

New Data, New Directions: A Commentary on Emerging Big Geospatial Data for Population Research

Ngā Raraunga Hou, Ngā Aronga Hou: He kōrero mō te Raraunga Rarahi Mokowā ā-Nuku e Maiea ana mō te Rangahau Taupori

MALCOLM CAMPBELL*

Abstract

This commentary primarily discusses new data sources featuring fine-grained spatio-temporal data that have emerged in recent years from a variety of sources that previously did not exist or that were not easily accessible either from public or commercial entities. The focus here is on mobile phone location data (MPLD) as one avenue for potential new directions that have emerged and become increasingly relevant to a constellation of thematic areas across population research. I also discuss how new data sources, principally MPLD, aid in addressing some of the longstanding challenges and limitations of existing secondary data sources, while simultaneously creating both opportunities and limitations for researchers.

Furthermore, I discuss some salient potential applications and pitfalls of big geospatial data when these are applied to population research internationally, and also in Aotearoa New Zealand. It could be argued that there is an explicit challenge to the longstanding conceptualisation of static residence-based measures applied to a range of thematic areas across population research, such as understanding movement or migration. I further posit that it is timely to (re)consider the role that big geospatial data, specifically MPLD, plays in understanding central questions in both geography and demography, such as where, when, why and how do people move?

* Malcolm Campbell is an associate professor in quantitative human geography in the School of Earth and Environment, Faculty of Science at the University of Canterbury | Te Whare Wānanga o Waitaha. He is the director of the GeoHealth Laboratory | Te Taiwhenua o te Hauora at UC and the founder of the Regional Analytics Lab. Malcolm.Campbell@Canterbury.ac.nz

It can be demonstrated that there is a wealth of big geospatial data that now can be exploited leading to opportunities for better understanding of the dynamic processes related to people and places. It could be argued that it is time for a more fulsome engagement with MPLD and other big geospatial data sets and techniques to grapple with both the opportunities and challenges for understanding population dynamics, as well as the patterns and processes that give rise to inequalities and inequities between people and places in Aotearoa New Zealand and further afield.

Keywords: Aotearoa New Zealand, big geospatial data, population dynamics, GIS

Whakarāpopotonga

Ko te nuinga o ngā kōrero i roto nei e pā ana ki ngā puna raraunga hou e whakaatu ana i ngā momo raraunga mokowā ā-nuku kua puta mai i ēnei tau tata nei i te huhua o ngā mātāpuna kāore i reira i mua, kāore rānei i tareka māmātia i ngā hinonga tūmatanui, arumoni rānei. Kei te arohia nei ko ngā raraunga tauwāhi waea pūkoro (MPLD) hei ara tohu i ngā aronga hou kua puea mai, ā, e tipu mai ana tōna hāngai ki tētahi kāhuinga o ngā wāhi tāhuhu puta noa i te rangahau taupori. E matapakina ana hoki e au ngā āhua e whakatutuki ai ētahi puna raraunga hou, inarā te MLPD, i ngā wero me ngā whakatikinga tauroa o ngā puna raraunga tuarua o ēnei wā, ā, i taua wā tonu e hanga ana i ngā arawātea me ngā kōpiritanga mō te hunga rangahau.

I tua atu, e kōrero ana ahau mō ētahi whakahāngaitanga me ngā raruraru hira ina hoatu ērā ki te rangahau taupori puta noa i te ao, ā, ki Aotearoa nei anō hoki. Ka taea e te tangata te tohe kei reira he wero tūturu ki te whakaaturanga mauroa ā-ariā waitara o ngā inenga aro ki ngā wharenoho pateko ki te whānuitanga o ngā wāhi tāhuhu puta noa i te rangahau taupori, pērā i te noho mārama ki ngā nekeneke, ngā hekenga rānei a te tangata. Ko tāku e kī nei anō hoki, koia te wā tika ki te whai whakaaro (anō) ki te tūnga tō te raraunga mokowā ā-nuku rarahi i te mārama haere ki ngā pātai matua i te matawhenua me te hangapori, arā, ka nuku te tangata ki hea, āhea nuku ai, he aha i nuku ai, ā, ka pēhea te tangata e nuku ai?

Kāore e kore he nui rawa te raraunga mokowā ā-nuku rarahi e wātea ana ka taea te whakamahi e ara mai ai he arawātea e pai ake ai te māramatanga ki ngā tukanga hihiri e pā ana ki te tangata me te wāhi. Ka taea te tāpae te kōrero, kua tae mai te wā kia kaha ake te whai wāhi ki te MLPD me ētahi atu huinga me ngā tikanga raraunga mokowā ā-nuku rarahi ki te whakaanga ki ngā arawātea me ngā wero kia mārama ai ki ngā āhuatanga taupori, me ngā taurua me ngā tukanga e puta mai ai ngā manarite-kore me ngā tautika-kore i waenga i ngā tāngata me ngā wāhi i Aotearoa, i tāwāhi hoki.

Ngā kupu matua: Aotearoa, raraunga mokowā ā-nuku rarahi, ngā āhuatanga taupori, GIS.

This commentary primarily discusses new data sources featuring fine-grained individual spatio-temporal data that have emerged in recent years from a variety of sources that previously did not exist or that were not easily accessible either from public or commercial entities. Additionally, this commentary – which is situated within a special issue on geospatial techniques on population research – focuses on a specific data source, mobile phone location data (MPLD), as one critical avenue for potential new directions that have emerged. MPLD is likely to become increasingly relevant to a constellation of thematic areas across population research. I discuss how new data sources, principally MPLD, aid in addressing some of the longstanding challenges and limitations of existing secondary data sources, while simultaneously creating both opportunities and limitations for researchers.

Furthermore, the commentary discusses some salient potential applications and pitfalls of big geospatial data when applied to population research internationally, and also here in Aotearoa New Zealand. It could be argued that there is an explicit challenge to the longstanding conceptualisation of static residence-based measures applied to a range of thematic areas across population research, such as understanding movement or migration. I posit that it is timely to (re)consider the role that big geospatial data, specifically MPLD, play in understanding central questions in both geography (Christaller & Baskin, 1966) and demography, such as: where, when, why and how do people move? The main aim of this commentary is, therefore, to discuss MPLD and outline both the opportunities and challenges when it is applied to population research.

What is mobile phone location data?

Briefly, mobile phone location data (MPLD) can be defined as containing two primary components, as a minimum: location (e.g., latitude and longitude) and time. That is, MPLD is an example of *spatio-temporal data*. In addition, there can be aggregations or anonymisations of raw MPLD provided as a derived product, often by commercial entities. Examples of such products include Safegraph,¹ NEAR mobility,² and the Dataventures population density product in New Zealand, used during the COVID-19 response but since disestablished (Campbell, Marek, Wiki, et al., 2021). These MPLD are updated on a regular basis: for example, daily, weekly or monthly updates

of data. Some of the data are temporally sporadic, while other sources are curated into regular hourly, daily or other regularly sampled temporal intervals. Thus, MPLD is, or could be described as, a relatively straightforward and rudimentary spatio-temporal data source. The mobile phone in this instance is used as the platform for collection of individual spatio-temporal data, though I concede that terminology may evolve and progress. MPLD can also be linked to other sensed or imputed variables such as altitude or speed, which adds to the complexity and dimensionality of the data source. Furthermore, I suggest that 'big geospatial data' collected from mobile applications, such as social networks (e.g., Facebook, twitter/X, Instagram), tourism/travel (e.g., Airbnb, Flightradar24), navigation (e.g., Uber, Lime, Google Maps) or fast-food delivery (e.g., Uber Eats), can also be usefully categorised into the orbit of MPLD that could be useful for population research. This is primarily due to the same underlying structure of the data from mobile applications often containing the core components of location and time – though occasionally MPLD is linked to objects (e.g., taxis, scooters) rather than individuals. Such big geospatial data may include additional linked data features; for example, social, economic and demographic data. There is some debate around the ontological dimensions of big data (i.e., volume, velocity and variety (Kitchin & McArdle, 2016)) and the specific composition and meaning of big geospatial data, hence my aim to focus more specifically on MPLD.

After defining MPLD and outlining the key aim of this commentary, I move to outline the nature of the potential challenge or complementarity of MPLD with more conventional secondary sources of data for population research. Here, I seek to extend engagement with MPLD, or indeed other big geospatial data, as emergent data sources that, if properly understood, can significantly aid in understanding population trends and processes in almost real-time. Furthermore, if MPLD is derived from objects (e.g., planes, ships, taxis), we can discover and observe potential patterns and trends in the flows of both people and goods (Erlström et al., 2022).

Next, I point to the tension between the rudimentary aspects of MPLD, that when combined with volume, velocity and variety – the 3 Vs of big data – can add considerable complexity to the research process. Previous research has demonstrated that an absence of MPLD data had important consequences for the conceptualisation of place and produced significant misclassification (Campbell, Marek, Hobbs, et al., 2021), where the research

demonstrated that occupation and age of participants was an important determinant of exposure misclassification using MPLD. My commentary extends this ongoing debate by discussing how similar issues may apply to a broad range of population research topics. The census is considered the benchmark data set in terms of data analysis for population research and official statistics. Furthermore, the census is a critical indicator of potential bias in other emergent data sources such as MPLD, so it is necessary to understand the precise limitations of MPLD, and where and for whom bias may be present. A limitation is that as census data are collected at variable intervals internationally, intra-country comparison can be a challenge. Moreover, there is often a considerable lag between collection and publication as well as ongoing debates around the need for a census (Baffour et al., 2013; Boyle & Dorling, 2004) which are problematic. A very specific challenge is that certain phenomena, such as disasters or wars (Leasure et al., 2023), occur in close to real-time and as such, need a policy or governmental response matching the immediacy of the event. Population trends and processes are dynamic. It follows that more rapidly available and precise spatio-temporal data is critical in better understanding and responding to these challenges and that MPLD is one source of potential 'solutions'. To help further illuminate the importance of the emergent data sources discussed, I now thematically explore some of the myriad ways in which new data could, or perhaps should, lead to new directions in population research.

Selected thematic applications of mobile phone location data in population research

Demographic patterns and processes

First, U-turn the focus to how MPLD could address two key demographic processes: ageing and *fertility changes*. *Ageing*, the drift upwards in the median age of the population, is now well established in many economically developed countries – for example, in parts of Western Europe (Zheng et al., 2022) and in Japan (Higo, 2022) – but is also occurring globally. Ageing can lead to important changes in daily travel patterns or environmental impacts (Zheng et al., 2022), to frequency of health care access, and often changing preferences in residential location. Simultaneously, alongside ageing, the

number of children is declining, with profound potential changes predicted for the economy from modelled demographic projections (Liu & McKibbin, 2022). In Aotearoa New Zealand, and many places worldwide, demographic changes have the potential to alter how places are shaped as movement patterns can be fundamentally altered by family structures during periods of childhood and as we age (Yan et al., 2022). MPLD may give early indications as to the nature of changing movement patterns in populations, whether at work, education or play, and alone or in groups.

Interactions between environment and populations

Turning the focus to themes in the *environment* (built or natural), several overlapping salient issues could alter population trends and processes: trends in urbanisation, climate change and environmental impacts on population health and wellbeing. An important process that impacts unequally across places is the movement of populations to urban areas. More than 50 per cent of the world's population is now urbanised (United Nations, 2019), a proportion which is important in terms of how places are shaped, and which alters how populations interact with each other and travel through the environment. Understanding how populations live, work and play on a daily basis requires MPLD or fine spatio-temporal data to address research questions more accurately. Researchers may also utilise techniques such as agent-based models (Fazio et al., 2022) or spatial microsimulation (Campbell, 2011; Campbell & Ballas, 2016) in combination with MPLD to better understand and advance our understanding of such patterns. It is possible that movement by populations towards urban areas could have benefits for emissions, transport patterns and, therefore, climate change (Luqman et al., 2023). Linked to the urbanisation process is the challenge of environmental sustainability and climate change. That it is possible to link our precise travel behaviour, mode, duration and carbon emissions should afford new opportunities to better understand the impacts of travel behaviour on the environment. Moreover, as the environment changes, interactions between population health and wellbeing and the environment are also affected (Bowie et al., 2013). As previously argued, understanding the differences in exposure to salutogenic or harmful aspects of the environment requires precise spatio-temporal location information to determine exposure (Campbell, Marek, Hobbs, et al. 2021). It could be

argued that using MPLD affords new data and leads to new understandings of environmental considerations in population research. For example, we may be able to illuminate how travel changes alter environmental exposure(s), leading to changes in population health and wellbeing, while simultaneously being attentive and able to measure the unequal impacts created by differing levels of uptake of alternative transport modes, whether e-bike or electric vehicle.

Population movement patterns and processes

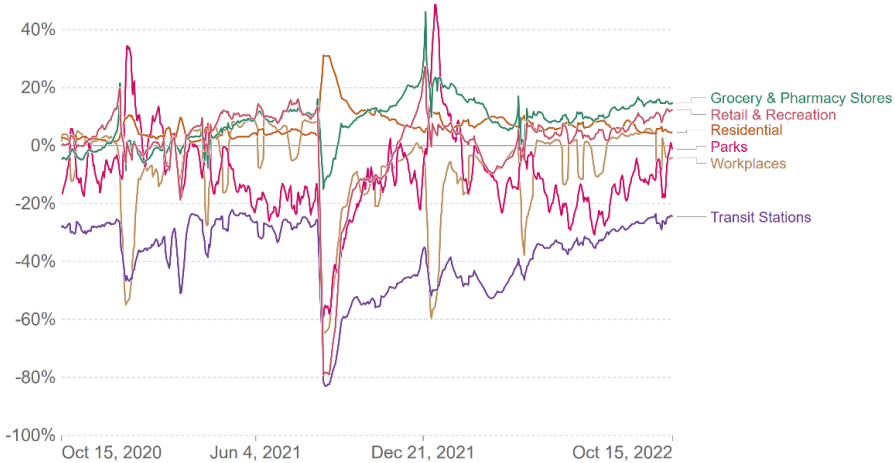
Arguably, the most intuitively applicable area of application for MPLD is in the area of understanding *daily movement(s)*, *travel behaviour(s)* and *migration*. First, there is a difference between longer- and shorter-term migration; for example, compare the movement of refugees or temporary movement of people following a disaster with longer-term more permanent and planned migration behaviour. However, even with more permanent migration, this change can occur at any point during the intercensal period. Data on long-term movement is collected – for example, through the IDI,³ rental bond data, or the electoral roll, to name a few data sources in New Zealand – allowing understanding of relocation during the intercensal period. However, MPLD can give a rapid assessment when such official sources of data lag, take time to release, or are unavailable (Deville et al., 2014). For example, the case of refugee movements (Alencar, 2020; Beine et al., 2021) can be understood using MPLD, highlighting specific routes and patterns, which allows humanitarian intervention and planning to occur. Refugee movements are often transnational, cyclical and unpredictable; in other words, unlikely to be captured by ‘official’ statistics. When disruptions to longer-term trends and patterns occur – for example, during disasters – this can have an immediate impact on movement by population(s), which is often unable to be detected using only administrative data for quite some time. Specifically, when ‘normal’ movements are temporality disrupted, such as changes observed during the COVID-19 pandemic (Campbell, Marek, Wiki, et al., 2021; Conrow et al., 2021). Temporary disruptions may also have unequal impacts across population groups or for particular places, destinations or environments (see Figure 1). As the climate changes, we may further see the confluence of climate change and refugees (Lu et al., 2016), creating movement unequally around the planet.

Figure 1: Example of daily movement and environmental interactions shift

How did the number of visitors change since the beginning of the pandemic?, New Zealand



This data shows how community movement in specific locations has changed relative to the period before the pandemic.



Source: Google COVID-19 Community Mobility Trends - Last updated 24 July 2023

Note: It's not recommended to compare levels across countries; local differences in categories could be misleading.

OurWorldInData.org/coronavirus • CC BY

In a similar vein, travel patterns that may alter on a daily basis – for example, commuting behaviour as roads are open, closed or altered – and data like these are also captured in a snapshot way. Thus, MPLD allows the detection of important trends in a much finer spatio-temporal way than the census can – and at a fraction of the cost. For example, MPLD could be used to investigate how remote working could be altering commuting patterns or to identify those who have multiple residences. In New Zealand, a subset of interest for movement patterns is tourism. There is a paucity of information on where tourists go, and when and how. Some select examples exist (Shoval & Ahas, 2016); for example, in transport (Buning & Lulla, 2021) or tracking of wine tourists (Lewis et al., 2021). And we also have both financial transactions and MPLD to give us a clearer picture, with much larger sample sizes. Studies that can estimate the whole of population are likely to be even more useful for economic development and infrastructure planning. They are also likely to give an early indication of shifts in established patterns that lead to ‘overtourism’. In New Zealand, tourism electronic card transactions (TECTs) data gave a reasonably rapid indication of tourism spend and where it is occurring, whether from domestic or international

sources. However, especially since the release of TECTs data has been recently discontinued, there is further room to utilise MPLD in this area and explore further fundamental questions on daily movement patterns across a range of applications and domains.

Understanding emerging innovation and disruption to population trends and processes

Data from innovative mobile platforms and applications (e.g., scooters, Uber, Airbnb) that impact on populations and places have been remarkably absent from official or administrative sources during their initial emergence. While understandable, it means that the specific social and spatial impacts on populations are often poorly understood for quite some time and based on ‘anec-data’ rather than an official register. I argue that MPLD is one avenue for grappling with understanding how innovations and disruptions can be understood promptly – whether from a new mobile application, social media platform or a pandemic. Adding MPLD to the population researchers’ toolkit may help in any response as a critic and conscience when, rather than if, the next trend or innovation arises.

Where to for population research?

A key aspect of big geospatial data is that there is a paucity of understanding of the nature and extent of bias. Furthermore, bias cannot be, or is not, well understood without a reference point such as the census. Future research should also consider the development of MPLD standards as an overlooked but important component of creating and maintaining social licence for the use of MLPD. I argue the census or administrative data is critical in terms of both understanding and correcting for any bias in MPLD. A key proposition is that MPLD can be deployed usefully to understand changing patterns and processes in the population, particularly where there is an intersection between thematic domains of application in population research as outlined above. In closing, four questions that could benefit from a more fulsome engagement with MPLD and other forms of big geospatial data are articulated.

- Where are people (currently) located?
- Where, when and how do people move?
- Where, when, how and with which environment(s) do people interact?
- Where, when, how and with which population(s) do people interact?

A key conclusion is that any of the questions outlined above could be more rapidly and precisely answered using big geospatial data such as MPLD.

Conclusion

It can be demonstrated that there is a wealth of big geospatial data, particularly MPLD, to be exploited, leading to opportunities for better understanding of key population processes impacting on both people and places. I argue that it is time for a more fulsome engagement with big geospatial data sets and techniques to understand both the opportunities and challenges such data may present for understanding population dynamics, as well as the patterns and processes that give rise to inequalities or inequities between people and places, both in Aotearoa New Zealand and further afield.

Notes

- 1 <https://www.safegraph.com/>
- 2 <https://near.com/platform/>
- 3 <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>

References

- Alencar, A. (2020). Mobile communication and refugees: An analytical review of academic literature. *Sociology Compass*, 14(8): e12802.
<https://doi.org/10.1111/soc4.12802>
- Baffour, B., King, T., & Valente, P. (2013). The modern census: Evolution, examples and evaluation. *International Statistical Review*, 81(3): 407–425.
<http://dx.doi.org/10.1111/insr.12036>
- Beine, M., Bertinelli, L., Cömertpay, R., Litina, A., & Maystadt, J-F. (2021). A gravity analysis of refugee mobility using mobile phone data. *Journal of Development Economics*, 150, Article 102618.
<https://doi.org/10.1016/j.jdeveco.2020.102618>

- Bowie, C., Beere, P., Griffin, E., Campbell, M., & Kingham, S. (2013). Variation in health and social equity in the spaces where we live: A review of previous literature from the GeoHealth Laboratory. *New Zealand Sociology* 28(3): 164–191. <https://core.ac.uk/download/pdf/56364141.pdf>
- Boyle, P., & Dorling, D. (2004). Guest editorial: The 2001 UK census: Remarkable resource or bygone legacy of the ‘pencil and paper era’? *Area*, 36(2), 101–110. <https://doi.org/10.1111/j.0004-0894.2004.00207.x>
- Buning, R. J., & Lulla, V. (2021). Visitor bikeshare usage: Tracking visitor spatiotemporal behavior using big data. *Journal of Sustainable Tourism* 29(4), 711–731. <https://doi.org/10.1080/09669582.2020.1825456>
- Campbell, M. (2011). *Exploring the social and spatial inequalities of ill-health in Scotland: A spatial microsimulation approach* [PhD thesis, University of Sheffield]. White Rose eTheses Online. <https://etheses.whiterose.ac.uk/1942/>
- Campbell, M., & Ballas, D. (2016). SimAlba: A spatial microsimulation approach to the analysis of health inequalities. *Frontiers in Public Health*, 4(230). <https://doi.org/10.3389/fpubh.2016.00230>
- Campbell, M., Marek, L., & Hobbs, M. (2021). Reconsidering movement and exposure: Towards a more dynamic health geography. *Geography Compass* 15(6): e12566. <https://doi.org/10.1111/gec3.12566>
- Campbell, M., Marek, L., Wiki, M. Hobbs, M. Sabel, C. E., McCarthy, J., & Kingham, S. (2021). National movement patterns during the COVID-19 pandemic in New Zealand: The unexplored role of neighbourhood deprivation. *Journal of Epidemiology and Community Health*, 75(9), Article 903. <https://jech.bmj.com/content/75/9/903>
- Christaller, W., & Baskin, C. W. (1966). *Central places in southern Germany* (C. W. Baskin, Trans.). Prentice Hall.
- Conrow, L., Campbell, M., & Kingham, S. (2021). Transport changes and COVID-19: From present impacts to future possibilities. *New Zealand Geographer* 77(3), 185–190. <https://doi.org/10.1111/nzg.12315>
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* 111(45): 15888–15893. <https://doi.org/10.1073/pnas.1408439111>
- Erlström, A., Grillitsch, M., & Hall, O. (2022). The geography of connectivity: A review of mobile positioning data for economic geography. *Journal of Geographical Systems*, 24(4): 679–707. <https://doi.org/10.1007/s10109-022-00388-4>
- Fazio, M., Pluchino, A., Inturri, G., Le Pira, M., Giuffrida, N., & Ignaccolo, M. (2022). Exploring the impact of mobility restrictions on the COVID-19 spreading through an agent-based approach. *Journal of Transport & Health*, 25, Article 101373. <https://doi.org/10.1016/j.jth.2022.101373>
- Higo, M. (2022). A “death-laden society”: The next stage of a hyper-aged Japan and health challenges ahead. *Ageing and Health Research*, 2(4), Article 100110. <https://doi.org/10.1016/j.ahr.2022.100110>

- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), Article 2053951716631130. <https://doi.org/10.1177/2053951716631130>
- Leasure, D. R., Kashyap, R., Rampazzo, F., Dooley, C. A., Elbers, B., Bondarenko, M., Verhagen, M., Frey, A., Yan, J., Akimova, E. T., Fatehkhia, M., Trigwell, R., Tatem, A. J., Weber, I., & Mills, M. C. (2023). Nowcasting daily population displacement in Ukraine through social media advertising data. *Population and Development Review*, 49(2), 231–254. <https://doi.org/10.1111/padr.12558>
- Lewis, G. K., Hardy, A., Wells, M. P., & Kerslake, F. L. (2021). Using mobile technology to track wine tourists. *Annals of Tourism Research Empirical Insights*, 2(2), Article 100022. <https://doi.org/10.1016/j.annale.2021.100022>
- Liu, W., & McKibbin, W. (2022). Global macroeconomic impacts of demographic change. *The World Economy*, 45(3), 914–942. <https://doi.org/10.1111/twec.13166>
- Lu, X., Wrathall, D. J., Sundsøy, P. R., Nadiruzzaman, M., Wetter, E., Iqbal, A., Qureshi, T., Tatem, A., Canright, G., Engø-Monsen, K., & Bengtsson, L. (2016). Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change*, 38, 1–7. <https://doi.org/10.1016/j.gloenvcha.2016.02.002>
- Luqman, M., Rayner, P. J., & Gurney, K. R. (2023). On the impact of urbanisation on CO2 emissions. *npj Urban Sustainability*, 3(1): Article 6. <https://doi.org/10.1038/s42949-023-00084-2>
- Shoval, N., & Ahas R. (2016). The use of tracking technologies in tourism research: the first decade. *Tourism Geographies*, 18(5), 587–606. <https://doi.org/10.1080/14616688.2016.1214977>
- United Nations. (2019). *World Urbanization Prospects: The 2018 Revision* (ST/ESA/SER.A/420). Department of Economic and Social Affairs, Population Division, United Nations. <https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf>
- Yang, Y., Sasaki, K., Cheng, L., & Tao, S. (2022). Does the built environment matter for active travel among older adults: Insights from Chiba City, Japan. *Journal of Transport Geography*, 101, Article 103338. <https://doi.org/10.1016/j.jtrangeo.2022.103338>
- Zheng, H., Long, Y., Wood, R., Moran, D., Zhang, Z., Meng, J., Feng, K., Hertwich, E., & Guan, D. (2022). Ageing society in developed countries challenges carbon mitigation. *Nature Climate Change*, 12(3), 241–248. <https://www.nature.com/articles/s41558-022-01302-y>