023

Spatial databases:

Generating new insights on office design and human behaviours in the workplace

Petros Koutsolampros
Spacelab, London; UCL
petros.koutsolampros.11@ucl.ac.uk

Kerstin Sailer
Space Syntax Laboratory, The Bartlett School of Architecture, UCL
k.sailer@ucl.ac.uk

Ros Pomeroy
Spacelab, London
ros.pomeroy@spacelab.co.uk

Martin Zaltz Austwick
Centre for Advanced Spatial Analysis (CASA), The Bartlett, UCL
m.austwick@ucl.ac.uk

Andy Hudson-Smith
Centre for Advanced Spatial Analysis (CASA), The Bartlett, UCL
a.hudson-smith@ucl.ac.uk

Rosie Haslem
Spacelab, London
rosie.haslem@spacelab.co.uk

Abstract

Space Syntax research has shown how human behaviours in the workplace are shaped by spatial configuration; in turn, evidence-based design practices have highlighted ways in which this data can be used to inform tailor-made solutions in office design. Yet, existing research focuses on either single case studies or comparisons of a few cases on a small scale. Also, each study uses its own methods and metrics which makes it difficult to establish wider patterns beyond single datasets. This paper presents a larger than usual data set on workplaces, which has been collected by Spacelab, a design and consultancy practice based in London. This dataset includes spatial and space usage information such as syntactic analysis and desk occupancy on client companies. It resides in a spatial relational database, allowing for systematic combination of the collected data, useful for doing either deeper analysis, or generating benchmarks and baselines. These insights are not only highly relevant to clients but also give rise to opportunities to generate new insights on office design and human behaviours in the workplace from a research perspective. Two main research questions relating to the size of samples are discussed: Firstly, whether large samples are necessary to fully understand phenomena, and secondly, whether behavioural patterns vary across cases. Observation data and syntactic analysis are combined to understand in which areas of an office different activities take place. Observation data is also brought together with the functional allocations of space in order to ask whether activities follow the programme introduced by functions such as meeting rooms, *kitchens, workspaces, etc. It is shown that observation data only becomes robust and reliable with longer periods of observations than previously recommended. Three to four full days seems to produce reasonably stable results for desk occupancy, while five full days seemed required for percentages of people walking and interacting. Some surprising findings were revealed regarding the distribution of activities in space, for instance dispelling the myth that interactions happen in corridors and highlighting that interactions tend to occur in rather segregated spaces. While it is argued that predictive power of the analysis varies, first steps towards establishing generic patterns have clearly been taken.*

Keywords

Workplace, spatial datasets, spatial databases, digitised collection of spatial data.

1. Introduction

Space Syntax research can look back on a tradition of 25 years of research into the relationship between workplace layout and human behaviours. Studies have highlighted that more integrated buildings show higher rates of movement and encounter (Hillier and Grajewski, 1990); that movement flows in offices follow both spatial integration and the placement of attractors (Sailer, 2007); that interactions among staff take place predominantly around desks and workstations (Markhede and Koch, 2007; Rashid et al., 2006; Steen, 2009; Steen et al., 2005); that proximity among co-workers drives frequency of interaction (Sailer and Penn, 2009) but also the structure of interaction networks (Sailer and McCulloh, 2012); that a more intelligible and integrated layout creates denser interaction networks and supported task performance (Peponis et al., 2007); and that integrated spaces enable innovation due to fostering unplanned face-to-face interaction (Penn et al., 1999; Toker and Gray, 2008; Wineman et al., 2013).

This growing understanding of the mechanisms and effects of spatial layout on organisational outcomes is increasingly used in evidence-based design practices (Sailer et al., 2008; 2010), which have highlighted ways in which these insights can be used to inform tailor-made solutions in office design. It has been argued that often cases are unique and generic patterns of influence are difficult to establish (Sailer, 2010). Therefore, the ability to successfully predict likely behavioural outcomes just by analysing a floor plan of an office is still not feasible. This is also due to the lack of comparability across different studies. Each study uses its own methods and metrics, which makes it difficult to establish wider patterns beyond single datasets. Also, the majority of existing research focuses on either single case studies or compares a few cases on a small scale.

Hence, this paper presents a larger than usual data set on workplaces, which has been collected by Spacelab, a design and consultancy practice based in London. Data from individual consultancy projects, containing data on both the spatial layout of offices and its organisational behaviours, has been used to create a Spatial Relational Database, which not only offers a digital storage solution, but enables the bringing together of data on two levels: firstly, data within a single project can be combined more systematically, for instance observation data, the syntactic analysis of floor plans and the structure of organisational networks gathered through online surveys. This allows for deeper analysis. Secondly, data from different projects can be brought together to create benchmarks and baselines, which are not only highly relevant to clients in order to understand how their space and behaviours compare to their competitors; they also give rise to opportunities to generate new insights on office design and human behaviours in the workplace from a research perspective. In addition, the database allows for consistency and minimises human error in handling increasingly large and diverse data sets. This paper will use samples currently contained in a spatial relational database, which has been established as part of the Workplace Consultancy of Spacelab.

Two main research questions relating to the size of samples will be discussed: Firstly, it will be asked whether large samples are necessary to fully understand phenomena, or, at which point saturation of data sets in. This question will be analysed using observation data – asking to what degree results change if fewer snapshots are combined. It will also be investigated which metrics are more robust and generic across cases and which ones are more volatile and prone to change from case to case. Secondly, it will be asked whether behavioural patterns vary across cases. Observation data and syntactic analysis will be combined to understand in which areas of an office different activities take place. This is interesting, since a strong variation in patterns would underline the uniqueness of cases, whereas a weak variation in patterns could point towards first steps of a more predictive model and understanding of space and organisation.

The paper is structured as follows: section 2 will contextualise the work of the paper, by reviewing existing approaches which have worked with larger samples of building analysis. Section 3 will introduce the Spatial Relational Database of Spacelab, while section 4 presents the dataset, types of data, methods and metrics used in this paper. Two exemplary questions arising from working with large samples will be answered in the following sections 5 and 6, namely, how resilient is observation data to shorter observation periods, and how are activities in offices distributed across space. The final section of the paper will reflect on the state of the art of workplace research and draw conclusions on limitations and next steps.

2. Spatial databases – Working with large samples

In contrast to research investigating single cases or small samples of workplaces (as mentioned above), most recently scholars have begun to collate bigger samples of buildings in order to generate new research insights. Comparing 50 different offices, Shpuza and Peponis (2008) argued that configurational metrics were systematically affected by the shape of the floor plate.

Analysing a sample of 62 workplaces, Sailer et al. explored overarching patterns that differentiated generative buildings from conservative buildings (Sailer et al., 2012). The sample consisted of floor plans, which were investigated using Visibility Graph Analysis (VGA), however, data on generativity was only available for a limited number of cases and drew on a variety of methods and metrics, highlighting the difficulties of creating larger samples from a diverse evidence base.

Two recent papers investigated generic spatial properties of layouts of different building types, among them offices, but also museums, libraries, hospitals, shops and religious buildings. Using a sample of 67 floor plans, Peponis (2012) analysed patterns of intelligibility and its potential effects on cognition, while Abshirini and Koch (2013) focused on visibility properties of building layouts and typological explorations using a sample of 98 buildings.

Despite not working with a large sample, another piece of research seems noteworthy, since it combines different types of data in an innovative fashion. Using syntactical metrics alongside organisational analysis, Derix and Jagannath (2014) combine occupational affordances and morphology into a single representation and profile of a building. This is an interesting approach in the context of this paper, since it uses visualisation tools to combine different metrics in order to gain new levels of understanding of the complex interplay of different spatial and organisational variables.

In summary, existing research has followed either one of these strategies: 1) Single or few cases bringing together spatial and organisational data in traditional correlational or other statistical analysis (see references mentioned in the introduction); 2) Single or few cases bringing together spatial and organisational data in new ways (Derix and Jagannath, 2014); 3) Large samples, but syntactical analysis only (Shpuza and Peponis, 2008; Peponis, 2012; Abshirini and Koch, 2013); 4) Large samples, but incomplete datasets (Sailer et al., 2012).

Hence this paper provides a novel approach of working with a larger than usual sample, bringing together both spatial and organisational data in a complete, consistent and coherent manner.

3. Spatial relational database

The dataset examined has been generated over the course of two years at Spacelab as a Knowledge Transfer Partnership project between Spacelab and UCL's Space Syntax Laboratory. It currently resides in a PostgreSQL database with the spatial extension PostGIS. This database brings together spatial information (floor plans, syntactical analysis, functional distribution, team allocations), behavioural and space usage information (desk and meeting room occupancy, locations of people standing, walking, sitting and interacting) and organisational information (networks of interaction on individual and team level, organisational structure, organisational and team cultures).

4. Big Spatial Data – Examined sample, methods and metrics

The sample of this analysis is 27 cases. The study's focus is behaviour and usage in office space; thus, different offices in different locations belonging to the same company have been considered standalone cases. In total, data from 14 companies in a variety of industries such as Media, Advertising, Legal, Technology, Retail and Financial Services is analysed. The cases vary in size, from 400 to 15000 m² of office area and 40 to 1700 staff. This paper discusses three different types of data.

Firstly, activity and occupancy data derived from participant observations (Grajewski, 1992) is used. Using mobile devices, observers recorded positions and activities of staff (standing, sitting, walking, interacting) in the company space, in hourly intervals (typically) 8 times a day over the course of 5 days.

Secondly, a functional categorisation of a company's different rooms and spaces was undertaken, distinguishing open plan and cellular workspaces, alternative workspaces (e.g. breakout spaces), meeting rooms, primary circulation and other facilities (e.g. kitchens, canteens, tea points).

Thirdly, floor plans were investigated producing a Visibility Graph Analysis (VGA) with depthmapX (Varoudis, 2012). The main metric used is Mean Depth (MD) based on visibility, i.e. an eye-level analysis.

5. How resilient is observation data to shorter observation periods?

Observation of space usage is one of the key methods of the Space Syntax analysis of buildings. The standard Observation Manual (Grajewski, 1992) defines five time periods (morning, mid-morning, lunch, early afternoon, later afternoon) and suggests observing each building area twice in each period and over two distinct working days. This procedure results in 20 distinct snapshots for each area. Since snapshots represent one moment in time and are rather volatile to variation (for instance due to an event, like a large staff meeting), it makes sense to collect larger samples. However, it has never been tested systematically how large this sample size should be in order to be both valid, but also resource-efficient.

Spacelab's database provides an excellent opportunity to analyse variations in observation data as a result of varying lengths of observation. The aim is to establish how many snapshots (or how many days of observation) will be needed to provide a robust picture of activities. In detail, three different metrics are explored: 1) desk occupancy (% of desks occupied at any point in time), 2) people walking (% of people walking of all people present) and 3) people interacting (% of people interacting of all people present).

19 cases held the full set of 5 days of observation data. For each case, the overall result of desk occupancy, people walking and people interacting (using the aggregate from 5 days of observations) was compared to the results, if only 4, 3, 2 or 1 days of data was used. All possible combinations of 1, 2, 3 and 4 days of data were produced and the coefficient of variation (CV, Standard deviation divided by the mean) for each metric was calculated to highlight the difference from the aggregated 5 day result. In the case of people walking and interacting, weighted standard deviation was used for the calculation of CV to account for the varying number of people in space.

Figure 1 plots the coefficient of variation across projects for desk occupancy. As expected, the results indicate rather large variations of for one day of observation data, with decreasing variations, when more days are aggregated. The deviation for using only one day of data ranges from 2.9% to 30.8% of the mean, for two days, 1.8% to 18.7%, for three 1.0% to 10.0% and for four days 0.7% to 7. 7%.

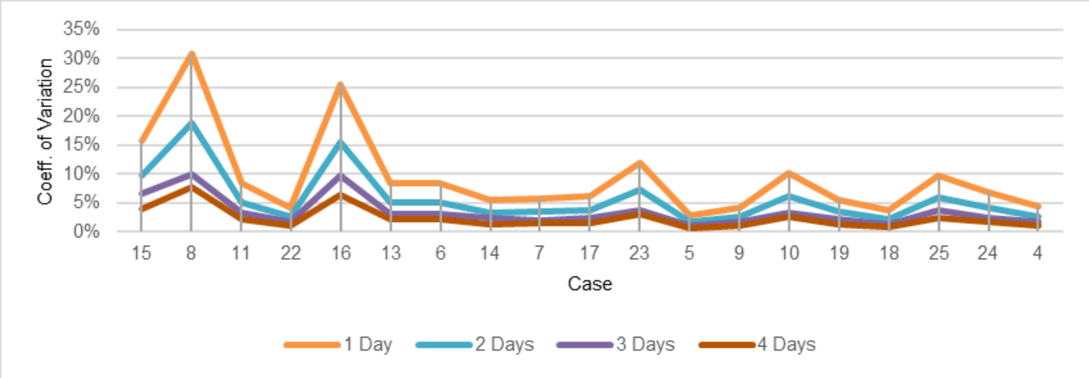

Figure 1: Coefficient of variation of the desk occupancy across projects for different number of days. The cases are in ascending order according to the number of desks.

Two cases in particular show rather large deviations of more than 25% of the mean (8 and 16) with three more cases (23, 10 and 25) showing a CV of more than 10%, if only one day of observation data was considered. The standard approach suggested in the Space Syntax Observation Manual (i.e. 2 days) can still yield rather large deviations (up to 19% of the mean). Only with a minimum of 3 days aggregated, a converging of results appears.

Case 8 is the example with the largest variations when doing the observation for less than 5 days. This case specifically had lower general occupancy across two of the five days in the week, for unknown reasons (see Table 1). Unless there is a hidden pattern that relates to company culture, such as “Friday and Thursday are work from home days”, doing the observation on any of these two days would have yielded a very different understanding of how the company works.

Case	No. of Desks	Mon.	Tue.	Wed.	Thu.	Fri.	VC	Mean
8	44	40%	36%	37%	21%	17%	31%	30%
22	62	39%	40%	36%	38%	41%	4%	39%
19	564	35%	35%	34%	34%	30%	6%	34%

Table 1: Sample of single-day occupancy rates for 3 different cases for each day including the Coefficient of Variation (CV) and the Mean; Case 8 (marked in red) has a high CV due to the drop in occupancy in the last two days.

The analysis for the other two measures - the percentage of people walking and interacting returns similar results with comparable CV values of up to 38% (walking) and 23% (interacting) as shown in an overview in table 2. The percentage of people walking seems even more volatile to a shorter observation period due to the range of the percentage of people walking itself being low, thus even with a relatively low deviation, the data collected can vary greatly.

	Desk Occupancy				Walking				Interacting			
Case	1 Day	2 Days	3 Days	4 Days	1 Day	2 Days	3 Days	4 Days	1 Day	2 Days	3 Days	4 Days
15	15.8%	9.7%	6.6%	4.0%	18.5 %	11.6%	8.4%	4.8%	5.0%	3.2%	2.5%	1.4%
8	30.8%	18.8%	10.0%	7.7%	27.8 %	14.7%	9.7%	5.7%	12.3 %	7.2%	3.5%	2.9%
11	8.4%	5.2%	3.3%	2.1%	33.2 %	20.3%	14.2%	8.3%	21.5 %	13.1%	9.4%	5.3%
22	4.2%	2.6%	1.9%	1.0%	37.9 %	23.3%	18.1%	9.5%	6.8%	4.1%	2.5%	1.7%
16	25.5%	15.6%	9.8%	6.4%	11.3 %	6.7%	4.5%	2.7%	23.3 %	14.3%	8.8%	5.9%
13	8.4%	5.2%	3.1%	2.1%	20.8 %	12.6%	8.3%	5.1%	9.0%	5.6%	4.2%	2.3%
6	8.5%	5.2%	3.1%	2.1%	25.8 %	15.6%	9.5%	6.3%	12.2 %	7.6%	3.8%	3.1%
14	5.5%	3.4%	2.5%	1.4%	12.9 %	7.9%	5.5%	3.2%	13.2 %	7.8%	4.7%	3.1%
7	5.8%	3.6%	1.9%	1.5%	30.7 %	19.1%	13.3%	7.9%	5.3%	3.3%	2.5%	1.4%
17	6.2%	3.8%	2.3%	1.6%	13.7 %	8.4%	4.4%	3.4%	7.6%	4.7%	3.2%	1.9%
23	12.0%	7.3%	3.8%	3.0%	14.4 %	8.9%	6.1%	3.6%	13.0 %	8.0%	4.7%	3.2%
5	2.9%	1.8%	1.0%	0.7%	16.8 %	10.4%	6.6%	4.3%	5.9%	3.6%	2.3%	1.4%
9	4.3%	2.6%	1.8%	1.1%	20.7 %	12.7%	6.5%	5.2%	9.5%	5.8%	4.0%	2.4%
10	10.3%	6.3%	3.3%	2.6%	38.7 %	23.5%	12.3%	9.5%	8.6%	5.3%	3.0%	2.1%
19	5.6%	3.4%	2.1%	1.4%	10.0 %	6.0%	3.8%	2.4%	10.5 %	6.5%	4.2%	2.7%
18	3.7%	2.3%	1.4%	0.9%	16.0 %	9.7%	5.5%	3.9%	8.6%	5.2%	3.2%	2.1%
25	9.7%	5.9%	3.7%	2.4%	10.8 %	6.6%	4.1%	2.7%	4.5%	2.8%	1.8%	1.1%
24	6.9%	4.2%	2.5%	1.7%	1.9%	1.2%	0.9%	0.5%	6.6%	4.1%	2.4%	1.7%
4	4.5%	2.8%	1.8%	1.1%	13.5 %	8.4%	5.1%	3.4%	3.2%	1.9%	1.4%	0.8%
Aver age	9.4%	5.8%	3.5%	2.4%	19.8 %	12.0%	7.7%	4.9%	9.8%	6.0%	3.8%	2.4%

Table 2: Coefficient of variation across projects, and three different measures: Desk occupancy, Percentage of people walking and Percentage of people interacting.

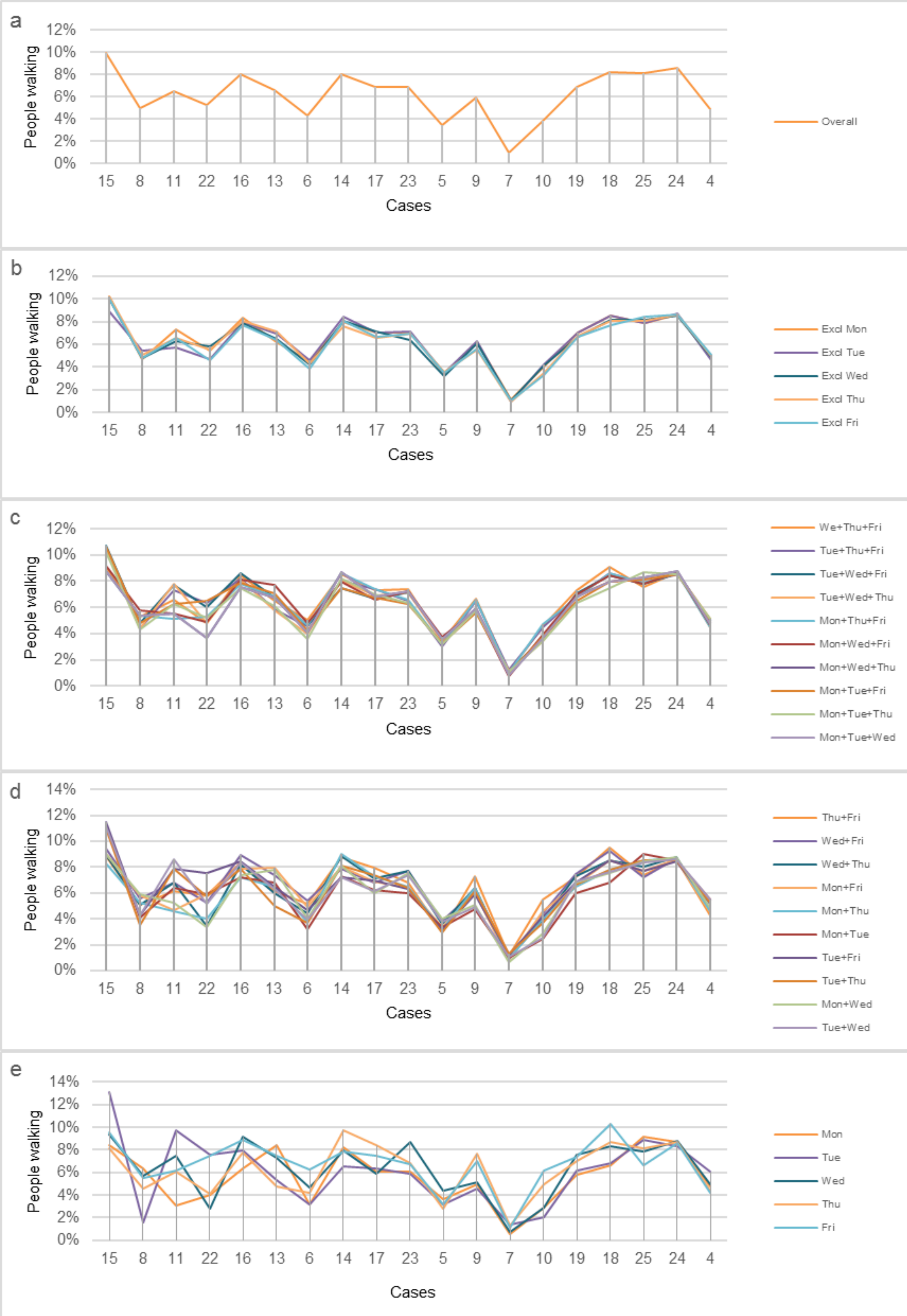

Figure 2: Coefficient of variation of the percentage of people walking: a) Actual (5 days), b) 4 days, c) 3 days, d) 2 days and e) single day. Each line represents one possible combination of days.

This analysis showed that the resilience of the dataset varies with the number of days of observation, as well as the metric in question. Desk occupancy and percentage of people interacting are high-volume, and suffer less than the percentage of people walking (Table 2), which is a less common observation.

Overall, it can be concluded that in order to comment confidently on desk occupancy and percentage of people interacting at least 3-4 days of observation are required, where the average deviation is less than 5% of the actual mean and the worst case 10%. For percentages of people walking 5 days are needed to reach valid conclusions as even for 4 days the average deviation is 4.9% of the mean.

6. How are activities in offices distributed across space?

The distribution of social behaviours in space has been at the heart of Space Syntax research, where it has long been argued that integrated spaces attract more activities such as movement or encounter (Hillier, 1996). Specifically in offices is has been argued that the introduction of a programme and organisational rules (Hillier and Penn, 1991) may deviate movement flows and occupation away from configurational logic (Sailer, 2007). This raises two questions: firstly, which activities seek rather integrated or rather segregated locations? And secondly, which functional areas attract activities due to their programme?

The advantage of answering those types of questions on the basis of a large database lies in the ability to compare across cases and understand variations. If only minor variations were found between cases, it could be argued that a generic pattern emerged, which would lend itself to predictions. Simulating and modelling likely social behaviours from a floor plan is one of the hallmarks of the Space Syntax approach, yet in buildings, particularly offices, the lack of consistent outcomes still renders this impossible. Therefore, making steps towards a more powerful predictive model is an important contribution.

In order to answer the first question, average MD values from the syntax analysis of floor plans were retrieved for each observed activity (standing, walking, interacting and sitting). Since sitting highly coincides with the locations of desks, only people sitting outside of pre-allocated desks were taken into consideration for the analysis.

Since a VGA produces a 2D grid of pixels at discrete intervals, it had to be transformed in order to be matched with observation data, which is continuous, i.e. can happen anywhere in space. Therefore, the discrete VGA values went through a resampling algorithm in order to generate the values at the points of the observation data. The resampling algorithm uses the Lanczos kernel with a window of 2 (see an example in Figure 3). Turkowski and Gabriel (1990) claimed that this kernel and window combination provide the "best compromise in terms of reduction of aliasing, sharpness, and minimal ringing" for decimation and interpolation of 2-dimensional image data.

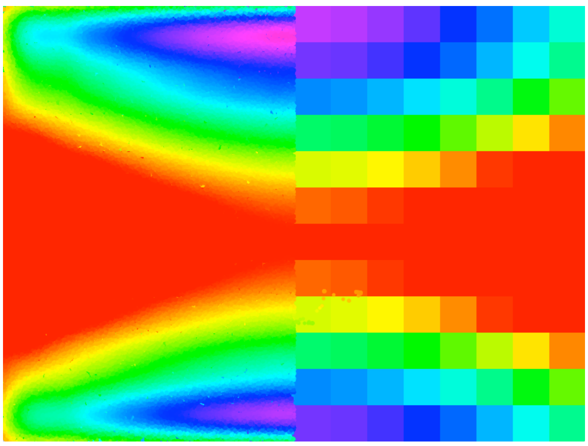

Figure 3: Implementation of the Lanczos resampling algorithm

14 cases were analysed: in a first step, average MD of each single observed activity within a project was compared in an Analysis of Variance (ANOVA) to test whether significant differences in depth could be attributed to activity type. Table 3 illustrates that all results were highly significant, hence it could be confirmed that in each case sitting, standing, walking and interacting showed distinct integration patterns. However, effect sizes as measured by R2 were rather low due to only small variations in depth values between the different activities.

Case	Sitting (not own desk)	Standing	Walking	Interacting	Average	Prob>F	R^2	No of Observations
6	7.84	7.31	7.11	7.55	7.49	<.0001	0.02	2141
13	6.89	6.68	5.81	6.82	6.52	<.0001	0.07	1360
16	6.33	6.01	5.77	6.25	6.16	<.0001	0.03	1558
17	9.28	8.94	8.30	9.04	8.94	<.0001	0.06	2766
4	6.44	6.15	5.79	6.33	6.29	<.0001	0.02	4391
22	3.83	3.64	3.50	3.79	3.75	<.0001	0.03	1676
8	1.88	1.80	1.66	1.83	1.84	<.0001	0.09	858
9	3.42	3.24	2.94	3.28	3.26	<.0001	0.02	3175
25	8.92	7.74	7.44	8.12	8.30	<.0001	0.08	21903
3	4.88	4.38	4.62	4.65	4.75	<.0001	0.02	2176
11	6.01	7.12	5.55	6.85	6.14	<.0001	0.16	868
10	4.13	4.30	3.79	4.22	4.14	<.0001	0.02	5254
15	2.83	3.39	2.92	3.09	2.96	<.0001	0.03	2608
7	6.40	6.04	6.83	6.45	6.28	<.0001	0.01	2415

Table 3: Average Visual Mean Depth for each activity and general average for each project. The table also contains the results of significance from the ANOVA. The coloured cells indicate the ranking of each activity. Darker cells indicate a higher MD. Cases 6 and 15 (in red) are examined in Figure 8.

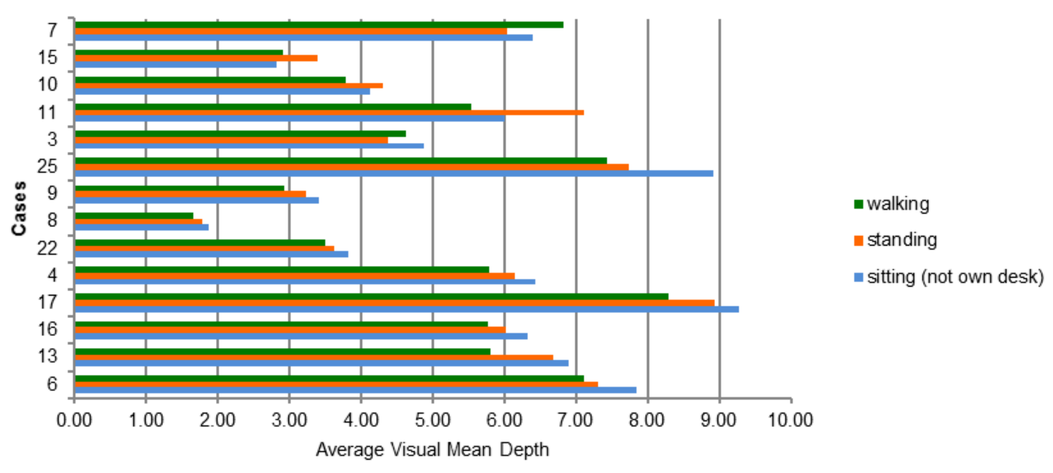

Figure 4: Average Visual Mean Depth for each activity across all cases. The bottom 9 cases are the ones where sitting is more segregated than standing, and standing more than walking.

It can be seen from the colouring of cells in table 3 that a pattern of ranking by depth emerges among the different activities in each of the cases. To see this more clearly, figure 4 plots average MD values for walking, standing and sitting of each case.

It can be seen that sitting is ranked as the activity with the highest mean depth (i.e. most segregated) in 10 cases out of 14, second highest in 3 cases, and lowest in 1 case. Standing is usually second most segregated (in 9 out of 14 cases with deepest activity in 3 cases and shallowest in 2 cases) while walking is usually ranked as the one with the lowest average mean depth in 11 out of 14 cases. In 9 out of the 14 cases there is a specific pattern appearing with sitting having a higher mean depth than standing and standing higher than walking.

When interacting is added to the comparison, it is consistently and across all projects ranked second most segregated. This is most likely due to the fact that it overlaps with all the other activities and is thus closest to the average. The overlap with sitting may be the reason for the consistent second position. As mentioned above, sitting is counted as people not at their own desks, that is, not in their normal 'working mode'. Possible scenarios include: A) if they are sitting in a meeting room, B) if they are sitting in a different facility, such as a canteen, kitchen or tea point, C) if they are using an extra chair in order to cooperate with someone who is sitting at their own desk and D) if they are sitting and working in an alternative space. These scenarios may cover both sitting and interacting at the same time.

While clearly one would expect walking to be most integrated, the fact that interactions are more segregated than the average activity on the whole may come as a surprise. Traditional Space Syntax theory would argue that encounter stems from movement flows (Hillier, 1996), hence it could be expected to see interactions in rather integrated spaces, similarly to walking. Yet this is consistently not the case. It seems that other factors come into play when determining the locations of interactions rather than purely configurational logic. Penn et al. (1999) reach a similar conclusion when confronted with two opposing trends, for two different floors as to whether interaction follows movement.

Two single cases will be explored in more depth to understand the interplay between mean depth and densities of activities. To do so, a visualisation technique was developed plotting both variables in a single floor plan using four colours in two dimensions (see figure 5): increasing activity (in cyan) and decreasing MD (in red).

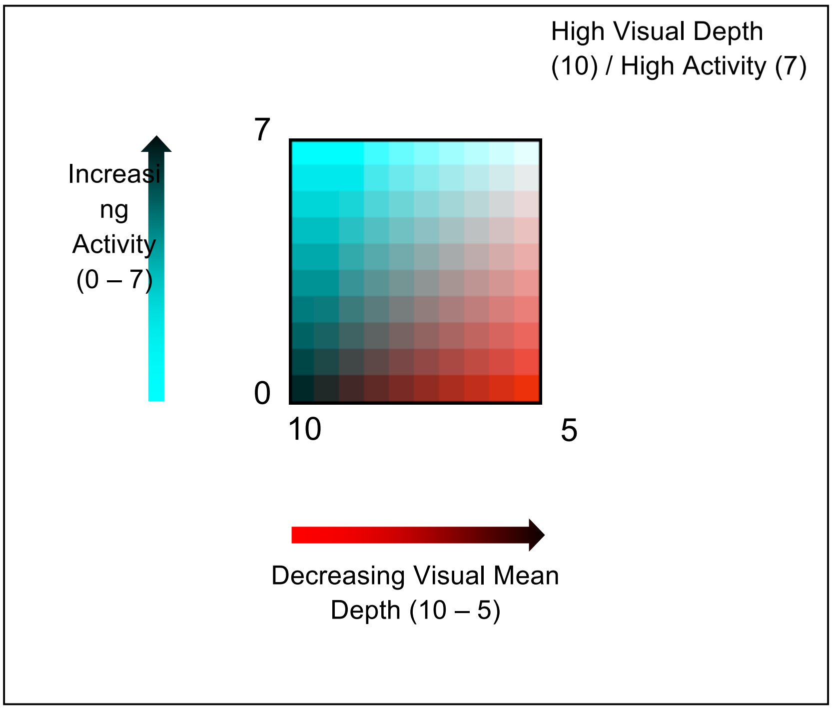

Figure 5: Legend explanation for Figure 6 where black pixels are low on activity and mean depth, red only high mean depth, cyan only high activity and white means both variables are high

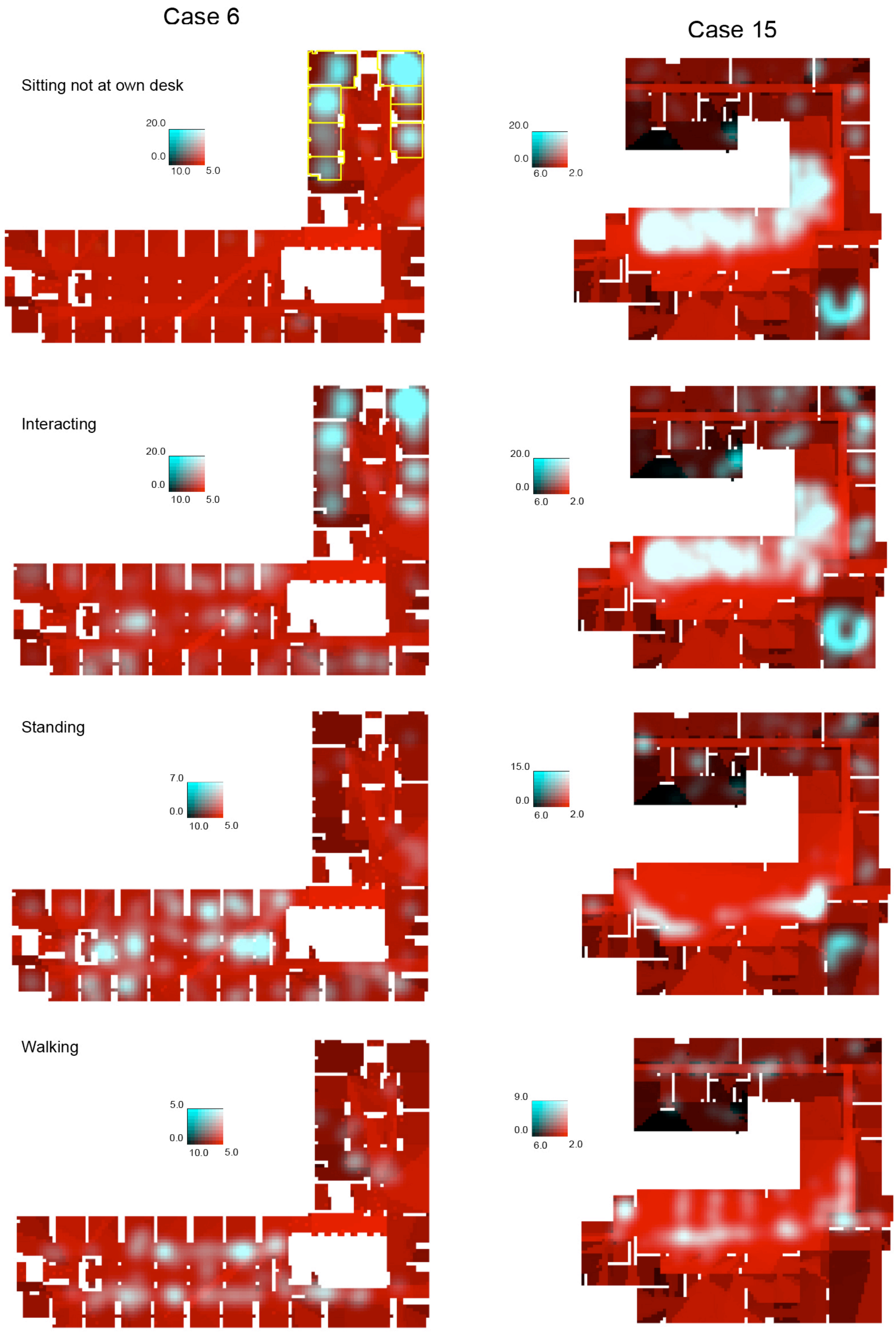

Figure 6: Activities and Visual Mean Depth, cases 6 (left) and 15 (right). Case 15 includes a canteen in a highly integrated location (centre of plan); the meeting rooms in case 6 are marked with a yellow outline.

Figure 6 displays two of the cases, 6 and 15, which represent two different rankings of activities. Case 6, a Legal firm, is the most common scenario of sitting as more segregated than standing, and standing more than walking, while in case 15, a technology company, standing is the most segregated activity with interacting, walking and then sitting ranked second, third and last.

The difference in the rankings of sitting and standing is most likely due to the configuration of space. The highly integrated space (very bright red) in the centre of the plan of case 15 is a canteen with tables, in effect causing an aggregation of people sitting, and people standing when queuing to be served (white areas). The difference in number of people in the canteen in relation to the rest of the plan is likely affecting the measurement. This difference is larger for the sitting activity, forcing the relevant sample to lean towards a lower mean depth. The meeting room at the bottom right of the plan in case 15 may also help reinforce this change in ranking, as the seats are placed on a more integrated part of the room than the presenter position, where standing aggregates (cyan areas due to low integration).

Both examples show how interaction follows sitting closely with some influence from standing. In case 6 people sitting aggregate mostly in meeting rooms (outlined in yellow, top right of the plan), which are also very segregated, while standing happens mostly in open-plan offices and kitchens (bottom centre of the plan). Interaction includes the points from both these activities, leaning mostly towards sitting due to the sheer number of people captured doing so. Walking in both cases follows circulation, thus showing least depth.

In summary, the analysis of the spatial distribution of activities shows that in the majority of cases (9 out of 14) a pattern emerges, in line with Penn et al. (1999), where sitting is more segregated than standing, and standing in turn is more segregated than walking. The fact that interacting is always ranked second when added to the mix is likely an artefact of its overlap with sitting. Thus, given a new case study, one would expect a similar pattern to emerge with relatively high confidence.

Finally, in the last empirical section of this paper, observation data is brought together with the functional allocations of space. It can be asked whether activities follow the programme introduced by functions such as meeting rooms, kitchens, workspaces, etc. For each of the 6 main types of spaces - open plan workspaces, cellular workspaces, alternative spaces, meeting rooms, primary circulation and other facilities - the numbers of interacting people per snapshot per 100 m² of provided area of each type was calculated for each case. The sample size for this analysis is 24 cases. Interactions were chosen as a particularly interesting example, but the analysis could easily be done for all other activities.

Programme may be involved in distributing activities differently by industry. It could be hypothesized that cases from the same industry would follow similar patterns of distributing interactions across functions. For instance organisational cultures relying heavily on meetings might more likely cluster in the same industry, such as Legal Firms.

In order to test whether differences between industries exist, all cases were aggregated by industry as well as for the overall benchmark and an ANOVA was carried out to test whether numbers of people interacting differ by type of space. The results are presented in table 4.

On average across all projects, interactions tend to prefer the following places (in this order): Meeting rooms (5 people per round per 100 m²), Alternative spaces (3), Open plan workspaces (2.43), Other facilities (1.17), Cellular workspaces (1.13) and Primary circulation (0.56). An analysis of variance shows that these differences results are significant at the 0.0001 level with a moderate overall size effect (R^2=0.43).

The same analysis is repeated by industry, allowing for a possibility to more accurately predict the results of a new case, as long as it fits into one of the existing industries and typical organisational cultures. There is only one case in Academia, thus it is omitted from the breakdown.

While table 4 presents the results from the ANOVA as well as mean values, standard deviation (SD) and standard error of the mean (SEM), figure 7 shows the percentage split of all activities by location case by case, by industry and for the overall benchmark.

Industry	Sample size	No. of Observ.		Open Plan	Cellular	Primary Circulation	Meeting	Alternative Spaces	Other Facilities	R^2	P-Value
Overall	24	144	Mean	2.43	1.13	0.56	5.06	3.00	1.17	0.43	<.0001
			SD	0.92	0.93	0.33	2.72	2.84	1.19		
			SEM	0.19	0.19	0.07	0.56	0.58	0.24		
Legal	3	18	Mean	1.70	1.10	0.31	3.71	1.20	0.52	0.89	<.0001
			SD	0.52	0.12	0.12	0.49	0.86	0.27		
			SEM	0.30	0.07	0.07	0.29	0.50	0.15		
Technology	4	24	Mean	3.21	0.87	0.53	3.82	2.70	2.41	0.36	0.1264
			SD	1.50	0.72	0.27	3.00	1.48	2.50		
			SEM	0.75	0.41	0.13	1.50	0.74	1.25		
Media	4	24	Mean	2.59	2.59	0.75	7.84	3.45	1.03	0.67	0.0007
			SD	0.58	0.84	0.50	3.22	3.10	0.57		
			SEM	0.29	0.42	0.25	1.61	1.79	0.28		
Retail	3	18	Mean	2.38	0.60	0.54	5.03	4.05	1.00	0.75	0.0026
			SD	0.51	0.75	0.20	1.16	2.57	0.65		
			SEM	0.29	0.53	0.12	0.67	1.48	0.37		
Financial Services	6	36	Mean	1.99	0.65	0.56	5.56	3.73	1.17	0.48	0.0010
			SD	0.71	0.50	0.30	1.92	4.58	0.67		
			SEM	0.29	0.23	0.12	0.78	1.87	0.27		
Advertising	3	18	Mean	3.20	0.84	0.68	4.43	2.53	0.75	0.43	0.1894
			SD	0.11	1.05	0.54	4.22	2.20	0.74		
			SEM	0.06	0.74	0.31	2.44	1.27	0.43		

Table 4: Numbers of people interacting per type of space overall (top row) and by industry; red cells pinpoint insignificant differences with p values above the acceptable 0.05, yellow highlights the strong meeting room usage in the Legal industry, and green shows cases with high mean (usage) and low error.

In two of the industries studied, Technology and Advertising, the results became insignificant, with a large SEM appearing across all types of spaces. The sample in both cases leads to inconclusive results, failing to capture the industry typical preferred places for interactions, or the cultures of different companies. This is very likely a result of dealing with smaller numbers and fewer cases. In the case of Technology the results may stem from the ambiguous definition of the industry itself, containing rather diverse office interaction cultures. Legal workplaces show a particularly high coefficient of determination (R^2=0.89). SEMs for all but open plan types of spaces are lower than the general average, even with only 18 observations, pointing perhaps to a stricter definition of working processes and cultures in addition to strong programmatic rules what each space is for and whether one may interact there or not. Meeting rooms in this sample seem to be the spaces where most interaction is taking place with a mean of 3.71 people per round, per 100 m². Legal also seems to be the industry with most interactions taking place in meeting rooms (43%), which makes it the single biggest category (see Figure 7).

A similar pattern is found in other industries. In the Media, Retail and Financial Services samples higher than average R^2 values are found with highly significant differences among the locations (0.67, 0.75 and 0.48 respectively). Interaction in open plan workspace is less prevalent than in meeting rooms, the low SEM reveals that it is likely to be close to reality and accurately defined. Cellular and alternative workspaces suffer from high SEM in comparison to their respective means. The definition of alternative spaces might be too broad, too diverse, or it might entail too much variation on usage depending on office culture.

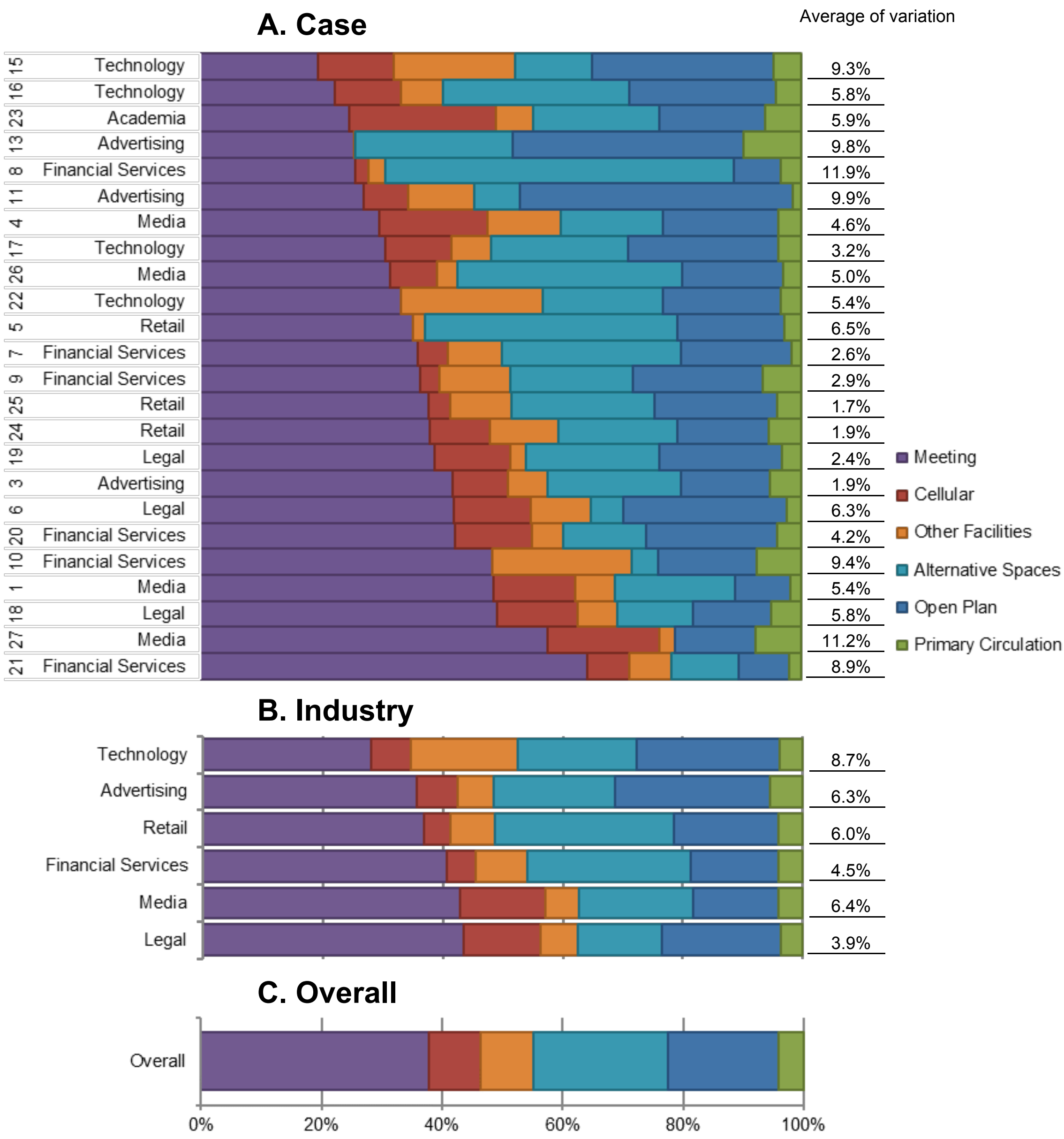

Figure 7: Locational split of the percentage of people interacting per round per m² in each type of space: case by case (top), by industry (middle) and overall benchmark (bottom); average of variation as compared to the overall split is calculated following the method introduced by Capille and Psarra (2013).

Primary circulation seems to be the least preferred place to interact as it retains a generally low average across projects with the SEM following suit. This is an interesting and possibly slightly surprising finding, since often myths are purported that colleagues bump into each other in the corridor. This seems not supported by evidence.

Finally, it can be analysed to which degree the locational preference of interactions in each case as well as each industry differs from the overall benchmark. For this the average of variation of the percentage distribution of locations is calculated following the method developed by Capille and Psarra (2013), which aggregates the average deviation from a baseline, in this case the overall benchmark. It can be seen that individual cases vary from 1.7% (case 25) to 11.9% (case 8). Out of 24 cases, 15 show lower variations than 6%, hence it could be argued that new cases could be predicted rather confidently. Interestingly, aggregating by industry does not necessarily lead to a clearer picture. While legal and financial firms show relatively low variations overall, some of the other industries show higher variations.

7. Conclusion

This paper presented first steps of bringing together different data sets and comparing spatial and social phenomena across larger than usual sample sizes of office buildings.

It was shown that observation data only becomes robust and reliable with longer periods of observations than previously recommended. 25-30 single snapshots (3-4 full days) seemed to produce reasonably stable results for desk occupancy, while 40 snapshots (5 full days) seemed required for percentages of people walking and interacting.

Future research could analyse further metrics, as well as look into the specific reasons that introduce volatility, in order to acquire an overall picture of the accuracy of the analysis.

New insights were also established regarding the distribution of activities in space. Some surprising findings were revealed, for instance dispelling the myth that interactions happen in corridors and highlighting that interactions tend to occur in rather segregated spaces. It was argued that the predictive power of the analysis varies, yet first steps towards establishing generic patterns have clearly been taken.

Next steps could focus on additional combinations of spatial, social and organisational data, for instance looking more deeply into network patterns of interaction and organisational cultures and structures.

References

Abshirini, E. and Koch, D. (2013), *Visibility analysis, similarity, and dissimilarity in general trends of building layouts and their functions.* In: Kim, Y. O., Park, H. T. and Seo, K. W.(eds.), Proceedings of the Ninth International Space Syntax Symposium, Seoul: Sejong University.

Boyd, D. and Crawford, K. (2012), Critical Questions for Big Data. Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Capille, C. and Psarra, S. (2013), Space and planned informality: Strong and weak programme categorisation in public learning environments. In: Kim, Y. O., Park, H. T. and Seo, K. W.(eds.), Proceedings of the Ninth International Space Syntax Symposium, Seoul: Sejong University.

Derix, C. and Jagannath, P. (2014), Digital intuition – Autonomous classifiers for spatial analysis and empirical design. *The Journal of Space Syntax, 5*(2), 190-215.

Grajewski, T. (1992), Space Syntax Observation Manual (2001 unpublished revised edition: L. Vaughan). London: UCL Bartlett and Space Syntax Ltd.

Hillier, B. (1996), Space is the machine. A configurational theory of architecture. Cambridge: Cambridge University Press.

Hillier, B. and Grajewski, T. (1990), The application of space syntax to work environments inside buildings: second phase: towards a predictive model. London: Unit for Architectural Studies, The Bartlett School of Architecture and Planning, University College London.

Hillier, B. and Penn, A. (1991), Visible Colleges: Structure and Randomness in the Place of Discovery. Science in Context, 4(1), 23-49.

Markhede, H. and Koch, D. (2007), *Positioning Analysis: Social Structures in Configurative Modelling.* In: Kubat, A. S., Ertekin, Ö., Güney, Y.I. and Eyübolou, E. (eds.), Proceedings of the Sixth International Space Syntax Symposium, Istanbul: ITU Faculty of Architecture.

Penn, A., Desyllas, J. and Vaughan, L. (1999), The space of innovation: interaction and communication in the work environment. *Environment and Planning B: Planning and Design,* Vol. 26(2), p.193-218.

Peponis, J. (2012), Building layouts as cognitive data: purview and purview interface. *Cognitive Critique, 6*, 11-50.

Peponis, J., Bafna, S., Bajaj, R., Bromberg, J., Congdon, C., Rashid, M., Warmels, S., Yan, Z., Zimring, C. (2007), Designing space to support knowledge work. *Environment and Behavior,* Vol. 39(6), p.815-840. doi: 10.1177/0013916506297216.

Rashid, M., Kampschroer, K., Wineman, J. and Zimring, C. (2006), Spatial layout and face-to-face interaction in offices - a study of the mechanisms of spatial effects on face-to-face interaction. *Environment and Planning B: Planning and Design, 33*, p.825-844.

Sailer, K. (2007), *Movement in workplace environments - configurational or programmed?* In: Kubat, A. S., Ertekin, Ö., Güney, Y.I. and Eyübolou, E. (eds.), Proceedings of the Sixth International Space Syntax Symposium, Istanbul: ITU Faculty of Architecture.

Sailer, K. (2010), *The space-organisation relationship. On the shape of the relationship between spatial configuration and collective organisational behaviours.* Doctoral Dissertation PhD, Technical University of Dresden, Dresden. Retrieved from http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-38427.

Sailer, K., Budgen, A., Lonsdale, N., Turner, A. and Penn, A. (2008), *Evidence-Based Design: Theoretical and Practical Reflections of an Emerging Approach in Office Architecture.* Paper presented at the Undisciplined! Proceedings of the Design Research Society Conference 2008, Sheffield/UK.

Sailer, K., Budgen, A., Lonsdale, N., Turner, A. and Penn, A. (2010), Pre and Post Occupancy Evaluations in Workplace Environments. *The Journal of Space Syntax,* Vol. 1(1), p.199-213.

Sailer, K. and McCulloh, I. A. (2012) Social Networks and Spatial Configuration - How Office Layouts Drive Social Interaction. *Social Networks, 34*(1), 47-58. doi: http://dx.doi.org/10.1016/j.socnet.2011.05.005.

Sailer, K. and Penn, A. (2009), *Spatiality and transpatiality in workplace environments*. In: Koch, D., Marcus, L. and Steen, J. (eds.), Proceedings of the Seventh International Space Syntax Symposium, Stockholm: Royal Institute of Technology.

Sailer, K., Pomeroy, R., Raheem, A., Budgen, A. and Lonsdale, N. (2012), The Generative Office Building. In: Greene, M., Reyes, J. and Castro, A. (eds.), Proceedings of the Eighth International Space Syntax Symposium, Santiago de Chile: PUC.

Shpuza, E. and Peponis, J. (2008), The effect of floorplate shape upon office layout integration. *Environment and Planning B: Planning and Design,* Vol. 35(2), p.318-336.

Steen, J. (2009), *Spatial and Social Configurations in Offices.* In: Koch, D., Marcus, L. and Steen, J. (eds.), Proceedings of the Seventh International Space Syntax Symposium, Stockholm: Royal Institute of Technology.

Steen, J., Blombergsson, M. and Wiklander, J. (2005), Useful buildings for office activities. *Facilities, 23*(3/4), 176-186.

Toker, U. and Gray, D. O. (2008), Innovation spaces: Workspace planning and innovation in U.S. university research centers. *Research Policy, 37*, p.309-329.

Turkowski K. and Gabriel S. (2009), *Filters for Common Resampling Tasks.* In Andrew S. Glassner., Graphics Gems I. Academic Press. P.147–165.

Varoudis, T. (2012), depthmapX: Multi-Platform Spatial Network Analysis Software, OpenSource (Version 0.30). London: University College London. Retrieved from http://varoudis.github.io/depthmapX/

Wineman, J., Hwang, Y., Kabo, F., Owen-Smith, J. and Davis, G. F. (2013), *Spatial Layout, Social Networks and Innovation in Organizations.* In: Kim, Y. O., Park, H. T. and Seo, K. W.(eds.), Proceedings of the Ninth International Space Syntax Symposium, Seoul: Sejong University.