

Chapter 47

Transportation Modeling



Eric J. Miller

Abstract Informatics are rapidly and radically transforming urban transportation in ways not seen since the introduction of the automobile over a hundred years ago. Near-ubiquitous smartphone usage, pervasive cellular and Wi-Fi connectivity, powerful and cost-effective computing capabilities, advanced GIS software and databases, advanced platforms for managing and scheduling service operations, etc., are combining to enable the introduction of new mobility services and technologies that are increasingly disrupting conventional trip-making behavior and the “rules of the game” in terms of transportation network operations and the regulation of system performance. The implications of these major informatics-driven changes for transportation modeling are equally disruptive and major. These include changes in: travel behavior; transportation system performance; the data available for model development and application; and modeling methods. Each of these broad areas of impact are discussed in this chapter.

47.1 Introduction

Use of large, computer-based models of travel demand and transportation system performance is standard practice in urban regions worldwide for transportation planning and decision-support purposes (Meyer and Miller 2013). They enable planners to estimate quantitatively the likely future impacts of a wide variety of policy options, including investment in major new transportation infrastructure (roads, transit, etc.), land-use policies, pricing/fare policies, new technologies, population and employment growth trends, etc. Detailed discussion of these models is well beyond the scope of this chapter, but the state of the art is extensively documented in the literature (see, for example, Ben-Akiva and Lerman 1985; Train 2009; Ortuzar and Willumsen 2011; Castiglione et al. 2015). Rather, this chapter explores current and emerging impacts of urban informatics on transportation modeling needs, capabilities, opportunities, and challenges.

E. J. Miller (✉)
University of Toronto, Toronto, Canada
e-mail: miller@ecf.utoronto.ca

© The Author(s) 2021
W. Shi et al. (eds.), *Urban Informatics*, The Urban Book Series,
https://doi.org/10.1007/978-981-15-8983-6_47

Informatics are rapidly and radically transforming urban transportation in ways not seen since the introduction of the automobile over a hundred years ago. Near-ubiquitous smartphone usage, pervasive cellular and Wi-Fi connectivity, powerful and cost-effective computing capabilities, advanced GIS software and databases, advanced platforms for managing and scheduling service operations, etc. are combining to enable the introduction of new mobility services and technologies that are increasingly disrupting conventional trip-making behavior and the “rules of the game” in terms of transportation network operations and the regulation of system performance.

The implications of these major informatics-driven changes for transportation modeling are equally disruptive and major. These include:

- Changes in travel behavior.
- Changes in transportation system performance.
- Changes in the data available for model development and application.
- Changes in modeling methods.

Each of these topics are discussed in detail in the following four sections. Looming over this discussion of technology-driven changes in the transportation system and associated modeling needs is the potential for the introduction into widespread usage within a currently ill-defined but still foreseeable future of electric vehicles (EVs) and connected and autonomous vehicles (CAVs), which may also be electrified (CAVEs). Full discussion of these technologies and their potential impacts goes well beyond the topic of urban informatics per se. But some possible impacts of eventual CAV impacts on travel behavior and transportation network performance are briefly discussed in Sects. [47.2](#) and [47.3](#).

47.2 Informatics and Travel Behavior

The primary impacts of informatics on travel behavior to date derive from two related informatics-based services:

- Real-time travel-related information.
- New mobility services and technologies.

These are discussed in the following two sub-sections. As becomes clear in this discussion, the driving technology enabling all these services are cellular- and Web-based apps running on smartphones and other computing devices, tied to centralized computing platforms that receive and send massive amounts of data and that process customer data requests for information and services, match customers with service providers, etc. The evolution and widespread adoption of smartphones among a broad segment of trip-makers, in particular, has been fundamental to the development and implementation of these various services.

47.2.1 Real-Time Travel-Related Information

A veritable plethora of Web- and smartphone-based apps exist that trip-makers can use to plan their trip destination, mode, and route choices prior to traveling and to dynamically choose their travel route during their trip. Many of these apps are provided by private companies, but public-sector apps also exist. For example, most public transit agencies provide some form of route guidance, as well as schedule and fare information.

Perhaps the most pervasive and impactful of these apps are the wide range of route-guidance apps based on the Global Positioning System (GPS) and available either on-board many automobiles or as apps for smartphones or other mobile devices such as tablets. These sense the current location of the device (and, hence, vehicle) and provide real-time estimates of current traffic conditions on the roadway being used. They also provide estimates of current travel times to a user-specified destination, along with recommended best routes to take to this destination. The definition of best route may be based either on shortest distance or shortest expected travel time, with the latter being the preferred and, increasingly, the most common option. Link and route travel times are determined based on crowd-sourced information on speeds gathered from all the users of the service, as well as possibly other information that may be available to the service provider (police/traffic center advisories, other roadway sensor data, etc.). They also depend critically on access to very precise and accurate geographic information system (GIS) representations of the road network, including speed limits and other road attributes. Huge effort over the past several decades has gone into developing such detailed maps for much of the world, particularly, in urbanized areas. Thus, these route-guidance apps represent an advanced marriage of GPS tracking and GIS mapping and analysis capabilities.

Both real-time and historical data are used in the calculations. The quality of the travel-time and route-selection calculations obviously depends on the number of users in the system at any one time, the depth and relevance of the available historical information, and, critically, the quality and accuracy of the (typically proprietary) algorithms used by the service provider to do these calculations. Machine learning methods (running on powerful cluster/cloud computing platforms) play a key role in sifting through the massive real-time and historical data to identify traffic patterns and to make short-term predictions of best routes to recommend. While these algorithms still are not 100% perfect under all conditions and in all places, their accuracy in making short-run predictions of roadway performance is typically quite impressive.

In addition to on-board route-guidance apps, conventional variable message signs on roadways and radio traffic reports have for decades provided a certain amount of high-level, real-time information concerning current travel conditions on major roadways, although these rarely provide route guidance. That is, a variable message sign might indicate that the roadway is congested ahead, but will not actually suggest or advise to take an alternative route. This is both due to legal concerns (if a driver takes a suggested alternative route and gets into an accident, who is liable?) and to

minimize the potential for introducing instability into the system (what if everyone took the alternative route?).

Many apps also exist for providing static or real-time information concerning public transit routes, schedules, fares, and travel times. Most transit agencies now provide such an app, but many private and open-source public apps also exist. Such apps may provide information concerning: when the next transit vehicle is expected to arrive at a given stop; assistance for planning a trip from a given origin to a given destination at a given time of day; fare policies and payment options; service disruptions notices, etc. In addition to mobile-device-based apps, many transit agencies also provide real-time information at transit stops and stations concerning expected next-vehicle arrival times, by transit line. Various apps also exist to help bicyclists track their bike usage and routes are taken. Personal fitness apps for tracking distance walked also exist.

Although not generally thought of as being particularly travel-related, a vast array of Web sites provides information concerning every form of activity imaginable—restaurants, stores, entertainment venues, hotels, etc. These activity locations are potential destinations for trip-making that is not related to work or school, and the ubiquitous and voluminous availability of such data may well influence trip-makers' decision-making, especially regarding trip destination.

In general, most of these apps and services can be used for pre-trip planning (“Where should I go for dinner tonight?” “Should I drive or take transit for this trip?”) as well as for on-route dynamic decision-making (“Accident ahead; let’s get off the freeway”). While usage of these various apps is clearly very widespread, the actual impacts of this usage on travel behavior are not at all well understood. What percentage of the population are using what kinds of apps? Does this usage significantly influence choice of mode or destination, or timing of trips? Route-guidance apps must be affecting route choices, given their widespread use, but how great are the resulting deviations from the routes that drivers would have chosen in the absence of the app? To what extent is congestion being reduced (or increased?) through extensive use of these apps? These issues are discussed in greater detail below.

47.2.2 New Mobility Services and Technologies

Current and emerging information and communications technology (ICT) is not only dramatically increasing and improving the information available to trip-makers to help them in their travel decision-making, it is also revolutionizing the services available to them by which they may travel. New ICT-based mobility services and technologies are emerging virtually daily that provide new travel options for trip-makers. As with the new information services, these critically depend on smart mobile devices for communicating with potential customers of the service and on powerful computing platforms to manage the service.

As discussed in detail by Calderón and Miller (2019, 2020), a *mobility service* can be defined as an operation that enables a person to complete a trip from an origin to a destination by means of a given mode (technology) and service process. Public transit and conventional taxis are traditional mobility services. But a wide range of informatics-enabled mobility services has emerged in recent years. These take many forms, including:

- *Ridehailing*: Services such as Uber and Lyft (also conventional taxi), in which a service provider connects drivers with passengers to provide passengers with a door-to-door trip from their origin to their destination. Ridehailing can be further sub-divided into *single-user* and *shared-ride* services, with the latter involving passengers sharing the vehicle with other passengers and, as a result, experiencing some amount of trip deviation from a direct origin-to-destination trip in order to accommodate the pickups and drop-offs of the other passengers sharing the vehicle.
- *Vehicle-sharing*: These services provide short-term rentals of vehicles to customers who pick up the vehicle from where it is parked, use it to execute one or more trips, and then leave the vehicle safely parked once they are finished with it. Different services use different types of vehicles, including: automobiles (car-share), bicycles (bike-share, using both conventional bicycles and e-bikes), and, most recently, e-scooters. Vehicles usually are parked at designated stations (parking lots, bike-share docking stations, etc.), but dockless systems increasingly exist, in which the car, bike, e-scooter, etc., can be left anywhere, and is picked up by the next customer from wherever it was last left. Such dockless systems obviously depend on GPS tracking of the vehicle so that its location is known at all times. Vehicle-sharing services are usually provided by a for-profit company, but examples of peer-to-peer systems also exist in which private individuals offer their vehicle for usage by others when they do not need it for their personal use.¹
- *Demand-responsive transit (DRT)/microtransit*: A wide variety of transit services exist (or can be imagined) that deviate from conventional fixed-route, fixed-schedule (typically large-vehicle) transit operations, including various combinations of route deviation, flexible stop location, on-demand scheduling of vehicle routing, and, usually, use of smaller vehicles that are cost-effectively matched to travel demand levels. Various forms of DRT have operated basically as long as public transit has existed. In particular, in much of the world jitney operations (along with other forms of privately operated informal transit services) are critical components of urban transportation, especially for lower-income trip-makers. In addition, DRT (often referred to as paratransit) services are a standard means of providing on-demand transit to mobility-impaired trip-makers who are unable to use conventional transit services. Platform-based informatics systems are redefining and enhancing the capabilities and potential applications of such

¹Examples of peer-to-peer shared-ride systems also exist in which a platform connects private individuals who are willing to share rides with other individuals. A common example of such a system occurs on many university campuses, in which students offer rides to other students to travel back and forth between the university and nearby home cities during holiday weekends, etc.

services by significantly improving both the quality of service that can be offered to customers (through improved real-time scheduling and more efficient routing) and the cost-effectiveness with which the service can be provided.

While a wide diversity of mobility services exists, they all involve some combination of a generic set of operating functions (Calderón and Miller 2019, 2020). These consist of:

- *Matching* trip-maker requests for service with drivers and vehicles.
- *Rebalancing* vehicle fleets to maintain an appropriate spatial distribution of vehicles available for service.
- *Trip pricing* and payment.
- *Pooling* customers within vehicle tours for shared-ride operations.

Clearly not all operations pertain to all services. Bike-share services, for example, only provide real-time information concerning the current availability of bicycles by location, leaving customers to find their way to and rent one of these available bicycles. They do, however, have to deal with rebalancing, since usage patterns often result in large numbers of bicycles at popular destinations and too few bicycles at some origin locations. Ridehailing operators, on the other hand, primarily are concerned with matching customers to vehicles so as to both maximize the customer experience (usually meaning minimizing service wait times) and minimizing operating costs (e.g. avoiding very long dead-heading of vehicles). They may or may not engage in active attempts to rebalance the locations of the vehicles currently in service.² Pooling, of course, only pertains to shared-ride operations, but is a very critical component of the service, since the classical weakness of shared-ride services has been poor customer experiences: long wait times and circuitous routing (and hence long travel times relative to a more direct origin–destination journey).

Pricing levels and policies vary from one service to another and vary to the extent that prices dynamically vary with demand levels (so-called surge pricing) and, possibly, other factors (such as weather). Online payment systems based on credit cards are, however, an important feature of all new mobility systems. The convenience of this automated payment system should not be underestimated. At the end of the day, differences between a conventional taxi and an Uber are arguably not that great,³ but the convenience of being able to simply step out of the car at the end of the trip (as well as the convenience of booking the trip with a few key-strokes on a smartphone) appears to be a significant factor in the success of new mobility services.

The role of informatics-based platforms, involving an integrated of GPS, GIS, real-time cell- and Web-based communications, combined with high-capacity computing and data processing and analytics based on artificial intelligence (AI) is fundamental

²Since ridehailing services currently depend on independent driver contractors, the ability of the ridehailing platform provider to influence their locations when not in service tends to be indirect at best.

³Although differences clearly exist, particularly, perceptual differences. Taxis, for example, are often criticized as being “dirty”. Safety/security differences also exist, as do price differentials.

to all such mobility services. It is such platforms that have allowed both conventional taxi and transit services to be re-invented and for new technologies and services such as bike- and e-scooter share services to emerge.

The concept of *mobility as a service* (MaaS) generalizes mobility services by extending the platform concept to integrate two or more mobility services to provide seamless, and door-to-door mobility solutions that dynamically mix and match mobility services customer by customer to optimize their travel experience within a one-stop-shopping process. MaaS is seen by many as the future of transportation, with MaaS platforms acting as brokers that piece together different mobility services to best meet the trip-maker's needs and preferences. In such a future, a trip-maker may be picked up at her door in a suburb by a ridehailing company, taken to a commuter rail station just in time to board her train, and then have an e-bike waiting for her at her downtown egress station to complete her journey to her office, all for one fare automatically charged to her credit or debit card (perhaps with various loyalty points as well).

Such complete mobility solutions do not generally currently exist, although many companies and organizations are working toward their implementation. A particularly important policy question exists concerning the extent to which MaaS solutions can be integrated to improve the cost-effectiveness and attractiveness of public transit, so as to maintain it as a primary mass mover of trip-makers in high-density corridors. Urban areas worldwide are currently overwhelmed by auto congestion, and it is essential, however MaaS plays out, that it enables more efficient usage of transportation networks through the promotion of transit (where appropriate) and congestion reductions, while still accommodating the growth in travel that is inevitable as urban regions continue to grow. Notably, there is a growing literature that indicates that current mobility services are both adversely impacting conventional transit usage and increasing the amount of congestion (at least in central areas) in many cities (Li et al. 2019; Graehler et al. 2019; Rayle et al. 2016).

While an academic literature exists that explores the potential impact of route-guidance information on travel behavior, most of this is based on stated preference surveys or hypothetical simulation experiments rather than real-world data. A major barrier to investigating these questions is that the vast bulk of data concerning app usage and subsequent behavior is proprietarily held by private companies who are usually unwilling to share it with public agencies or academic researchers.

Enormous speculation currently exists concerning the potential impacts on travel behavior of the ubiquitous availability of fully autonomous vehicles. Exploration of this issue is well beyond the scope of this chapter. We simply note that CAVs potentially might dramatically alter auto ownership levels (people may simply rent mobility on a per-trip basis), public transit usage, and roadway congestion levels, among many possible other impacts. Transit ridership impacts are a particularly important policy question. CAVs might be used to support the use of higher-order transit by providing first- and last-mile solutions for getting to and from transit in low-density suburban neighborhoods. Or ubiquitous automated ridesharing services might decimate transit usage, likely leading to increased, rather than decreased, congestion on urban streets. In any event, increasing connectivity and automation of

the transportation system will further increase the availability of massive, dynamic real-time information concerning travel and the associated need for advanced informatics methods for the storage and analysis of these data for transportation planning and operations purposes.

47.3 Informatics and Transportation Network Performance

Transportation network performance is the emergent outcome of a short-run (day-to-day, hour-by-hour, minute-by-minute) demand–supply interaction, in which the performance of a network link (road or transit line segment) depends on the volume of flow (cars, passengers, etc.) using the link at a given time. That is, the travel time required to traverse the link (and associated congestion level) depends on the level of link usage, while the number of users of the link depends (at least in part) on the travel time experienced on the link.

Route-guidance apps surely have an impact on the route choices of individual trip-makers (otherwise, why would they use them?), and, hence the distribution of flows across links and paths within the network, and ultimately on link and path travel times. Such apps are used both for pre-trip planning (What’s the best way of getting there? What’s a good time to leave to avoid traffic?) and dynamic on-route guidance. The actual impacts of such route-guidance apps on trip-makers’ route choices, however, are typically unknown, since only the app companies usually see the data and they are generally not telling.

Note that a major impact of CAVs is likely to be to take route choice decisions largely out of the hands of the trip-maker and place them under control of the vehicle and its associated automated route-guidance system. This should help improve roadway performance since vehicles will be more likely to be spread across network paths so as to minimize overall congestion. But this may also involve an ethical issues of whether it is appropriate to impose a longer trip on one user so that other users may benefit from shorter travel times (which is usually what is required in order to reduce overall delay in the system).

Informatics-based connectivity (whether in an automated or conventional vehicle) offers the potential for ubiquitous road pricing, in that if every vehicle’s location is known and local roadway congestion levels are also known at each point in the network, then usage of the road system can be dynamically priced to encourage more system-optimal route choices by trip-makers, or, at least, to charge trip-makers the actual social cost of their trip. Such a system addresses the ethical issue raised above by creating the potential of offering multiple route choices to trip-makers: for example, a quicker but more expensive route (since it involves higher social marginal costs associated with the trip) or a slower but less expensive one (in which socially beneficial behavior is encouraged or rewarded by a discounted travel cost).

Parking could be similarly monitored and dynamically charged to reduce on-street parking on congested streets, direct cars to vacant parking spaces, etc. Parking lots and garages take up an enormous amount of valuable space, on-street parking very

significantly reduces the capacity of our streets to carry traffic of all sorts (i.e. bicycles, transit, etc. in addition to cars and trucks), and drivers cruising to find (cheap) parking is a major source of congestion in its own right in most urban centers. Even with conventional cars, informatics-based parking apps and usage monitoring systems in parking lots can reduce these impacts considerably, as is being demonstrated, for example, by the SF Park demand-responsive parking pricing experiment in San Francisco (<https://sfpark.org/>). A major asserted benefit of CAVs is that they may eliminate most on-street parking as well as significantly reduce parking lot needs, especially in urban cores. As with all aspects of CAVs, these benefits are at the moment speculative, but are the subject of considerable research (Nourinejad et al. 2018).

Informatics is also extensively (and increasingly) used in transportation network operational control. Traditionally, roadway performance (volumes, speeds, congestion levels) has been monitored by electromagnetic loop detectors embedded in roadways that detect vehicles passing over the detector by the magnetic signature of the vehicle. While useful, such loop-detector systems are expensive to install and maintain and are often subject to failure. Numerous other technologies now exist for monitoring roadway traffic, including video cameras (which require advanced image-processing methods for automated data gathering from the video images), Bluetooth detectors (which detect the unique MAC addresses of vehicles, smartphones, and other Bluetooth-enabled devices, thereby being able to trace the paths and average speeds of these vehicles as they pass a sequence of detectors within the network), and purchasing of on-board route-guidance and other passive location-detection app data from third-party providers. In the case of public transit, many agencies have automatic vehicle location (AVL) systems for tracking transit vehicles in real time and automatic passenger counting (APC) systems for measuring real-time passenger boardings and alightings per vehicle at each stop along a given transit route.

47.4 Informatics and Data Support for Travel-Demand Modeling

The informatics-based services and apps discussed in Sect. 47.2 are generating tremendous amounts of data, day after day, concerning millions of trips being made within a given metropolitan region.

Travel-demand modeling has always depended heavily on large cross-sectional surveys of trip-makers within an urban region. Such surveys are expensive and time-consuming to undertake, subject to various sampling and other biases, and often facing increasing challenges in terms of being able to generate representative samples (Miller et al. 2012; Srikuenthiran et al. 2018). While traditional large household travel surveys are likely to continue be undertaken for the foreseeable future (Miller et al. 2018), current and emerging informatics methods offer promising alternatives and complements to traditional surveys in terms of both new modes and technologies

for conducting surveys and new passive (non-survey) methods for observing travel-related behavior, which are discussed in the following two sub-sections. Common to all these sources of data is the problem of imputing missing attributes of the trip or the trip-maker, which requires advanced statistical data fusion and modeling methods, which are briefly discussed in the third sub-section.

47.4.1 Informatics-Based Survey Methods

The primary two informatics-based survey methods are Web-based surveys and smartphone-app-based surveys and trackers. Web-based surveys have become a de facto standard method for undertaking travel surveys, replacing or complementing more traditional methods such as telephone interviews, self-completed mail-back surveys, and face-to-face interviews.⁴ Web-based surveys can be very cost-effective since they eliminate the need to hire interviewers, and the marginal cost per survey completion is very low once the up-front cost of the survey development and implementation is accounted for. On the other hand, establishing and contacting a representative sample can be challenging, response rates can be low, and the quality of responses can also be sometimes problematic given the lack of supervision and assistance provided by an interviewer. This last problem, however, can be significantly mitigated by very careful software design to maximize the clarity of the questions being asked and to minimize respondent burden (Loa et al. 2015; Chung et al. 2020; Srikukenthiran et al. 2018).

Similarly, many custom smartphone apps exist that have been explicitly designed to track persons' trip-making and to gather information concerning trip and trip-maker attributes. These generally involve a brief up-front survey to gather key demographic and socio-economic information concerning the trip-maker (and, ideally, the trip-maker's household). The app then is designed to actively track all movements by the person over multiple days, or even possibly weeks, using the smartphone's on-board GPS and other tracking capabilities. This generates space-time traces of the person's movements while carrying the smartphone (assuming that it's turned on!). The potential to gather detailed information concerning personal travel behavior is considerable. In particular, route choice and information concerning active modes, both of which are typically challenging to gather with conventional survey methods, are readily gathered by such apps (Grond and Miller 2016; Lue and Miller 2019). Numerous technical issues, however, are not fully resolved, thus limiting their current widespread usage. These include issues of phone battery life versus the precision of the route tracking (the more precise the tracking, the greater the drain on the battery); the ability to impute travel mode and trip purpose purely from the trip trace; and the representativeness of the smartphone-based samples and sample recruitment methods (Rashed et al. 2015a; b).

⁴Even for these traditional survey modes, tablet-based Web software is being used to conduct and record the interviews. See, for example, Chung et al. (2020) and Harding et al. (2017).

Considerable processing of the raw traces also needs to be undertaken in order to identify the end (stop) point of a trip in space and time (e.g. has the person stopped for a quick shopping activity in a store or is she or he just waiting a long time at a bus stop?), the purpose of the trip (i.e. the type of activity engaged in at the trip end), and the mode of travel used to undertake the trip. Location, purpose, and mode are all essential trip attributes if these data are to be useful for travel-behavior analysis and modeling. Ideally, these attributes should be imputable from the trace data themselves, combined with additional available data, notably GIS datasets concerning land use and points of interest (POI—schools, stores, etc.) and transportation network data concerning road and transit networks. That is, the respondents are passively tracked, without having to explicitly query them concerning their trip-making. If sufficient multiple-day data for enough trip-makers are available, then machine learning methods can, in principle, be used to impute trip stop, mode, and purpose. The current state of practice, however, is such that it is generally required to actively gather at least some information concerning the trips being made, either on the fly as the trips are being detected or at the end of a day through retrospective questioning of the respondents. This active questioning allows labels to be attached to the detected trips (this trip was by car to go shopping) that greatly enhances the ability to train the automated attribute imputation models, at the price of imposing an on-going response burden on the survey participants. Thus, active questioning is often undertaken for a few days at the beginning of the survey period and then turned off with the tracking app running totally passively for the remainder of the survey under the assumption that the imputation apps can be sufficiently trained with the sample of active data obtained (Faghih Imani et al. 2020; Harding et al. 2020; Harding et al. 2016a, b).

47.4.2 *Passive Trip Tracking*

Numerous informatics-based methods exist to gather information concerning trip-making behavior. These include (Miller et al. 2012):

- Passive smartphone-based location trackers.
- Cellphone traces.
- Transit smartcard transaction data.
- Bluetooth sensors.
- Credit card transaction data.

Passive Location Trackers: As discussed in Sect. 47.2.1, vast quantities of information concerning trip-making are being collected by route-guidance apps, as well as other apps that track smartphone locations for a variety of purposes. In addition to facilitating route guidance, the data collected by such apps can be used to identify origin-destination trips by time of day. These data can be distinguished from the smartphone-app data discussed in the previous section in that they do not require

involvement of the phone user in any way and they are completely anonymized (and generally aggregated in one way or another).

Cellphone Trace Data: Whenever turned on, all cellphones are in constant communication with their cellular network. Movements of cellphones (and, hence, their owners) can thus be tracked through time and space. These cellphone traces require significant processing in order to be useful for the analysis of travel behavior, but many analysts are working with such processed data to develop datasets on origin-destination trips by time of day in many urban regions (see Faghieh Imani and Miller (2018) for a comprehensive review). The primary attraction for cellphone trace data is its ubiquity in providing massive amounts of travel data, day after day, in virtually every urban region worldwide. Also, given the very deep penetration of cellphones in today's society, these traces can likely be treated as being reasonable representative of the trip-making public. The major limitation of these, data, however, is that the spatial-temporal resolution of the traces is inherently limited by the spacing of the cell towers receiving the cellphone transmissions. Achievable resolutions vary considerably within an urban region. The relatively gross resolution generally achieved poses significant challenges with respect to imputing trip mode (which generally requires good speed measurements) and trip destination activity type (Caceres et al. 2013; Faghieh Imani et al. 2018).

An interesting special use of cellphone tracking data is to identify intercity trips. When a cellphone is detected in a city other than its home city, one can impute that an intercity trip has occurred. Intercity travel is a particularly difficult travel market to survey effectively, and so use of cellphone data for this purpose is a promising avenue of research (Bekhor et al. 2013; Janzen et al. 2017).

Transit Smartcard Transaction Data: Another major informatics-enabled source of travel data are data from smartcard transactions collected by public transit agencies. Most major cities worldwide employ some form of smartcard for riders to use to pay their fares, with these cards becoming almost universal in usage. These data thus provide a near-complete record of transit usage in a city. These smartcard systems vary in technical sophistication, but they generally involve one of two primary designs: tap-on systems, in which transit riders tap into the system when they first board a transit vehicle or enter a transit station; and tap-on-and-off systems, in which riders must also again tap the card when they exit the system. These latter systems obviously provide a complete record of all trips made from a first-boarding stop or station to a last-alighting stop or station, by time of day. Tap-on systems require extensive processing to impute trip-alighting locations (typically by observing the boarding location of the next transit trip), but still provide very usable information concerning transit usage (Trépanier et al. 2007; Munizaga and Palma 2012; Parada and Miller 2017).

Bluetooth Sensor Data: As noted in the previous section, Bluetooth detectors can be used to track the passage of Bluetooth-enabled vehicles and personal devices as they pass by detectors mounted along the side of a road. Using records from multiple antennas makes it possible to derive travel times between antenna locations. Hence, depending on the setting, data could be used to derive O-D matrix and partial route choice of a sample of vehicles (cordon setting). While the available data have mostly

been used to provide information on vehicle movements, it is also becoming possible to study pedestrian behavior. Malinovskiy et al. (2012) investigated the feasibility of using Bluetooth for pedestrian studies using two separate sites. Their results suggest that “given sufficient populations, high-level trend analysis can provide insights into pedestrian travel behavior.”

Credit Card Transaction Data: Although not currently widely used due to lack of access to the data, credit card transaction records can provide detailed information concerning travel for a wide variety of purposes (basically any activity that involves paying with a credit card at an out-of-home location for a good or a service). It also provides expenditure data along with the activity/travel data, something which is not generally gathered in conventional surveys, but could be very useful in modeling not just time but monetary budget allocations. Further, it could provide information concerning in-home versus out-of-home shopping/recreation expenditures, again, something that is of considerable interest for understanding travel behavior. The major limitations of this data source, of course, are whether access to such data can be obtained, and the protection of the confidentiality of the data.

While each of these passive data types have their individual strengths and weaknesses, they share common strengths in terms of:

- Providing a continuous stream of data over days, weeks, and even longer periods of time, thereby permitting time-series analysis of travel trends and dynamics (as opposed to the typically one-day cross-sectional snapshots obtained through conventional surveys).
- Generating massive amounts of data, potentially for thousands or even millions of trip-makers in a large urban region (as opposed to the small samples that can typically be observed in conventional surveys); they truly are big data.
- Being total passive—they require no effort (or perhaps even awareness) on the part of the trip-maker for the data to be collected.

They also, however, share common, significant challenges in their usage in travel-behavior analysis and modeling:

- The data are inevitably anonymized to preserve confidentiality, and, thus, no personal attributes of the trip-makers are known.
- The data are individual-based, not household-based. That is, we generally know nothing about the other members of the trip-maker’s household. Household interactions and constraints, however, generally significantly affect an individual’s travel behavior.
- As with passive smartphone-app survey data, trip attributes beyond origin, destination, and trip start and end times are generally unknown. That is, trip mode⁵ and purpose need to be imputed.
- The spatial-temporal precision of the trace data can vary considerably from one type of data source to another, and even from one trip to another within a given data

⁵Except, of course, in the case of transit smartcard data, where the travel model obviously is transit.

type. Cellphone traces are particularly problematic in this regard, often making mode and purpose imputation challenging.

47.4.3 Data Fusion and Imputation

As discussed above, there are many sources of information concerning travel behavior, ranging from traditional surveys to various informatics-based passive data streams. Virtually all such datasets are incomplete in one way or another in terms of missing one or more attributes of the trip-maker or the trip that are desirable for travel analysis and modeling purposes. This may range from trip-makers' incomes not being collected in a household travel survey to a complete lack of information concerning trip-maker characteristics in most passive datasets. Passive location-tracking data also often lack explicit information concerning key trip attributes such as travel mode and trip purpose. In all such cases, it is desirable to impute the missing information through the fusing of two or more datasets to create a new, combined dataset that contains a richer set of attributes than either original dataset. A common, relatively simple example of this is using census data to impute missing income information in a household travel survey. This is done by using the correlation between income and other household attributes observed in the census data to impute the missing incomes for households observed in the survey, based on the household attributes that are observed in both the census and survey datasets (Bonnell et al. 2009).

A wide typology of data fusion and imputation use cases exist, with many methods available for addressing these cases. Detailed discussion of these use cases and methods is well beyond the scope of this chapter, but can be found in a range of sources, including the work of Miller et al. (2012) and Srikukenthiran et al. (2018). Only two observations are included here. The first is that a particularly important type of data needed for many data fusion exercises that have not yet been mentioned herein are data based on GIS concerning the spatial distributions of people (and their attributes), jobs, and other economic and social activities (stores, schools, etc.). These may be stored at various levels of spatial aggregation (traffic zones, census tracts, etc.), but are also often available in increasingly accurate and comprehensive POI datasets from a variety of commercial and open-source providers. POI data provide information concerning land uses at the very fine level of detail of the individual building, parcel, or geocoded point in space. They thus enable highly disaggregated analysis of point-to-point travel behavior, which is increasingly the level of detail at which travel-demand models are being developed.

Second, as in virtually every sphere of data analysis today, machine learning methods are being increasingly applied to a wide variety of transportation data fusion problems (Gao et al. 2017). One such example involves the use of transit smartcard transaction data, combined with conventional household survey travel data, to train a deep neural network model to predict travel mode. This model is then applied to cellphone trace data to impute the travel mode for the trips represented by these traces (Vaughan et al. 2020).

47.5 Informatics and Modeling Methods

As noted at the beginning of the chapter, a thorough discussion of travel-demand modeling methods is well beyond the chapter's scope. A few characteristics of the current state of best practice, however, include those by Miller (2018, 2019):

- Essentially, all best-practice models are based on activities and tours, in which: (a) travel is the emergent outcome of the need to participate in out-of-home activities; and (b) individual trips are modeled within the context of the overall tours or trip-chains that people engage in throughout their daily activity pattern, so that within-tour decision-making interactions can be accounted for (e.g. if a car leaves the driveway it must eventually return home).
- Travel behavior is largely modeled using sophisticated discrete-choice models based on random utility theory, which provides a very strong behavioral foundation for operational models.
- Increasingly, these activity- and tour-based models are implemented within an agent-based microsimulation modeling framework (see Chap. 44).
- The development of such models has been based on sophisticated, but classic, econometric parameter-estimation techniques (typically maximizing log-likelihood functions).
- Even very complex model systems for large urban regions are developed based on relatively small, cross-sectional samples of a region's trip-making population.

Modern informatics is providing both challenges to the current modeling status quo and opportunities for the development of next-generation models. As noted in Sects. 47.1 and 47.2, informatics-based apps are providing enhanced information and influencing travel choices in ways that are not completely understood and that definitely are not being captured in currently operational models. However, it might also be noted that current models typically assume implicitly that trip-makers have perfect information concerning their travel options and attributes. Hence, it might be argued that these new information sources are actually bringing behavior more in line with modeling assumptions since trip-makers now do have much better information to use in their decision-making!

While the future is perhaps more uncertain than ever before, a few important, specific, and informatics-related observations concerning the current and emerging state of the art in travel-demand modeling can be made with reasonable confidence and are provided below.

First, current best-practice models definitely are not well suited for analyzing new mobility systems, let alone CAVs (Miller 2019). These models need to be redesigned and rebuilt to much better represent both demand decisions and the performance and supply characteristics of these new services (Calderón and Miller 2019, 2020). As data concerning the performance and usage of a wide variety of mobility services become available, the potential for developing improved models increases. New informatics-enabled survey methods also provide the opportunity to gather data on trip-maker preferences and attitudes that will assist in this endeavor.

Second, the increasing availability of massive and passive big data is going to profoundly change how we model travel behavior. While significant technical issues remain, they will provide the opportunity to:

- Develop dynamic models of travel-behavior evolution, freeing us from the tyranny of infrequent, cross-section survey datasets as a basis for model building.
- Establish much more comprehensive and complete representations of travel in an urban region, freeing us from dependency on small-sample surveys which, despite their richness in socio-economic information, inevitably contain significant sampling and response biases.

Third, machine learning and other AI-based methods are rapidly being applied to travel-demand modeling (Yin et al. 2016). While such methods often produce better fits to base data than conventional econometric methods, whether they actually represent improved models for policy analysis and forecasting is very much an open question. A very interesting panel session was held at the US Transportation Research Board Annual Meeting in 2017 titled “Machine Learning Is from Venus, Econometric Modeling Is from Mars: Two Different Travel Forecasting Perspectives.” The very strong consensus coming out of this session was that the two modeling approaches are primarily complementary, and that travel-demand modeling needs to optimize its exploitation of both modeling disciplines if it is to meet the profession’s modeling needs. In particular, the notion that the advent of big data and AI-based analysis methods will mean the death of (travel demand) models does not appear to be either a likely or attractive alternative. Longer-term, strategic forecasting requires models that can generate emergent, out-of-sample, extrapolated behavioral responses to new scenarios, policies, etc. They cannot just extrapolate current patterns. Further, the interpretability of model sensitivities, elasticities, etc., is a critical component of travel-demand modeling, something that machine learning methods are notoriously poor.

More speculatively, two final questions concerning how informatics-based data and methods might fundamentally change travel-demand modeling in the coming years are the following.

First, can the relatively rich theory of travel behavior that the field has developed over the past sixty years, combined with advanced simulation, data fusion, and machine learning methods be used to both bridge the socio-economic information gaps typical in big data and to merge complementary data sets together to create much more comprehensive representations of travel behavior? Vaughan et al. (2019) provide one example of this approach, in which cellphone traces, transit smartcard transactions, and conventional home-interview travel survey datasets are merged to create a more comprehensive representation of base-year travel than it is possible to achieve from any of the three datasets independently.

Second, is there a quantum theory of travel behavior out there? That is, is there a more explicitly statistical (as opposed to behavioral) approach to modeling that is better suited to the strengths (and weaknesses) of the new datasets? But such a theory or model would still need to be predictive to answer what-if questions. In physics, prediction is the ultimate proof of a theory: Einstein’s theories of special

and general relativity were accepted, not because of their elegance, but because they are capable of predicting actual behavior. And, indeed, quantum theory's acceptance rests on its ability to predict real-world phenomena (and despite the objections of Einstein on philosophical grounds). The great question facing travel behavior theorists and modelers going forward is how urban informatics-based data and methods will enable us to obtain deeper understanding of actual travel behavior, and, building on this understanding, to develop more powerful and compelling theories and models of travel behavior that enable us to better predict travel behavior in support of transportation policy analysis and forecasting.

47.6 Chapter Summary

This chapter has examined the many ways in which informatics has been changing transportation modeling. These include disruptive changes to: travel behavior, transportation system performance, the data available for model development and application, and modeling methods themselves.

Travel behavior is being influenced primarily by two types of informatics-based services. The first is travel-related Web- and smartphone-based apps that provide a wide range of real-time information, including roadway route guidance, transit service information, and information concerning alternative activity locations. This information is used in both trip preplanning and on-route dynamic decision-making. The second disruptor of travel behavior is the wide variety of new informatics-enabled mobility services that provide trip-making alternatives to conventional travel modes such as public transit, taxis, and even the privately owned car. Most notable are the Uber and Lyft ridehailing services. Other mobility service types include ridesharing (UberPool), car-sharing, bike-sharing, e-scooters, and various forms of demand-responsive transit and microtransit. The mobility service field is evolving rapidly, and the final steady state with respect to these services and their impacts on travel behavior is very difficult to predict. It is clear, however, that travel-demand models will need to evolve considerably if they are to be adequate tools for modeling these impacts and to provide the level of policy guidance needed to ensure socially beneficial outcomes with respect to these services.

These changes in travel behavior and mobility service options are also impacting transportation network performance, notably in terms of roadway congestion and transit usage. Informatics also can support improved real-time control of road and transit operations, implementation of road pricing schemes, and managing parking supply and pricing.

Informatics technologies are also dramatically changing the data available to support travel-demand modeling. Web-based and custom app-based survey methods are complementing, and increasingly replacing, conventional survey methods for collecting travel-behavior information. In addition, a wide variety of sources for passively tracking trips are available, where by passive is meant that the trip-maker is

not required to interact with the tracking device or answer any questions. Passive trip-tracking data sources include: smartphone-based location-tracking apps (the route-guidance apps discussed above, but many other apps routinely track the phone's location); cellphone traces; transit smartcard transaction data; Bluetooth sensors; and credit card transaction data. All these data sources offer massive amounts of information, gathered continuously over time concerning trip-making in a given region. They also share common issues concerning lack of socio-economic information about the trip-makers, as well as lack of key trip attributes such as travel mode and trip purpose. A variety of data fusion and imputation methods (including machine learning methods), however, can often be used to augment the passive data, thereby enhancing their utility for modeling.

Given the increasing availability of large, passive datasets, travel-demand modeling will inevitably evolve to exploit these data. Continuous time-series streams of data should support the development of more dynamic (adaptive) models. The very large samples of trip-makers observable within these datasets should lead to models that are more representative and comprehensive relative to current models, which have relied on relatively small-sample survey data for their development. Machine learning and other AI-based methods will continue to play an increased role in model development and application. And, finally, it is possible that travel-demand models may adopt a more explicitly statistical approach to modeling travel behavior (as opposed to the current emphasis on a more behavioral approach) as the optimal way of exploiting the massive, passive datasets with which modelers will be increasingly working.

The challenges facing transportation modelers in the emerging informatics-enriched and informatics-enabled world are large. But the opportunities to develop significantly improved and more powerful models for policy analysis and decision support are also great. It is an exciting time to be a transportation modeler!

References

- Bekhor S, Cohen Y, Solomon C (2013) Evaluating long-distance travel patterns in Israel by tracking cellular phone positions. *J Advan Transp* 47(4):435–446
- Ben-Akiva M, Lerman SR (1985) *Discrete choice analysis: theory and application to predict travel demand*. MIT Press, Cambridge MA
- Bonnel P, Morency C, Bayart C (2009) Survey mode integration and data fusion: methods and challenges. *Keeping Up with a Changing World*. Emerald, *Transport Survey Methods*, pp 587–612
- Caceres N, Romero LM, Benitez FG (2013) Inferring origin-destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data. *J Advan Transp* 47(7):650–666
- Calderón F, Miller EJ (2020) A literature review of mobility services: definitions, modelling state-of-the-art, and key considerations for a conceptual modeling framework. Forthcoming, *Transp Rev* <https://doi.org/10.1080/01441647.2019.1704916>
- Calderón F, Miller EJ (2019) A conceptual framework for modeling the supply side of mobility services within large-scale agent-based models. Submitted to *Transportation Letters*

- Castiglione J, Bradley M, Gleibe J (2015) Activity-based travel demand models: a primer. Report S2-C46-RR-1. Transportation Research Board, Washington, DC
- Chung B, Srikukenthiran S, Miller EJ, Nurul Habib K (2020) An inductive experimental approach to developing a web-based travel survey builder: developing a guideline to design efficient web-survey platform. Presented at the 99th annual meeting of the Transportation Research Board, Washington, DC, January
- Faghih Imani A, Miller EJ (2018) Preliminary analysis of Montevideo cellular data records. Report 4: iCity-South: Urban Informatics for sustainable metropolitan growth in Latin America. Report to CAF, the Development Bank of Latin America. University of Toronto Transportation Research Institute, Toronto, January
- Faghih Imani A, Vaughan J, Yusuf B, Miller EJ (2018) Final Project Report, Report 6: iCity-South: Urban informatics for sustainable metropolitan growth in Latin America. Report to CAF, the Development Bank of Latin America. University of Toronto Transportation Research Institute, Toronto, December
- Faghih Imani A, Harding C, Srikukenthiran S, Miller EJ, Nurul Habib K (2020) Lessons from a large-scale experiment on the use of smartphone apps to collect travel diary data: the “City Logger” for the Greater Golden Horseshoe (GGH) area. Presented at the 99th annual meeting of the Transportation Research Board, Washington, DC, January
- Gao D, Srikukenthiran S, Nurul Habib K, Miller EJ (2017) Data fusion: techniques and applications. Transportation Tomorrow Survey 2.0, research report. University of Toronto Transportation Research Institute, Toronto, November
- Graehler M, Mucci RA, Erhardt GD (2019) Understanding the recent transit ridership decline in major US cities: service cuts or emerging modes? 98th annual meeting of the Transportation Research Board. pp 1–19
- Grond K, Miller EJ (2016) Analysis of GPS smartphone data for a Toronto cycling route choice model. Presented at the 5th symposium of the European Association for Research in Transportation, Delft, September
- Harding C, Srikukenthiran S, Zhang Z, Nurul Habib K, Miller EJ (2016) Smartphone app and trace processing assessment, vol 2. Transportation Tomorrow Survey 2.0 research report. University of Toronto Transportation Research Institute, Toronto, November
- Harding C, Srikukenthiran S, Zhang Z, Nurul Habib K, Miller EJ (2016) Smartphone app and trace processing assessment, vol 1. Transportation Tomorrow Survey 2.0 research report. University of Toronto Transportation Research Institute, Toronto, November
- Harding C, Srikukenthiran S, Nurul Habib K, Miller EJ (2017) Evaluating face-to-face surveys to augment regional travel surveys. Transportation Tomorrow Survey 2.0 research report. University of Toronto Transportation Research Institute, Toronto, December
- Harding C, Faghih Imani A, Srikukenthiran S, Miller EJ, Nurul Habib N (2020) Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. Presented at the 99th annual meeting of the Transportation Research Board, Washington, D.C., January
- Janzen M, Vanhoof M, Smoreda Z, Axhausen KW (2017) Closer to the total? Long distance travel of French mobile phone users. *Travel Behav Soc* 11:31–42
- Li W, Shalaby A, Nurul Habib K (2019) UTTRI technical support for the City of Toronto vehicle for hire bylaw review, report no. 5, The relationship between PTC and Public transit: a descriptive analysis. University of Toronto Transportation Research Institute, Toronto, May
- Loa P, Srikukenthiran S, Nurul Habib K, Miller EJ (2015) Current state of web-based survey methods. Transportation tomorrow survey 2.0 research report. University of Toronto Transportation Research Institute, Toronto, September
- Lue G, Miller EJ (2019) Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behav Soc* 14:34–42
- Malinovskiy Y, Saunier N, Wang Y (2012) Pedestrian travel analysis using static bluetooth sensors. Paper 12–3270 in Transportation Research Board annual meeting compendium of papers
- Meyer MD, Miller EJ (2013) Urban transportation planning, A decision-oriented approach. 3rd Edn. <https://mtpplan.com/services.html>

- Miller EJ (2018) Agent-based activity/travel microsimulation: What's next? In: Briassoulis H, Kavroudakis D, Soulakellis N (eds) *The practice of spatial analysis*. Springer, pp 119–150
- Miller EJ (2019) Travel demand models, the next generation: Boldly going where no-one has gone before. In: Goulias KG, Davis AW (eds) *Mapping the travel behavior genome, the role of disruptive technologies, automation and experimentation*
- Miller EJ, Nurul Habib K, Lee-Gosselin M, Morency C, Roorda MJ, Shalaby A (2012) Changing practices in data collection on the movement of people. Final report. Report to the Transportation Association of Canada, Île d'Orléans, Lee-Gosselin Associates Ltd., Quebec
- Miller EJ, Srikukenthiran S, Chung B (2018) Workshop synthesis: household travel surveys in an era of evolving data needs for passenger travel demand. In: Bonnel P, Munizaga M, Morency C, Trepanier M (eds) *Transportation research procedia, transport survey methods in the era of big data: facing the challenges vol 32*. pp 374–382
- Munizaga MN, Palma C (2012) Estimation of disaggregate multimodal public transport origin-destination matrix from passive smart card data from Santiago, Chile. *Transp Res Part C: Emer Technol* 24:9–18
- Nourinejad M, Bahrami S, Roorda MJ (2018) Designing parking facilities for autonomous vehicles. *Transp Res Part B: Methodological* 109:110–127
- Ortuzar JdeD, Willumsen LG (2011) *Modeling transport*, 4th edn. Wiley, New York
- Parada Hernandez C, Miller EJ (2017) Analysis of montevideo smartcard data, preliminary report. Report 3: iCity-South: Urban informatics for sustainable metropolitan growth in Latin America. Report to CAF, the Development Bank of Latin America. University of Toronto Transportation Research Institute, Toronto, September
- Rashed M, Srikukenthiran S, Nurul Habib K, Miller EJ (2015a) Current state of smartphone survey methods. *Transportation Tomorrow Survey 2.0 research report*. University of Toronto Transportation Research Institute, Toronto, September
- Rashed M, Srikukenthiran S, Nurul Habib K, Miller EJ (2015b) Current state of smartphone survey methods: annotated bibliography. *Transportation Tomorrow Survey 2.0 research report*. University of Toronto Transportation Research Institute, Toronto, September
- Rayle L, Dai D, Chan N, Cervero R, Shaheen S (2016) Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transp Policy* 45:168–178
- Srikukenthiran S, Loa P, Hossain S, Chung B, Nurul Habib K, Miller EJ (2018) Final report. *Transportation Tomorrow Survey 2.0 research report*. University of Toronto Transportation Research Institute, Toronto, December
- Train K (2009) *Discrete choice methods with simulation*, 2nd edn. Cambridge University Press, Cambridge
- Trépanier M, Tranchant N, Champleau R (2007) Individual trip destination estimation in a transit smart card automated fare collection system. *J Intell Transp Sys* 11(1):1–14
- Vaughan J, Faghieh Imani A, Yusuf B, Miller EJ (2020) Modelling cellphone trace travel mode with neural networks using transit smartcard and home interview survey data. *Forthcoming, Eur J Transp Infrastrut Res*
- Yin M, Sheehan M, Feygin S, Paiement JF, Pozdnoukhov (2016) A generative model of urban activities from cellular data. *IEEE Transactions in ITS* 1–16



Eric J. Miller has been a faculty member in the Department of Civil and Mineral Engineering, University of Toronto since 1983, where he is currently Director of the University of Toronto Transportation Research Institute. Research areas include activity-based travel modeling, integrated transport-land-use modeling, and agent-based microsimulation. He is the recipient of the 2018 International Association for Travel Behavior Research Lifetime Achievement Award.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

