Extracting Key Urban Footfall Signatures using Principal Component Analysis

Jack Liddle[1], Wenhua Jiang[2] and Nick Malleson*[1, 2]

[1]School of Geography, University of Leeds, Leeds, UK
[2]Alan Turing Institute, London, UK

February 7, 2025

Summary

This study applies Principal Component Analysis (PCA) to pedestrian count data from Melbourne, Australia, to extract core temporal 'signatures' that underpin aggregate urban footfall patterns. By reducing complex, noisy data into a small number of interpretable components, our analysis sheds light on how activity patterns have changed (and continue to change) in various neighbourhoods following the immense disruption caused by COVID-19. Our findings highlight the utility of PCA as a tool for systematic, data-driven analysis of pedestrian mobility.

KEYWORDS: Principal Component Analysis; Footfall; Urban Dynamics; Geographic Information Science.

1 Introduction

As urbanization continues to accelerate – an estimated 68% of the global population expected to reside in cities by 2050 (United Nations 2018) – understanding pedestrian dynamics has become increasingly important. The role of pedestrian activity in fostering vibrant urban environments has long been recognised (Jacobs 1961), yet only recently has the advent of the "golden age of data" (Arribas-Bel & Tranos 2018) facilitated large-scale quantitative studies of pedestrian movement. This shift has also highlighted the necessity of data-driven empirical evidence (Philp et al. 2022) to inform urban development.

Despite this, relatively little focus has been given to the *temporal signatures* that arise from pedestrian activity in different locations. These signatures, which highlight fluctuations in the number of pedestrians present in a given place over a specific time period, offer insights into how the built environment is utilised by citizens over varying time scales. For example, consider Figure 1 that illustrates a hypothetical temporal footfall pattern for an urban location over the course of a week. Qualitatively, this patten appears to be driven by traditional '9–5' employment so predominantly

*Corresponding author n.s.malleson@leeds.ac.uk. ORCID: 0000-0002-6977-0615

consists of a "commuting" signature. However, there may be other signatures that are obscured by the large numbers of commuters (e.g. limited lunchtime pedestrians?) and so the purpose of this work is to unpack this single signal into a number of distinct components ('signatures').

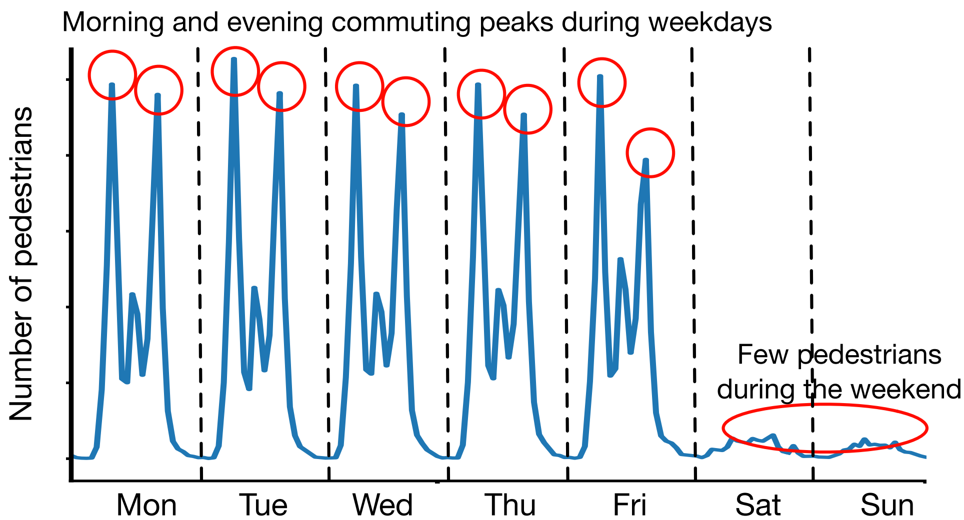


Figure 1: Hypothetical hourly footfall pattern for an urban location over a week.

In an attempt to better understand these signatures, this study applies Principal Component Analysis (PCA) to extensive pedestrian count data from Melbourne, Australia. We demonstrate that PCA effectively reduces complex, noisy pedestrian count data into a small number of interpretable components that reveal distinct, recurring trends. Specifically, we aim to (i) examine the interpretability of principal components and their correspondence with known behavioural trends and (ii) investigate how urban footfall dynamics evolve over time, particularly in response to external shocks such as COVID-19. Our findings highlight the utility of PCA as a tool for systematic, data-driven analysis of pedestrian mobility, with implications for urban planning, transport modelling, and policy development.

2 Previous Work

Quantifying pedestrian footfall is crucial for applications as diverse as urban planning (Cooper et al. 2021), environmental health (Park & Kwan 2017), and public safety (Malleson & Andresen 2016, Tucker et al. 2021). Despite the emergence of a growing literature on "ambient" (Whipp et al. 2021) or "day time" (Boeing 2018) populations – commuters, shoppers, tourists, etc – much less attention has been paid to the fine-grained temporal signatures of pedestrian activity. New data sources, such as telecommunications (Bogomolov et al. 2014), street view images (Chen et al. 2022), social media (Malleson & Andresen 2015, Liu et al. 2022), and pedestrian sensors (Crols & Malleson 2019, Philp et al. 2022), have enabled more detailed footfall analysis, and yet the core temporal components that underpin footfall counts remain unexplored.

PCA is a powerful tool for extracting these patterns, and yet one that is almost entirely absent from research in this area. The closest study is that of Kim (2020), which uses cell phone activity counts (aggregated to a $50m^2$ grid, hourly) as a proxy for pedestrian dynamics, applying functional PCA to explore urban vitality. While some components, like the "3 peaks" align with ours, key differences exis, possibly due to only two components being presented by Kim. This study extends previous work by systematically uncovering urban footfall patterns using PCA.

3 Methodology

We leverage pedestrian count data from 94 sensors across Melbourne, Australia, each capturing hourly footfall (the total number of pedestrians per hour who pass the sensor). All data are publicly available from the Melbourne Open Data Portal[1]. Sensors have come online gradually and some have been recording as far back as 2010. We use data from 1st Jan 2018 to 31st December 2019 to draw on the largest number of sensors possible without having the components impacted by COVID-19. Our data are structured such that each row represents the footfall counts over a sensor-day combination, with 24 columns to represent the counts of pedestrians for each hour in that day. PCA is applied to these time-series data and the output is a series of 24-item vectors (each one a principal component) which represent the core underlying patterns that underpin the aggregate sensor-day footfall patterns. The components are organised such that the first explains most of the variation in the points, with latter components explaining iteratively less variation. The 'explained variance ratio' demonstrates that with only the first three components it is possible to explain 95% of the variation in the daily sensor data, so these are the three components that we predominantly focus on in our analysis.

The source code for the analysis is available in full from the project GitHub repository[2]

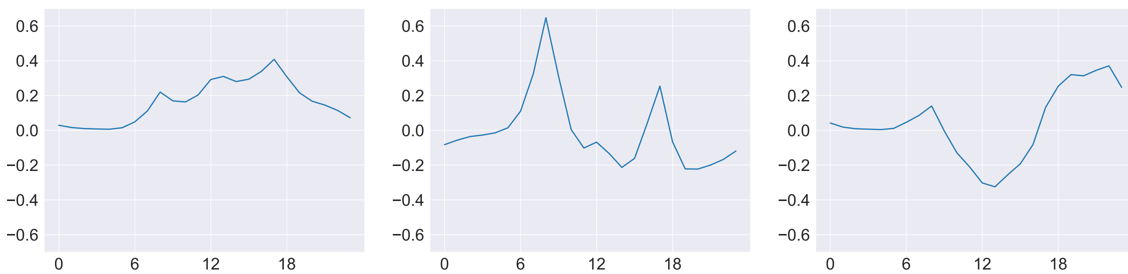
4 Results and Discussion

4.1 Interpreting the Principal Components

The first intriguing result is that the first three components appear to be representative of distinguishable aspects of urban dynamics. Although this is not entirely unexpected, we were surprised that some exhibited such interpretable patterns. Figure 2 illustrates the shapes of those components. We also assign our own labels as the components appear to represent (i) general 'busyness' (very similar to the mean footfall pattern, so indicative of how busy a 'typical' area is); (ii) 'commuter rush' (peaks in the morning and afternoon are reminiscent of typical '9-5' commuting behaviour); (iii) 'lunchtime suppression' (reflects a dip in footfall during lunchtime).

[1] `https://www.melbourne.vic.gov.au/open-data`

[2] `https://github.com/nickmalleson/melbourne-timeseries/blob/main/03d%20PCA%20Analysis%20-%20As%20a%20timeseries-CUPUM.ipynb`



(a) Component 1: 'busyness'

(b) Component 2: 'commuter rush'

(c) Component 3: 'lunchtime suppression'

Figure 2: The first three principal components; together they explain 95% of the variance in the footfall data

4.2 Evolution of Urban Usage Patterns

Beyond identifying static behavioural signatures, PCA also enables an analysis of how pedestrian dynamics evolve over time. To reconstruct the footfall during an individual sensor-day, 'loadings' (weights) are applied to the principal components and these loadings demonstrate how much each component contributes to the observed footfall pattern in that place and time. For example, if a particular sensor-day exhibited commuting patterns but relatively small numbers of people, it would have a negative weight for component 1 (busyness) and a positive weight for component 2 (commuter rush).

Figure 3 examines the loadings for two location across a longer time period to explore potential changes in how residents use their neighbourhoods, particularly after the COVID-19 pandemic. Both areas experience less than average footfall, hence component 1 ('busyness') is negative. In both cases we also see busyness drop considerably at the start of 2020, caused by the start of state-wide lockdown policies. In the first area, Collins Place, not only do we see footfall start to recover but we also see a recovery in component 2 ('commuter rush') which suggests that the location continues to be largely used as a commuting hub, albeit with lower overall volume. Victoria point, on the other hand, might be undergoing wider changes. Footfall was largely driven by commuting before the pandemic, but following March 2020 component 2 ('commuter rush') has yet to recover with component 3 ('lunchtime suppression') becoming better able to explain the variation in footfall counts. This suggests that footfall in the region may no longer be dominated by footfall; potentially indicative of a wider change of use.

5 Conclusions

Whilst further, more focussed, analysis is necessary before drawing firm conclusions about changes in footfall in these neighbourhoods, the findings illustrate how PCA can serve as a diagnostic tool for assessing long-term changes in urban activity patterns. By systematically quantifying shifts in pedestrian behaviour, planners and policymakers can gain insight into the evolving demands on different urban spaces, informing targeted interventions to support economic recovery and urban

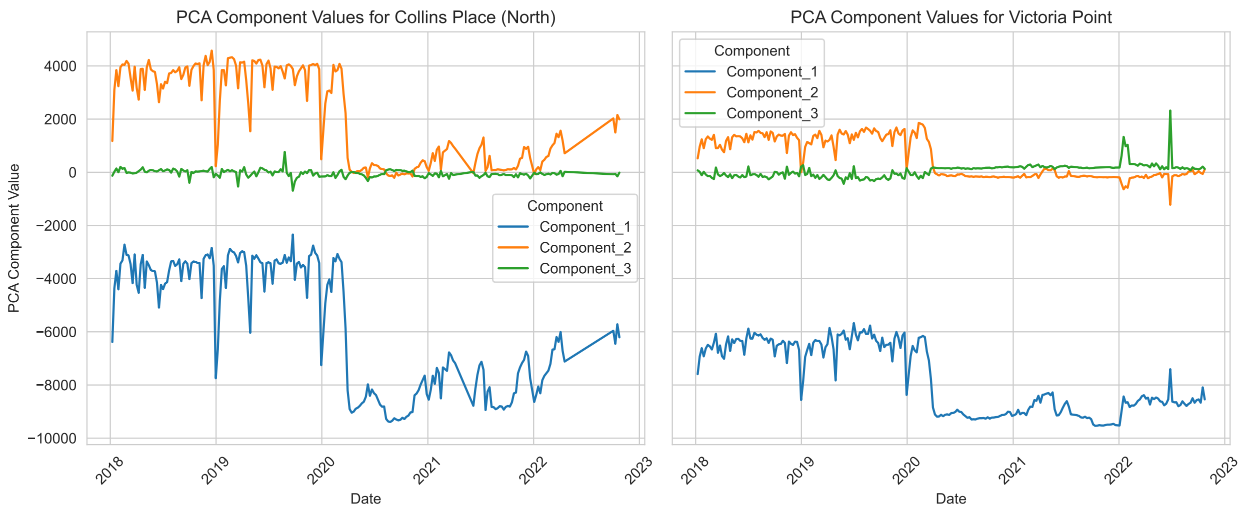


Figure 3: Change in component loadings over time in two locations

resilience.

Future research will expand this approach to multiple cities to assess the generalisability of the identified components and to leverage additional data sources. Additionally, integrating PCA with other machine learning techniques may offer more sophisticated tools for modelling pedestrian behaviour.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 757455).

References

Arribas-Bel, D. & Tranos, E. (2018), 'Big Urban Data: Challenges and Opportunities for Geographical Analysis', *Geographical Analysis* **50**(2), 123–124.

Boeing, G. (2018), 'Estimating local daytime population density from census and payroll data', *Regional Studies, Regional Science* **5**(1), 179–182.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F. & Pentland, A. (2014), Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data, *in* 'ICMI '14 Proceedings of the 16th International Conference on Multimodal Interaction', ACM, New York, NY, USA, pp. 427–434.

Chen, L., Lu, Y., Ye, Y., Xiao, Y. & Yang, L. (2022), 'Examining the association between the built environment and pedestrian volume using street view images', *Cities* **127**, 103734.

Cooper, C. H., Harvey, I., Orford, S. & Chiaradia, A. J. (2021), 'Using multiple hybrid spatial design network analysis to predict longitudinal effect of a major city centre redevelopment on pedestrian flows', *Transportation* **48**(2), 643–672.

Crols, T. & Malleson, N. (2019), 'Quantifying the ambient population using hourly population footfall data and an agent-based model of daily mobility', *GeoInformatica* .

Elhaik, E. (2022), 'Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated', *Scientific Reports* **12**(1), 14683.

Jacobs, J. (1961), *The Death and Life of Great American Cities*, Random House, New York.

Kim, Y.-L. (2020), 'Data-driven approach to characterize urban vitality: How spatiotemporal context dynamically defines Seoul's nighttime', *International Journal of Geographical Information Science* **34**(6), 1235–1256.

Liu, L., Lan, M., Eck, J. E., Yang, B. & Zhou, H. (2022), 'Assessing the Intraday Variation of the Spillover Effect of Tweets-Derived Ambient Population on Crime', *Social Science Computer Review* **40**(2), 512–533.

Malleson, N. & Andresen, M. A. (2015), 'The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns', *Cartography and Geographic Information Science* **42**(2), 112–121.

Malleson, N. & Andresen, M. A. (2016), 'Exploring the impact of ambient population measures on London crime hotspots', *Journal of Criminal Justice* **46**, 52–63.

Park, Y. M. & Kwan, M.-P. (2017), 'Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored', *Health & Place* **43**, 85–94.

Philp, S., Dolega, L., Singleton, A. & Green, M. (2022), 'Archetypes of Footfall Context: Quantifying Temporal Variations in Retail Footfall in relation to Micro-Location Characteristics', *Applied Spatial Analysis and Policy* **15**(1), 161–187.

Tucker, R., O'Brien, D. T., Ciomek, A., Castro, E., Wang, Q. & Phillips, N. E. (2021), 'Who 'tweets' where and when, and how does it help understand crime rates at places? Measuring the presence of tourists and commuters in ambient populations', *Journal of Quantitative Criminology* **37**, 333–359.

United Nations (2018), *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*, United Nations, Department of Economic and Social Affairs, Population Division, New York.

Whipp, A., Malleson, N., Ward, J. & Heppenstall, A. (2021), 'Estimates of the Ambient Population: Assessing the Utility of Conventional and Novel Data Sources', *ISPRS International Journal of Geo-Information* **10**(3), 131.