Building and Environment 270 (2025) 112544

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Environmental and spatial dynamics in a flexible workspace for hybrid work: A data-driven design framework

Jiayu Pan [a], Tze Yeung Cho [b], Maoran Sun [a], Koen Steemers [a], Ronita Bardhan [a,*]

[a] Department of Architecture, University of Cambridge, Cambridge, UK
[b] Department of Engineering, University of Cambridge, Cambridge, UK

ARTICLE INFO

Keywords:
Co-working workspace
Flexible working
Indoor environment comfort
Daylight
Seat choice
Spatial configuration

ABSTRACT

The evolving landscape of hybrid working necessitates a corresponding change in workspace design, yet the understanding of the spatial and environmental design of flexible workspaces remains relatively limited. This study investigates the association between workspace design and occupants' behaviour and preferences and builds a data-driven model to predict seat occupancy based on design features. Case study data collection process spans over a year in a hybrid co-working space in London, forming a multi-domain integrated dataset including sensor-detected occupancy, on-site measured and simulated environmental variables, computed spatial metrics and human-voted preferences. The combined effects of thermal comfort (temperature and humidity), visual comfort (artificial light and daylight exposure) and spatial metrics (connectivity and isovist field) are analysed to understand their collective impacts on space occupancy and user preferences. Geo-spatial regression and auto-correlation models are employed to explore associations at a grid level. Through cluster analysis, 61 seats are categorised to four levels based on their occupancy and utilisation rates, and a supervised classification model predicts occupancy levels using design variables. The study aims to present a data-driven analytical framework for evaluating flexible workspace utilisation and demonstrate predictive capabilities by identifying associations between environmental and spatial variables and occupancy patterns, offering a tool to guide evidence-based design and management of flexible workspaces. The analysis underscores the significance of spatial and environmental metrics, revealing that areas or seats with high spatial connectivity and larger visual fields tend to attract more occupants. Additionally, environmental variables, such as temperature and daylight level, exhibit positive effects on increased occupancy.

1. Introduction

The design of workspace has a significant impact on workers' comfort, satisfaction, productivity and well-being [1–3]. In order to improve space use and planning in office environments, it is valuable to understand the relationships between occupants' seat choice and design elements such as view, layout, seat type, environmental conditions and spatial configuration.

Hybrid working patterns are expected to be mainstream for the future of work [4–6], and as a result design adaptations to optimally utilise office space are required. Hybrid work refers to a work model that blends remote and on-site work, granting employees the autonomy to choose their location based on task requirements and personal preferences. Flexible workspaces, characterised by adaptable layouts, shared resources and a departure from traditional fixed-seating arrangements with hot-desking and desk-sharing rules to accommodate dynamic work patterns, have become popular options to accommodate hybrid working [7]. Activity-based work provides a strategic design principle that divides office space to a variety of settings tailored to specific activities. As a result, occupants benefit from shifting open-plan office to activity-based setups with an increased overall satisfaction level [8].

The change in work patterns and preferences has led to increasing challenges in space planning for office spaces, resulting in more uncertainties in design. While users switch to a flexible schedule of working, they are also granted the freedom to choose among different seats. As a result, users will exhibit space preferences in such scenarios. An alternative office model that accommodates hybrid working includes co-working space that has: easy local access; a dynamic and creative environment; an adaptable and sharable layout; with flexible arrangements and reduced costs [9,10]. Such co-working office spaces provide a

---

* Corresponding author.
E-mail address: rb867@cam.ac.uk (R. Bardhan).

https://doi.org/10.1016/j.buildenv.2025.112544
Received 9 September 2024; Received in revised form 13 December 2024; Accepted 6 January 2025
Available online 10 January 2025

Table 2.1
Summary of key literature about the relationship between space choice behaviour and design elements.

Reference	Type of space	Occupants' behaviour-related factors (dependent factor)	Environment/design factors (Independent factor)	Objective	Major finding(s)
[19]	Office	• System-related operation of light and shade • Occupancy level	• Outdoor and indoor environmental conditions (solar radiation, temperature, wind speed, relative humidity and illuminance)	Investigate the relationship between occupancy, environmental control behaviour and environmental conditions	- Large differences in occupancy patterns among different office buildings are found. - Shading and lighting control behaviours are associated to the internal and external environmental conditions, especially with the occupancy level, illuminance level and solar radiation.
[20]	Library	• Spatial choice and space use (collected by survey)	• Environmental: temperature, air movement, air quality, amount of light, visual comfort, noise level, sound privacy • Spatial: accessibility to facilities, amount of space, crowdedness, distance from an entrance, visual privacy • Furniture: adjustability of furniture, comfort of furnishing • Human/social: ease of interaction with friends • Others: aesthetic appearance, cleanliness	Analyse how space attributes influence students' spatial choice	- 'Amount of space', 'Noise level', 'Crowdedness', 'Comfort of furnishing' and 'Cleanliness' are the most important factors that influence spatial choice behaviour
[21]	Office	• Knowledge sharing behaviour (recorded in interaction diaries)	• Spatial: collocation, proximity, inter-visibility, hearing distance and position in the building	Analyse the relationship between layout metrics and meeting behaviours	- Close proximity relates to an awareness of each other and promotes knowledge sharing.
[22]	University campus	• Group work activity (collected by paper-based survey)	• Environmental: noise level, light level, distance from window • Spatial: distance from the nearest neighbour, walking time, degree of enclosure • Furniture: comfort of furnishing, desk size	Develop a space-preference model for group work activities	- Noise level is a significant attribute for both individual and group work, while lighting level, window view and comfort of furnishing are more important for individual work.
[23]	Customer lounge	• On-site observed occupants' behaviour (seat preferences)	• Spatial: integration, isovist area, connectivity • Furniture type	Analyse the relationship between spatial structure and observations of space usage	- Seat preference is a complex phenomenon, depending on the degree of control, furniture type and furniture directness.
[15]	Office	• User behaviour: interaction and eating (collected by observation)	• Spatial: connectivity, distance to circulation	Analyse the relationship between where people choose to eat and interact and the spatial characteristics of the space	- There are no effects of local visibility and distance to circulation on interaction and eating activities.
[12]	Higher education buildings	• Observations of spatial choice	• Environmental: noise level, light level • Spatial: distance from the nearest neighbour, distance to space, degree of enclosure • Furniture: comfort of furnishing, desk size	Predict space-choice behaviour with a focus on individual work-related activities.	- The model provides good explanation of spatial preferences, with the main predictors of 'noise level', 'desk size' and 'distance from nearest neighbour'.
[24]	Library	• Space choice behaviour and user satisfaction (collected by questionnaire)	• Environmental: air movement, air quality, amount of light, noise level, sound privacy, temperature, visual comfort, window view • Spatial: access to facilities, amount of space, crowdedness, distance from entrance, visual privacy • Furniture: comfort of furnishing • Human/social: ease of interaction Others: adjustability to physical conditions, aesthetic appearance, cleanness	Investigate the impacts of space attributes on space-choice behaviour and user satisfaction	- The effects of space attributes are different for space choice and satisfaction. 'Amount of light', 'amount of space' and 'cleanliness' have higher influences on spatial choice.
[25]	University campus	• Spatial choice and activities (collected in time-use diaries)	• Environmental: noise level, lighting level, distance from window • Furniture: desk size, comfort of furnishing • Other: walking time/distance	Validate the spatial-choice model and examine the attributes of the space-options for rejection cases.	- Close distance method could not accurately predict the use of space. - The application of six factors in space-choice models could effectively improve the accuracy of the prediction. - Temperature is an extra factor demonstrated in the rejection case.

(*continued on next page*)

Table 2.1 (*continued*)

Reference	Type of space	Occupants' behaviour-related factors (dependent factor)	Environment/design factors (Independent factor)	Objective	Major finding(s)
[26]	Business centre	• Knowledge sharing behaviour (collected by questionnaire)	• Types of facilities and workspace	Analyse the influence of the physical work environment on social networking and knowledge sharing behaviours	- Lounge, meeting spaces and flexibly used workspaces are important for knowledge sharing.
[27]	Office (hybrid working)	• Preferred seating in hybrid workspace (collected by survey)	• Environmental: closeness to window, quietness, thermal comfort, indoor air quality • Spatial: openness, visual privacy, closeness to facilities, closeness to entry, seating arrangement • Human/social: closeness to manager, team member and friends, interaction with colleagues	Understand the 'preferred seating' and productivity constraints at work	- Interaction with people and closeness to window are key considerations in the preference of seat.
[28]	Office	• User behaviour: interaction and movement (collected by observation)	• Spatial: visual mean depth, travel concentration	Introduce the travel concentration metrics to measure targeted movement.	- The combination of travel concentration and visual mean depth have a significant predictive power on the movement and interaction activities
[13]	Public cultural centre and public library	• User behaviour (collected by sensor)	• Window view • Size of social group	Associate indoor design elements with social behaviour	- The change of window view has led to a change in the seat preference behaviour
[29]	University campus	• Occupants' activity (Wi-Fi data)	• Thermal comfort	Develop an agent-based model of building occupants' activities and thermal comfort	- Some buildings are preferable due to their proximity to other amenities and desirable indoor thermal environment
[30]	Office	• Movement of occupants (simulation)	• Environmental: outdoor and indoor temperature, acoustic performance	Develop an integrated platform to enable the joint simulation of building-occupant interactions	- Occupant-centric data of temperature and noise level recorded in the movement simulation
[31]	Offices and institutional building	• Space use and activity in flexible setup	• Spatial: space type, space number, Environmental: sound insulation, visual privacy, Furniture: furniture arrangement, equipment arrangement, equipment feature	Establish a framework for automating space-use analysis, which predicts and updates the utilisation of flexible spaces by incorporating user activities	- The study presents an automated analysis framework based on the performance of suggested design factors

valuable resource to examine occupants' seat and space preferences in order to reveal insights for the design of future workspaces. Additionally, the data-driven methods hold immense potential to help form evidence-based design solutions in this setup. The exploration of various empirical data can contribute to the systematic evaluation of the office space [11] and provide unique insights to improve the design.

This study assesses the design elements behind seat choice and occupancy level with quantitative evidence, drawing on a case study of a hybrid flexible workspace in London. The investigated design indicators include environmental measurements, spatial attributes and activity preference votes, while sensor-detected occupancy represents the space and seat preferences. The study starts by quantifying the grid-based spatial correlation of different design parameters (including environmental parameters of temperature, humidity and daylight as well as spatial parameters such as isovist area) and occupancy level. The data of seat choices are clustered based on the occupancy indicators, then a supervised classification model is applied to predict the level of occupancy for the seats.

This study intends to develop a spatial analytical pipeline with machine-learning-oriented approaches to utilise quantitative data to predict space preferences and occupancy. The novelty of this study lies in:

- the unique and comprehensive dataset collected in a hybrid flexible workspace for over a year, through occupancy sensors, environmental measurements, simulation and computational spatial metrics;
- integrating the datasets to understand and interpret the patterns of occupancy;
- the application of geo-spatial and machine-learning-based methods in the development of seat choice modelling and prediction.

The model helps predict the seat preferences and occupancy level and understands the role of diverse design features behind it, especially in the open seating area of a flexible and hybrid-working case. It investigates the interaction between workspace design and occupants' behaviour and contributes to the improvement of space efficiencies [12] and subsequently inform the space design and planning in such kind of unique workspace.

2. Background: examining activities and seat choices in office space

There are multiple studies of the association between seat choices, human activities and different design factors in office environment [12–17]. The spatial features of seats in open plan offices, like the number of people in visual field and control of the environment, are found to have an impact on the perceived teamworking and focused working, especially for technology company [18]. The visibility features like isovist area and connectivity are found to have a high predictive effect on the distribution of movement activities [16]. Table 2.1 provides a summary of the relationship between the space choice behaviour and design elements in the previous studies, to provide a thorough review of the background. These studies demonstrate different levels of interactions between environmental and spatial design elements and occupants' behaviour in offices.

Occupancy prediction models are classified to utilisation-based and schedule-based types [12]. The analysis and prediction of schedule-based occupants' presence in office space mainly serve the purpose of energy modelling [12,32,33–35]. Occupants' space choice is a complicated decision-making process. The spatial-choice models are developed with multiple factors, based on users' activities, behaviour, the characteristics of space (e.g. distance, location, size and furniture)

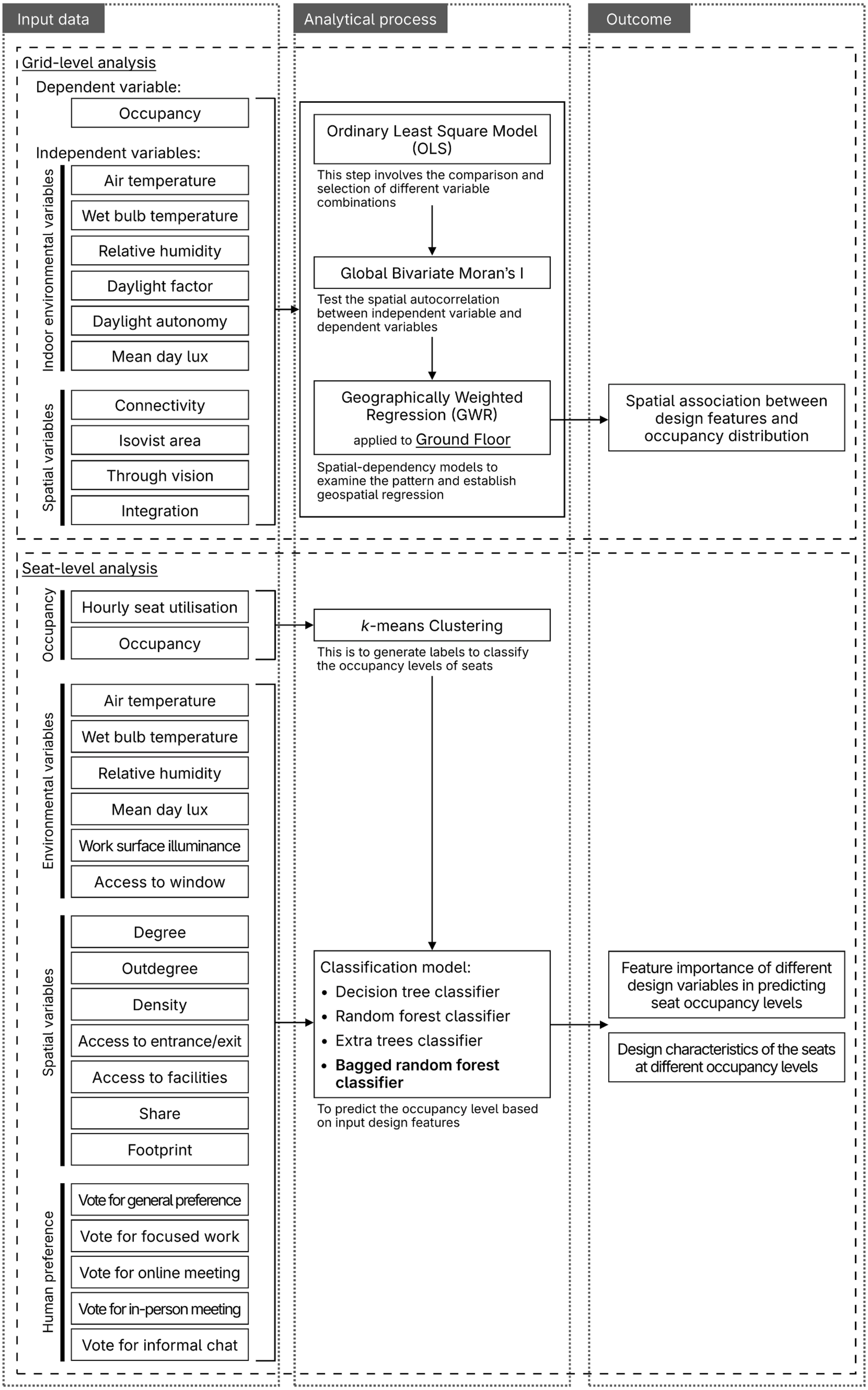
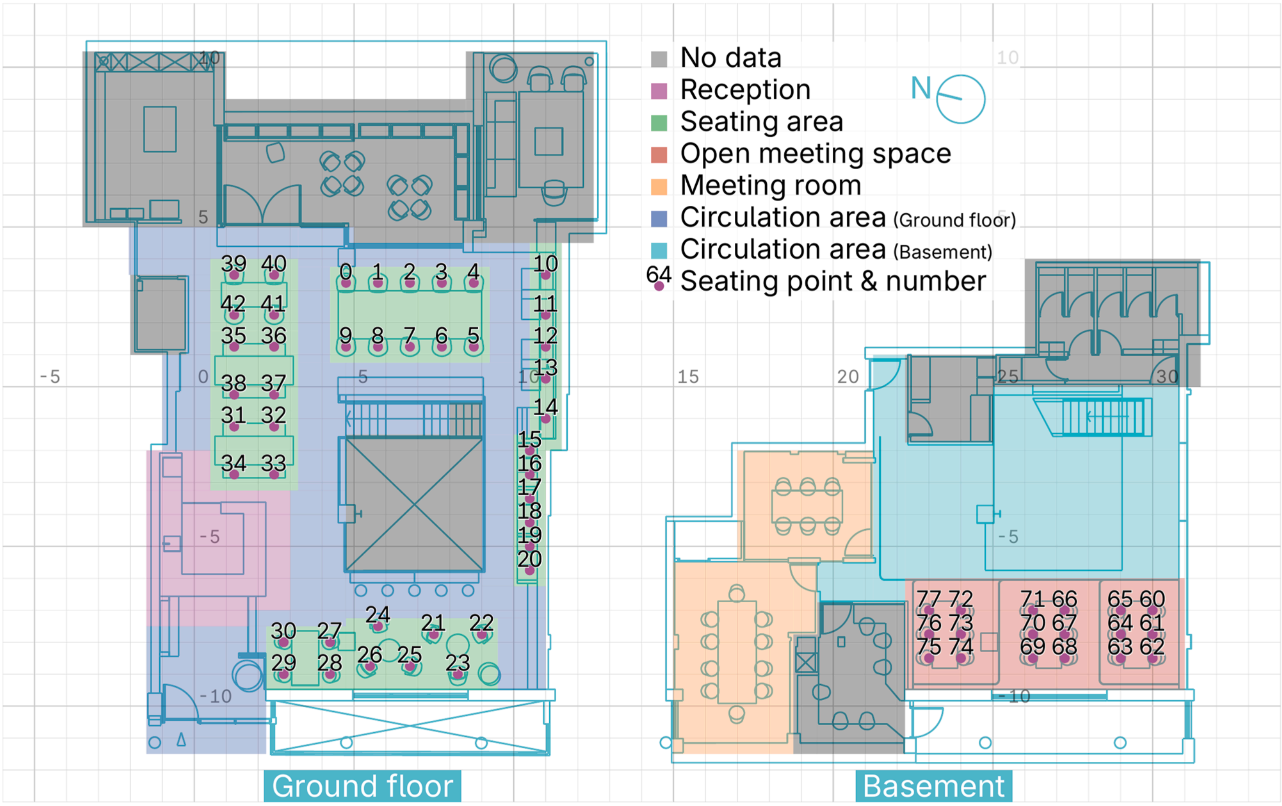


Fig. 1. Framework of data and methods.

a) Site plan



b) Site photos

Fig. 2. Basic site information: (a) site plan and (b) site photos.

[36], spatial visual attributes [16], human cognition and environmental psychology [37].

Previous studies emphasise the importance of developing a systematic way to collect and analyse data and understand occupants' space preference [12,11]. The common data collection methods to investigate indoor occupancy and design mainly include sensor-detection and measurement [12,13] and subjective survey [7,11]. The analytical methods are diverse. The statistical models like multiple ordinal regression [18] and the discrete choice model [12] are utilised to investigate and derive the preference for space and seats. Multi-agent models are also an option to encode the decision-making rules in user behaviour and activities [34,37-40]. The geospatial analysis has been widely performed at the urban scale, while the potential of creating, managing and mapping indoor geospatial data to analyse and utilise the interior space use have not been fully realised [41]. Particularly for the office environment, indoor spatial analysis could be applied to understand the space use and environmental conditions, create data-driven design insights, improve space efficiency, enhance the user experience and optimise space operation and facility maintenance (ArcGIS [42,43]). At the same time, there have been fewer reported applications of machine-learning models in the prediction of space choice.

Additionally, while the previous studies about workspace design are mainly conducted in the work or study space like open plan [18,44] and private offices [13,3] and public library [20,37,13], the hybrid and flexible work setup is a relatively new concept that is less studied. As it is an emerging type of workspace as a response to the advancement in technology and the impacts of the pandemic, the study of occupancy and activity pattern is essential to inform the design of new types of workspace [45] as well as the refurbishment of existing offices. Therefore, a potential gap for the application of novel approaches in a novel

Table 3.1
Definitions of all parameters in the analysis.

Category	Analysis level	Feature	Acronym	Source	Note and Explanation
Occupancy	**Grid/seat level**	**Kernel density estimation of the occupancy level**	**O_KDE**	**Sensor-detection**	The distribution of occupancy level around the space after processing by kernel density estimation [45]
	Seat level	**Hourly seat utilisation rate**	**SU**	**Sensor-detection**	The percentage of occupied time over a total period, at a granularity level of hour [45].
Environment	**Grid/seat level**	**Air temperature**	**AT**	**Measurement**	Unit: degree celsius (°C), measured by Kestrel 5700
	Grid/seat level	**Wet bulb temperature**	**WBT**	**Measurement**	Unit: degree celsius (°C), measured by Kestrel 5700
	Grid/seat level	**Relative humidity**	**RH**	**Measurement**	Unit: %RH, measured by Kestrel 5700
	Grid level	**Daylight factor**	**DF**	**Simulation**	Daylight factor is a measure to quantify the amount of natural daylight received inside a room relative to the available external daylight under overcast sky conditions [50].
	Grid level	**Daylight autonomy**	**DA**	**Simulation**	Daylight autonomy is a metric assessing the proportion of occupied hours during which natural daylight meets the target illuminance at a specific point within a space [51].
	Grid/seat level	**Mean day lux**	**MDL**	**Simulation**	Mean illuminance is the average illuminance across the occupied floor space during all hours [52].
	Seat level	**Work surface Illuminance**	**Illuminance**	**Measurement**	Illuminance indicates the amount of light received on the surface, measured by light meter. Unit: lx.
	Seat level	**Access to window**	**Dist_window**	**Computation**	The access to window is represented by the calculated Euclidean distance from the seat to the nearest window.
Spatial	**Grid level**	**Connectivity**	\	**Computation**	(Visual) Connectivity indicates the count of other cells within a isovist polygon (UCL [53]).
	Grid level	**Isovist area**	**IA**	**Computation**	Isovist area refers to the area enclosed by the isovist polygon, with its origin point at the center of the grid [54].
	Grid level	**Through vision**	**TV**	**Computation**	Through vision is the number of lines of visibility between all other visible cells [55].
	Grid level	**Visual Integration**	**VI**	**Computation**	Visual integration represents the visual distance from all spaces to all others (UCL [53]).
	Seat level	**Degree**	\	**Computation**	Degree is the count of people/seats potentially situating in someone's proximity and directly visible, calculated as the number of desks/seats within a seat's 360° isovist field [18].
	Seat level	**Outdegree**	\	**Computation**	Outdegree is the number of desks/seats within a seat's forward facing (170°) isovist field [18].
	Seat level	**Density**	\	**Computation**	Density represents the relative number of seats in someone's desk neighbourhood, calculated as the ratio of degree and 360° isovist area [18].
	Seat level	**Access to entrance/exit**	**Access_E**	**Computation**	The distance of travel path from the seat to entrance and exit, computed based on network analysis.
	Seat level	**Access to facilities**	**Access_F**	**Computation**	The distance of travel path from the seat to toilets (located in basement), computed based on network analysis.
	Seat level	**Share**	\	**Computation**	Inverse of the number of seats around a table.
	Seat level	**Footprint**	\	**Computation**	The size of a seating area, with a unit of m2
Human preference	**Seat level**	**Vote for general preference**	**V_Pref**	**Survey**	Votes for the seat as the general preference
	Seat level	**Vote for focused work**	**V_Work**	**Survey**	Votes for the seat for focused work
	Seat level	**Vote for online meeting**	**V_OM**	**Survey**	Votes for the seat for online meeting
	Seat level	**Vote for in-person meeting**	**V_IM**	**Survey**	Votes for the seat for in-person meeting
	Seat level	**Vote for informal chat**	**V_Chat**	**Survey**	Votes for the seat for informal chat

workspace setup is identified for the presented investigation in this manuscript.

3. Data and methods

Fig. 1 illustrates the framework of data and methods applied in this study. This section introduces the site information, data collection process and analytical methods and models. In this study, the major hypothesis is that, environmental, spatial and behavioural factors significantly influence workspace occupancy patterns. Specifically, the study proposes: 1) environmental qualities (such as air temperature humidity and daylight) and spatial factors (like isovist and connectivity) can affect overall space occupancy; 2) environmental condition, spatial metrics and human preference for seat type can influence the seat utilisation. These hypotheses guide the analysis framework and inform the interpretation of our findings. Therefore, two sets of analysis were performed at grid level and seat level. The grid-level analysis is performed with 1307 half-meter grids across the whole space to detect the spatial pattern, while seat-level analysis involves the prediction of occupancy level of seats around the site with a supervised classification model.

3.1. Site information

The case study in this manuscript is a co-working space in London, UK, with the site plan and photos shown in Fig. 2. This is a multi-functional space operating from 9am to 5pm on weekdays, which serves as a co-working hub, an exhibition and event venue and a local café. The space is composed of two floors, with a maximum capacity of 70 to 80 people: the ground level is a vibrant café area, featuring an open workspace and a reception, while the basement level is designed to cater to diverse meeting requirements, with two enclosed meeting rooms and flexible open spaces suited for various gatherings. The case study space has a reservation system for meeting rooms but does not have a reservation system for individual workstations.

As the first step, our previous studies [46,45] focused on predicting occupancy schedules and analysing occupancy patterns based on sensor-detected occupancy and on-site activity observation datasets.

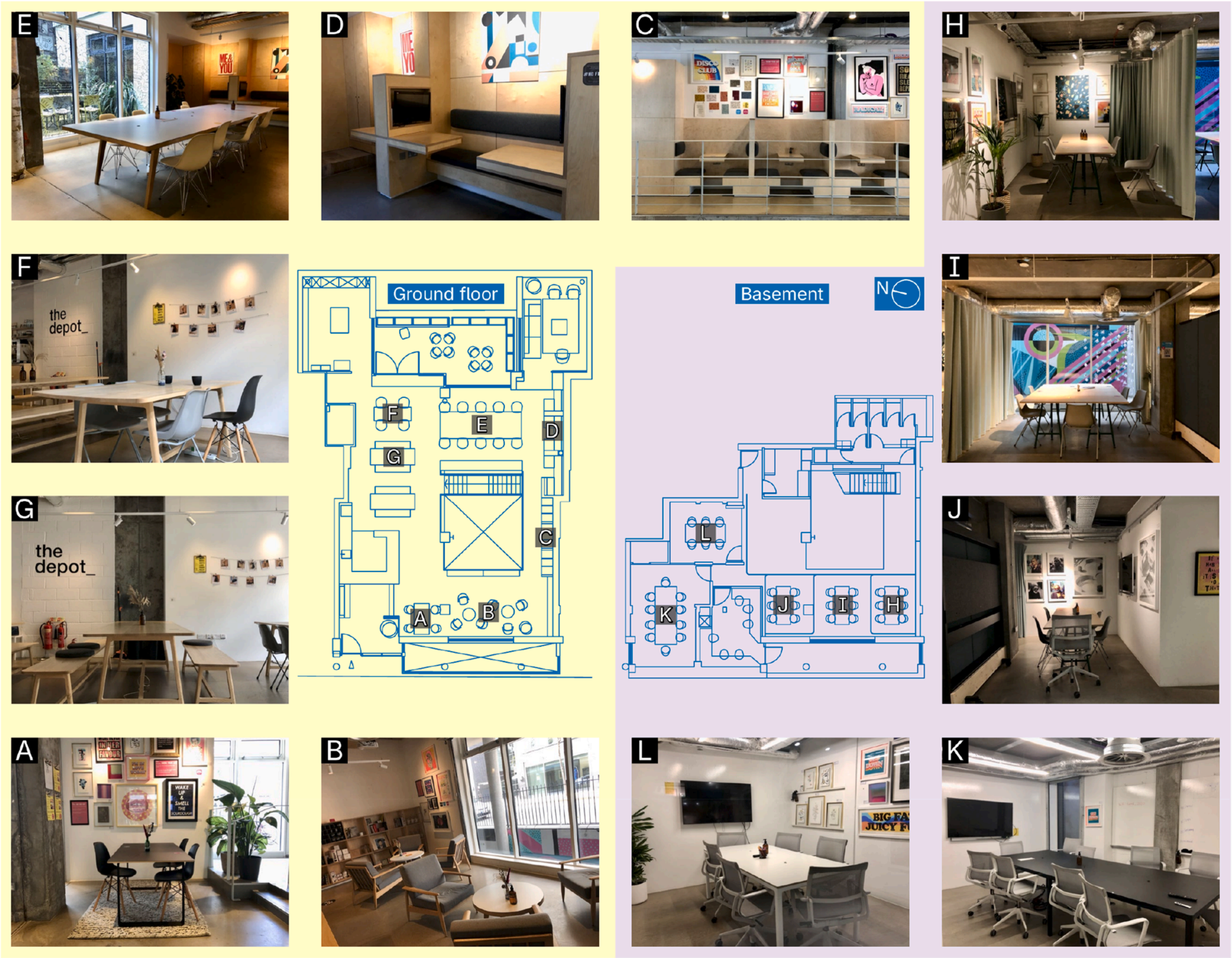


Fig. 3. Seat types listed in the survey.

These studies revealed that peak office use occurs on Tuesdays, Wednesdays and Thursdays and established a basic connection between activities and seat preferences. Additionally, as the second step, a longitudinal survey with on-site measurements was conducted to understand subjective environmental perceptions, highlighting a strong preference for seats with good daylight, ventilation and reduced noise levels reported by survey participants [47,48]. This study builds upon the previous research, which separately examined the occupancy pattern and subjective environmental preferences. The current work acts as a synthesis, integrating the evidence into a comprehensive framework to model their combined effects on workspace utilisation. Unlike prior studies that focused on individual dimensions, this study uniquely employs multi-level analyses (grid-level and seat-level) to predict occupancy patterns, offering novel insights into the interplay between design factors and user behaviour in the case study setting.

3.2. Data collection and data curation

Four types of quantitative parameters in the site are combined and applied in the analysis. These data range from the sensor-detected occupancy, on-site environmental measurements, computed spatial metrics and survey-collected human preferences. Table 3.1 summarises the data involved with explanations and level of analysis. The selection of features is mainly informed by previous studies [12,14,16,18,49], also determined depending on the availability of information from the site.

3.2.1. Sensor-detected occupancy data

The occupancy data is collected through the 17 installed PointGrab sensors around the case study area, which records the timestamped x-y coordinates of each occupant at one-second intervals. The sensors have a high accuracy level of over 95 % [56]. The data was collected from around mid-April 2021 to early February 2023, encompassing a period of approximately one year and nine months. The exploratory analysis of the occupancy data has been presented in previous works, in which the more detailed description about data collection process were described [46,45]. Additionally, the location of occupancy sensors is mapped in the Appendix A.

Kernel density estimation (KDE) is applied to smooth the localised density of occupancy and represent the concentration of occupancy [14, 57], as the original recordings of occupants' locations in the x-y coordination system are continuous and need to be transformed to generate occupancy values for seats and grids. The computed KDE values (O_KDE) are extracted for both each grid and each seat for further analysis. Meanwhile, the utilisation rate of each seat (SU) is calculated as the percentage of occupied time over total time duration at the hour level [45,58]. A seat is classified as 'occupied' when an occupant is detected by the sensor within a 0.5-meter radius of the seat at any time instance [45].

3.2.2. Environmental variables

The on-site measurements of environmental conditions were monitored once a month from March 2022 to February 2023. Five Kestrel 5700 environmental instruments (Kestrel [59]) were placed around the

Table 3.2
OLS model comparison.

	Variables	R-squared	Adjusted R-squared	Akaike information criterion (AIC)	Bayesian information criterion (BIC)	Log-likelihood	Jarque-Bera (JB)	Condition number
OLS Model 1	**10 variables: AT, WBT, RH, DF, DA, MDL, Connectivity, IA, TV, VH**	0.094	0.087	−1497	−1441	759.73	9963.933	1190
OLS Model 2	**6 variables: AT, RH, DA, MDL, IA, TV**	0.070	0.066	−1471	−1435	742.55	8915.293	23.2
OLS Model 3	**5 variables: AT, RH, DA, MDL, IA, TV**	0.063	0.060	−1463	−1432	737.62	9216.588	20.6
OLS Model 4	**4 variables: AT, RH, MDL, IA**	0.068	0.065	−1472	−1446	741.00	8740.164	17.7

space and record readings for a whole day from 9:30 to 17:30, and the example of sensor location is illustrated in Appendix A with the descriptive data of initial recordings and heatmaps. The sampling frequency of the original dataset is 30 s, which means the environmental parameters are recorded once in every half-minute. The measured environmental data, mainly including air temperature (AT), wet-bulb temperature (WBT) and relative humidity (RH), was then interpolated with spatial kriging methods [60] to derive grid-specific and seat-specific values for analysis.

Work surface illuminance (referred to as ‘illuminance’) was measured throughout a single day around the March equinox 2022, for four times for each seat with a URCERI light meter, capturing a combined effect of artificial lighting and natural light conditions. The average value was then calculated to represent the illuminance level for each seat. Additionally, the daylight level around the space was simulated in Climate Studio v1.9 [52], via Grasshopper in Rhino 7. The simulation uses Typical Meteorological Year (TMY) weather files for London City Airport, the nearest weather station, with standard building material properties like concrete exterior walls, beige flooring, white walls and clear glass window. The mean illuminance level (MDL), daylight autonomy (DA) and daylight factor (DF) were extracted from simulation for the grids and seats. Meanwhile, the Euclidean distance to window (Dist_window) for seats is computed to indicate the accessibility to window.

3.2.3. Computed spatial metrics

Several spatial metrics were computed based on the network analysis and visibility field analysis (VGA) [61,55]. Visibility graph analysis explores the characteristics of a visibility graph created from a given spatial environment [55]. VGA is an extension of an earlier concept of isovist, which is defined as the collection of all points visible from a specific vantage point in space [62,63]. The shape and size of isovist normally change with position in the space. The analysis of isovist helps understand the spatial characteristics of environments, especially regarding the aspects of view control and privacy, and how these influence human behaviour and preferences [64]. Isovist polygon is formed by the edges of the visible area, while isovist area represents the total area within an isovist polygon. In this study, the computed metrics from grid-based VGA (0.5 × 0.5 m grid) in DepthMapX [65] include connectivity, isovist area (IA), integration (VI) and through vision (TV), at eye-level (1.5 meter) to examine what individuals can see (UCL [53]). The detailed definitions and explanations given in Table 3.1.

In addition, seats are characterised by the number of people sharing a table (referred to as ‘share’) and the size of the seat (‘footprint’). Share is the inverse of the number of people or seats around one table. Footprint represents the size of seat by dividing the occupied area of a table with the number of seats around the table. The seat-specific spatial metrics are computed based on the metrics applied in a previous study, with a reference to the social network analysis [18,66], including degree, outdegree and density (see Table 3.1 for the specific definitions of each feature). Meanwhile, the access to the entrance/exit (Access_E) and key facilities (mainly toilets) (Access_F) are represented by the computed travel distance to the places, based on the network analysis package [67]. As the reception is very close to the entrance, the access to reception is not calculated separately.

3.2.4. Survey-collected preference of seats and activities

Previous studies discuss the importance of activities in space choice and management [12,68]. A questionnaire survey about the work patterns and workspace preferences was conducted in summer 2022, while seat preferences regarding different office activities were enquired in the questionnaire. The ethics approval for the survey was granted by the Faculty of Architecture and History of Art Research Ethics Subcommittee at the University of Cambridge. A total of 456 people responded to the questions about ‘Which seat would you normally use/want to choose (for different activities)’, with the photos of seats around the case study site (Fig. 3) shown in the questionnaire (Questions attached in Appendix B). The activities suggested in the questions include ‘in general’, ‘do focused work’, ‘have online meetings’, ‘have in-person meetings’ and ‘chat with colleagues’. The votes for different types of seats were collected and assigned to each seat to represent the attractiveness of seats for different activities. Each respondent could choose a maximum of five and a minimum of three seat types in each question.

3.3. Analytical methods

The study utilised both grid-level and seat-level analyses to capture space use patterns at different scales. The grid-level analysis looks at the occupancy pattern and features across the whole space, at a 0.5 × 0.5 m grid structure, investigating collective spatial trends and environmental impacts across the space and attempting to take proximity effects into account with geo-spatial models. The seat level analysis examines the seat preference by synthesising the utilisation rate and occupancy level of each seat, focusing on granular details of individual seating preferences and highlighting localised interactions between users and the immediate environment.

Prior to the analysis, some essential data processing steps like normalisation (max-min normalisation) and Pearson correlation matrix are applied. Most of the analysis and computation process are realised through python 3 [69], with the packages of scikit-learn [70], geopandas [71], statsmodel [72] and pysal [73].

3.3.1. Grid-based analysis: preference for space

The distribution of occupancy with the corresponding design features is investigated across the space with a grid structure, with a total of 1307 half-meter square grids across the whole site. The grid-structure is commonly applied in the spatial analysis of the concentration of occupancy and activities in interior environment, especially for office spaces [14]. The 0.5 × 0.5 meter grid size is selected as it provides a good level of detail without over-generalisation. Two explanatory regression models are applied to construct the relationship between spatial and environmental features and occupancy level of grids: one is the typical multivariate linear regression model, and the other is geographically weighted regression (GWR) model. Prior to the GWR, a set of spatial tests is performed to examine the bivariate spatial autocorrelation with Global Moran’s I [74].

The explanatory ordinary least squares (OLS) regression is first applied to examine the correlation between the independent variables of environmental and spatial features and the dependent variable of KDE occupancy and identify suitable variables for further analysis [75]. OLS model minimises the sum of squared differences between the observed dependent variable and the output of the linear function of the independent variable. There are ten tested independent variables, including AT, WBT, RH, DF, DA, MDL, connectivity, IA, TV and VI. Pearson correlation matrix is used to examine the strength of the linear relationship between the variables and help narrow down the range of variables.

The tested combination for best-fit model and the statistical metrics, such as R-squared, Akaike information criterion (AIC), Bayesian information criterion (BIC) and Jarque-Bera (JB), are reported in Table 3.2. For the first model with all ten variables, the presence of multicollinearity is detected with a high condition number of 1190, indicating a high possibility of model error caused by the change of constant. Therefore, Model 1 is considered not acceptable, and the variables need to be reduced. Model 2, 3 and 4, with six, five and four independent variables respectively, show relatively low R-squared and adjusted R-squared values with significantly decreased condition numbers compared to Model 1. As a result, the selected model (Model 4) has four independent variables of AT, RH, MDl and IA with a constant variable. This model has the advantage of the lowest AIC, BIC and JB values with acceptable R-squared and adjusted R-squared values among Models 2, 3 and 4. However, the relatively low R-squared value of the model indicates further exploration may be required regarding the regression relationships, while spatial correlations are introduced in the analysis as the subsequent step.

The spatial autocorrelation index Moran's I involves the identification of spatial clusters. The univariate Global Moran's I is used to examine the spatial pattern for occupancy level. The spatial weight is constructed by identifying the polygons or grids that share at least one vertex. While the null hypothesis is the spatial randomness, the significance level is determined based on the permutation test, with 999 permutations applied in this analysis. Global Moran's I is expressed by the equation [76]:

$$I = \frac{\sum_i \sum_j \left(w_{ij} z_i \cdot z_j / S_0 \right)}{\sum_i \left(z_i^2 / n \right)}$$

w weight
z variable derivation from the mean
S_0 sum of all weights
n number of observations
i, j locations

Bivariate Global Moran's I is also applied to examine the spatial correlation between occupancy level and other design variables. It explores the correlation between one variable and the spatial lag of the other variable. Bivariate Global Moran's I is given as follows [74]:

$$I_B = \frac{\sum_i \sum_j \left(w_{ij} y_j \cdot x_i \right)}{\sum_i \left(x_i^2 \right)}$$

w weight
x, y variables derivation from the mean
n number of observations
i, j locations

This step also assists the decision of carrying out the analysis with GWR [75]. While OLS regression assumes a universal constant relationship among all variables, the GWR takes the spatial variations into the regression modelling to develop spatially varying relationships [77]. At an analysed location i, the dependent variable y_i can be expressed as:

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \epsilon_i$$

k k-th dependent variable
β_{i0} intercept coefficient
β_{ik} correlation coefficient
ϵ_i random error the inputs of $n \times k$ matrix $\boldsymbol{X}$ of variables and $k \times 1$ matrix $\boldsymbol{y}$ of dependent variables, the above correlation equation can be rewritten in a matrix form to account for the spatial variation as:

$$\widehat{\beta}_i = \left[X^{\mathrm{T}} W_i X \right]^{-1} X^{\mathrm{T}} W_i y$$

W_{i} is weight of the variable at each location correlate to the current location i, and determined by bandwidth.

It recognises and accounts for spatial variations by assigning different weights to observations based on their proximity to the location of interest [77]. This analysis is progressed with the 748 grids on the ground floor, as the ground floor exhibits more spatial variations with diverse seating types and design features. A bandwidth of 51 is determined with 'gaussian', 'AIC' and 'corrected Akaike information criterion (AICc)' as the selection criterion. The use of geo-spatial methods like GWR and Moran's I is particularly effective in this case because they address the spatial heterogeneity of the relationships between variables, which is highly relevant in spatially complex environments.

3.3.2. *Seat-based analysis: preference for seats*

There are a total of 61 seats included in the seat-level analysis. The seats in enclosed meeting rooms are excluded, as they are only available with the reservation in some cases. The use of meeting rooms may not reflect the free selection of seats. The combination of unsupervised clustering model and supervised classification model has proved to be an effective way to develop feature-oriented prediction model [78]. In this study, a clustering model helps to label the seats based on the level of occupancy, and a classification model is applied to formulate the prediction of occupancy.

As the first step, the unsupervised machine learning method of *k*-means clustering model, with features of KDE occupancy and utilisation level, is used to assign the label for occupancy levels. *k*-means clustering is a straightforward method that creates clusters by computing the distance between points to maximise the between-cluster within-group similarity and minimise the between-group differences [79]. The clustering process generates four clusters of seats with different occupancy levels, from high occupancy to low occupancy, while the number of clusters is selected based on the inertia values plotted by the elbow curve.

Four different supervised classification algorithms, including a decision tree classifier, an extra trees classifier, a random forest classifier and a bagging classifier based on the random forest classifier [80,81], are applied to predict the seat preference and choice with the seat-level features. Decision tree classifier constructs a tree-like structure through a process of selecting the best features to split the data at each node in the tree, derive the corresponding outcomes at each branch with a label at the leaf node. Random forest classifier combines multiple decision trees and merges their prediction to improve accuracy and robustness. Each tree in the forest is constructed with a random subset of the training data and a random subset of features for decision-making, while the final prediction is generated by synthesising the results from individual trees. Extra trees classifier further randomises the tree-building process compared to Random Forest. Bagging is a further ensemble process to train the models independently in parallel on different random subsets of the data. It is applied on the random forest model to improve the stability and accuracy of the model while reducing model variances and overfitting. In this case, a train-test-split of 0.2 is applied, with 20 % of test size and 80 % of train size for the models. The performances of the models are compared to select the appropriate prediction model. Additionally, the feature importance is then extracted to analyse the contribution of different design factors in predicting the occupancy labels.

Table 4.1
Descriptive data (Grid-level).

Category	Features	Mean	Standard deviation (std)	Min	Max	25 % (Q1)	50 %(Q2)	75 % (Q3)
Occupancy	**O_KDE**	0.003	0.006	0.000	0.040	0.000	0.001	0.004
Environment	**AT (°C)**	22.24	0.37	21.86	22.88	21.86	22.24	22.63
	WBT (°C)	15.15	0.26	14.86	15.49	14.86	15.27	15.38
	RH (%RH)	46.83	0.53	45.25	47.71	46.45	47.10	47.10
	DF	0.008	0.018	0.000	0.176	0.002	0.003	0.005
	DA (%)	0.14	0.22	0.00	0.91	0.01	0.05	0.18
	MDL (lx)	285.69	753.33	0.00	7545.82	53.41	86.64	179.51
Spatial metrics	**Connectivity**	1872.32	905.01	105.00	3042.00	1321.00	2348.50	2729.00
	IA	117.90	57.10	6.29	191.13	83.13	146.99	172.09
	TV	38,844.17	41,646.63	0.00	148,969.00	7105.75	30,239.50	66,067.50
	VI	32.22	16.03	5.81	84.23	19.62	35.11	45.62

Table 4.2
Descriptive data (Seat-level).

Category	Features	Mean	Std	min	max	25 % (Q1)	50 %(Q2)	75 % (Q3)
Occupancy	**O_KDE**	0.010	0.010	0.000	0.044	0.004	0.007	0.011
	SU	0.47	0.24	0.01	0.87	0.28	0.48	0.67
Environment	**AT (°C)**	22.37	0.39	21.86	22.88	21.86	22.52	22.71
	WBT (°C)	15.20	0.23	14.86	15.45	14.86	15.35	15.38
	RH (%RH)	46.62	0.67	45.25	47.64	46.14	46.67	47.10
	MDL (lx)	281.05	460.64	17.43	2688.87	59.08	121.35	286.21
	Illuminance (lx)	120.25	71.27	27.80	360.70	68.15	96.77	149.48
	Dist_window (m)	2.94	1.87	0.48	7.18	1.66	2.51	3.88
Spatial metrics	**Degree**	27.23	9.54	10.00	40.00	19.00	27.00	37.00
	Outdegree	14.93	8.02	3.00	31.00	9.00	14.00	20.00
	Density	0.19	0.04	0.11	0.30	0.17	0.21	0.22
	Access_E	16.16	6.44	4.04	27.32	11.84	15.04	23.09
	Access_F	12.28	3.18	7.01	18.38	9.62	11.84	15.23
	Share	0.30	0.24	0.10	1.00	0.17	0.25	0.33
	Footprint (m2)	0.96	0.33	0.45	1.60	0.67	1.00	1.23
Human voted preferences	**V_Pref**	133.56	48.41	66.00	206.00	90.00	141.00	153.00
	V_Work	113.80	39.85	45.00	194.00	103.00	121.00	144.00
	V_OM	96.49	25.89	54.00	138.00	79.00	92.00	126.00
	V_IM	113.49	39.47	45.00	169.00	76.0	110.0	149.0
	V_Chat	140.30	82.05	38.00	292.00	60.0	172.0	199.0

4. Results

The result section of the study is composed of three parts, the descriptive data and correlation analysis, the grid-level regression analysis and the seat-level occupancy classification and predictions.

4.1. Descriptive data

4.1.1. Descriptive data by features

The descriptive data of grid-level and seat-level features are presented in Tables 4.1 and 4.2. At the grid level, there are a total of 1307 grids involving in the analysis, with 748 on the ground floor and 559 on the basement. A total of 61 seats are included in the seat-level analysis with the seats in the two enclosed meeting rooms excluded from the analysis.

At the grid level, the mean KDE occupancy is relatively low at 0.00306, with a range from 0 to 0.0399, while O_KDE at the seat level has a higher mean value at 0.0103 as the seats have more concentrated occupancies. The highest seat utilisation rate is at 87 %, while the lowest is less than 1 %.

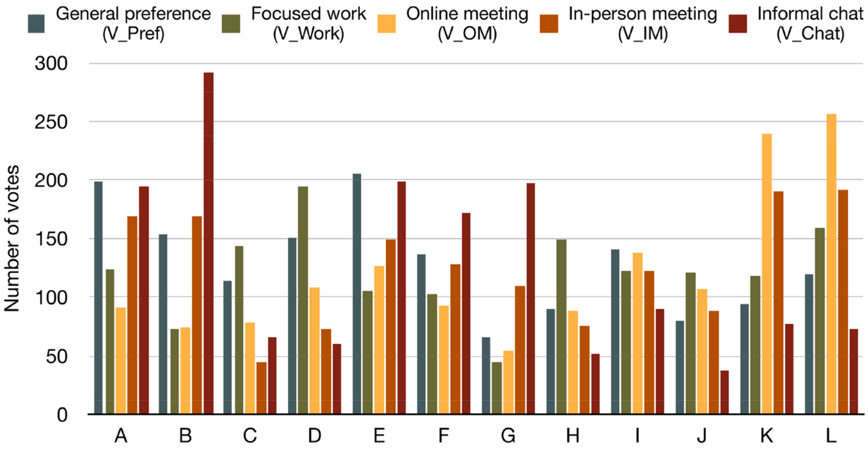
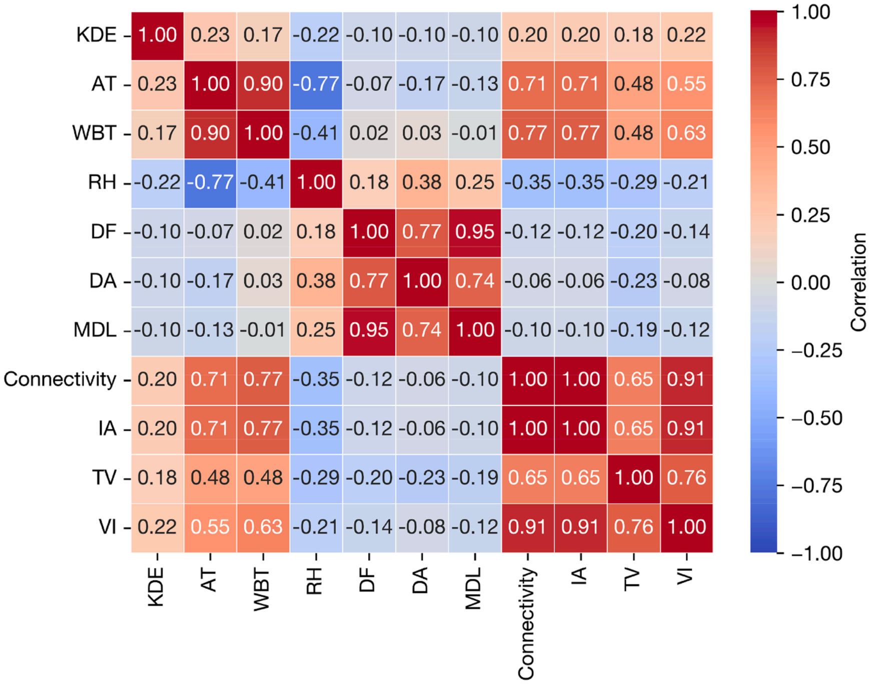


Fig. 4. Distribution of preference votes across different seat types.



	KDE	AT	WBT	RH	DF	DA	MDL	Connectivity	IA	TV	VI
KDE	1.00	0.23	0.17	-0.22	-0.10	-0.10	-0.10	0.20	0.20	0.18	0.22
AT	0.23	1.00	0.90	-0.77	-0.07	-0.17	-0.13	0.71	0.71	0.48	0.55
WBT	0.17	0.90	1.00	-0.41	0.02	0.03	-0.01	0.77	0.77	0.48	0.63
RH	-0.22	-0.77	-0.41	1.00	0.18	0.38	0.25	-0.35	-0.35	-0.29	-0.21
DF	-0.10	-0.07	0.02	0.18	1.00	0.77	0.95	-0.12	-0.12	-0.20	-0.14
DA	-0.10	-0.17	0.03	0.38	0.77	1.00	0.74	-0.06	-0.06	-0.23	-0.08
MDL	-0.10	-0.13	-0.01	0.25	0.95	0.74	1.00	-0.10	-0.10	-0.19	-0.12
Connectivity	0.20	0.71	0.77	-0.35	-0.12	-0.06	-0.10	1.00	1.00	0.65	0.91
IA	0.20	0.71	0.77	-0.35	-0.12	-0.06	-0.10	1.00	1.00	0.65	0.91
TV	0.18	0.48	0.48	-0.29	-0.20	-0.23	-0.19	0.65	0.65	1.00	0.76
VI	0.22	0.55	0.63	-0.21	-0.14	-0.08	-0.12	0.91	0.91	0.76	1.00

Fig. 5. Correlation matrix of occupancy, environmental and spatial parameters (grid-level).

Table 4.3
OLS results (4 variables), for all grids.

	Coefficient	Standard error	t	P-value ($P>|t|$)
Constant	0.107*	0.024	4.506	0.000
AT	−0.0031	0.025	−0.125	0.900
RH	−0.1056***	0.032	−3.304	0.001
MDL	−0.0664*	0.040	−1.667	0.096
IA	0.0681***	0.020	3.422	0.001
R-squared	0.068			
Adjusted R-squared	0.065			

**$p < 0.05$.
* $p < 0.1$.
*** $p < 0.01$.

The air temperatures within the space range from 21.86 °C to 22.88 °C, with an average of 22.24 °C. Average RH values span from 45.25 % to 47.71 %. Daylight factor analysis yields an average of approximately 1 %, suggesting relatively insufficient daylight across the space, except for specific areas near windows that receive much higher daylight exposure. Worksurface illuminance is relatively low, averaging 120.25 lx, falling below the recommended 300–500 lx for a typical office workstation. The average distance of seats to windows is 2.94 m, ranging from 0.48 m for the closest seat to 7.18 m for the furthest one.

The grid-level spatial metrics, including connectivity, IA, TV and VI, exhibit significant variation before normalisation. At the seat level, the mean degree is 27.23, while the average outdegree is around half of the degree value at 14.93, as visual field size of outdegree is approximately halved compared to the degree. Travel-related metrics offer insights into spatial accessibility. The mean travel distance to the entrance, averaging 16.16 m with a range of 7.01 m to 27 .32m. The mean distance to toilets is 12.28 m, with variations from 7.01 m to 18 .38m. The average footprint of seats is $0.96m^2$, with diverse seat sizes ranging from the largest at $1.6m^2$ to the smallest at $0.45m^2$.

Human-voted preferences reflect people's inclinations when selecting seats for various activities. The site features twelve seat types, and Fig. 4 illustrates the voting results. Notably, Seat Types A and E emerge as the most popular choices, garnering the highest overall votes. These two are the tables shared by four people and ten people respectively. For focused work, type D, the semi-enclosed seat for only one person to use, gains the highest vote. Type B, the sofa seats, stands out as the preferred choice for informal chats.

4.1.2. Correlation matrix

Understanding the correlation between features is important to reveal the inherent relationship between features and select features in regression models. The Grid-level correlation matrix is shown in Fig. 5. The correlations between occupancy level to other variables are relatively low with absolute values less than 0.25, indicated that the use of single variable may be not sufficient to predict the occupancy level. At the same time, it could be seen that many of the variables within the same feature categories are closely related, as expected. For example, in the environment group, WBT and AT show a high correlation at 0.9, while WBT is less negatively correlated to RH compared to AT because the humidity level is relatively low around the space. DF, DA and MDL are highly correlated with each other, and the correlation is particularly high between DF and MDL at 0.95. For spatial metrics, there is a nearly perfect correlation relationship between connectivity and IA. The correlations between VI and IA/Connectivity are also high at 0.91, while TV and VI are slightly less correlated with the value of 0.76. These correlation results indicate the potential multi-collinearity existing when performing the regression analysis, and the number of input independent variables could be reduced as some variables could replace each other.

4.2. Grid-level analysis

4.2.1. Ordinary least square regression

The results of OLS model with four features for all grids are demonstrated in Table 4.3. The environmental parameters of AT, RH and MDL have negative coefficients, indicating the spaces with high temperature, high humidity and high daylight level may be less occupied. However, the coefficient of AT is not considered significant with a high p-value. Meanwhile, the spatial feature, represented by IA, has a positive coefficient with a low p-value ($p < 0.01$). It shows that the high

Table 4.4
Results of Global Moran's I and bivariate Moran's I.

	x	y	Moran's I	P-value	Z-score
Univariate	**O_KDE**		0.5078	0.0000	35.9260
Bivariate	**O_KDE**	**AT**	0.2248	0.0010	11.6904
	O_KDE	**WBT**	0.1694	0.0010	9.2737
	O_KDE	**RH**	−0.2132	0.0010	−11.5892
	O_KDE	**DF**	−0.0786	0.0010	−4.2066
	O_KDE	**DA**	−0.0947	0.0010	−5.0765
	O_KDE	**MDL**	−0.0865	0.0010	−4.5300
	O_KDE	**Connectivity**	0.1939	0.0010	10.5280
	O_KDE	**IA**	0.1935	0.0010	10.3247
	O_KDE	**TV**	0.1716	0.0010	8.9281
	O_KDE	**VI**	0.2072	0.0010	11.0045

spatial connectivity is associated with high occupancy level. The R-squared value of the model is 0.068, which means the four independent variables could only explain 6.8 % of variations in KDE occupancy level.

4.2.2. Detection of spatial pattern

Global Moran's I and bivariate Moran's I are applied to examine the spatial autocorrelation patterns for the grid-level dataset. Table 4.4 demonstrates the global spatial autocorrelation results with the Moran's I values, pseudo p-value and z-score with 999 permutations. The single-variate Moran's I is applied to examine the spatial autocorrelation of the occupancy level, with a positive value of 0.508, indicating a relatively strong and significant autocorrelation. The bi-variate Moran's I values between the KDE-occupancy level and design features detect the existence of significant spatial correlations. The environmental variables like AT and WBT and spatial metrics of connectivity, IA, TV and VI have positive spatial correlations with the occupancy level, while the daylighting and humidity features have negative spatial correlations with the occupancy level. The high occupancy grids are situated in the area that is more spatially connected and has a higher temperature.

4.2.3. GWR for ground floor

GWR is applied to capture the spatial variations in the regression, as the explanatory power of OLS regression results are relatively limited. Table 4.5 demonstrates the result of the global model and the GWR model with only the grids on the ground floor. The global model for ground floor shows a slightly different tendencies compared to the previous OLS model for the whole space. The major difference is the positive coefficients of AT and RH, indicating the grids with higher temperature and humidity level tend to attract more occupants on the ground floor, but the coefficient of RH is not statistically significant.

The application of GWR has significantly improved the explanatory power of the model, from a R-squared value of 0.067 in OLS model to 0.582 in GWR model. The performance of the GWR model also improves with the decreased AIC and BIC values. The sum of squares of residuals in the GWR model is reduced to 8.893 from 19.839 in the OLS model. Fig. 6 shows the variations of local R-squared values and the intercept values in the space. Fig. 7 visualises the local spatial variation of coefficients around the space from the GWR model, which enables a closer

Table 4.5
Results of GWR, ground floor.

	Global model (OLS)				Geographically weighted regression (BW=51)				
	Coefficient	Standard error	t	P-value ($P>\|t\|$)	Mean	STD	Min	Median	Max
Constant	−0.161	0.114	−1.416	0.157	4.630	22.281	−99.242	1.442	94.633
AT	0.187*	0.099	1.884	0.060	−5.006	19.438	−88.897	−2.052	95.883
RH	0.031	0.081	0.379	0.704	−4.555	13.412	−52.869	−1.936	54.670
MDL	−0.011	0.089	−0.122	0.903	2.656	30.548	−130.474	−0.111	202.109
IA	0.146***	0.032	4.501	0.000	1.587	9.398	−72.558	0.015	72.200
R-squared	0.067				0.582				
Adjusted R-squared	0.062				0.522				
AIC	−582.300				−1002.035				
BIC	−559.200				−562.212				
Log-Likelihood	296.170				596.271				
Residual sum of squares	19.839				8.893				

** $p < 0.05$.
* $p < 0.1$.
*** $p < 0.01$.

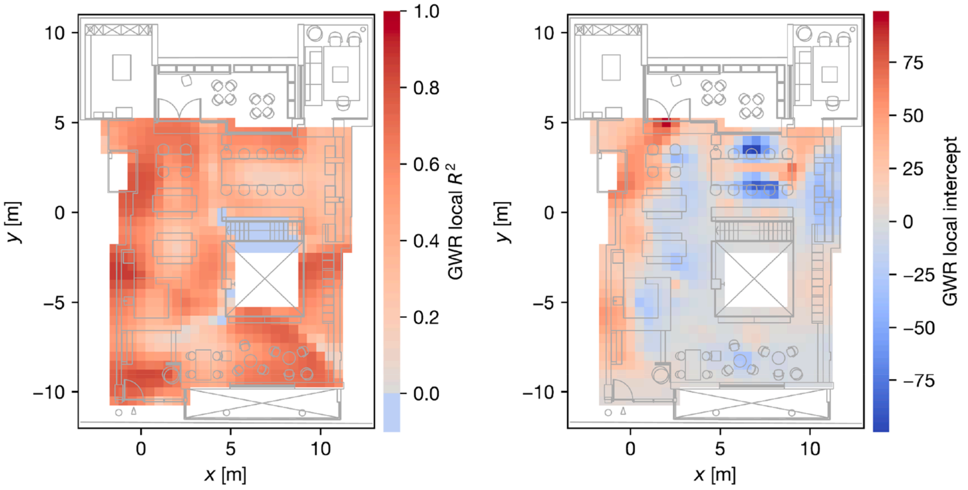


Fig. 6. Local R-squared values and intercept values from GWR model.

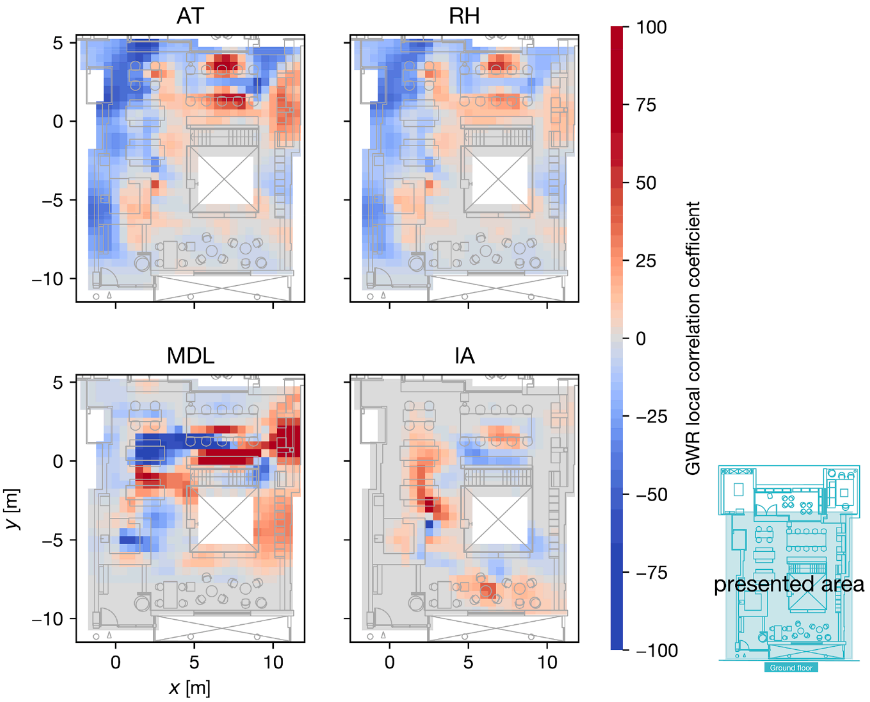


Fig. 7. Local coefficients for each explanatory variable in GWR.

Table 4.6
Occupancy level of k-means clusters.

Cluster	Occupancy level	Count	O_KDE		SU	
			mean	std	mean	std
1	High	16.0	0.016	0.012	0.779	0.050
2	Medium	20.0	0.005	0.002	0.299	0.055
3	Medium high	17.0	0.015	0.011	0.542	0.064
4	Low	8.0	0.001	0.001	0.101	0.081

Table 4.7
Comparison of classification models.

Model	Test-set accuracy	Train-set accuracy
Decision Tree classifier (DT)	0.615	1.000
Random Forest classifier (RF)	0.692	0.854
Extra Trees classifier (ET)	0.769	1.000
Bagging Random Forest classifier (Bagging RF)	0.769	0.875

look to the distribution of positive and negative coefficients for different variables. Although the local coefficients may vary significantly, the mean coefficients of four variables reveal some patterns. Both the mean coefficient values of AT and RH are negative, while the mean values of MDL and IA are positive.

4.3. *Predicting seat occupancy*

4.3.1. *k-means clustering*

k-means clustering is applied to form four groups of seats with different occupancy levels based on their KDE occupancy features and hourly utilisation rate, as shown in Table 4.6. The seat clusters are characterised by their occupancy levels, from ‘High’, ‘Medium high’ to ‘Medium’ and ‘Low’. Cluster 1 is the high occupancy cluster with the highest mean KDE value and the highest average utilisation rate, followed by Cluster 3, which features a high KDE value and a lower utilisation rate. Cluster 2 is the largest cluster with 20 seats and medium occupancy level, while Cluster 4 includes only 8 seats with the lowest occupancy level.

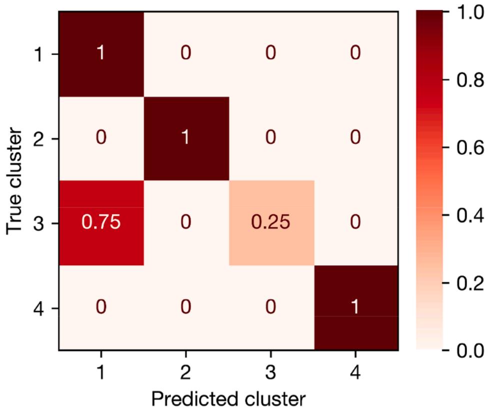


Fig. 8. Confusion matrix of the test set, bagging RF.

4.3.2. *Prediction with the classifiers*

After comparing four models (refer to Table 4.7), the bagging model based on the random forest proves to be the optimal choice, exhibiting high accuracy in both the training and test sets. The training set accuracy for the bagging Random Forest (bagging RF) stands at 76.9 %, while the test set accuracy reaches 87.5 %. These accuracy levels are deemed satisfactory, and there is no significant risk of overfitting. The confusion matrix for the test set results is shown in Fig. 8. Notably, the primary misclassifications occur between Cluster 1 and 3, which is acceptable given the similarity in occupancy levels, categorised as ’high occupancy’ (Cluster 1) and ’medium-high occupancy’ (Cluster 3).

In this study, the feature importance help interpret the contribution of design factors in the prediction of occupancy level, as represented by Fig. 9. In this case, spatial and environmental features take a higher importance compared with the human voted preferences. The feature of ‘Degree’, which represents the number of other seats that is directly visible from the seat, accounts for the highest importance level of 0.117.

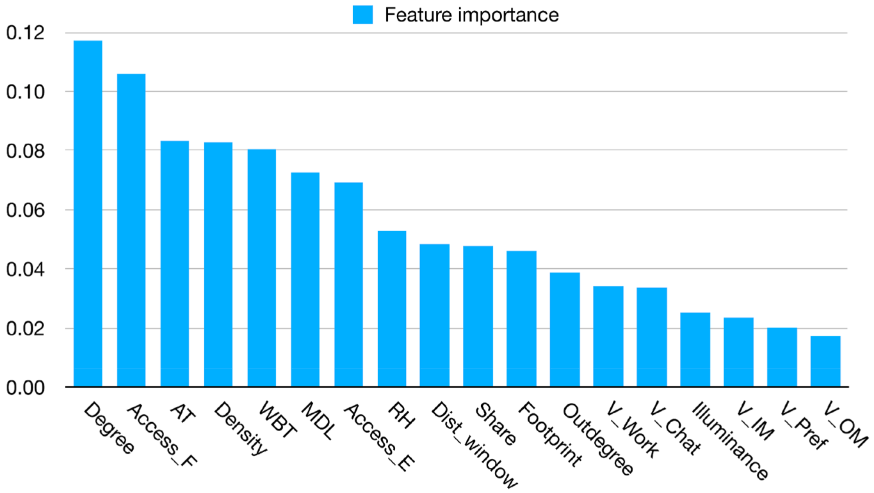


Fig. 9. Feature importance of the bagging RF model.

The distance to facilities (Access_F) is the second most important feature with a value of 0.106, followed by temperature variables (AT and WBT), density and daylight. The predictive effect of human preference votes is relatively limited with the range of importance between 0.017 and 0.034.

5. Discussion

As a result, the exploratory regression models demonstrate the relationship between the design variables and occupancy level. The interior spatial relationship between variables is captured through the spatial autocorrelation indicators and GWR. The spatial regression model is more suitable than the OLS model, with a significant increase in the explanatory power as R-squared values indicated. Although some differences are found when comparing the OLS and GWR models for the whole site and for the ground floor, all models agree that the variable of IA contribute positively to the high occupancy level. This finding corresponds to some conclusions from previous studies, that the more office activities are observed in places with large visual area or high visual integration [16,82].

The bagging random forest model outlines several important factors regarding the prediction of seat occupancy. By looking at the design feature, it is possible to gain an understanding of seat characteristics across different occupancy levels. In Fig. 10, the normalised values of the top significant design features in each cluster (Fig. 10(a)), along with the seat locations (Fig. 10(b)), are depicted, while the values of all features are detailed in Appendix C. In the high-occupancy cluster (Cluster 1), seats exhibit features like high degree, relatively short walking distance to facilities, higher temperature and lower relative humidity with less daylight exposure. The medium-high occupancy cluster (Cluster 3) shares similar features but with seats positioned closer to facilities. The medium- and low-occupancy seats locate in areas characterised by low temperature and high humidity, situated close to windows with increased daylight exposure. Cluster 2, with medium occupancy, displays the lowest average degree among all clusters, with the majority of seats located in the basement, featuring shorter travel distances to facilities but longer distances to entrance and exit. The lowest-occupancy cluster (Cluster 4) showcases some seats close to windows positioned at the edge of each floor.

As a result, several insights for design could be summarised from the analysis:

1. Despite the common preference for more daylight and proximity to windows found in subjective surveys [48,27], it is intriguing that seats with the highest daylight exposure are not necessarily the most occupied ones. This suggests a potential mismatch in the subjective environmental preferences and the real seat selections. The non-incidental association between subjective and objective evaluations of the office space were also identified in the previous study [11]. Nevertheless, clusters exhibiting higher illuminance levels (a combination of both daylight and artificial light) tend to show higher occupancy (as demonstrated in Appendix C), despite the relatively lower predictive importance of the illuminance feature. While daylight enhances perceived workspace quality, glare from direct sunlight can negatively impact visual comfort and task performance. Users engaged in screen-based tasks may find window seats less appealing and opt for seats in more uniformly lit environments [83].
2. The variations of temperature and humidity level around the interior space are subtle, yet having some impacts on the space preferences. Generally, seats or areas with relatively higher temperatures and lower relative humidity tend to be preferable, correlating with higher occupancy rates. As the previous study has reported a neutral temperature of 21.54 °C in a longitudinal thermal comfort survey, indicating the overall environment is slightly cool and survey participants adapt to the environment [48], the finding in this study demonstrates that the occupants tend to look for slightly warmer areas in the case study site.
3. Spatial metrics computed based on visual fields contribute positively to the occupancy, at both the grid level and seat level. Highly spatially integrated spaces tend to attract more occupants, and there’s a clear preference for seats that are visually connected to others in a 360-degree isovist field.
4. As the variation of seat types (type A to J) are primarily represented by the human-voted features and share, the limited importance of these features suggests that seat type arrangements may not be the dominant factor in seat selection decisions. Participants, despite demonstrating tendencies to choose different seats for various activities, seem to prioritise other factors in their decision-making process. Additionally, the size of seat, as represented by the footprint values, is not considered a significant feature too.

6. Conclusion

While the flexible and activity-based workspace setup becomes a popular option to reduce economic cost, improve space efficiency without harming employees’ perceived satisfaction and productivity [7,8], this study further investigates the association between occupancy and design features in a flexible workspace setup, demonstrated by the analysis at grid-level and seat-level. It establishes the spatial correlations between occupancy level and the environmental and spatial design

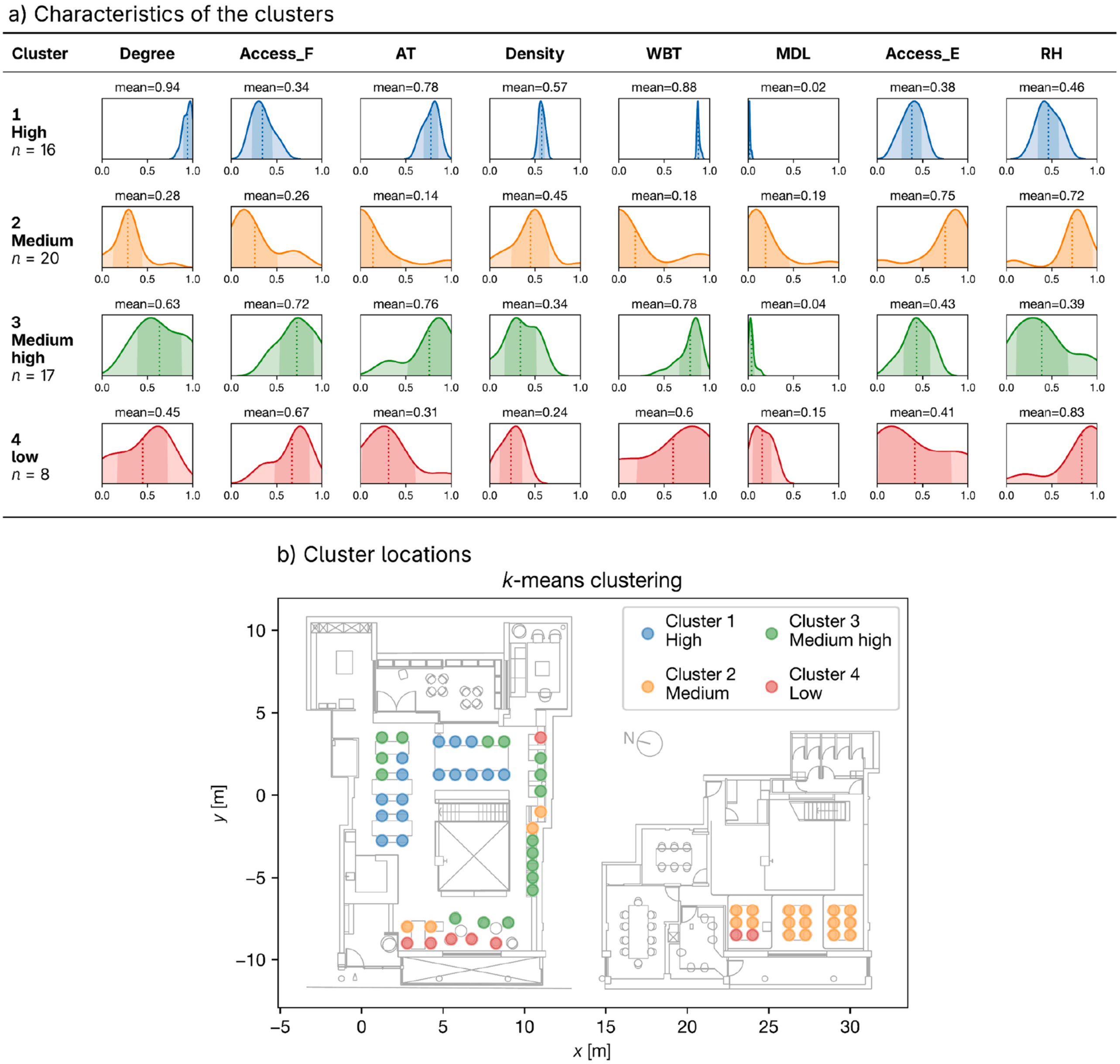


Fig. 10. Design characteristics of the clusters with different occupancy levels.

features at the grid level, while building a prediction model for seat-level occupancy with the spatial, environmental and human preference variables. The analytical models capture the design characteristics of seats and spaces that occupants may prefer more when they are granted with the freedom to choose around a hybrid office setup. The importance of spatial design metrics, such as isovist area and degree, is validated in the analysis. The area or seats with a high spatial connectivity and larger visual field may attract more occupants. At the same time, the environmental variables like temperature, humidity and daylight level are found to have diverse effects on the occupancy. High daylight exposure, which is normally seen as the desirable feature, was not found to attract more occupants, though the seats with higher occupancy generally receive a good amount of lighting. These findings could be a useful guidance for the office design and planning in a new scenario of hybrid working and flexible offices.

The study also demonstrates a systematic approach, through collecting multi-channel post-occupancy empirical data from sensors, simulation and surveys to understand the office design [11]. The analytical framework is set up to demonstrate the application of geospatial models and machine learning toolkits in interior spaces. The development of this analysis and prediction pipeline could significantly benefit the interior space use analysis and establish the data-driven design insights to inform architects, interior designers and facility managers.

This study is limited to a single case study, which inherently restricts the immediate generalisability of the results. While the developed framework effectively explores the relationship between spatial and environmental features and user behaviour in this context, future research could enhance its applicability by incorporating additional contextual variables. These include external climatic conditions, organisational culture and industry-specific work patterns. Expanding the framework to multiple case studies could also improve its reliability and provide a more comprehensive predictive model for diverse office settings.

Although the case study site features a diversity of space and seats, the scale of the space is relatively limited with only around 60 seats involved in the prediction model. Moreover, as the study involves a variety of data collected from different sources, some of the datasets are not collected concurrently. To address this challenge, we included long time spans for environmental and occupancy attributes and averaged the data to represent general patterns. While this approach allowed us to highlight relative differences among spaces and integrate diverse parameters into a consistent spatial and temporal framework, it also introduced limitations. The measured parameters could not fully reflect seasonal variations, and the averaged values may obscure transient fluctuations in workplace configuration and usage patterns. Consequently, the values reflect only relative differences rather than the preferable range of conditions. Future studies could incorporate more granular, temporally synchronised datasets with a dynamic model to capture these fluctuations more effectively. Those environmental variables are interpolated from five measurement points around the space, which may not be sufficient to ensure high-resolution and accurate results. Also, the study could be benefited from incorporating repeated work surface illumination measurements. The collection of human-voted preferences for various activities might carry a degree of bias, as certain participants may choose seats solely based on visual impressions from the photos without physically visiting the site. However, this data was primarily used to complement sensor-detected occupancy patterns, rather than as a standalone determinant of seating behaviour. Furthermore, the spatial regression model excludes the basement due to limitations in the GWR model and variations in the usage patterns and rules of the meeting rooms.

In subsequent investigations, the prediction model could be employed in renovation scenarios to anticipate variations in occupancy levels resulting from alterations in layout, design features and seat arrangements in the case study space. The prediction of the use of meeting space can be embedded by identifying usage patterns and specifying a distinct set of design features that directly impact the meeting room experience. Additionally, the datasets applied in this study are not exhaustive. More factors could be included in such kind of data-driven framework for workspace design, depending on the context and availability - such as mean radiant temperature, noise levels, air quality, carbon dioxide levels, high-resolution air movement measurements, comfort of furnishings, social networks among workers, work activities, task requirements, individual schedules, group size and seat reservation systems. Future research could integrate these variables to expand the framework's scope and applicability. This would significantly enhance the depth and breadth of analysis, offering more comprehensive insights into workspace design and utilisation.

CRediT authorship contribution statement

Jiayu Pan: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – review & editing, Writing – original draft. **Tze Yeung Cho:** Data curation, Software, Visualization, Writing – original draft. **Maoran Sun:** Writing – review & editing, Methodology, Validation. **Koen Steemers:** Writing – review & editing. **Ronita Bardhan:** Writing – review & editing, Visualization, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to thank the lab_ collective, sense_ and the depot_ for enabling the data collection for this study. This work was supported by The Alan Turing Institute's Enrichment Scheme, received by JP (https://www.turing.ac.uk/work-turing/studentships/enrichment). RB would like to thank British Council for funding the MINE – Machine learning Intelligent Network for Epidemics (G111411; 2020-RLWK12–10195). All opinion, findings and/or conclusion are that of the authors and do not necessarily reflect the views of the funding organisation.

Appendix A. Sensor locations and measurements

See Figs. A1, A2 and Table A1

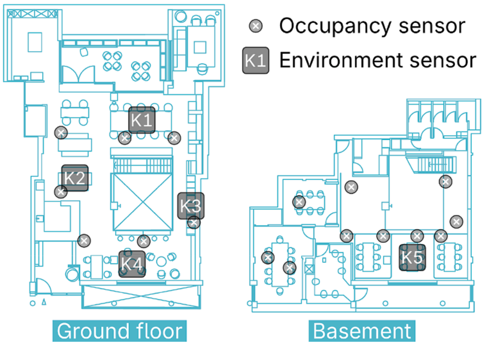


Fig. A1. Location of sensors.

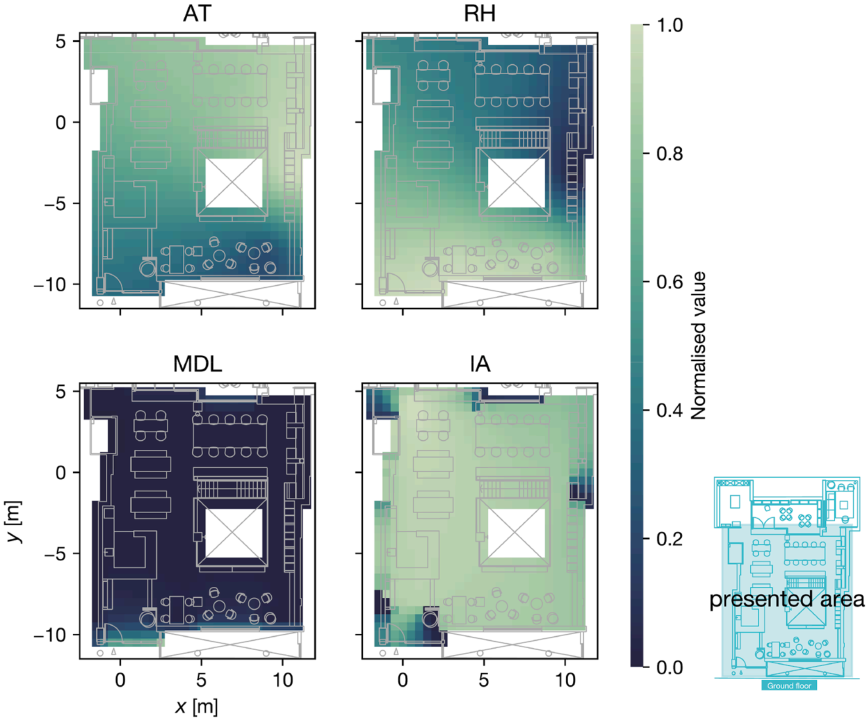


Fig. A2. Heatmaps of environmental and spatial metrics.

Table A1
Descriptive data of environmental recordings.

Date	**Zone**	**Temperature °C**				**Wet Bulb Temp °C**				**Relative Humidity %**			
		Mean	**Max**	**Min**	**Std**	**Mean**	**Max**	**Min**	**Std**	**Mean**	**Max**	**Min**	**Std**
2022–03–10	**1**	21.5	21.9	21.7	0.218	13.7	14.2	12.9	0.245	42.0	43.2	39.8	0.732
2022–03–10	**2**	20.9	21.2	20.2	0.161	13.6	14.0	12.8	0.181	44.0	46.6	41.0	1.209
2022–03–10	**3**	21.0	21.1	20.8	0.100	13.5	13.9	12.9	0.218	42.7	44.5	39.6	1.039
2022–03–10	**4**	20.8	21.2	20.6	0.128	13.6	14.1	13.2	0.166	44.1	46.6	41.2	0.923
2022–03–23	**1**	23.6	24.8	20.9	0.701	15.2	16.3	13.1	0.738	40.9	45.9	34.2	3.829
2022–03–23	**2**	22.1	22.8	20.2	0.664	14.6	16.0	12.7	0.964	44.4	50.5	37.4	3.611
2022–03–23	**3**	22.7	23.7	20.1	1.067	14.6	15.9	12.4	1.049	41.5	45.5	35.2	2.264
2022–03–23	**4**	21.7	22.3	19.4	0.699	14.2	15.4	12.3	0.918	43.5	48.4	35.8	3.287
2022–03–23	**5**	20.4	21.0	19.4	0.433	13.1	14.3	12.3	0.507	42.8	48.7	39.0	1.504
2022–04–25	**1**	22.2	22.9	20.8	0.600	13.6	14.2	12.8	0.433	37.3	39.4	34.9	1.134
2022–04–25	**2**	22.0	22.5	20.8	0.544	13.5	14.2	12.7	0.422	37.8	39.9	35.2	1.055
2022–04–25	**3**	22.2	22.8	21.2	0.460	13.5	14.3	12.7	0.428	37.1	39.1	34.6	1.008
2022–04–25	**4**	21.7	22.6	20.6	0.532	13.4	14.2	12.4	0.507	38.5	40.8	36.1	0.927
2022–04–25	**5**	21.7	22.7	21.1	0.336	13.4	13.9	12.8	0.277	38.3	40.8	36.1	1.205
2022–05–18	**1**	22.2	24.9	20.1	1.017	14.9	16.8	13.2	0.785	45.1	50.1	41.7	1.644
2022–05–18	**2**	22.7	24.4	21.8	0.675	15.1	16.5	14.3	0.600	44.0	46.8	40.5	1.627
2022–05–18	**3**	23.0	24.1	22.3	0.467	15.1	16.2	14.2	0.457	42.7	46.3	39.9	1.373
2022–05–18	**4**	22.0	23.3	20.6	0.829	14.6	15.8	13.4	0.715	44.8	47.2	42.7	0.968
2022–05–18	**5**	21.3	23.6	19.3	1.240	14.5	16.1	12.6	0.768	47.9	54.9	43.3	3.098
2022–06–15	**1**	23.1	24.9	21.8	0.669	14.9	16.4	13.8	0.647	40.7	46.3	36.2	2.296
2022–06–15	**2**	23.5	24.8	22.2	0.552	15.0	16.4	14.0	0.574	39.9	45.5	35.2	2.093
2022–06–15	**3**	23.4	24.6	22.8	0.377	14.9	17.9	13.9	0.534	39.9	53.9	36.0	2.259
2022–06–15	**4**	23.0	25.3	22.1	0.444	14.8	19.8	13.9	0.653	40.8	66.5	36.4	2.915
2022–06–15	**5**	23.6	24.0	23.2	0.226	15.3	16.0	14.7	0.350	40.9	45.2	36.6	2.376
2022–07–13	**1**	24.4	26.0	22.5	0.867	16.7	18.2	14.7	0.822	45.8	51.6	41.7	2.130
2022–07–13	**2**	24.4	25.2	22.6	0.513	16.8	23.6	14.9	0.730	46.3	62.0	41.0	1.839
2022–07–13	**3**	24.6	26.4	23.2	0.673	16.8	19.8	15.1	0.695	45.2	56.6	41.4	1.937
2022–07–13	**4**	23.8	24.8	21.8	0.756	16.5	17.5	14.7	0.639	47.3	53.9	43.3	2.246
2022–07–13	**5**	22.9	24.3	20.8	0.868	15.8	19.7	13.8	0.893	47.7	67.5	41.8	3.382
2022–08–22	**1**	24.2	25.4	21.9	1.050	17.3	19.2	14.9	0.976	50.5	58.8	47.2	1.538
2022–08–22	**2**	24.7	25.6	23.0	0.666	17.4	18.7	15.7	0.803	48.8	52.0	43.3	2.026
2022–08–22	**3**	24.7	26.2	22.9	0.980	17.4	18.9	15.6	0.895	48.6	51.5	43.9	1.544
2022–08–22	**4**	24.0	26.0	21.4	1.062	17.3	18.9	15.0	0.961	51.6	59.0	47.5	1.173
2022–08–22	**5**	24.2	25.2	23.2	0.341	17.2	19.0	15.9	0.637	50.4	55.7	46.4	2.338
2022–09–14	**1**	23.6	24.6	22.4	0.653	17.3	18.1	16.7	0.358	54.4	60.1	48.1	3.066

(continued on next page)

Table A1 (*continued*)

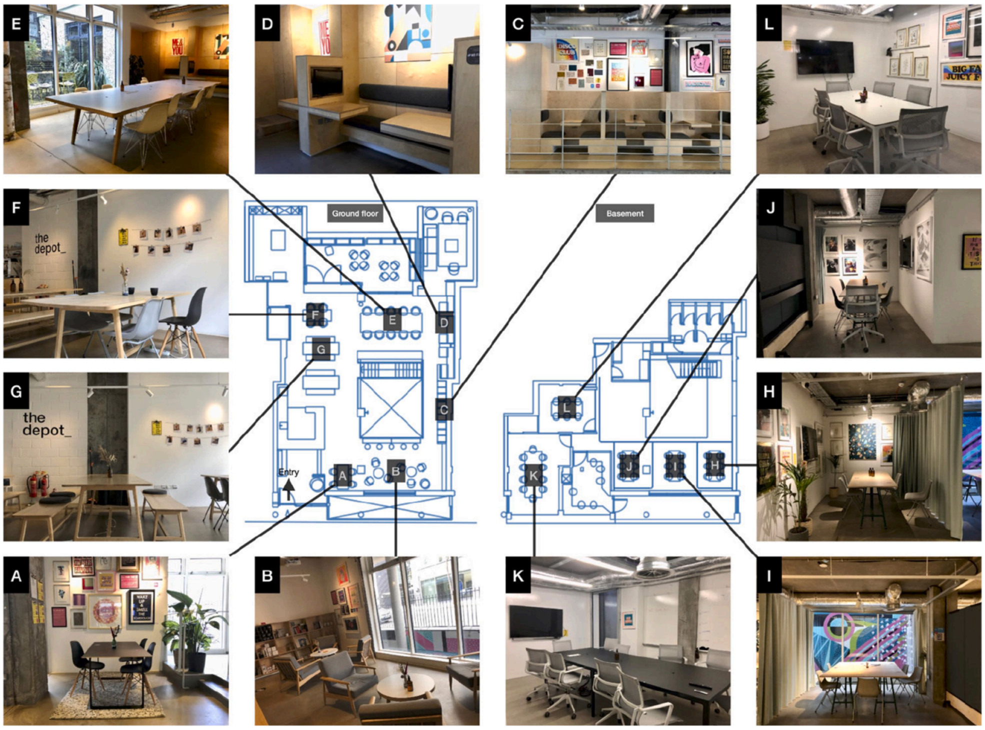
Date	**Zone**	**Temperature °C**				**Wet Bulb Temp °C**				**Relative Humidity %**			
		Mean	**Max**	**Min**	**Std**	**Mean**	**Max**	**Min**	**Std**	**Mean**	**Max**	**Min**	**Std**
2022–09–14	**2**	23.6	24.8	21.7	0.430	17.3	18.2	16.3	0.373	54.1	59.7	48.8	2.494
2022–09–14	**3**	23.4	24.2	22.5	0.547	17.1	17.8	16.4	0.365	53.4	58.6	48.0	2.856
2022–09–14	**4**	22.5	23.6	21.3	0.797	16.7	17.5	16.1	0.404	56.0	61.0	49.0	3.830
2022–09–14	**5**	22.5	23.0	22.1	0.282	16.9	17.2	16.2	0.273	57.0	60.7	51.9	2.348
2022–10–14	**1**	23.3	24.1	22.5	0.376	16.8	17.3	16.2	0.192	52.0	54.0	49.7	1.041
2022–10–14	**2**	23.2	23.7	22.8	0.188	16.8	17.2	16.3	0.164	52.5	54.8	51.5	0.437
2022–10–14	**3**	24.1	24.1	24.0	0.047	16.9	17.0	16.8	0.040	48.9	49.8	48.4	0.464
2022–10–14	**4**	22.3	22.4	22.2	0.051	16.3	16.6	16.2	0.096	54.4	56.0	53.7	0.713
2022–10–14	**5**	22.8	23.2	22.3	0.221	16.4	17.1	16.1	0.135	52.6	56.7	51.4	0.654
2022–11–08	**1**	23.2	24.6	19.8	0.380	17.2	17.9	15.5	0.236	56.1	73.6	51.8	1.444
2022–11–08	**2**	23.2	23.5	22.7	0.176	17.3	17.6	16.8	0.181	56.8	58.3	55.3	0.578
2022–11–08	**3**	24.3	25.2	23.5	0.564	17.8	18.7	17.1	0.420	53.3	58.8	49.4	1.646
2022–11–08	**5**	22.7	23.2	22.2	0.232	16.7	17.1	16.4	0.183	55.0	57.1	53.1	0.897
2022–12–20	**1**	21.2	22.8	18.1	1.089	14.6	15.8	12.6	0.753	49.3	66.7	46.6	1.479
2022–12–20	**2**	21.5	23.4	17.2	1.398	14.6	18.6	12.2	0.870	47.9	66.9	43.9	2.218
2022–12–20	**3**	21.6	23.2	18.0	1.326	14.9	16.1	12.7	0.884	48.8	66.5	46.0	1.729
2022–12–20	**4**	21.1	23.6	17.8	1.072	14.5	18.1	12.4	0.759	49.4	61.2	46.1	1.372
2022–12–20	**5**	19.7	21.7	17.4	0.494	13.5	15.8	12.3	0.343	50.2	55.4	47.9	0.922
2023–01–31	**1**	21.1	22.4	17.5	1.124	14.2	15.2	12.0	0.874	46.9	55.9	45.0	0.987
2023–01–31	**2**	21.0	22.3	18.1	0.956	14.2	15.2	12.1	0.847	47.3	52.4	44.3	1.110
2023–01–31	**4**	20.9	21.9	17.4	0.937	14.2	15.2	12.1	0.776	47.9	54.4	45.4	1.086
2023–01–31	**5**	20.5	21.2	17.7	0.648	13.5	14.1	12.0	0.498	45.6	56.9	43.6	0.915
2023–02–28	**1**	22.2	23.2	18.6	0.817	13.9	14.8	11.4	0.643	39.0	41.5	36.9	0.951
2023–02–28	**2**	21.0	21.8	18.7	0.626	13.4	14.2	11.3	0.634	41.5	43.6	38.9	1.277
2023–02–28	**3**	21.6	22.2	19.1	0.705	13.5	14.2	11.4	0.641	39.4	41.7	37.6	0.860
2023–02–28	**4**	21.0	21.9	18.6	0.815	13.4	14.2	11.3	0.720	42.0	43.8	39.7	0.782
2023–02–28	**5**	20.0	21.2	18.6	0.563	12.4	13.5	11.0	0.575	39.9	42.5	37.6	1.193

Appendix B. Survey

Case study: a hybrid co-working space - which space you would prefer?

25. Have you been to the depot_ (18 Wenlock Road, London)?

○ Yes ○ No ○ Not sure



26. Which seat you would you **normally use/would you choose**?
(minimum selection = 3, maximum selection = 5)

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H ☐ I ☐ J ☐ K ☐ L

27. Which seat would you normally use/want to choose if you want to **do some focused work**?
(minimum selection = 3, maximum selection = 5)

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H ☐ I ☐ J ☐ K ☐ L

28. Which seat would you normally use/want to choose when you **have online meetings**?
(minimum selection = 3, maximum selection = 5)

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H ☐ I ☐ J ☐ K ☐ L

29. Which seat would you normally use/want to choose **when you have in-person meetings**?
(minimum selection = 3, maximum selection = 5)

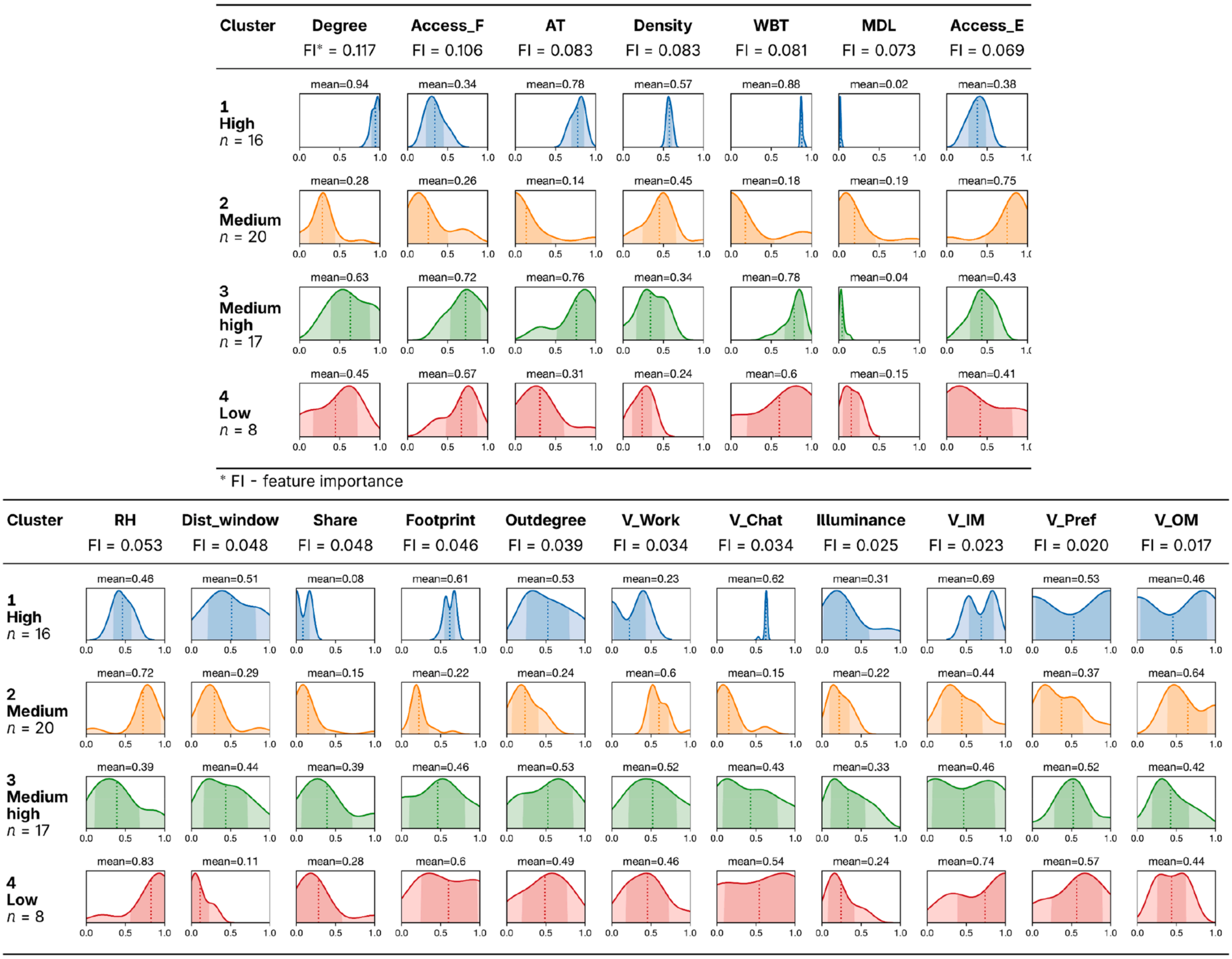
☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H ☐ I ☐ J ☐ K ☐ L

30. Which seat would you normally use/want to choose **when you chat with colleagues**?
(minimum selection = 3, maximum selection = 5)

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H ☐ I ☐ J ☐ K ☐ L

31. If you have any comments or thoughts about this survey, or your workspace and work experience, please share with us:

Appendix C. Characteristics of clusters



* FI - feature importance

Data availability

Data will be made available on request.

References

[1] R. Appel-Meulenbroek, V. Danivska, A Handbook of Theories on Designing Alignment Between People and the Office Environment, 1st ed., Routledge, London, 2021 https://doi.org/10.1201/9781003128830.

[2] D. Clements-Croome, Creating the Productive Workplace: Places to Work Creatively, 3rd edition, Routledge, New York, 2017, https://doi.org/10.4324/9781315658834, 3rd ed.Routledge, 2017. “First edition published by E & FN Spon 2000. Second edition published by Taylor & Francis 2006.”.

[3] S. Leder, G.R. Newsham, J.A. Veitch, S. Mancini, K.E. Charles, Effects of office environment on employee satisfaction: a new analysis, Build. Res. Inf. 44 (2016) 34–50, https://doi.org/10.1080/09613218.2014.1003176.

[4] R. Appel-Meulenbroek, A. Kemperman, A. Van De Water, M. Weijs-Perrée, J. Verhaegh, How to attract employees back to the office? A stated choice study on hybrid working preferences, J. Environ. Psychol. 81 (2022) 101784, https://doi.org/10.1016/j.jenvp.2022.101784.

[5] K. Sailer, M. Thomas, R. Pachilova, The challenges of hybrid work: an architectural sociology perspective, Build. Cities 4 (2023) 650–668, https://doi.org/10.5334/bc.350.

[6] The Economist, The working-from-home illusion fades, Econ. Finance Econ. (2023).

[7] J. Kim, C. Candido, L. Thomas, R. De Dear, Desk ownership in the workplace: the effect of non-territorial working on employee workplace satisfaction, perceived productivity and health, Build. Environ. 103 (2016) 203–214, https://doi.org/10.1016/j.buildenv.2016.04.015.

[8] L. Rolfö, J. Eklund, H. Jahncke, Perceptions of performance and satisfaction after relocation to an activity-based office, Ergonomics 61 (2018) 644–657, https://doi.org/10.1080/00140139.2017.1398844.

[9] J. Berbegal-Mirabent, What do we know about co-working spaces? Trends and Challenges Ahead. Sustainability 13 (2021) 1416, https://doi.org/10.3390/su13031416.

[10] I. Kojo, S. Nenonen, Typologies for co-working spaces in Finland – what and how? Facilities 34 (2016) 302–313, https://doi.org/10.1108/F-08-2014-0066.

[11] A. Mahdavi, U. Unzeitig, Occupancy implications of spatial, indoor-environmental, and organizational features of office spaces, Build. Environ. 40 (2005) 113–123, https://doi.org/10.1016/j.buildenv.2004.04.013.

[12] S.H. Cha, K. Steemers, T.W. Kim, Modeling space preferences for accurate occupancy prediction during the design phase, Autom. Constr. 93 (2018) 135–147, https://doi.org/10.1016/j.autcon.2018.05.001.

[13] A. Khoudi, Using Digital Twin of Indoor Workspaces to Attain a Quantitative Understanding and Improvement of the User Behaviour and Experience, Aarhus School of Architecture, Denmark, 2022. Aarhus School of Architecture, Exners Plads 7, 8000 Aarhus C.

[14] P. Koutsolampros, Human Behaviour in Office environments. Finding patterns of Activity and Spatial Configuration in Large Workplace Datasets, University College London, London, 2021.

[15] Koutsolampros, P., Sailer, K., Haslem, R., Austwick, M.Z., Tasos Varoudis, 2017. Big data and workplace micro-behaviours: a closer inspection of the social behaviour of eating and interacting, in: Proceedings of the 11th Space Syntax Symposium. Presented at the Space Syntax Symposium.

[16] Koutsolampros, P., Sailer, K., Varoudis, T., Haslem, R., 2019. Dissecting visibility graph analysis: the metrics and their role in understanding workplace human behaviour, in: 12th International Space Syntax Symposium.

[17] Tabak, V.(Vincent), 2009. User simulation of space utilisation:system for office building usage simulation. https://doi.org/10.6100/IR640457.

[18] K. Sailer, P. Koutsolampros, R. Pachilova, Differential perceptions of teamwork, focused work and perceived productivity as an effect of desk characteristics within a workplace layout, PLOS ONE 16 (2021) e0250058, https://doi.org/10.1371/journal.pone.0250058.

[19] A. Mahdavi, A. Mohammadi, E. Kabir, L. Lambeva, Occupants' operation of lighting and shading systems in office buildings, J. Build. Perform. Simul. 1 (2008) 57–65, https://doi.org/10.1080/19401490801906502.

[20] S.H. Cha, T.W. Kim, What matters for students' use of physical library space? J. Acad. Librariansh. 41 (2015) 274–279, https://doi.org/10.1016/j.acalib.2015.03.014.

[21] R. Appel-Meulenbroek, B. De Vries, M. Weggeman, Knowledge sharing behavior: the role of spatial design in buildings, Environ. Behav. 49 (2017) 874–903, https://doi.org/10.1177/0013916516673405.

[22] S.H. Cha, K. Steemers, T.W. Kim, Modelling building users' space preferences for group work: a discrete-choice experiment, Archit. Sci. Rev. 60 (2017) 460–471, https://doi.org/10.1080/00038628.2017.1363706.

[23] Sailer, K., Psathiti, C., 2017. A prospect-refuge approach to seat preference: environmental psychology and spatial layout, in: Proceedings of the 11th International Space Syntax Symposium. Instituto Superior Tecnico, Departamentode Engenharia Civil, Arquitetura e Georrecursos. Lisbon, Portugal, p. 137.1-137.16.

[24] S.H. Cha, T.W. Kim, The role of space attributes in space-choice behaviour and satisfaction in an academic library, J. Librariansh. Inf. Sci. 52 (2020) 399–409, https://doi.org/10.1177/0961000618794257.

[25] T.W. Kim, S.H. Cha, Empirical validation of the spatial-choice modelling approach to user simulation, Archit. Sci. Rev. 62 (2019) 313–322, https://doi.org/10.1080/00038628.2019.1625299.

[26] M. Weijs-Perrée, R. Appel-Meulenbroek, T. Arentze, G. Romme, The influence of the physical work environment of business centres on social networking and knowledge sharing in the Netherlands, Intell. Build. Int. 11 (2019) 105–125, https://doi.org/10.1080/17508975.2019.1574705.

[27] Rahaman, M.S., Kudo, S., Rawling, T., Ren, Y., Salim, F.D., 2020. Seating preference analysis for hybrid workplaces. Presented at the NFW '20: Symposium on New Future of Work, New York, NY, USA.

[28] Koutsolampros, P., Sailer, K., Haslem, R., 2020. Travel Concentration: the effects of attractor-bound movement on workplace activity, in: Future Workspaces. Proceedings of the Transdisciplinary Workplace Research (TWR) Conference 2020.

[29] M. Mosteiro-Romero, M. Quintana, R. Stouffs, C. Miller, A data-driven agent-based model of occupants' thermal comfort behaviors for the planning of district-scale flexible work arrangements, Build. Environ. 257 (2024) 111479, https://doi.org/10.1016/j.buildenv.2024.111479.

[30] D. Schaumann, S. Moon, M. Usman, R. Goldstein, S. Breslav, A. Khan, P. Faloutsos, M. Kapadia, JOIN: an integrated platform for joint simulation of occupant-building interactions, Archit. Sci. Rev. 63 (2020) 339–350, https://doi.org/10.1080/00038628.2019.1662767.

[31] X. Chen, T.W. Kim, J. Chen, B. Xue, W. Jeong, Ontology-based representations of user activity and flexible space information: towards an automated space-use analysis in buildings, Adv. Civ. Eng. 2019 (2019) 3690419, https://doi.org/10.1155/2019/3690419.

[32] S. Arvidsson, M. Gullstrand, B. Sirmacek, M. Riveiro, Sensor fusion and convolutional neural networks for indoor occupancy prediction using multiple low-cost low-resolution heat sensor data, Sensors 21 (2021) 1036, https://doi.org/10.3390/s21041036.

[33] Goldstein, R., Tessier, A., Khan, A., 2010. Schedule-calibrated occupant behavior simulation, in: Proceedings of the 2010 Spring Simulation Multiconference. Presented at the SpringSim '10: 2010 Spring Simulation Conference, Society for Computer Simulation International, Orlando Florida, pp. 1–8. https://doi.org/10.1145/1878537.1878725.

[34] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simul. 4 (2011) 149–167, https://doi.org/10.1007/s12273-011-0044-5.

[35] D. Wang, C.C. Federspiel, F. Rubinstein, Modeling occupancy in single person offices, Energy Build 37 (2005) 121–126, https://doi.org/10.1016/j.enbuild.2004.06.015.

[36] Goldstein, R., Tessier, A., Khan, A., 2011. Space layout in occupant behavior simulation, in: Proceedings of Building Simulation 2011. Presented at the 12th Conference of International Building Performance Simulation Association, Sydney Australia, pp. 1073–1080.

[37] Chun, J., Gen.Schieck, A.F., Psarras, S., Koutsolampros, P., 2019. Agent based simulation for "Choice of Seats": a study on the human space usage pattern, in: Proceedings of the 12th Space Syntax Symposium. Presented at the Space Syntax Symposium.

[38] Y. Chen, T. Hong, X. Luo, An agent-based stochastic occupancy simulator, Build. Simul. 11 (2018) 37–49, https://doi.org/10.1007/s12273-017-0379-7.

[39] Zimmermann, G., 2010. Agent-based modeling and simulation of individual building occupants, in: Proceedings of SimBuild 2010 Conference. New York, pp. 269–276.

[40] Zimmermann, G., 2007. Modeling and simulation of individual user behavior for building performance predictions, in: Proceedings of the 2007 Summer Computer Simulation Conference. Society For Computer Simulation, San Diego, CA.

[41] H.-K. Kang, K.-J. Li, A standard indoor spatial data model—OGC IndoorGML and implementation approaches, ISPRS Int. J. Geo-Inf. 6 (2017) 116, https://doi.org/10.3390/ijgi6040116.

[42] ArcGIS Indoors, Create a Hybrid Workplace with Indoor Mapping, ArcNews. (2021). URL, https://www.esri.com/about/newsroom/arcnews/create-a-hybrid-workplace-with-indoor-mapping/ (accessed 11.20.23).

[43] ESRI, The Smart Workplace: Indoor GIS For Smart Buildings, Environmental Systems Research Institute, 2023.

[44] A. Shafaghat, A. Keyvanfar, M.S. Ferwati, T. Alizadeh, Enhancing staff's satisfaction with comfort toward productivity by sustainable Open Plan Office Design, Sustain. Cities Soc. 19 (2015) 151–164, https://doi.org/10.1016/j.scs.2015.08.001.

[45] J. Pan, T.Y. Cho, M. Sun, R. Debnath, N. Lonsdale, C. Wilcox, R. Bardhan, Future workspace needs flexibility and diversity: a machine learning-driven behavioural analysis of co-working space, PLOS ONE 18 (2023) e0292370, https://doi.org/10.1371/journal.pone.0292370.

[46] Pan, J., Cho, T.Y., Bardhan, R., 2022b. Occupancy level prediction based on a sensor-detected dataset in a co-working space, in: Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. Presented at the BuildSys '22: The 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, ACM, Boston Massachusetts, pp. 340–347. https://doi.org/10.1145/3563357.3566133.

[47] Pan, J., Bardhan, R., Chen, S., 2022a. How do occupants perceive thermal comfort in a hybrid office space? A case study of a co-working space in London. Presented at the 3rd Conference on Comfort at the Extremes Resilient Comfort: Designing to Survive (CATE2022).

[48] J. Pan, S. Chen, R. Bardhan, Reinventing hybrid office design through a people-centric adaptive approach, Build. Environ. (2024), https://doi.org/10.1016/j.buildenv.2024.111219.

[49] K. Jens, A. Khoudi, Using computer-vision sensors to study the impact of window views on occupancy and self-assessed productivity in flexible working environments: an intervention study, Intell. Build. Int. 15 (2023) 78–90, https://doi.org/10.1080/17508975.2022.2084012.

[50] R.G. Hopkinson, Architectural Physics: Lighting, Her Majesty's Stationery Office, London, 1963.

[51] C.F. Reinhart, Daylight Availability and Manual Lighting Control in Office buildings: Simulation studies and Analysis of Measurement, Fraunhofer IRB Verl, Stuttgart, 2001.

[52] Solemma L.L.C., 2023. Climate studio.

[53] U.C.L. Space Syntax, Space syntax glossary [WWW Document], Space Syntax Online Train. Platf. (2024). URL, https://www.spacesyntax.online/glossary/.

[54] Turner, A., 2004. Depthmap 4 — a researcher's handbook.

[55] A. Turner, M. Doxa, D. O'Sullivan, A. Penn, From isovists to visibility graphs: a methodology for the analysis of architectural space, Environ. Plan. B Plan. Des. 28 (2001) 103–121, https://doi.org/10.1068/b2684.

[56] PointGrab, 2024. PointGrab technology [WWW document]. URL https://www.pointgrab.com/technology/.

[57] B.W. Silverman, Density Estimation For Statistics and Data analysis, Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, 1998.

[58] C. Tagliaro, Y. Zhou, Y. Hua, A change in granularity: measure space utilization through smart technologies, Facilities 39 (2021) 64–79, https://doi.org/10.1108/F-08-2019-0093.

[59] Kestrel Instruments, 2023. Kestrel 5700 elite weather meter with applied ballistics and LiNK [WWW Document]. Kestrel Instrum. URL https://kestrelmeters.com/products/kestrel-elite-weather-meter-with-applied-ballistics (accessed 8.4.23).

[60] P.A. Burrough, Principles of Geographical Information Systems For Land Resources assessment, Monographs on Soil and Resources Survey, Clarendon Press ; Oxford University Press, Oxford [Oxfordshire] : New York, 1986.

[61] A.-L. Barabási, Network Science, Cambridge University Press, 2016.

[62] M. Batty, Exploring Isovist fields: space and shape in architectural and urban morphology, Environ. Plan. B Plan. Des. 28 (2001) 123–150, https://doi.org/10.1068/b2725.

[63] Davis, L.S., Benedikt, M.L., 1979. Computational models of space: isovists and isovist fields. Comput. Graph. Image Process. 11, 49–72. https://doi.org/10.1016/0146-664X(79)90076-5.

[64] M.L. Benedikt, To take hold of space: isovists and isovist fields, Environ. Plan. B Plan. Des. 6 (1979) 47–65, https://doi.org/10.1068/b060047.

[65] depthmapX development team, 2017. depthmapX.

[66] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, 1st ed., Cambridge University Press, 1994 https://doi.org/10.1017/CBO9780511815478.

[67] Hagberg, A.A., Schult, D.A., Swart, P.J., 2008. Exploring network structure, dynamics, and function using NetworkX, in: Proceedings of the 7th Python in Science Conference (SciPy2008). Pasadena, CA USA, pp. 11–15.

[68] A. Pennanen, User Activity Based Workspace Definition As an Instrument For Workplace Management in Multi-User Organizations, Haahtela-kehitys, Helsinki, 2004.

[69] G. Van Rossum, F.L. Drake, Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009.

[70] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: machine learning in python. https://doi.org/10.48550/ARXIV.1201.0490.

[71] Jordahl, K., Bossche, J.V.D., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A.G., Farmer, C., Hjelle, G.A., Snow, A.D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., Maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L.J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, Leblanc, F., 2020. Geopandas/geopandas: v0.8.1. https://doi.org/10.5281/ZENODO.3946761.

[72] Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python, in: 9th Python in Science Conference.

[73] S.J. Rey, L. Anselin, PySAL: a python library of spatial analytical methods, Rev. Reg. Stud. 37 (2007) 5–27.

[74] L. Anselin, I. Syabri, Y. Kho, GeoDa: an introduction to spatial data analysis, Geogr. Anal. 38 (2006) 5–22, https://doi.org/10.1111/j.0016-7363.2005.00671.x.

[75] S. Mehrotra, R. Bardhan, K. Ramamritham, Urban form as policy variable for climate-sensitive area planning under heterogeneity: a geographically weighted regression approach, Area Dev. Policy 5 (2020) 167–188, https://doi.org/10.1080/23792949.2019.1609368.

[76] P.A.P. Moran, The interpretation of statistical maps, J. R. Stat. Soc. Ser. B Methodol. 10 (1948) 243–251, https://doi.org/10.1111/j.2517-6161.1948.tb00012.x.

[77] A.S. Fotheringham, C. Brunsdon, M. Charlton, Geographically Weighted regression: the Analysis of Spatially Varying Relationships, Wiley, Chichester, England ; Hoboken, NJ, USA, 2002.

[78] Y.Q. Ang, Z. Berzolla, C. Reinhart, Smart meter-based archetypes for socioeconomically sensitive urban building energy modeling, Build. Environ. (2023) 110991, https://doi.org/10.1016/j.buildenv.2023.110991.

[79] scikit-learn, 2.3. Clustering [WWW document], Scikit-Learn 132 (2023). URL, https://scikit-learn.org/stable/modules/clustering.html#k-means.

[80] scikit-learn, 1.11. Ensembles: gradient boosting, random forests, bagging, voting, stacking [WWW Document], Scikit-Learn 132 (2023). URL, https://scikit-learn.org/stable/modules/ensemble.html#forest.

[81] scikit-learn, 1.10. Decision trees [WWW Document], Scikit-Learn 132 (2023). URL, https://scikit-learn.org/stable/modules/tree.

[82] M. Rashid, K. Kampschroer, J. Wineman, C. Zimring, Spatial layout and face-to-face interaction in offices—a study of the mechanisms of spatial effects on face-to-face interaction, Environ. Plan. B Plan. Des. 33 (2006) 825–844, https://doi.org/10.1068/b31123.

[83] I. Konstantzos, S.A. Sadeghi, M. Kim, J. Xiong, A. Tzempelikos, The effect of lighting environment on task performance in buildings – A review, Energy Build 226 (2020) 110394, https://doi.org/10.1016/j.enbuild.2020.110394.