Analytics for Sensing Urban Space and Experiences

"A smart sustainable city is an innovative city that uses information and communication technologies (ICT) and other means to improve quality of life, efficiency of urban operations, and services, and competitiveness, while ensuring that it meets the needs of present and future generations with respect to economic, social, and environmental aspects."
(ITU-T Study group on Smart Cities, https://www.itu.int/en/ITU-T/ssc/Pages/info-ssc.aspx)

Doina Olaru, UWA Business School

Summary

How can we best understand the underlying urban rhythm of cities? This has been a key question for transport and planning over many decades. Measurements in real-time are needed to design solutions that will improve the quality of experiences for individuals, ensuring the efficient, wise, and safe use of urban infrastructure and spaces. Recent improvements in instrumentation, networking technologies and modelling approaches have resulted in new ways to 'sense' the city.

Sensors and data integration allow us to construct multi-dimensional views of urban life, informing the urban planning and operation of these spaces and providing researchers with critical data to iterate and improve how places are built and operate.

The *smart city* concept has been discussed for many years, with *smart mobility* at its heart. From ultrasonic detectors, magnetometers, cameras, radars, lidars, and weather, noise and air pollution sensors to WiFi, Bluetooth, aerial imagery, drones and satellites, the literature (both scientific and popular) is rich in presenting their benefits to address congestion, safety, health, economic growth and the sustainable use of resources. Mobility is fundamental to the functioning of a city and thus continuous monitoring of transport activities is an integral part of modern *smart city* systems.

However, *smart cities* require an understanding of human behaviour and of the interactions between sensors and individuals. Engaging in different activities informs planners of the individual needs and preferences and of the likely acceptance of new technologies and services, reflecting our social readiness. Thus, regardless of the purpose of the application, a key component of the evolution of *smart cities* is gathering behavioural data and its integration with other data sources to profile human behaviour.

This presentation highlights current trends and solutions within the scope of *smart mobility* (e.g. smart freeways, smart card, mobility as a service (MaaS)) and technologies that enable understanding behaviour, with examples of developments in data analytics, visualisation and machine learning applied to Perth as a case study. The paper reports on several research projects that have assisted professionals in making sense of data flow collected from infrastructure, vehicles (private, public) and individuals (as data-gathering sensors).

City decision-makers can use these integrated data and tools to determine their priorities and challenges in a wide range of applications (e.g., traffic flow optimisation, road maintenance, walkability and promotion of active travel, improved logistics and freight delivery, support for autonomous vehicles) that could lead to urban transformation and improve citizens' quality of life.

This challenging, promising, and cross-disciplinary field of research is even more relevant in the current environment. The global COVID-19 pandemic has radically changed how we interact within our cities and required us to adapt to new lifestyles. Analytics on urban activities and transport from various sensors help us understand the dynamics and identify new alternatives for activities and mobility.

Introduction

Smart transport technology has been heralded as a novel solution for addressing many global transport problems (Mangiaracina *et al*., 2017; Araral, 2020; Boussaid *et al*., 2020; Butler *et al*., 2020), with the *smart city* concept extensively discussed in the literature (both scientific and popular). For example, since 2014, more than *"17,000 papers have been published or presented in conferences, mostly in computer science, engineering, and social sciences."* (Araral, 2020: 1). A Google web search for 'smart city' returns more than 3 billion hits, 'intelligent transport system' 174 million, 'smart mobility' 304 million, 'sensing city' 485

million, in less than 0.65 seconds. Indeed, *"While smart is the order of the day, it appears to be a contested notion that rather defies clarity of definition."* (Lyons, 2018: 4).

Here, I sketch the main components of the *smart city* and offer a few examples where connectivity and integration of data from transport sensors and secondary data add value to the provision of urban services. As a preamble, it is useful to remind readers that a *smart city* includes smart infrastructure, water and sanitation, energy, buildings, smart healthcare and education, smart governance and security, smart mobility solutions and city planning.

This paper focuses on smart transport solutions arising from an analysis of data from sensors (e.g., smart freeways, SmartRider) and enabling technologies that support monitoring, detection, management of operations and services, reinforcement, communication, and planning (Batty *et al*., 2012; Bibri & Krogstie, 2020). Their primary benefits are addressing congestion, safety, and the sustainable use of resources. Hence, steady monitoring of transport activities is an important part of modern *smart city* systems.

However, *smart cities* require an understanding of the behaviour of citizens within them. Engaging in different activities informs planners of the needs and preferences of the citizens, and understanding behaviour is essential for public acceptance of new technologies and services. Thus, regardless of the purpose of the application (e.g., traffic regulation, health, energy use), an essential component of the evolution of *smart cities* is gathering behavioural data and integrating it with other data sources.

This paper examines developments in data analytics, visualisation and machine learning to understand mobility behaviour, using Perth as a case study. City decision-makers can use the integrated data and tools to determine their priorities and challenges in a range of applications that could lead to urban transformation and improve citizens' quality of life.

First, I overview sensors in the city, incorporating their type, role, and impact. Next, I present some examples of integrated data projects and their findings, describing the analytical tools supporting these projects (e.g., econometric techniques, spatial analysis, activity spaces, compositional analysis, machine learning, data integration and visualisation in the form of smart dashboards), with supporting material included in appendices. Finally, I draw the main conclusions and suggest new directions for future research.

Research methods
Transport is by nature a multidisciplinary field that requires many fields to answer pressing questions of its role in shaping/changing an environment. Transport research relies heavily on empirical data and the intensive use of an arsenal of analytical methods to address a variety of problems, such as operations research techniques for optimisations (road allocation problems, maintenance scheduling, routing), statistical techniques (parametric and non-parametric) for describing and explaining/modelling various mobility processes, time series and econometrics for understanding temporal processes, discrete choice models for understanding decision processes related to selecting modes of transport, routes and destinations, spatial analysis and network analysis for examining transport network forms and structures and their changes or modelling the spatial distribution of outcomes, and various simulations to understand flows and agent actions. Many of these methods have been combined into more complex models (e.g., four-step land-use-interaction models, developed in the 1950s - Ortúzar & Willumsen, 2011) that have become commonplace for many transport agencies and professionals with the advent of computational techniques.

More recently, the adoption of mixed methods (merging qualitative insights with quantitative evidence) has pervaded transport research, recognising the multi-faceted nature of transport phenomena. While primary data collection, experimental conditions, or observations have fed more sophisticated models over time, data that is 'sensed' ('big data') requires new analysis methods. Transport field experiences the adoption of advanced data analysis and statistical methods, combined with traditional techniques, and cross-validating their results. This presentation conveys several modelling techniques specific to transport to better understand mobility.

Sensors and technologies
Smart transport technology for 'clever' cities and Intelligent Transport Systems (ITS) are interchangeably used in the literature, identifying technology-enabled solutions relevant to the transport sector (Butler *et al*., 2020). Their overall function is to support transport network controllers in decision-making and improve overall user

satisfaction with the transport system (Gordon, 2016; Perallos *et al*., 2016). The technology applies to monitoring and detection, traffic management, operation and control, enforcement, trip planning and communication, fares and revenue management, emerging mobility services, and other vehicle technologies, such as autonomous vehicles (AV). They play a crucial role in improving traffic flow and safety and consequently are beneficial for reducing energy/fuel consumption and improving air quality. Sensing equipment could be considered as two main types—fixed- and mobile-location sensors—or, as presented here, infrastructure-based, vehicle-based, and individual-based (Table 1).

A notable aspect is that most sensor types—ultrasonic detectors, magnetometers, cameras, radars, lidars, weather, noise and air pollution sensors, WiFi, aerial imagery, drones and satellites—are not used individually. Their richness lies in being combined and presenting distinct aspects of the daily mobility (Batty *et al*., 2012) (e.g., associating individual vehicle technologies using statistical methods between frames and sensors - El Faouzi *et al*., 2016; combining CCTV and plate recognition for monitoring compliance in terms of speed or illegal use of mobile devices, and using together traffic volumes, speeds, weather and incidents for prediction).

Boussaid *et al*. (2020) detailed the guidelines for the use of these systems, which depend on factors such as data requirements (accuracy and format of data output), application requirements (purpose of deploying the sensors), site requirements (practical considerations due to the physical infrastructure and power/communications networks), budgetary requirements, ease of maintenance and ease of integration with traffic management systems.

Another important aspect of smart transport systems is their multiple effects that cascade towards a more sustainable and equitable transport system. As indicated, management and operation systems that use inputs gathered by monitoring and detection systems to improve traffic flow, reduce congestion and improve efficiency have a flow-on effect for reducing emissions, thus benefiting the environment. In addition, management and operation systems, through enforcement functionality, contribute to improved safety outcomes.

Table 1: Examples of sensors and their use

Sensor	Type	Purpose	Measurements and functions	Source
Inductive loops	Infrastructure-based	Monitoring/detection Adapting signalisation and control of traffic	Counts, location of distinct road users, waiting time; assist traffic control (smooth traffic, prioritise modes, lanes); speed can be detected when tubes used together or with induction loops	Mangiaracina *et al*., 2017 Boussaid *et al*., 2020
Magnetometers				
Pneumatic tubes				
Microwave radar	Infrastructure-based & vehicles		Counts and speed	
Active acoustic	Vehicles	Monitoring/detection Warning systems	Speed, counts; assist drivers/vehicles in making decisions when combined with in-vehicle navigation systems,,	Boussaid *et al*., 2020 Mahrez *et al*., 2021
LIDAR Passive infrared		Detection	Identify fixed and mobile possible hazards	Boussaid *et al*., 2020
In-ground piezo sensors	Infrastructure-based & vehicles	Detection	Location, size (weigh-in-motion), speed	Boussaid *et al*., 2020 Bibri & Krogstie, 2020
Wireless (WiFi and Bluetooth): 'sniffers' to detect vehicles		Enforcement (e.g., bus-only and clearway)	Counts, location; incomplete and challenging to map devices to individual traffic users, vehicles	
Radio frequency tags	Vehicles	Detection and revenue management (e.g., tolls)	Counts, location	
Remote sensing (e.g., aerial photography or satellite-based infrared sensors)	Infrastructure-based & vehicles	Increase operational efficiency of the transport system, considering	Location, speed	Mangiaracina *et al*., 2017 Boussaid *et al*., 2020

Sensor	Type	Purpose	Measurements and functions	Source
		interactions between multiple categories		
Global Navigational Satellite System (GNSS), GPS tracking (Automatic Vehicle Location System, AVLS)	Vehicles	Fleet monitoring; reinforcement and asset management; feeds to live ITS (e.g., travel time); providing online map navigation	Speed, location; monitor compliance (heavy vehicles, buses); traffic management and automatic provision of priority to buses	Bibri & Krogstie, 2020
Floating cars		Forecasting travel time	Speed, counts, densities; increase spatial resolution of data	Bibri & Krogstie, 2020 Boussaid *et al.*, 2020
CCTV digital video cameras	Infrastructure-based & vehicles	Live monitoring for immediate intervention; surveillance; reinforcement	Monitor traffic conditions, incidents, weather; improve safety, personal security and provide counts; combined with Automated Number Plate Recognition (ANPR)	
Variable Message Signs (VMS)	Infrastructure-based	Reinforcement Traveller Information System (TIS); Planning and communication	Provide user information on time, availability, incidents, alternative routes and modes, speed limits, lane closures, weather conditions, events, available parking (audio-visual)	
Smart parking	Infrastructure	Management and operation	Improve management of parking spaces and reduce search time	Jun, 2014 Zheng *et al.*, 2015
Smart cards	Individual-based (portable)	Planning and operation; fare and revenue management	Location, density, fare, individual behaviour; provides a powerful "social microscope"	Batty *et al.*, 2012 Birenboim *et al.*, 2021 Toh *et al.*, 2020
Mobile phones, wearables		Better understanding of use of space, traffic patterns, routing	Multiple digital traces (location, speed, weather, in-built sensors, crowdsensing – social media)	
Drones	Vehicles	Monitoring; management and operation	Most of the above	Campbell *et al.*, 2018 Butler *et al.*, 2020
Incorporation of sensor technology in autonomous vehicles (AV) V2V, V2I, V2X				

Note 1: Performance of sensors is subject to many environmental factors (weather, lighting, radio frequency interference and high volume of traffic – microwave radar and wireless); thus, integration of sources also serves as cross-validation and imputation (Batty et al., 2012).

Note 2: Communications networks are key to connecting various sensors and other ITS components to central sites, including private fibre optic cabling, use of public cellular networks (3G to 5G) and private radio networks, such as those operated by Transperth.

Note 3: Most ITS sensors allow for real-time monitoring (and intervention) and asynchronous analysis of long-term changes in the city. Some systems are permanently installed, others are 'sampling' information. Due to the numerous publications, we refer readers to Batty et al., 2012; Mangiaracina et al., 2017; Bibri, 2019; Araral, 2020; Bibri & Krogstie, 2020.

Benefits and challenges

The remarkable possibilities afforded by sensing the city arise from data fusion for traffic management, planning and decision support systems, which requires data analytics and adequate capacities and competencies to capture, process and interpret results.

An example of integration is the Australian Integrated Multimodal EcoSystem (AIMES), developed at the University of Melbourne and installed in Carlton. The 'smart intersection' has more than 200 sensors connecting the whole transport environment within the area, analysing data to improve traffic conditions, while offering safety, reduced congestion, increased sustainability, and a better user experience for traffic participants. Using artificial intelligence (AI), AIMES profiles intersections to identify potentially dangerous activities of all road users—vehicles, cyclists, and pedestrians—developing a deeper understanding of safety

measures and actions for preventing collisions (Juleff, 2019). Artificial neural network (ANN) algorithms are trained/tuned to determine the best way to identify vehicles with sufficient labelled data to identify vehicles for network management and for AV to detect other vehicles in the future.

In Perth, the Smart Freeway technology deployed on Kwinana Freeway is an integrated package of ITS interventions—network intelligence (from infrastructure-based sensors, radars, CCTV), coordinated ramp signalling, speed and lane use management—to monitor traffic conditions and adjust them in real-time, while informing travellers via variable message signs (VMS). The Main Roads WA online *trafficmap* aggregates information from a wide range of systems, such as the Integrated Road Information System (IRIS) that includes data from the SCATS intersection management system.

Another example is the development of MaaS systems, fuelled by deploying a myriad of new mobility services (Shaheen *et al*., 2019), allowing transport users to plan their journey and transport providers to make transport simpler, more efficient and 'just in time'. Putting together emerging transport modes and other transport technologies represents the true value of smart mobility (Hensher *et al*., 2020). Integration into a single user-friendly service/application promotes efficiency, encompassing all aspects of customer journeys: planning, booking, ticketing, payment, updates.

MaaS offers multimodal journey planners to help increase public transport uptake by allowing transport users to plan an efficient journey easily. MaaS includes rideshare services (technically on-demand ride services, such as Uber, Ola and Didi), where drivers provide transport services in exchange for payment, orchestrated through smartphone applications for booking, monitoring and payment. Other notable sharing services include car share (e.g., Car Next Door or Flexicar systems) and bike/e-scooter sharing (docked and dock-less), enabled by mobile applications and information systems that match 'owners' and 'users' (car share) or service booking and payment (e.g., Lime, Spin, Whoosh companies).

Traveller Information Systems (TIS) applications have also advanced in recent decades, allowing users to tap into the data collected from *smart city* sensors and interact with these networks due to their interface between users and wider management, monitoring and planning networks. Users can access many types of information and 'live' travel times, understand the current state of the network (traffic, public transport, parking) and receive future predictions or recommendations for improved use of transport services through VMS, smart devices and broadcasted information (Butler *et al*., 2020). In addition, as illustrated by MaaS, these systems facilitate ride-sharing or mode-of-transport sharing and consumption of integrated door-to-door services, reducing physical and 'cognitive and affective' efforts of travel via information services (Lyons, 2018).

The widespread use of mobile/smartphones has opened other opportunities. Current mobile technologies offer a microcosm of sensors that manufacturers continuously upgrade to satisfy the increasing demands of users and allow them to interact with the environment. They include an impressive number of sensor technologies, such as GPS, one or more cameras and microphones, accelerometer, gyroscope, and pedometer, ambient light, magnetometer and proximity sensor, and newer models even have barometers, temperature and humidity sensors, and heart rate sensors (Butler *et al*., 2020; Birenboim *et al*., 2021). Via smartphones, users act as 'human sensors' and offer data and reports to crowdsourced systems (Birenboim *et al*., 2021).

Finally, AVs are the pinnacle of *smart mobility*, requiring various technologies and 'smart' features for implementation (Campbell *et al*., 2018; Butler *et al*., 2020). They pose several challenges in terms of both technology and user acceptance. A major challenge in the *smart city* context is ICT because AVs rely on the parallel development of ICT and other ITSs. Higher AV levels will require the ability to communicate with other vehicles (V2V), traffic participants (V2X) and infrastructure (V2I). Given the importance of data sharing and communication, 5G, IoT, cloud computing, and interoperable data standards are critical for AV development. Some forms of 5G can provide low latency (1 ms), but there are network coverage problems due to geographical disparities of internet accessibility. Current applications of connected vehicles primarily focus on asset tracking, condition-based maintenance and remote control of auxiliary systems (Boussaid *et al*., 2020).

Connected AVs cooperate by exchanging information on their position and operation, leading to better use of transport infrastructure by coordinating 'platoons' of vehicles travelling closer together. This has fuel efficiency and safety benefits by knowing the behaviour of the surrounding vehicles in more detail.

V2I communication and integration with other network sensors and systems helps AVs sense their environment and drive without human intervention for all or part of the journey (Campbell *et al*., 2018). These systems also offer vehicles a partial automation function (e.g., lane-keeping, warnings, assisting the driver in controlling the vehicle in dangerous situations, Butler *et al*., 2020). Systems such as Autonomous Emergency Braking have existed for many years and deliver tangible safety benefits. Adaptive Cruise Control allows drivers to follow the car in front of them at a safe distance, slowing down or speeding up as required, offering safer, more efficient and enjoyable driving. In all V2- systems, vehicles respond to other vehicles and the broader network, requiring timely and accurate communication.

On the way towards automation, an important milestone is the Electric Vehicle (EV) and electrification of the transport system (Batty *et al*., 2012; Butler *et al*., 2020). EVs have clear environmental benefits when used with electricity generated from non-fossil fuel sources (Bibri & Krogstie, 2020; Butler *et al*., 2020). As the take-up of EVs increases worldwide, it is important to consider their impact on the wider electrical infrastructure network in terms of total demand for electricity and the 'size' of the transmission infrastructure required for quick-charging systems (*smart energy*). New smart surfaces (solar and piezoelectric roads) to recharge EVs have been tested in many countries, including China, France and the USA for solar and USA, UK and Japan for piezoelectric (Toh *et al*., 2019). In Sweden, electrified roads exist, where car and truck batteries recharge as they traverse the road.

However, these benefits do not come without a cost. The most prevalent risks and challenges associated with smart cities include:
- Data privacy (information security) and cybercrime, as many advanced *smart city* technologies require the gathering and use of diverse sources of data regarding individuals (Batty *et al*., 2012; Araral, 2020; Bibri & Krogstie, 2020). These concerns erode consumer confidence and discourage users from embracing *smart mobility*. Recommended remedial actions are investing in high-speed secure communications and ensuring coverage, along with establishing security standards and protocols (Batty *et al*., 2012; Butler *et al*., 2020; Mahrez *et al*., 2021);
- Harmonised supporting communication networks is a key component behind *smart cities* (Mahrez *et al*., 2021). As above, the 'currency' of an ITS is accurate, timely, relevant information, thus ICT needs to be aligned to the ITS deployments, providing flexible, fast, reliable, dependable, scalable communication and interaction;
- Public acceptance of technologies—the literature on technology adoption suggests similarities between individuals' propensity to adopt various technologies, such as mobile telecommunications for work, education, health and entertainment, and the uptake of smart transport technologies. Specific to MaaS, ITS Australia (2018) surveyed Australians to gain customer insights, with 46% of respondents predicted to be ready to use a 'pay-as-you-go' scheme, noting that younger respondents were more interested in using MaaS than those over 65. Government oversight, live information, and personalisation within a seamless platform were desirable features of the system (Reed *et al*., 2021);
- Governance issues (Araral, 2020)—variation in governance and smart policies could explain variations in the technology adoption in a *smart city*; limited understanding of the motivation for promoting the technologies may hinder implementation (Batty *et al*., 2012; Eden Strategy Institute, 2021);
- Data quality, sample size and representativeness (e.g., Bluetooth), and integration with legacy systems also pose challenges (Butler *et al*., 2020);
- Capacity and competency for 'datafication'—the repertoire of tools, methods and techniques needed to match the *"radical expansion in the volume, range, variety, and granularity"* of data generated by the *smart city* (Bibri & Krogstie, 2020: 5) requires specialised knowledge for *smart cities* to reach their potential (Bettencourt, 2014; Mahrez *et al*., 2021). Massive data sets from mobility sensors have leveraged the development of predictive analytics/modelling to design and implement solutions as soon as issues appear or to enhance planning. Yet, data-driven solutions address a variety of purposes, requiring data analysis skills, specialised domain knowledge, and data communication. For operational (short-time) and 'live' updates, these solutions are often disseminated through automated changes in the signalling, adaptation and VMS, with planning solutions communicated through data portals and dashboards, allowing data interpretation at an appropriate level of detail suitable for the end-users' circumstances.

Advances in Data Processing, Analysis and Integration Methods: Case Study—Perth

> *"The wave of the datafication of cities, as mainly enabled by the IoT technology, is giving rise to a new phenomenon-known as the data-driven city."* (Bibri & Krogstie, 2020: 3)

The growing interest in *smart cities*, patterns of urban mobility and the role of sensors to improve urban rhythms goes hand-in-hand with methodological developments that provide researchers with unprecedented access to data on daily lives (Batty *et al*., 2012; Bibri & Krogstie, 2020). Fusing and processing large amounts of data from various sources requires analytical skills beyond the traditional multivariate analyses, and the potential of big data technology is "*still largely untapped*" (Bibri & Krogstie, 2020).

Methodological developments discussed in the presentation include spatial analysis, activity spaces, compositional analysis, econometric techniques and multilevel modelling. An overview is also provided of how data is integrated and visualised in smart dashboards (e.g., RailSmart Planning Support Platform), augmenting the information provided to planners by characterising the main attributes of mobility within a *smart city* ecosystem. This application of quantitative, computational, visual methods is considered now under the umbrella of 'urban analytics'.

As general descriptive and predictive models are part of the standard toolkit of data science researchers, along with many supervised and unsupervised pattern recognition techniques, I focus more on techniques specifically used by transport modellers to understand travel behaviour using examples from four different projects, with some techniques presented in the Appendix. All of the examples use *big data* in many of their dimensions: volume, variety, veracity, velocity, vagueness, or value.

Whenever possible, cross-validation by machine learning algorithms (Murphy, 2012) was used to benchmark the results obtained through multivariate data analysis. Feedforward ANN and self-organising maps (Kohonen networks) were applied to cross-sectional and longitudinal data to identify patterns automatically or make predictions or classifications. The volumes of data allowed for training the models with 70–80% of the records, evaluating the performance on the remainder as test data.

Projects
Railsmart
The RailSmart project embodies the *smart city* concept, extending earlier paradigms:

> *"... from an initially technology-driven focus (with high emphasis on ICT) aimed at maximizing efficiency of the hard urban infrastructure (i.e., transport, communications, waste, energy, water, etc.) to adopting a more comprehensive approach, wherein the central role of people and the soft infrastructure (i.e., institutions, citizen engagement, data, social innovation, knowledge economy, justice, etc.) is acknowledged. By integrating multiple dimensions, smart cities intend to function as agents of strategic transformation."* (Sharifi, 2019: 1270).

It is an example of developing urban intelligence functions, integrating data and models to improve efficiency/productivity, resilience and sustainability. The project provides a planning support system that relies on more than 60 input data sources, has a backend (API) platform for dynamic modelling analysis, and a frontend web application for users to analyse the transport features of the city (built around railway corridors), characterising the city using socio-demographics and employment, and testing various scenarios.

Four main tools support the platform (http://railsmart.patrec.org/login):
Tool 1: Activity Centre Node-Place Analysis and Activation Tool provides a detailed analysis of the railway stations in Perth with respect to *Place* (quality and vibrancy of built environment), *Node* (transport features and functions of the station), and *Background Traffic* characteristics (station accessibility via park-and-ride/kiss-and-ride, feeder buses, taxis, or active travel and interaction with road traffic).
Cluster analysis and profiling of the stations was undertaken to generate groups with similar Place, Node, and Background Traffic characteristics, which fed a regression analysis to predict patronage based on key station and catchment characteristics. The analysis identified patterns common to various station types and then predicted patronage and mode choice for new planned stations.
Tool 2: Railway Usage Analysis and Travel Behaviour is a data-driven tool centred on the analysis of SmartRider data. While Tool 1 examined transport from the built form and aggregated flows, Tool 2 analysed the way people move through the public transport system. This tool focused on the intervals between journeys and examined time-use patterns between when passengers leave the system and when they re-enter it during

the day. For example, gaps between 9 am and 3 pm (corresponding to school attendance), 9 am and 5 pm (work), and 7 pm and 8 am (home) tell us about activities and land-uses around the transport hubs.

Tool 3: Enhanced Employment Self Sufficiency Tool explores ways to maximise employment potential for the newly proposed Alkimos, Eglinton, and Yanchep railway stations. The employment creation follows distinct processes, with some jobs across a range of skill levels naturally occurring once the population grows, creating the demand for services (e.g., nurses, teachers, and retail personnel), and others, more strategic in nature, being created due to a competitive advantage (e.g., mining occurs where exploration identifies resources or the computer industry develops in areas of concentrated high-level computer skills). Tool 3 mapped the proportions of the two types of occupation categories and location quotients using three distinct analytical bases (Perth, WA and Australia). The tool allows planners to direct employment creation to strategic jobs, based on competitive advantage, committing the land, services and infrastructure needed to support them.

Tool 4: Dynamic Accessibility Planning Tool calculates and displays city-wide accessibility (by public or private transport) from any station to the centroid of all STEM zones (derived from the Department of Transport's Strategic Transport Evaluation Model). The tool measures accessibility to jobs, houses and population.

The project also includes a future-oriented scenario builder based on projections. While the metropolitan analysis is factual and historical, the scenario builder anticipates/forecasts patronage changes from land-use and transport changes. The value of the scenario builder lies in broad trends. The system tests for different development options, so strategies can be compared to achieve optimal outcomes. The RailSmart Dashboard generates visualisations that assist expert and non-expert users to interpret and analyse information and understand the city on its urban fabric, strategic advantages of various areas, public transport options, and performance of places.

Patronage prediction

Using SmartRider data from 2009 to 2019, this project investigated factors contributing to changes in patronage to understand spatial and temporal variations in mobility and to provide guidance on service features that may be needed to increase the mode share of public transport. An important feature of the SmartRider system in Perth is that the card enables the electronic collection of the fares and information gathering from the first 'tagging on' to the last 'tagging off', when the user completes their travel by public transport in their linked journey. This provides information on public transport use and demand, both spatially (origin-destination, O-D, pairs) and by time-of-day. In such studies, privacy is always considered and the data anonymised through several scrutiny levels and confidentiality checks.

The analysis integrated (again) multiple data sources (e.g., Census, land-use, economic indicators, traffic) and included descriptive and predictive analytics accounting for the multilevel nature and different spatial and temporal scales of the various datasets. Several sets of descriptive, diagnostic, and predictive techniques have highlighted the role of accessibility and macro-level socio-economic factors in explaining the patronage variation over the study period. An alternative modelling methodology was used to compare predictions to benchmark the effectiveness of the spatial regression models for determining patterns between patronage and other factors over space and time or the results of latent growth models for patronage (details provided in the Appendix). Firstly, ANN was used to compare the statistical fit to the data, i.e. accuracy of prediction. Secondly, the flexibility of ANN was exploited to complement/enrich the findings of the multivariate models and estimate the contribution of fares and service levels to patronage. Then, ANNs were applied to undertake a forecasting exercise considering the impact of distinct fare strategies and the removal of some services (bus*km) at certain locations in the network. While the ANNs did not outperform the detailed spatial regression models in terms of predicting observations in the hold-out/test sample (both predictions had normalised RMSE close to 0.07), transport professionals could interpret the contribution factors of ANN more easily than non-linear analytical models. In addition, the ANN residual maps provided clear spatial dependence patterns and identified more outliers with high volumes of journeys than the spatial combined models.

Active travel and activity spaces

Big data on human mobility is a fundamental component of *smart city* analytics, thanks to the tremendous penetration of user-based technology, allowing data collection with high spatial and temporal resolutions. When combined with accurate location readings (i.e., through location tracking technologies), portable devices can generate information with a precise spatiotemporal stamp. That is, combining passive and active data sources can help us better understand *smart mobility* users and offer insights into their daily routines, pointing out the benefits of active travel (often incidental) for human physical activity. While the project was sparked

by the idea of identifying aspects of transport that induce behavioural changes for individuals with health and wellbeing benefits (Bauman *et al*., 2019; Harms *et al*., 2018), it revealed and reiterated the potential of wearables to portray travel patterns more accurately. Our expectation of consistency between the detailed activity diaries and video recording and GPS tracks turned out to be false, with substantial variation in frequency, duration and timing of episodes. Acknowledging the missing data and potential/lack of reliability of records provided by smartwatches and video cameras (mainly due to technological constraints posed by their size and limited power), the corroboration of data sources served as a validation, which we argue is imperative for ensuring the credibility of the results.

The analysis highlighted three distinct typologies of travellers, with different activity spaces (extent of urban areas covered to satisfy daily mobility needs). Similar heterogeneity was highlighted in a separate project, examining the association between the quality of food environments and accessibility using mobile data (see Appendix for more detail).

One of the main take-aways was that wearables/portable sensing, including 'sparse' mobile traces, allows researchers to mitigate several methodological issues related to the delineation of the context and the environmental impact assessment on daily human activities, including (1) uncertain geographic context due to using predefined geographical units (e.g., census mesh blocks, neighbourhoods) as areas for capturing the role of environmental factors on mobility and their high aggregation level; (2) improved ecological validity of the study (i.e., whether the study findings can be generalised to real-life situations).

Autonomous Vehicles (AV)
A resounding success of the UWA engineering students provided the opportunity to further reflect on the importance of analytics for offering AV technology solutions that the public would trust and accept (https://www.uwa.edu.au/news/article/2021/june/uwa-students-first-in-australia-to-build-brains-of-autonomous-bus). The nUWAy shuttle bus is equipped with sensors to guide navigation and prevent collisions (eight Lidar sensors, two cameras, radar, GPS and an inertial measurement system), with an independent hardware system monitoring vehicle movements, automatically stopping if the proximity perimeter is 'invaded'.

Undoubtedly, sensors play a key role in AV functioning and the public's confidence to adopt these vehicles. While there is no V2V or V2I communication to feed the nUWAy, AVs do not (or will not) solely rely on local sensors to respond to stimuli. Connected AVs are more likely to reap the benefits for traffic congestion and air quality; they can also communicate with nearby vehicles and/or infrastructure and use the information to make decisions, reducing vehicle reaction time (to near zero) as the required information is directly transmitted, rather than sensed.

The two main forms of vehicle connectivity, V2V and V2I, have distinct roles, which were explored in detailed simulation models by colleagues at UWA's Planning and Transport Research Centre, PATREC (www.patrec.org). The AV research considered the simplest V2I communication, in which traffic signals broadcast their status to queueing CAVs so that they can react when lights turn green (https://www.uwa.edu.au/news/article/2021/february/driverless-vehicles-could-create-more-traffic-congestion). Other more advanced applications, such as using V2I for optimising network operations, will be investigated. In terms of V2V, vehicles only communicate with their immediate neighbours in the same lane and require the leading and following vehicles to be connected, with the following CAV acquiring the speed and acceleration of the CAV in front with no communication delay. However, when a CAV follows a non-CAV, the follower must rely on its sensors (and their associated delays) and switch its driving behaviour to an unconnected model. The simulation scenarios confirmed the challenges of mixed AV and non-AV traffic and emphasised that AV and CAV perform well at high market penetration rate.

Conclusion
Current *smart cities* provide enormous amounts of data from fixed and portable sensors waiting to be exploited for 'smart' decision-making and planning. Although there is no clear definition of a *smart city*, and multiple, complementary views are offered, the research community converges on the idea that *smart cities* are not only an issue of hardware and software, ignoring applications, but emphasise the criticality of data-driven 'smartness' and consider the human dimension. Consequently, our ability to harness data—collect, clean,

fuse/combine/integrate, and analyse primary and secondary data—from distinct sources is expected to lead to actionable insights.

The project examples provided here show a more human-centred approach, where data allows us to identify the underlying urban rhythm at different temporal scales and understand the transformations triggered by mobility processes. This is a step forward in achieving a key objective of *smart cities*—understanding activity engagement and offering citizens improved safety, efficiency, reliability, and sustainability while addressing a wide range of environmental and socio-economic concerns supported by advanced technologies and data-driven solutions.

With the exception of the AV project, all projects combined 'organic' data from sensors (smartcards, traffic flows from various sources, wearables, mobile phones) with many secondary sources (Census, land-use/points of interest, economic indicators, surveys, thus purposeful data) to understand mobility needs and patterns in Perth. All made use of visualisation (and GIS mapping) and attempted to cross-validate results from the traditional predictive models and of segmentation with machine learning techniques.

Whereas the *RailSmart* and *Patronage* projects used only quantitative data at various spatial and temporal scales, the *Active travel and activity spaces* highlighted the role of the additional visual images in understanding the context of the mobility decisions (zooming into daily decisions). The mix of data types (and of 'big' and 'small' data) and their increasing volume confirm the need for analytics to gain insights into the urban processes and inform policy and practice. In addition, our applied solutions provide some grounding for more in-depth investigation of human mobility and suggest some sequel ideas for research, some of which are underway: examining the role of changing travel on transport and communication technologies in an ongoing or post-pandemic world; assessing technological advancements, but more importantly, data-intensive initiatives and policies to improve the smartness/readiness level of our city; shaping pathways towards AV mobility, through pilot programs and trials, as part of the smart mobility strategy for Perth.

Considering the rapid changes in technologies and methods, assessing the impact of *smart city* developments calls for intensified efforts to continue data integration and analytics in multidisciplinary teams to understand enablers and barriers and prepare cities to become more adaptable with enhanced foresight. This requires funding projects, supporting programs and creating 'ecosystems' with a wide range of stakeholders to sustain innovation and catalyse city initiatives, approaches, and policies, with a genuine focus on the individual.

The shift from efficacy and outcomes of technological solutions towards individual needs is timid, but visible; nevertheless, it has intensified data privacy and cybersecurity concerns, which require special consideration.

Note: Supplementary material (presentation file, including output examples from the models) is available at: https://patrec.org/other-publications/.

References

1. Aitchison, J. (2003) *The statistical analysis of compositional data*, London: Blackburn Press.
2. Araral, E. (2020) Why do cities adopt smart technologies? Contingency theory and evidence from the United States, *Cities*, 106, 102873, https://doi.org/10.1016/j.cities.2020.102873.
3. Axhausen, K.W., Botte, M. & Schönfelder, S. (2004) Measuring the spatial reach of persons, cities or organisations, Presented at the *Focus Group 3 Meeting of the European Union Network Sustainable Transport in Europe and Links and Liaisons to America (STELLA)*, January, Arlington, TX, USA.
4. Batty, M., Axhausen, K. W., Gianotti, A., Pozdnoukhov, A., Bazzani, M., Wachowicz, M., & Portugali, Y. (2012) Smart Cities of the Future, *European Physical Journal Special Topics*, 214(1), 481-518, https://doi.org/10.1140/epjst/e2012-01703-3.
5. Bauman, A., Bittman, M., & Gershuny, J. (2019) A short history of time use research; implications for public health, *BMC Public Health*, 19 (Suppl. 2), 607, https://doi.org/10.1186/s12889-019-6760-y.
6. Bettencourt, L.M. (2014) The Uses of Big Data in Cities, *Big Data*, 2(1), 12-22, https://doi.org/10.1089/big.2013.0042.

7. Bibri, S.E. & Krogstie, J. (2020) The emerging data-driven Smart City and its innovative applied solutions for sustainability: the cases of London and Barcelona, *Energy Informatics*, 3, 5, https://doi.org/10.1186/s42162-020-00108-6.

8. Birenboim, A., Helbich, M., & Kwan, M.-P. (2021) Advances in portable sensing for urban environments: Understanding cities from a mobility perspective, *Computers, Environment, and Urban Systems*, 88, 101650, https://doi.org/10.1016/j.compenvurbsys.2021.101650.

9. Bollen, K.A. & Curran, P.J. (2006) *Latent curve models: A structural equation perspective* (Vol. 467), Hoboken, NJ: Wiley.

10. Botte, M. (2015) *The connection of urban form and travel behaviour: a geo-spatial approach to measuring success of transit oriented developments using activity spaces*, PhD Thesis, The University of Western Australia.

11. Botte, M. & Olaru, D. (2012) Geo-spatial analysis of Activity Spaces in a TOD environment – Tracking impacts of rail transport policy using kernel density estimation, *Road and Transport Research*, 21(1), 64–81.

12. Boussaid, F., Ward, I., Bennamoun, M., Sun, C., Pierau, E., & Biermann, S. (2020). *Enhanced Vehicle Detection at Traffic Signals and Smart Freeways* (unpublished).

13. Burt, W.H. (1943) Territoriality and Home Range Concepts as Applied to Mammals, *Journal of Mammalogy*, 24(3), 346–352, https://doi.org/10.2307/1374834.

14. Butler, L., Yigitcanlar, T., & Paz, A. (2020) Smart Urban Mobility Innovations: A Comprehensive Review and Evaluation, *IEEE Access*, 8, 196034-196049, https://doi.org/10.1109/ACCESS.2020.3034596.

15. Cagney, K.A., Cornwell, E.Y., Goldman, A.W., & Cai, L. (2020) Urban Mobility and Activity Space. *Annual Review of Sociology*, 46, 623–648, https://doi.org/10.1146/annurev-soc-121919-054848.

16. Campbell, S., O'Mahony, N., Krpalcova, L., Riordan, D., Walsh, J., Murphy, A., & Ryan, C. (2018) Sensor technology in autonomous vehicles: A review. In *29th Irish Signals and Systems Conference, ISSC 2018*. https://doi.org/10.1109/ISSC.2018.8585340.

17. Casas, I. (2007) Social Exclusion and the Disabled: An Accessibility Approach, *The Professional Geographer*, 59(4), 463-477, https://doi.org/10.1111/j.1467-9272.2007.00635.x.

18. Cervigni, E., Renton, M., Haslam-McKenzie, F., Hickling, S., & Olaru, D. (2020) Describing and mapping diversity and accessibility of the urban food environment with open data and tools, *Journal of Applied Geography*, 125,102352, https://doi.org/10.1016/j.apgeog.2020.102352.

19. Duncan, T.E. & Duncan, S.C. (2009) The ABC's of LGM: An Introductory Guide to Latent Variable Growth Curve Modeling, *Social and Personality Psychology Compass*, 3(6), 979-991. https://doi.org/10.1111/j.1751-9004.2009.00224.x.

20. Eden Strategy Institute (2021) *Top 50 Smart City Governments*, https://www.smartcitygovt.com.

21. Egozcue, J.J. & Pawlowsky-Glahn, V. (2005) Groups of parts and their balances in compositional data analysis, *Mathematical Geology*, 37(7), 795-828, https://doi.org/10.1007/s11004-005-7381-9.

22. El Faouzi, N.E. & Klein, L.A. (2016) Data fusion for ITS: techniques and research needs, *Transportation Research Procedia*, 15, 495-512.

23. Golledge, R.G. (1997) *Spatial Behavior: A Geographic Perspective*, USA: Guilford Press.

24. Gonzalez, M.C., Hidalgo, C.A., & Barabasi, A.-L. (2008) Understanding individual human mobility patterns. *Nature* 453, 779–782, https://doi.org/10.1038/nature06958.

25. Gordon, R.L. (2016) *Intelligent Transportation Systems*, Switzerland: Springer.

26. Grifoll, M., Ortego, M.I., & Egozcue, J.J. (2019) Compositional data techniques for the analysis of the container traffic share in a multi-port region, *European Transport Research Review*, 11:12, https://doi.org/10.1186/s12544-019-0359-z.

27. Hägerstraand, T. (1970) What about people in regional science, *Papers of the Regional Sciences Association*, 24, 6–21. https://doi.org/10.1111/j.1435-5597.1970.tb01464.x.

28. Hensher, D.A., Mulley, C., Ho, C., Wong, Y., Smith, G., & Nelson, J.D. (2020) U*nderstanding Mobility as a Service (MaaS). Past, Present, and Future*, Kidlington, UK: Elsevier.

29. Juleff, S. (2019) *Making sense of Australia's smartest intersection*. https://about.unimelb.edu.au/newsroom/news/2019/november/making-sense-of-the-smartest-intersection-in-australia.

30. Harms, T., Gershuny, J., & Olaru, D. (2018) Using time-use data to analyse travel behaviour: Findings from the UK, *Transportation Procedia*, 32, 634-648, https://doi.org/10.1016/j.trpro.2018.10.007.

31. Harms, T., Olaru, D., & Pattison, C. (2019) Using wearable technology to analyse daily travel behaviour, *Australasian Transport Research Forum (ATRF)*, Canberra, 30 Sep-2 Oct.

32. Horton, F.E. & Reynolds, D.R. (1971) Effects of Urban Spatial Structure on Individual Behavior, *Economic Geography*, 47(1), 36-48.

33. Jennrich, R.I. & Turner, F.B. (1969) Measurement of non-circular home range, *Journal of Theoretical Biology*, 22, 227-237.

34. Kamruzzaman, M. & Hine, J. (2012) Analysis of rural Activity Spaces and transport disadvantage using a multi-method approach, *Transport Policy*, 19(1), 105–120, https://doi.org/10.1016/j.tranpol.2011.09.007.

35. Lyons, G. (2018) Getting smart about urban mobility – Aligning the paradigms of smart and sustainable, *Transportation Research A*, 115, 4-14, http://dx.doi.org/10.1016/j.tra.2016.12.001.

36. Mahrez, Z., Sabir, E., Badidi, E., Saad, W., & Sadik, M. (2021) Smart Urban Mobility: When Mobility systems Meet Smart Data, *IEEE Transactions on Intelligent Transportation Systems*, 1558-0016, https://doi.ord/10.1109/TITS.2021.3084907.

37. Mangiaracina, R., Perego, A., Salvadori, G., & Tumino, A., (2017) A comprehensive view of intelligent transport systems for urban smart mobility, *International Journal of Logistics Research and Applications*, 201(1), 39-52, https://doi.org/10.1080/13675567.2016.1241220.

38. McArdle, J.J. (2009) Latent variable modeling of differences and changes with longitudinal data, *Annual Review of Psychology*, 60, 577-605, https://doi.org/10.1146/annurev.psych.60.110707.163612.

39. Murphy, K.P. (2012). *Machine Learning: A Probabilistic Perspective*, Cambridge, MA: MIT Press.

40. Olaru, D., Moncrieff, S., McCarney, G., Sun, Y., Reed, T., Pattison, C., Smith, B., & Biermann, S. (2019) Place vs. Node Transit: Planning Policies Revisited, *Sustainability*, **11**, 477, 1-14, https://doi:10.3390/su11020477.

41. Olaru, D., Robson, L., Biermann, S., Cardell-Oliver, R., Smith, B., Sun, C., & Reed, T. (2019) The Role of TODs in Achieving Smart and Sustainable Cities, *State of Australian Cities (SOAC) Conference*, Perth, 3-5 Dec.

42. Olaru, D., Smith, B., & Biermann, S. (2021) Enhancing land-use inputs to strategic transport models through advanced econometrics, *Planning and Transport Research Centre (PATREC) Report* for the WA Department of Planning, Lands, and Heritage (DPLH), Main Roads Western Australia, 107pp, iMove Project CRC 1-023, 55pp, https://resources.patrec.org/reports/projects2021/cubeland.pdf.

43. Ortúzar, J. de D. & Willumsen, L.G. (2011) *Modelling Transport*, 4th Edition, John Wiley and Sons: Chichester.

44. Perallos, A.,, Hernandez-Jayo, U., Onieva, E., & García Zuazola, I.J. (2016) *Intelligent Transport Systems: Technologies and Applications*, Chichester, UK: John Wiley & Sons.

45. Rai, R.K., Balmer, M., Rieser, M., Vaze, V.S., Schönfelder, S., & Axhausen, K.W. (2007) Capturing Human Activity Spaces New Geometries, *Transportation Research Record*, 2021, 70–80.

46. Reed, T., Ji, Y., Sun, C., & Biermann, S. (2019) *Pulse of Perth: SmartRider Data Visualisation* https://resources.patrec.org/reports/projects2019/RacPop-FinalReport.pdf.

47. Robson, L., Biermann, S., Cardell-Oliver, R., Olaru, D., Martinus, K., Smith, b., Chi, S., Sun, Y., Reed, T., Jie, F., Standing, C., Parsons, C., ARUP (2019) Railsmart Technical Report, *Planning and Transport Research Centre (PATREC) Report* for the WA Department of Transport and the City of Wanneroo, Smart Cities and Suburbs Program, https://resources.patrec.org/reports/projects2019/RailSmart-TechnicalReport.pdf.

48. Schönfelder, S. & Axhausen, K.W. (2010) *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*, London: Taylor & Francis.

49. Shaheen, S., Cohen, M., & Farrar, E. (2019) Carsharing's impact and future, *Advances in Transport Policy and Planning*, 4, 87-120, https://doi.org/10.1016/bs.atpp.2019.09.002.

50. Smith, B., Olaru, D., Reed, T., Cardell-Oliver, R., & Biermann, S. (2020) Factors Influencing Public Transport Patronage Trends: Perth 2009 to 2019, *Planning and Transport Research Centre (PATREC) Report* for the WA Department of Transport (DoT), Public Transport Authority (PTA), iMove Project CRC 3-007, Milestone 2.1, 112pp, https://resources.patrec.org/reports/projects2020/uptp1.pdf.

51. Smith, B., Olaru, D., Reed, T., Cardell-Oliver, R., & Biermann, S. (2020) Factors Influencing Public Transport Patronage Trends: Perth 2015 to 2019, *Planning and Transport Research Centre (PATREC) Report* for the WA Department of Transport (DoT), Public Transport Authority (PTA), iMove Project CRC 3-007, Milestone 3.1, 107pp, https://resources.patrec.org/reports/projects2020/uptp2.pdf.

52. Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010) Limits of Predictability in Human Mobility, *Science*, 327, 1018–1021, https://doi.org/10.1126/science.1177170.

53. Sun, Y., McCarney, G., Sabri, S., Amirebrahimi, S., Agunbiade, M., Chen, Y., Keshtiarast, A., Kalantari, M., Bright, C., Olaru, D., Reynolds, M., Rajabifard, A., & Biermann, S. (2017) A flexible accessibility analysis tool for enhanced urban analytics, Paper presented at the *Australasian Transport Research Forum (ATRF)*, Auckland, New Zealand, 27-29 Nov.

54. Sun, Y., Olaru, D., Smith, B., Greaves, S., & Collins, A. (2017) Road to autonomous vehicles in Australia: an exploratory literature review, *Road and Transport Research*, 26(1), 34-47.

55. Toh, C.K., Sanguesa, J.A., Cano, J.C., & Martinez, F.J. (2020). Advances in smart roads for future smart cities. *Proceedings of the Royal Society A*, 476: 20190439, http://dx.doi.org/10.1098/rspa.2019.0439.

56. Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015) Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach, *Transportation*, 42, 625–646, https://doi.org/10.1007/s11116-015-9597-y.

Appendix

Activity spaces (AS)

The theory of activity spaces (AS) stems from biological research undertaken on habitat use, territoriality behaviour, and mammalian home range studies from as early as the 1950s and 1960s (Burt, 1943; Jennrich & Turner, 1969) and applied in many areas since (e.g., analysis of crime incident locations, accessibility to health care services, impact of transit oriented-development - see Botte & Olaru, 2012; or food deserts, Cervigni *et al*., 2020).

AS are an interaction of the individual with the surrounding environment; thus, it considers travel demand and supply and time limitations and institutional constraints. AS represent the spaces needed to cover required daily activities (e.g., work, recreation, health care, food purchase or consumption), accounting for transport services that enable users to meet those needs (Figure A1). AS are geospatial, statistical measures that describe spatial use (realised AS) (Axhausen *et al*., 2004; Kamruzzaman & Hine, 2012; Cagney *et al*., 2020) or spatial perception, and spatial awareness of travellers (Horton & Reynolds, 1971; Casas, 2007), at the individual or aggregate level.
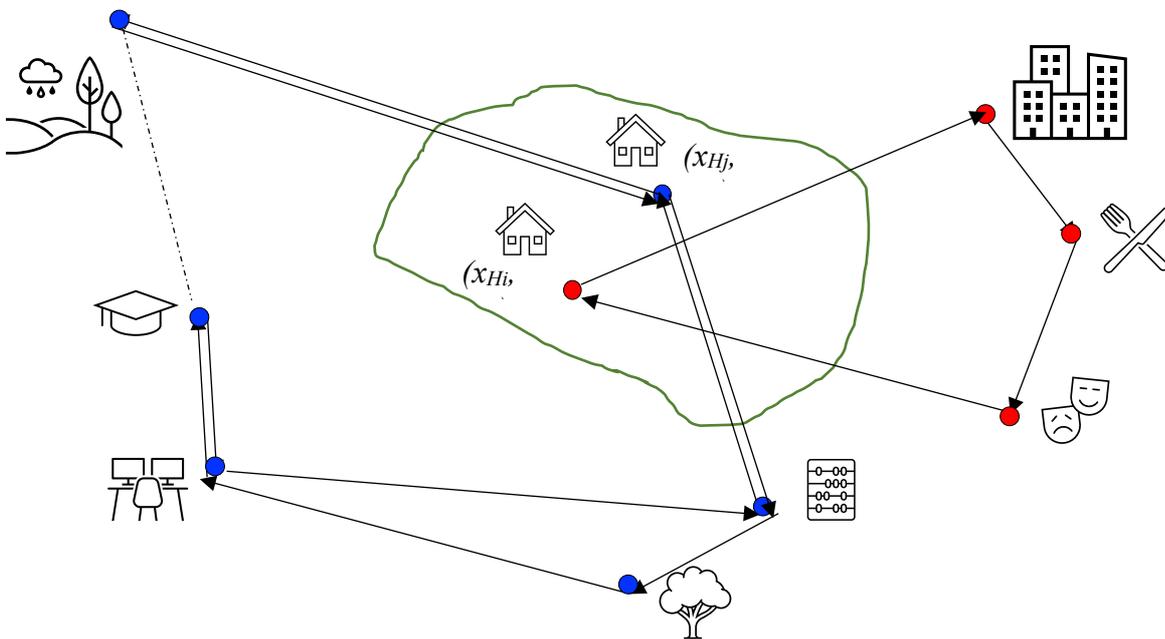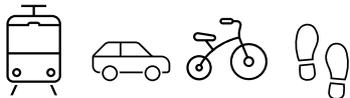

Figure A1: Activity locations visited by individuals in two households (red and blue) living in the same area and using distinct modes of transport (below)

Activity spaces have much in common with the space-time paths and prisms, demarcating possible locations for activities within a given time 'budget' for travel and considering the speed of transport services (Hägerstrand, 1970; Axhausen *et al*., 2004). By reflecting the space in which individuals conduct their daily activities and the areas they traverse, AS allow us to consider the effects of exposure to various environments facilitated by travel. A significant improvement in the quality of services, resulting in greater potential to reach further destinations, may be reflected in larger AS or in time savings embedded in stationary activity times. Similarly, AS adjust dynamically due to changes in the built environment and transport services and may increase with more access to activities further from home or decrease if the opportunities are closer to home.

Different formulations have accentuated the various dimensions of travel-related choice decisions or the potential to reach opportunities while connecting spatial information and temporal dimensions (e.g., time spent on activity and travel; Botte, 2015). Initial work applied simple elliptical measures. Later testings incorporated kernel densities, minimum spanning tree (Axhausen *et al*., 2004; Schönfelder & Axhausen, 2010), or alternative geometries (Rai *et al*., 2007; Botte & Olaru, 2010). Rai *et al*. (2007) compared new geometries, such as superellipses, Cassini ovals and bean curves with confidence ellipses, providing insight and

recommendations on the appropriateness of various shapes for certain conditions. As these measures are determined by the household basic places/pegs (home, work, school, other frequently visited activity centres) and accessibility provided by the transport network, they highlight various travel behaviour aspects, with no single all-encompassing measure. Botte (2015) tested multiple measures and geometries and concluded that the bivariate kernel density estimator is preferred when more detailed data are available (particularly GPS tracks).

The kernel is a non-parametric raster-based approach that captures the activity intensity/density landscape without setting predetermined geometrical boundaries (beyond an initially specified raster cell size and a bandwidth based on a search radius specified by the analyst). The sequential application of a preselected bivariate density distribution over individual activity locations (point densities) in the raster, followed by a smoothing algorithm, produces individual, converging, continuous and/or clustered density surfaces, reflecting various activity patterns and identifying areas of intensive use. The measure can incorporate activity frequency or duration in particular locations (or varying spatial awareness) and consider both footprint and surface areas at 100% (all non-zero activity density area) or any other specified confidence level (e.g., 95%, 50% interval). This allows the analyst to discard parts of AS that may include non-typical activities in the life of individuals, determined by some special conditions or events and not representative of their daily routines.

The application presented here for food deserts uses mobile data from Place Intelligence (2019) and applies kernel density measures. AS are compared against clusters of food environment with various richness/abundance and diversity levels and socio-economic indicators (e.g., ABS Index of Education and Occupation, IEO). The results suggest significant associations between AS and food environment, with small AS for mobile devices with homes in the inner city and suburban areas with high food access, and large AS towards the urban fringes (mobile devices assigned to the last three clusters of poor food access must travel further to access stores where they can purchase fresh food). Unsurprisingly, IEO top deciles (8–10) correspond to the lowest AS, confirming that those living in the inner suburbs of Perth can get most of their nutrition needs within 1.2 km of their homes. Conversely, deciles 1–7 have relatively large AS, suggesting high mobility, high spatial coverage, and/or limited food outlets. Thus, living in affordable areas, well-catered with food outlets, or working in activity centres further from home, but accessing food en route, ensure access to good quality food, even if the overall AS may not reveal substantial differences between these two categories. In terms of spatial heterogeneity in the food environment, both the Perth CBD and suburban activity centres are important locations of high food abundance and diversity for all socio-economic groups.

Compositional analysis (CoDA)
Similar to the compositions expressed as molar concentrations of each component in chemistry, mineral structure of rocks in geology, or distinct % of ethnicities in various parts of a city in demography, data on time spent in various daily activities is compositional and thus adds up to a constant value, 24 h. Therefore, changing the time spent in one part will inevitably lead to a change in time for at least one of the remaining parts. As per many time-use researchers and activity-travel modellers, our interest is to identify trade-offs between travel and other daily activities.

Compositional data has different properties from unconstrained data and thus requires other analytical approaches that account for its constrained structure. CoDA is a family of techniques addressing the mathematical nature of the 'fixed-sum' data (compositional vectors with relative information of the parts to their whole), with the absolute values in the composition transformed into sets of log-ratios.

A composition is defined as a vector of strictly positive components x_i, adding to a constant c and carrying only relative information (proportional vectors provide the same compositional information).
For example, a D-part composition can be written as $x=[x_1, x_2, ..., x_D]$, and the c (closure) can be a physical raw measurement or converted to percentages ($c = 100$).

The suitable sample space for the representation of compositional vectors is the Simplex, and the so-called Aitchison geometry for the Simplex includes distance between compositions. The distance between two compositions $x=[x_1, x_2, ..., x_D]$ and $y=[y_1, y_2, ..., y_D]$ is defined as:

$$d(x, y) = \sqrt{\frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left(ln \frac{x_i}{x_j} - ln \frac{y_i}{y_j} \right)^2} \qquad [1]$$

And the corresponding norm is: $\|x\| = \sqrt{\frac{1}{2D}\sum_{i=1}^{D}\sum_{j=1}^{D}\left(\ln\frac{x_i}{x_j}\right)^2}$ [2]

CoDA principles include three conditions that need to be met by the statistical methods applied to compositional data: scale invariance, permutation invariance, and coherence (Aitchison, 2003). We take advantage of these properties to analyse time spent in sedentary activities, recreational physical activity, and travel time, including active travel.

There are several types of algorithms for log transformation of compositional data:

- Additive log-ratios (*ADL*) refer to the $\ln\frac{x_i}{x_j}$, where x_j is a reference category for all the other parts and treated as a common denominator; *ADL* enables mapping of a composition in the *D*-part Aitchison Simplex to a *D-1* dimensional Euclidean vector. The *D* parts can be reordered, giving *D* possible sets of *ADL*.

- Centred log-ratios (*CLR*) are log-ratios between part x_i and the geometric mean of all parts $\ln\left(\frac{x_i}{(\prod_{i=1}^{D}x_i)^{\frac{1}{D}}}\right)$; unlike *ALR*, the transformation can map a composition in the *D*-part Aitchison Simplex isometrically to a *D-1* dimensional Euclidean vector; yet, the covariance matrix is singular due to the zero-sum of the transformed vectors, and the orthogonal references in its subspace are not obtained in a straightforward manner.

- Isometric log-ratios (*ILR*) build *D-1* orthonormal coordinates for *D* parts in the composition, preserving all metric data properties and resulting in coordinates with a non-singular covariance matrix (Egozcue & Pawlowsky-Glahn, 2005). Transformed data can be analysed by all standard statistical methods; the only problem is the difficulty in the interpretation, since there is no one-to-one relationship between the original parts and transformed variables. *ILR* contrasts two subsets of parts J_1 and J_2 by defining the log ratio of their respective geometric means, with a scaling factor; the coordinates are called balances, with the form:

$$ILR_j = \sqrt{\frac{J_1+J_2}{J_1J_2}}\ln\frac{g_m(x+)}{g_m(x-)}$$ [3]

where J_1 and J_2 are the number of parts in the positive and negative groups, respectively.

In the time-use and active travel studies, we applied the following ILRs for data with five parts (travel active and non-active, contracted time, physical activities, and other sedentary time):

$$ILR_{Non-active-travel} = \sqrt{\frac{1}{2}}\ln\left(\frac{Other\ travel}{Active\ travel}\right); \qquad ILR_{Contracted} = \sqrt{\frac{2}{3}}\ln\left(\frac{Contracted\ time}{(\prod t_j)^{\frac{1}{2}}}\right);$$

$$ILR_{Other-sedentary} = \sqrt{\frac{3}{4}}\ln\left(\frac{Sedentary}{(\prod t_j)^{1/3}}\right); \quad ILR_{Physical-activity} = \sqrt{\frac{4}{5}}\ln\left(\frac{PA}{(\prod t_j)^{1/4}}\right)$$

For example, *ILR_Physical-activity* expresses the ratio of total time in daily physical activities, except for travel to the geometric mean of time in all other behaviours, while *ILR_Non-active-travel* considers the ratio of all travel that is non-active to active travel, i.e., the relative occurrence of the two types of behaviours within the total travel. Note that any other part can represent the 'starting' point for *ILR*s. The set presented above represents one of 105 possible contrasts for five parts—a data set of modest dimensionality (Figure A2). Thus, the interpretation must account for this aspect.
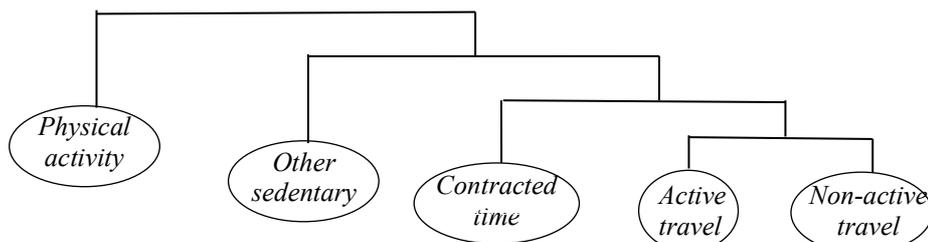

Figure A2: 'Dendrogram'-type representation of a combination of parts, based on the 'starting' point

The interpretation of *ADL* and *CLR* values is easier (e.g., negative *ADL* values show much shorter time than the reference category, and negative *CLR* values show shorter durations than the geometric mean for all

categories) than *ILR* values. Changes in active travel time will inevitably lead to changes in time for at least one other activity, but all *ILR*s will be modified. When used as exposure, model coefficients correspond to changes associated with increases in active travel relative to the compensatory decrease in other components.

Our time-use analysis of the Heathway project data used the following steps: normalising the geometric means of all activities/behaviours to add to 100%, deriving the *ILR*s, and applying MANOVA and other general linear model techniques to separate the effects of socio-demographics and circumstances in time allocation, health (MET, calories expenditure, heart rate) and self-reported wellbeing outcomes (enjoyment).

Consistent with prior scholarly work, inferential statistics and effect sizes depended on the approach taken, with more significant effects obtained with the isometric transformation measures than the standard approach. The analysis demonstrated the role of sensors in providing a more accurate account of activities.

Latent growth modelling (LGM)
Longitudinal data can yield invaluable information on trends in socio-economic and natural phenomena, as well as specific differences (across individuals, clusters, groups) in aspects of change over time. Initially developed in psychology (to model development of vocabulary in children) and then extended to education (e.g., growth of science achievements and attitudes) and health research (to evaluate clinical interventions), LGM is a broad class of statistical methods that can express an outcome as a function of a predictor variable that captures the passage of time (Bollen & Curran, 2006; Duncan & Duncan, 2009). LGM can examine intra-individual (within entity—either person, public transport station or train corridor/line in this project) changes over time and inter-individual (between entities—persons, destinations, zones) changes in individual trajectories. LGM is appealing due to its ability to model change and investigate antecedents and consequents of change, i.e., they represent a dynamic variation of structural equation models, SEM (Duncan & Duncan, 2009; McArdle, 2009).

For the patronage project, the models enabled us to reveal and describe the shape of the mean trend of patronage between 2009 and 2019 at various spatial levels and identify significant differences between stations, O-D pairs, zones/corridors in the shape of the trajectories, and to link variables systematically associated with changes in patronage over time (macroeconomic indicators, such as changes in economic growth/employment, land-use/transport investments, petrol price, fares, and travel distance). LGM can be estimated using two distinct approaches (multilevel and LGM–SEM), with specific benefits and limitations: multilevel models permit spatially invariant covariates (e.g., macroeconomic variables, such as CPI, unemployment) but are less flexible for functional form and cannot readily investigate antecedents and consequents of change.

Given its flexibility, the LGM–SEM structure was predominantly used in this project. Historically, this model is rooted in exploratory factor analysis and principal components (PCA) literature, as covariances between time periods are common elements of change or chronometric unobserved factors (Duncan & Duncan, 2009; McArdle, 2009). Consequently, the loadings may be interpreted as parameters representing the dependence of the repeated measures. The change is set as a series of loadings representing time in a linear, quadratic or other fashion.

Figure A3 presents a typical LGM–SEM model, including intercept *I*, slope *S* (modelled as latent variables) and some predictors. By SEM convention, circles/ovals represent latent variables, squares/rectangles represent measured variables (patronage over time), single-headed arrows represent regression weights, and double-headed arrows represent covariances.

I captures information about the starting point of patronage, representing the history, whereas *S* captures information about the rate of change over time. The depicted yearly model uses 11 indicators for the latent constructs (monthly model uses 132 indicators for change), and the time loadings (arrows from *S* to *y9 … y19*) delineate the strength and direction of an average pattern of change (i.e., linear, quadratic, cubic).
The model also includes one predictor (distance) affecting the rate of change (arrow from the square *dist* to the circle marked *S*) and one moderator (weekday), indicating that separate models are estimated for weekdays and Saturday, Sundays, and public holidays.

Loadings, means, and variances are interpreted to understand dynamics. Variances are particularly important because with no variance in the intercept, every line/route/O-D pair would have the same starting point, and

with no variance in the slope, every line/route/O-D pair would follow the same trajectory, which is not the case with the patronage data. Covariances between *I* and *S* are estimated to identify positive or negative correlations between starting points and acceleration/deceleration of the trajectories over time.
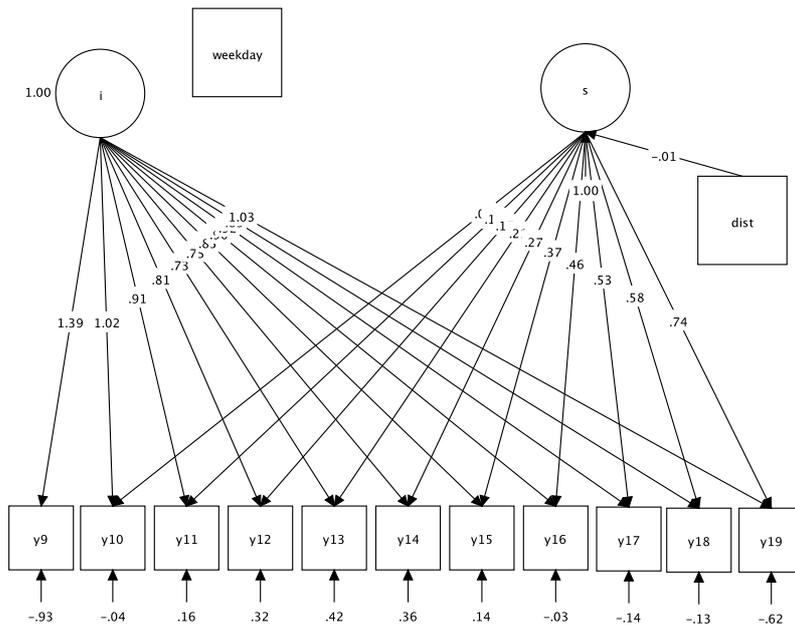

Figure A3: Example structure LGM model

More advanced models also included arrows between successive time periods to capture the autoregressive aspect of the time series. Models have also been estimated with free parameters or a pre-set functional form (linear, piecewise linear, quadratic trajectories).

Models for all types of journeys, separated by SmartRider user/ticket tupe, were estimated at the O-D pair level (40+ million records), station level (300,000 records) and the corridor or zone level. The best-fitting unconditional growth models for various parallel processes were combined into one model (+/– autoregressive effects) to examine whether the growth factors for predictors influenced the growth factors for patronage.

The results confirmed a piecewise trend over the 11 years, with a modest increase in patronage until 2013, decreasing from 2014–2016, then increasing slightly after 2016 (S-shape). In contrast, fares continuously increased over the 11 years, albeit at different rates and reflecting the distance gradient. Consistent with the literature on economic effects and travel mode choice, CPI and petrol price had a positive effect on overall patronage, whereas unemployment had a negative effect.

An interesting finding was the negative effect of distance on patronage, suggesting that shorter to moderate trips are preferred over longer trips requiring transfers, especially when the frequency of the services decreases during the off-peak period. Another significant effect was of changing the number of mining jobs on the rate of change for patronage, but shifting direction over time. Negative effects on the speed of developmental processes usually reflect plateauing effects, explained by the peak of the economic cycle/boom in this project. However, the reduction in mining activity after 2014 made the decrease in patronage more prominent.

In terms of pricing, the significant negative correlation between the intercept of patronage and the intercept of fares suggests that high fare levels are a deterrent for public transport patronage, confirming the long-term sign of price elasticities. Also of note is the significant positive correlation between the rate of change in patronage and the rate of fare change between 2009 and 2013, but negative correlation after 2014.

Finally, consistent with the distinct trajectories of various types of users (Standard, Student, Senior and Concession), separate sets of models were estimated using ticket type as the moderator, as the change processes could not be described accurately by a single trajectory (Wald test significant at 0.001 level confirmed the superiority of the free model). While Standard and Concession journeys decreased between 2013 and 2016, Student and Senior categories had modest increases. In the last three years of the time series, the patronage recovery was due to Standard riders, with all other categories recording negative slopes.