

246

Interpreting the configuration of micro-urban environments:

Line-based Analyses vs. Visibility Graph Analyses for estimating pedestrian flows

Federico Mara; Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Italy; federico.mara@phd.unipi.it

Özge Öztürk Hacı; Department of Geomatic Engineering, Yildiz Technical University; Knowledge Discovery and Data Mining Laboratory, ISTI-CNR; ozgeo@yildiz.edu.tr

Fatih Gülsen; Department of Geomatic Engineering, Yildiz Technical University; fgulgen@yildiz.edu.tr

Diego Altafini; Welsh School of Architecture, Cardiff University, UK; Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Italy; altafinid@cardiff.ac.uk

ABSTRACT

The existence of strong correlations between pedestrian flows and Space Syntax metrics, whether in form of Axial Analysis (AA), Angular Segment Analysis (ASA), or Visibility Graph Analysis (VGA), has been extensively discussed. While AA and ASA tend to perform these estimations well, at both architectural and urban scales, critical aspects emerge in the VGA, especially when applied to micro-urban scales. Those can be summarized in the grids' non-randomness and their spacing, the impact of variations in altitude and the heavy computational workloads required to model visual fields' relationships in larger areas. Still, VGA is a valuable resource to micro-urban analyses, as it depicts details within areas characterized by wide open spaces or higher morphological complexity that are beyond line-based models' capabilities, potentially enabling more precise insights into local movement logics. Considering this, the paper proposes an initial methodological discussion on the methods for estimating pedestrian flows at micro-urban scales, comparing different environment extensions, techniques, metrics, and extraction methods. Tests are performed within a 10.000m² historical district at Pisa, Italy, which aligns with the spatial and morphological characteristics deemed suitable to VGA. Using spatial statistics, measures associated to pedestrian flows are correlated to observed pedestrian traffic passing through 10 gates in a span of three days to highlight what measure gives a better response. This research encourages the exploration of VGA within micro-urban contexts and can be useful to enable a finer-grained comprehension of social logics in space, as many VGA measures seem to be suitable to comprehend pedestrian flows hierarchies.

KEYWORDS

Micro-urban scale analysis, Pedestrian Flows, Visibility Graph Analysis, Isovist-based metrics, Spatial Correlations.

1 INTRODUCTION

Space Syntax has demonstrated excellent capabilities in estimating both to-movement and through-movement across various scales, proving to be an inherently multiscalar tool. Its applications span from the supra-regional scale, estimating vehicular flows (Hillier et al 2007, Turner 2009, Altafini Cutini 2022), to the urban and neighbourhood scale (Hillier Iida 2005, Van Nes Yamu 2021, Baran et al 2008, Van Nes Lopez 2007, Lopez Van Nes 2007), down to the architectural scale (Hillier et al 1993, Turner Penn 2002) by estimating pedestrian natural movement (Sharmin Kamruzzaman 2018, López Baeza et al 2021).

In particular, regarding the estimation of pedestrian behaviour, various techniques have been employed over time, which have demonstrated strong correlations—reaching peaks ranging between 50% and 80%—with pedestrian movement variance (Penn Turner 2002, Hillier et al 1993, Jiang 2009, Penn et al 1998). Choice, integration, and connectivity metrics have demonstrated the most robust correlations with pedestrian movement, as highlighted by Sharmin Kamruzzaman (2018), indicating a significant association with street network properties (Jiang 2009). Sharmin and Kamruzzaman's (2018) review explores the most robust correlations between various analytical techniques (Axial Analysis and Angular Segment Analysis) by using a quantitative approach. From this review emerges that ASA yield superior correlations, confirming the findings of Hillier and Ida (2005) and Turner (2007). The latter, in particular, introduces the potential of the Road-Centre Line (RCL) base as an alternative to the traditional axial map. It is noteworthy how ASA has proven particularly suitable for estimating pedestrian movement across various scales of analysis – from urban and neighbourhood scales to micro-urban and architectural scales – through the utilization of both axial maps and ASA RCL-based maps.

Meanwhile, the Visibility Graph Analysis (VGA) has proven to be functional in exploring dynamics at the neighbourhood, micro-urban, and architectural scales, due to its higher granularity and reliability in estimating pedestrian movement (Turner Doxa Sullivan Penn 2001, Penn Turner 2002, López Baeza et al 2021), with the added potential to investigate otherwise undetectable metrics (see Koutsolampros 2019 for a review).

Therefore, while AA and ASA perform these estimations well at different scales, some critical aspects of VGA emerge, especially in analysing urban scales scenarios. These issues are represented by: the impact of grid spacing and the non-randomness of the grid (Dalton et al 2022) – which incur in differences in depth among analyses and in analysis strictly related to the vertices grid –; the impact of variations in altitude – a shared trait with the Axial Analysis: variations in altitude can distort the fields-of-view, thus affecting significantly small scales analyses, which require a more granular and precise overview –; and the heavy computational workloads required to calculate the visual field relationships among high amount of vertices in large areas, which limits the granularity of the visual fields. Nevertheless, VGA represents a valuable resource to micro-urban analyses, as it allows to depict

details within areas characterized by wide open spaces, higher morphological complexity that, more than often, are set beyond the line-based models capabilities, potentially enabling more precise insights into movement logics at such scales (López Baeza et al 2021).

At the neighbourhood and micro-urban scale, there appears to be a potential overlap in the analysis techniques applicable to investigating pedestrian behaviour. These investigations, in particular, may prove important for exploring social phenomena in the context of criminal dynamics (Hillier Sahbaz 2012, Van Nes et al 2016, Summers Johnson 2017, Nubani Wineman 2005, Nubani Wineman 2022, Mara Cutini 2023), and it is intriguing to understand the type of response and accuracy that each technique can provide in estimating pedestrian movement and behaviour. This calls for more studies to corroborate the limitations of AA and ASA abstractions for open spaces at neighbourhood and micro-urban scales, as well as to recognise the issues or potentials for improvement in VGA.

With these challenges in mind, this paper proposes a discussion by comparing different methods for estimating pedestrian flows at micro-urban scales. In particular, we compare different environment extensions, techniques, metrics, and extraction methods, analysing their impact in the search for correlation with the actual flows. In accordance with the stated objectives, the paper is organized as follows: Section 2 introduces the case study, datasets, and the methodology utilized, encompassing analysis techniques, adopted SS metrics, the methodology for extracting such metrics, and the size of the analysed environment. Section 3 delves into a comprehensive discussion of the results. Lastly, Section 4 presents the conclusions along with the main findings and suggestions for further developments.

2 DATASETS AND METHODS

2.1 Case Study – the historical centre of Pisa, Italy

The case study is located within the historical centre of Pisa, Italy. The city is characterized by a vibrant university community that populates the historic centre. There, numerous historical buildings are dedicated to university centres or student residences. Moreover, tourism serves as a pivotal economic driver, with notable attractions such as the renowned *Piazza dei Miracoli* and its iconic leaning tower drawing visitors from afar. The focus of our investigation lies in a central area of the historical centre, accessible exclusively to pedestrians.

This area is composed of *Piazza delle Vettovaglie*, *Largo Sant'Omobono*, and the streets leading to it. The medieval conception of this area offers a unique and irregular spatial configuration, with abrupt transitions from wide open spaces with great visibility, to the self-referential square, to the narrow and winding connecting streets that reveal glimpses of tower houses and historical buildings. This generates very different environmental conditions and perceptions of space, even within a few dozen metres. Given its location in the heart of the historic centre, this area is traversed and inhabited

by a significant number of people (locals, workers, students, and tourists) throughout the day, reflecting the area's particular transformation. In the early morning on workdays, a small local market takes place here; it transitions into a hub for tourists or workers during the late morning and early afternoon; and evolves into a passage or stopover point in the evening. Finally, it represents the main nightlife hub, especially on weekends.



Figure 1. Configuration of the study area, gates location and buffer considered for spatial correlations.

2.2 Datasets

For this study three distinct datasets were utilized. The first two were essential for constructing the basic grid for space syntax analyses, while the third was utilized to explore spatial correlations. Spatial modelling relied on data from the Carta Tecnica Regionale (CTR 2K) provided by Geoscopio Toscana (Regione Toscana, 2019), which formed the base for axial and angular analysis. Additionally, data concerning the RCL, extracted from Geoscopio Toscana, were utilized.

Furthermore, data regarding the pedestrian flows were collected by counting the number of people traversing the area during the analysed days: Wednesday (17.05.23), Friday (19.05.23), and

Sunday (21.05.23). Specifically, a manual remote count was performed on pedestrians going in and out through 10 gates (see fig.1), which essentially represent all access points to the area, except for gate number 10. In this manner, the aggregated number of people entering and exiting through each gate was recorded every 24 hours, for a total of 30 observations (see Table 3). The counting process involved reviewing surveillance camera footage from the *Polizia Municipale*, as automatic counting through the use of computerized systems was not feasible due to privacy concerns. Nevertheless, this manual counting process enabled the footage to be slowed down during peak traffic times, facilitating accurate counting.

2.3 Methodology

As mentioned, the study aims to compare the gates hierarchy in terms of real flow with the gate hierarchy in terms of configurational metrics potentially related to movement. To conduct these analysis, four variables were considered: I) environment extent, II) analysis technique, III) analysis metrics, IV) extraction technique (see figure 2).

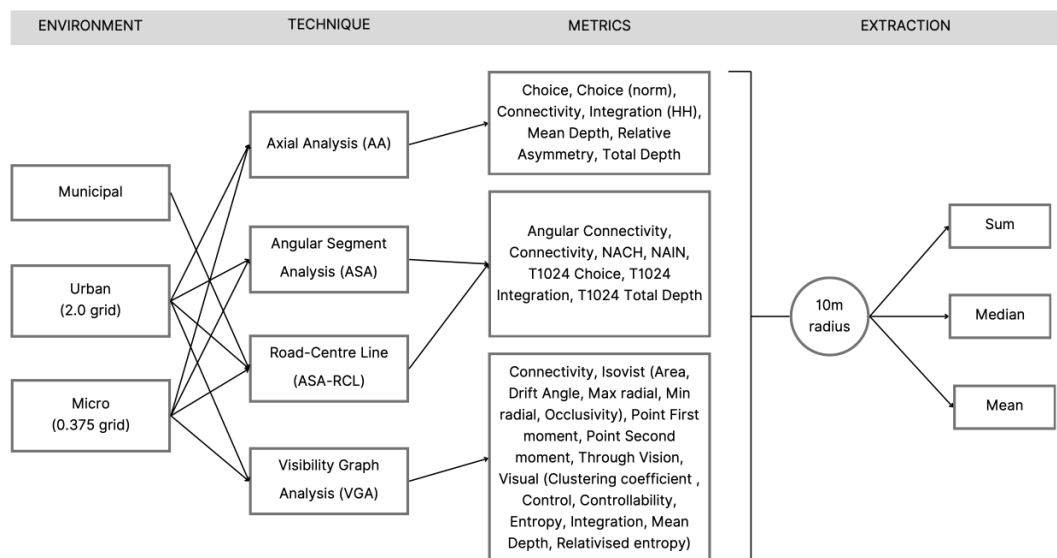


Figure 2. Scheme of the variables tested: I) environment extensions, II) analysis techniques, III) metrics, and IV) extraction approach.

2.3.1 Environment extension

Three environment extensions were selected. The first, corresponding to the urban space within the historic walls, covering approximately 2.6km², referred to as the 'urban' environment. The second, with a more limited area of approximately 24,800m² (157.5m*157.5m), labelled as the 'micro' environment. Additionally, for the ASA and RCL analysis (as detailed below), the entire municipal area of Pisa was also considered as the starting grid, termed the 'municipal' environment.

2.3.2 SS Analysis techniques and configurational metrics selection and extraction

Four analysis techniques were tested: Axial Analysis (AA), Angular Segment Analysis (ASA), Angular Segment Analysis using a Road-Centre Line dataset (ASA RCL), and Visibility Graph Analysis (VGA). The grids used for the analysis at the ‘micro’ scale are depicted in Figure 3.



Figure 3. Grids for the analyses: categorized by length – lines for AA (a), segments for ASA (b), RCL segments for ASA (c); and categorized by size – vertices for VGA (d).

AA was computed following the methods outlined by Turner et al. (2005, p.428), reducing the environment to a grid and utilizing the minimal subset (fewest and longest lines). ASA was conducted using the same procedure as AA, but also considering angular relationships described in Turner, (2001). The ASA RCL used the RCL map extracted from *Geoscopio Toscana* as the base grid, following the methodologies outlined by Turner (2007). Finally, for VGA, efforts were made to use the highest granularity possible to obtain a densely packed grid and thus mitigate the effect of a non-randomized grid highlighted by Dalton et al. (2022). Thus, two different granularities were used based on the size of the environment, considering the limitations posed by DepthmapX 0.5 software. Specifically, a 2 meters grid was adopted for the urban environment – so to not exclude any road spaces – and a 0.375

meters grid for the micro environment; even though further reductions from 0.375 to 0.2 were possible, the computational workloads proven themselves as rather prohibitive.

Both urban and micro environment extensions were considered for each of the 4 analysis techniques, except for ASA RCL, which has three, since it was also computed also at the municipal scale (Figure 2). The analyses considered 4 topological radii (Rn, R3, R5, and R7) for all three scales, but the micro-urban (*Piazza della Vettovaglie* – Figure 3). The variables used in each technique are summarized in Table 1.

Table 1. Metrics calculated for each of the investigated configurational analysis technique, in grey.

Metric	AA	ASA	ASA-RCL	VGA
Angular Connectivity				
Choice				
Choice (norm)				
Connectivity				
Integration (HH)				
Isovist Area				
Isovist Compactiveness				
Isovist Drift Angle				
Isovist Drift Magnitude				
Isovist Max. Radial				
Isovist Min. Radial				
Isovist Occlusivity				
Isovist Perimeter				
Line Length				
Mean Depth				
NACH				
NAIN				
Node Count				
Point First Moment				
Point Second Moment				
Relative Asymmetry				
Segment Length				
T1024 Choice				
T1024 Integration				
Through Vision				
Total Depth				
Visual Clustering Coefficient				
Visual Control				
Visual Controllability				
Visual Entropy				
Visual Integration (HH)				
Visual Integration (P-Value)				
Visual Integration (Tekl)				
Visual Mean Depth				
Visual Relativised Entropy				

2.3.3 Approach for extracting the configurational metrics

The same approach for all the analysis techniques was adopted to associate the Space Syntax configurational metrics to the gates. From each gate, a buffer radius of 10 meters was established to determine a general field-of-view (FOV). Then, this FOV was clipped in relation to the visual

obstructions (buildings), establishing a restricted FOV (Figure 1), to which configurational metrics were associated as sum, median and mean values, subsequently used for correlations and descriptive statistics.

2.3.4 Statistical Analysis of Pedestrian Data

In order to robustly analyse pedestrian flows within the micro-urban context of Pisa, a series of statistical methods were employed to both summarize the dataset and to reveal patterns in pedestrian movement. We initiated our investigation by calculating fundamental descriptive statistics, such as means, standard deviations, and quartiles. These are essential to delineating the central tendencies and dispersions within pedestrian counts across varied gates and time frames, providing a foundation for discerning the distribution of pedestrian traffic and pinpointing distinctive trends or outliers. For the identification of outliers, which might skew our analysis, the Interquartile Range (IQR) method was employed. This approach is esteemed for its robustness, designating outliers as those data points existing beyond 1.5 times the IQR from the quartiles, thereby facilitating a more precise appraisal of pedestrian patterns. Advancing this foundational analysis, the study investigated the relationships between pedestrian counts and spatial metrics utilizing Pearson correlation analysis. This method measures the linear association between two continuous variables, with correlation coefficients ranging from -1 (indicating a perfect negative correlation) to +1 (indicating a perfect positive correlation), and values nearing 0 implying no linear correlation (Stigler 1989). This analysis is instrumental in identifying spatial metrics that exhibit significant associations with pedestrian flows, thereby yielding insights into the spatial determinants influencing pedestrian behaviour. In the expanded dataset used for analysis, which encapsulates data gathered over three distinct days, the dynamic temporal effects were captured, thus augmenting the reliability of the resultant correlations. All statistical analyses were conducted with Python, a programming language lauded for its potent libraries suitable for data analysis. Utilizing libraries such as Pandas for data manipulation and SciPy for executing statistical functions, we confirmed the assumptions of normality and linearity via appropriate tests prior to the Pearson correlation analysis. Configuration settings for the Pearson correlation included parameters for computing a 95% confidence interval, facilitating the precise computation of correlation coefficients and ensuring the replicability and reliability of our findings. The integration of pedestrian counts collected under the same gate number with space syntax metrics – 9 datasets: micro and urban AA, ASA, ASA RCL, VGA and municipal ASA RCL – into a cohesive dataset is a novel approach that extends the scope of analysis. This integration not only captures the dynamic effects of time on pedestrian behaviour but also offers a methodological enhancement, enabling a comprehensive evaluation of time-related variables, such as seasonal effects, special events, or variations across different days of the week. This methodological strategy substantially refines the accuracy and generalizability of the correlation analysis, permitting a deeper understanding of the interplay between pedestrian mobility and space syntax. Conclusively, the application of Pearson correlation analysis in this study underscores the method's efficacy in quantitatively evaluating the

interactions between space syntax measures and pedestrian counts, thereby emphasizing the significance of such data in spatial analysis.

3 RESULTS AND DISCUSSION

This section traces the development of the work. Firstly, the dataset related to pedestrian flow is elaborated and discussed. In particular, variations in people count between the studied days and among different gates are analysed, attempting to quantitatively examine how the flows distribute in the area. Subsequently, the study shifts to the analysis of correlations with the extracted SS values, through the comparison of environment extents, analyses, metrics, and extractions adopted. This has been done to understand which methodologies and metrics offer the best correlations at this scale, with careful considerations of VGA due to its use at a non-architectural scale.

3.1 Pedestrian count analysis and descriptive statistics

From the analysis of the distribution of pedestrian counts data, flows emerge as heterogeneous among gates. From a total of 30 observations (3 days, 10 gates), the average count stands at approximately 3,333.80 people. The data exhibits a standard deviation of 3,921.96, indicating variability in the counts.

The smallest recorded people count is 210, while the highest reaches 16,964. When examining the data's quartiles, the first quartile (Q1) is at 893.25 people, the median (or the second quartile) is at 1,853 people, and the third quartile (Q3) is at 4,133 people. These quartiles provide a clearer view of the data's spread and central tendency (Table 2).

Table 2. Descriptive statistics of pedestrian flows

Statistic	Value
Count	30
Mean	3,333.80
Standard Deviation	3,921.96
Minimum	210
1st Quartile (Q1)	893.25
Median	1,853
3rd Quartile (Q3)	4,133
Maximum	16,964

To determine the outliers in the "People Count" column, the Interquartile Range (IQR) method was utilized (Tukey 1977). In this approach:

1. The IQR is calculated as the difference between the third quartile (Q3) and the first quartile (Q1).
2. The lower bound is defined as $Q1 - 1.5 \times IQR$
3. The upper bound is defined as $Q3 + 1.5 \times IQR$

Values outside of these bounds are considered outliers. Based on this method, four outliers were identified in the dataset (Table 3, Figure 4):

- On Wednesday at gate 8: 10,788 people
- On Friday at gate 2: 9,172 people

- On Friday at gate 8: 16,964 people
- On Sunday at gate 8: 9,008 people

These values are notably higher compared to the other observations. This is due to the existence of preferential paths by pedestrians, as qualitatively appreciated from numerous on field observations. From *Borgo Largo*, the main street on the right (Figure 1), pedestrians pour into *Piazza della Vettovaglie* by crossing gate 8, which indeed appears as an outlier for each day. Similarly, gate 2 also appears particularly busy, especially with students crossing the area on Wednesday and Friday (Table 3).

Table 3. People count by gate and day.

		Gates									
		1	2	3	4	5	6	7	8	9	10
DAY	WED	3,143	8,451	582	936	484	2,331	2,487	10,788	1,567	515
	FRI	6,045	9,172	906	1,365	1,001	4,463	2,240	16,964	1,983	889
	SUN	2,359	5,214	555	726	528	2,195	1,723	9,008	1,184	210

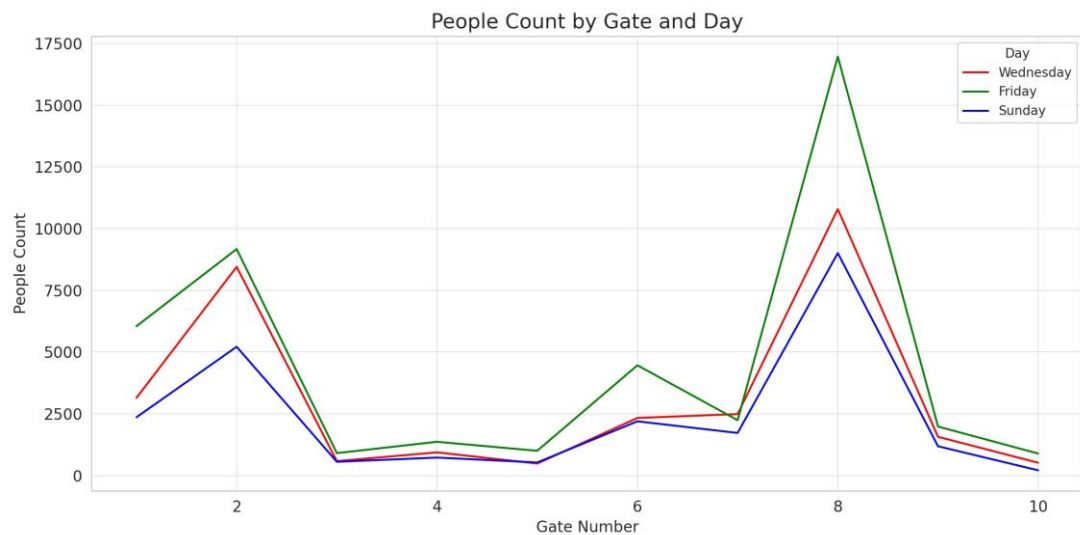


Figure 4. People count by gate and day.

The line chart (Figure 4) illustrates what are the variation in the pedestrian counts across different gates, organized by days of the week. Several observations can be drawn from the chart:

Gate Variability: There's a noticeable fluctuation in the number of pedestrians across different gates. Some gates, as in example gate 2 and gate 8, tend to consistently experience higher foot traffic compared to others. This suggests that these gates tend to be more popular or strategically located where more people tend to pass through.

Day-to-Day Variation: The pedestrian traffic also varies depending on the day of the week. For example, Fridays appear to have a surge in pedestrian counts for certain gates, which might be attributed to specific events or weekly patterns.

Consistent patterns vs. outliers: While there's variability, most of the gates tend to maintain a somewhat consistent number of pedestrians irrespective of the day, indicating a steady flow of people using these gates daily. Nevertheless, there are some days (such as Friday) and gates (such as 2 and 8) with exceptionally high pedestrian counts. Those can be associated with working routines or the night-time activities taking place in the square, to be investigated in further studies.

Overall, decoding these patterns can be important for local policies that require controlling or that are dependent on pedestrian flows, planning maintenance services, or scheduling events around these areas.

3.2 Pedestrian count analysis and descriptive statistics

The top ten SS measures associated with pedestrian movement among all the 9 datasets are examined. To find the most observed variable in the top 10, we counted the number of times the variables in each table occur and identified the variable with the highest frequency. Table 4 shows the most frequent measures detected among all 9 datasets.

By examining Table 4, we observe that, for example, the "Connectivity_mean" variable appears 5 times in the top 10 correlation values. This could suggest that the "Connectivity_mean" variable exhibits a significant correlation across most of the 9 datasets examined. Similarly, choice measures also show high frequency. However, considering the best 10 measures might create a misunderstanding, as significant quantities could be selected but not necessarily with the highest correlation indices. Additionally, some quantities, such as connectivity, are statistically favoured as they appear in many analysis techniques. For these reasons, the ratio (frequency/n. time calculated) was calculated, according to which Table 4 is ranked.

Table 4. Frequencies of the 10 best correlated measures in the 9 datasets.

Variable	Frequency	N of times calculated	Ratio (Frequency/n.times)
Point_Second_Moment_sum	2	2	1,00
Choice_Norm__sum	2	2	1,00
T1024_Choice_R500_metric_sum	2	2	1,00
Choice_sum	2	2	1,00
Point_Second_Moment_mean	2	2	1,00
Point_First_Moment_sum	2	2	1,00
Through_vision_sum	2	2	1,00
Point_First_Moment_mean	2	2	1,00
Through_vision_mean	2	2	1,00
Isovist_Area_mean	2	2	1,00
NACH_R7_mean	2	3	0,66
NACH_R7_median	2	3	0,66
T1024_Choice_sum	3	5	0,60

Connectivity_mean	5	9	0,55
T1024_Choice_mean	2	5	0,40
T1024_Integration_median	2	5	0,40
NACH_sum	2	5	0,40
Connectivity_sum	2	9	0,22

The cluster of the most 5 correlated measures across the datasets was also considered (Table 5). This visualization confirmed that connectivity often achieves a good correlation score but frequently does not rank among the best, unlike measures related to betweenness, which tend to be among the most prominent. However, this is just a side result, which largely confirms existing literature studies. Our main goal is to understand the different correlations obtained when varying the considered variables. The following paragraph delves into this topic in depth.

Table 5. Frequencies of the 5 best correlated measures overall, considering the 9 datasets.

Variable	Frequency
T1024_Choice_sum	3
Point_Second_Moment_sum	2
T1024_Choice_R500_metric_sum	2
Connectivity_mean	2
NACH_R7_mean	2

In Figure 5, the histograms show the distribution of correlation coefficients between variables in each data set. The histograms of each dataset reveal varied correlation distributions with unique skewness characteristics.

Micro AA and urban AA both exhibit a limited number of modes and a negative skew, indicating that the bulk of the data is concentrated on the right, but with micro AA having a broader distribution. Urban AA measures reach higher correlations (more than 0.8), and also the preceding bar of the histogram highlights how urban AA has many measures with a high correlation. Moreover, urban AA reaches the highest negative correlation values among all the analysis.

Micro ASA stands out with very prominent modes and significant negative skewness, implying a strong concentration of specific correlation values towards the higher end, with a longer tail extending towards lower values. Similarly, urban ASA presents several modes with a negative skew, indicating more values clustered at the higher end. ASA, for both the environment, don't reach a 0.8 correlation value, reached instead by the local ASA RCL despite of the low number of measures calculated at this scale (no local radii were calculated). In general, micro RCL has few modes and a slight negative skew, suggesting a subtle concentration of higher values. Urban ASA RCL's almost symmetric distribution indicates an even spread of values on both sides of the mean, with closely aligned mean and median values. Municipal ASA RCL is characterized by numerous modes with a relatively symmetric distribution and a narrower spread of values, indicating an equal balance of correlation values on both sides of the mean.

Urban VGA shows a variety of relationships, with some variables being more strongly correlated than others. The presence of multiple modes suggests complex interactions between the variables, rather than a single, unified pattern of correlation. The slight negative skewness could mean that there are some pairs of variables with a moderately negative correlation, but the majority of the correlations are positive or near zero. Micro VGA, characterized by two distinct peaks and a notable negative skew, suggests a concentration of values in the negative range, with a longer tail on the left side. The slight negative skew indicates that there are fewer instances of strong negative correlations compared to positive ones. The presence of bimodal peaks might suggest that there are two dominant types of relationships within the variables in the dataset. Micro VGA has also higher frequency of strong positive correlations than other datasets. In particular, micro VGA reaches both high positive (more than 0.8) and high negative (more than 0.5) correlation values.

However, these histograms are not informative in a comparative logic, given the different number and types of measures that have been calculated with each type of analysis. On the other hand, they effectively quantify the correlation between the measurements and the flow, revealing the variability and concentration of these relationships.

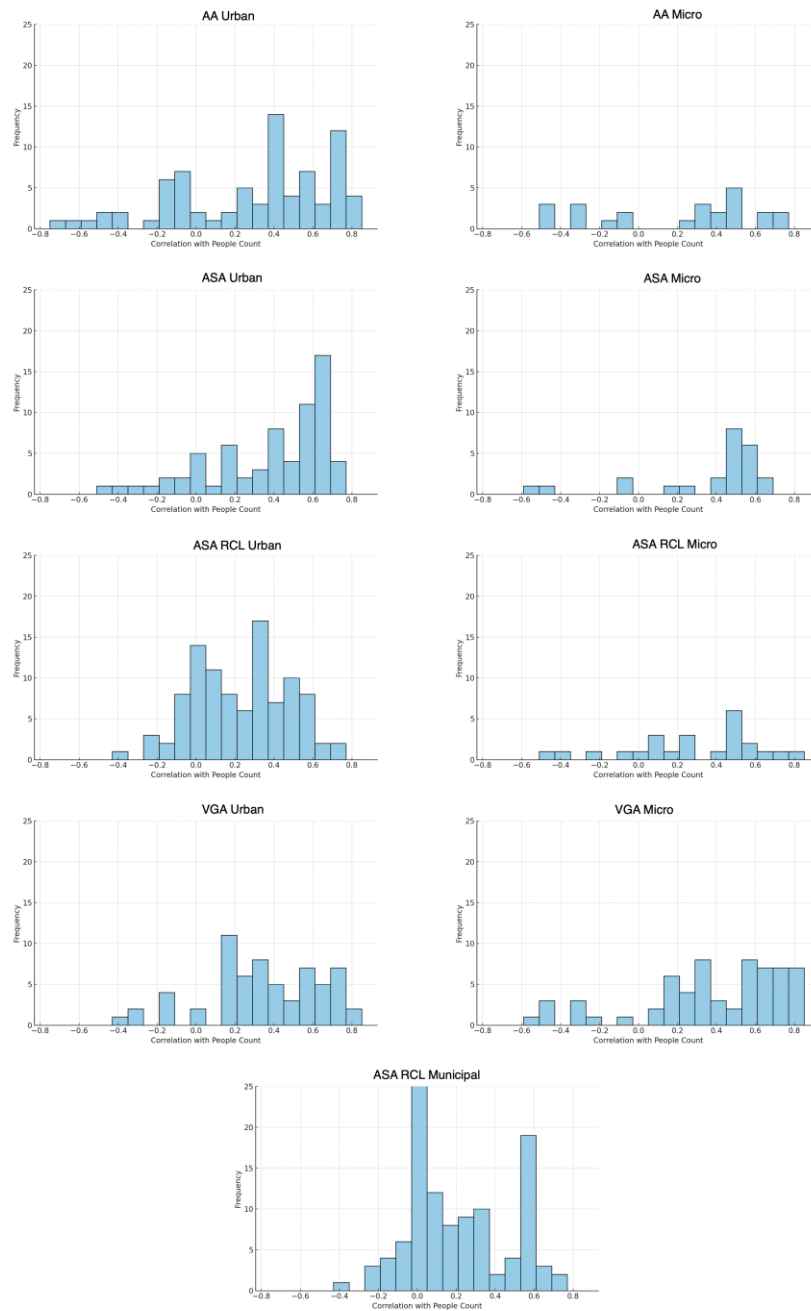


Figure 5. Histograms depicting datasets for all used techniques in all scales.

Figure 6 presents the correlation of Space Syntax metrics with pedestrian count, for each analysis technique (rows) and environment extension (columns), except for the municipal ASA RCL (alone in the last row). The correlations are illustrated using a threshold value of 0.4 for positive values and 0.6 for negative values to represent significant variables.

Overall, both AA and ASA yield a higher number of measures achieving a good positive correlation (>0.6) through the urban environment extension compared to the micro environment extension. However, this result is influenced by the greater number of measures calculated at the urban scale (various metric and topological radii, see Section 3). Nevertheless, this trend is maintained

for negative correlation only by AA, while for ASA, the number of measurements with appreciable correlation is equivalent, albeit much lower than the positive one (>0.4). In the case of ASA RCL, the trend is confirmed – 6, 4, and 3 measures positively correlated with the flow respectively for municipal, urban, and micro environment extensions – while only ASA RCL micro is able to achieve good negative correlations. Conversely, VGA reverses this trend. In fact, VGA micro is able to obtain both a greater number of good positive correlations – 21 measures positively correlated with pedestrian flow with a threshold of 0.6 – and negative correlations compared to urban VGA. This result appears to be quite significant, especially considering the calculation of the exact same measures on both environment extensions.

Another significant aspect emerges from the analysis of the measures. Choice confirms itself as a very good movement estimation metric, being the most frequently present measure among those positively correlated with pedestrian movement with significant value (> 0.6). Specifically, concerning axial and segment analysis, Choice measures are present 9 out of 9 times when calculated on the micro environment, and still present 27 out of 47 times when calculated on the urban grid extension. On the contrary, Choice measures are not that frequent in the case of urban ASA: Integration ranks first with 11 out of 23 occurrences, followed by Choice with 8 out of 23, and then Connectivity and Depth with 2 out of 23 respectively. However, still for ASA, the first two measures to provide the best positive correlation are Choice measures (Choice_sum and Choice_mean). The situation is more varied for VGA, which offers a significant number of measures positively correlated with pedestrian flow: as already said, 21 with correlation greater than 0.6, but 12 with correlations greater than 0.7, and 4 with correlations greater than 0.8. Micro VGA proves to be the second analysis technique with a higher number of measures positively correlated with pedestrian flow significantly (> 0.6) with 21 measures, preceded by urban ASA with 23 and followed by urban AA with 20 measures. Regarding negative correlation, the most prevalent measure across the datasets is Depth (mean or total), present 15 out of 21 times.

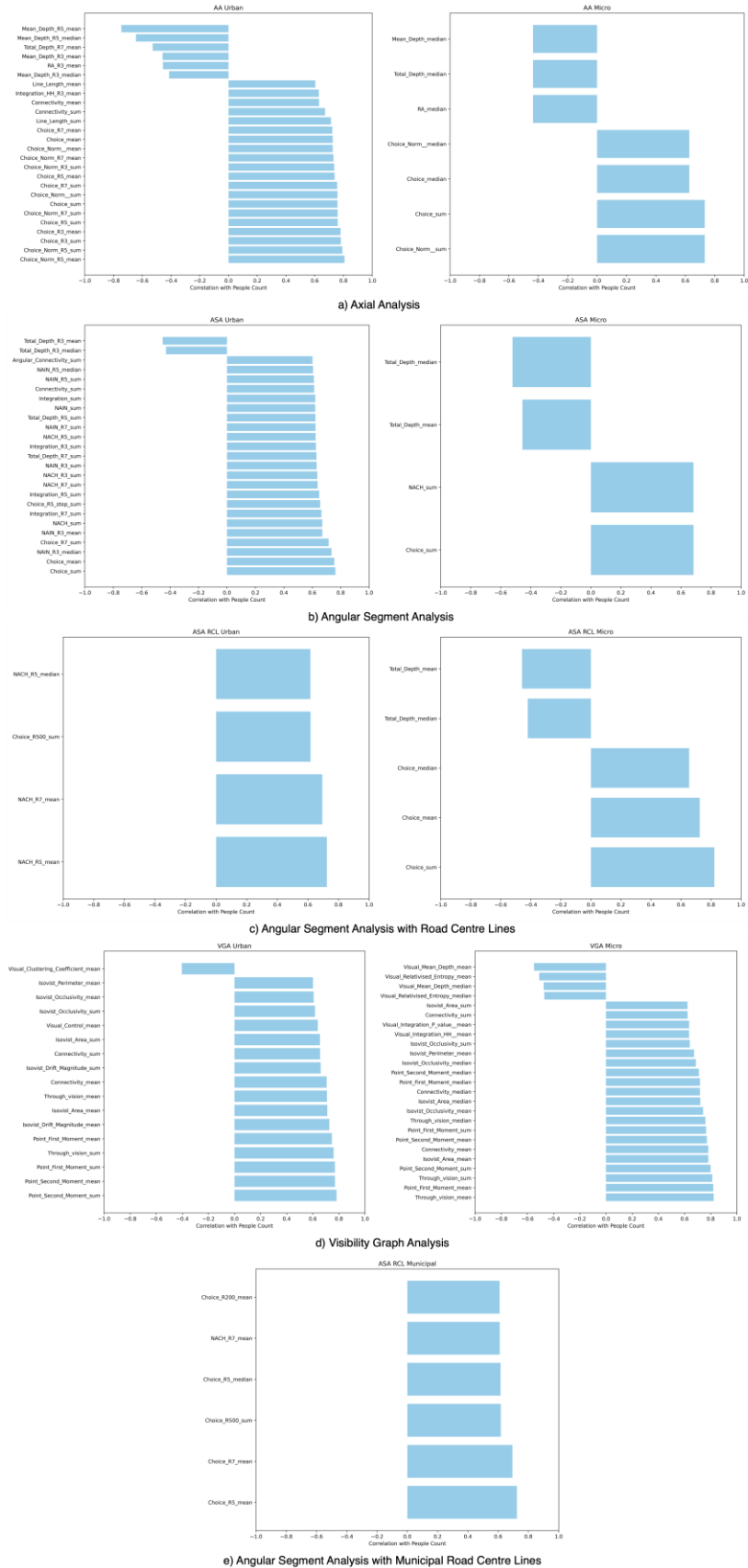


Figure 6. The Histograms with the best correlated measures for all the datasets: 0.6 positive threshold and 0.4 negative threshold.

In Figures 7 and 8, the heatmap provided illustrates the correlation coefficients between pedestrian count and a variety of measures across nine distinct datasets. Correlation is a statistical measure that quantifies the degree and direction of the relationship between two variables. The coefficients range from -1 to 1, where values near 1 or -1 indicate a strong relationship, and values close to 0 suggest a weak relationship. Positive values indicate a direct correlation, while negative values denote an inverse correlation. Figure 7 presents positive correlations, indicating that as pedestrian counts increase, the associated measures also tend to increase. Figure 8 presents negative correlations instead.

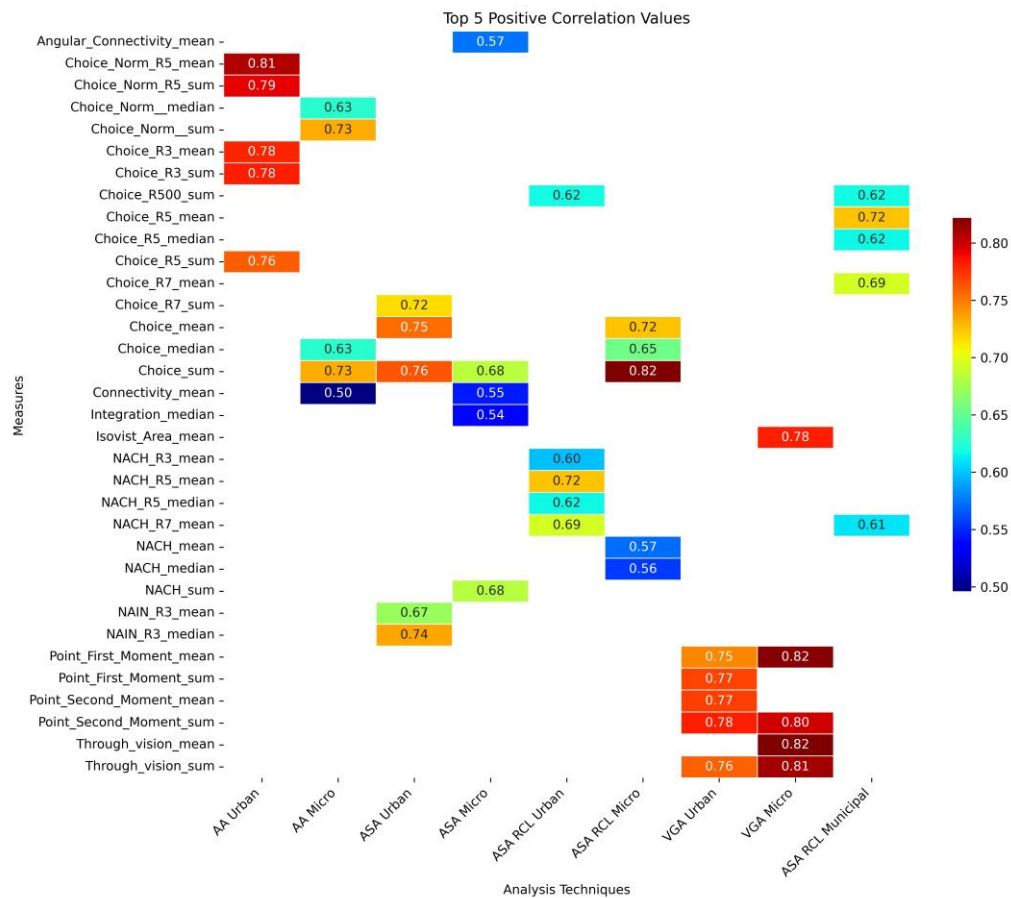


Figure 7. The top 5 positive correlation values for all dataset.

In Figure 7, it can be observed that among all the line-based analysis, namely AA, ASA, and RCL ASA, the measure with the highest degree of correlation is a Choice measure: Choice_Norm_R5_mean for urban AA (+0.81), Choice_sum for micro AA (+0.73), Choice_sum for urban ASA (+0.76) and micro ASA (+0.68), NACH_R5_mean for urban ASA RCL (+0.72), Choice_sum for micro ASA RCL (+0.82), Choice_R5_mean for municipal ASA RCL (+0.72). For VGA, however, the measures most strongly positively correlated are Point_Second_Moment_sum (+0.78) for urban VGA, while Point_First_Moment_mean and Through_vision_mean (+0.82) showed the best correlation for micro VGA. Overall, the most highly correlated values were thus found to be, with a value of 0.82, Point_First_Moment_mean and Through_vision_mean [micro VGA] and Choice_sum [micro ASA RCL].

It is interesting to note that correlation values greater than 0.75 were obtained only from analyses performed on a micro environment extension or, alternatively, from local measures calculated on the urban environment extension.

Regarding negative correlation (Figure 8), less significant values were found. The Mean_Depth_R5_mean AA urban emerges as the most strongly correlated measure (-0.75). However, a similar trend to that observed in positive correlation is reproduced: higher correlations (exceeding -0.4) are found either in analyses conducted on a micro-extended environment or, alternatively, on local measures calculated on an urban-extended environment. Concerning negative correlations as well, micro VGA proves to be highly effective, albeit with values much lower than the peak reached by the AA urban Mean_Depth_R5.

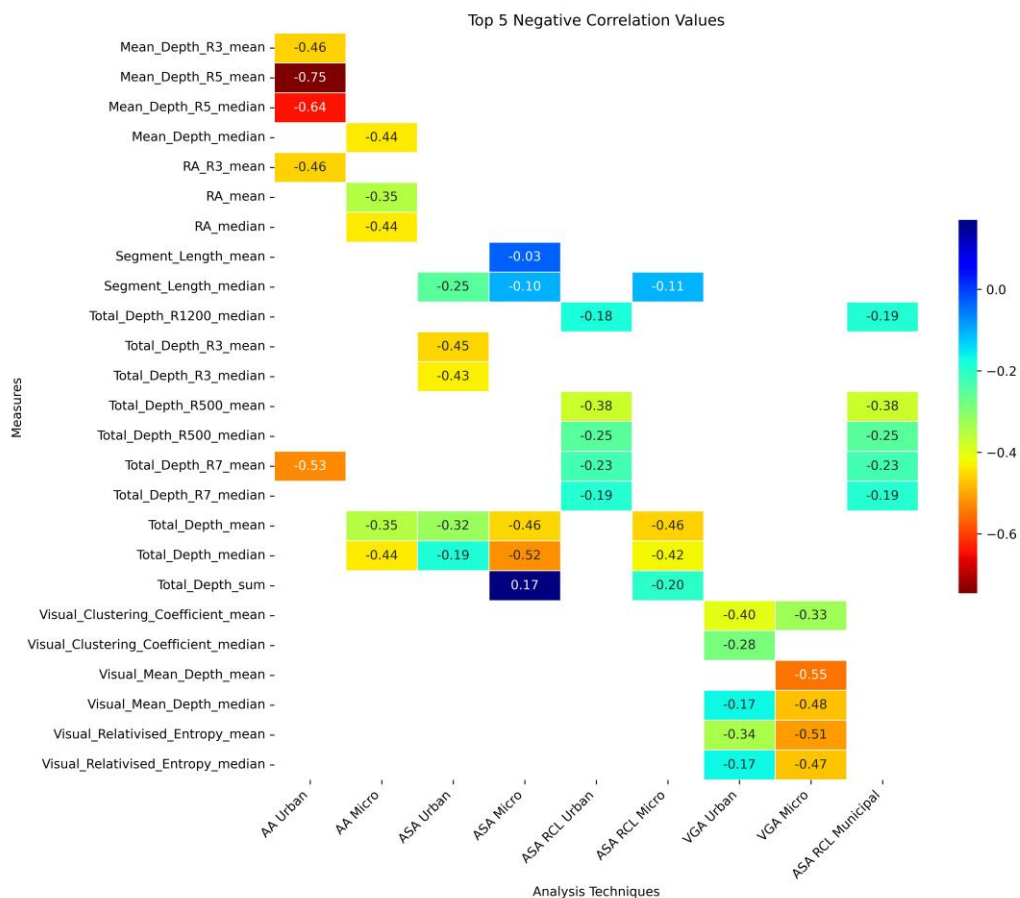


Figure 8. The top 5 negative correlation values for all dataset.

Lastly, Figure 9 presents the spatialization of measurements that, for each analysis technique, have shown the strongest correlations, with a correlation value of at least 0.7, alongside the highest negative correlation.

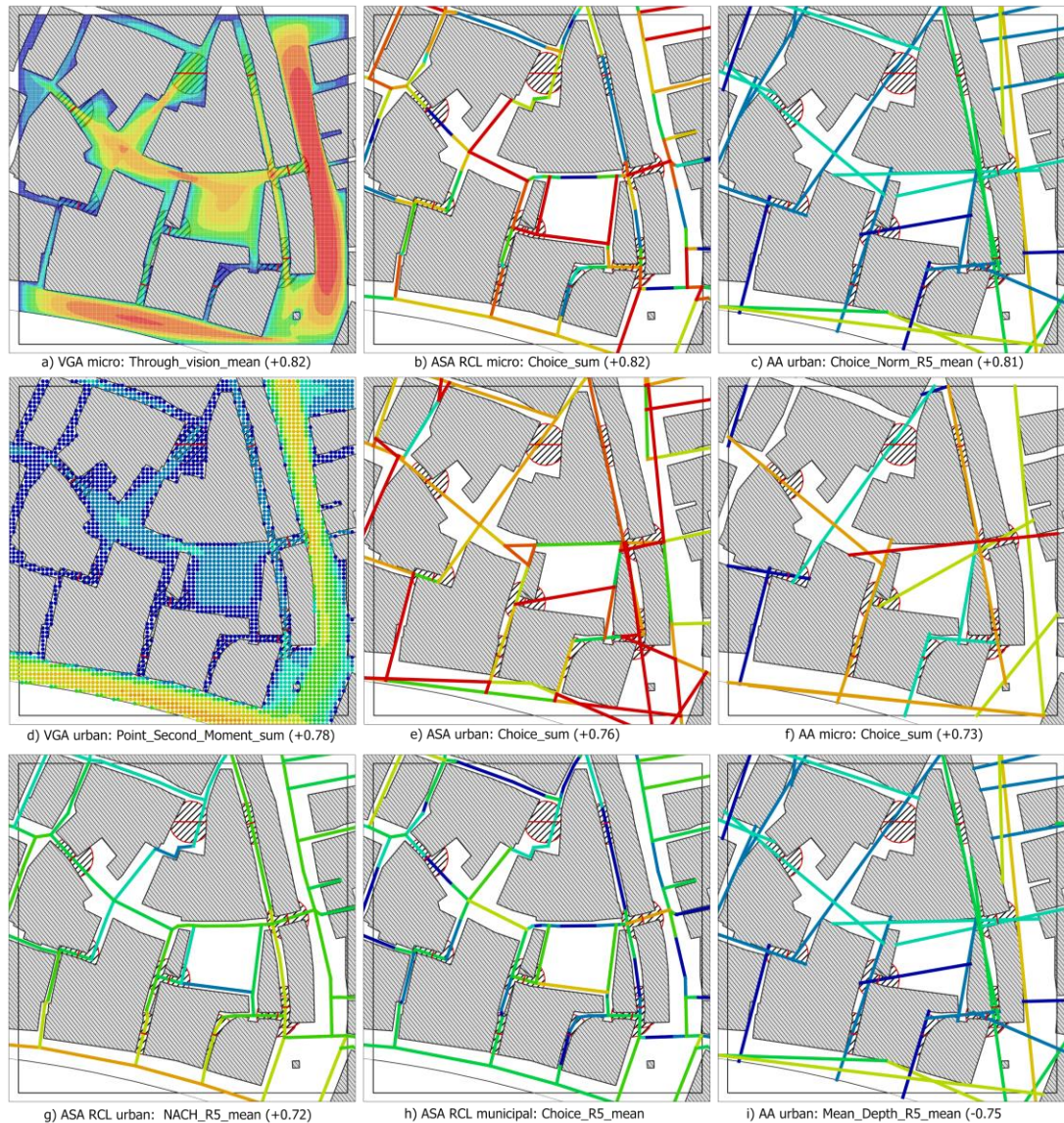


Figure 9. Best correlated measures with at least 0.7 correlation values.

4 CONCLUSIONS

This paper proposed a comparison among various methodologies for estimating pedestrian movement at the micro-urban scale. In particular, several Space Syntax techniques and metrics have been evaluated through the combination of different parameters to test which methodology could prove most effective at this scale, and potentially highlight further opportunities to be explored.

The variables tested included: I) environment extent, II) analysis technique, III) analysis metrics, IV) extraction technique. The ability of SS to intercept pedestrian flows has once again emerged, with correlation values even exceeding 0.8. It has also emerged how VGA can be an excellent tool for micro-urban scale analysis, as it has reported the highest correlation values with pedestrian flows both at the micro- and urban-extended environments, particularly with mean and sum values, and with measures of *Through Vision*, *Point First Moment*, and *Point Second Moment*.

Several significant aspects emerged, in particular:

- Micro VGA analysis provided the best positive correlations with pedestrian movement with two measures: *Through Vision* and *Point First Moment*. Measures such as *Point Second Moment* and *Isovist Area* have also shown strong correlations with pedestrian movement. Overall, in this study and with the techniques employed, VGA emerged as a highly sensitive analysis method because it contains metrics that are highly correlated with pedestrian flow, also its measures can be used in a prediction model to estimate pedestrian flow accurately.
- VGA is the technique with the highest number of measures strongly correlated to pedestrian flow: *Through Vision*, *Point First Moment*, *Point Second Moment*, *Isovist Area*, *Connectivity*.
- The best correlations (above 0.75), both positive and negative, were found with analyses conducted on micro-extended environments or, alternatively, with local analyses on urban-extended environments.
- Choice has stood out as the measure most strongly correlated with pedestrian flow regarding line-based analyses (AA, ASA, RCL ASA), confirming literature findings.
- The best correlations were found by calculating the mean or sum value. Rarely did the use of the median value offer good correlations: only once it reported a correlation value exceeding 0.65, in the case of ASA Urban NAIN R3 (+0.74).

These results will need to be tested on additional case studies to confirm their robustness and potential generalizability, in order to develop a shared methodology that could represent the preferred approach for studying pedestrian movement at the urban and micro-urban scale, with significant implications for urban planning, safety, and sustainable strategies. The aforementioned findings represent, according to the authors, interesting initial points for further investigation. In particular, the main directions identified are as follows:

- Quantitatively assess the impact of varying the size of the environment chosen as the base grid for conducting the various analyses and, within the constraints imposed by the calculation software, conduct the various analyses on several environment extents;
- Regarding VGA, develop methodologies to mitigate the impact of the non-randomness of the grid, especially in the case of non-high-density vertices maps;
- Test additional methods of extracting measures, such as comparing different radii for extraction and experimenting with other approaches based on configurational theory;
- Improve the dataset related to pedestrian flow, both in terms of temporal extension – therefore, having more observation days than the three considered in this study – and in terms of granularity – thus counting the number of pedestrians crossing a bigger number of virtual gates in the investigated area, in addition to access and exit gates – thereby obtaining

additional parameters to evaluate the potential of each measure to detect pedestrian flow and estimate movement in the urban environment;

- Finally, explore the impact of the distribution of activities and POIs on pedestrian flow, perhaps through the use of Place Syntax or Agent-based Modelling.

ACKNOWLEDGEMENTS

This research was made in collaboration with Cardiff University, which has received funding from the **United Kingdom Research and Innovation Post Doctoral Fellowship Guarantee Scheme, set over the European Union's Horizon Europe – Marie Skłodowska Curie Actions Post Doctoral Fellowships. UKRI Grant no. 101107846-DECIDE/EP/Y028716/1**. Views and opinions expressed are those of the authors only, and do not necessarily reflect those of the United Kingdom or European Union. Neither the United Kingdom or European Union nor the granting authority can be held responsible for them.

REFERENCES

- Altafini D; Cutini, V. (2022) Configurational properties and the internal geography of local-regional urban spaces: findings of a multiscale-based analysis of the Tuscany's road-circulation networks. In: Van Nes A, De Koning R (eds.) *Proceedings of the 13th Space Syntax Symposium*. Western Norway University of Applied Sciences (HVL), Bergen, Norway.
- Baran, P. K., Rodríguez, D. A., & Khattak, A. J. (2008). Space syntax and walking in a new urbanist and suburban neighbourhoods. *Journal of Urban Design*, 13(1), 5-28.
- Conroy Dalton, R., Dalton, N. S., McElhinney, S., & Mavros, P. (2022, August). Isovist in a Grid: Benefits and limitations. In *Proceedings 13th International Space Syntax Symposium* (p. 551). Western Norway University of Applied Sciences, Department of Civil Engineering.
- Hillier, B., & Iida, S. (2005, September). Network and psychological effects in urban movement. In *International conference on spatial information theory* (pp. 475-490). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hillier, B., & Sahbaz, O. (2012). Safety in numbers: high-resolution analysis of crime in street networks. In *The urban fabric of crime and fear* (pp. 111-137). Dordrecht: Springer Netherlands.
- Hillier, B., A. Turner, T. Yang and H. T. Park. 2007. Metric and topo-geometric properties of urban street networks. In *Proceedings Space Syntax. 6th International Symposium, Istanbul, 2007*, ed. A. S. Kubat.
- Hillier, B., Penn, A., Hanson, J., Grajewski, T., & Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: planning and design*, 20(1), 29-66.
- Jiang, B. (2009). "Ranking Space for Predicting Human Movement in an Urban Environment." *International Journal of Geographical Information Science*, 23(7), 823–837.
- Koutsolampros, P., Sailer, K., Varoudis, T., & Haslem, R. (2019, July). Dissecting Visibility Graph Analysis: The metrics and their role in understanding workplace human behaviour. In *Proceedings of the 12th International Space Syntax Symposium* (Vol. 12). International Space Syntax Symposium.
- Lopez Baeza, J., Carpio-Pinedo, J., Sievert, J., Landwehr, A., Preuner, P., Borgmann, K., ... & Noennig, J. R. (2021). Modeling pedestrian flows: Agent-based simulations of pedestrian activity for land use distributions in urban developments. *Sustainability*, 13(16), 9268.
- López, M. J., & Van Nes, A. (2007). Space and crime in Dutch built environments: macro and micro scale spatial conditions for residential burglaries and thefts from cars. In *Proceedings of the 6th Space Syntax Symposium (6SSS), Istanbul, Turkiye, June 12-15, 2007*.
- Mara, F., & Cutini, V. (2023, June). Modelling Crime Dynamics and Risky Places: A Basic Combination of Crime Theories and Space Syntax to Evaluate Spatial Crime Impedance. In *International Conference on Computational Science and Its Applications* (pp. 227-243). Cham: Springer Nature Switzerland.
- Nubani, L., & Wineman, J. (2005, June). The role of space syntax in identifying the relationship between space and crime. In *Proceedings of the 5th space syntax symposium on space syntax, delft, holland* (pp. 13-17).

- Nubani, L., & Wineman, J. (2022). Using Isovists in Measuring Surveillance and Expected Guardianship in Residential Neighborhood Property Crimes. *ISPRS International Journal of Geo-Information*, 11(11), 544.
- Penn, A., Hillier, B., Banister, D. and Xu, J. (1998). "Configurational Modeling of Urban Movement Networks." *Environment and Planning B: Planning and Design*, 25, 59 – 84.
- Penn, A., Turner, A. (2002). *Space syntax based agent simulation*. Springer-Verlag (2002).
- Regione Toscana (2019). Direzione Urbanistica e Politiche Abitative - Sistema Informativo Territoriale e Ambientale – SITA.: Grafo stradale e ferroviario della Regione Toscana.
- Sharmin, S., & Kamruzzaman, M. (2018). Meta-analysis of the relationships between space syntax measures and pedestrian movement. *Transport Reviews*, 38(4), 524-550.
- Stigler, S. (1989). Francis Galton's Account of the Invention of Correlation. *Statistical Science*, 4, 73-79. <https://doi.org/10.1214/SS/1177012580>.
- Summers, L., & Johnson, S. D. (2017). Does the configuration of the street network influence where outdoor serious violence takes place? Using space syntax to test crime pattern theory. *Journal of quantitative criminology*, 33, 397-420.
- Tukey, John W. *Exploratory data analysis*. Vol. 2. 1977.
- Turner, A. (2001) Angular Analysis, in: In: Peponis J, et al. (eds.) *Proceedings of the 3rd International Space Syntax Symposium*. Atlanta, pp. 30.1-30.11.
- Turner, A. (2007). From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. *Environment and Planning B: planning and Design*, 34(3), 539-555.
- Turner, A., (2009) *Stitching Together the Fabric of Space and Society: An Investigation into the Linkage of the Local to Regional Continuum*, in: Koch, D; Marcus, L; Steen, J. (Ed.), *Proceedings of the 7th International Space Syntax Symposium*. KTH, Stockholm, pp. 1–12.
- Turner, A., & Penn, A. (2002). Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. *Environment and planning B: Planning and Design*, 29(4), 473-490.
- Turner, A., Doxa, M., O'sullivan, D., Penn, A. From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and design*, 28(1), 103-121 (2001).
- Turner, A., Penn, A., and Hillier, B. (2005). "An Algorithmic Definition of the Axial Map." *Environment and Planning B: Planning and Design*, 32, 425-444.
- Van Nes, A., & López, M. J. (2007). Micro scale spatial relationships in urban studies: the relationship between private and public space and its impact on street life. In *Proceedings of the 6th Space Syntax Symposium (6SSS)*, Istanbul, Turkiye, June 12-15, 2007.
- Van Nes, A., & Yamu, C. (2021). Introduction to space syntax in urban studies (p. 250). Springer Nature.
- Van Nes, A., López, M., de Bonth, L., Verhagen, D., & Waayer, S. (2016). Spatial tools for diagnosing the degree of safety and liveability, and to regenerate urban areas in The Netherlands. In S. Nijhuis, S. Zlatanova, E. Dias, F. van der Hoeven, & S. van der Spek (Eds.), *Geodesign: Advances in bridging geo-information technology, urban planning and landscape architecture* (pp. 139-156). (Research in Urbanism Series; Vol. 4). TU Delft OPEN Publishing. <https://doi.org/10.7480/rius.4.853>