Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 257 (2025) 225–232

Procedia Computer Science

www.elsevier.com/locate/procedia

The 16th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 22-24, 2025, Patras, Greece

Link analysis through time series decomposition and clustering

Henrik Fredriksson[a,*], David Gundlegård[b], Anna Danielsson[b], Clas Rydergren[b]

[a] *Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, Karlskrona, Sweden*
[b] *Department of Science and Technology, Communications and Transport System, Linköping University, Norrköping, Sweden*

Abstract

We propose a methodology to identify and analyze similarities in long-term trends of road links using travel speed data. The methodology employs seasonal trend decomposition by LOESS (STL) to extract trend curves from travel speed time series, clearly representing underlying long-term patterns and behavior. These trend curves are then analyzed using the k-means clustering algorithm to group road links based on long-term trends. The resulting clusters offer valuable insights for long-term planning in traffic management, infrastructure development, and identifying potential bottlenecks within the road network. To demonstrate the proposed methodology, we applied it to travel speed data from the European road E4, focusing on the route between Södertälje and Stockholm. The analysis reveals distinct trend characteristics and behaviors, highlighting the diverse nature of traffic patterns in different road links.

Keywords: link analysis; time series decomposition; trend curve; clustering

1. Introduction

Understanding travel speed variations over time in transportation networks is crucial for effective traffic management and planning. Analyzing the temporal patterns of traffic states, such as travel speed, can reveal critical insights into the dynamics of the transportation system and congestion trends. The generation of traffic and mobility data has significantly increased due to advancements in digital technologies such as the Internet of Things (IoT) and Intelligent transportation systems (ITSs). The increase in traffic data availability can enhance transportation decision-making through data-driven methods and analytics [1].

Long-term planning and trend analysis plays a critical role in traffic analysis. Trend analysis provides insights into how traffic states evolve, allowing road infrastructure planners to make informed decisions about future infrastructure needs. Traffic states are not static—they change due to various factors like population growth, urban development,

* Corresponding author. Tel.: +46-455-385548
E-mail address: henrik.fredriksson@bth.se

10.1016/j.procs.2025.03.031

and shifts in commuting habits. By studying historical data, it is possible to identify recurring patterns such as peak periods, seasonal fluctuations, or long-term behavior of the traffic states. Understanding the long-term behavior of the traffic states helps prepare for future challenges, such as the risk of more severe congestion, which could lead to breakdowns, longer travel times, increased vehicle emissions, and road safety issues. Incorporating long-term trend analysis into planning also helps optimize infrastructure investments. For instance, expanding a highway can prevent bottlenecks and reduce the need for costly last-minute fixes. Trend analysis also supports adopting new technologies, such as smart traffic lights or autonomous vehicle lanes, that may enhance efficiency and safety in the future.

A clustering algorithm aims to group data points with similar characteristics together. In recent years, clustering algorithms have been applied to spatio-temporal traffic data to reveal similar patterns in the traffic state in both spatial and temporal domains. This approach helps to understand the historical and prevailing traffic conditions, and clustering is also used for short-term prediction purposes [2, 3, 4, 5]. In particular, clustering of time series of traffic data is a commonly used approach to reduce the amount of data and to find representative links and days capturing the temporal variations of traffic data [6, 7, 8]. For a longer time horizon, clustering has been used together with regression to estimate the Annual Average Daily Traffic (AADT) volumes [9].

The complexity and variability of the vast amounts of available traffic data pose significant challenges in storing, managing, and processing all the data on a large and operational scale [10, 11]. Moreover, finding representative traffic patterns and regularities from large amounts of data is essential as it may serve as the basis for predicting future states [7]. Furthermore, when dealing with very long time series with high dimensionality, some clustering algorithms become impractical or inefficient [12]. This paper addresses these challenges by proposing a methodology to cluster links in road networks by extracting trend curves from time series data of travel speeds on links and subsequently clustering links with similar travel speed trends. The purpose is to find representative long-term behavior and identify meaningful patterns and trends of the links. The extraction of trend curves helps to reduce data complexity, enabling a more precise visualization and understanding of long-term patterns in traffic behavior. Clustering road links with similar trend characteristics makes it possible to group links that exhibit comparable speed dynamics, which is potentially valuable for targeted investments, infrastructure planning, and performance assessment of traffic control measures.

The remainder of the paper is structured as follows. Section 2 outlines the methodology, including data preprocessing, trend extraction, and clustering approaches. Section 3 presents the results and discusses the implications of the identified clusters. Finally, Section 4 concludes the paper with final remarks and directions for future research.

2. Methodology

In this section, we describe the steps of our approach to cluster links to reveal different trend characteristics of the links using travel speed data.

2.1. Data processing

The first step is to process the travel speed data to find a representative time series to extract trend curves. We represent a time series for each link as an ordered n-tuple

$$Y(t) = [Y(1), Y(2) \dots, Y(n-1), Y(n)], \tag{1}$$

where $Y(k)$ is the travel speed value at timestamp $k = 1, \dots n$. We use the median value from travel speed observations between 6 a.m. to 10 p.m. from Monday to Thursday for each week to create a time series (1) for each link. Using real-world travel speed data can introduce errors and faulty values due to data collection inaccuracies, which can affect the reliability of the collected information. Therefore, we identified and removed outliers to find a representative time series. To this end. we use the interquartile range (IQR) method. The IQR method identifies outliers as data points that fall outside the range

$$[Q_1 - 1.5 \times \mathrm{IQR}, Q_3 + 1.5 \times \mathrm{IQR}]\,. \tag{2}$$

In (2), Q_1 represents the lower quartile, Q_3 represents the upper quartile, and the IQR is calculated as IQR= $Q_3 - Q_1$. Outliers and missing data points were replaced and filled using linear interpolation.

2.2. Time series decomposition

Decomposing a time series attempts to find separate components, including the trend component and seasonal component [13]. The trend component represents the long-term changes in the time series' level, while the seasonal component describes periodic fluctuations caused by factors such as the time of year. For example, the travel demand is usually consistently lower on weekends than on weekdays due to less commuting to workplaces, affecting travel speeds.

Additive time series decomposition assumes that (1) can be written in the form

$$Y(t) = T(t) + S(t) + E(t), \tag{3}$$

where $Y(t)$ represents the actual data at period t, while $T(t)$, $S(t)$ and $E(t)$ correspond to the trend, seasonal, and irregular components, respectively, at the same period t.

This paper focuses on the long-term perspective, and therefore, we want to find the trend component $T(t)$ from travel speed data $Y(t)$. To this end, we use seasonal trend decomposition using LOESS (STL). STL decomposition is a method that breaks down a time series into trend, seasonal, and irregular components. In STL decomposition, locally weighted smoothing (LOESS) estimates the trend and seasonal components by fitting smooth curves to the data [14]. The smoothing process gives more weight to neighboring data points close to the estimated point, while farther points receive less weight. The smoothing allows the method to capture localized patterns in the time series. STL is an iterative process where the trend and seasonal components are refined over multiple cycles. First, the trend is estimated by smoothing the data, then removed to estimate the seasonal component. The seasonal part is subtracted from the original data, leaving residuals used to refine trend and seasonality in subsequent iterations. The process repeats until the estimates converge. Fig. 1 shows a time series $Y(t)$ of weekly travel speed medians for a link and the trend-, seasonal- irregular component obtained by the STL decomposition method.

2.3. Clustering

This paper uses the well-known k-means clustering algorithm [15]. The k-means algorithm partitions data points into a fixed number of k clusters. The algorithm iteratively assigns data points to clusters to minimize the within-cluster variance. In this paper, Euclidean distance is the chosen metric to measure the variance between data points and the centroids. The within-cluster sum of squares (WCSS) and elbow method are used to derive the optimal number of clusters. WCSS measures how tightly the data points in each cluster are grouped around their respective centroids [16]. WCSS is calculated as

$$\text{WCSS} = \sum_{i=1}^{N} \|x_i - c_{j(i)}\|^2 \tag{4}$$

where $c_{j(i)}$ is the centroid of the cluster assigned to data point x_i and N is the total number of data points. Using a plot of WCSS against the number of clusters and highlights the elbow point where the rate of decrease slows significantly.

In time series clustering, *shape level similarity* refers to comparing the overall form or pattern of data, focusing on, e.g., peaks and valleys, whereas *structure level similarity* involves comparing the underlying relationships or patterns, such as periodicity or seasonality, rather than the exact shape of the data [12]. Our focus in this paper is on the former, shape-level similarity. Therefore, the extracted trend curve is transformed to deal with different posted speed limits on the links along our studied road. The data points in the trend curves are linearly scaled into the range $[0, 1]$ to make them comparable, regardless of the posted speed limit.

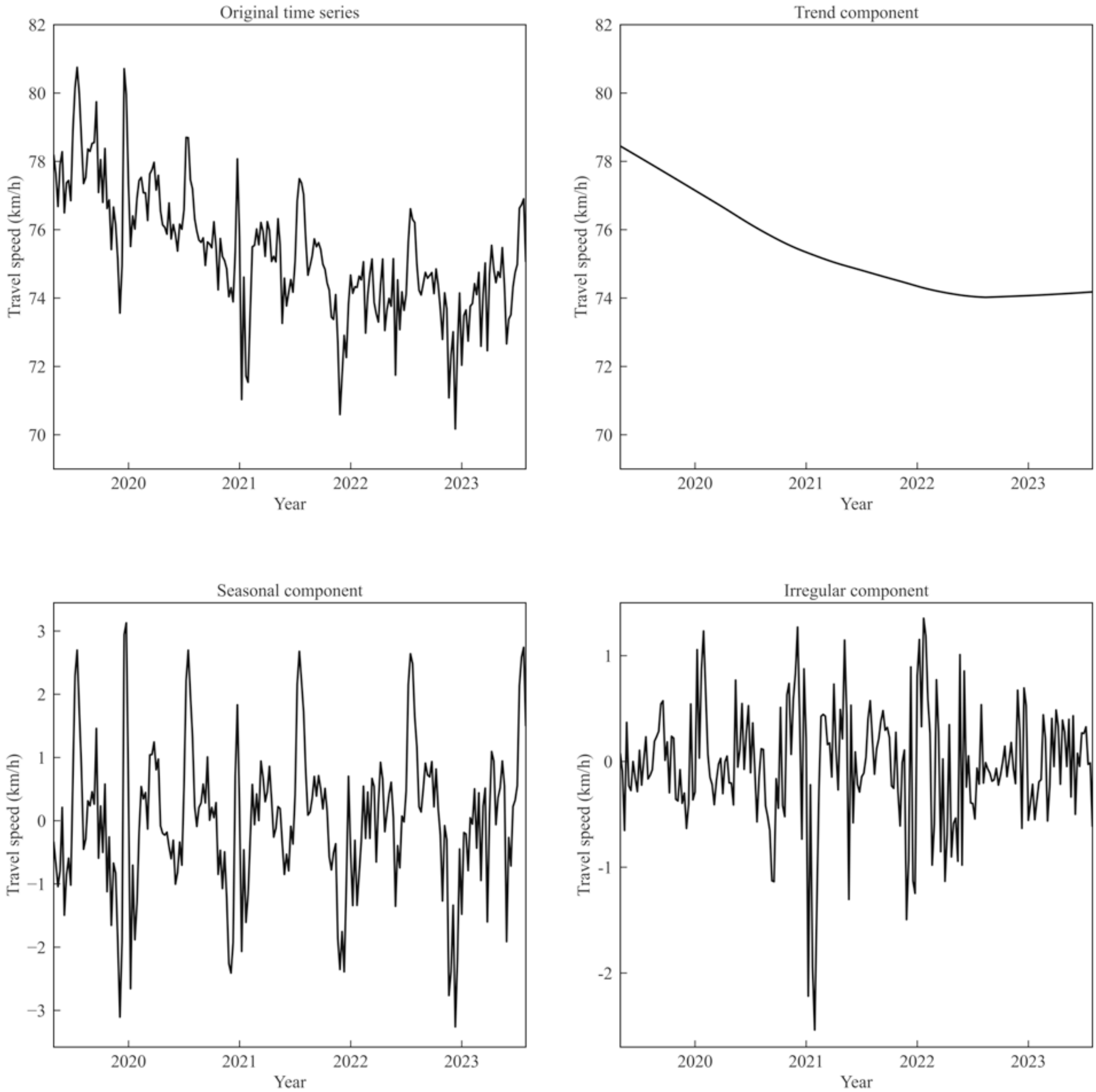

Fig. 1. Time series of weekly median travel speed for a link and the trend-, seasonal- irregular component obtained by the STL decomposition method.

3. Case study

3.1. Setup

The results from our proposed methodology are derived from travel speed data collected on the E4 European route between Södertälje and Stockholm. The road comprises 38 links in the southbound direction and 39 links in the northbound direction, each containing two to four driving lanes. The link lengths vary between 234 and 859 meters, with an average length of 661 meters. The dataset spans from 2019-05-06 to 2023-07-31, giving $n = 222$ data points (weeks) for each time series (1).

3.2. Results

The optimal number of clusters, determined by the WCSS (4) and elbow method, for our studied road is four. The resulting clusters are visualized in Fig. 2, showing the trend curves of the link in each cluster and their mean value (centroid). Fig. 3 shows the spatial distribution of the links in each cluster.

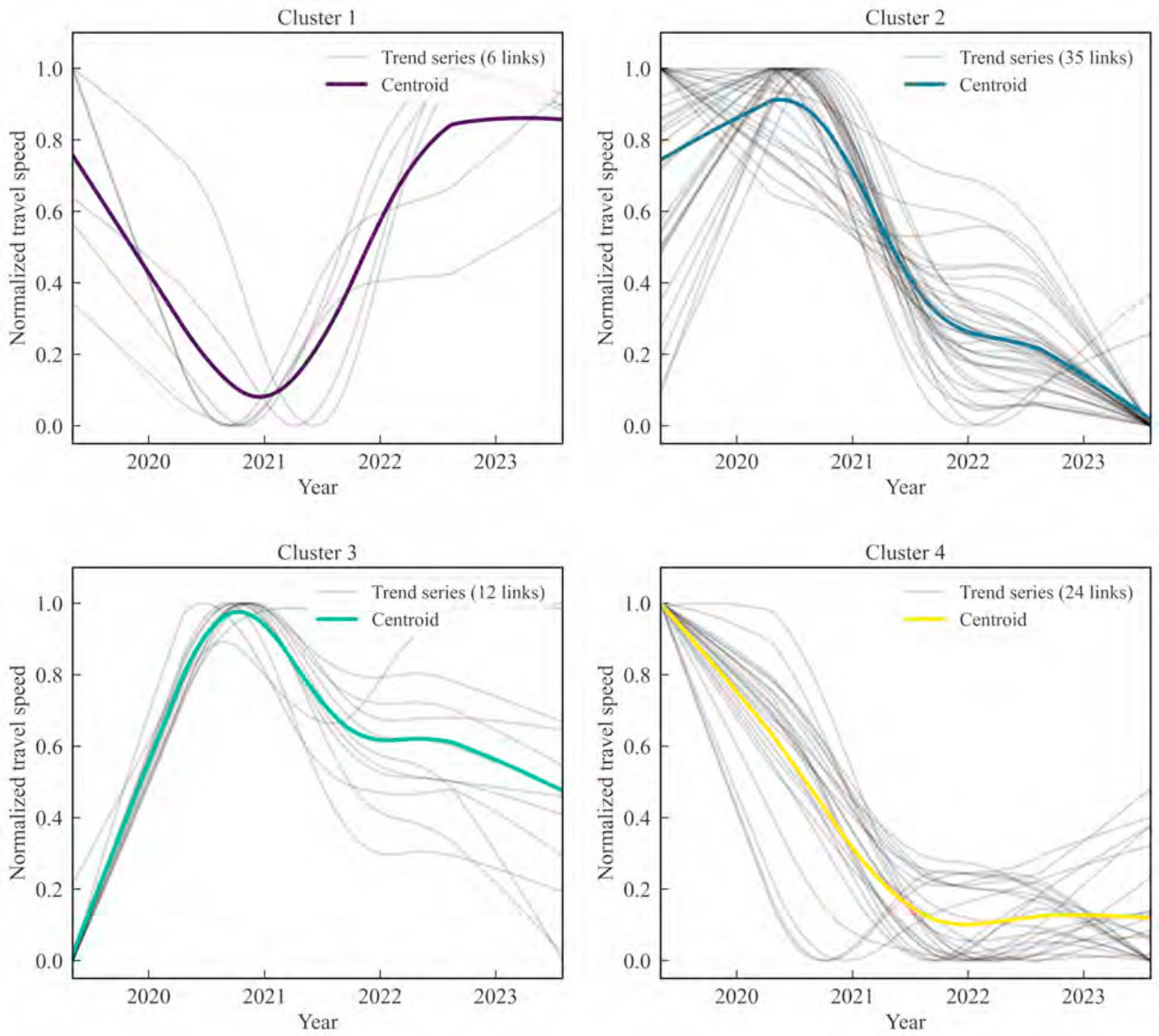

Fig. 2. Clusters of trend curves for weekly median travel speed on a specific link, obtained through STL decomposition.

The links in *Cluster 1* exhibit distinct trend curve shapes compared to those in the other clusters. Specifically, they show a noticeable decrease in travel speed around December 2020, possibly influenced by specific traffic states or conditions during that period. This decrease is followed by a rebound and a steady increase in travel speed throughout 2021, indicating a recovery phase and possibly an adaptation to new traffic conditions. It is important to note that this cluster contains only six links, with free-flow travel speeds ranging from 69.3 km/h to 80.8 km/h. The relatively

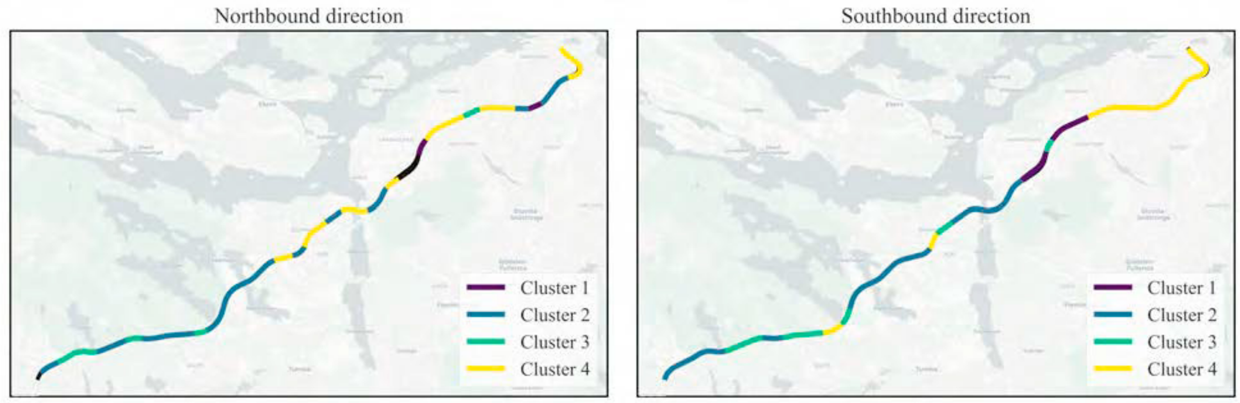

Fig. 3. Spatial distribution of the links in the resulting clusters.

small size of this cluster and the specific range of free-flow speeds and their spatial locations suggest that this group of links may be more sensitive to certain external factors, such as localized traffic patterns, or roadworks, which could possibly explain the pronounced fluctuation in their travel speed trends. Additionally, the unique characteristics of these links might make them more susceptible to variations in traffic volume and driver behavior, further contributing to the observed trends.

Cluster 2, the largest cluster, consists of 35 links. The main characteristic of the trend curves in this cluster is a consistent decrease in travel speed, starting from mid-2020 and continuing onward. A notable aspect of this cluster is that the majority of links (22) have a free-flow travel speed of around 100 km/h, while the remaining 13 links exhibit free-flow travel speeds ranging from 70.2 km/h to 81.2 km/h. This difference in free-flow speeds may indicate varying road conditions or traffic state patterns. However, the overall trend of decreasing speeds across the cluster suggests a widespread factor affecting most of these links, such as increasing travel demand.

The trend curves in *Cluster 3* are characterized by a gradual increase in travel speed from the start of the study period, followed by a decline in speeds from late 2020 onward. This cluster primarily contains links with free-flow travel speeds of around 100 km/h. The consistent free-flow speed across these links suggests they share similar characteristics and usage patterns. The observed decline in travel speeds from late 2020 onward could be influenced by external factors such as the impact of the COVID-19 pandemic on travel behavior and traffic patterns.

Cluster 4 is the second-largest cluster, comprising 24 links. The trend curves in this cluster show a steady decline in travel speeds from 2019 until around 2022, at which point the speeds level off. This extended period of decline could suggest that these links experienced long-term, cumulative effects from factors like increased travel demand, which gradually slowed down traffic over several years. The leveling off in 2022 indicates that traffic conditions may have stabilized, though at lower speeds than in the earlier period.

Scaled trend curves are used to find clusters based on shape similarity, accounting for differences in travel speed values. Fig. 4 shows histograms showing the travel speed range, the difference between the highest and lowest travel speed value, for each trend curve within each cluster. This highlights the variations in absolute travel speed changes over the studied period for links sharing the same trend characteristics.

3.3. Discussion of the results

Besides the trend curves in Cluster 1, a notable characteristic of the studied period is the overall decrease in travel speeds. A possible explanation for this trend is the effects of the COVID-19 pandemic. Due to restrictions and concerns about virus transmission, commuters may have been less inclined to use public transport, opting instead for private cars to maintain social distance and reduce exposure. This shift in commuting behavior likely led to increased travel demand, contributing to higher traffic volumes and congestion. The increased travel demand during this period may have negatively impacted travel speeds, as roads that were previously able to accommodate traffic efficiently became more congested. Additionally, changes in work patterns, such as the rise of remote work and flexible schedules, could have altered peak travel times, further complicating traffic flow. The pandemic also prompted changes in

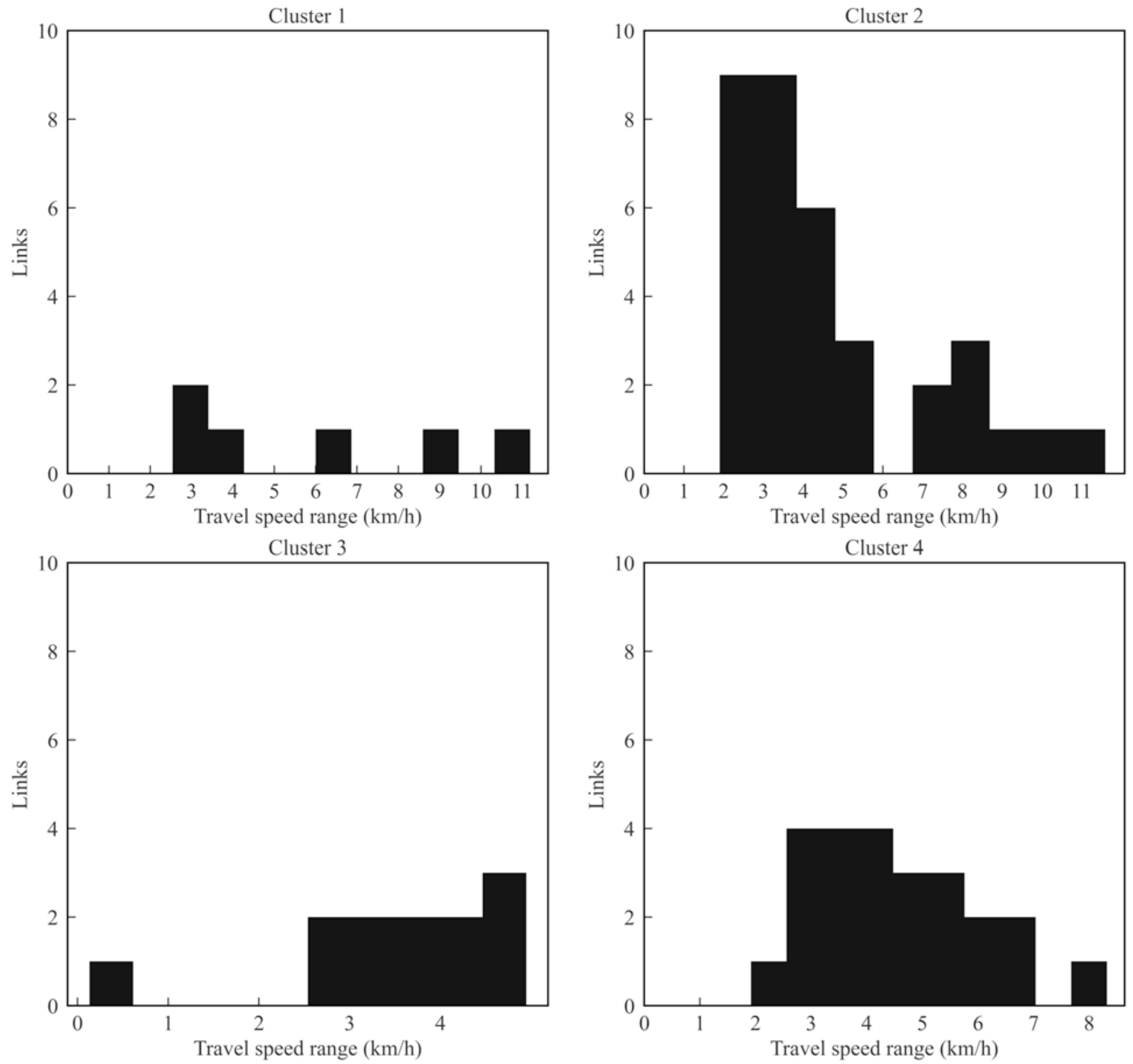

Fig. 4. Distribution of travel speed range of trend curves in the resulting clusters.

delivery and logistics operations, with more goods being transported by road to meet the surge in online shopping and home deliveries, adding to the overall travel demand. Overall, the combination of these factors during the COVID-19 pandemic likely played a significant role in the observed decrease in travel speeds, highlighting the complex interplay between public health measures, commuter behavior, and traffic dynamics.

It is also worth emphasizing that not all trips starting in Södertälje have their destination in Stockholm and vice versa. Vehicles will leave and enter the studied road on off-ramps and on-ramps, meaning the travel demand may vary across the links. This may be one of the reasons neighboring links have different trend characteristics.

4. Conclusions and future work

Long-term trend analysis is essential in traffic planning as it helps identify how traffic states evolve, driven by factors like population growth and urban development. By studying historical data, traffic planners can forecast future road demand, mitigate congestion risks, and optimize infrastructure investments. A proactive approach supports the adaptation of emerging technologies in transportation, such as the IoT and ITSs, to improve the efficient and safe mobility of people and goods. This paper presents a methodology to detect trend characteristics for links in a road network. The proposed methodology employs time series decomposition techniques to extract trend curves from time series data and utilizes the k-mean clustering algorithm to group links with similar trend characteristics. Using travel speed data from the European road E4, the case study reveals diverse trend characteristics across different links. Notably, the study highlights the potential impact of the COVID-19 pandemic on overall travel demand, which has led to shifts in traffic patterns and travel behaviors. By identifying and understanding these trends, traffic planners can make informed decisions to enhance road network performance. These decisions include implementing targeted measures to alleviate congestion, planning for future infrastructure needs, and integrating new technologies to adapt to changing traffic dynamics. The insights gained from this analysis are crucial for developing resilient and adaptive transportation systems that can respond effectively to current and future challenges.

Future research includes analyzing additional data sources, such as travel demand and weather conditions, to study how external factors affect the trends of travel speeds.

References

[1] M. Chowdhury, K. Dey, A. Apon, Data analytics for Intelligent Transportation Systems, Elsevier, 2017.

[2] R. Asadi, A. Regan, Spatio-temporal clustering of traffic data with deep embedded clustering, in: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility, 2019, pp. 45–52. doi:10.1145/3356995.3364537.

[3] R. Asadi, A. Regan, Clustering of time series data with prior geographical information, arXiv preprint arXiv:2107.01310 (2021). doi:10.48550/arXiv.2107.01310.

[4] Z. Cheng, W. Wang, J. Lu, X. Xing, Classifying the traffic state of urban expressways: A machine-learning approach, Transportation Research Part A: Policy and Practice 137 (2020) 411–428. doi:10.1016/j.tra.2018.10.035.

[5] M. Y. Ansari, A. Ahmad, S. S. Khan, G. Bhushan, Mainuddin, Spatiotemporal clustering: a review, Artificial Intelligence Review 53 (2020) 2381–2423. doi:10.1007/s10462-019-09736-1.

[6] T. Erdelić, S. Vrbančić, L. Rošić, A model of speed profiles for urban road networks using G-means clustering, in: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015, pp. 1081–1086. doi:10.1109/MIPRO.2015.7160436.

[7] M. Cebecauer, E. Jenelius, D. Gundlegård, W. Burghout, Revealing representative day-types in transport networks using traffic data clustering, Journal of Intelligent Transportation Systems 28 (2024) 695–718. doi:10.1080/15472450.2023.2205020.

[8] D. Toshniwal, N. Chaturvedi, M. Parida, A. Garg, C. Choudhary, Y. Choudhary, Application of clustering algorithms for spatio-temporal analysis of urban traffic data, Transportation Research Procedia 48 (2020) 1046–1059. doi:10.1016/j.trpro.2020.08.132, recent Advances and Emerging Issues in Transport Research – An Editorial Note for the Selected Proceedings of WCTR 2019 Mumbai.

[9] A. Sfyridis, P. Agnolucci, Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling, Journal of Transport Geography 83 (2020) 102658. doi:10.1016/j.jtrangeo.2020.102658.

[10] A. I. Torre-Bastida, J. Del Ser, I. Laña, M. Ilardia, M. N. Bilbao, S. Campos-Cordobés, Big data for transportation and mobility: recent advances, trends and challenges, IET Intelligent Transport Systems 12 (2018) 742–755. doi:10.1049/iet-its.2018.5188.

[11] G. Badu-Marfo, B. Farooq, Z. Patterson, A perspective on the challenges and opportunities for privacy-aware big transportation data, Journal of Big Data Analytics in Transportation 1 (2019) 1–23. doi:10.1007/s42421-019-00001-z.

[12] X. Wang, K. Smith, R. Hyndman, Characteristic-based clustering for time series data, Data mining and knowledge Discovery 13 (2006) 335–364. doi:0.1007/s10618-005-0039-x.

[13] S. Makridakis, S. C. Wheelwright, R. J. Hyndman, Forecasting methods and applications, John Wiley & Sons, Inc., 2008.

[14] C. RB, STL: A seasonal-trend decomposition procedure based on loess, Journal of Official Statistics 6 (1990) 3–73.

[15] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 1967, pp. 281–297.

[16] K. M. Kumar, A. R. M. Reddy, An efficient k-means clustering filtering algorithm using density based initial cluster centers, Information Sciences 418-419 (2017) 286–301. doi:10.1016/j.ins.2017.07.036.