#170

TOWARDS STATISTICAL SIGNIFICANCE OF CONFIGURATIONAL MODELS:

New evidence of variance and bootstrapping

JOÃO PINELO SILVA
Department of Architecture & Interior Design, University of Bahrain, Kingdom of Bahrain
jpinelo@uob.edu.bh

ABSTRACT

Configurational modelling involves simple but powerful methodologies that seamlessly integrate the design process and has high adherence by professionals. However, a traditionally intuitive approach, rather than a statistically informed one, occasionally compromises such models. As a consequence, the models often do not reach statistical significance and therefore are of limited efficacy. Were they to reach statistical significance, configurational models would increase their validity but also the potential for data sharing and therefore their economic feasibility. In this paper, I discuss the aspects of the process of creating configurational models that are crucial towards statistical validity. After introducing the methodological framework, I focus on the two stages for measuring traffic: finding a representative sample for each spatial unit and measuring sufficient units to form the model. I present new evidence of the expected variability of data that dismisses common but false assumptions that often lead to statistical insignificance. I demonstrate how the introduction of variance breaks down a model. I argue that reaching statistical significance can be achieved with the use of basic statistics, which are within reach of designers. Finally, I introduce Bootstrapping, an advanced but straightforward method to provide statistical significance in cases of a small sampling.

KEYWORDS

Space syntax, configurational models, spatial configuration, covariance analysis, bootstrapping.

1. INTRODUCTION

According to Karimi (2012), the configurational approach, "... provides a reliable evaluation system that can lead the design process by bringing together creativity and research into a single framework." (Karimi 2012, p. 299). Perhaps this is part of the reason for its wide acceptance by planners and architects. Often, such studies are used by professionals to 'get a grasp' on the role of the configuration in social and economical phenomena.

Observing social dynamics within a spatial framework is rooted in the architectural tradition through the works of some of the greatest thinkers, who advocate direct observation as a way to understand city life. Examples are the works of Jacobs (1961), Alexander (1977, Alexander 2002), Whyte(1980, Whyte, Whyte 1988), Hall (1980, Hall 1990), Lynch (1964), Gehl (1999, 2000, 2006, 2006, 2010). While these authors refer to observation as the means to an intuitive understanding of a place, others, such as Gehl, use more systematic and quantified practices. Gehl and Svarre (2013) compiled a list of methodologies for the observation and measurement of behaviour and traffic. Architects and planners often extend the approach to model pedestrian and vehicular traffic, by linking configurational properties to traffic volume. However, this is sometimes done in an intuitive way, following tradition, rather than a statistically sound methodology. As a result, the models may lack statistical significance, and therefore their use may be limited.

Architects are no longer the exclusive drivers for the systematic consideration of pedestrians in planning. Non-motorized counts are a key element in calibrating multimodal models (Ryus 2014, p.19), therefore, suggesting opportunities for data sharing between architects, planners and 'traditional' transportation modellers. Pedestrian and cycling data tend to be used for several purposes simultaneously (Ryus 2014, p.86). A practitioner survey identified the lack of staff time and funding limitations as the most common barriers to collecting more non-motorized traffic data (Ryus 2014, p.25). Lack of sufficient counting data is a limitation of configurational-based models. However, in a context of wide interest for such data, the lack of resources may be overcome by cost-sharing. Data sharing would imply statistical significance, which in this context, is a new concern for architects, and hence worth covering now.

The paper focuses on the less-documented aspects of data management and sampling issues, which, for the benefit of a broader audience, can be addressed through basic statistics. Firstly, I briefly describe the configurational approach; secondly, I discuss covariance analysis and the issues of spatial distribution and temporal aggregation of data. Thirdly, I present new evidence of the variance of pedestrian traffic and demonstrate the potential negative impact of the large variance of flow on models. Fourthly, I discuss the procedure for sampling statistically significant traffic data and introduce an advanced method to increase the reliability of small samples; which is followed by conclusions.

2. DATASETS AND METHODS

The configurational approach is based on the analysis of the relationship between network properties of the spatial configuration, such as measures of centrality (e.g integration or choice), and socio-economic phenomena (e.g. traffic or land use).

There is a long history of axial and segment maps being used to model traffic. Reports of significant correlations between traffic and network centrality extend back to the nineteen eighties. An early criticism was that studies conducted in England were geographically bound. While the first major studies did take place in London (UK), several followed from around the world. To name only a few, Peponis et al. studied six Greek cities (Peponis 1989), reporting high correlations in all cases. Read (1999) did similarly in Amsterdam, reporting correlations of 60-70%. There were also studies made in North Africa (Loumi 1988), Brazil (Pereira 2012), Hong Kong (Chu 2005), Seoul (Park 2015), the United States (Raford 2003), and Israel (Lerman, Rofè et al. 2014).

Some of these are major studies, with large samples and statistically significant results. For example, Hillier (1993) measured traffic in 379 street sections. Penn's (1998) study involved pedestrian traffic in 466 locations and vehicular in 397 locations. In both cases, traffic was measured all day, with measurements repeated 20-30 times. In Penn's work, confounding variables were building height, predominant land use and road width. Penn reported correlations between $r = 0.82$ and $r = 0.58$ at $p < 0.0001$ between traffic and integrated centrality measures. Other variables such as land-use, building height and road hierarchy correlated at 0.12, 0.22, and 0.81, respectively. The work by Penn (1998) is of increased relevance because it also reports on a quasi-experimental approach, rather than the traditional purely correlational approach, and thus it has high scientific validity.

The performance of the configurational model can be explained by the hypothesis advanced by the natural movement theory (Hillier, Penn et al. 1993). The theory postulates that asymmetries in the topology of the spatial configuration create unevenness in the accessibility of street segments. Other things being equal, increased availability is likely to result in higher traffic, and land use locates along the network according to this. In turn, land use reinforce traffic patterns.

In most cases, other variables (such as land-use, building height, the number of street lanes, residential and employment density, proximity to public transit, road width, etc.) can increase the performance of the model. However, the cost of including the other variables is often prohibitive. Traditional transport models use such a variety of inputs and are therefore data and labour intensive, and as a consequence rather expensive to build (Weber 2012). On the other hand, Pereira (2012) observes that even traditional data-intensive transportation models (such as the Atkins Saturn), would benefit from including configurational variable(s).

The creation of a configurational model follows five steps. First, draw the axial/segment map with any CAD/GIS package (or within UCL DepthmapX software). Second, import the map to UCL DepthmapX (or access this directly via Quantum GIS Space Syntax Toolkit (Gil et al, 2015)). Third, compute the centrality measures. Fourth, measure traffic and input data. Fifth, plot the relationships in UCL DepthmapX or statistical package. During iterative sessions, changes to the configuration (edited in the software) lead to the repetition of steps 3 and 4.

2.1 CONFIGURATIONAL APPROACH

The configurational approach is based on the analysis of the relationship between network properties of the spatial configuration, such as measures of centrality (e.g integration or choice), and socio-economic phenomena (e.g. traffic or land use).

There is a long history of axial and segment maps being used to model traffic. Reports of significant correlations between traffic and network centrality extend back to the nineteen eighties. An early criticism was that studies conducted in England were geographically bound. While the first major studies did take place in London (UK), several followed from around the world. To name only a few, Peponis et al. studied six Greek cities (Peponis 1989), reporting high correlations in all cases. Read (1999) did similarly in Amsterdam, reporting correlations of 60-70%. There were also studies made in North Africa (Loumi 1988), Brazil (Pereira 2012), Hong Kong (Chu 2005), Seoul (Park 2015), the United States (Raford 2003), and Israel (Lerman, Rofè et al. 2014).

Some of these are major studies, with large samples and statistically significant results. For example, Hillier (1993) measured traffic in 379 street sections. Penn's (1998) study involved pedestrian traffic in 466 locations and vehicular in 397 locations. In both cases, traffic was measured all day, with measurements repeated 20-30 times. In Penn's work, confounding variables were building height, predominant land use and road width. Penn reported correlations between $r = 0.82$ and $r = 0.58$ at $p < 0.0001$ between traffic and integrated centrality measures. Other variables such as land-use, building height and road hierarchy correlated at 0.12, 0.22, and 0.81, respectively. The work by Penn (1998) is of increased relevance because it also reports on a quasi-experimental approach, rather than the traditional purely correlational approach, and thus it has high scientific validity.

The performance of the configurational model can be explained by the hypothesis advanced by the natural movement theory (Hillier, Penn et al. 1993). The theory postulates that asymmetries in the topology of the spatial configuration create unevenness in the accessibility of street segments. Other things being equal, increased availability is likely to result in higher traffic, and land use locates along the network according to this. In turn, land use reinforce traffic patterns.

In most cases, other variables (such as land-use, building height, the number of street lanes, residential and employment density, proximity to public transit, road width, etc.) can increase the performance of the model. However, the cost of including the other variables is often prohibitive. Traditional transport models use such a variety of inputs and are therefore data and labour intensive, and as a consequence rather expensive to build (Weber 2012). On the other hand, Pereira (2012) observes that even traditional data-intensive transportation models (such as the Atkins Saturn), would benefit from including configurational variable(s).

The creation of a configurational model follows five steps. First, draw the axial/segment map with any CAD/GIS package (or within UCL DepthmapX software). Second, import the map to UCL DepthmapX (or access this directly via Quantum GIS Space Syntax Toolkit (Gil et al, 2015)). Third, compute the centrality measures. Fourth, measure traffic and input data. Fifth, plot the relationships in UCL DepthmapX or statistical package. During iterative sessions, changes to the configuration (edited in the software) lead to the repetition of steps 3 and 4.

2.2 COVARIANCE ANALYSIS

Covariance analysis is a statistical method used to check if two variables are associated, which is at the root of configurational-based models. Essentially, it analyses the covariance of the variables, as in do higher values of variable A correspond to higher values of variable B, and lower values of variable A to lower values of variable B - a positive correspondence - direct relationship. Or the opposite - inverse relationship. If the variables show consistency, then values of the dependent variable can be extrapolated from those of the independent variable. Typical independent variables are descriptors of the built environment, such as land use, building height, and measures of network centrality, among others. The dependent variable is (typically) measured traffic.

The strength, or magnitude, of the relationship, is the correlation coefficient (r). A strong relationship[1] is said to be of high covariance, with the value of r approaching 1, while 0 represents the absence of a relationship. The sign describes the direction. A positive r implies a direct relationship, while a negative r implies an inverse relationship. For the purpose of interpretation, it is more useful to calculate the coefficient of determination R2 (R2 = r2). R2 represents the variability shared between the two variables.

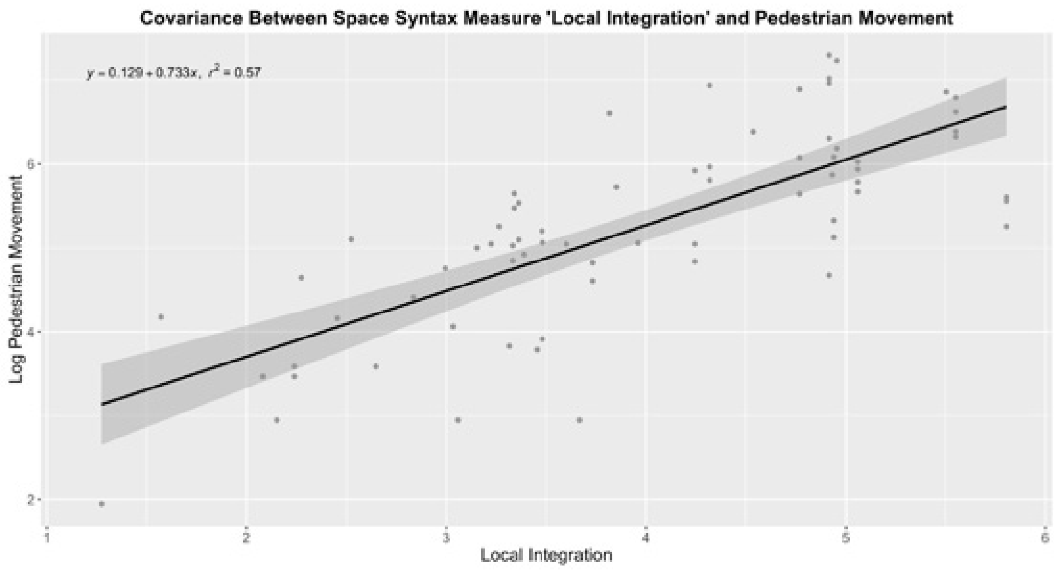

Figure 1 - Correlation and r2 between centrality measure 'local integration' and measured pedestrian traffic for the South Kensington Museum area in London.

The relationship of covariance is typically visualised through a scatterplot (as in Fig. 1), where the x-axis and the y-axis represent the independent and the dependent variables, respectively. Each point represents a spatial unit (street segment). Fig. 1 illustrates a positive correlation between two variables. Without implying a causal relationship, R2 suggests how much of the variability in local integration is shared by the logarithm of pedestrian movement (I explain the reason for using the logarithm of the measurement below). R2 can be multiplied by 100 to obtain a percentage. One can read the result as: local integration can account for 57% of the variation of the logarithm of the pedestrian movement; which implies that 43% of traffic cannot be explained by local integration alone. Notice that for the coefficient of determination (R2) of 0.57, the correlation coefficient (r), = $\sqrt{0.57}$ = 0.75. R2 is a measure of the substantive importance of an effect, not simply the correlation coefficient.

Though the values of r and R2 represent the magnitude of the correlation and can safely be interpreted as mentioned above, this, per se, does not mean that, statistically, the covariance is significant. One can only measure the statistical significance of r by testing a hypothesis; which is done using probabilities. The hypothesis to test is: what is the probability that the correlation is different from zero? In statistical terms, this is called the null hypothesis. Statistical packages

1 Cohen (1988, 1992) suggests 0.5 and larger as a large effect.

do this and present the result as the p-value, which is a measure of the weight of evidence against the null hypothesis. The universally accepted value for declaring statistical significance is over 95%. Therefore, confirming the null hypothesis requires a likelihood > 95%. Consequently, not confirming it requires < 5%. Therefore, the threshold value of p is < 0.05 (0.05 probability or < 5%). In the example of Fig. 1: R2 = 0.57, p = 4.825e-14 (or 0.00000000000004825), and therefore p ≤ 0.0012[2] (statistically highly significant); this means that the probability of the null hypothesis (meaning absence of a relationship) is less than 5%.

Conceptually, it is important to understand that, statistically, this is not a confirmation that the relationship exists. It only means that it is extremely unlikely that it does not exist. Since one cannot know the statistical significance of the effect measured without the p-value, this should always be reported. The R2 and p values are as important as the scatterplot itself. Without these, one cannot be confident that the 'cloud' of points follows the correlation line. Urban planners often use scatterplots as quick diagnosing tools. By identifying outliers, practitioners immediately flag places that are underused or overused about their 'natural' configurational potentials. This flagging system turns out to be rewarding because these points concentrate high potential for improvements. I have often seen little additional local analysis yielding a high level of objectivity for 'surgical' interventions.

Having analysed variance and measured a moderate to high R2 (> 0.3) with acceptable significance (p < 0.05), it is legitimate to use the linear model fitted to the correlation to predict values of y based on x. To do that, one can use the equation of the model (as in the top left corner of Fig. 1). Such process is called regression; it implies working backwards using the model fitted to the data. Regression can be used to estimate (within the probability framework of the model) the amount of traffic on a new street.

The calculation of the correlation coefficient must take into account the type of distribution of the data correlated. For normally distributed data, or large samples, Pearson's correlation coefficient should be used. For not normally distributed data (or for ordered variables), then one should employ Spearman's correlation coefficient. In cases of small datasets and many tied ranks, Kendall's tau correlation should be measured.

Factors to consider when selecting sites to make measurements are: a) choose sites that are at both extremes of the centrality measure, b) add some randomly selected sites, c) include places within the area of the project, and d) include control locations (sites outside of the scope of the project). For how many sites should data be collected? As for any sampling, the principle is the more data, the better. On a more descriptive note, I suggest following the rule of thumb from Green (1991), which basis the recommendation on the number of predictors (k) involved. To test only the R2, the author recommends a minimum of 50 + 8k sites; to test the individual predictors, 104 + k sites. Alternatively, Miles' (2001) suggestions could be followed: a minimum of 30 places for a large effect (r ≈ 0.5)3 and one predictor, and for a medium effect (r ≈ 0.3), 60 sites. The numbers increase for more predictors. Smaller effects (r ≈ 0.1) would require a minimum of 400 sites for one predictor.

2.3 SPATIAL AND TEMPORAL AGGREGATIONS OF DATA

Models of pedestrian movement try to discern regularities in the spatial distribution of a population. They focus on relationships between phenomena on the basis of their location; that is, phenomena are assumed to be related because they co-exist geographically.

A brief analysis of the spatial distribution of traffic data makes it obvious that the movement of population is not evenly distributed. Rather, it is asymmetrically distributed (positively skewed), with the majority of places having comparatively little circulation and very few places having much larger quantities of movement, resembling a logarithmic frequency distribution (Fig. 2). This asymmetry is relevant because to measure the covariance between variables it

2 The American Psychological Association (APA) recommends that the exact p value be reported, unless this is < 0.001, in which case it should be reported as such (APA, 2010). Note that p < 0.001 means rejecting the null hypothesis at a 99.9% confidence level. That is, that there is less than 0.1% chance that the null hypothesis is confirmed.

is important that they follow a similar pattern of distribution. Because the centrality measure of integration approximates a normal distribution, the traffic variable is transformed towards a normal distribution through the application of a logarithmic transformation. Therefore the logarithm of traffic is plotted in Fig. 1. Data transformations such as this are common procedures in mathematics (Games 1984). Fig. 2 illustrates the frequency distribution of the raw values of movement (left) and integration (centre), as well as the effect of the log transformation of movement (right).

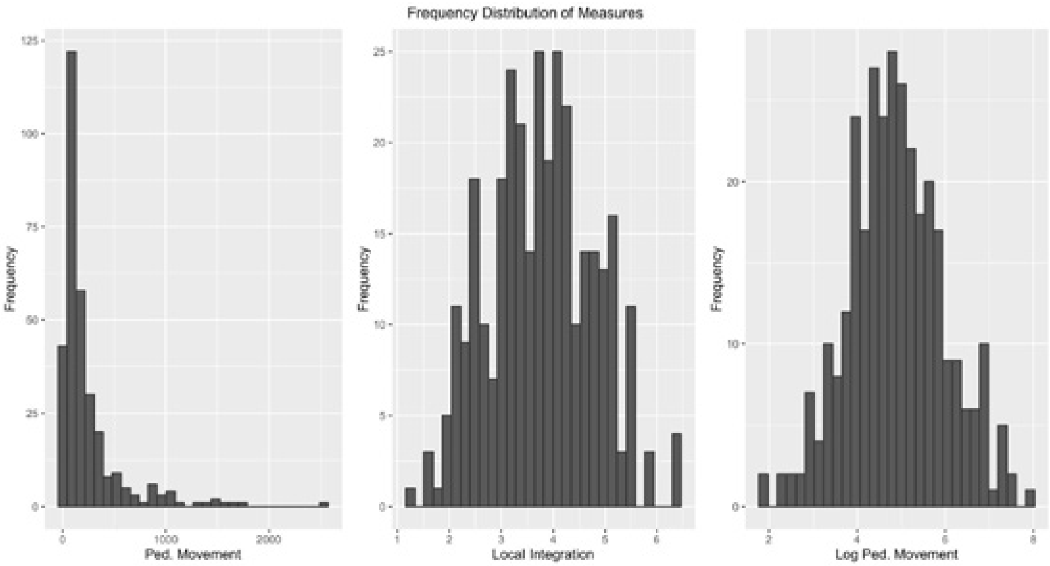

Figure 2 - Left: variable pedestrian movement follows a logarithmic distribution. Center: Local Integration follows a normal distribution. Right: the log transform of pedestrian movement follows a normal distribution.

While the features of the built environment are relatively stable, the flow of movement on the spatial unit empirically varies over time. Therefore, the value that represents the traffic variable is the result of a time-wise aggregation of circulation in that spatial unit. As Fig. 1 illustrates, each point on the scatterplot is represented by one value of y. It is, therefore, necessary to find the one value that describes the overall traffic in each spatial unit. Traditionally this is the mean of the hourly traffic measured throughout the day, as in (Penn 1998, p.70). Stage two is to collect data on enough spatial units to reach statistical representativeness for the overall model. Stage two was explained above; stage one is discussed below.

2.4 MEASURING PEDESTRIAN TRAFFIC

Time-cycles that are empirically observable (e.g., daily and weekly) seem to be seductive regularities that researchers might be tempted to lean on to discard the need for repeating measurements. However, as will be demonstrated, these time-cycles are not as numerically regular as one would infer by simple observation. Therefore, relying on them would compromise the statistical significance of the model. Nonetheless, even when faced with large differences in measured volumes, researchers often dismiss the need for further measurements. The dismissal of the variation is likely to originate with the uncertainty generated by the whole process, where data tell one story, but observable daily patterns seem to suggest another. Furthermore, because traffic is sampled, the differences in measurements can be interpreted as the result of the interpolation process rather than an actual variation of flow. Furthermore, the traditional data collection process, conducted with pen and paper annotations with quick sums on site and posterior data input to spreadsheets, has itself abundant opportunities for data management errors. Overall, the only assumptions made must be very basic or broad, and directly confirmed within the data. These might involve such things such as the shape of the frequency distribution of a population or samples.

In statistical terms, the problem of stage one is one of creating samples to represent a population. The first step is to define the population, which would be the number of pedestrians who move through the spatial unit within the timespan that one aims at characterising. To portray one day of traffic, for example, measurements should be repeated on 30 days. The repetition implies the statistical notion of sampling. The statistical aspect implies the Central Limit Theorem (CLT) (Feller 1968, Feller 1971), which, in simple terms, postulates that given sufficient measurements of a phenomenon, their distribution tends to be normal. This is significant because the properties of normal distributions are well known. Therefore, demonstrating that the distribution of the samples tends to normality makes the inferences made from the samples to the population well documented, not only regarding specific values but critically also for the margin of error. There is a certain agreement among statisticians that "enough measurements" is approximately 30 (Field, Miles et al. 2012, p.43). With a smaller number of samples, it is likely that normality might not completely reveal itself; therefore, for smaller samples, a t-distribution must be assumed.

For explanatory purposes, I am assuming measuring 30 'typical weekdays', which (to me) begins to be reasonable from an ethical standpoint as it gains statistical validity. For example, to characterise weekdays, one would take measurements on different weekdays (30 in this case). To describe one particular day, e.g., Wednesdays, then 30 Wednesdays should be measured.

Another issue to consider when planning counts is how long to count for, since typically it is not possible to count for the entire period of interest. As a matter of principle, one could assume that the longer the sampling time, the more reliable the count. As reported by Turvey (1987), this direct relationship is the case only up to a certain threshold, after which the increase in time does not provide a proportional increase in reliability. Turvey's data suggest a minimum of ten minutes for sampled counting. However, in low traffic areas (especially where zero counts could be recorded) longer periods should be used. Counts for smaller periods are often used to infer hourly rates. This is done either by using a blind expansion factor (e.g. six for ten-minute counts), or one based on locally measured data. Despite their common use, expansion factors imply error and should only be used if absolutely unavoidable. For more on this practice see Turvey's (1987) comprehensive study. Hankey (2014) discusses scaling factors.

In statistical terms, one measurement of the variation of a sample is the standard deviation (SD), which is an indication of how well the mean (as a model) represents the population. A large SD indicates a large variation from the mean and therefore using the mean to represent the population may imply a significant misrepresentation. A small SD suggests otherwise. If one collects several samples, the measure to which the mean of the sample means represents the mean of the population is the standard error (SE). In general terms, it can be stated that:

Outcome (inferred population) = model + error

If the mean is the model, then the error is the SD, and therefore:

Outcome = mean + SD (or SE[3] as described above)

Knowing the nature of the curve of the distribution of values of a sample is critical when assessing the quality of fit of a model. It is known that, for a normal distribution, 99% of the samples of a population will have their means within ± 2.58 times SD. For a 90% confidence level (CL) this factor is 1.64. The mean plus and minus these factors provide the upper and lower boundaries of the confidence interval (CI). A larger CI range implies a less good fit of the mean to represent the population.

Representing the CI can be done numerically or graphically, as per Fig. 3, which displays error lines and the counts as points. This example includes some equivalent-hour combinations that have been counted only once per gate (for example Friday at 14 hours). Notice that in such cases, the mean coincides with the count, but crucially, there are no error bars. The absence of such is a sign not of the lack of error, but of the insufficiency of data to estimate the error.

3 As previously mentioned, if using the means of several samples, the standard error should be used rather than the standard deviation.

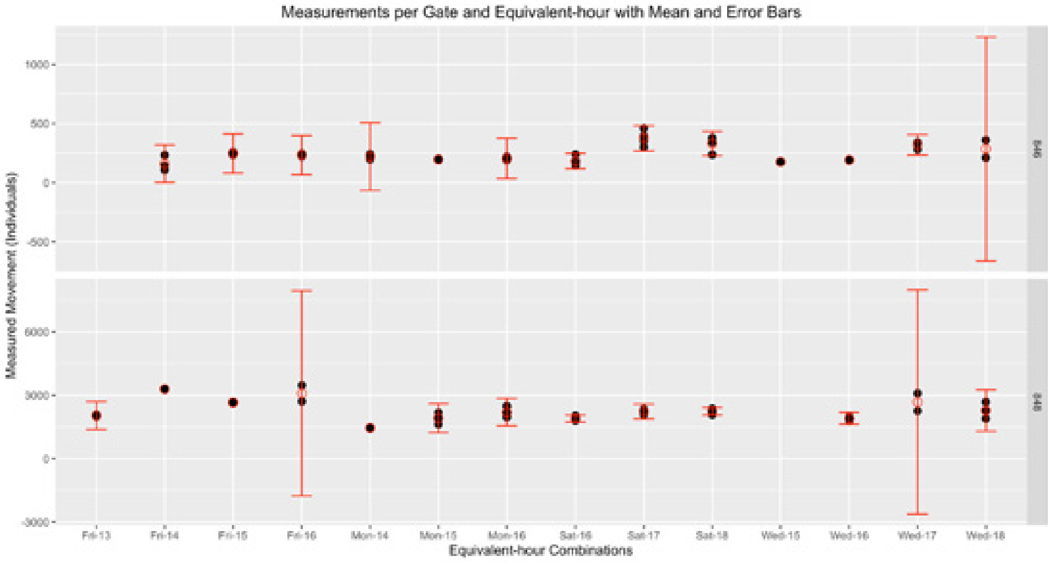

Figure 3 - Count data per equal-hour combination, grouped by gate (846 and 848). Black circles represent count values (three solid circles per equivalent-hour per gate mean that three counts were made. The value of each count can be read on the y axis.). Circumferences represent the means of counts. The lines represent the error for a 95% CL, i.e., the range of the CI for the 95% CL.

3. VARIABILITY OF MEASURED FLOW

Despite several studies referring to pedestrian counts, the perspective is traditionally one of reporting the data and not informing the methodology. One exception is a study by Turvey (1987) focusing on the method of counting and sampling. The study explores data of 15 locations in the United Kingdom to find an ideal sampling duration. The method implied full-time counts for the stipulated periods and an analysis of the variation of different sampling periods from five to forty minutes in five-minute increments. The variation between samples of the same time-span within the same counting interval was as high as ± 50% in the pilot study and reached 40.2% in one case of the main study. Even with the recommended sampling length of 20 minutes, Turvey (1987, p.24) concludes "a coefficient of variation of at least 25% must be assumed".

The considerations about pedestrian counts do not stop with their inherent variability within a particular period. The same study reports that when comparing twenty-minute sample volumes for two different days, each for three periods (for eleven sites, mostly on various weekdays) the highest error measured was -59% (Turvey 1987, p.40). The mean of the absolute values of the error was 48%; the minimum error was 0% and the mean of the absolute error 18.24, with a SD of 16.72. These values suggest a significant variation between movement volumes for different days at the same site and time of observation. The finding seems to defy a common assumption: that traffic at one site repeats itself in precise daily cycles.

Turvey's study reports changes in the same period in several days, but not necessarily the same weekday. Therefore, it is not absolutely conclusive that variation exists among equivalent time periods, i.e., the same weekday at the same time, henceforth called equivalent-hours. A dataset of counts made at two sites in the Kingdom of Bahrain during January and February 2015 can help to clarify this point, as it provides evidence of the variance of measurements within equivalent-hours. The Bahraini dataset does not contain sampling but 57 continuous hours of observations in 20 combinations of equivalent-hours in two sites at different times of the day and on different days of the week. The analysis reveals that measurements of equivalent-hours (taken in the same place on consecutive weeks for the same time-period) yield different results, with considerable variation. Table 1 summarises the results.

Gate ID	Weekday	Hour	N	Week Diff.	Min.	Max.	Mean	SD	% Diff.
848	Wed	16	3	4	1778	1992	1903.33	111.61	7.71
848	Wed	18	3	3	1896	2678	2277.33	391.36	24.17
848	Wed	17	2	2	2258	3092	2675.00	589.73	21.92
848	Fri	13	2	6	1987	2090	2038.50	72.83	3.40
848	Fri	16	2	6	2706	3470	3088.00	540.23	17.20
846	Fri	14	3	6	108	232	161.00	63.93	55.36
846	Fri	15	2	6	234	260	247.00	18.38	7.14
846	Fri	16	2	6	221	247	234.00	18.38	7.55
848	Sat	16	4	4	1792	2028	1893.25	98.29	8.41
848	Sat	17	3	2	2079	2350	2234.33	139.79	8.33
848	Sat	18	4	4	2084	2364	2237.25	115.39	8.57
846	Sat	16	4	4	145	239	181.75	40.26	35.54
846	Sat	17	4	4	301	459	376.25	67.67	29.78
846	Sat	18	4	4	236	378	330.25	64.70	33.41
848	Mon	15	3	3	1627	2176	1914.33	275.40	20.22
848	Mon	16	3	3	1971	2486	2201.00	261.87	16.02
846	Mon	14	2	2	198	243	220.50	31.82	14.08
846	Mon	16	2	2	191	218	204.50	19.09	9.00
846	Wed	17	3	3	279	342	319.00	34.77	14.00
846	Wed	18	2	3	211	360	285.50	105.36	38.11

Table 1 - A summary of the differences in measurements between the twenty equivalent-hour combinations. N is the number of weeks that particular combination was measured; Week Diff. is the number of weeks between first and last measurements; Min. and Max. are the min. and max. values measured, respectively; Mean, SD and Var. are the mean, standard deviation and variance of the measurements, respectively; % Diff. is the percentage difference between the minimum and maximum values.

If the SD is not sufficiently intuitive as an indication of the variability of the data, I include a column with the maximum percentage difference (% Diff.) registered for each equivalent-hour (row). Note that this is not exactly equivalent to SD because it involves only the minimum and maximum values, unlike SD, which comprises all values and their distances to the mean; however, the value as a percentage might provide a more intuitive measure of the range of variation observed. The summary statistics for percentage difference (% Diff.) are as follows: Min. 3.4%, Max. 55.36%, Mean 19.00% SD 13.49%.

The variability of measurements reported suggests that such variation should be expected between measurements in the same place, at equivalent-hours. However, because not all cases were measured the same number of times (between 2 and 4, column N in Table 1), it was plausible (though not likely) that the variation was a function of the number of measurements. A simple correlation suggests no such relationship (Fig. 4). As Fig. 4 and the r2 of 0.068 suggest, the number of times each equivalent-hour was measured does not seem to influence the percentage difference.

In summary, the case study strongly suggest that a high variability between counts of the same phenomenon is to be expected, whether on consecutive days, or equivalent-hours, thus disproving the myth that movement repeats itself in precise daily cycles. The findings also suggest that the distribution curve of samples will be wide, tending to platykurtic, with large standard deviations. These findings imply: a) high variability encourages the use of extensive sampling, and b) when using the mean for a model, one must disclose the error of the distribution, quantified by SD or SE.

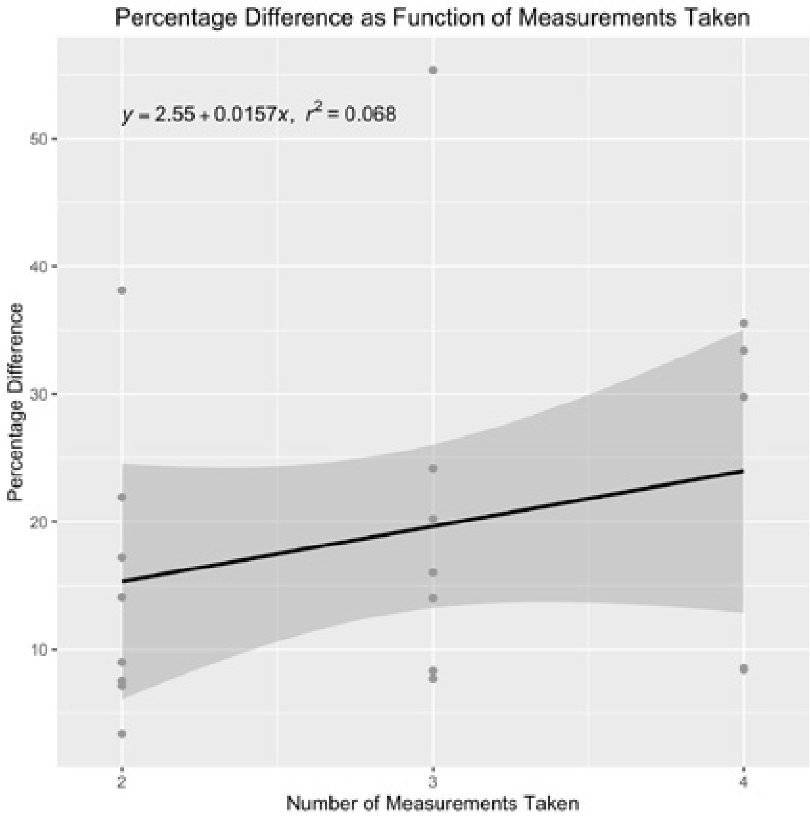

Figure 4 - Scatterplot of the covariance between the number of times each equivalent-hour combination was measured and the percentage difference between the minimum and maximum values. The shaded area represents the 95% CI around the regression line.

3.1 THE POTENTIAL IMPACT OF LARGE VARIATION IN MEASUREMENTS

The wide variance strongly suggests that a reliable account of the traffic that is representative of a place must reflect several repeated measurements. To illustrate this point, I use a benchmark dataset (Jiang 2009) for configurational models. I adopt it for its public availability (at http://discovery.ucl.ac.uk/1232/), and the fact that it has been extensively studied (Carvalho and Penn 2004, Hillier and Iida 2005, Jiang 2009). It contains centrality measures and traffic for four London areas: Barnsbury, Brompton, Calthorpe and South Kensington Museum. The areas were originally chosen to include a “range of different predominant land use types and mixes, densities of development, and street grid morphologies.” (Penn et all, 1998, p.63). All present highly significant correlations between traffic and local integration (Figure 5). The traffic data was gathered through 20-30 repeated measurements, and the four areas have R2 of 0.66, 0.45, 0.59 and 0.57, respectively, all at $p < 0.001$ (left column of Fig. 5). The scatterplots in the right-hand column of Fig. 5 represent the correlations with hypothetical (synthesised) data based on the multiplication of measured values by a random factor based on the percentage difference found in the Bahraini dataset (a random normal distribution with mean = 18.99709 and SD = 13.49427). As a result, the R2 decreases, simulating the potential results of a less-well-documented measurement.

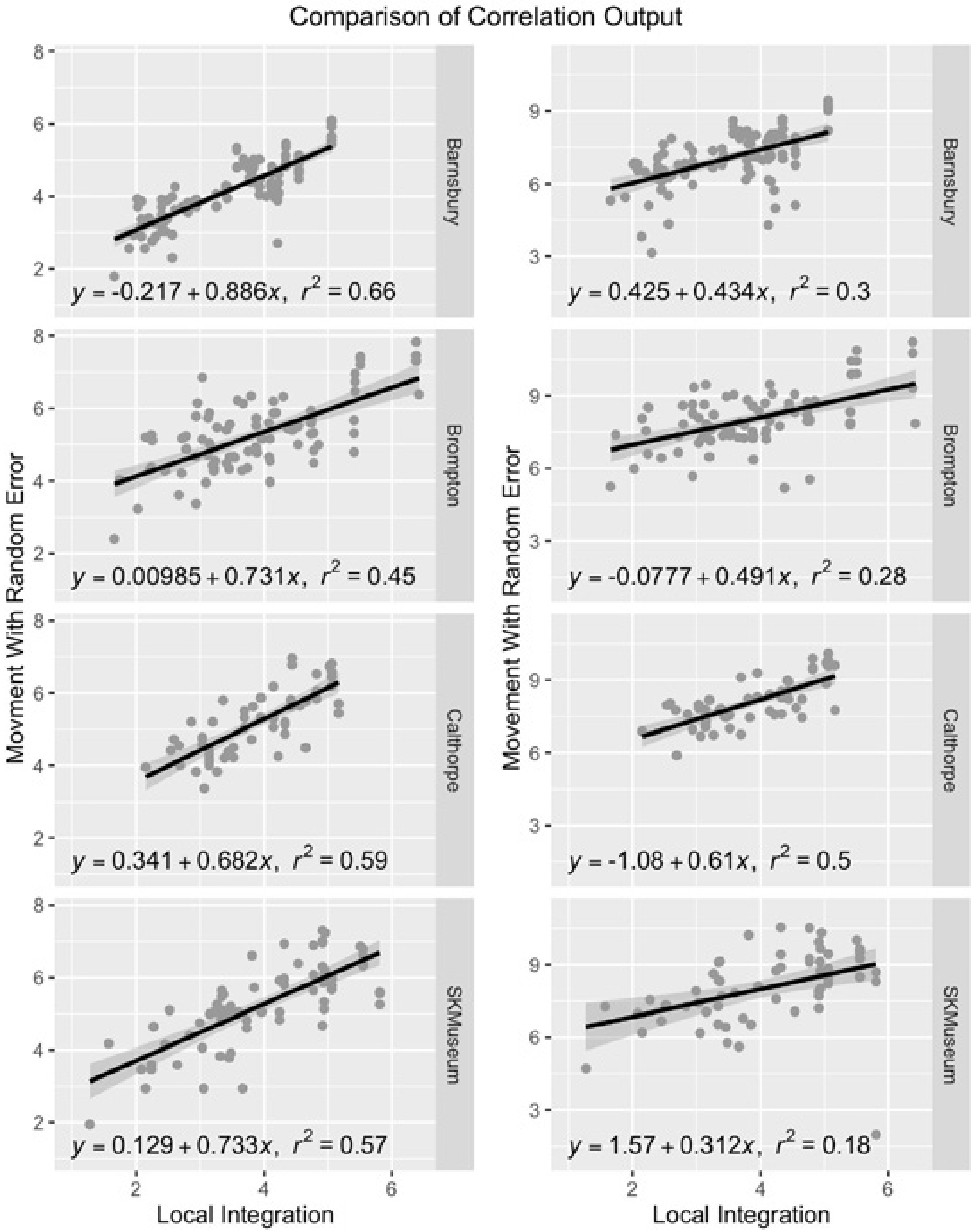

Figure 5 - London dataset. Correlation and r2 for the four London areas. Real values are on the left. On the right are synthesized data based on real traffic but altered by multiplication by a random normal distribution with mean and SD as reported in the study in the Kingdom of Bahrain. Note the shaded areas around the regression lines; they represent the 95% CI.

One can argue that a high R2 achieved between such variables (as centrality measures and traffic) represents a high level of order that is phenomenological relevant; implying that the traffic values have 'captured the essence' of circulation through the network. Such argument would be a reasonable assumption; after all, notice that with syntactic data (Fig. 5), none of the initial correlations improved when randomness was introduced. Nonetheless, if the supporting data are not statistically significant, then the assumption might be invalidated.

4. SUGGESTIONS FOR REACHING A REPRESENTATIVE MEASUREMENT

The mean of the population is inferred from the mean(s) of the sampled movement. The inference implies uncertainty, and therefore a confidence level (CL). To measure the confidence interval (CI), it is necessary to know the shape of the distribution. The relevance of the number of samples is obvious in the formula (1) for standard deviation (similar for SE):

$$s = \sqrt{\frac{\sum (x_i - \bar{x})}{N - 1}} \quad (1)$$

Where is standard deviation, are sampled values, is the mean of the sample and the number of elements in the sample. As per the equation, a larger denominator yields a smaller standard deviation. One obvious way of achieving this is to increase the sampling (N). The magnitude of the effect depends on the tendency to the mean.

If, despite a sample of 30 counts, the SD is still large, one must accept that the mean is not a good fit for the data, and that should be made clear as an indication of the limitations of the model. This can be shown graphically, as illustrated in Fig. 3, or by calculating the boundaries for a certain confidence level. As an illustrative example, I take a case reported in Table 1 (row 2). The measurements imply that it can be assumed that, if 100 samples were taken, in 99 of those samples the estimated mean of the population would be the mean of the sample ± 2.575 times its SD, or 1903.33 ± 287.40 = 16015.93 and 2190.73, a difference of approximately 575 persons. This is approximately 25% of the mean - this is a CI (also called margin of error) of ± 12.5% (at 99% CL). A lower CL would imply a lower range CI.

Note that with a non-normal sampling distribution (or less than 30 measurements), the boundaries of the CI can be calculated based on a t-distribution. Factors for this can be found in tables that are easy to find in statistics books. To choose the correct factor, one uses the degrees of freedom (DF = N-1). As illustrated by such tables, the higher the DF, the lower the factor, and therefore the narrower the confidence interval for a CL, again highlighting the importance of repeated measurements.

Alternatively, when it is not possible to make 30 measurements per event, a more advanced statistical method can be used to simulate a larger number of samples - Ordinary Nonparametric Bootstrap, or Bootstrapping for short, which can be easily performed with statistical software. Bootstrapping is a resampling technique. It implies repeatedly using the existing samples to simulate new samples. For more on bootstrapping methods please see (Boos 2003).

If one aims at characterising weekly movement at each of the two gates previously mentioned in the Bahrain dataset, one could calculate the mean for each gate based on all the samples available, and use that value on the model. Since the measurements were taken at different times and in different days, it can be argued that they represent the weekly movement. In this case, gates 846 and 848 have 31 and 32 measurements, respectively, and therefore the sample sizes are adequate to, through descriptive statistics only, classify movement on each gate. To illustrate how bootstrapping works, if there were only ten samples for each gate (rather than the 31 and 32 available), one could simulate more by creating bootstrap samples. Bootstrap randomly draws (with repetition) from our ten real samples, for each gate.

Traditionally, this is done thousands of times, and the result aggregated by a statistic, the mean in this case. In other words, rather than calculating the mean of the ten (real) samples, one calculates the mean of thousands of (bootstrap) samples. Fig. 6 illustrates both bootstrapped and non-bootstrapped results for 10,000 simulations (bootstrap samples) for each gate, considering several (3, 5, 10 and 30) initial sample sizes. Note in Fig. 6 that, as the number of 'real' samples increases, so does the 'ability' of the bootstrap to compensate for the lack of data. As Fig.6 suggests, its performance degrades rapidly as the number of real samples diminishes. When bootstrapping is computed, the standard error is also included, and this should also be reported.

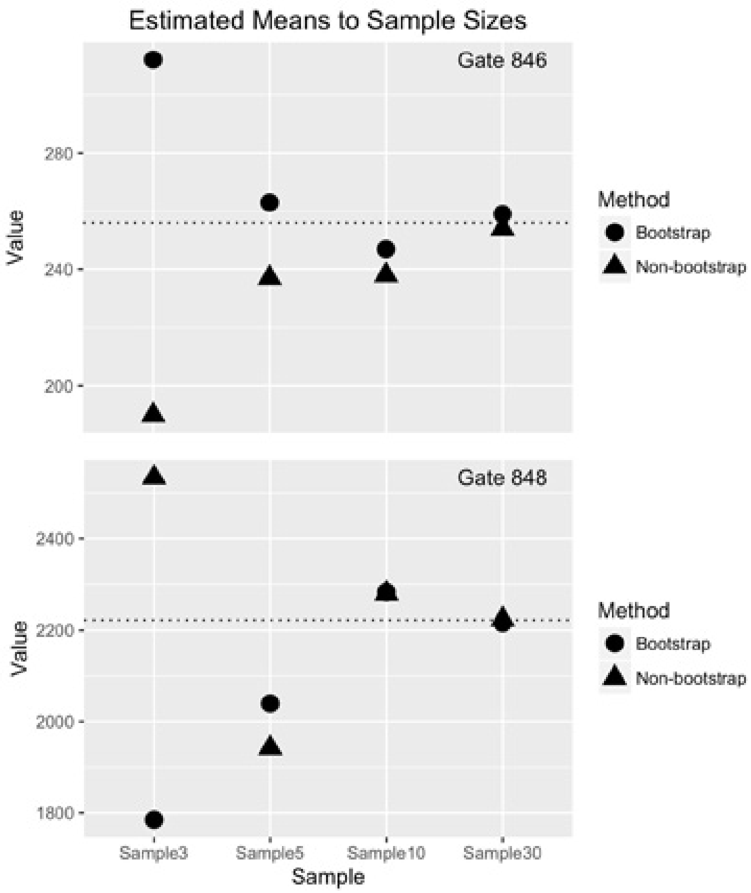

Figure 6 - Plot showing the regression to the mean (dotted lines) of the 32 and 31 samples of gates 846 and 848 respectively. On the X axis, Sample, indicates N number of real samples from where 10,000 bootstrap samples were created. Round markers represent the means of each group of bootstrap samples. Triangular markers dots represent the means of only the N real samples.

5. CONCLUSION

In this paper, I suggest that transforming informal observation methods into formal ones can have a significant impact on the validity of configurational models, which would improve their reliability and widen their usage, and could increase their economic feasibility while expanding the territory covered. I suggest that the minimum threshold for such a transition would be reaching statistical significance. I demonstrate that this is a reachable goal with a resource to basic statistics only, and therefore can be widely applied by designers.

Furthermore I disprove the assumption that traffic flow in equivalent-hours in any one place shows little variability is incorrect, and demonstrate the fallout of a robust model when randomness is introduced by simulating poor traffic sampling. Based on the central limit theorem, I suggest that to reach a statistically significant model, a minimum of 30 repeated measurements are used to characterise a spatial unit. I also suggest several rules of thumb anchored in the literature for establishing the critical minimum number of sites on which traffic is measured. I describe why models should document their limits through disclosing the R2, r, p value, CI and CL for traffic data. Finally, I provide an example of the use of bootstrapping as means to overcome small sampling.

ACKNOWLEDGEMENTS

This work was funded by the Deanship of Scientific Research at the University of Bahrain, grant 12/2014.

REFERENCES

(2010). Publication manual of the American Psychological Association. Washington, D.C., American Psychological Association (APA).

Alexander, C. (2002). The nature of order an essay on the art of building and the nature of the universe. New York, Oxford University Press.

Alexander, C., S. Ishikawa and M. Silverstein (1977). A pattern language towns, buildings, construction. New York, Oxford University Press.

Batty, M. (2001). "Agent-based pedestrian modeling." Environment and Planning B-Planning & Design 28: 321-326.

Batty, M. (2003). Agent-Based Pedestrian Modelling, Centre for Advanced Spatial Analysis , University College London, 1-19-Torrington Place, London WC1E 6BT, UK. 61.

Boos, D. D. (2003). Introduction to the Bootstrap World. *Statistical Science*, 18(2), 168–174. http://doi.org/10.1214/ss/1063994971

Carvalho, R., Penn, A., 2004, Scaling and Universality in the micro-structure of urban space. Physica A, 332, pp. 539-547.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, N.J., L. Erlbaum Associates.

Cohen, J. (1992). "A power primer." Psychological Bulletin 112(1): 155-159.

CHU, S. C. H. When and Why do People Walk in the City: The influence of urban elements on time-pattern of pedestrian movement. Sixt International Walk 21 Conference, 22-23 September 2005 2005 Zurich.

Dalton, R. C. (2003). "The secret is to follow your nose - Route path selection and angularity." Environment and Behavior 35(1): 107-131.

Desyllas, J. (1999). When downtown moves: Quantifying, representing and modelling the spatial variable in office rents, Brazilia, 1999, Second International Symposium on Space Syntax.

Feller, W. (1968). Introduction to probability theory and its applications volume 1. [S.l.], John Wiley and Sons Inc.

Feller, W. (1971). Introduction to Probability Theory and Its Applications Volume II. [S.l.], John Wiley and Sons.

Field, A. P., J. Miles and Z. Field (2012). Discovering statistics using R. London, SAGE.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness." Sociometry (40): 35 - 41.

Games, P. A. (1984). "Data transformations, power, and skew: A rebuttal to levince and Dunlap." Psychological Bulletin 95(2): 345.

Gehl, J. (2006). Life between the buildings : using public space. Copenhagen, Danish Architectural Press.

Gehl, J. (2006). New city life. Copenhagen, Danish Architectural Press.

Gehl, J. (2010). Cities for People. Washington, Island Press.

Gehl, J. and L. Gemzøe (1999). Public spaces - public life : Copenhagen 1996. Copenhagen, Danish Architectural Press.

Gehl, J. and L. Gemzøe (2000). New city spaces. Copenhagen, Danish Architectural Press.

Gehl, J. and B. Svarre (2013). How to study public life. Washington, Island Press.

Gil, J., Varoudis, T., Karimi, K., Penn, A. (2015). The space syntax toolkit: Integrating depthmapX and exploratory spatial analysis workflows in QGIS; Proceedings of the 10th International Space Syntax Symposium; London.

Green, S. B. (1991). "How Many Subjects Does It Take To Do A Regression Analysis." Multivariate Behavioral Research 26(3): 499-510.

Hall, E. T. (1980). The silent language. Westport, Conn., Greenwood Press.

Hall, E. T. (1990). The hidden dimension. New York, Anchor Books.

Hankey, S. L., G.; Marshall, J. (2014). Day-of-Year Scaling Fcators and Design Considerations for Non-motorized Traffic Monitoring Programs. Transportation Research Borad 93rd Annual Meeting, Washington, D.C.

Hankin, B. D. and R. A. Wright (1958). "Passenger Flow in Subways." OR 9(2): 81-88.

Hillier, B. (1996). Space is the machine : a configurational theory of architecture, New York : Cambridge University

Press.

Hillier, B. and J. Hanson (1984). The social logic of space, Cambridge : Cambridge University Press.

Hillier, B. and S. Iida (2005). Network and Psychological Effects in Urban Movement. Spatial Information Theory. A. Cohn and D. Mark, Springer Berlin Heidelberg. 3693: 475-490.

Hillier, B., A. Leaman, P. Stansall and M. Bedford (1976). "Space syntax." Environment and Planning B: Planning and Design - Pion Ltd 3: 147-185.

Hillier, B., A. Penn, J. Hanson, T. Grajewski and J. Xu (1993). "Natural Movement - Or, Configuration and Attraction in Urban Pedestrian Movement." Environment and Planning B: Planning and Design - Pion Ltd 20(1): 29-66.

Hoogendoorn, S. P. (2003). Pedestrian Travel Behavior Modeling. 10th International Conference on Travel Behavior Research, Lucerne, Switzerland.

Hoogendoorn, S. P. and P. H. L. Bovy (2004). "Pedestrian route-choice and activity scheduling theory and models." Transportation Research Part B: Methodological 38(2): 169-190.

Jacobs, J. (1961). The death and life of great American cities. New York, Vintage Books.

Jiang, B. (2009). "Ranking spaces for predicting human movement in an urban environment." International Journal of Geographical Information Science 23(7): 823-837.

Karimi, K. (2012). "A configurational approach to analytical urban design: /`Space syntax/' methodology." Urban Des Int 17(4): 297-318.

Kim, Y. O. and A. Penn (2004). "Linking the spatial syntax of cognitive maps to the spatial syntax of the environment." Environment and Behavior 36(4): 483-504.

L., H. F. (1971). "The statistics of crowd fluids." Nature(229): 381 - 383.

Lerman, Y., Y. Rofè and I. Omer (2014). "Using Space Syntax to Model Pedestrian Movement in Urban Transportation Planning." Geographical Analysis 46(4): 392-410.

Loumi, A. (1988). Spatial Configuration in Towns in North Africa. PhD, University College London.

Lynch, K. (1964). The image of the city. [S.l.], MIT Press.

Miles, J. and M. Shevlin (2001). Applying regression & correlation : a guide for students and researchers. London, SAGE.

Modesti, P. and A. Sciomachen (1998). "A utility measure for finding multiobjective shortest paths in urban multimodal transportation networks1." European Journal of Operational Research 111(3): 495-508.

Navin, F. P., R. J. Wheeler (1969). "Pedestrian flow characteristics." Traffic Engineering, Inst Traffic Engr 39.

Park, H.-T. (2015). The spatio-temporal routine of pedestrian movements in large cities: the case of Seoul. 10th International Space Syntax Symposium, London.

Penn, A., Hillier, B., Banister, D., Xu, J. (1998). "Configurational modeling of urban movement netwroks." Environment and Planning B-Planning & Design 25: 59.

Peponis, J. H., E.; Livieratos, C.; Fatouros, D. A. (1989). "The spatial core of urban culture." Ekistics 334(335): 43-55.

Pereira, R. H., Frederico; Medeiros, Valério; Barros, Ana (2012). The use of space syntax in urban transport analysis: limits and potential. 8th International Space Syntax Simposium, Santiago Chile.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for

Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raford, N. (2003). Looking both ways: Space Syntax for pedestrian exposure forecasting and collision risk analysis. 4th International Space Synatx Conference, London.

READ, S. 1999. Space syntax and the Dutch city. Environment and Planning B: Planning and Design - Pion Ltd, 26, 251-264.

Ryus, P. (2014). Guidebook on pedestrian and bicycle volume data collection. Washington, DC, Transportation Research Board of the National Academies.

Ryus, P. F., Erin; Laustsen, Kelly; Schneider, Robert; Hull, Tony; Proloux, Frank; Miranda-Moreno, Luis; (2014). Methods and Technologies for Pedestrian and Bicycle Volume Data Collection. USA, NCHRP - Transportation Research Board.

Schadschneider, A. (2002). "Traffic flow: a statistical physics point of view." Physica A: Statistical Mechanics and its Applications 313(1–2): 153-187.

Turner, A. (2001). Angular analysis. 3rd International Symposium on Space Syntax, Georgia Institute of Technology, Atlanta, GA.

Turner, A. (2004). Depthmap 4. Univeristy Colege London.

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 & Design 34(3): 539-555.

Turner, A. P., Alan; Hillier, Bill (2005). "An algorithmic definition of the axial map." Environment and Planning B-Planning & Design 32: 425-444.

Turner, A. P. 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 & Design 29: 473-490.

Turvey, I. G., May, A.D., Hopkinson, P.G. (1987). Counting Methods and Sampling Strategies Determining Pedestrian Numbers. Leeds, UK, Institute of Transport Studies, University of Leeds.

Weber, R. L., J. (2012). Modelling urban systems. The Oxford Handbook of Urban Planning. R. L. Weber, J. Oxford, UK, Oxford University Press.

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Whyte, W. H. (1988). City: Rediscovering the center. New York, Doubleday.

Whyte, W. H. and F. Conservation (1980). The social life of small urban spaces, Washington, D.C. : The Conservation Foundation.