Data-driven Methods for Identifying Travel Conditions Based on Traffic and Weather Characteristics

Georgia Ayfantopoulou[1], Evangelos Mintsis[1], Zisis Maleas[1], Evangelos Mitsakis[1], Josep Maria Salanova Grau[1], Vassilis Mizaras[1], Panagiotis Tzenos[1]
[1] Hellenic Institute of Transport, Centre for Research and Technology Hellas, Thessaloniki, Greece

Abstract

Accurate and reliable traffic state estimation is essential for the identification of congested areas and bottleneck locations. It enables the quantification of congestion characteristics, such as intensity, duration, reliability, and spreading which are indispensable for the deployment of appropriate traffic management plans that can efficiently ameliorate congestion problems. Similarly, it is important to categorize known congestion patterns throughout a long period of time, so that corresponding traffic simulation models can be built for the investigation of the performance of different traffic management plans. This study conducts cluster analysis to identify days with similar travel conditions and congestion patterns. To this end, travel, traffic and weather data from the Smart Mobility Living Lab of Thessaloniki, Greece is used. Representative days per cluster are determined to facilitate the development of traffic simulation models that typify average traffic conditions within each cluster. Moreover, spatio-temporal matrices are developed to illustrate time-varying traffic conditions along different routes for the representative days. Results indicate that the proposed clustering technique can produce valid classification of days in groups with common characteristics, and that spatio-temporal matrices enable the development of traffic management plans which encompass routing information for competing routes in the city of Thessaloniki.

Keywords: cluster analysis, congestion identification, traffic data, travel conditions, weather data

1 Introduction

Traffic congestion has huge societal, environmental and economic implications. Hence, congestion identification is critical for the development of efficient traffic control methods and management plans that can alleviate the aforementioned adverse impacts. Several studies have proposed methods for traffic state estimation that make use of different data types. A speed performance index (SPI) that accounts for traffic flow speeds and road speed limits was proposed for the identification of congestion patterns on urban freeways [Che21]. Floating car data (FCD) were used for the estimation of congestion on large scale road networks and travel times within congestion zones in the form of time-varying Travel Time Indexes [Erd21]. Model-driven and data-driven methods were combined for traffic state estimation [Shi21]. A data fusion methodology that considers different sensor types was developed for accurate traffic state estimation [Gen22]. Vehicle trajectory data were used for investigating the possibility that turning movements at intersections constitute bottlenecks while accounting for congestion spreading [Wei22]. This study adopted data-driven methods to identify travel conditions and develop corresponding traffic simulation models and traffic management plans. It applies spectral clustering to group days of similar traffic and weather characteristics, identifies representative days per cluster for the development of traffic simulation models, and estimates spatio-temporal matrices (STMs) for alternative routes (all representative days are considered).

2 Methodology

Cluster analysis was conducted to group days with similar travel conditions. The cluster analysis method implemented in the context of this study was based on the principles proposed by FHWA for identifying travel conditions [Wun19]. In specific, spectral clustering was applied to group days based on similar traffic and weather characteristics [Lux07]. Spectral clustering performs a low-dimension embedding of the affinity matrix between samples, followed by k-means clustering of the components of the eigenvectors in the low dimensional space. It is especially computationally efficient when the affinity matrix is sparse and the algebraic multigrid (AMG) solver is used for the eigenvalue problem. The version of spectral clustering adopted by this study requires the number of clusters to be specified in advance and works well for a small number of clusters. Additionally, it receives as input a pseudo random number generator used for the initialization of the eigenvectors decomposition and the k-means initialization, and the strategy for assigning labels in the embedding space (k-means). The adopted spectral clustering technique examines different possible outcomes in terms of cluster numbers. The number of clusters that provides a good description of the data and prevents over-fitting is selected with the use of the elbow method.

Subsequently, representative days per cluster are determined based on the relevant methodology proposed by FHWA [Wun19]. Representative days exemplify system dynamics within a specific cluster of days, since they exhibit minimum difference between observed time-variant performance measures (pertaining to the representative day) and mean time-dependent observed measures (considering all days in the cluster). Therefore, development and calibration of traffic simulation models is feasible based on observed rather than synthetic traffic data which are characterized by the unrealistic smoothing of time dynamic performance measures. The identification process of representative days encompasses the following steps:

1. Estimation of the average time-variant speed for each 15-minute time interval across all days in the cluster for multiple routes.

2. Estimation of the difference between the average speed and the speed observed on a particular day (expressed as a percentage of the mean value).

3. Selection of the individual day that minimizes the difference between the individual day and the average speeds considering all routes.

Moreover, speed measures (15-minute time interval) are also used to generate spatio-temporal matrices (STMs) for multiple routes of the examined network and each representative day. STMs facilitate the automatic identification of traffic congestion and bottlenecks [Hal16]. To this end, cut-off speeds that indicate transitions among free-flow traffic, mild congestion and severe congestion are estimated as a function of free-flow speeds (FFS) on a cluster basis. Specifically, 2/3 of FFS is considered the threshold between free-flow traffic and mild congestion, while 1/3 of FFS is considered the threshold between mild and severe congestion. FFS is picked as the 90th percentile speed from the cumulative density function (CDF) of speed measures pertaining to all days in a cluster. At last, speed measures from the representative days are compared with average cut-off speeds (considering all days in each cluster) for the generation of STMs. The illustration of congestion patterns on STMs enables the development of specific traffic management plans (e.g. routing information in the case of alternative paths) that can effectively mitigate the identified congestion issues.

3 Results and Discussion

The city center of Thessaloniki, Greece was selected as the study site. FCD, Bluetooth data, and weather data collected from the Smart Mobility Living Lab of Thessaloniki were used for the identification of travel conditions [Sal18]. The following data (in 15-minute time interval) was input to the adopted spectral clustering algorithm:

- Number of taxi rides encompassing at least one passenger,

- Number of vehicles detected via Bluetooth sensors,

- Speed measures along multiple routes or the city center of Thessaloniki,

- Mean temperature,

- Precipitation probability.

The clustering analysis considered 365 days (year 2019) and the study period was selected between 07:00 am – 12:00 pm. The raw data were processed in a format that renders each day as the dependent variable, while the aforementioned characteristics are the explanatory variables the characterize it. The average value of the 10-nearest neighbors was used for data imputation purposes. To obtain the optimal number of clusters the elbow curve is estimated (Figure 1). According to the elbow curve, the case of 7 clusters is a cut-off point beyond which the addition of an extra cluster does not provide a significantly improved explanation of the variation in the dataset. However, for simplicity reasons the case of 4 clusters is presented in this study, which still exhibits a good performance in terms of explained variation. A comprehensive description of each cluster is provided below:

1. Typical weekdays with 18.7 °C average temperature, 4 % precipitation probability, 26 % increase of taxi rides and 5 % reduction of average speed on multiple routes compared to the base cluster.

2. Typical weekdays with 18.7 °C average temperature, 17 % precipitation probability, 38 % increase of taxi rides and 14 % reduction of average speed on multiple routes compared to the base cluster.

3. Weekends and holidays with 17.8 °C average temperature, 4 % precipitation probability, very few taxi rides and free-flow traffic conditions (Base Cluster).

4. Typical weekdays with 13.65 °C average temperature, 48 % precipitation probability, 53 % increase of taxi rides and 26 % reduction of average speed on multiple routes compared to the base cluster.

Figure 2 depicts the allocation of days (per month) in the aforementioned clusters. It can be observed that summer (August), Christmas (December), and Easter (April) holidays are grouped in the same cluster (No. 2) with weekends. Typical weekdays with good weather and moderate-to-heavy traffic (Cluster No. 0 & 1) appear mainly between March – September, while typical weekdays with low temperatures, increased precipitation probabilities and heavy traffic (Cluster No. 3) appear during the winter months. Days belonging to Cluster No. -1 were omitted from the analysis due to reduced availability of traffic data.

Finally, representative days were identified and STMs were created for specific routes with the use of speed measures from the latter days. Figure 3 shows STMs for two competing routes crossing the city center of Thessaloniki (Tsimiski St. & Egnatia St.) and the representative day of Cluster No. 0. In the 3rd and 4th columns of the STMs appear the cut-off speeds that dictate transitions among different traffic states (free-flow traffic, mild and severe congestion). According to the STMs, Tsimiski St. is rather congested during the AM peak period, while traffic flows more efficiently along Egnatia St. for the same time period. Thus, personalized or generic routing advice can be provided to vehicular traffic upstream of the east entrances to the city center. Accordingly, congestion issues can be identified on other routes and appropriate traffic management plans can be developed.

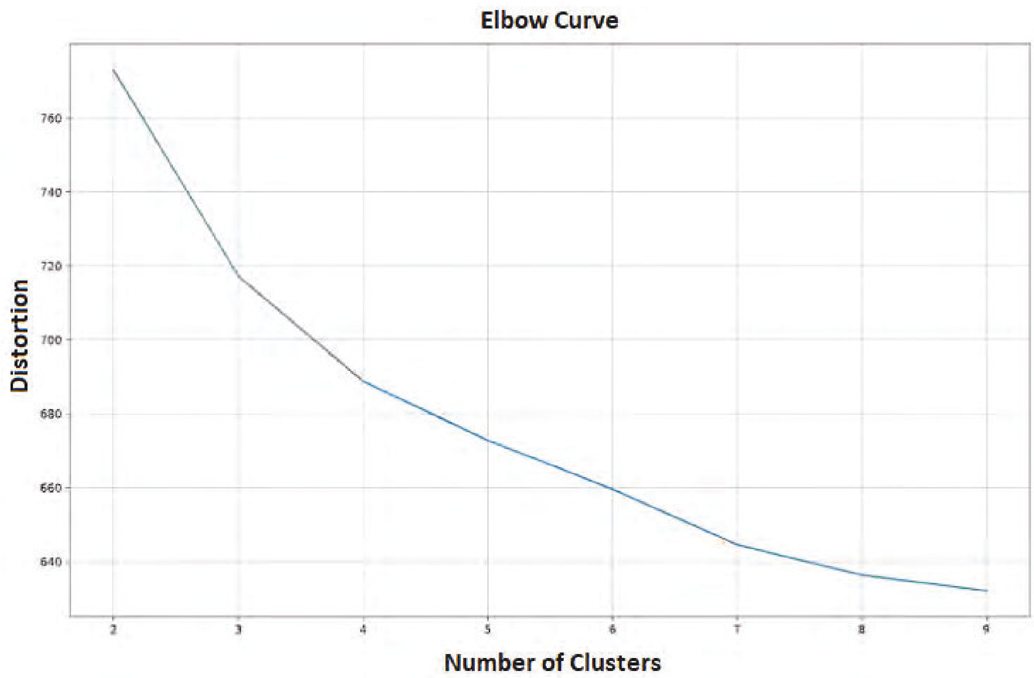

Figure 1: Distortion values for different number of clustes (elbow curve).

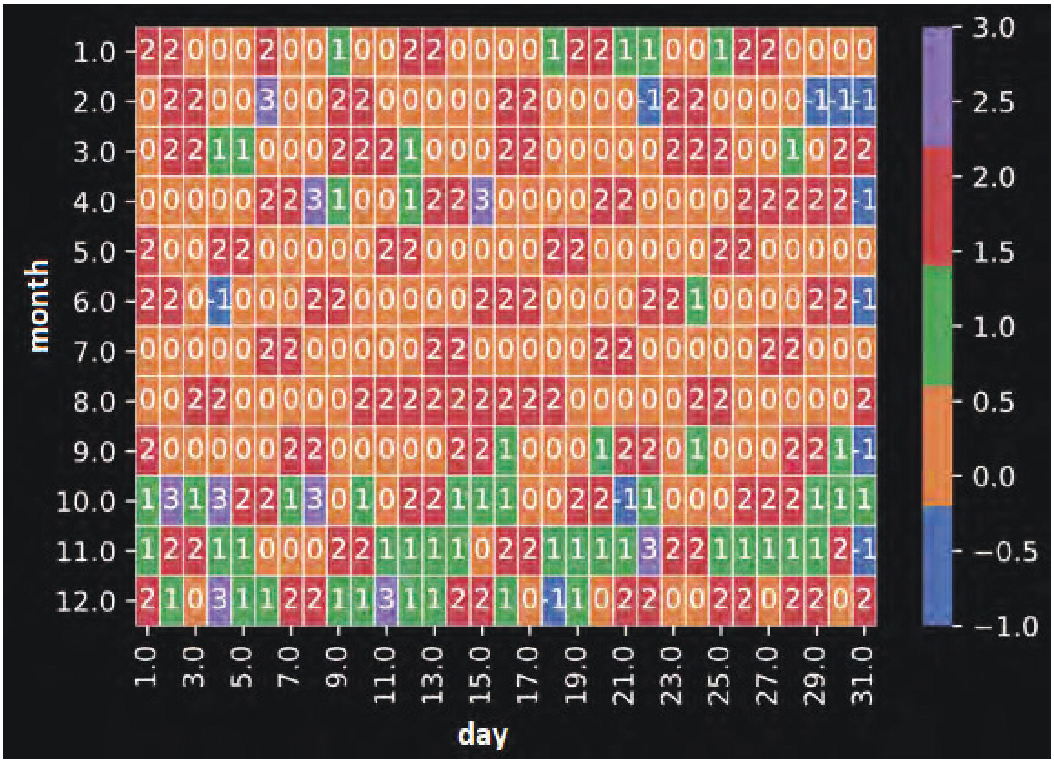

Figure 2: Allocation of days (per month) to different clusters (case of 4 clusters).

Road Name	Link	Upstream Node	Downstream Node	Upper Limit	Lower Limit	07:00	07:15	07:30	07:45	08:00	08:15	08:30	08:45	09:00	09:15	09:30	09:45	10:00
Tsimiski	23265	XANTH	E. Aminis	35	17		46	33	29	25	20	29	26	26	31	28	33	28
	23127	E. Aminis	P. Mela	33	17		36	37	19	18	16	23	17	19	27	25	22	29
	22875	P. Mela	P.P. Germanou	34	17	48	32	22	11	9	10	8	13	17	15	15	12	20
	8414	P.P. Germanou	A. Sofias	35	17	42	34	28	7	3	15	15	13	20	22	24	21	22
	22603	A. Sofias	E. Venizelou	30	15	35	26	17	15	14	16	16	19	17	18	17	16	19
	22385	E. Venizelou	I. Dragoumi	26	13		6		26		25	31	16	23	15	16	21	17
	22329	I. Dragoumi	Dodekanisou	25	12	20	19	10	16	19	26	29	20	21	18	22	20	24
	22181	Dodekanisou	N. Limnou	35	18		39	9	29	30	30	30	29	26	31	28	22	40
	21973	N. Limnou	26 Oktovriou	38	19	32	49	48	36	44	34	35	37	37	39	33	38	33
Egnatia_Westbound	23309	E. Aminis	L. Iasonidou	27	14	31	40	30	29	30	24	25	32	21	27	25	17	15
	23099	L. Iasonidou	A. Sofias	31	15	41	35	30	34	32	31	28	30	31	34	31	31	29
	22818	A. Sofias	E. Venizelou	25	13	30	27	23	28	21	27	25	23	23	19	24	23	27
	22512	E. Venizelou	I. Dragoumi	25	13	22	25	19	15	17	33	21	14	18	15	21	21	28
	22443	I. Dragoumi	Antigonidon	29	14	43	32	38	34	30	27	37	23	30	27	28	33	29
	22295	Antigonidon	Dodekanisou	27	14	44	22	28	20	26	22	22	22	17	28	25	22	27

Figure 3: STMs for two competing routes in the city of Thessaloniki, Greece.

4 Conclusions

This study conducted a spectral clustering analysis to categorize days in groups that exhibit similar traffic behavior and weather characteristics. The elbow method was used to determine the optimal number of clusters, and results are presented for the case of 4 clusters. It can be seen that the spectral clustering algorithm groups days in an accurate and meaningful manner. Representative days are identified per cluster, and STMs are developed for alternative routes (all representative days are considered). The identification of representative days will be used for the development of traffic simulation models that are typical of average traffic dynamics within each cluster, while STMs enable the development of traffic management plans that encompass routing advice (generic or personalized) and travel time information along alternative routes. Moreover, the results of this analysis will be used for the development of trigger-based traffic signal management strategies. In the future, traffic count, incident and micromobility data will be combined with the data of this study to achieve an improved classification of days in groups with similar characteristics. The inclusion of incident data in the aforementioned analysis will enable the identification of safety related issues and the development of relevant real-time traffic management strategies that could be conveyed to multiple road users via robust connectivity technologies. Safety-related data from the deployment of micromobility modes in the city of Thessaloniki will enrich the results of the clustering analysis and provide the opportunity to develop traffic management strategies that account for multimodality. Finally, data from related to micromobility modes will facilitate the identification of patterns pertaining to trip making and mode choice.

Acknowledgments

This study was supported by the Greek Public Investments Plan under project “Intelligent Traffic Management System - Integration of Traffic Control Centres” and the EIT Urban Mobility Initiative under project “RideSafeUM”.

References

[Che21] Y. Chen, C. Chen, Q. Wu, J. Ma, G. Zhang, and J. Milton: “Spatial-temporal traffic congestion identification and correlation extraction using floating car data”. In: *Journal of Intelligent Transportation Systems* 25.3 (2021), pages 263–280. ISSN: 1547-2442. DOI: 10.1080/15472450.2020.1790364.

[Erd21] T. Erdelić, T. Carić, M. Erdelić, L. Tišljarić, A. Turković, and N. Jelušić: “Estimating congestion zones and travel time indexes based on the floating car data”. In: *Computers, Environment and Urban Systems* 87 (May 2021). ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2021.101604.

[Gen22] A. GENSER, N. HAUTLE, M. MAKRIDIS, and A. KOUVELAS: “An Experimental Urban Case Study with Various Data Sources and a Model for Traffic Estimation”. In: *Sensors* 22.1 (Aug. 2022). ISSN: 1424-8220. DOI: 10.3390/s22010144.

[Hal16] D. HALE, R. JAGANNATHAN, M. XYNTARAKIS, P. SU, X. JIANG, J. MA, J. HU, and C. KRAUSE: *Traffic bottlenecks: identification and Solutions*. Technical report No. FHWA-HRT-16-064). United States. Federal Highway Administration. Office of Operations Research and Development., 2016.

[Lux07] U. von LUXBURG: “A tutorial on spectral clustering”. In: *Statistics and Computing* 17.4 (Aug. 2007), pages 395–416. ISSN: 1573-1375. DOI: 10.1007/s11222-007-9033-z.

[Sal18] J. M. SALANOVA GRAU, E. MITSAKIS, P. TZENOS, I. STAMOS, L. SELMI, and G. AYFANTOPOULOU: “Multisource Data Framework for Road Traffic State Estimation”. In: *Journal of Advanced Transportation* 2018 (June 2018), pages 1–9. ISSN: 0197-6729. DOI: 10.1155/2018/9078547.

[Shi21] R. SHI, Z. MO, K. HUANG, X. DI, and Q. DU: *Physics-Informed Deep Learning for Traffic State Estimation*. 2021. URL: http://arxiv.org/pdf/2101.06580.

[Wei22] L. WEI, P. CHEN, Y. MEI, and Y. WANG: “Turn-level network traffic bottleneck identification using vehicle trajectory data”. In: *Transportation Research Part C: Emerging Technologies* 140 (July 2022). ISSN: 0968-090X. DOI: 10.1016/j.trc.2022.103707.

[Wun19] K. E. WUNDERLICH, M. VASUDEVAN, and P. WANG: *TAT volume III: guidelines for applying traffic microsimulation modeling software 2019 update to the 2004 version*. Technical report No. FHWA-HOP-18-036. United States. Federal Highway Administration, 2019.

Corresponding author: Evangelos Mintsis, Hellenic Institute of Transport, Centre for Research and Technology Hellas, Thessaloniki, Greece, e-mail: vmintsis@certh.gr