ITM Web of Conferences **88**, 01006 (2026) https://doi.org/10.1051/itmconf/20268801006
AIBDCS 2026

Short-Term Forecasting and Anomaly Detection of Melbourne CBD Pedestrian Flows Using Hourly Sensor Counts

Zemu Wang*

School of Computing and Information Systems, Faculty of Science, The University of Melbourne, Melbourne, Australia

Abstract. This research aims to analyze the open-source pedestrian counts from the city of Melbourne to identify temporal patterns, forecastability, and anomalous pedestrian activity within the Melbourne central business district (CBD). This research utilizes hourly pedestrian counts from 23rd January 2024 to 22nd January 2026, which then form the basis for city-level time series indicators. Three main analytical approaches will be adopted for the purpose of this research: a weekly diurnal heatmap, a series representing the total pedestrians per day, and an hourly anomalous detection using a z-score method. A simple model will also be adopted to identify forecastability. The results clearly indicate the presence of temporal patterns within the data set, including strong intraday rhythms and notable differences between weekdays and weekends. The simple model was effective in capturing the main temporal patterns observed. It had good overall forecastability for the test set. The results from the anomaly detection highlighted a small number of extreme deviations in the data. These deviations were mainly positive short-term spikes. It is evident that the results demonstrate that time series analysis is a viable tool for understanding normal and abnormal pedestrian activity in the Melbourne CBD.

1 Introduction

The activity of pedestrians within the central business district (CBD) is viewed as a significant signifier of urban functionality, and utilization of urban spaces. Pedestrian activity patterns may have effects on crowd control, transport, and the retail trade. There are excessive number of pedestrians which may indicate the high crowd density and also service demands. A significantly reduced pedestrian traffic might indicate the presence of atypical conditions and poor data quality in general. As pedestrian activity can undergo significant changes in a short time, it is essential to understand both normal and unusual trends.

The earlier work has indicated that the tendencies of pedestrian displacement across the city follow considerable temporal trends instead of random fluctuations. Standard patterns are usually made up of daily commutes, weekday vs. weekend variation, along with particular forms of activity in the city centre. Pedestrian sensor data collected during pedestrian activity in Melbourne has led to the conclusion that spatiotemporal analysis could be used to make great contributions to the level of intensity and demand of activities in the central business district [1,2]. Simultaneously, the trends of pedestrians are influenced by the external factors such as weather and special events that may result in substantial changes in the short run [3]. Those findings have indicated that atypical pedestrian behavior should not be perceived purely as statistically anomalous, but as possibly meaningful.

The current study uses the openly accessible data on hourly pedestrian count recorded by the City of Melbourne that will help to analyse the pedestrian count in the Melbourne Central Business District. This study has two primary goals, namely identifying regular trends in pedestrian count through aggregated time series indicators and visualizations, and identifying anomalous pedestrian count and its prediction accuracy based on the temporal data. By combining descriptive analysis, anomaly detection and forecasting, the research aims at enhancing the comprehension of the circumstances under which the pedestrian counts do not behave as expected and the implications of such occurrences to the city.

2 Methods

2.1 Data sources

The research is based on the publicly accessible pedestrian count data of the City of Melbourne Pedestrian Counting System [4]. This paper will use an hour-by-hour pedestrian count dataset that measures the flow of pedestrians at specific locations within the Melbourne Central Business District (CBD). Per sensor per hour, the dataset also has the counting date, hour of the day, counts of people moving in two different directions. Total counts of the pedestrians in every sensor during any given hour of the day were obtained by adding together the numbers of pedestrians going in both directions.

* Corresponding author: zemuw@student.unimelb.edu.au

The supplementary data that was used to enable a more efficient interpretation of the abnormal data by providing sensor identifiers and coordinates has also been used to enhance interpretation [5]. The research is conducted within 23 January 2024 - 22 January 2026 and it contains 1,451,812 records of 100 sensors. Because the research involves looking at the trends of time on the level of the city, the original data at the sensor level have been summarized into various indicators.

2.2 Indicator construction

To understand the volume of the people on the move in the city, three aggregate indicators have been made with the help of hourly sensor data.

Initially, an hourly total time series was formed by grouping together all the data on the number of pedestrians walking by every sensor at every hour. The indicators of this kind preserve the short term dynamics in the data, which is helpful when detecting anomalies in the data as well as when forecasting.

The second step is that a daily total time series was formed by summing up all the hour-by-hour time series data on each calendar day. It will be easier to comprehend the long term trends present in the data by the fact that there is less short-term volatility in the data when it is aggregated into daily totals compared with the hourly total time series.

In conclusion, a mean hourly time series by day of the week was built through the summation of all the combinations of day of the week and hour of the day. Such an indicator serves well in the understanding of the cyclical nature of the week in the data (morning and evening peaks, activity during lunch, differences between the working days and the weekend).

These kinds of indicators were selected as they summarize the degree of pedestrian movement at different time scales.

2.3 Data processing and visualization

This was done before analysis by standardizing the temporal fields so that every record has its own hourly timestamp. Those records which were not providing valid time data were excluded as they cannot be used in a time-series aggregation. Hourly pedestrian counts are therefore aggregated across the sensors to present the city level data in the next step of analysis.

Visual analysis was one of the important pieces of descriptive analysis in the study. Line charts gave the general line of the time of pedestrians and the heatmap gave the average line of the number of pedestrians in every hour of the day and a weekday.

2.4 Short-term forecasting (baseline)

A simple seasonal baseline model was applied to estimate the predictable behavior of pedestrians in the short-term periods. To achieve that, the series with the hourly city-wide totals were divided into two sets; a training set made up of the initial 80 percent of the available data and a test set of the rest of the 20 percent. The split ensures that the data is kept in order throughout the time and avoids leaking information between the future data and the model.

The prediction model computes each test case the pedestrian volume by taking the mean hourly reading recorded in the training phase on the individual day of the week and the time of the day. To be more precise, the forecast model takes it as granted that the pedestrian flow follows the weekly cycle and it uses the average value in the particular day and hour of the day. The rationale behind choosing the average value is that this is a common practice to apply such simple methods in time-series forecasting issues, and the walking behavior of people within central business district (CBD) zones generally exhibits high predictability with respect to the weekly trends. This method is also straightforward and easy to comprehend and can therefore be applied in solving the issue at hand [6, 7].

To measure the precision of the forecast, three usual measures were employed - namely, mean absolute error (MAE), root mean squared error (RMSE) and the coefficient of determination (R^2). Measurement of MAE is based on the average forecast error whereas the RMSE puts greater weight on bigger errors. The measurement of R^2 is the ratio of variation in the series which is explained by the model. Such three metrics offer a more comprehensive image of the performance of the selected model in short-term predictions of the series [6].

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}i| \ \boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}i)^2} \quad (1)$$

where y_i is the observed hourly total at index i and $\hat{y}_i$ is the corresponding forecast.

2.5 Anomaly identification

Identifying abnormal pedestrian activity on the city-wide hourly total series is done with a use of a z-score-based technique. The z-score of every hourly observation was calculated in the following manner:

$$z_t = \frac{x_t - \mu}{\sigma} \quad (2)$$

where x_t is the hourly total count at time t, and μ and σ are the average and standard deviation of the hourly total series.

This technique relies on the fact that standardized deviation of every hourly measurement against mean can be used to easily identify those hours, where pedestrian activity is either very high, which is unusual. In this research anomaly screening was performed using the maximum absolute z-scores. It implies that both unusually high and low pedestrian volumes were viewed as possible anomalies. The z-score method can be used as the first-stage screening of anomalies particularly under the condition the interpretability and transparency of the method are the first priority concerns [8]. The z-score method is uncomplicated and works effectively to detect extreme temporal anomalies within a combined urban time series.

This section alone provides the procedure to detect them. In the results section both ranked anomalous hours are displayed as well as their corresponding z-scores.

3 Results and Discussion

3.1 Forecasting performance

The baseline forecasting model captures the fundamental time trends of pedestrians in the centre of Melbourne business sector. The baseline forecasting model obtained an MAE of 5, 459.60, RMSE of 8,819.36, and R^2 of 0.8845 on the test set. It indicates that there is a large percentage of variation in hourly pedestrian being explained by the baseline forecasting model. These results indicate that pedestrians have very strong weekly time dynamics. As per this study, the fact is consistent with findings of other researchers who reported the presence of spatiotemporal rhythms of pedestrians in Melbourne [1].

The comparison of the observed and predicted hourly total pedestrians in the test period is illustrated in Fig. 1. It is evident that the cyclical trend of the forecasted line is consistent with the overall trend of the observed data. It is also evident that the general rise-fall trend has been adequately depicted by the forecasted line. But one can see that the peaks in the observed data are also underestimated at times. This is expected in performance of a simple benchmark model [6]. The difference between the forecasted line and the observed data may be explained by the different temporary effects, such as special events, weather conditions, etc. [3].

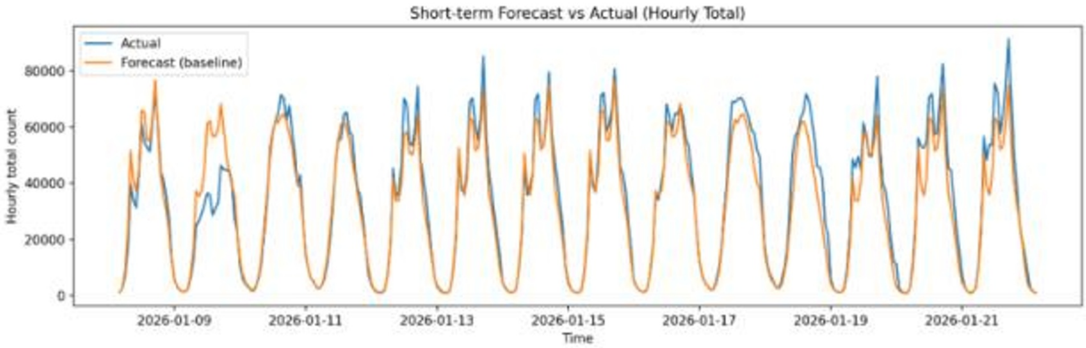

Fig. 1. The difference between short term estimation and real hourly total pedestrian number in the test period (Photo/Picture credit: Original).

3.2 Weekly–diurnal structure

Fig. 2 shows the average pedestrian counts per hour of day and day of week. Diurnal variation is well present across the week. Early morning and late night hours have fewer pedestrians, which increase on the arrival of day, and persist until daytime and early evening hours. Differences between weekdays and weekends in the number of pedestrians are also observed. On weekdays, the pedestrian counts reveal that there is a close concentration of numbers at the so-called daytime hours - the time related to work activity and business centers. In case of weekdays, the pedestrian counts have even distributions of numbers at midday and evening hours, which relate to non-work activity. The daily variation of pedestrian counts is also mostly consistent with earlier documented patterns of pedestrian movements within the city of Melbourne [1, 9].

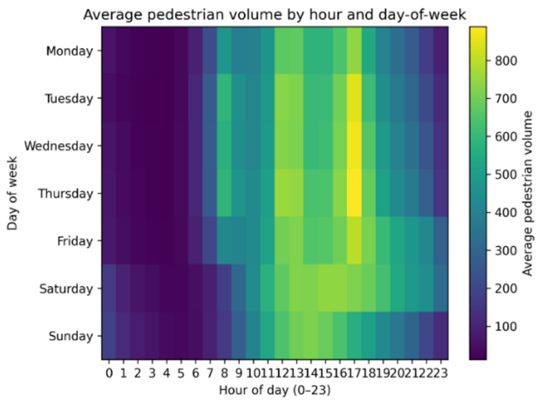

Fig. 2. Mean pedestrian count per hour of the day and day of the week averaged over all sensors (Photo/Picture credit: Original).

3.3 Day-level variation

Fig. 3 shows the overall volume of pedestrians measured by all sensors combined. The aggregated data is also simpler to interpret than the hourly data. Much of the dates fall in a steady interval. It means that the system in the city can handle an equivalent amount of traffic. But there are some days on which pedestrian numbers are small.

Such low volumes of people walking may be due to actual traffic on the CBD. Nevertheless, the low volume can also be due to lack of sensor coverage [10]. These dates should thus be handled with a degree of caution. However, the combined results remain an efficient source of finding out dates to analyze further, particularly when it comes to the comprehension of citywide interruptions, or abrupt shifts in pedestrian demand [3].

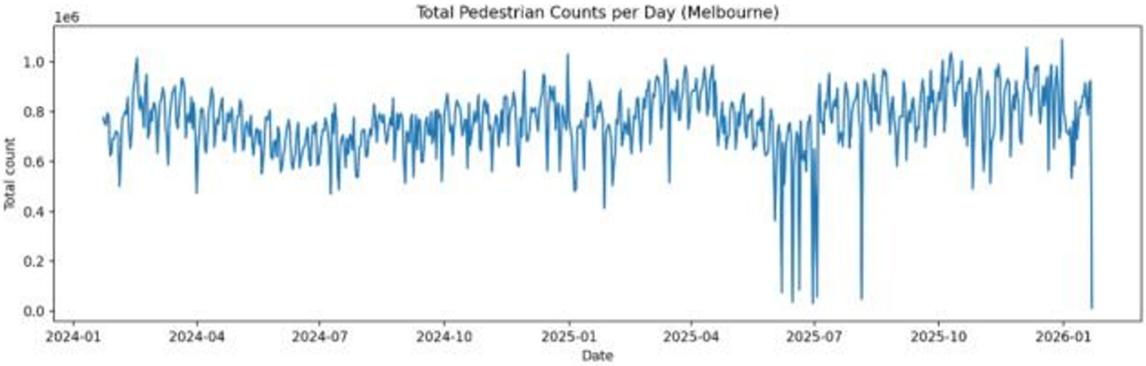

Fig. 3. Total daily number of pedestrians per sensor in the system (Photo/Picture credit: Original).

3.4 Extreme hourly deviations

In order to detect abnormal pedestrian behavior, these hourly counts were standardized by using z-score standardization and then ordered according to the amount of their difference above the overall average of all hourly pedestrian counts [8]. Even though the anomaly detection algorithm is intended to give more importance to the highest ranked anomalies due to the absolute value of the calculated z-score, the highest ranked anomalies in the present dataset are largely positive spikes rather than negative ones. It shows that the most extreme deviations in the overall hourly counts are primarily made up of short intervals of extremely high pedestrian counts.

The hourly city-wide pedestrian counts are shown along with the highlighted anomalous counts as indicated in Fig. 4. Most of the known anomalies include a sharp increase in the number of people walking. This implies that the anomalies represent brief intervals of increased pedestrian movement. In a central business district (CBD), these abrupt population surges of pedestrians can be caused by the movement of crowds and other mass events or by a synchronized trend of arrivals and departures [3]. Whereas most of the statistics peaks tend to reflect real-life occurrences, there is no way to compare the statistically identified anomalies with the real-life happenings. Certain of the recognized anomalies might be associated with the quality of data that occurs due to the acquired data (Table 1) [6, 10].

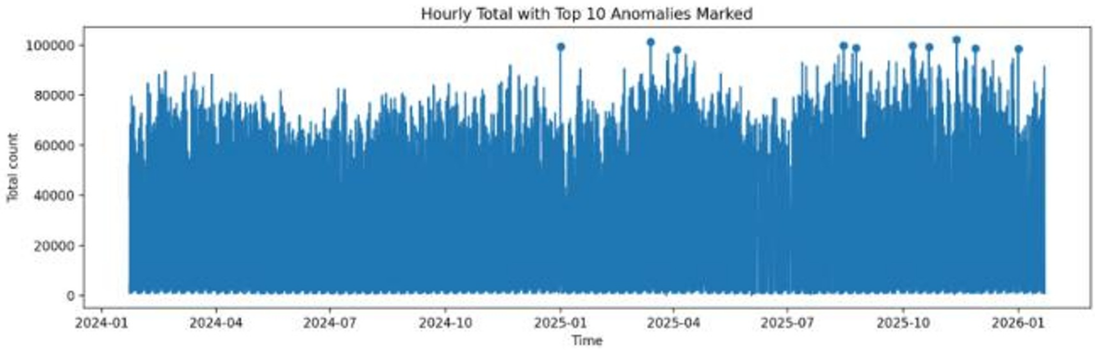

Fig. 4. Combined hourly total series with outliers (Photo/Picture credit: Original).

Table 1. Top 10 anomalous hours identified from the aggregated hourly total series.

Rank	Date-time (dt)	Hourly total (all sensors)	Z-score
1	2025-11-12 17:00	102,091	2.881
2	2025-03-13 17:00	101,273	2.848
3	2025-08-14 17:00	99,816	2.788
4	2025-10-08 17:00	99,789	2.787
5	2024-12-31 22:00	99,412	2.771
6	2025-10-21 17:00	99,127	2.759
7	2025-08-24 14:00	98,926	2.751
8	2025-11-27 17:00	98,631	2.739
9	2025-12-31 22:00	98,399	2.729
10	2025-04-03 17:00	98,012	2.714

4 Conclusion

The research paper focuses on the use of open-source pedestrian counts of the City of Melbourne to examine the temporal trends, prediction, and irregularity of pedestrian counts in Melbourne CBD. The findings are drawn on a basis of per hour pedestrian counts (1,451,812 counts) of 100 sensors during the time period between 23rd January 2024 and 22nd January 2026. As shown by the results, there has been a certain temporal pattern in pedestrian counts. In the weekly diurnal heatmap, the pedestrian counts are low in the late-night and early-morning hours, moderate in the day and early-evening hours, and relatively more in the weekend than weekdays.

When it comes to the forecasting aspect, even a simple model involving seasonal factor can be used as an explanatory variable of the largest periodicity of the pedestrian data and can have a MAE of 5,459.60, a RMSE of 8,819.36, and R^2 of 0.8845. As can be seen, the variability between the number of pedestrians is largely explained by the weekly periodicity. Nevertheless, high data points in the data are a serious problem when it comes to forecasting the numbers of pedestrians. On the other hand, the process of anomaly detection by using the z-score method helps in detecting the extreme values in the data, which take the form of positive spikes in the data, mostly seen during late afternoon.

To sum up, it is concluded that the notion of aggregated time-series analysis is a significant instrument in order to comprehend normal and abnormal movement of pedestrians in Melbourne Central Business District. Nevertheless, statistical anomalies should be regarded with some caution because they might point to genuine changes in behavior or alternatively issues with the sample. Future research will be enhanced when more additional variables like weather and events are included that influences the behavior of pedestrians so as to be able to predict and have a better understanding on the abnormal pedestrian behavior.

References

1. R. Pfiester, R.G. Thompson, Y. Zhang, Spatiotemporal exploration of Melbourne pedestrian demand. J. Transp. Geogr. **95**, 103151 (2021)
2. F. Jiang, J. Ma, Z. Li, Pedestrian volume prediction with high spatiotemporal granularity in urban areas by the enhanced learning model. Sustain. Cities Soc. **79**, 103653 (2022)
3. S. Eom, Y. Nishihori, How weather and special events affect pedestrian activities in Melbourne. Int. J. Sustain. Transp. **16**, 462–475 (2021)
4. City of Melbourne, Pedestrian Counting System: Monthly counts per hour, City of Melbourne Open Data Portal, https://data.melbourne.vic.gov.au/explore/dataset/pedestrian-counting-system-monthly-counts-per-hour/table/ (accessed 3 February 2026)
5. City of Melbourne, Pedestrian Counting System - Sensor Locations, City of Melbourne Open Data Portal, https://data.melbourne.vic.gov.au/explore/dataset/pedestrian-counting-system-sensor-locations/table/ (accessed 3 February 2026)
6. R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice (OTexts, Melbourne, 2021)
7. D. Lehmberg, F. Dietrich, G. Köster, Modeling Melburnians—Using the Koopman operator to gain insight into crowd dynamics. Transp. Res. Part C Emerg. Technol. **133**, 103437 (2021)
8. V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey. ACM Comput. Surv. **41**, 15 (2009)
9. M. Zaouche, N.W.F. Bode, Bayesian spatio-temporal models for mapping urban pedestrian traffic. J. Transp. Geogr. **111**, 103647 (2023)
10. V. Bakhshi, M.S. Tushar, S. Bayldon, S.W. O'Brien, R. Buehler, S. Hankey, K. Nordback, Scalable validation of automated bicycle and pedestrian counters: Evidence from Washington, D.C. Transp. Res. Interdiscip. Perspect. **34**, 101661 (2025)