Exploring alternative data sources for estimating traffic index variations, traffic flow ranks, and the Annual Average Daily Traffic (AADT)

Insights and outcomes from a workshop

Ary P. Silvano

VTI PM 2024:7A
Published 2024
vti.se/publications

VTI PM 2024:7A

Exploring alternative data sources for estimating traffic index variations, traffic flow ranks, and the Annual Average Daily Traffic (AADT)

Insights and outcomes from a workshop

Ary P. Silvano

Translated title: Utforskning av alternativa datakällor för skattning av trafikindex variationer, rangkurvor och årsmedelsdygnstrafik (ÅDT). Insikter och resultat från en workshop.
Author: Ary P. Silvano, VTI (orc-ID-nr 0000-0002-7080-5176)
Reg. No., VTI: 2022/0459-8.3
Publication: VTI PM 2024:7A
Published by VTI 2024

Summary

A workshop on enhancing traffic variation estimation, led by VTI, was carried out focusing on alternative data sources for estimating traffic index variations over the year, so-called traffic indices (i.e., monthly, weekday, hourly) and traffic flow ranks (i.e., ranking the hours with the highest indices), as critical metrics for understanding traffic patterns. Moreover, these alternative data sources can potentially improve and complement the estimation of the Annual Average Daily Traffic (AADT).

The workshop highlighted the limitations of current methodologies relying on data from fixed counting points and the need for integrating a wider array of data sources to capture the complex dynamics of traffic more accurately. Therefore, the primary objective was to explore alternative data sources for estimating traffic index variations over the year to better capture and understand regional and local traffic variations. Discussions centered around the potential of mobile cellular network data, satellite imagery, and vehicle probe data, each presenting unique characteristics and uncertainties. Mobile cellular network data showed promising in providing continuous traffic data but raised concerns about data representation and privacy. Satellite data offered high spatial resolution yet faced challenges in data management and necessitated integration and validation against traditional methods. Vehicle probe data, providing mostly urban coverage, highlighted limitations of market penetration, but has the potential of integrating additional sensor technologies.

The conclusion emphasized the crucial role of diverse data sources in advancing traffic analysis, pointing out the need to address uncertainties like data representativeness, privacy concerns, technical integration, and sound validation methods in harnessing these alternative data sources. Furthermore, the workshop advocated for collaborative efforts among stakeholders (i.e., authorities, data providers, academia, end-user's consultancy firms) to refine these alternative data sources and develop new analytical tools for a better understanding of traffic variations.

Keywords

Traffic variations, traffic flow ranks, Annual Average Daily Traffic (AADT), mobile cellular network, satellite imagery, probe data.

Kort sammanfattning

I december 2023 genomfördes en workshop om alternativa datakällor för att skatta trafikvariation och trafikintensitet. Workshopen leddes av VTI och fokuserade på hur alternativa datakällor skulle kunna användas för att förbättra och komplettera dagens beräkningar av trafikvariation över året (månadsvis, veckodagvis och timvis), rangkurvor och årsmedelsdygnstrafik (ÅDT).

Workshopen betonade begränsningarna av nuvarande metoder som förlitar sig på data från punktmätningar och behovet av att integrera ett bredare spektrum av datakällor för att fånga komplexa trafikmönster. Därför var det primära målet att utforska alternativa datakällor för att uppskatta trafikvariationer över året och för att bättre fånga och förstå regionala och lokala trafikvariationer. Diskussionerna fokuserade på potentialen hos data från mobila cellnätverk, satellitbilder och data från fordonen (eng. probe data). Data från mobila cellnätverk visade sig lovande för att tillhandahålla kontinuerliga trafikdata men det finns osäkerheter vad gäller datarepresentation och integritet. Satellitdata erbjuder hög rumslig teckning men det finns utmaningar i datahantering och ett behov av validering med fasta markdata. Probe data från fordonen täcker mestadels urbana områden och har begränsningar i marknadspenetration och i glesbygd. Men en fördel är att det finns potential att integrera med ytterligare sensordata.

Sammanfattningsvis betonade deltagarna i workshopen vikten av att använda en mångfald av datakällor inom trafikanalys för att förbättra framtagandet av viktiga trafikmått. Samtidigt finns en hel del osäkerheter med aspekter som datarepresentation, integritetsfrågor, teknisk integration och tillförlitliga valideringsmetoder för att utnyttja dessa alternativa datakällor. Vidare förespråkade workshopen samarbete mellan olika intressenter (till exempel myndigheter, dataleverantörer, akademi, slutanvändare och konsultföretag) för att förbättra dessa alternativa datainsamlingsmetoder och utveckla nya analytiska verktyg för en bättre förståelse av trafikvariationer.

Nyckelord

Trafikvariation, rangkurvor, årsmedelsdygnstrafik (ÅDT), mobila cellnätverk, satellitbilder, fordons probe data.

Preface

This document reports on the results and conclusions of the Workshop from the project ‘Traffic Variation Updates’ which aims to update the traffic variation indices and traffic flow ranks, based on the methodology described in VTI Notat 31-2005 (Björketun & Carlsson, 2005). The Workshop aims to explore alternative data collection techniques that can enrich and enhance the estimation of traffic variations over the year. The Workshop was prepared and led by the author with the support and expertise of Johan Olstam from VTI and Hung Nguyen from the Swedish Transport Administration.

The author would like to thank all persons involved in making this workshop possible. With special thanks to the group leaders for introducing the data types to the participants and guiding the group discussions. Special thanks to Christian Lewenhaupt from Telia, Magnus Andersson from Malmö University, and Steffen Knab from NiraDynamics.

The project was funded by the Swedish Transport Administration and the contact person has been Hung Nguyen.

Linköping, March 2024

Ary P. Silvano
Project leader

Granskare/Examiner

Åsa Forsman, VTI.

Innehållsförteckning

1. Introduction

1.1. Background

Traffic variation indices, traffic flow ranks, and the Annual Average Daily Traffic (AADT) are key metrics in understanding traffic patterns, providing valuable insights into various aspects of road usage. For example, the AADT describes the average traffic over the year but does not say anything about levels of congestion or temporal traffic variations. On the other hand, traffic variations do not say anything about average traffic conditions but shed light into the temporal variations of traffic throughout the year. This variation is critical in distinguishing different types of traffic, such as local traffic, through traffic (commuting), and tourist traffic. Furthermore, this variability is also crucial in calculating typical traffic variation indices, which include hourly, weekday, and monthly index (i.e., how traffic varies through these periods of time), as well as defining traffic flow ranks, which indicate different levels of traffic intensity. For example, hours with the highest variation indices (e.g., morning and afternoon peak hours) are categorized into the highest flow rank. The estimation of these metrics (i.e., traffic indices, traffic flow ranks, and AADT) is critical for a comprehensive understanding of how traffic dynamics change over the course of a year. However, the current methodology presents limitations because it relies on data from a limited number of measurement sites, utilizing primarily inductive loops and pneumatic tubes. These data are collected from both point-fixed (measured throughout multiple years) and point-varying (measured only in a single year) traffic counting points that operate year-round. Therefore, a wider data source scope is needed to better capture the diverse and complex traffic patterns more accurately.

The workshop aims to explore the possibilities of integrating a wider array of data sources, including traditional cross-sectional data collection methods (e.g., pneumatic tubes, inductive loops) and alternative data collection methods. Hence, the workshop seeks to address the challenges of currently limited traditional datasets to estimate traffic variation indices, traffic flow ranks, and the AADT by exploring alternative traffic counting collection approaches.

1.2. Objective

The aim of this workshop is to facilitate a comprehensive discussion on exploring alternative data sources for traffic variation, traffic flow ranks, and AADT estimation. The objectives are twofold:

- **Exploring Potential Data Sources:** The primary focus is to explore and discuss alternative data sources that can potentially enhance the estimation of traffic variations, including monthly, weekday, and hourly variations. This exploration will encompass data sources such as mobile cellular network data, satellite imagery data, and vehicle probe data.
- **Understanding Regional and Local Traffic Variations:** Another key goal is to discuss how these alternative data sources can help in capturing more detailed regional and local traffic variations. The workshop will explore the potential of these data sources to provide a finer understanding of traffic dynamics.

The focus is on engaging in a dialogue about the possibilities these alternative data sources present by bringing together experts and stakeholders from various fields, e.g., practitioners, researchers, end-user's consultancy firms, and authorities. This exploratory phase is crucial for setting the stage for future advancements. The insights and ideas generated from this workshop are expected to be instrumental in guiding the subsequent phases of method development and data integration.

1.3. Methodology

This section outlines the methodology employed in identifying potential data sources, participant recruitment, workshop preparation, and execution.

1.3.1. Identification of Data Sources:

Based on experience and expertise from the project board, three types of data sources, with the potential to enrich the estimations of traffic variations, traffic flow ranks, and the AADT, were identified:

- Mobile Cellular Network Data: Utilizing data from mobile network operators to estimate vehicle movement patterns based on detected cellphone geographical movements along the mobile cellular network.
- Satellite Imagery Data: Applying high-resolution remote sensing technology from satellites to capture vehicular traffic and road usage patterns.
- Vehicle Probe Data: Leveraging data from GPS-equipped vehicles to track and analyze traffic flows.

1.3.2. Participant Recruitment:

The recruitment process for workshop participants involved two key strategies:

- Collaboration with the Swedish Transport Administration (TrV): Assistance was sought from TrV to gain access to contacts of potential participants, particularly focusing on data providers and consultancy firms using traffic variations and traffic flow ranks as input for traffic forecasts.
- Academic Network Utilization: Colleagues across various academic institutions were contacted to extend invitations to potential participants from the academic sector.

Institutions that participated and were represented in the workshop include VTI[1], TrV[2], SMHI[3], Linköping University, Malmö University, Telia, NiraDynamics, MOVEA, Calås Innovation[4], and WSP[5]

1.3.3. Workshop Invitation and Preparation:

Upon finalizing the list of potential participants, formal invitations were sent out. These invitations detailed the background, the specific problem being addressed, objectives, and the expected contributions from the participants (see Appendix 1). The final number of experts who participated was 24. Based on their expertise and interest, the participants were assigned to a specific group for in-depth discussions during the workshop. The mobile group comprised 8 participants, the satellite group included 9 participants, and the probe data consisted of 7 participants.

Prior to the workshop, participants were provided with a set of preparatory materials, which were distributed to guide the discussion and ensure focused outcomes (see Appendix 2). Additionally, expert leaders were appointed for each alternative data source. These leaders were responsible for presenting an introductory overview of their respective data source and facilitating the related group discussions.

[1] https://www.vti.se/en

[2] https://www.trafikverket.se/

[3] https://www.smhi.se/en

[4] https://calasinnovation.se/

[5] https://www.wsp.com/en-gl/

1.3.4. Workshop Program:

A comprehensive program for the workshop was drafted and disseminated to all participants in advance (see Appendix 3). This program outlined the schedule, topics of discussion, and the format of the workshop.

1.3.5. Workshop Execution:

The workshop lasted for 3 hours and was conducted digitally on December 15, 2023. It employed a combination of presentations, group discussions, and takeaway sessions to explore the potential of the identified data sources in enhancing the estimation of traffic variations, traffic flow ranks, and the AADT.

2. Results

2.1. Mobile Cellular Network

2.1.1. Data Representation

This section outlines Telia's utilization of mobile cellular network data, termed 'Crowd Insights'[6]. Unlike GPS data, Crowd Insights comprises signal data exchanged between mobile handsets and Telia network of cell towers, with each tower covering a specified area, referred to as a grid. A significant volume of data is processed; the Telia network handles approximately 3.3 billion signals daily. This data enables the analysis of population distributions covered by Telia and the detection of cellphone movement patterns between the grids. It can depict where individuals reside, their dwelling patterns, and their movement trajectories within Sweden. The network coverage comprises over 20 000 grids of varying sizes, covering Sweden. The grid sizes are determined by population density in different areas, such as city centers, where grids tend to be smaller. One of the smallest grid sizes is 250 square meters. This granularity allows for detailed insights into population dwelling times and movements between grids. Furthermore, these grids can be aggregated to form larger geographical classification areas, e.g., municipalities, and regions, providing a comprehensive view of population dynamics (Lewenhaupt, 2023).

The data collection and processing ensure anonymity and general data protection regulation (GDPR[7]) compliance. A rigorous privacy-by-design approach is employed, with anonymization processes repeated every hour. These processes require a minimum of five individuals in each grid for any given hour to maintain privacy. The data extrapolation methods extend beyond subscriber behavior, aiming to represent the total population. This extrapolation results in statistics that reflect overall population movements and travel patterns. The data sets produced encompass various aspects, including location activity, visit patterns in catchment areas, and trip data. The trip data includes route points and prediction of the travel mode used, although it does not specify the type of road. Telia data has found commercial application and academic collaboration, including collaborative projects with Linköping University[8] and other institutions. These collaborations focus on developing statistical methods and refining data extrapolation techniques. See for example Gundlegård (2018) and Breyer (2019). Currently, similar data sets are being utilized in Finland for traffic analysis and related purposes (Lewenhaupt, 2023).

2.1.2. Discussions' takeaways

The discussion on mobile cellular network data, led by Telia, focused on various aspects such as data granularity, temporal aspects, and applications for road network analysis. Key takeaways include:

1. *Data Granularity and Temporal Aspects:* Data is collected 24/7 and can be aggregated to present information at an hourly level. The coverage areas determine the granularity, influencing the capture of road network data.
2. *Socio-Economic Data and Market Representation:* Concerns were raised about the representation of different socio-economic groups in the data, considering Telia's market share (~35%). The data may not completely represent the Swedish population, and deviations from other data sources like the population register from Statistics Sweden (SCB[9]), were discussed.

[6] https://business.teliacompany.com/crowd-insights

[7] https://gdpr.eu/what-is-gdpr/

[8] https://liu.se/en

[9] https://www.scb.se/en/About-us/

SCB has also conducted a study on the suitability of anonymized and aggregated mobile phone position data for producing official statistics from Telia data (Vlag, et al., 2022).

3. *Data Types and Privacy Concerns:* The type of data collected includes position of the mobile handset, time stamp, and cell tower information. Discussions highlighted privacy considerations, ensuring that only anonymized data is used, preventing the tracking of individuals or vehicles.
4. *Data Utilization and Accessibility:* Questions about special software or hardware requirements for data utilization were addressed. Data can be delivered in formats like CSV or Excel and is accessible via an application program interface (API). Data processing, from raw data to statistics (e.g., anonymization, aggregation, extrapolation), is carried out by Telia within its cloud infrastructure, involving proprietary software in the process, i.e., only aggregated results are delivered to customers with Telia owning the algorithms to estimate the statistics of interest.
5. *Network Granularity and Urban-Rural Differences:* In densely populated areas with many roads, the coverage areas (i.e., the grids) are smaller, posing challenges in accurately identifying specific links due to the high density of roads. Conversely, rural areas with larger grids allow for much easier identification of general movement patterns, yet fail to capture movement patterns in finer details, as too much information is combined into a single grid. This highlights the difficulties of using grid-based systems to analyze movement patterns, presenting drawbacks in different environments.
6. *Coverage Areas and Trip Analysis:* The discussion touched on the ability to analyze complete trips, considering the varying coverage areas that change over time. This analysis can provide insights into traffic patterns, although the accuracy in urban areas with complex road networks may be low.
7. *Data Validation and Quality:* The importance of data validation was emphasized, highlighting the need to ensure data quality and reliability. Historical comparisons and collaborations, such as those with TrV, were mentioned as methods to validate and enhance the data.

2.2. Satellite Imagery

2.2.1. Data representation

This data consists of high-resolution images collected from satellites, showing the number of vehicles on a selected road network area at a specific point in time. This high-resolution data is often acquired by paid services rather than freely available sources. For instance, SkySat[10] data from the European Space Agency (https://www.esa.int/), while available for research projects, may not be sufficient for all needs, necessitating additional paid services. Satellite imagery and video data from various sources, including Iconos[11], QuickBird[12], WorldView[13], and SkySat, are extensively utilized. These satellite constellations, with varying lifespans, provide essential data; however, some, such as QuickBird, has been inactive since 2015. This aspect of satellite longevity is crucial when considering data sources for research. Additionally, satellite data can be complemented with the use of aerial data, obtainable from land survey departments. High-resolution satellite imagery data, especially sub-meter resolution

[10] https://earth.esa.int/eogateway/missions/skysat

[11] https://www.satimagingcorp.com/satellite-sensors/ikonos/

[12] https://www.satimagingcorp.com/satellite-sensors/quickbird/

[13] https://earth.esa.int/eogateway/missions/worldview-3

necessary for vehicle detection, significantly impacts the data volume and subsequent computational requirements. Google's Aerial Data[14] is one such source utilized in research (Andersson, 2023).

Temporal resolution is another critical factor. Newer satellites, like those in the SkySat constellation, offer higher temporal resolutions, enabling more frequent data capture, potentially several times a day. However, data might only be available for specific times of the day, an important consideration in traffic research applications. Remote sensing data from satellites and aerial sources are increasingly accessible and affordable, providing a comprehensive overview of traffic situations. In research, convolutional neural networks (CNNs) are commonly used for image classification, including vehicle detection (Koga, et al., 2020), (Stuparu, et al., 2020). These methods, combined with observed traffic counts and spatial interpolation techniques, can potentially allow for extensive coverage over larger areas and extended timeframes (Andersson, 2023).

Challenges in using satellite data for traffic analysis include limited temporal resolution and the impacts of weather conditions, e.g., cloud cover. Additionally, distinguishing between different vehicle types (e.g., cars, trucks, buses) and the extensive processing time for high-resolution data are notable concerns. Techniques to mask out (e.g., hide or remove) roads in images can help reduce processing time. The potential to estimate traffic variations for various vehicle types at a segmented level, using satellite imagery, remains a topic of active discussion and exploration in the field (Andersson, 2023). See for example the research work by Sakai et. al., (2019) and Resksten & Salberg (2021).

2.2.2. Discussions' takeaways

The discussion on satellite data revolved around the use of satellite images and remotely sensed measurements for traffic analysis. Key points from this discussion include:

1. *Spatial and Temporal Resolution:* Satellite images provide high spatial resolution, with some measurements offering resolutions as fine as 0.5 meters. This spatial resolution is critical for detecting vehicles. In terms of temporal resolution, certain satellite constellations allow for daily imagery, which can potentially enable the measurement of daily traffic variations, given the observations are at the same time of the day. Some satellite constellations (e.g., SkySat consisting of 21 satellites) can capture imagery up to seven times a day, depending on the weather conditions on the location of interest.
2. *Coverage and Scale of Application:* The coverage of satellite imagery varies based on the desired scale of analysis, whether it be national, regional, local, or city level. The choice of scale influences the type of satellite product used.
3. *Use of Commercial Products:* There are various commercial products available for remotely sensed data. The discussion acknowledged a gap in reviewing these products, especially in their application to traffic counts and estimation of traffic patterns.
4. *Combining Data Sources:* Remotely sensed data often needs to be combined with ground data (i.e., other traditional methods) for accurate interpretation. The Transport Administration fixed year-round ground data was mentioned as a potential source for interpolating and calculating traffic flow and traffic density.
5. *Data Storage and Management:* Managing and storing satellite data requires substantial technical capacity due to the volume and frequency of data collected. This is primarily a technical matter but is essential for effective data utilization.
6. *Regulatory and Institutional Considerations:* The discussion touched upon the regulatory aspects, such as space laws and the lifespan of satellites. It emphasized the need for

[14] https://earthengine.google.com/

continuous adaptation of methods and collaboration with various actors, including state and national agencies, for the effective use of satellite data in traffic analysis.

2.3. Probe data

2.3.1. Data representation

NiraDynamics[15] collects detailed vehicle travel information such as GPS location, vehicle speed (in millimeters per second), vehicle heading (i.e., travel direction in degrees relative to the compass), and the accuracy of these values. The data is streamed and updated at intervals of a few seconds, providing near real-time data. The GPS data, however, may sometimes be inaccurate, showing vehicles driving on the side of the road, which is then adjusted based on accuracy assessments. To comply with legal data protection requirements, NiraDynamics ensures that the start and end points of each journey are anonymized, preventing the tracing back to a single vehicle. The data is collected from Audi[16] vehicles with an active Audi connect function[17], which is typically available free of charge for the owner for the first three years of the life of the vehicle. After this period, vehicle owners can choose to continue using this function for a fee. Data analysis is limited to passenger vehicles, as NiraDynamics does not provide data for commercial or heavy traffic vehicles. The data allow to analyze traffic patterns, vehicle speeds, and congestion in near real-time, offering insights into urban traffic management and planning (Knab, 2023).

2.3.2. Discussions' takeaways

The discussion on vehicle probe data focused on the current capabilities and future potential of vehicle fleet data in traffic analysis. Key takeaways from this discussion include:

1. *Coverage and Data Evaluation:* As for now, the current vehicle fleet for data collection is estimated at around 50 000 vehicles in Sweden (only Audi vehicles up to 3 years-old), with varying penetration rates (percentage share of the total vehicle fleet) and with high coverage in major cities. However, coverage in rural areas may be less consistent. It was suggested that for annual evaluations, the coverage should be sufficient, but this needs further investigation. There are ongoing discussions within the Volkswagen Group[18] to potentially expand the vehicle fleet, which would enhance data collection capabilities.
2. *Use of Vehicle Fleet Data:* NiraDynamics has conducted basic investigations with the vehicle fleet data, and there is interest in exploring this data more deeply. Ideas for evaluation include comparing vehicle speeds from Nira with known vehicle speeds from other sources, such as fundamental diagrams or traffic flow counting systems. See Forsman et al., (2018) for such a comparison using Bluetooth data. Addressing challenges like standing vehicles and missing driving directions was recognized as essential for accurate analysis.
3. *Data on Traffic Flow and Vehicle Density:* The discussion also covered the application of probe data in estimating traffic flow and vehicle density. For accurate flow estimation, it is necessary to understand the proportion of probe vehicles in the total traffic flow. Questions were raised about the ability to identify the driving direction of vehicles and how to treat individual data points, as the speed data might be provided by one or few vehicles. In a previous study in Sweden, probe data was found to have good correlation with the speed and flow from stationary counting radar sensors (Ahlberg, et al., 2021)

---

[15] https://niradynamics.se/

[16] https://www.audi.se/

[17] https://www.audi-mediacenter.com/en/the-audi-rs-6-avant-and-the-new-audi-rs-6-avant-performance-12345/i/nfotainment-and-audi-connect-12356

[18] https://www.vwgroup.se/

4. *Other Data Sources and Sensor Technology:* The possibility of integrating other types of sensor data for traffic analysis was discussed. For instance, onboard computer data (e.g., lidar) could provide additional insights into vehicle proximity and density. However, utilizing such data might require software changes in vehicles to consider these additional information sources.
5. *Challenges and Potential Solutions:* The challenges of low market penetration and how it affects data accuracy were acknowledged. The potential use of radar sensors, parking aid sensors, and camera data as additional data sources was discussed. These could provide indirect measurements, such as the distance and speed of surrounding vehicles, thereby enhancing the overall quality of traffic data analysis.

3. Conclusions

The workshop discussions provided invaluable insights into the potential and challenges of using diverse data sources for the estimation of traffic variation indices, traffic flow ranks, and the AADT. The key findings from each data type offer a multifaceted view of current capabilities and future directions in traffic data collection and analysis. See Table 1 for a summary of the key findings.

3.1. Mobile Cellular Network Data:

The exploration of mobile cellular network data highlighted its value in capturing traffic variations with fine temporal granularity. However, challenges such as data representation across socio-economic groups, privacy concerns, market share, and the specificity of network granularity in urban versus rural settings were noted. The discussion highlighted the importance of balancing data accessibility with privacy and representativeness. It also emphasized the need for robust data validation methods to ensure reliability and the difficulties of converting detected cellphone movements between grids into vehicle counts/movements on the roads (e.g., several cellphones or persons can be in the same vehicle) and the type of travel mode (e.g., cellphone in car, bus, tram).

3.2. Satellite Imagery Data:

Satellite data, with its high spatial and increasing temporal resolution, emerges as a powerful tool for traffic analysis. Its ability to cover large areas is particularly valuable. However, the discussions revealed challenges in data storage and management, and the necessity of combining satellite data with ground data for accurate interpretations, due to its much lower updating frequency. This highlights the importance of interdisciplinary collaboration, technical capacity building, and regulatory compliance for effective utilization of satellite imagery in traffic studies.

3.3. Vehicle Probe Data:

The probe data from vehicle fleets, currently limited to a specific brand and age of vehicles in Sweden, presents both opportunities and challenges. While offering higher coverage in major cities, its applicability in rural regions and for diverse traffic scenarios remains to be explored. The potential expansion of the vehicle fleet and integration with other sensor technologies indicates promising future capabilities, yet the market penetration and technical limitations must be addressed to fully leverage on this data source.

3.4. Overall Implications:

The collective insights from these discussions point towards a future where traffic analysis is based on a combination of traditional and alternative data sources. This integrative approach is crucial for developing a more comprehensive understanding of traffic patterns and addressing specific challenges in traffic management and infrastructure planning. However, the effective use of these diverse data sources centers on addressing key challenges such as data representativeness, privacy, technical capacity, sound validation approaches, and methodological extrapolation and integration for harnessing these data sources for meaningful insights.

3.5. Future Directions:

It is imperative to foster collaborative efforts among technology providers, data analysts, traffic experts, and policy makers. This collaboration is essential for refining data collection methods, enhancing data accuracy, and developing new analytical tools. Additionally, continuous monitoring of technological advancements and regulatory frameworks will be crucial in adapting these methods to evolving traffic scenarios and data landscapes.

Table 1 Summary of the workshop key findings.

Data type	**Benefits & Potential**	**Challenges & Limitations**
Mobile Cellular Network	• Continuous data 24/7. • Essentially, coverage for the whole country. • Finer granular aggregation up to the hour level. • Data accessible via API. • Data processing infrastructure in place.	• Only part of movements and only part of users (Telia’s share). • Flow counting is estimated based on the market share of the operator. • Needs methodology to impute where the movements take place (i.e., road type). • Needs further validation to ensure data quality and reliability for vehicle flow and imputed travel mode (e.g., car, bus, tram). • Proprietary software is involved. • Difficulties due to grid system of varying sizes. Smaller in urban areas and bigger in rural areas.
Satellite Imagery Data	• Some constellations, several times a day, depending on weather at location of interest. • Some measurements up to 0.5-meter resolution. • Varied spatial coverage, from national to city or area/zone level.	• Low updating frequency. • Heavily dependent on weather conditions, e.g., cloudy, rainy. • May need substantial technical and storage capacity. • Needs development of methodologies for traffic applications, e.g., vehicle classification, vehicle counting. • Needs data validation to ensure quality and reliability. • May need development of regulatory mechanisms/frameworks. • May need the collaboration with other national agencies, e.g., Space Agency.
Vehicle Probe Data	• Near real-time data collection • Detailed vehicle data, e.g., speed, GPS location, heading. • Can offer good coverage in major cities. • Possibility to integrate other types of onboard sensor data, e.g., lidar to estimate vehicle headways, density.	• As for now, only a limited number of vehicles (i.e., 3 years-old or newer). • Low coverage in rural areas • Proprietary software is involved. • Recognized issues include standing vehicles, missing driving directions and mis map-matching vehicles outside the road. • Needs further development of methodologies for traffic applications. • Needs agreement with other vehicle manufactures to increase vehicle probe fleet. • Needs further data validation to ensure quality and reliability.

References

Ahlberg, J. et al., 2021. *Probe data - förstudie kirng användning av GPS-baserad probedata för skattning av hastigheter, länkflöde och ruttval (In Swedish with Summary in English),* TRV 2019/98384: Linköping University & Ramböll.

Andersson, M., 2023. *Remote sensing data for traffic density estimations [PowerPoint slides],* s.l.: Malmö University.

Björketun, U. & Carlsson, A., 2005. *Trafikvariation över året. Trafikindex och rangkurvor beräknade från mätdata (in Swedish),* Linköping: VTI.

Breyer, N., 2019. *Analysis of Travel Patterns from Cellular Network Data.* Licenciate Dissertation ed. Linköping: Linköping University.

Forsman, Å., Vadeby, A., Gundlegård, D. & Ringdahl, R., 2018. *Utvärdering av hastighetsmätningar med blåtandssensorer: Jämförelse med data från MCS (Motorway Control System) (In Swedish with Summary in English),* Linköping: VTI.

Gundlegård, D., 2018. *Transport Analytics Based on Cellular Network Signalling Data.* Doctoral Dissertation ed. Linköping: Linköping University.

Knab, S., 2023. *NIRA Dynamics FCD - Vehicle Speed / Traffic Flow [PowerPoint slides],* s.l.: NiraDynamics.

Koga, Y., Miyazaki, H. & Shibasaki, R., 2020. A method for detection in high-resolution satellite images that uses a region-based object detector and unsupervised domain adaptation. *Remote Sensing,* 12(575).

Lewenhaupt, C., 2023. *Telia Crowd Insights. Mobile Network Data [PowerPoint slides],* s.l.: Telia.

Reksten, J. H. & Salberg, A.-B., 2021. Estimating Traffic in Urban Areas from Very-High Resolution Aerial Images, International Journal of Remote Sensing. *Engineering, Environmental Science, Computer Science,* 42:3(Taylor & Francis), pp. 865-883.

Sakai, K., Seo, T. & Fuse, T., 2019. *Traffic density estimation method from small satellite imagery: Towards frequent remote sensing of car traffic.* Auckland, New Zealand, IEEE Intelligent Transportation Systems Conference (ITSC).

Sayers, M. W., Gillespie, T. D. & Paterson, W. D., 1986. *Guidelines for Conducting and Calibrating Road Roughness Measurements,* Washington, D.C., USA: The International Bank for Reconstruction and Development - The World Bank.

Stuparu, D.-G., Ciobanu, R.-I. & Dobre, C., 2020. Vehicle detection in overhead satellite images using a one-stage object detection model. *Sensors,* 20(6485).

Vlag, P., Durnell, U. & Malmros, J., 2022. *Mobil phone position data and official statistics.* s.l., Nordic Statistical Meeting - NSM 2022.

Appendix 1

Workshop Invitation

Workshop on alternative data sources for estimating traffic variations: Annual Average Daily Traffic (AADT), traffic variation indices, and traffic flow ranks.

Location: Digital

Duration: 09:00 – 12:00 (3 hours)

Date: 15th of December 2023

Bakground

Annual Average Daily Traffic (AADT) is a central aspect of traffic analysis. There are also other important aspects of the AADT, such as its seasonal variations over the year. The AADT is used to estimate the type of traffic (e.g., local traffic, through traffic, tourist traffic), traffic variation indices (e.g., per hour, per weekday, per month) and traffic flow ranks (i.e., levels of traffic). These estimations are crucial for creating a complete picture of how traffic varies throughout the year. However, these estimations are currently based on data from fixed year-round traffic counting points, which have inherent limitations.

The current dataset, based on fixed year-round counting points from 2005, is insufficient to provide a comprehensive overview of traffic flows throughout the country (covering nearly 98,000 km of national roads). With only 80 fixed counting points representing the whole country's variations, there is a clear need to include more and varied data sources.

Objective

Our goal is to explore other types of data sources for two main purposes: to expand the database for the estimation of traffic variations throughout the year and to include more detailed regional and local variations. The goal is to have a more detailed and more reliable traffic variation estimation which can integrate different data types.

Your contribution

If you work with or have knowledge of data that can contribute to our goal, we would be very grateful for your participation. How can your data be used in this context? In what ways can we together improve the process for estimating traffic variation indices and traffic flow ranges?

It is important to note that our purpose is to generate traffic flow variations that extend down to the hour and road link level. We encourage you to suggest how your type of data can be used in this context (e.g., hourly flows per road link). At the same time, please share how easy or complicated it would be to include your data in the process of estimating AADT, traffic indices and traffic flow ranges.

Your expertise and contribution are crucial to us. By collaborating, we can both expand the database and include more detailed regional/local traffic variations.

Appendix 2

List of preparation questions

How is the data collected?

- For mobile network data: Are GPS or mobile network data used?
- For satellite data: What kind of imaging or sensing technology is involved?
- For probe data: How are the probes (like vehicles with GPS devices) tracking and transmitting data?

What is the sampling frequency?

- How often is the data collected (e.g., every second, minute, hour)?
- Where is the data hosted, locally or in the cloud?

Spatial and Temporal Coverage:

- What geographical areas are covered, and for what duration? Specific parts of the road network? For a complete year?

What specific data is collected?

- For mobile network data: Is it location data, movement patterns, app usage?
- For satellite data: Is it images, heat maps, traffic congestion visuals?
- For probe data: Is it speed, direction, vehicle type?

Data Detail Level:

- Does the data distinguish between different types of vehicles, like cars, trucks, buses?
- Can the data be aggregated per hour, per weekday, per month at link level?

Can the data be shared?

- Are there privacy concerns or data sharing restrictions?
- How can they be solved? Aggregation? anonymization?

What are the costs involved in data acquisition and maintenance?

- Are there subscription fees, equipment costs, or data processing charges?

Cost-Benefit Analysis:

- How do the costs compare with the potential benefits or insights gained? Equipment/infrastructure cost: low, medium, high?

What type of data processing is needed?

- Is specialized software or hardware required to process and analyze the data? Open source or Proprietary?
- Is there a need to develop ‘new methods’ or modify/adapt existing methods for the purpose of estimating traffic variations and traffic flow ranges? Stage of research: infancy, mature, ready?

Complexity of Data Integration:

- How easy is it to integrate and correlate data from these data sources with already available data, e.g., data from pneumatic tubes (Traffic Measurement System - TMS), data from automatic plate recognition (Electronic Congestion Charging - Stockholm)?

Real-Time Processing:

- Can the data be processed in real-time for immediate traffic analysis?

Who is responsible for data collection and management?

- Is it a government body, private company, or a collaboration?
- What would be the eco-system (collaboration) for this type of data to achieve our purpose?
- Any idea of a Business Model?

Data Security and Privacy:

- What measures are in place to ensure data security and user privacy?

Regulatory Compliance:

- Are there any legal or regulatory standards that need to be met?

Appendix 3

Workshop Program

Workshop on alternative data sources for estimating traffic variations: Annual Average Daily Traffic (AADT), traffic variation indices, and traffic flow ranks.

Location: Digital

Date: 15th of December 2023

Time: 09.00 - 12.00

Block 1: Introduction and Background (09.00 - 09.30)

09.00 - 09.05: Welcome and technical check-in

A brief introduction of participants (name + institution), the agenda, and technical instructions for effective participation via Teams.

09.05 - 09.15: Presentation of background, purpose, and goals of the workshop

Overview of the background, the purpose of the workshop, and the desired end goals.

09.15 - 09.30: Overview of current methods for developing traffic variation indices

Presentation of the current methods for calculating AADT, traffic variation indices and traffic flow range curves.

Discussion about the limitations of existing data sources.

Block 2: Exploration of alternative data sources (09.30 - 10.45)

09.30 - 09.45: Presentations of new potential data sources

Short presentations (5 min) by invited speakers who possess relevant data.

- Mobile network data
- Satellite data
- Probe data

09.45 - 09.50: Coffee break

09.50 - 10.50: Group Discussions and Idea Generation

Participants are divided into smaller groups to discuss the opportunities and challenges with the presented data sources. Each group brainstorms on how these data can be integrated to improve existing methods of traffic variation estimation.

Block 3: Compilation and Action Plan (10.50 - 12.00)

10.50 - 11.00: Coffee break

11.00 - 11.30: Compilation of Group Discussions

Each group presents their thoughts and ideas to the other participants (Each group 10 has minutes).

11.30 - 11.50: Workshop on Data Integration and Method Development

Joint discussion on how we can proceed with integrating the new data sources and developing the methods for improving/enriching the estimation of AADT, traffic variations and traffic flow ranges.

11.50 - 12.00: Next Steps and Conclusion

Summary of the day and discussion on the next steps in the process.

Thanks to the participants and information on how to continue to engage in the project.

The Swedish National Road and Transport Research Institute (VTI) is an independent and internationally prominent research institute in the transport sector. We conduct research and development to advance the state of knowledge within infrastructure, traffic, and transport. Through our work we contribute to the attainment of Sweden's transport policy goals related to accessibility, safety, environment, and health.

We conduct commissioned research within all modes of transport and work in an interdisciplinary organisation. Knowledge that we develop provides important information for stakeholders in the transport sector and in many cases leads to direct applications within both national and international transport policies.

As well as research we also undertake investigations, provide counselling, and perform various services related to measurement and testing. At VTI we have a wide range of advanced research equipment along with world-class driving simulators. We also have accredited laboratories for road material testing and crash safety testing.

The library at VTI is a national resource that collects and disseminates information in the field of Swedish transport research. As well as answering queries and lending publications the library also offers services such as information searches, monitoring developments within the field, and maintaining a website with a structured catalogue of publications and projects.

In Sweden, VTI collaborates with universities that conduct related research and education. We participate regularly in international research projects, primarily in Europe, and are active within international networks and alliances. We have about 240 employees and are located in Linköping, Stockholm, Gothenburg and Lund.

vti

Swedish National Road and Transport Research institute • www.vti.se • vti@vti.se • +46 (0)13–20 40 00