Texas A&M Transportation Institute

Task 2.2

Emerging Air Quality Issues Analysis –

Emerging Data and Applications for Transportation Air Quality Analysis

INTERIM MEMORANDUM- DRAFT

Prepared for the Texas Department of Transportation

August 2017

Environment and Air Quality Division

Inter-Agency Contract (Contract No: IAC 0000007200)

Sub-Task 2.2 Emerging Air Quality Issues Analysis – Emerging Data and Applications for Transportation Air Quality Analysis

DATE: August 14, 2018

TO: William Knowles, P.E.
Texas Department of Transportation (TxDOT)

COPY TO: Janie Temple, TxDOT
Laura Norton, TxDOT

FROM: Reza Farzaneh, Ph.D., P.E.
Tara Ramani, Ph.D., P.E.
Rohit Jaikumar, Ph.D.
Robert Huch, P.G., CPESC
Joe Zietsman, Ph.D., P.E.
Environment and Air Quality Division,
Texas A&M Transportation Institute

FOR MORE INFORMATION:

Reza Farzaneh
(512) 407-1118
Reza.Farzaneh@tti.tamu.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

BACKGROUND

Advancements in data science and data visualization have impacted the transportation sector significantly in recent years. Several studies have explored the topics of "big data" (emerging data) and advancements related to data, and the implications for transportation agencies (*1–5*). These include:

1. Technological advances that allow for large amounts of data to be collected cheaply and at a more disaggregate level than before,
2. Commercial availability of more powerful data handling platforms and software; and,
3. Increased emphasis on data visualization for better communication of data analysis results.

The transportation community has recognized the importance of data-related topics, with the Transportation Research Board (TRB) sponsoring research projects and addressing data issues through the activities of its standing committees (*6*, *7*). The Texas Department of Transportation (TxDOT) and other entities in Texas have also conducted several studies focused on Texas-specific data for transportation planning (*8*). TxDOT is also moving towards integrating data sources and making data publicly available (*9–11*). The federal performance measurement requirements, first instituted by MAP-21 and carried forward by the FAST Act, have also supported data-driven and performance-based planning. Resources such as the National Performance Management Research Dataset (NPMRDS) also provide opportunities for more advanced use of data by transportation agencies. However, these types of data sources and applications also require a transition to using different tools and approaches to data handling and analysis. The traditional tools used by transportation agencies are often not sufficient to handle the magnitude of data efficiently, and a more integrated and holistic approach to data management is needed moving forward (*12*).

EMERGING DATA FOR AIR QUALITY ANALYSES

Transportation and air quality planning have always relied on assembling the best available data for modeling and decision-making purposes. Traditionally, data for air quality analyses derived from the travel demand modeling process, which itself provided data at a fairly aggregated level, and relied on expensive travel surveys and similar data collection exercises. As with other transportation applications, transportation air quality analyses also stand to benefit from advancements with regards to emerging data and data applications. Potential advantages could include:

- Use of larger amounts of data
- Use finer-scale/disaggregate data
- Ability to conduct real-time and/or predictive analyses
- Ability to overlay and integrate multiple datasets, including spatial analyses.
- Ability to visualize and display data in innovative ways

However, there are limited studies available that specifically discuss "big data" or emerging data specifically in the context of transportation air quality, and discuss the skills and tools needed for practitioners to more efficiently leverage data and advancements in this area.

Study Need and Goal

A consensus exists that advancements in the area of data can impact transportation analyses, including air quality analyses. However, the available literature and information resources are lacking a perspective tailored to transportation air quality. Specifically, the current state of practice does not address skills and tools needed for transportation air quality practitioners to handle the area of data as an emerging discipline.

The aim of this study is to fill this gap by exploring the following:

- Data handling processes, considerations, and relevant tools
- Relevant datasets and data sources available to transportation practitioners
- Examples of air-quality analyses that have used innovative/emerging sources of data, or examples of analyses that could be extended into the air quality realm

This study was performed by Texas A&M Transportation Institute (TTI) staff under Subtask 2.2 (Emerging Air Quality Issues) of the TTI-TxDOT Air Quality and Conformity Interagency Contract (IAC 0000007200).

TERMINOLOGY AND CONCEPT

In this study, we use the term “emerging data and applications” to collectively refer to advancements in data science that cover: 1) new sources of data (including data that is being newly-leveraged in the transportation sector and data collected through new technologies, including finer scale or fine resolution data), 2) tools and methods for advanced data handling, 3) tools, approaches and applications for data visualization. The term “big data” is often used to describe these emerging sources. For example, a Federal Highway Administration (FHWA) white paper discussed big data for transportation applications as covering data capture, data management, and data analysis elements (*13*). Generally, the term “big data” is taken to refer not just to the size of the data (i.e. data volume), but also data of increasingly diverse types and categories (i.e. data variety) and frequency of analyzing and extracting information (i.e. data velocity). Any combination of these parameters can be viewed as a big data application. Figure 1 shows this conceptual interaction between the data volume, velocity, and variety in the context of big data applications.

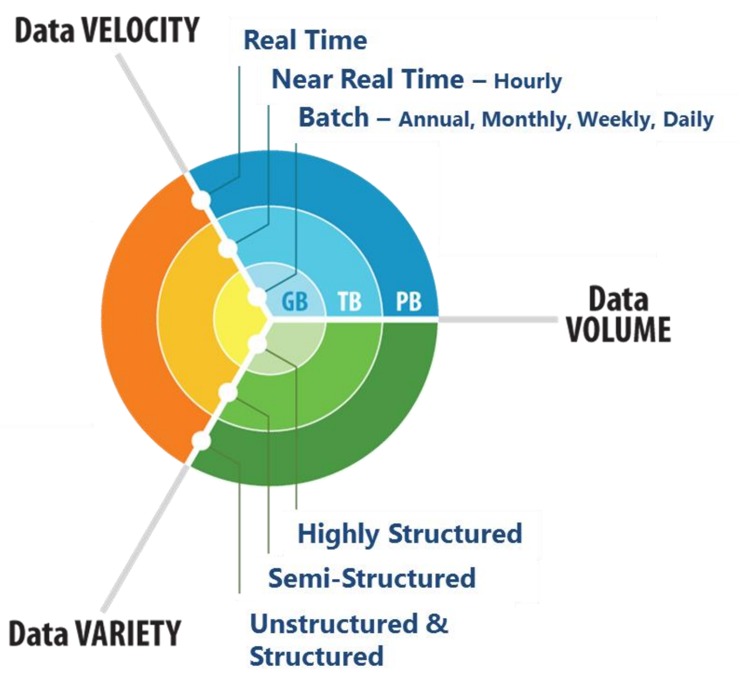

Figure 1. Conceptualization of Big Data. Source*: (14)*

However, the term big data can be viewed as be somewhat limiting, referring specifically to the context of business intelligence and analytics usually undertaken in a corporate setting (*15*). This report therefore does not explicitly use the term, and covers issues related to data sources, handling practices, and potential applications under the umbrella of “emerging data applications”.

Data analytics is another important term in the context of using emerging data sources. Data analytics involve a variety of quantitative techniques to extract actionable information from data. The traditional data analysis approach focuses primarily on descriptive information on activities and trends of the past. Modern data analytics techniques, by contrast, use large volumes and varieties of data at very high speeds (i.e. velocity) to accurately forecast trends, identify significant events and anomalies, and discover underlying patterns.

Historically, private businesses and large corporations have been at the forefront of using data analytics and big data systems to improve their operations, reduce cost, and ultimately increase profitability. Many public agencies have also started using big data, though they are found to be lagging behind the public sector in integrating and benefitting from these applications. A survey of public agencies pointed to staffing challenges, Information Technology (IT) infrastructure and analytics tools, and budget shortfalls as the main challenges faced (*16*).

THE DATA LIFE CYCLE

In the discussion of the implications of big data for transportation, the FHWA defined big data as a "*process of knowledge generation*" that covered the following elements (*13*):

- Data capture - that includes massive datasets encompassing all or most of the population being studied (as opposed to small samples)
- Data management - that features storage in decentralized and virtual locations (i.e., the cloud) and handles both structured and unstructured data.
- Data analysis - that is often automated, with computers doing more of the work to find complex patterns among a large number of variables.

From a systems thinking perspective, it has also been emphasized that data is the first step in a hierarchy, from which it needs to be translated to information (i.e. data organized in a form that is useful for analysis), followed by knowledge (i.e. information that is combined with experience, context and interpretation to provide an understanding of the problem at hand (*17*) . The application of data, from its collection to its use to generate knowledge, is often termed as the "data life cycle". The exact steps as defined vary from source to source, but they broadly cover the elements of data acquisition, management and analysis (*13*, *17*, *18*). This report discusses the elements of the data life cycle as follows, consistent with the steps used by Jagadish et al. (*18*):

- Data acquisition
- Information extraction and cleaning
- Data integration, aggregation, and representation
- Modeling and analysis
- Interpretation.

It should be noted that these steps represent distinctions along a continuum that relates to the use of data to inform decision making-or policy, as shown in Figure 2. However, breaking down the data life cycle can help understand challenges and needs raised by a changing paradigm in which data science is viewed as a standalone area of expertise separate from technical or subject matter expertise in other fields.

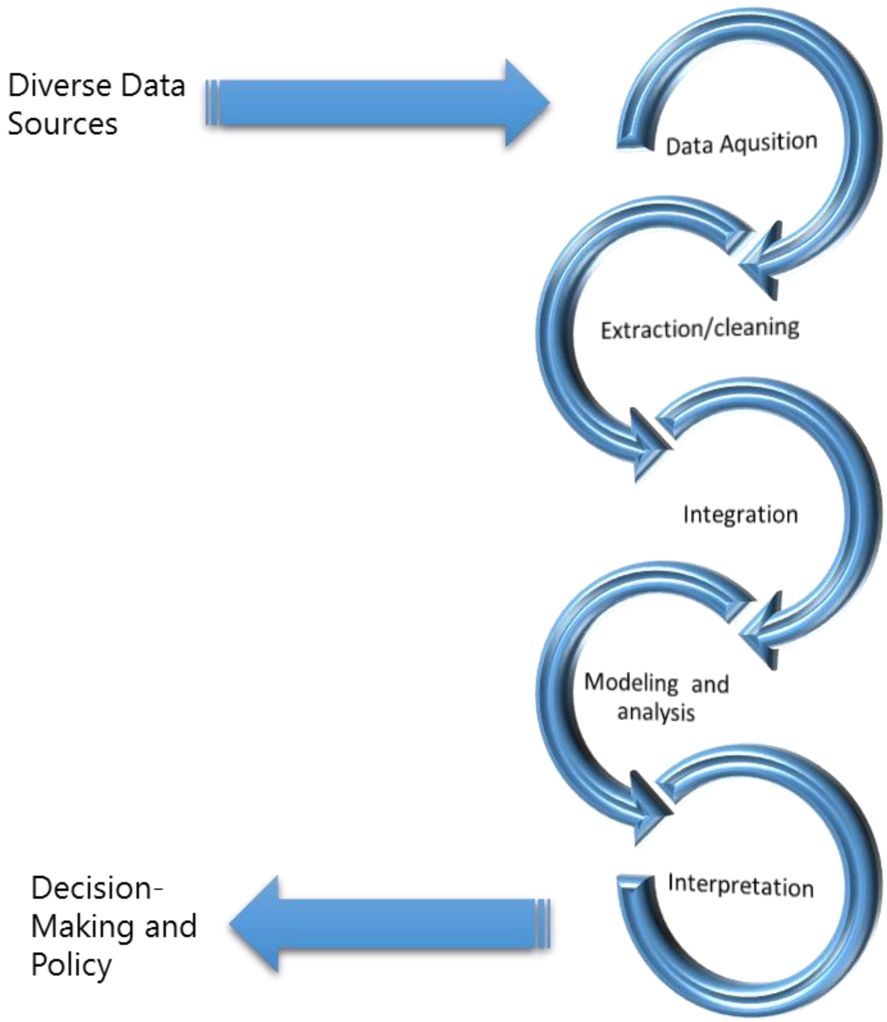

Figure 2. Elements of Data Handling Process. Adapted from (*18*).

CONSIDERATIONS FOR AN EMERGING DATA PARADIGM

As noted in the introductory section, public agencies (including transportation agencies) face challenges in the area of emerging data due to constraints related to staffing and skills, IT infrastructure, and budgets. These are described further below:

- *Staffing challenges:* the growing volume and complexity of emerging data requires capacity and expertise to translate data into actionable information. Traditionally, data analysts focused on specific types and scopes of data and most of them lack the skills and experience to leverage the volumes and varieties of available data. Further, public agencies looking to hire additional employees with advanced data analytics skills face stiff competition from the private sector.
- *Lack of IT infrastructure and analytics tools:* Traditional analysis and visualization methods and tools which are commonly used by public agencies often fail to handle large amount of data and effectively ingest multiple data formats and structures. Working with emerging data requires a new set of tools that have not been traditionally used by public agencies and require specific expertise to use them. For example, the new data files can no longer be stored on personal computers and processed using common tools such as MS Excel and Access.
- *Budget limitations* – It is widely recognized that emerging datasets and advanced analytics can bring value to an agency. However, this requires initial investment and capacity-building to develop a working data system. Public agencies operating under budget constraints find it challenging to allocate resources to support these efforts.

It light of these challenges, it is important to view the advances in emerging data as an ongoing issue that transportation agencies must address. A review of literature discussing emerging data, including in the context of transportation agencies illustrates the need for a paradigm shift in how data is handled and processed (*4*, *6*, *13*, *18*–*21*). Figure 3 illustrates the key elements of this paradigm shift. It includes the use of larger samples and volumes of data at more disaggregated levels, the ability to link multiple datasets, and for spatial and temporal disaggregation. Improved data management and data warehousing practices, the use of business intelligence (BI) and automation (through artificial intelligence and machine learning approaches), and the emergence of data science as a discipline of study in itself.

Old Paradigm	New Paradigm
Aggregated data	Spatial and temporal disaggregation
Smaller samples	Larger samples
Static approaches – spreadsheets/small databases	Dynamic approaches– data warehousing and BI
Single source	Multiple sources, ability to overlay and integrate
Smaller file sizes	Larger file sizes
Data analysis by subject matter experts	Data science as a discipline
Manual or effort-intensive analyses	Automation – artificial intelligence and machine learning

Figure 3. Paradigm Shift due to Emerging Data and Analysis Approaches

In this changing landscape, transportation agencies must approach the use of emerging data in a systematic manner. This is important in order to leverage the potential benefits from emerging data and data analysis methodologies.

Table 1 provides an overview of considerations, skills and tools required, structured along the elements of the data life cycle identified previously. A generic description of each step of the data lifecycle is also provided. The table is developed based on a review of the literature cited previously, and supplemented by the experience of the TTI team in dealing with advancements in data.

Table 1. High-Level Considerations for an Emerging Data Paradigm

Step	Description	Challenges/Considerations	Applicable Skills and Tools
Data Acquisition	Process of identifying relevant data and filtering them out from larger datasets based on user needs.	Privacy and data ownership in light of individual and disaggregate data that can be personally identifying. Data volume, timeliness of analysis (in case of real-time data.	Skills – Data intuition, data querying, pattern recognition Tools - Data logging and sensor platforms such as GPS, on-board vehicle diagnostics, cellphone data, Apache hive, Sqoop, power query, power BI
Data Extraction, Cleaning and Storage	Defining the ways in which usable information is extracted from an acquired data set, checking data for quality and organizing and storing data for further use.	Inconsistency and incompleteness in data Reliability of sources, uncertainty, errors, and missing values must be managed. Use of appropriate database structures and software to handle data.	Skills – Data wrangling, database management, data mining, pattern recognition, Tools – Python, SQL, NoSQL, Power BI etc. Advances in data storage technologies.
Data Integration, Aggregation, Representation	Analytical methodologies to integrate multiple data sources and enable the use of the data for further modeling and analysis.	Heterogeneity of data including unstructured and structured data. Representativeness of the data and possibility of biased samples. Use of appropriate database structures and software to handle data.	Skills – Data mining, data fusion Tools – Apache Hadoop, Spark, etc.
Modeling and Analysis	Process of using the data for purposes such as building and running models or conducting analytical processes to quantify or represent issues or trends of interest.	Emerging data are fundamentally different from traditional modeling data. Need to understand implications of heterogeneity, noise in the data, representativeness, etc.	Skills - Machine learning, data mining, statistical and quantitative analysis, programming. Tools - R, SAS, SPSS, Python, GIS
Interpretation	Using the results of the data combined with practitioner knowledge and understanding of the context to support decisions and frame policy.	Include human perspective and experience, not purely data-oriented. Visualization and collaboration must support input from multiple human experts, and shared exploration of results.	Skills – Data visualization, data shaping Tools – R, Power BI, Tableau, Qlikview, GIS

DATA FOR TRANSPORTATION AIR QUALITY ANALYSES

A main element of transportation air quality analyses is the need for data related to the characterization of traffic or traffic activity. Traditionally, the traffic data for air quality analysis come from regional travel demand models (TDM) and case-specific traffic analysis based on short-term (e.g., 24 or 48 hours) observations. However, non-traditional sources of traffic data have steadily gained ground in the past few years, both in terms of quantity/coverage and quality. These data sources are easier to access (e.g., web-based) and provide hourly or sub-hourly details of the traffic on a section of the road.

Traffic data, at a minimum, includes traffic volumes and speeds, and fleet composition at the roadway link level. The traffic data must be consistent with the location and timeframe of the desired analysis. Currently, state DOTs including TxDOT have access to data from federal datasets, state and local public sector entities, and the private sector. Several ancillary datasets also allow for extension of traffic and air quality data for additional analyses, linkage with other spatial datasets, etc.

Figure 4 provides a summary of selected sources of data. These do not represent an exhaustive list of data sources, but rather a summary of “typical” traditional and emerging data sources relevant in the transportation air quality analysis landscape. A summary and further discussion of these sources is presented in the Appendix.

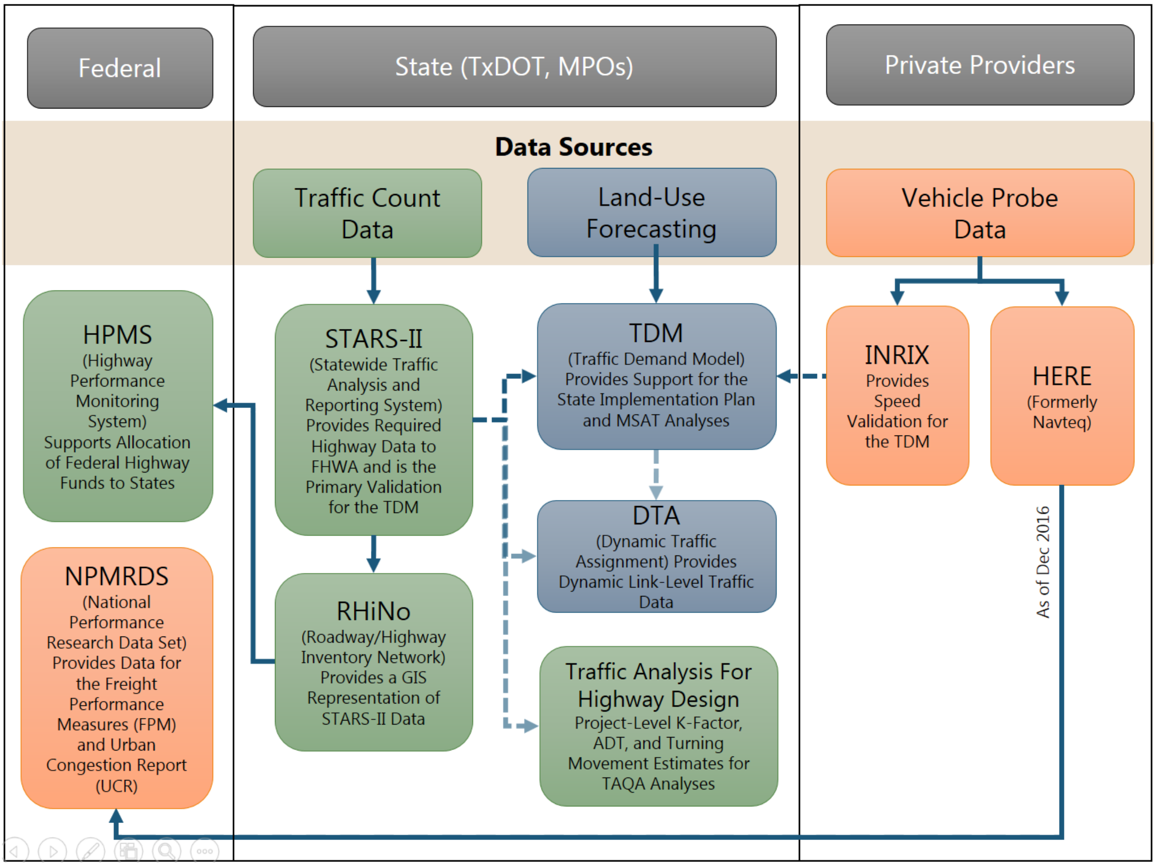

Figure 4. Traffic Data Sources and Uses

TRAFFIC AND AIR QUALITY-RELATED APPLICATIONS OF EMERGING DATA

As mentioned in the introductory section, emerging data and its applications have several potential benefits. This includes larger samples of data at finer scales, real-time analyses, ability to overlay and integrate data from multiple data sources, and visualize and display data in innovative ways. This section describes applications specifically in the context of traffic data that can support emissions and air quality analyses, or direct applications for air quality analyses. These include:

- GPS data with GIS-based location matching - to enhance traffic modeling and drive cycle development (*22*), and to provide trip information to support travel demand modeling (*23*).

- GPS and On-Board Diagnostics (OBD) for emissions and energy estimates – this includes the use of GPS and OBD data for emissions estimation (*24*, *25*), identification of eco-driving techniques (*26*), and for using real-world data as a substitute for traffic microsimulation results in emissions estimates (*27*).

- GPS Data and Cellular Phone Data Records in Travel Demand Modeling and Traffic Modeling – this currently represents the largest applications of emerging data (*28–30*). This includes use of GPS and cellphone data as a substitute or supplement to travel surveys (*31*, *32*), for assessing origins and destinations (O-D) and travel demand (*33–37*), examining individual mobility patterns (*38*) and behavioral analyses (*39*), tracking of truck routes (*33*) as well as for detection of travel mode(*40–46*).
- Advances in Data Handling and Visualization – advances discussed in the literature include techniques for automated data quality checks and data cleaning (*47*) and automated classification of data such as driving modes (*48*). Data warehousing and storage techniques for improved data handling include the development of customized platforms and combined datasets (*49*), and moving away from traditional databases to more efficient systems (*50*, *51*). Finally, advanced data visualization systems allow for spatiotemporal/space-time visualizations of traffic or transportation, with linkage to GIS datasets (*52–54*).

Some of the examples discussed in this section directly relate to transportation air quality analyses, while others provide inputs that can be leveraged for producing air quality analyses that rely on more disaggregated data, real-time data, or finer-scale data in terms of vehicle trajectories and speed profiles. The ability to combine information from multiple datasets (for example individual vehicles' behavior in relation to overall prevailing traffic patterns) and better assessment and spatial classification of data using GIS and other techniques can also support improved emissions analyses as an extension of improved modeling processes. It can also support the visualization of emissions and dispersion through GIS overlays, and the visualization of spatial and temporal changes in regional emissions.

CONCLUSIONS AND NEXT STEPS

This report provides an overview of advancements in data science and potential implications for transportation air quality analyses. Key conclusions and observations include the following:

- Turning data into useful knowledge and information has emerged as a field in its own right. Handling emerging data requires an understanding of the entire data life cycle, from acquisition to the interpretation of data after analysis.
- The availability of more disaggregated, real time data from multiple sources presents both challenges and opportunities. It requires specialized skills and resources that transportation agencies may not currently have, as well as a paradigm shift in how data analyses are viewed.
- Some key skills that are relevant to handle emerging data needs include GIS skills, statistical analysis skills, programming skills, knowledge of business intelligence and analytics, and data visualization expertise. It is also important to acknowledge that sound subject-matter and technical knowledge is also required to ensure that the emerging data applications fit within the context in which TxDOT operates.
- A review of selected federal, state and private sector data sources currently used by TxDOT and its partners for transportation air quality analyses indicate that emerging data sources are being considered and used, such as in the examples of AirSage data, data from INRIX and NMPRDS, etc.
- A review of literature on emerging data-related applications relevant to air quality analyses show that there are several opportunities to advance traffic and travel demand modeling to better support air quality analyses.

Further, TTI has also explored the use of emerging data for air quality analyses, including the STARS-II application conducted as part of a previous IAC, and investigation of the use of NPMRDS data for air quality analyses in discussion with Texas MPOs. Additional opportunities and applications can include:

- Identification of congestion and emissions hot-spots in Texas or Texas metropolitan areas

- Improvement of the regional emissions estimation (conformity analysis) process and inventory development

- Using spatial analyses for identification of off-network and idling events for improved emissions modeling.

TTI has also incorporated the knowledge gained from this study into ongoing TxDOT and non-TxDOT initiatives, including TxDOT research projects. TTI can also work with TxDOT to provide training, guidance, or develop best-practices for the use of emerging data in transportation air quality analyses. TTI will work with TxDOT to identify priority areas of interest and investigation for FY2018.

REFERENCES

1. Kanniyappan R, McQueen B. What's the Big Deal about Big Data in Transportation? In: Data Symposium, Florida Department of Transportation. Orlando, Florida: Teradata; 2014.
2. Olavsrud T. 17 Steps to Implement a Public Sector Big Data Project | CIO. http://www.cio.com/article/2368491/big-data/144854-17-Steps-to-Implement-a-Public-Sector-Big-Data-Project.html#slide1. Published 2014. Accessed August 17, 2017.
3. Kochhar D. Big Data in Public Transportation - Hortonworks. hortonworks.com. https://hortonworks.com/blog/big-data-public-transportation/. Published 2016. Accessed August 17, 2017.
4. USC. TransDec: Big Data for Transportation. University of Southern California. http://imsc.usc.edu/intelligent-transportation.html. Published 2017. Accessed August 17, 2017.
5. Grogan T. A Brief History of Big Data (in Federal Transportation) - The Eno Center for Transportation. ENO TRANSPORTATION WEEKLY. https://www.enotrans.org/article/brief-history-big-data-federal-transportation/. Accessed August 17, 2017.
6. Higgins, N., R. Basile, S. Van Hecke, J. Zissman, and S. Gilkeson. *Data Visualization Methods for Transportation Agencies*. National Academies Press, 2016.
7. TRB. TRB Committees | Data and Information Technology. TRB.org. http://www.trb.org/DataInformationTechnology/TRBCommittees.aspx. Published 2017. Accessed August 17, 2017.
8. Quiroga C, Li Y, Koncz N, Overman J. Analysis and Integration of Spatial Data for Transportation Planning. Vol FHWA-TX-09.; 2009.
9. TxDOT. TxDOT One-Stop Data Analysis. http://idserportal.utsa.edu/txdot/onestop/Sources.aspx. Published 2017. Accessed August 17, 2017.
10. TxDOT. OneDOT Data Shop. http://www.txdot.gov/inside-txdot/division/planning/one-data.html. Published 2017. Accessed August 17, 2017.
11. TxDOT. Statewide Traffic Analysis and Reporting system (STARS II). In: TxDOT Transportation Planning Conference. Corpus Christi; 2014. http://txdot.ms2soft.com/%0A.
12. Jain M. Transportation Data Management and Analysis (TDMA). In: Southeast Florida FSUTMS Users Group. ; 2014.

13. Burt M, Cuddy M, Razo M. White Paper - Big Data's Implications for Transportation Operations: An Exploration.; 2014.

14. Phemi Inc. Big Data Is Not Just About "Big." https://phemi.com/products/phemi-central/why-big-data/.

15. Chen H, Storey VC. Business Intelligence and Analytics: From Big Data to Big Impact. Mis Q. 2012;36(4):1165-1188. doi:10.1145/2463676.2463712.

16. *Unisys, Big Data: The Path to Mission-Centric Analytics. 2015. http://assets.unisys.com/Documents/Federal/Report_20151005_BigDataThePathtoMissionCentricAnalytics.pdf*

17. Vandervalk, A. Turning Data Into Information for Transport Decision Making. *Association for European Transport and Contributors*, 2012, pp. 1–20.

18. Jagadish, H., J. Gehrke, and A. Labrinidis. Big Data and Its Technical Challenges. *Communications of the*, 2014.

19. PCAST. Big Data and Privacy: A Technological Perspective. *Science and Technology*, No. May, 2014, p. 76.

20. Khaleghi, B., A. Khamis, F. O. Karray, and S. N. Razavi. Multisensor Data Fusion: A Review of the State-of-the-Art. *Information Fusion*, Vol. 14, No. 1, 2013, pp. 28–44. https://doi.org/10.1016/j.inffus.2011.08.001.

21. McKinsey & Company. Big Data: The next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*, No. June, 2011, p. 156. https://doi.org/10.1080/01443610903114527.

22. Barth, M., E. Johnston, and R. Tadi. Using GPS Technology to Relate Macroscopic and Microscopic Traffic Parameters. *Transportation Research Record*, Vol. 1520, No. 1, 1996, pp. 89–96. https://doi.org/10.3141/1520-11.

23. Axhausen, K., SchöUnfelder S., J. Wolf, M. Oliveira, and U. Samaga. Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information. *Transportation Research Record*, Vol. 1870, No. August, 2004, pp. 46–54. https://doi.org/10.3141/1870-06.

24. Eisinger, D. S., D. a. Niemeier, T. Stoeckenius, T. P. Kear, M. J. Brady, a. K. Pollack, and J. Long. Collecting Driving Data to Support Mobile Source Emissions Estimation. *Journal of Transportation Engineering*, Vol. 132, No. 11, 2006, pp. 845–854. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:11(845).

25. Alessandrini, a, F. Filippi, F. Orecchini, and F. Ortenzi. A New Method for Collecting Vehicle Behaviour in Daily Use for Energy and Environmental Analysis. *Proceedings*

of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, Vol. 220, No. 11, 2006, pp. 1527–1537. https://doi.org/10.1243/09544070JAUTO165.

26. Barth, M., and K. Boriboonsomsin. Energy and Emissions Impacts of a Freeway-Based Dynamic Eco-Driving System. *Transportation Research Part D: Transport and Environment*, Vol. 14, No. 6, 2009, pp. 400–410. https://doi.org/10.1016/j.trd.2009.01.004.
27. Beckx, C., L. I. Panis, D. Janssens, and G. Wets. Applying Activity-Travel Data for the Assessment of Vehicle Exhaust Emissions: Application of a GPS-Enhanced Data Collection Tool. *Transportation Research Part D: Transport and Environment*, Vol. 15, No. 2, 2010, pp. 117–122. https://doi.org/10.1016/j.trd.2009.10.004.
28. Horn, C., S. Klampfl, M. Cik, and T. Reiter. Detecting Outliers in Cell Phone Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2405, 2014, pp. 49–56. https://doi.org/10.3141/2405-07.
29. Huntsinger, L. F., and K. Ward. Using Mobile Phone Location Data to Develop External Trip Models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2499, 2015, pp. 25–32. https://doi.org/10.3141/2499-04.
30. Herrera, J. C., D. B. Work, R. Herring, X. (Jeff) Ban, Q. Jacobson, and A. M. Bayen. Evaluation of Traffic Data Obtained via GPS-Enabled Mobile Phones: The Mobile Century Field Experiment. *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 4, 2010, pp. 568–583. https://doi.org/10.1016/j.trc.2009.10.006.
31. Nitsche, P., P. Widhalm, S. Breuss, and P. Maurer. A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys. *Procedia - Social and Behavioral Sciences*, Vol. 48, 2012, pp. 1033–1046. https://doi.org/10.1016/j.sbspro.2012.06.1080.
32. Nitsche, P., P. Widhalm, S. Breuss, N. Br??ndle, and P. Maurer. Supporting Large-Scale Travel Surveys with Smartphones - A Practical Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 212–221. https://doi.org/10.1016/j.trc.2013.11.005.
33. Wang, M., and S. D. Schrock. Feasibility of Using Cellular Telephone Data to Determine the Truckshed of Intermodal Facilities. *Cell*, No. August 2009, 2012.
34. Cheng, X., W. Li, F. Jia, D. Yang, and Z. Duan. Analyzing Human Activity Patterns Using Cellular Phone Data: Case Study of Jinhe New Town in Shanghai, China. 2013.
35. Chen, C., L. Bian, and J. Ma. From Traces to Trajectories: How Well Can We Guess Activity Locations from Mobile Phone Traces? *Transportation Research Part C: Emerging Technologies*, Vol. 46, 2014, pp. 326–337. https://doi.org/10.1016/j.trc.2014.07.001.

36. Berlingerio, M., F. Calabrese, G. Di Lorenzo, R. Nair, F. Pinelli, and M. Luca Sbodio. AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data. In *Machine Learning and Knowledge Discovery in Databases*, pp. 668–684.

37. Fang, J., M. Xue, and T. Z. Qiu. Anonymous Cellphone-Based Large-Scale Origin-Destination Data Collection: Case Studies in China. 2014.

38. Calabrese, F., M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti. Understanding Individual Mobility Patterns from Urban Sensing Data: A Mobile Phone Trace Example. *Transportation Research Part C: Emerging Technologies*, Vol. 26, 2013, pp. 301–313. https://doi.org/10.1016/j.trc.2012.09.009.

39. Ythier, J., J. L. Walker, and M. Bierlaire. The Influence of Social Contacts and Communication Use on Travel Behavior: A Smartphone-Based Study. 2013.

40. Reddy, S., J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Determining Transportation Mode on Mobile Phones. 2008.

41. Gonzalez, P. A., J. S. Weinstein, S. J. Barbeau, M. A. Labrador, P. L. Winters, N. L. Georggi, and R. Perez. Automating Mode Detection for Travel Behaviour Analysis by Using Global Positioning Systems-Enabled Mobile Phones and Neural Networks. *IET Intelligent Transport Systems*, Vol. 4, No. 1, 2010, p. 37. https://doi.org/10.1049/iet-its.2009.0029.

42. Lu, H., J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell. The Jigsaw Continuous Sensing Engine for Mobile Phone Applications. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems - SenSys '10*, 2010, p. 71. https://doi.org/10.1145/1869983.1869992.

43. Stenneth, L., O. Wolfson, P. S. Yu, and B. Xu. Transportation Mode Detection Using Mobile Phones and GIS Information. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2011, p. 54. https://doi.org/10.1145/2093973.2093982.

44. Jahangiri, A., and H. Rakha. Developing a Support Vector Machine (SVM) Classifier for Transportation Mode Identification Using Mobile Phone Sensor Data 2. *Transportation Research Board 93rd Annual Meeting*, No. 93, 2014.

45. Lari, A. Automated Transportation Mode Detection Using Smart Phone Applications via Machine Learning: Case Study Mega City of Tehran. *Transportation Research Board 94th Annual Meeting*, Vol. 6147, 2015.

46. Das, R. D., N. Ronald, and S. Winter. A Simulation Study on Automated Transport Mode Detection in near-Real Time Using a Neural Network. *CEUR Workshop Proceedings*, Vol. 1323, No. March, 2015, pp. 46–57.

47. Stopher, P., C. FitzGerald, and J. Zhang. Search for a Global Positioning System Device to Measure Person Travel. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 3, 2008, pp. 350–369. https://doi.org/10.1016/j.trc.2007.10.002.
48. Belz, N. P., and L. Aultman-Hall. Analyzing the Effect of Driver Age on Operating Speed and Acceleration Noise. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2265, No. 1, 2012, pp. 184–191. https://doi.org/10.3141/2265-21.
49. Ma, X., Y.-J. Wu, and Y. Wang. DRIVE Net. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2215, No. 1, 2011, pp. 37–49. https://doi.org/10.3141/2215-04.
50. Yu, M.-C., T. Yu, S.-C. Wang, C.-J. Lin, and E. Y. Chang. Big Data Small Footprint. *Proceedings of the VLDB Endowment*, Vol. 7, No. 13, 2014, pp. 1429–1440. https://doi.org/10.14778/2733004.2733015.
51. Sun, S., and C. Liu. Application of Improved Storage Technology in Intelligent Transportation System. *Advances in Transportation Studies*, Vol. 3, 2016, pp. 51–60. https://doi.org/10.4399/978885489937705.
52. Demiryurek, U., F. Banaei-Kashani, and C. Shahabi. TransDec: A Spatiotemporal Query Processing Framework for Transportation Systems. *IEEE International Conference on Data Engineering*, 2010, pp. 1197–1200. https://doi.org/10.1109/ICDE.2010.5447745.
53. Imawan, A., and J. Kwon. A Timeline Visualization System for Road Traffic Big Data. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2015, pp. 2928–2929. https://doi.org/10.1109/BigData.2015.7364125.
54. Zhang, J., Z. Chen, Y. Liu, M. Du, W. Yang, and L. Guo. Space–time Visualization Analysis of Bus Passenger Big Data in Beijing. *Cluster Computing*, 2017. https://doi.org/10.1007/s10586-017-0890-8.

APPENDIX – DESCRIPTION OF SELECTED DATA SOURCES

TRAFFIC FROM COUNT DATA

Statewide Traffic Analysis and Reporting System (STARS-II)

STARS-II is TxDOT's database of traffic activity across the state. The traffic data is presented through a web-based geographic information system (GIS) interface that can be used to search for and download traffic count data. TxDOT's STARS-II Traffic Count Database System (TCDS) interface (http://txdot.ms2soft.com/tcds/) is hosted using MS2 software[1] and is accessible by the public. The Transportation Planning and Programming Division (TPP) of TxDOT manages the STARS-II program.

The traffic data presented in STARS-II is collected using permanent and temporary traffic counters strategically distributed across the state. TxDOT collects traffic data at approximately 362 permanent traffic count locations, and from temporary locations that are used within a dynamic sampling plan that covers approximately 75,000 to 80,000 locations annually, including Texas-Mexico border bridges. On-system roadways are monitored annually statewide. Urban areas are monitored on a five-year rotating cycle in accordance with TxDOT's Five Year Count Program (previously called urban saturation count program).[2] Manual traffic counts are used to establish axle-to-vehicle ratio for pneumatic tube axle counts.

The results of these traffic count operations are stored in STARS II where the data are validated, analyzed, and made available for decision support. Data includes volume, vehicle classification, speed, weight-in-motion in detail reports (e.g., AADT by year, AADT by day of week by month for year, average hourly traffic by day of week for year) or listing reports (e.g., AADT by day of week by direction for month or year). Data may be downloaded as CVS, PDF, Excel, or TIFF file. Data headers in CVS file are not always labeled clearly, will require cleanup of data. Data is available through the MS2 website is

[1] Midwestern Software Solutions, LLC. (MS2; http://www.ms2soft.com/) provides customized cloud-based transportation data management system software for a fee. MS2 does not provide data, but a way to manage traffic data. In Texas, TxDOT, the Houston-Galveston Area Council (H-GAC), the City of Frisco, the City of McKinney, and City of Richardson use MS2 software.

[2] TxDOT, 2001, Traffic Data and Analysis Manual, http://onlinemanuals.txdot.gov/txdotmanuals/tda/tda.pdf.

available for years 1996 through 2015. Annual Urban Saturation Maps (pdf) are available by district for the years 2010 through 2015.[3] The STARS-II traffic information is used to monitor and analyze traffic to support decision making within TxDOT and other agencies, and submitting annually to the FHWA for the Highway Performance Monitoring System (HPMS).

RHiNo (Roadway/Highway Inventory Network)

RHiNo is part of the Texas Reference Marker System (TRM) that was implemented in 1995. TPP is currently in the process of updating the TRM data system with a more advanced data system known as Geospatial Roadway Information Database (GRID). The RHiNo dataset includes 96,000 state highway records that cover 137 attributes and represent a wide range of items such as reference marker displacement, highway status and type, functional class, maintenance responsibility, AADT for the previous 10 years, truck percentage, urban/rural status, shoulder width, median width, right-of-way width, roadbed width, posted speed limit, surface type and characteristics, and load limits. The TxDOT end of year Roadway Inventory routed linework for 2015 includes data associated with On-System highways, County Roads, Functional Classified City Streets, Toll Roads and Local Streets. Data is downloadable from TxDOT as a spreadsheet, KLM file, or shapefiles. They are updated annually for the TxDOT end of year roadway inventory required by the FHWA, and are available here:
http://gis.txdot.opendata.arcgis.com/datasets?q=RHiNO

Highway Performance Monitoring System

The Highway Performance Monitoring System (HPMS) is a national level highway information system that includes data on the extent, condition, performance, use, and operating characteristics of the nation's highways. The HPMS contains administrative and extent-of-system information on all public roads. Information on other characteristics is represented in HPMS as a mix of universe and sample data for arterial and collector functional systems[4].

[3] TxDOT, Transportation Planning Maps: Urban Saturation Maps, http://www.txdot.gov/inside-txdot/division/transportation-planning/maps.html.

[4] FHWA. Highway Performance Monitoring System (HPMS). Office of Highway Policy Information. https://www.fhwa.dot.gov/policyinformation/hpms.cfm

The HPMS, initially implemented in 1978, is used to serve the data and informational needs of the FHWA. HPMS is a federally-mandated program used by the FHWA to provide data to Congress on the nation's streets and highways. Congress uses the data for allocation of funds to states. HPMS is a cooperative effort among state departments of transportation, local governments, and metropolitan planning organizations to assemble and report the necessary information. Every state collects, maintains, and reports certain data to the FHWA each year according to the methods prescribed in the *Highway Performance Monitoring System Field Manual* for the continuing analytical and statistical database[5]. The roles of FHWA headquarters, FHWA field offices, and state departments of transportation are defined within the manual.

The HPMS data are designed to provide an inventory of all on-system roads and other public roads that are functionally classified. Specific data collected by TxDOT under the HPMS program include location by jurisdiction, the number of lanes, median widths, shoulder widths, and other basic road attributes. The HPMS data model has been developed within a Geographic Information System (GIS) framework, to take full advantage of the spatial relationships that exist between data elements which are both internal and external to HPMS. The data model is designed to be flexible in terms of compatibility with other data sources and expandable as additional data becomes available. In addition, the data model is designed to achieve independence with respect to the way in which the various data components relate to one another. This approach allows for future modification to a particular area of the model (e.g. a dataset, or data item) with little or no impact on other datasets or data items. For instance, if a change is needed to a roadway section's surface type (e.g. changing it from a code 1-unpaved to code 2-conventional asphalt concrete), it can be done so without impacting the value that is coded for that section's annual average daily traffic (AADT).

The data model's design is structured in a way that allows external data sources to be used to populate the various data fields in HPMS. For example, NAAQS boundary spatial data can be used to assign a pollutant standard to each roadway section for the purpose of generating area-wide totals (e.g. vehicle miles of travel).

---

[5] TxDOT, 2001, Traffic Data and Analysis Manual, http://onlinemanuals.txdot.gov/txdotmanuals/tda/tda.pdf

This data model is organized conceptually into a group of six catalogs. Each catalog groups the various datasets by type and/or function. The types of data can be categorized as:

- geospatial data, representing various highway systems, geographic boundaries etc.,
- roadway attribute data that can be linked to a related GIS dataset, which allows the attribute data to be represented spatially via linear referencing or
- metadata, which provides additional global information about the data[6].

TxDOT district offices collect, update, and submit the required information for roadways within their district to TxDOT's TPP. The data are collected between September 1 and December 31 each year and are submitted to TPP by December 31[7]. TxDOT prepares an annual report to FHWA on or before June 15th each year in accordance with FHWA's *HPMS Field Manual*[8].

Traffic Analysis for Highway Design ("Corridor Process" Reports)

The *Traffic Analysis for Highway Design Memorandum* is prepared by the Transportation Planning and Programming Division (TPP) of TxDOT and includes project-level information that is typically used for NEPA environmental impact studies for noise, Traffic Air Quality Analysis (TAQA), and PM hot spot analysis. The *Traffic Analysis for Highway Design* report typically includes the AADT, K-Factor, and the directional distribution related to peak ADT distributions (the 30th highest hourly volume) as well as turning movement diagrams illustrating the 20-year, and 30-year design period traffic projections. The project-level turning movement diagrams provide a summary of base year and forecasted year node-to-node turning movement for each intersection in the project area. This memorandum is generated by TPP's corridor analyst on request. The

[6] FHWA. Highway Performance Monitoring System Field Manual- Chapter 3: Data Model & Required Datasets. Office of Highway Policy Information. https://www.fhwa.dot.gov/policyinformation/hpms/fieldmanual/chapter3.cfm#chapt3_3_12017.

[7] FHWA Highway Performance Monitoring System Field Manual- Section 5: Highway Performance Monitoring System Traffic Data. Office of Highway Policy Information.

[8] TxDOT. Traffic Data and Analysis Manual.In, Transportation Planning and Programming Division (TPP), 2001

traffic data is based on STARS-II data, project-specific traffic counts, and/or professional judgement.

Travel Demand Modeling

Travel demand modeling is a travel forecasting method used to predict travel characteristics and usage of transport services based on alternative socio-economic, and land-use configurations for a modeled area. TDM is considered a traditional traffic data source for air quality analyses including conformity and NEPA. TDMs are used to evaluate transportation system alternatives that help transportation decision makers at the local and state levels improve the overall function of the transportation system.

TDM can vary in complexity and scale to fit the need of the decision makers. Smaller areas, such as travel related to a corridor or facility can be modeled as well as macro-level TDMs to model regional travel behavior. Broadly, TDM can be approached using trip-based components or activity-based model components.

TxDOT's Transportation Planning and Programming Division (TPP) uses the Texas Package Suite of Travel Demand Models (referred to as the Texas Package) to prepare travel forecasts for urban areas in Texas. The Texas Package is a trip-based set of computer modules based on a traditional four-step travel demand forecasting process that includes trip generation, trip distribution, mode choice, and traffic assignment; however, the Texas Package only uses three of the four traditional steps (Table 2). A mode choice model is not used in the Texas package due to the modest public transportation systems in most Texas urban areas, although Houston, Dallas-Fort Worth, San Antonio, and Austin have developed mode choice models for use in their areas[9].

[9] TxDOT. Traffic Data and Analysis Manual- Section 2: Texas Travel Demand Model Package. http://onlinemanuals.txdot.gov/txdotmanuals/tda/texas_travel_demand_model_package.htm2017.

Table 2. Travel Demand Model Modules.

Modules	Traditional Four-Step Forecasting Process	Texas Package
1. Trip Generation	x	x
2. Trip Distribution	x	x
3. Traffic Assignment	x	x
4. Mode Choice	x	Hou, DFW, SA, Austin only

The TDM traffic predictions are developed based travel between traffic analysis zones (TAZs) for metropolitan areas with populations of over 200,000. The TDM uses TAZs, zone centroids, and centroid connectors developed by TPP, and the appropriate MPO and TxDOT's district office for the area to provide link-level data. The variables used by the TDM include comprehensive travel survey data, U.S. Census data, current and projected socio-demographic data, existing and projected transportation system data, and current traffic data. The TDM is also referred to as a Static Traffic Assignment (STA) since it provides fixed projections of link-level volumes based on model inputs.

This data is validated using TxDOT's STARS-II traffic count data. TDM data files are provided as TransCAD files or as CSV files presenting speed, capacity, area type, functional class, and facility class, etc. Files are typically provided by the MPO for the estimated time of project completion (ETC year) and ETC year plus 20 years. The TDM is validated by comparison of TDM-predicted, base-year traffic to replicate observed traffic counts from the HPMS. Use of TDM data requires coordination between TxDOT, the MPO, and the user to obtain the data. Generally, all modeled links within the metropolitan area are included in the dataset. User must determine the appropriate links for a given project.

TRAFFIC MODELING AND DYNAMIC TRAFFIC ASSIGNMENT (DTA)

Travel demand models provide static traffic assignment that is widely for a range of transportation and air quality analyses. However, the STA model does not account for the time-varying travel conditions of a transportation network. Additionally, STA is unable to model dynamics of traffic flow, such as congestion, queue buildup, bottlenecks, and spillovers in a network.

Traffic modeling to support the macroscopic/deterministic approaches from a TDM can range from microscopic analyses (that model individual vehicles at the corridor or facility level) to mesoscopic modeling that can be undertaking at a regional scale. Dynamic traffic assignment (DTA) can be incorporated into the traditional 4-step modeling process[10] to provide a representation of network-wide traffic flow patterns that go beyond static link flows.

VEHICLE PROBE DATA

INRIX

INRIX Inc. (INRIX) provides privately-owned real-time traffic data (probe data) collected from mobile cellular GPS-based devices (iPhone, Android, BlackBerry and Windows Phone phones), connected cars (Ford SYNC and Toyota Entune), commercial fleet vehicles (trucks, delivery vans, and other fleet vehicles equipped the GPS locator devices), and cameras. INRIX collects freeways and arterial traffic data in the US, Canada, and much of Europe, South America, and Africa (approximately 47 countries) from over 275 million devices over 5 million miles of road at a 100m granularity. Temporal resolution of the data is at 1, 5, 15, 30 and 60 minutes. INRIX data is useful at providing accurate speed data at a high resolution. Speed data can indicate congestion, and is sensitive enough to identify temporary traffic events such as traffic accidents or lane closures. Through geospatial data processing, probe data can identify trip data, such as population origin and destination zones, diversion routes during peak time and incidents, corridor usage, and more.

The data is a subscription/fee based data source, and the company also developed and distributes free mobile applications such as INRIX Traffic and INRIX ParkMe, to provide users with real-time traffic data and information on their commutes.

HERE (Formerly NAVTEQ)/NPMRDS

The National Performance Management Research Data Set (NPMRDS) is a vehicle probe-based travel time data set acquired by the FHWA to support its Freight

[10] Investigating Regional Dynamic Traffic Assignment Modeling for Improved Bottleneck Analysis: Final Report http://library.ctr.utexas.edu/ctr-publications/0-6657-1.pdf

Performance Measures (FPM)[11] and Urban Congestion Report (UCR)[12] programs. Probe data for passenger vehicles is obtained from several sources including mobile phones, vehicles, and portable navigation devices. Freight data is obtained from the American Transportation Research Institute (ATRI) leveraging embedded fleet systems.

The NPMRDS consist of average travel times reported every five minutes on the National Highway System (NHS) as defined in MAP-21[13] and on a five-mile radius of arterials at border crossings. It is monthly archived data. While the data is primarily for FHWA's use, FHWA is making the data available to states and MPOs, as well as to the Canadian and Mexican national governments, border provinces and States, to use to do performance measures and help grow the use and application of performance measures more locally and consistently. There is no charge for the data.

The contractor providing the data is HERE traffic (formerly Nokia/Navteq). HERE is using the American Transportation Research Institute (ATRI) truck probe data for the freight data. HERE is providing the data in three ways: freight truck, passenger vehicle, and all vehicles. Data is downloaded directly from HERE. FHWA maintains NPMRDS data to October 2011.

The NPMRDS data set includes a static file, a monthly data file, and a shape file of the National Highway System (NHS). Agency partners are encouraged to use the data set to develop systems performance measures for use in evaluating projects and operational strategies, as input into investment decision making, as calibration for models, and for producing publicly available reports.

The *static file* provides the roadway information such as the industry standard roadway segment ID (standardized Traffic Message Channel [TMC] location code), county, state, distance (length of TMC in miles), road number, road name, latitude, longitude, road

[11] Freight Performance Measures (FPM) help to identify needed transportation improvements and monitor their effectiveness and serve as indicators of economic health and traffic congestion https://ops.fhwa.dot.gov/freight/freight_analysis/perform_meas/#fhwa.

[12] The UCR is a quarterly snapshot of traffic congestion and reliability trends at the national and city level, developed using archived traffic operations data.

[13] Moving Ahead for Progress in the 21st Century Act (MAP-21) authorizes funds for Federal-aid highways, highway safety programs, transit programs, and for other purposes; and expands the National Highway System (NHS) to incorporate principal arterials not previously included. https://www.fhwa.dot.gov/map21/summaryinfo.cfm.

direction. The static file does not change every month and is updated only as needed to reflect the current roadway system. The *data file* is a .CSV file that includes the TMC code, date, epoch in 5 minute increments over a 24-hour period (0-287), travel time for all vehicles (seconds), travel time for passenger vehicles (seconds), and travel time for freight vehicles (seconds). The data file is provided on a monthly basis and range between 500MB and 4GB in size. The *shape file* includes the National Highway System (NHS) map with the TMC codes and attributes such as: street name, functional class, travel direction, a controlled access flag, and ramp identifiers. Similar to the static file, the shape files do not change every month and are updated only as needed to reflect the current roadway system.

Streetlightdata.com

Streetlightdata.com[14] data is archival, (not real time) and is between a day and a few years old that was processed from cellphone and INRIX data. The data is aggregated, anonymized, and de-identified. Available data starts in January 2014 to within one or two months of the current date. Streetlight metrics include: Origin/Destination Matrices, Select Link Analyses, Average Travel Times and Travel Time Distribution, Internal/External Studies, Commercial and Personal Travel Vehicle Comparisons. Streetlightdata.com uses INRIX data with a 5-meter spatial precision. Traffic can be differentiated by personal vehicles/medium trucks / heavy trucks. The data can be customized down to hourly data and can include distribution of speed (v/m/h). The data is cleaned and processed. Metrics can be customized to specific times of day, days of the week, and times of year. The outputs of Streetlightdata.com web application include visualizations, Traffic Message Channel (TMC) shapefiles, and CSV files so that you can look at results in the app, as well as manipulate the data independently.

Airsage

AirSage[15] uses data from wireless carriers on locations of mobile devices across the US. This allows for tracking of the movement of people across space and time as a means to supplement transportation surveys for modeling purposes. The data are packaged and sold commercially to organizations, and transportation-planning related products

[14] Streelightdata.com can be accessed at https://www.streetlightdata.com/.

[15] AirSage http://www.airsage.com/Industries/Transportation/

include trip matrices and other studies related to commute patterns and analysis of movement at specific locations.

NATIONAL TRANSPORTATION ATLAS DATABASE 2015 (NTAD2015)

The NTAD, available at https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2015/index.html provides an extensive set of GIS files that can be used to supplement data analyses including transportation air quality analyses. The table on the next page provides a summary of the key point, line and polygon shapefiles available in this database.

File Type	Name and Size	Description
Point	Airports (ZIP - 2.0MB)	Data provided and maintained by the Federal Aviation Administration. Data compiled in 2015.
Point	Alternative Fuels (ZIP - 2.9MB)	This spatial dataset, from the Department of Energy (DOE), replaces a similar data created by BTS in 2008. This is a geographic point dataset of fueling facilities that offer fuels other than gasoline in the United States. Data compiled by DOE, 2014.
Point	Amtrak Stations (ZIP - 78KB)	Data provided and maintained by the Federal Railroad Administration. Data compiled in 2015.
Point	Border Crossing Ports (ZIP - 23.6KB)	Points of entry for land modes along the U.S.-Canadian and U.S.-Mexican borders. Data created and maintained by U.S. Customs and Border Protection and spatially enhanced by the Bureau of Transportation Statistics, 2014.
Point	Data of dams 50 feet or more in height (ZIP - 829KB)	Data of dams 50 feet or more in height, or with a normal storage capacity of 5,000 acre-feet or more, or with a maximum storage capacity of 25,000 acre-feet or more, from the U.S. Army Corps of Engineers (USACE) National Inventory of Dams. Data compiled by USACE, 2006.
Point	Intermodal Terminal Facilities (ZIP - 460.7KB)	Data on locations where freight can be transferred between modes of transportation. Data compiled by the Bureau of Transportation Statistics (last updated 2003).
Point	Intermodal Passenger Connectivity Database (ZIP - 856.4KB)	A national dataset of over 7,100 facilities for passenger public transportation facilities. This dataset links the intermodal nature of passenger travel in transportation, particularly in urban areas. Data provided by the Bureau of Transportation Statistics, compiled in 2015.
Point	Crash characteristics and environmental conditions (ZIP - 25.7MB)	This dataset (Fatality Analysis Reporting System) contains information on over 30,000 crash characteristics and environmental conditions at the time of the crash (crashes occurring in 2013). Data provided by the National Highway Traffic Safety Administration, compiled in 2014.
Point	National inventory of navigable inland waterway locks (ZIP - 43.8KB)	Database contains select attribute information for each of the 229 locks inventoried. Data provided by the U.S. Army Corps of Engineers, compiled in 2013.
Point	N ational Bridge Inventory (ZIP - 71.7MB)	A database containing information on the more than 600,000 bridges located on public roads, including Interstate highways, U.S. highways, state and county roads, as well as publicly accessible bridges on Federal lands. Data provided by the individual state DOTs to the Federal Highway Administration, 2014.

Point	National Populated Places (ZIP - 2.5MB)	Place locations from the 2010 Master Address File/ Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) -with additional attribute information from the Geographic Names Information System, 2014.
Point	U.S. Army Corps of Engineers Ports (ZIP - 4.4MB)	Data provided and maintained by the USACE, 2015.
Point	Top 150 major ports in the United States for import and updated export activity (ZIP - 20KB)	The U.S. Army Corps of Engineers identified top 150 major ports in the United States for import and export activity. Data compiled by USACE, 2015.
Point	Railroad Grade Crossings - (ZIP - 16.0MB)	Data on use and physical characteristics of highway-rail crossings. Data provided and maintained by the Federal Railroad Administration, 2015.
Point	Travel Monitoring Analysis System (ZIP - 471KB)	Stations from the Federal Highway Administration (FHWA), replaces the atr and wim data. Data set contains information on over 6,500 stations. Data provided and maintained by state DOTs, compiled by FHWA, 2015.
Polyline	Freight Analysis Framework, version 3.4 (ZIP - 49.4MB)	Integrates data from a variety of sources to estimate commodity flows and related freight transportation activity among states, regions, and major international gateways. The FAF is maintained by the Federal Highway Administration, 2012.
Polyline	Hazardous Material Routes (ZIP - 9.2MB)	Data on highway routes for hazardous materials. Data provided and maintained by the Federal Motor Carrier Safety Administration (last updated 2006).
Polyline	Highway Performance Monitoring System (ZIP - 795MB)	Data provided and maintained by state DOTs and submitted to the Federal Highway Administration, 2014. Attributes represent calendar year 2012. Because of the attribute file size, HPMS data is in ZIP format. Be aware that the unzipped HPMS data is more than 3 GB.
Polyline	National Highway Planning Network (ZIP - 172.2MB)	Spatial network of principal roadways within the United States. Data provided and maintained by the Federal Highway Administration, 2014.
Polyline	Railway Network (ZIP - 35.8MB)	Spatial network of rail lines within the United States. Data provided and maintained by the Federal Railroad Administration, 2015. (Note: A nodal shapefile called “railwaynd” and AMTRAK lines are also provided.)

Polyline	Airport Runways (ZIP - 558KB)	Spatial data on airport runways within the United States. Data provided and maintained by the Federal Aviation Administration, 2015.
Polyline	Fixed-Guideway Transit Facilities (ZIP - 863.7KB)	Transit database including fixed guideway links and stations (last updated 2004).
Polyline	U.S. Army Corps of Engineers Navigable Waterway Network (ZIP - 1.3MB)	Data provided and maintained by the USACE, 2014. (Note: A nodal shapefile called "waterwaynd" is also provided.)
Polygon	Bureau of the Census Urbanized Area Boundaries (ZIP - 54.4MB)	Data provided and maintained by the U.S. Census Bureau, 2014.
Polygon	Core based statistical area representing Metropolitan & Micropolitan Statistical Areas (ZIP - 53.6 MB)	Defined by the Federal Office of Management and Budget and maintained by the U.S. Census Bureau, 2014.
Polygon	The 114th Congressional Districts Boundaries (ZIP - 43.6MB)	Data provided and maintained by the U.S. Census Bureau, 2014.
Polygon	U.S. County Boundaries (ZIP - 11 1.4MB)	Data provided and maintained by the U.S. Census Bureau, 2014. Data derived by clipping the county_pol (political boundary) shapefile using the NTAD 2014 state shapefile, which contains shorelines.
Polygon	U.S. County Boundaries representing the U.S. political boundaries (ZIP - 83.7MB)	Data provided and maintained by the U.S. Census Bureau, 2014.
Polygon	Freight Analysis Framework (ZIP - 24.8MB)	Version 3.4, domestic region level datasets and products provide information for states, state portions of large metropolitan areas, and remainders of states. This data supports the FAF network. Data compiled by the Federal Highway Administration, 2010.
Polygon	Hydrographic Features (ZIP - 223.3MB)	Data provided and maintained by the U.S. Census Bureau, 2006. (Note: A linear shapefile called "hydrolin" is also provided.)

Polygon	U.S. Military Installations (ZIP - 6.5MB)	Data provided by the Military Surface Deployment and Distribution Command-Transportation Engineering Agency, 2013.
Polygon	Metropolitan Planning Organization (ZIP - 10.6MB)	Database of MPO planning area boundaries collected by the Bureau of Transportation Statistics and Federal Highway Administration, 2014.
Polygon	Non-Attainment Areas (ZIP - 42.8MB)	Environmental Protection Agency-defined areas of the country where "air pollution levels persistently exceed the new national ambient air quality standards," 2015. Data replaces, and enhances non-attainment areas posted by EPA last year.
Polygon	National Park System Boundary Dataset (ZIP - 21.0MB)	Data provided and maintained by the National Park Service, 2015.
Polygon	U.S. State Boundaries (ZIP - 38.7MB)	Data provided and maintained by the U.S. Census Bureau, 2014. Data derived by clipping the state_pol (political boundary) shapefile using the NTAD 2014 state shape file, which contains shorelines.
Polygon	U.S. State Boundaries representing the U.S. political boundaries (ZIP - 9.8MB)	Data provided and maintained by the U.S. Census Bureau, 2014.