing of big data sources to improve situational awareness and to provide information to travellers. In the framework of the TT project, a model was developed to estimate the risk of tunnel closures in Tampere.

2. Methods

In this paper, we restrict our analyses to the westbound direction of the traffic on *Rantaväylä* (Road 12) and focus on the *Vaitinaro intersection* illustrated in Fig. 1 and in Fig. 3 as a black circle with label *tre802.* In the analyses, we adopt a probabilistic and statistical framework augmented with causal reasoning following Pearl, (2000).

We assume that, to some extent, it is possible to predict an increased risk of a tunnel closing event if the possible causes of such events are either i) observable before the event or ii) the causes are distinguishable and coexist with the event. Distinguishable means that the cause need not be directly observable but it is possible to distinguish it with data and sufficient background information. The sufficiency of the background information requires human knowledge and comprehension. A methodology to accommodate the background information is presented in Section 2.1.

The time granularity of the measured data is limited and it can contain both systematic and random errors as well as outliers. Therefore, it is an open question whether it is possible to deduce the causes of tunnel closing events from the observed data. Even in the positive case, one can ask whether it is possible to model and quantify the risk of the tunnel closing event. This depends on what kind of data is observed and what can be deduced from it. The available data is described in Section 2.2.

2.1. Background information

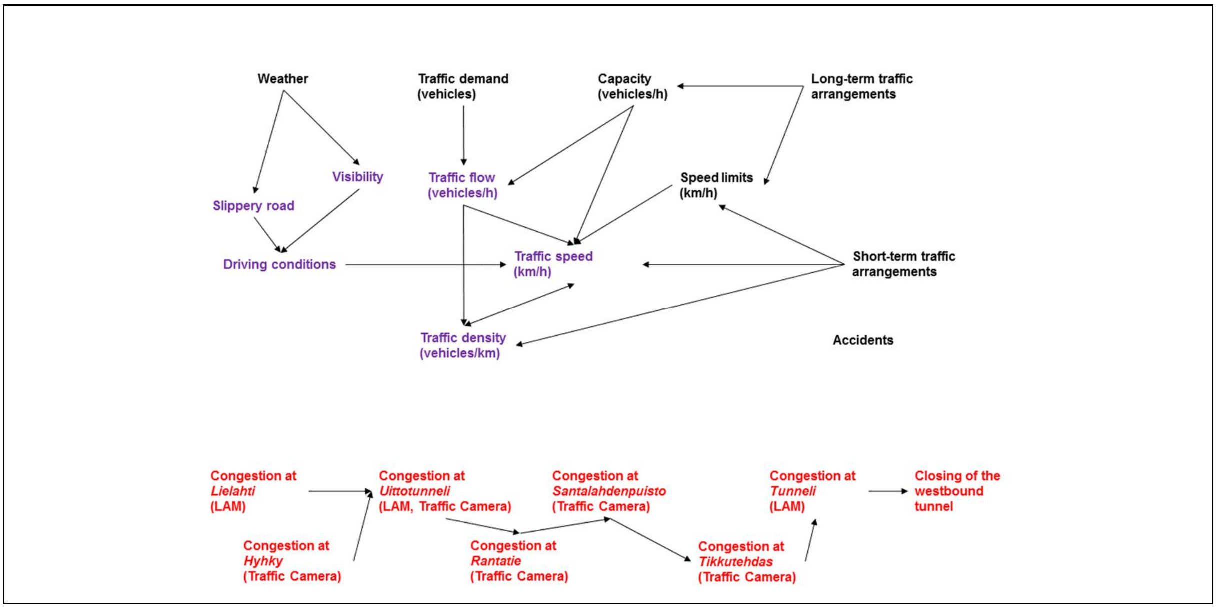

Fig. 2 Assumed causal relationships

Fig. 2 illustrates two *causal graphs* related to the problem. A causal graph is a *directed acyclic graph (DAG)*, i.e. a directed graph without cycles, in which a directed arrow illustrates a causal relationship. The upper graph is between the variables that are measured (purple) and some given factors (black). The lower graph is between the events that are observed. The variable traffic density is used to determine congestion, so the two graphs are closely related.

For example, the graph's *slippery road* is influenced by *weather*, whereas *driving conditions* are influenced by *slippery road* and *visibility*. *Driving conditions* influence *traffic speed*. In the lower graph, congestion at Uittotunneli may propagate backwards against the traffic flow and finally lead to tunnel closing. The upper graph of Fig. 2 represents one of the DAGs to be considered between the variables and it is assumed to be valid *before* a congestion causes problems. For example, if traffic density exceeds some threshold level, there can be feedback to the traffic speed and this situation may require a more specific DAG model. The *accidents* node is included without arrows on purpose, indicating that accidents are noticed in the data analysis process.

Our research task is to determine if congestion at or near Vaitinaro intersection is causing (via congestion at road Rantaväylä) the closing of the tunnel in the westbound direction. The causal graph indicates that there are confounding variables, i.e., variables that influence on the congestion at Vaitinaro and the tunnel closing event. If we fail to control for the confounders, the association between the congestion at Vaitinaro and tunnel closing is distorted, resulting to biased data analysis results. The variables in black and purple in Fig. 2 are confounding, but the influence on the observed congestion (large traffic density) takes place via driving conditions and speed, traffic flow and short-term traffic arrangements, unless there is an accident that immediately requires closing the tunnel. We need to control for these confounding variables and detect such accident events.

Drawing simple graphs like Fig. 2 is valuable in many ways. First, it communicates our assumptions and background information on the variables under which the data analysis part is performed. Second, the graph brings up the confounding variables. But most importantly, the graph allows applying a back-door criterion that determines what variables need to be controlled (Pearl, 2000). Third, the graph determines if the available data is sufficient to estimate the targeted causal relation. Thus, one should study the causal graphs and ask if there needs to be additional variables or arrows indicating causal relationship, or if variables or arrows can be removed. If the causal graphs become more complicated, one needs to assess if the amount of data is still sufficient.

2.2. Description of data

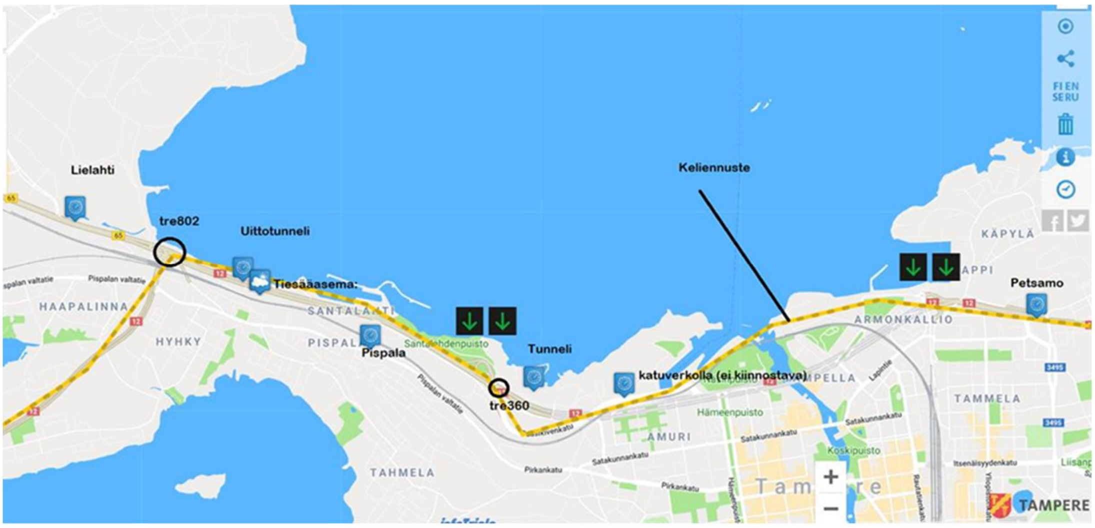

Fig. 3 The locations of the data measurement points

2.2.1. LAM data from the road 12

The acronym LAM means an automated traffic measurement point that is maintained by national authorities. There are several LAMs on the road 12 (from West to East): *Lielahti*, *Uittotunneli*, *Tunneli* and *Petsamo*. *Petsamo* is located near the east entry of the tunnel. The main advantage is that LAM data contains estimates for *the average traffic speed (km/h)*. Also, *traffic flow* values *(vehicles/h)* are available. The traffic flow divided by the traffic speed gives an estimate of *traffic density (vehicles/km)* that we consider very useful for this research problem. All these values are provided over 5–15 min intervals.

2.2.2. Traffic light data at Vaitinaro intersction (tre802)

The traffic light data is in the traffic flow format (*vehicles/h*), including also a daily profile. Since the LAM data covers the Road 65 direction towards west from *Vaitinaro* intersection, we use traffic light data of this intersection mainly for the southwest direction of Road 12. In this direction, we can divide the traffic flow value by the speed limit in order to obtain a traffic density estimate. This represents the traffic density immediately after the intersection.

2.2.3. Traffic camera data

There are about ten traffic cameras in the vicinity of the tunnel, depending on how far from the tunnel the traffic situation is being studied. Snapshot images retrieved every two minutes from these cameras were automatically processed by a neural network based classifier developed at VTT, see Kilpi, Koskinen, & Scholliers, (2019). It classifies traffic camera images into three categories: congested, fluent, or empty. The main purpose of the classifier is to support human operators in a TMC, suggesting high-priority camera feeds to follow. However, even if intended for operator support, the classification can produce additional information also for statistical analysis.

The accuracy for classifying single images may temporarily drop due to e.g. lighting conditions, and therefore the output cannot be trusted in the same way as other sensor data. The classification accuracy is approximately 95% in the current development phase. However, consecutive detections of congestion from the same camera give a strong indication on problems near the tunnel – or tunnel closure. Additionally, the tool enables saving snapshots showing congestion for visual inspection of the actual situation near the tunnel.

We combine LAM data based traffic density estimates and the camera-based congestion data. The origin of these two data sets are completely disjoint types of sensors, hence their information content is complementary of each other. Since they measure the same quantity, i.e., a momentary amount of vehicles on the portion of the road, their joint outcome is more informative than either of the data sources is alone.

2.2.4. Weather data

There is a road weather station (*Tiesääasema* in Fig. 3) very close to the *Vaitinaro* intersection. Its location is optimal and, therefore, the weather data is considered reliable and representative. It provides information about air temperature, road temperature, road condition (wet, dry, snowy, icy, …) and rain type (light, heavy, snow, sunshine, …). During the late springtime after March 18th, 2019, and until August 11th, 2019, the weather conditions have not caused too poor driving conditions although both the air and the road temperatures have been fluctuating around 0º Celsius and the road surface has been wet and icy.

2.3. Related work

Detection of *congestion propagation patterns* from data has been an active research subject during the recent years. The terminology is not fixed and sometimes the phrase *cascading pattern* is used with the same meaning. Usually, the congestion under the study is propagating in the backward direction, against the traffic flow, and the pattern has spatiotemporal dimensions both as a phenomenon and as a property of the data that is used for the analyses. Causal reasoning is naturally present in these studies.

The work of Nguyen et al. (2016) uses travel time data and introduces some explicit terminology, especially the concept of *causal congestion tree (CCT)*, that they use to describe the propagation pattern of a congestion. In our study case the CCT is very simple, nevertheless we plan to use the CCT concept more explicitly in a further study of our traffic tunnel case. In Xiong et al. (2018) the term *propagation graph* is used, but it seems to be essentially the same concept as the CCT. The work of He et al. (2018) is data mining oriented and it introduces useful spatiotemporal definitions. The work of Liang et al. (2017) reminds that while the observed traffic congestion is a continuous phenomenon, the data is discrete. This is a serious problem in our task since we would like to detect signs of serious congestion before the tunnel closing event occurs. The recent work of Di et al. (2019) is done in a much larger spatial scale in comparison to this work. Cascade effects caused by physical incidents is a topic in Basak et al. (2019). The approach to enumerate traffic congestion patterns in Inoue et al. (2016) is also data-mining oriented.

Methodologically relevant studies are the book of Pearl (2000), and the ways to apply it in the context of traffic related problems Queen and Albers (2009) and Blondel et al. (2017). In a further study, we aim to understand the most frequent reasons and root causes of the tunnel closing events and, if successful, study interventions and counterfactuals in these cases.

3. Case studies of tunnel-closing events

The data collection period for this research started on March 18th, 2019, and until August 11th, 2019, over 50 tunnel-closing events have occurred. Of these events, over 40 times the tunnel was closed towards the westbound direction. Time series views provide fast insights to the cases, as a possible direct cause must precede the closing event in quite a short time window in the data.

3.1. Sunday March 24th, 2019

One of the two lanes towards *Lielahti* at *Paasikiventie* road, near *Uittotunneli*, was temporarily out-of-use due to construction and measurement work. This apparently caused three short closings at 14:16, 15:12 and 15:31. The cause was reported at the local newspaper (https://www.aamulehti.fi/a/d2731b43-7d27-4dcc-b9f8-5ea1da5c1f05). Since we know the cause of this event, it is reasonable to look at the possible signs in the data before the event.

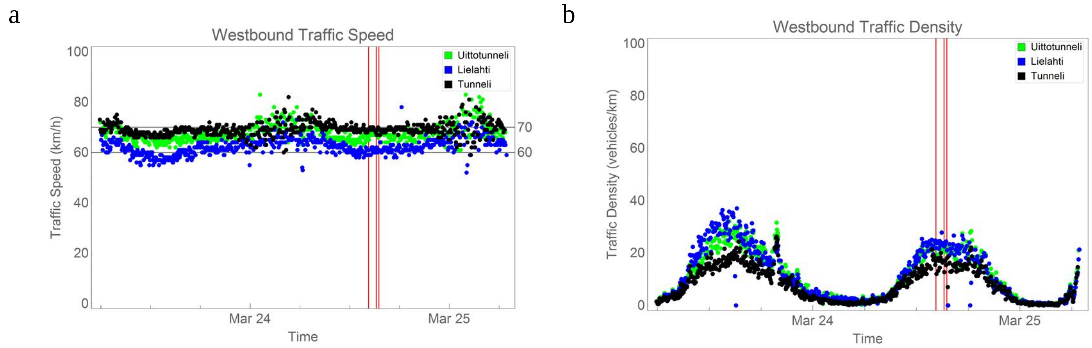

Fig. 4 (a) Traffic speeds during the case in March; (b) Traffic densities during the case in March

In the average speeds, shown in Fig. 4 (a), there is practically no signs of any problems prior to the event nor during the events. In Fig. 4 (b) there may be some prior signs in *Lielahti*, if one knows and takes into account the closed lane on *Paasikiventie*. Such a knowledge is not included in the LAM data. Traffic camera data from May, shown in Fig. 5 (a) shows that some of the cameras have captured congestion slightly before the first event time. The coding of the camera-based classes in Fig. 5 is 2 = congested, 1 = fluent, and 0 = empty.

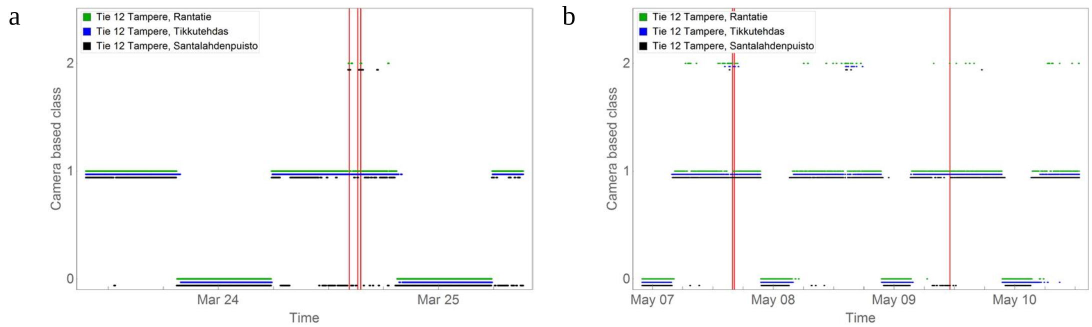

Fig. 5 (a) Traffic camera snapshot from March; (b) Traffic camera snapshot from May

3.2. *Tuesday May 7th, 2019*

On May 7th, the tunnel was closed twice: at 15:53 for 6 minutes and at 16:15 for 7 minutes. We consider these as the same event since it is reasonable to assume that they had the same root cause.

In this case, there were prior signs visible both in the LAM data and in the traffic camera data. The westbound traffic speed drops clearly before the first event at *Lielahti* and at *Uittotunneli*, see Fig. 6 (a). At *Lielahti*, there are some drops in the traffic speed also at other times. Traffic density increases also, Fig. 6 (b). However, the next two days have similar signs at the same time in the afternoon without a closing event.

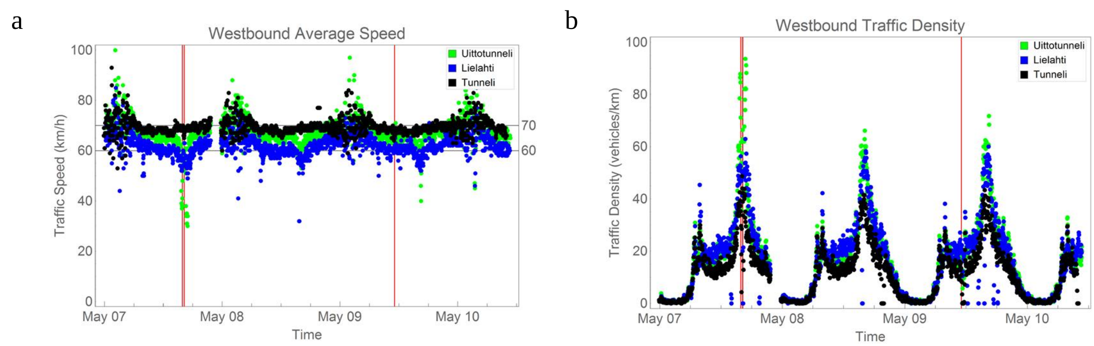

Fig. 6 (a) Average speeds during the two cases in May; (b) Traffic densities during the two cases in May

3.3. Thursday May 9th, 2019

According to the local newspaper (https://www.aamulehti.fi/a/4dafb2a6-4d1c-4dfa-b14d-cd62d67edf3a) the tunnel was closed on Thursday, May 9th, at 11:06 for 33 minutes due to an accident in the tunnel. The rightmost red vertical line in the plots of Fig. 6 and in Fig. 5 (b) indicate this event. Accidents are non-predictable events. The few speed drops of the *Tunneli* data (black dots) indicate that the accident occurred first and the tunnel was closed after the accident. If there is a delay between an accident and the tunnel-closing decision due to the accident, then signs of the closing event may be visible in the data before the event.

3.4. Conclusions from the case studies

We conclude the following from the visual detection of the time series of traffic speeds and densities:

1. In the case 1, prior signs were not visible in the LAM data although we expected that there would be prior signs. If the additional knowledge of the out-of-use lane were used, for example, by proper scaling of traffic density then the signs may become visible. Traffic camera image based classification indicated congestion and this shows the value of combining data from disjoint type of sensors.
2. In the case 2, prior signs were clearly present before the event. However, equal signs occur also other times without an associated closing event.
3. In the case 3, prior signs were not truly prior signs but the tunnel-closing event occurred after the accident. The accident may have caused some signs of congestion before the tunnel closing occurred.

These cases provide examples of different paths to the tunnel closing and hence issues that need to be considered in the statistical analysis of the data.

4. Capacity of Rantaväylä (Road 12) towards west and south-west

In this section we visualize and quantify the capacity of the westbound direction between the tunnel and *Vaitinaro* intersection or *Lielahti*. The amount of closing events captured so far is still too small for statistical analysis. Luckily, the amount of all data so far is sufficient to estimate capacity constraints at Road 12 after the tunnel.

4.1. Scatter plots of traffic intensities vs. the average speeds

In order to correctly interpret traffic densities, i.e, the ratio of *traffic flow/average speed,* it is feasible to look at the fundamental diagrams, in this case scatter plots of *traffic speed vs traffic flow,* to see their mutual dependence. The same value of the ratio can occur with different value pairs.

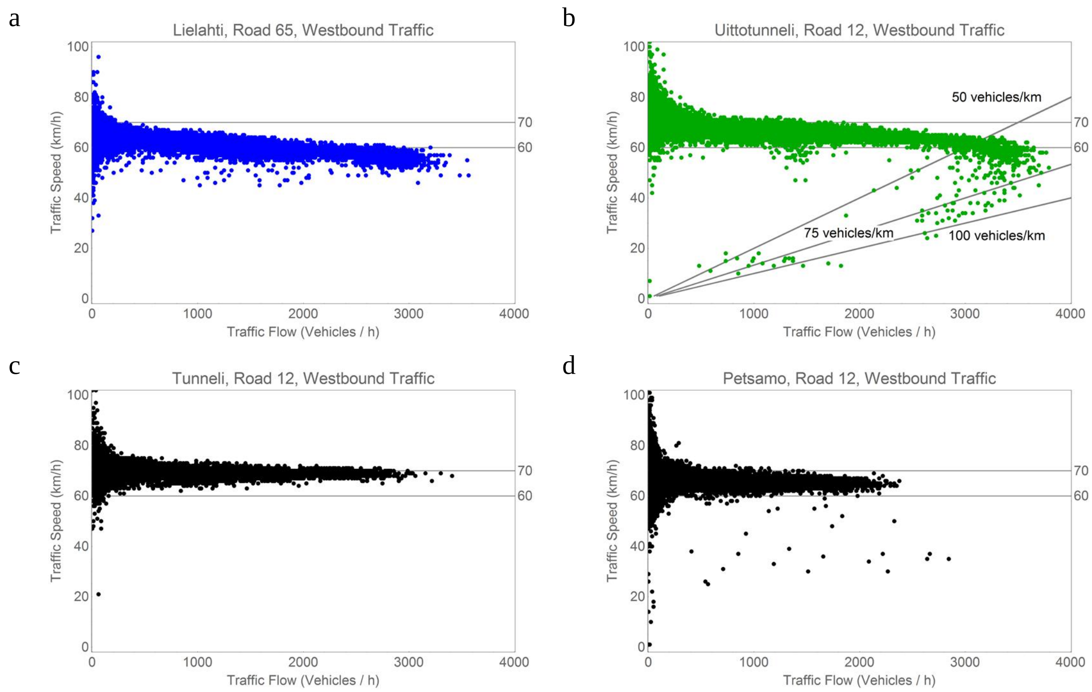

Fig. 7 Four fundamental diagrams of traffic speed vs. traffic flow. (a) Lielahti; (b) Uittotunneli; (c) Tunneli; (d) Petsamo;

The scatter plots of traffic intensity against the average speeds are shown in Fig. 7. The tunnel closing is a pre-emptive action, which can be confirmed when comparing the scatter plots of *Tunneli* and *Petsamo* against the two other data, *Uittotunneli* and *Lielahti*. The traffic must flow in and out of the tunnel steadily. On the other hand, the scatter plots of *Uittotunneli* and *Lielahti* have a shape that suggest that large traffic intensities at these locations indicate slower average speeds. The *Petsamo* should be almost identical with *Tunneli*, since practically the same vehicles are observed just before and right after the tunnel. The driving in the tunnel may have the effect of increasing the driving speeds.

In Fig. 7 (b), *Uittotunneli*, there are three diagonal lines where the traffic density is constant (the same lines apply to the other figures as well). The data suggests that if the traffic density at that measurement point is below 50 vehicles/km, there are hardly any problems. If the traffic density exceeds 75 vehicles/km, there is almost surely a congestion. Visual inspection of the traffic camera images and counting the vehicles in them gives similar numbers. The maximum observed traffic flow value at *Uittotunneli* was 4 140 vehicles/hour which is likely an outlier, all other values are smaller than 3 792 vehicles/hour.

The drops in the average speed of westbound traffic observed at *Lielahti*, see Fig. 6 (a), may be caused by congestion and, furthermore, the congestion may be due to insufficient capacity. Moreover, the drops in the traffic speed at *Lielahti* propagate backwards to *Rantaväylä*, since many such drops occur simultaneously at *Uittotunneli*.

Several, but not all, of the individual observations that are visible in Fig. 7 (b) and (d) that are associated with low traffic speed are related to some of the closing events. Usually afterwards, that is, the prior signs of the closing events are much harder to find from the observed data.

4.2. LAM traffic density vs. traffic camera congestion classes

Since traffic camera image classification uses an ordinal level scale, it is important to consider different ways to combine it with LAM traffic density. First, correlation in time is visualized in Fig. 8 by setting the camera image classification *empty* to mean 5 vehicles/km, *fluent* to mean 30 vehicles/km and *congested* to mean 50 vehicles/km. This non-linear choice of scale is adopted from the information of Fig. 7 (b).

Traffic camera analysis output has a coarser scale than LAM data, but it is obtained more frequently. The number of vehicles in a single image is more random since it represents an instant of time, while LAM data represents a time interval. It could happen, that several vehicles occur in the same image even if the traffic density is low or moderate, or the number of vehicles in a snapshot can be low even if the traffic density is moderate over a longer period. It seems therefore reasonable to combine a few consecutive images together with a simple majority rule: If most of the images from a single camera show congestion during a time interval, detect a congestion at that interval. This is the content of Fig. 8: camera-based information is collected over 10-minute intervals and this information is compared with the LAM traffic density averaged over the same 10-minute intervals.

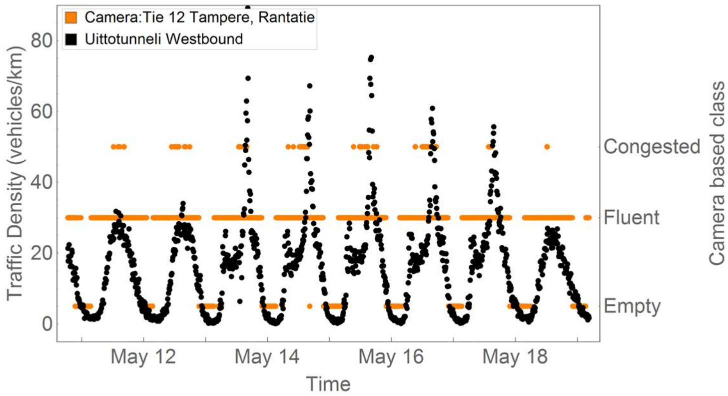

Fig. 8 Comparison of traffic density metrics

There is also a possibility of spatial combination of traffic image classifications: if two or more cameras represent the same view, from different angles of view, detect congestion, if more than one camera indicates congestion. It seems usual that camera views overlap somewhat and utilization of this overlap is feasible.

5. Results

The data collection period for this research started on March 18th, 2019, and lasted until August 11th, 2019. Over 50 tunnel-closing events had occurred. Of these events, over 40 times the tunnel was closed towards the westbound direction. From case studies we have identified several event paths from congestion to the tunnel closing. Three case studies was shown Section 3. Using the assumptions in Fig. 2 we have observed, for example, the path

Short-term traffic arrangements → Congestion at *Rantatie* → Congestion at *Santalahden puisto* → Closing of the westbound tunnel,

and the path

Congestion at *Lielahti* → Congestion at *Uittotunneli* → Congestion at *Tunneli* → Closing of the westbound tunnel.

The DAGs of Fig. 2 have evolved during the case studies indicating that we have learned and understood more during the case analysis. It is expected that the DAGs will still change when the analysis continues and our understanding of the possible cause's increases.

Data analysis has identified approximate capacity limits at *Lielahti* and at *Uittotunneli*. Early threshold levels for congestion at *Lielahti* and *Uittotunneli* can be estimated from the LAM data. Congestion alone may not cause closing of the tunnel, but some features of congestion may be used as indicators of increased risk, or probability, of a tunnel closing event. It is important to have enough data from non-closing and closing events given congestion.

We have studied methods to make image analysis comparable with the LAM data. Estimates of traffic density can be obtained from both data sources. Combining the image classification data with the traffic density estimates obtained from LAM data turned out to be a very fruitful approach. They offer two complementary ways to measure the congestion and, therefore, combining their information can be expected to detect early signs of congestion and quantify the risk of a tunnel closing due to congestion.

Acknowledgements

This study was funded by the European Union’s Horizon 2020 project Transforming Transport, grant agreement No 731932.

References

Basak, S., Ayman, A., Laszka, A., Dubey, A., & Leao, B. (2019). Data-Driven Detection of Anomalies and Cascading Failures in Traffic Networks. *Proceedings of the Annual Conference of the PHM Society*, *11*(1). https://doi.org/10.36001/PHMCONF.2019.V11I1.861

Blondel, G., Arias, M., & Gavaldà, R. (2017). Identifiability and transportability in dynamic causal networks. *International Journal of Data Science and Analytics*, *3*(2), 131–147. https://doi.org/10.1007/s41060-016-0028-8

Di, X., Xiao, Y., Zhu, C., Deng, Y., Zhao, Q., & Rao, W. (2019, August 8). *Traffic Congestion Prediction by Spatiotemporal Propagation Patterns*. 298–303. https://doi.org/10.1109/mdm.2019.00-45

He, Y., Wang, L., Fang, Y., & Li, Y. (2018). Discovering Congestion Propagation Patterns by Co-location Pattern Mining. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11268 LNCS*, 46–55. https://doi.org/10.1007/978-3-030-01298-4_5

Inoue, R., Miyashita, A., & Sugita, M. (2016). Mining spatio-Temporal patterns of congested traffic in urban areas from traffic sensor data. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 731–736. https://doi.org/10.1109/ITSC.2016.7795635

Kilpi, J., Koskinen, S., & Scholliers, J. (2019). Detection of anomalies in urban traffic from open data. *13th ITS European Congress, Brainport, the Netherlands, 3-6 June 2019*. ERTICO - ITS Europe.

Liang, Y., Jiang, Z., & Zheng, Y. (2017). Inferring Traffic Cascading Patterns. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, *2017-November*. https://doi.org/10.1145/3139958.3139960

Nguyen, H., Liu, W., & Chen, F. (2016). Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data. *IEEE Transactions on Big Data*, *3*(2), 169–180. https://doi.org/10.1109/tbdata.2016.2587669

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge University Press.

Queen, C. M., & Albers, C. J. (2009). Intervention and causality: Forecasting traffic flows using a dynamic bayesian network. *Journal of the American Statistical Association*, *104*(486), 669–681. https://doi.org/10.1198/jasa.2009.0042

Xiong, H., Vahedian, A., Zhou, X., Li, Y., & Luo, J. (2018). Predicting traffic congestion propagation patterns: A propagation graph approach. *IWCTS 2018 - Proceedings of the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, 60–69. https://doi.org/10.1145/3283207.3283213