Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

Defining archetypes of mixed-use developments for improved urban building energy modelling

Zhongming Shi [1], Heidi Silvennoinen [1], Arkadiusz Chadzynski [3], Aurel von Richthofen [1], Markus Kraft [2,3,4], Stephen Cairns [1], Pieter Herthogs [1]

released: November 12, 2021

[1] Singapore-ETH Centre
at CREATE
1 Create Way
CREATE Tower, #06-01
Singapore, 138602

[2] Department of Chemical Engineering
and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

[3] CARES
Cambridge Centre for Advanced
Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

[4] School of Chemical
and Biomedical Engineering
Nanyang Technological University
62 Nanyang Drive
Singapore, 637459

Preprint No. 285

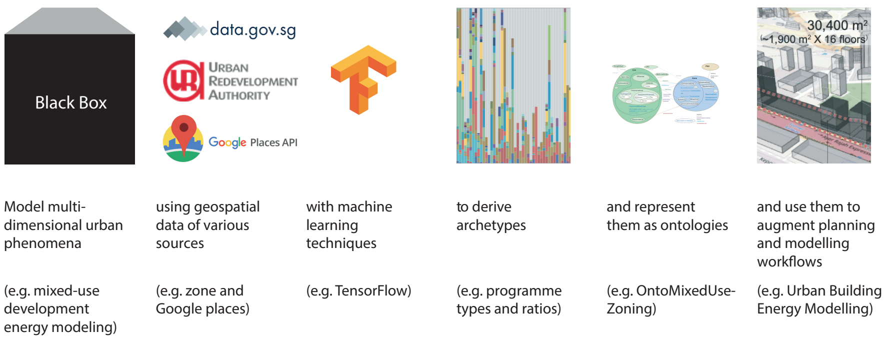
Keywords: Urban development; city planning; master plan; land-use; zoning; function; plot; GPR; GFA; knowledge graph; Semantic Web; Web Ontology Language; Google Maps; CEA; Semantic City Planning Systems; machine learning; Tensorflow; Singapore

Edited by
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

E-Mail: mk306@cam.ac.uk
World Wide Web: `https://como.ceb.cam.ac.uk/`

Abstract

Urban planning relies on the definition, modelling and evaluation of multidimensional phenomena for informed decision-making. Urban Building energy modelling, for instance, usually requires knowledge about a building's each use type's energy use profile and surface area. We do not have a detailed understanding of such information for mixed-use developments, which are gaining prominence in urban planning. In this paper, we developed a methodology to quantitatively define the characteristics of mixed-use developments using archetypes of programme profiles (ratios of each programme type) of a city's mixed-use plots. We applied our methodology in Singapore, resulting in 163 mixed-use zoning archetypes using Singapore's master plan data and Google Maps Place Type data. We also demonstrated how these archetypes augment the urban building energy modelling workflow for energy demand forecasts and energy supply system design. The archetypes definitions are represented and stored as a machine-readable ontology, improving adoption by other researchers and building towards an automated workflow. The archetypes have many potential urban planning applications beyond energy modelling.



Highlights

- We develop a methodology to understand mixed-use developments using master-planning data and Google Maps data.
- We formulate 163 mixed-use zoning archetypes in Singapore using machine learning methods.
- We demonstrate how these archetypes improve Urban Building Energy Modelling.
- We represent these archetypes as an ontology called ontoMixedUseZoning.

Contents

1 Introduction

Mixed-use developments feature in master plans of cities around the world [6, 12, 36]. These developments are zoned in such a way that they generally allow a variety of different land uses to coexist on the same (land) plot. For example, a plot assigned with a mixed-use *zoning type* in a city's master plan could combine a mix of *land use types*, such as residential, commercial, and office uses. Zoning legislation often specifies permitted and non-permitted combinations of uses at different levels of granularity (e.g. at the zoning, land use, or programme level).

Mixed-use developments can be implemented simply to increase conveniences for inhabitants (i.e. increased access to amenities), and mixing uses with various busy hours may also improve an urban quarter's liveliness throughout a day [17, 19]. More importantly, mixed-use developments improve urban sustainability in a number of ways. For example, certain combinations of mixed uses may significantly improve the efficiency and the cost-effectiveness of urban energy systems [30]. Mixed-use developments can reduce vehicular travel, as some trips are replaced by walking [33] and hence save urban transport-related energy and reduce carbon emissions [3]. As a result, we argue that mixed-use developments have become an important planning instrument for urban planning towards a sustainable future.

However, while mixed-use developments are ubiquitous in urban planning, we do not have a detailed understanding of what constitutes mixed-use, other than a combination of uses. A land-use zoning plan - often the only source of information available about future urban areas - does not provide uses and sizes at different granularities. We lack quantified definitions or archetypes to represent the different types of mixed-use developments that exist in terms of the numbers, kinds and distributions of uses that they contain.

Yet such information about the specific size and use of buildings is required for many different urban analyses and simulations, such as agent-based mobility modelling [e.g. 16] and urban building energy modelling [e.g. 9], the latter being the focus of this paper. An urban building energy modelling (UBEM) simulation needs, as inputs, the type and size of each UBEM use type in each unit of each building in the urban area under consideration. The quality of inputs has a large impact on the model's outcome, as different UBEM use types and their sizes can imply widely different energy use intensities and (peak) operating hours. Thus the lack of definitions and archetypes for mixed-use developments hinders our ability to analyse and simulate their benefits concerning urban sustainability targets.

To address this, we have developed an approach to derive typologies of *mixed-use zoning archetypes*, which can also be considered as "patterns" of mixed-use developments, from a collection of urban datasets. These typologies of existing mixed-use developments could help city planners envision and plan particular types of developments and help city scientists to more precisely model the impacts of mixed-use development. To facilitate the use of our typologies by such stakeholders, we represent them in a machine-readable ontology, which defines the characteristics, hierarchy and semantic relationships of these mixed-use zoning archetypes. This will enhance cross-domain interoperability and reusability of the derived mixed-use zoning archetypes, particularly within the context of Semantic City Planning Systems [40]. While our approach can be adopted for other cities and their mixed-use plots, in this paper, we specifically derive mixed-use zoning archetypes for Singapore and demonstrate how these archetypes improve a UBEM workflow in a case study.

Singapore is a fitting context for this study due to the prevalence of its mixed-use developments. In Singapore, zoning and land use planning is carried out by the Urban Redevelopment Authority (URA), mainly via the Singapore Master Plan. This is a statutory land use plan, updated every five years, which guides Singapore's development over 10 to 15 years [36]. Figure 1 shows a UML (Unified Modeling Language) Class Diagram of the zoning types included in Singapore's most recent 2019 Masterplan. Of these 32 zoning types, ten allow and encourage mixed uses, particularly commercial uses [32]. Mixed-use development is ubiquitous in Singapore, with these ten zoning types representing ~10% of all zoned plots in Singapore's current master plan (authors' calculation based on Singapore's 2019 Master Plan). In this work, we used data on *programme types*, representing a higher granularity than land use types, to link UBEM use types to zoning types.

Given the importance of mixed-use development in Singapore, an improved understanding of its nature could be key to meeting Singapore's ambition to significantly improve urban energy performance and reduce carbon emissions by 2030 [15]. Specifically, understanding the different uses that mixed-use development is made up of is necessary for carrying out UBEM analyses, which in turn are necessary for finding the most energy-efficient land-use distribution. Many UBEM tools require UBEM use types and their ratios as inputs. For example, EnergyPlus [7] and City Energy Analyst [9] directly utilise or adapt use types defined in standards published by ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) or SIA (Swiss Society of Engineers and Architects). These standards provide energy use profiles per unit floor area for each UBEM use type, such as multi-family residential or restaurants, which have specific energy use intensities and schedules.

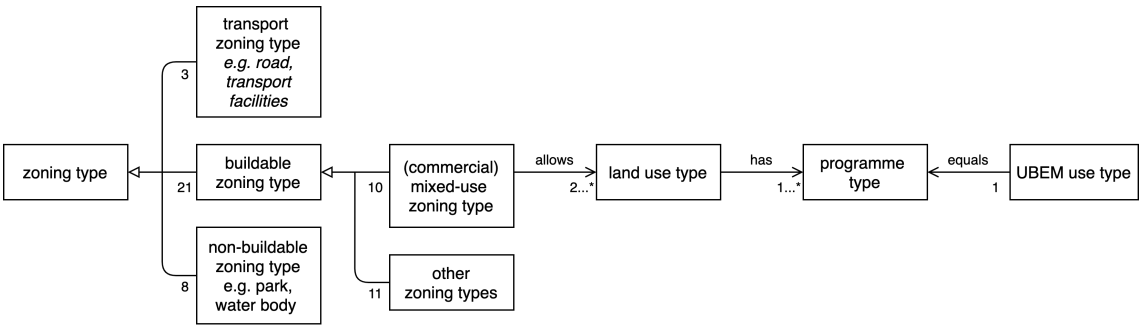


Figure 1: *UML Class diagram linking zoning types and UBEM use types. The intermediate classes and relations have been defined by the authors.*

The *purpose of this paper* is to link a master plan's zoning types and UBEM uses types to improve the accuracy of urban building energy performance assessment. Currently, to bridge this interoperability gap, urban planners and energy engineers may estimate a mixed-use plot's UBEM use types and their ratios based on experience or rules of thumb [29]. In this work, we first bridge the gap between zoning types and UBEM use types by formulating mixed-use zoning archetypes at the programme level using urban datasets and machine-learning techniques. Second, we create an ontology that links open zoning and UBEM-related data from different sources in a machine-readable format, providing a foundation for future applications that would be able to automatically assess the energy performance of different mixed-use zoning types. Third, we apply the formulated mixed-use zoning archetypes in a case study to demonstrate their impacts on UBEM for mixed-use developments in Singapore.

Section 2 provides necessary background information for what follows. Section 3 presents our five-step method for formulating each mixed-use plot's programme profiles and defining the mixed-use zoning archetypes, which are implemented as the ontoMixedUseZoning ontology. Section 4 presents the results and validates our results using measured data. Section 5 introduces a case study to demonstrate the impact of our work in urban building energy modelling. Section 6 reflects on the contributions and limitations of this work. Finally, Section 7 summarises our work and presents future research outlooks.

2 Background

This section provides background information on the main ideas combined in our methodology. It reviews the use of archetypes in research and modelling related to urban planning and design (Section 2.1), the data sources of zoning types and programme types (Section 2.2), the use of machine learning in urban morphology (Section 2.3) and the use of applied Ontology to improve data interoperability (Section 2.4).

2.1 Archetypes

In urban planning and design, archetypes are typically used to express commonalities between individual objects or concepts, such as buildings, areas or land uses [23]. Archetypes are commonly used in simulation-based studies on urban form for multiple urban qualities, such as urban daylighting [26], urban vitality [38], or on-site renewable energy use [31]. The use profiles required as inputs in UBEM analyses can also be considered archetypes. In UBEM tools, archetypes of different UBEM use types are used as inputs [1] and such archetypes summarise highly variable data related to building use and occupancy. Generally, these archetypes are formulated based on mass data of existing buildings and urban contexts. The archetypes developed in the present work differ from other archetypes through their focus on mixed-use plots, which are not accurately represented by existing archetypes.

2.2 Data sources

Our mixed-use archetypes are based on data from the Singapore government and Google Maps. Zoning data are collected from the Singapore government's open data platform, which provide the geolocation, zoning type and Gross Plot Ratio (GPR, the ratio of a plot's gross floor area to the plot area) for each plot in Singapore. Singapore's Master Planning Act documents the land use types allowed in each zoning type [32]. Programme data, i.e. data on the uses that take place in smaller units of the built environment, such as units in buildings or areas in parks, can be used as a proxy for UBEM use data. We considered obtaining programme data for each plot from two different sources: OpenStreetMap and the Google Place API service of Google Maps. We use Google data in this work due to its greater likelihood of accuracy, i.e. the inclusion of all existing programmes in operation. This accuracy is achieved thanks to regular updates on the operational status of Google Places, made both by Google itself and by business owners and casual Google Maps users. For these reasons, Google Maps data have also been used in the past to conduct UBEM simulations [14, 24, 43].

2.3 Machine learning to derive urban archetypes

Our method to formulate archetypes is similar to previous efforts to formulate quantitative archetypes of urban forms using unsupervised machine learning techniques such as clustering. For example, Schirmer and Axhausen [28], Vialard Vialard [39] and Shi et al. [31] have used clustering to formulate archetypes of street block typologies in Zurich, Atlanta and Singapore. In all these works, archetypes are based on geometry-related data, such as street block areas, block dimensions and GPRs. To our best knowledge, no studies have formulated archetypes based on the types of programmes and the ratios between each programme type's Gross Floor Area (GFA).

We use multivariate regression to estimate the GFA of each programme type. This technique is commonly used to analyse data with multiple unknown variables (in this case, the floor area of each programme type), and has been used to link characteristics of urban form to, for example, air pollution concentration [34], urban heat island effect [18] and energy demand [5].

2.4 Applied ontology

We created an ontology to create machine-readable links between conceptual classes such as programme types and zoning types, and instances of geospatial data. Applied Ontology implies the application of ontological approaches (from the philosophical branch of Ontology) to specific knowledge domains (in our case, land-use planning) and is commonly practised in Information Science and Computer Science (e.g. the Knowledge Representation and Reasoning field of Artificial Intelligence). Ontologies represent relationships between concepts within the same or between different knowledge domains, as well as the kinds of properties that objects or concepts can have. Such ontologies can then be used by Semantic Web Technology (SWT) applications that take diverse data as inputs and perform operations that take into account such relationships and properties. For example, our ontology could provide the basis for an application that performs automatic energy analyses for plots based on their zoning type.

There are two main benefits to basing applications on ontologies. Firstly, an application can reuse existing ontologies, while the new parts of the ontology can in turn be reused by other applications. This saves time in the application development process. The second benefit is their ability to be easily extended incrementally to include more concepts from the same or from different domains. These new concepts and relationships can be easily linked to the existing ontology without having to fundamentally alter the original data structure. For this reason, ontologies are often used to improve data interoperability, including in the urban domain [40]. For example, our ontology of mixed-use archetypes could be extended with concepts related to travel, such as defining an attractive destination as a plot that is the target of a certain number of trips and exemplifies a certain kind of archetype. With data of the profiles and trips, the ontology could then infer that a certain plot is such an attractive destination. In this way, the ontology allows linking data from two different domains and sources to derive new knowledge.

It is possible to construct ontologies manually, based on domain expertise, as well as through various automated or semi-automated means [2]. In the manual approach, the concepts and relationships in the ontology are modelled by a human based on their understanding of the domain. In the data-informed approach, the ontology is abstracted from data through various methods, including clustering and natural language processing techniques [2]. We adopt a hybrid approach.

3 Methodology

This section introduces our data-informed methodology for linking zoning types to UBEM use types via programme types, using Singapore as a demonstration. Figure 2 presents the methodology's five-step workflow. The first step (Section 3.1) is to collect the master plan data and Google Place data for all mixed-use plots. The second step (Section 3.2) is to group Google Places into Programme types, based on similarity from the point of view of building function and occupancy. The third step (Section 3.3) is to formulate, for each mixed-use plot, a *programme profile* (i.e. the floor area of each programme on the plot, as a percentage of the plot's GFA). In the fourth step (Section 3.4), similar programme profiles are grouped together to form the mixed-use zoning archetypes. In the fifth step, the archetypes and programme profiles are linked to other urban planning concepts in an ontology (Section 3.5).

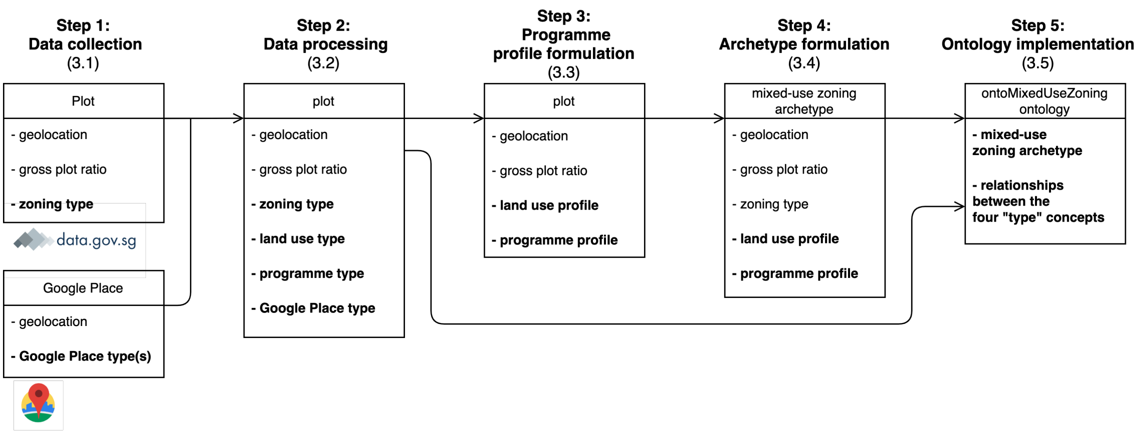


Figure 2: *The methodology to derive mixed-use zoning archetypes and develop the ontoMixedUseZoning ontology consists of five steps.*

3.1 Data collection

Our data collection began with a survey of Singapore's 2019 Master Plan from the Singapore Government's open data platform data.gov.sg. The master plan data contain each plot's geolocation, GPR and zoning type, as well as descriptions of the land use types allowed by each zoning type. Of the master plan's 32 different zoning types, our study focuses on the ten mixed-use zoning types: Commercial, Residential with Commercial at 1st storey, Commercial & Residential, Commercial & Institutional, Business Park, Hotel, White, Business 1-White, Business 2-White, Business Park-White. Each of these mixed-use zoning types allows commercial uses to some degree. Figure 3 shows the 10,961 mixed-use plots assigned one of these zoning types.

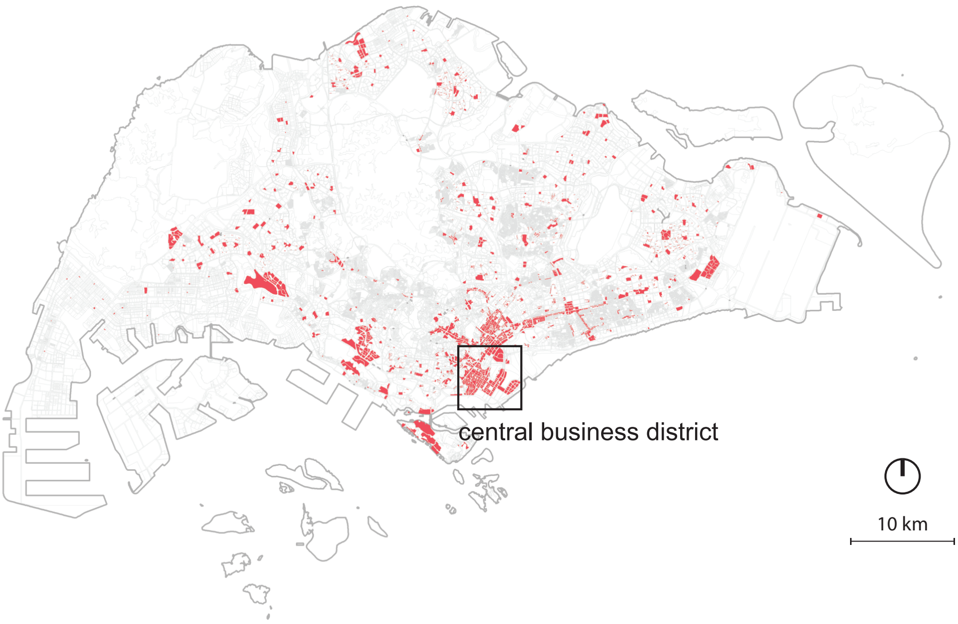


Figure 3: *Mixed-use plots in Singapore. The plots highlighted in red are assigned one of the ten mixed-use zoning types in the Singapore Master Plan 2019.*

Then, we collected data on Google Places, which are points of interest that appear on Google Maps. We filtered these data using two criteria: location and Place Type. The location filter simply included those Google Places located on mixed-use plots, using the Nearby Search requests of the Google Places API service [11]. We then excluded some of these places based on their Place Types, which are labels created by Google to describe the function of each Place (each Place can have one or more Types). A Place was included in our study if at least one of its Place Type labels corresponded to a land-use allowable in our ten mixed-use zones, according to the master plan data discussed above [32]. Overall, 46 such corresponding Place Types were found (documented in Appendix A). This filtering process resulted in a dataset with 57,730 Google Places and their attributes: geolocation (latitude and longitude), name of Place, address and Google Place ID (a unique identifier of the Google Maps platform). Data collection took place in January 2021. Of the 10,961 mixed-use plots in Singapore, 3,064 have a complete set of plot data, i.e. both a GPR and at least one Google Place data point.

3.2 Data processing

The data was processed by merging Place Types that are highly similar from the perspective of building function and occupancy. Specifically, many Google Place Types are similar in their energy use intensity and temporal distribution, which are the main factors impacting an UBEM outcome. For example, the Place Types clothing_store and shoe_store are similar in both of these respects (and were thus combined into a single category, apparel_store), while a nightclub and a locksmith likely differ significantly (and were thus kept separate). This merging process resulted in 36 distinct categories which are henceforth referred to as Programmes. Each programme is composed of between one and three original Place Types, as shown in Table 4 in Appendix A. This mapping was then used to re-classify each Google Place according to its Programme type.

Using this processed data, the next step was to simply count the occurrences of each Programme type in each mixed-use plot. Table 1 presents an example of the processed data for a plot with the zoning type Commercial.

Table 1: *An example data point in our processed data: for each plot, we have compiled the zoning type, plot area, GPR, and the frequency of occurrence of each Programme type in the plot. For example, in this particular Commercial plot (home to a large mall), we found one gym, seven banks, 104 apparel stores, amongst others.*

			programme type counts													
plot ID	**GPR** $[-]$	**plot area** $[m^2]$	**gym**	**bank**	**apparel_store**	**electronic_store**	**car_dealer**	**pet_store**	**doctor**	**beauty_service**	**...**	**school**	**bar**	**department_store**	**movie_theatre**	**florist**
#10377	1.6	84,826	1	7	104	9	1	3	2	24	...	2	31	9	1	3

3.3 Programme profile formulation

We then used the processed data to formulate a programme profile for each of the 3,064 mixed-use plots. The programme profile expresses the floor area of each programme type found on a plot as a percentage of the plot's total GFA. Deriving the programme profiles thus required solving multivariate equations with the floor areas of each programme type on the left-hand side and the plot's GFA on the right. Such equations are possible to solve by combining data from many plots. However, a challenge is that some mixed-use plots may contain uses that are not marked as Google Places and hence not represented by our programme types, such as residential or office use. For example, a plot with the commercial zoning type might only have a few Google places, with most of the floor area being occupied by office use, which would not be included in our Programme data. We have thus added three programme types based on the Master Planning Act but not included in the list of Google Place Types as an addition to the 36 programme types. These additional three programme types are office, residential and hotel.

To minimise the influence of "unknown" programme types from Google Maps in the regression, we built our models using commercial plots that are clustered into groups sharing the same unknown programme types or having no such types. This allowed us to estimate the typical floor area of each of our programme types, which could then be used to create programme profiles for all mixed-use plots. Figure 4 shows the method for formulating programme profiles for the commercial plots. Each of the four sub-steps is described in more detail below.

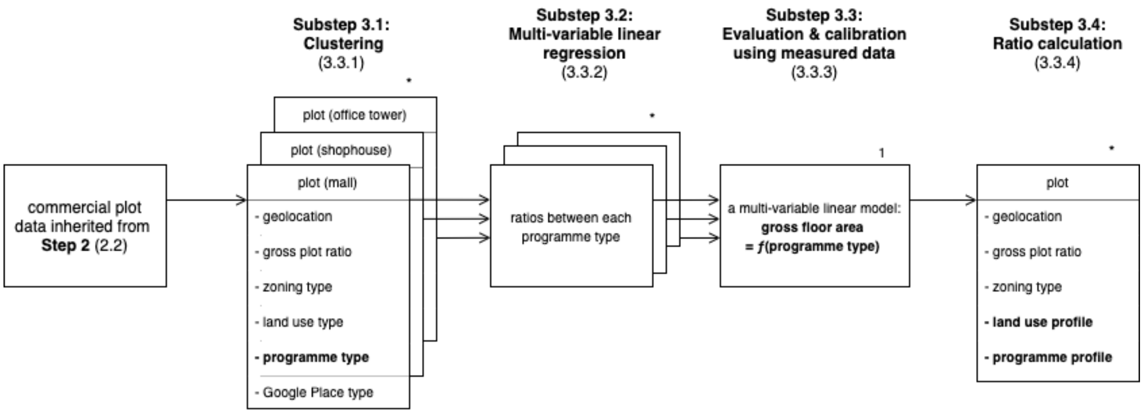


Figure 4: *Our method to determine each programme type's typical unit size has four substeps.*

3.3.1 Clustering

Even when focusing on the commercial plots, it proved difficult to find reasonable estimates of each programme type's floor area by solving the multivariate equations mentioned above. This is because commercial plots encompass vastly different kinds of developments, showing significant variability with respect to the presence of unknown programme types (i.e. residential, office, hotel room). Figure 5 shows three different types of built form commonly found in Singaporean commercial zones: office towers, malls and historical shophouses. A typical shophouse plot has a GPR of ~2.5 and contains a few commercial functions, but may also contain residential or office functions. A typical mall plot has a GPR of ~3 and a very high number of programmes, with no residential or office functions. An office tower plot typically has a GPR of ~10, with most of the floor area being devoted to office functions and only a small share to the programme types in our list in Appendix A.

We clustered the commercial plots to find more homogeneous groups of plots that contain similar or the same unknown programme type. After a comparison of different methods, we chose to cluster the plots using Ward's method for agglomerative hierarchical clustering [20, 27], based on each plot's GPR and programme count (the total number of all Google Places in that plot). This method was best able to distinguish between what are intuitively the main kinds of commercial plot types in Singapore, mentioned above: office towers, malls and shophouses. Our clustering ultimately resulted in five clusters, consisting of malls, mega malls, shophouses, office towers and a "leftover" cluster whose plots did not clearly fit any of these categories. Then, the multivariate equations to find programme profiles could be solved for those clusters of plots with little unknown floor area and a large plot count, in this case, the shophouse and mall clusters. In our data, these two types of commercial plots contain each of the 36 programme types in our ontology and combined they account for ~77% of all mixed-use commercial plots and ~28% of all the mixed-use plots that have full data (i.e. plots that also have a GPR listed in the master plan). All the non-commercial plots (the other nine mixed-use zoning types) contain a few of the programme types and are much less common than the commercial plots. These non-commercial plots were not used in the multivariate linear regression in the next sub-step.

(a) (b) (c)

Figure 5: *Three types of commercial plots in Singapore that differ significantly with respect to number of Google places and GPR: (a) malls; (b) office buildings in the central business district; and (c) historical shophouses.*

3.3.2 Multivariate linear regression

In the second sub-step, we fit multivariate linear regression models to the commercial plots of malls and shophouses to solve for the floor area of each of our 36 programme types. When sampling the data used for the regression (i.e. plot data sets containing 36 plots), we used the circular systematic sampling technique to ensure that all plot data had the same probability of being sampled. We applied it in two steps. First, we calculated the sampling interval, which equals the total sample size divided by $n = 36$. Second, by iterating over the plots in the dataset and selecting every nth plot in each iteration, we created a total number of plot data samples equalling the total number of plots. To create more samples, we shuffled the data and repeated these two steps 100 times for both the mall and shophouse clusters.

Multivariate regression could then be performed on this data with the assumption that each plot's built GFA equals the weighted sum of the floor areas of each programme type found on the plot, as follows:

$$GFA_j = \sum_{i=1}^{n} (S_i \times C_{i,j}) + U_j \tag{1}$$

Here, n is the number of programme types $[-]$, equal to 36 in our application; i is the ordinal number of a programme type in the list $[-]$; j ranges from 1 to the total number of plots $[-]$ (the total number of shophouse plots equals 750, and that of the malls equals 99); C, representing the variables in the linear regression model, is a programme type's count of occurrences $[-]$; S, representing the weights of the variables in the linear regression model, is a programme's typical unit size $[m^2]$; and U, representing the bias in the linear regression model, is the GFA of the unknown programmes outside the list of 36 programme types, such as the office programme type in an office tower. Each plot j has a permissible GFA ($pGFA$) $[m^2]$, calculated as

$$pGFA_j = A_j \times GPR_j \tag{2}$$

where j ranges from 1 to the total plot number $[-]$, A is the plot's area $[m^2]$, and GPR is the plot's GPR $[-]$ as listed in the master plan.

It is important to highlight that, as we did not have access to data on the actual built GFA for each plot, in this model, we assumed that each plot's built GFA equals its total permissible GFA set in the master plan (through the listed GPR). Built GFA could be both lower than permissible GFA (e.g. plots for which GPR has been increased, but where new or additional development has not taken place) or higher (e.g. through bonus GFA granted to developers when integrating certain policies, features, or amenities). Hence, this modelling assumption will introduce some inaccuracies, which are straightforward to resolve once accurate as-built GFA data becomes available.

Theoretically, feeding n sets of plot data into Equations 1 and 2 may inform the weights (S) and the bias (U) to fit the multivariate linear regression model. A multivariate regression model was fitted for each set of 36 mixed-use plot data using Tensorflow, an open-source software library for machine learning maintained by Google, which is often used to build classification or prediction models in urban science [e.g. 41, 42]. We divided the data into training and validation sets at a ratio of 0.8 and 0.2, respectively. We used the training data to build the multivariate linear regression model using the keras.Sequential model and the Adam Optimizer. The validation data was then used to test the model to avoid over-fitting. The tolerances for training and validating the model were kept at 0.2 and 0.03 in Mean Absolute Percentage Error.

In all, 884 models were created for the combined clusters of malls and shophouses. The weights of each linear regression model can be considered as the typical sizes of the programme type (R $[m^2]$) and the bias as the floor area of the unknown programme type. Many of these 884 models did not have a weight for all 36 programme types. This is because the sampled sets of 36 plots combined did not contain all 36 programme types. In such cases, we filled the missing weight of a programme type with that of a known similar programme type. Table 2 shows how we divided the 36 programme types into five groups of similar size, based on the authors' professional judgement. Whenever a programme type in a regression model had a missing weight, we replaced the missing value with the average weight of known programme types belonging to the same group in Table 2.

Table 2: *Five groups of programme types of similar unit sizes, used to replace missing programme types with similar alternatives.*

Group	36 programme types
A	beauty_service, pharmacy, liquor_store, laundry, convenience_store, movie_rental, locksmith, florist, hardware_store
B	restaurant, bar/cafe, doctor, veterinary_care, bank, embassy, home_goods_store, car_repair, bicycle_store, jewellery_store
C	gym, bowling_alley, night_club, art_gallery, museum, library, book_store, school, electronics_store, car_rental, furniture_store, apparel_store
D	supermarket, department_store, lodging, movie_theatre
E	casino

3.3.3 Model selection and calibration

In the next sub-step, we reduced the resulting 884 linear regression models down to a single one through processes of selection and calibration. We evaluated the models by comparing their predicted results for programme profiles (i.e. the mix of programmes and their typical floor areas) to a measured programme profile of a particular plot. This plot contained a mega mall and no unknown programme types, allowing us to eliminate the unknown types from the calibration. We derived its programme profile by tracing the mall's floor plan and calculating the floor areas of its 500 Google Places, grouped by programme type. Of our 36 programme types, 27 were represented in the measured data. Such a highly diverse mix of programme types was the reason for using this plot for calibration.

We used clustering to identify those multivariate regression models that yielded programme profiles most similar to the measured data. First, we merged the measured programme profile with the 884 modelled profiles, resulting in a dataset of 885 programme profiles. We then performed hierarchical clustering on the set. Finally, we selected all the regression models with programme profiles belonging to the same cluster as the measured programme profile.

Then, we calibrated the weights of the variables in the selected models relative to our benchmark plot's measured data using equation 3, scaling our selected regression model's GFA to equal that of benchmark mega mall plot.

$$R_{i,j} = \frac{GFA_x}{\sum_{i=1,j=1}^{n,m}(R_{i,j} \times C_{i,x})} \times S_{i,j} \tag{3}$$

Here, n is the number of programme types in the measured plot$[-]$; i is the ordinal number of the programme types in the list of types $[-]$; m is the number of linear models selected for calibration $[-]$; j is the ordinal number of the linear models in the list of models $[-]$; S is the weight of the variables in the linear models $[-]$; GFA_x is the measured plot's GFA

$[m^2]$, which is 135,721 is this demonstration; C_x is the count of each programme type $[-]$ in the measured plot as in Table 1.

Then, we merged these models and used the mean of the calibrated weights ($\overline{R_{i,j}}$ $[m^2]$) to produce the final multivariate linear equation, which models a plot's floor area ($gGFA$) for the 36 programme types based on the Google data, which should be the same or less than a plot's permissible GFA.

$$gGFA = \sum_{i=1}^{n=36} (\overline{R_{i,j}} \times C_i) \tag{4}$$

3.3.4 Programme ratio calculation

In the fourth and final sub-step, we formulated programme profiles for each plot in our data (including non-commercial plots). This was done by applying the calibrated linear regression model (i.e. Equation 4) to all 3,064 mixed-use plots and calculating the floor areas of each programme type found on the plot. When the GFA of the 36 known programme types ($gGFA$ $[m^2]$) is smaller than the plot's permissible GFA ($pGFA$ $[m^2]$), the difference equals the floor area of the plot's unknown programme type (uS $[m^2]$). When $gGFA_j$ is greater or equal to $pGFA_j$, the plot's unknown programme types are not considered, as shown in Equation 5 and 6.

$$A_{i,j} = \begin{cases} \dfrac{pGFA_j}{gGFA_j} \times (R_i \times C_{i,j}), & \text{when } gGFA_j \geqslant pGFA_j \\ \\ R_i \times C_{i,j}, & \text{when } gGFA_j < pGFA_j \end{cases} \tag{5}$$

$$uS_j = pGFA_j - gGFA_j \tag{6}$$

Here, i is a programme type's ordinal number in the list of 36 programme types, ranging from 1 to 36 $[-]$; j ranges from 1 to the total plot number $[-]$; A is a programme's GFA in a plot $[m^2]$; C is a programme type's count in a plot $[-]$.

Two programme types in particular, office and residential, represent the vast majority of unknown (i.e. other) programme types (uS) present in Singapore's ten mixed-use zoning types. A final modelling question was how to address casinos, which were not present in the plots used to create the linear regression model, as there are only two casinos in Singapore. We chose to set the floor area of casinos at 30,000 $[m^2]$, as both of Singapore's casinos have a floor area of ~30,000 m^2 [10, 25].

3.4 Archetype formulation

We then derived more precise mixed-use zoning archetypes (i.e. sub-types) based on the results of the first round of clustering described in Section 3.3.1. The method used was again hierarchical clustering.

In Section 3.3.1, the commercial plots were divided into five clusters based on GPR and programme count, resulting in clusters matching shophouses, malls, mega malls, office towers and mixes. We then divided each of these clusters further, based on the plots' programme profiles derived previously, and their GPRs, resulting in several sub-clusters. By selecting the medoid of the cluster (i.e. the programme profile, which has the shortest combined distance to all other profiles within the cluster), we selected the most representative plot for each sub-cluster, and they are defined as mixed-use zoning archetypes.

3.5 ontoMixedUseZoning ontology implementation

After defining the mixed-use zoning archetypes based on empirical data, we formally defined and represented the archetypes as an ontology, named ontoMixedUseZoning. This ontology contains classes that represent concepts relevant to mixed-use zoning, properties of these classes, as well as relationships between classes.

The ontology was created in three steps. First, we conceptualised the relationships between the key concepts of interest to us: plots, mixed-use zoning types, programmes allowed in each zoning type, and archetypes. In doing so, we consulted the URA 2019 Masterplan Written Document [36], which lists all zoning types in Singapore and specifies what kinds of land uses are allowed in each type. We then manually matched these land uses to our programme types as closely as possible, thereby creating an (indirect) link between each zoning type and their allowed and disallowed land uses and programme types. Each Archetype in the ontology is connected to the Programme types through the property containsProgramme, and to the geospatial plot objects that the Archetype represents. The second step was to formalise this conceptual diagram with Protégé, an ontology editor, in a machine-readable format using the Web Ontology Language (OWL2). Lastly, we used Protégé's in-built HerMit reasoner and the OntoDebug plugin to ensure that the ontology is consistent (lacks contradictions), and able to make correct inferences.

4 Results

Section 4.1 presents the programme profiles formulated for each mixed-use plot. Section 4.3 presents the mixed-use zoning archetypes made based on the programme profiles. Section 4.2 presents a validation of our results using measured data programme profiles of several plots. Section 4.4 presents the ontoMixedUseZoning ontology.

4.1 Programme profiles for individual plots

This section presents our approximations for the programme profiles in each of the 3,064 mixed-use plots in Singapore that have a complete set of plot data (both a listed GPR and at least one Google Place data point). The programme profiles of these 3,064 mixed-use plots, i.e. GFA distributions of different programmes, were formulated as described in Section 3.3. Figure 6 maps all the programme profiles for the 3,064 all plots. Figure 13 in

Appendix B illustrates the bars for these plots' formulated programme profiles by zoning type.

Figure 6: *Our total set of 3,064 plots and their formulated programme profiles mapped across Singapore.*

4.2 Validation

We then partially validated the derived mixed-use plot profiles. Due to the difficulty of finding programme-level data - which is of course the motivation for our study - it was not possible to validate all 3,064 programme profiles and 163 archetypes. To get a sense of the accuracy of our formulated archetypal programme profiles, we compared three of our results against data collected manually from three plots with malls; the malls are 321 Clementi, Takashimaya and City Square Mall (for details on each mall, see Appendix C). We focused on malls because their programmatic data (types and floor areas of programmes) is readily available and accessible in comparison to data on other types of mixed-use developments. In contrast to buildings containing residential and office functions, malls typically publish detailed directories with precise and current data on the areas occupied by each programme in the mall. The validation data was collected in mid-May

2021 from mall directory maps by tracing the floor plan of each level, classifying the different sections in the floor plan according to programme, and calculating the ratio of each programme type as a share of the mall's total floor area.

Figure 7 compares the formulated programme profiles to the empirical measurements for the three malls. Consult Table 6 in Appendix C for a detailed quantitative comparison between the formulated and the measured programme profiles. Considering the diversity of mixed-use plots in Singapore, our method produces fairly accurate programme profiles: for each of the three cases, our archetype programme profiles identify dominant programme types and their shares of the total floor area closely. The comparison also shows differences, which to a large extent can be explained by two factors: particular establishments with unusual (i.e. non-archetypal) floor areas and outdated Place Type data. For example, 321 Clementi has an unusually large gym, spread across two floors, occupying a larger portion of the mall than our model estimated. City Square Mall has less of its GFA dedicated to department stores than estimated by the model, as our Google Place data, on which the model was built, contains two department stores that permanently closed during the four-month period between the Google Place data collection and the measurement. The accuracy of the formulated programme profiles may thus improve when using updated Google Place data.

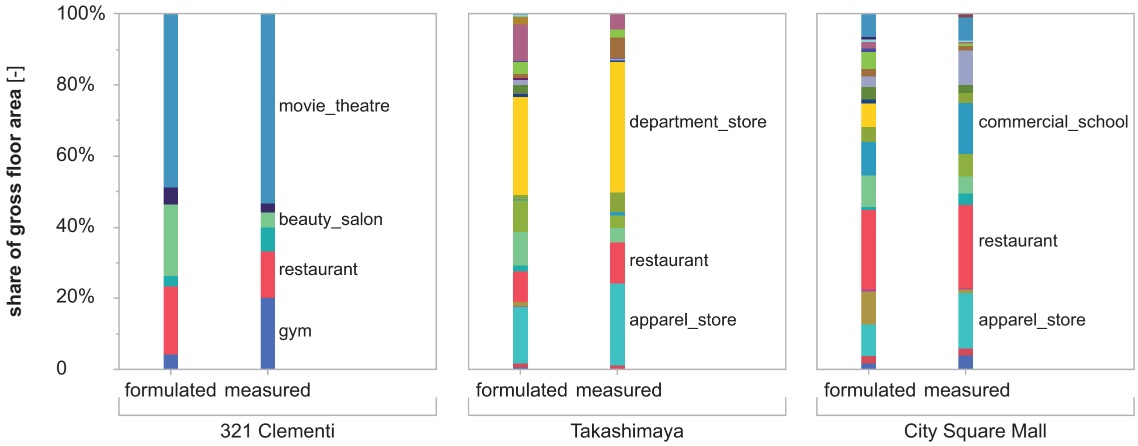


Figure 7: *Comparison of formulated programme profiles (left) to empirical measurements (right) for three malls in Singapore. Considering the diversity of mixed-use plots in Singapore, our method produces fairly accurate programme profiles, identifying dominant programme types and their shares of the total floor area. Consult Figure 6 for the full legend of the colours.*

4.3 Mixed-use zoning archetypes

This section presents the mixed-use zoning archetypes. Using the methods in Section 3.4, we derived 163 archetypes from the 3,064 mixed-use plots' programme profiles and their GPRs. We computed the medoids of each cluster to represent the whole group; these 163 representatives are our mixed-use zoning archetypes for Singapore. Figure 8 presents the processes of producing the mixed-use zoning archetypes, and Figure 9 plots the mixed-use zoning archetypes' GPRs and programme profiles.

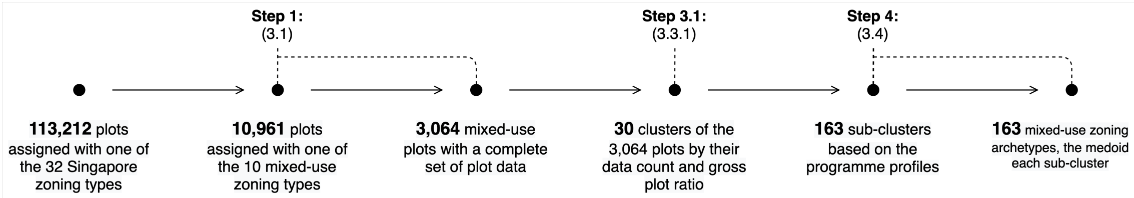


Figure 8: *Step 1, Step 3.1, and Step 4 produce 163 mixed-use zoning archetypes from 113,212 plots of the whole Singapore.*

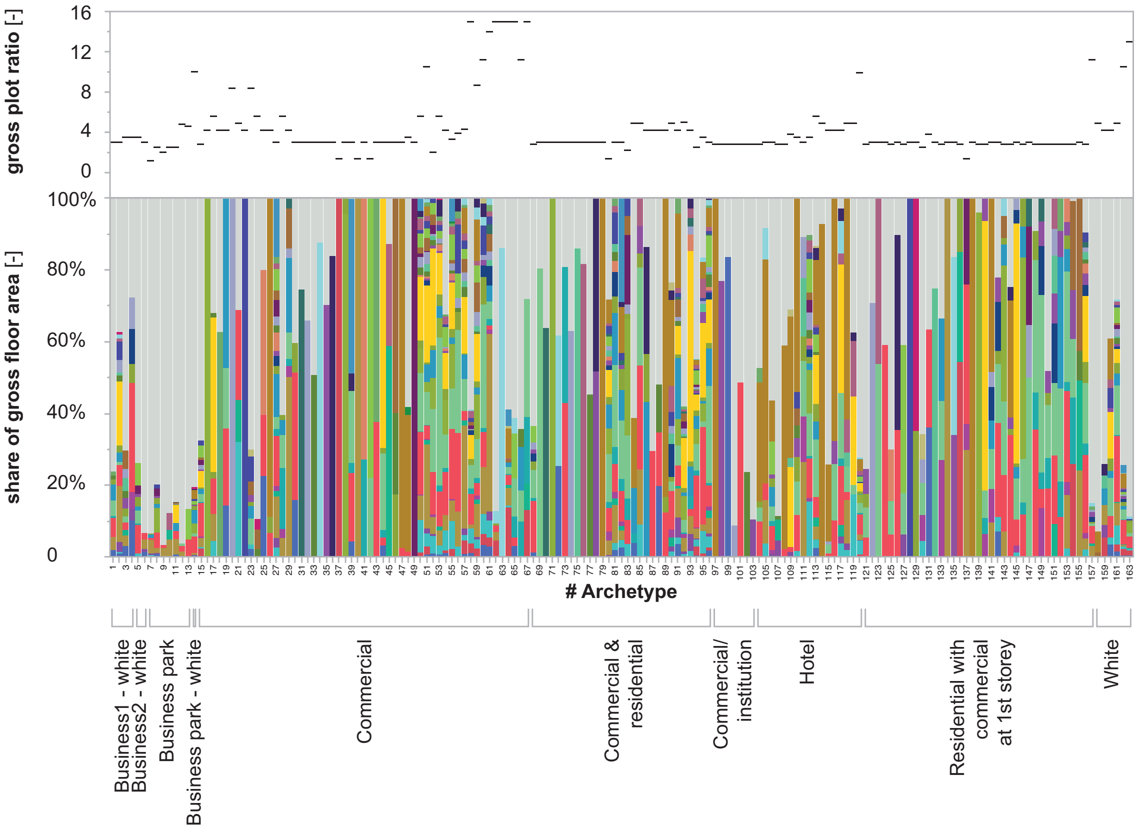


Figure 9: *The 163 mixed-use zoning archetypes found in Singapore's ten types of mixed-use zones. Each of the ten mixed-use zoning types is associated with one or more archetypes. Each archetype is characterised by a particular combination and distribution of programme types and programme floor areas, as well as a particular GPR. Consult Figure 6 for a legend of the programme colours.*

4.4 ontoMixedUseZoning ontology

As the final step of this study, we implemented the 163 mixed-use zoning archetypes as an ontology, called ontoMixedUseZoning. As discussed in Section 1, this will enhance cross-domain interoperability and reusability of the derived archetypes, particularly within the context of Semantic City Planning Systems [40]. OntoMixedUseZoning was implemented using Protégé, an ontology editor, using Web Ontology Language 2 (.owl format).

Figure 10 illustrates ontoMixedUseZoning diagrammatically, showing how the classes Plot, ZoningType, LandUseType, ProgrammeType, Archetype and DataSource are related to each other. These relations can be summarised as follows. Each Plot has a ZoningType. The ZoningType, in turn, allows certain land-use types, as specified in the URA 2019 Masterplan [32]. Each LandUseType corresponds to a ProgrammeType, which is based on the empirical data collected. The data collected in the present study covers Commercial and Hotel Programmes, which therefore have more specific subprogramme types. Finally, the ontology includes an Archetype class containing individual archetypes that each link to a mixed-use zoning type through the property hasArchetype. Each individual archetype also links to its constituent programme types via the property includesProgramme. In addition, each individual archetype has a value of GPR. The programme ratio of each archetype (e.g., that archetype 1 consists of 20% beauty service and 80% hotel programme) is not included in the ontology but can be added by an external application that uses the ontology.

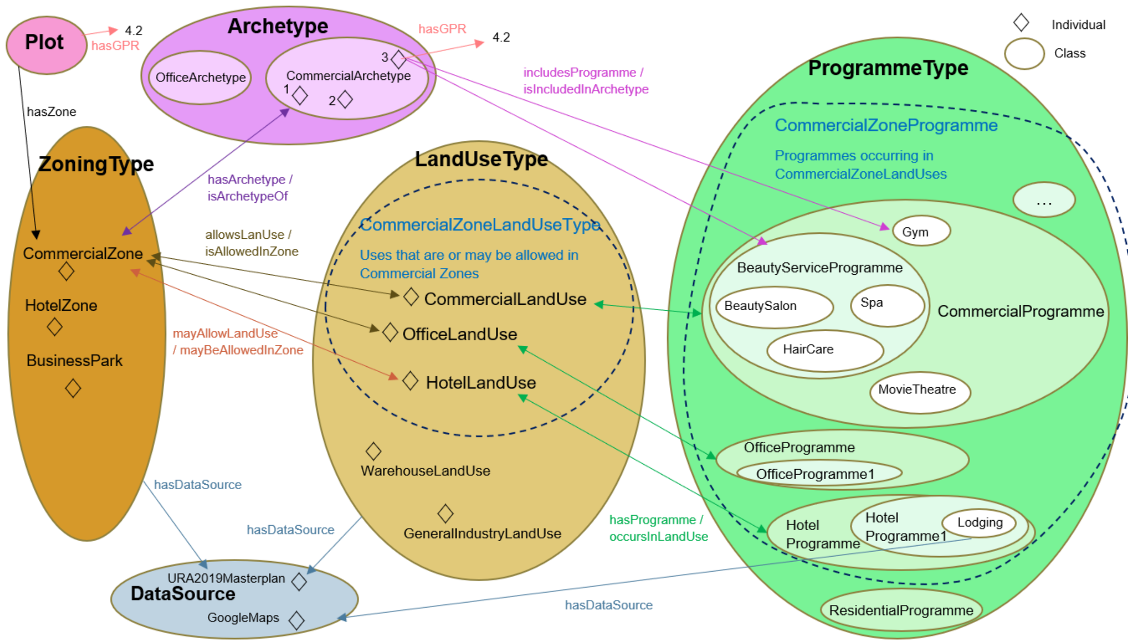


Figure 10: *Diagram illustrating the ontoMixedUsezoning ontology for Singapore, showing how the classes Plot, ZoningType, LandUseType, ProgrammeType, Archetype and DataSource are related to each other.*

5 Applying the archetypes in an UBEM workflow

This section presents a case study in which we use one of our mixed-use archetypes to conduct an urban energy analysis for a plot in Singapore, showcasing how they can be applied to improve an UBEM workflow. Section 5.1 describes the case study's plot and its context. Section 5.2 provides the metrics we use for our urban energy analysis. Section 5.3 presents the energy analysis results for our case study.

5.1 Case study's plot and its context

We chose a commercial plot with a GPR of 4.2 located close to Singapore's central business district. The plot, shown in Figure 11, is the site of a heritage railway station and the Cantonment MRT (mass rapid transit) station, which is currently under construction. This plot was chosen as it reflects and represents a trend of transit-oriented mixed-use development of Singapore in an age of fast MRT system expansion.

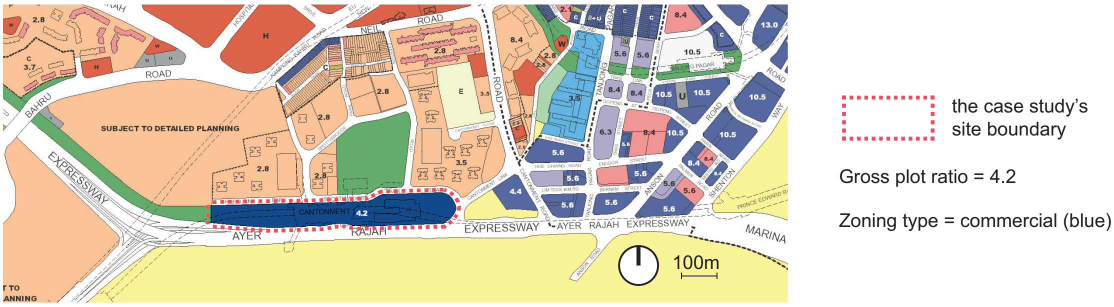


Figure 11: *Our case study plot (marked with a dashed red line) is the site of a heritage railway station and the future Cantonment MRT station. The base image shows the Singapore Master Plan 2019 [37].*

We populated the plot with a proposed building geometry and programme, representing a potential new development on the site for which we want to forecast the energy demand and on-site photovoltaic (PV) electricity yields; this is illustrated in Figure 12. To do this, we set up a UBEM simulation using the City Energy Analyst (CEA). UBEM simulations require building geometries and programmes as input. The buildings' forms (e.g. site coverage, number of towers and their footprint area) were determined based on a recent survey of high-density mixed-use urban form in Singapore [31], and the geometry was scaled to reach the target GPR of 4.2. The buildings were then assigned the programme profile of a mixed-use archetype that met the following two criteria: a commercial zoning type and a GPR of ~4.2. Figure 12 (b) illustrates the programme profile of Archetype #54, which is the only archetype that meets both criteria.

5.2 Urban building energy analysis

In this step, we modelled, simulated, and assessed the energy system of our proposed new development. Specifically, we executed UBEM simulations and conducted analy-

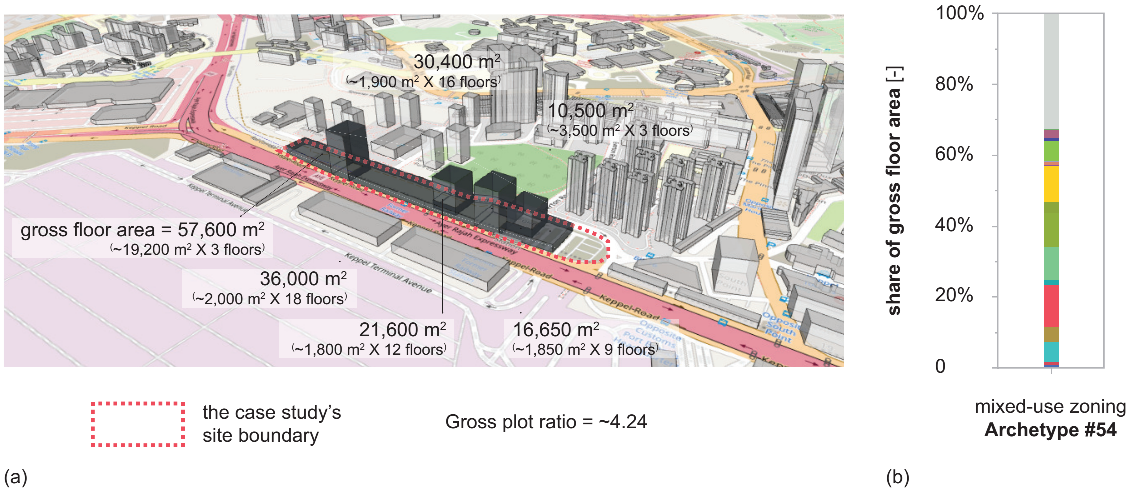


Figure 12: *(a) We added a potential new development to our case study site, with built volumes representing six buildings; (b) The building volumes are assigned the programme profile of Archetype #54. See Figure 6 for a legend of the programme colours. The geometry and the programmes are used as inputs for UBEM simulations using the CEA.*

ses assessing cooling supply system designs for the case study. The UBEM simulations were executed using the City Energy Analyst (CEA), an open-source Python-based toolbox capable of simulating urban solar radiation, forecasting building energy demand and designing thermal energy supply systems. All simulations in this work use CEA Version 3.4 [35]. Details are described in Appendix D.

We modelled a commonly-used centralised cooling supply system. Our simulation determines the size and the cost-effectiveness of the chillers. A measure for the cost-effectiveness is the chiller's Capacity Factor (CF), which measures to what extent the chillers are used to their maximum installed capacity - as chillers represent the largest overall expense in a cooling supply system, engineers aim to maximise their cost-effectiveness. The CF is calculated as

$$CF = \frac{\sum_{t=1}^{8,760} D_{qc}(t)}{QC_{ch} \times 8,760h} \tag{7}$$

where t represents all hourly time steps in a year, D_{qc} is the plot's hourly cooling demand in $[kWh]$, and QC_{ch} is the chiller's installed capacity in $[kW]$. CEA determines the installed size of the chillers based on the plot's peak cooling demand.

5.3 Comparison

To what extent does using our data-informed programme archetypes impact the energy simulation results in comparison to conventional inputs? To answer this question, we ran the simulations again using CEA's default use profile (i.e., retail use only) for a whole commercial site. This retail use also most closely matched the programme profile used in our data-informed simulation (i.e., Archetype #54, consisting mainly of the programmes department_store and apparel_store). The results of simulations are listed in Table 3. This comparison shows that using archetypes as an input has a significant effect on all the major outputs of our energy simulations, especially the peak cooling and final electricity demands, which are important for the design of both the PV and centralised cooling supply system. In Section 6.1, we discusses the impact of an archetype-based simulation.

Table 3: *The CEA simulation results for the case study when using archetype programme profiles and when using conventional (retail) use types.*

Metrics [*unit*]	**using Archetype programme profile**	**using conventional use type**
annual final electricity demand [MWh]	~51,683	~54,595
annual cooling demand [MWh]	~75,616	~60,365
peak hourly cooling demand [MWh]	~23	~17
installed annual chiller capacity [MWh]	~202,110	~148,114
chiller capacity factor [$-$]	~37%	~41%

6 Discussion

6.1 Impacts on urban building energy modelling results

In the energy simulations, the detailed programme profile was coupled with each programme type's corresponding CEA use type's energy use intensity and hourly occupancy schedules. When using the programme profile, the annual final electricity demand from the city grid is ~5% more than when using the default retail profile, while the peak hourly final electricity demand from the city grid is from ~35% to ~84% greater than in the simulation using retail use types. Such significant differences in energy demand forecast could impact the energy supply system design, as the temporal distribution of energy demand over time is crucial in energy system sizing and operation.

The different hourly cooling demand has a significant impact on the sizing and the operation of the centralised cooling energy supply system. The chiller capacity factors, measuring the cost-effectiveness of the chillers, are ~37% and ~41%, respectively. Although the results indicate the latter is seemly more cost-effective, the size of the chiller may not meet the peak cooling demand if the plot's programme profile ever develops following the path of Archetype #54. The required size of the chillers is ~36% greater.

6.2 Impacts on early stage master-planning

Aside from the impacts on UBEM analyses, the case study also demonstrates the mixed-use zoning archetypes introduced in the present paper can support the early stages of master-planning. The mixed-use zoning archetypes link programme types directly to the zoning types. In a greenfield project, urban planners can consult the archetypes for exemplary programme profiles when programming a mixed-use plot. The programme profiles provide much more detailed building occupancy information, which can inform not only building energy modelling but also other domains that rely on a detailed breakdown of programmatic GFA, such as urban mobility, real estate, or urban design, for example.

6.3 Potential impacts of the ontoMixedUseZoning ontology

The mixed-use archetypes presented in this paper are integrated into an information system for master planning - a Semantic City Planning System [40] - called Cities Knowledge Graph (CKG). The ontoMixedUseZoning ontology links a set of concepts describing particular master planning aspects (i.e. mixed-use programme profiles) to particular conceptual geospatial boundaries present in a city (i.e. plots). In the CKG, these plots are represented using the OntoCityGML conceptual schema, an ontology of the CityGML format [4].

By linking different domain ontologies, this approach enables us to execute multi-domain geospatial queries that combine geospatial characteristics of city models with land use planning characteristics, particularly of mixed-use zones. This allows us, for example, to query for city objects and programme profiles based on the target plot's geo-location, zoning type and GPR. This allows urban planners to query up-to-date programme profiles of any mixed-use plots that match their querying criteria.

As pointed out by Chadzynski et al. [4], the use of ontologies within dynamic geospatial knowledge graphs can help to address common issues related to keeping digital city models up to date. By representing our results in the ontoMixedUseZoning ontology, we provided the semantic architecture to automatically link the frequently updated Google Place data with the planning authority's master plan data. This introduces a degree of automation to the process of updating mixed-use zoning archetypes. Moreover, the default Open World Assumption (OWA) of knowledge graphs based on semantic web standards (i.e. that absent information is not necessarily false) adds the flexibility to link other planning or city-related domain aspects to the CKG as needed, incrementally, without jeopardising complex multi-domain inferences in the form of logical contradictions. This includes linking other types of Google data, such as the Google popular times and Google reviews, or data from other sources, such as OpenStreetMap or a city government's own data on building or land use, and thus helping the urban development professionals to assess the built urban environment more accurately and comprehensively.

Similarly, the CKG is also linked to other knowledge graph endeavours within The World Avatar (TWA) knowledge graph, enabling even broader queries and insights across various multi-domain interdependencies. Such integration not only allows sharing and querying linked datasets but also applying specific functionalities present in the TWA semantic information system, such as its Parallel World Framework capabilities [8], which, in case of master planning, would allow for scenario analysis and multi-factor optimisation of land use plans using various criteria.

6.4 Limitations

The first limitation concerns the validation, which is limited in scope. As building floorplan and programme datasets are not commonly available, our validation is limited to manually measuring the GFA of each programme in mixed-use plots for which such information is publicly available. In the case of Singapore, this limits us to shopping mall plots.

The second limitation is that our predictions of typical GFAs for particular programmes are solely based on commercial plots that feature either malls or historic shophouses, disregarding other types of mixed-use plots. This is the case because these two types of commercial plots represent the majority of the mixed-use plots in Singapore, and hence the largest dataset to apply a multivariate regression to. Nevertheless, applying the multivariate regression separately to each zoning type that allows mixed-use, if sufficient data were to be available, would likely lead to more archetype-specific (and hence context-specific) results.

The third limitation is the temporal accuracy and frequency of our base data and how it reflects changes. The mixed-use zoning archetypes were derived from data collected in January 2021, during the COVID-19 pandemic, when commercial programmes had likely already been impacted by major governmental restrictions and changes in usage patterns. Therefore, the archetypes at least partially reflect that particular state of affairs at that particular time. Hence, to more precisely inform the programming of future or existing mixed-use plots, it would be best to collect data at different times and repeatedly generate the programme profiles of mixed-use archetypes, enabling the monitoring of shifts in programmatic use patterns. The need for regular, preferably automated model updates is also illustrated by the fact that our empirical data used in the validation, obtained a few months after the Place Types data, already showed the effect of the pandemic on particular programme types (department stores). By defining our archetypes as a machine-readable ontology and integrating them into a Semantic City Planning System that, we aim to fully automate the derivation of mixed-use archetypes, addressing this limitation.

The fourth limitation is the universal applicability of our results and methodology. The mixed-use zoning archetypes presented in this paper and the related ontoMixedUseZoning ontology are specific to the Singaporean context and the temporal nature of the data, limiting the scope of application to Singapore and reducing the representativeness of the archetypes over time. However, the methodology presented in the present paper can be used to produce specific mixed-use archetypes for different urban contexts or for different points in time.

The fifth limitation concerns the UBEM tool used in our case study. The CEA provides 17 commonly-used pre-defined UBEM use types with the possibility to add user-defined extensions. This implies that if CEA had more pre-defined UBEM use types, we would have likely seen even more variation in the comparison between energy simulation results using the archetype-based programme profiles and the baseline.

7 Conclusions and outlooks

This work presents our methodology to quantitatively define the characteristics of mixed-use zones and plots in cities. Our methodology derives archetypes of programme profiles (the distributions of use types and their GFAs) of mixed-use plots in a particular city. The resulting set of archetypes can be used by built environment professionals as standard representations of mixed-use programme types commonly found in their particular city, for example in urban modelling applications such as UBEM.

Our methodology combines machine-learning methods (hierarchical clustering and multivariate analysis) to derive programme profile for each mixed-use plot. The most representative profile of each cluster becomes the mixed-use archetype. The archetypes definitions are represented and stored as a machine-readable ontology in .owl format. We applied our methodology to the city-state of Singapore, resulting in a total of 163 archetypes for the 10 types of mixed-use zoning used in Singapore's master plan. Based on Singapore's master plan data and Google Maps Place Type data, we also formulated a list of 36 different programme types.We applied the resulting archetypes for Singapore by demonstrating their use in an urban building energy modelling workflow, simulating energy demand forecasts of an urban development proposal and evaluating its particular energy supply system design. We discussed how the use of our mixed-use zoning archetypes affected simulated results when compared to conventional building use representations used in UBEM workflows.

This work contributes to the state of the art in three main ways. Firstly, it introduces a new concept, programme types, linking the zoning types used in land-use planning and the (building) use types used in UBEM and other urban modelling approaches. Secondly, these programme types are then used to create mixed-use zoning archetypes, which have many potential city planning or city science applications. For example, they can inform the master planning process by serving as examples of possible planning outcomes for a plot with a similar GPR and zoning type. Such archetypal programme profiles also improve various urban modelling workflows that rely on representations of use type and use surface area, UBEM being the example highlighted in this work. Our archetypes remove the need for each modelling effort to independently determine mixed-use distributions when simulating mixed-use urban areas, reducing modelling time and increasing comparability between simulations. Thirdly, the representation of these archetypes as the ontoMixedUseZoning ontology allows for robust interoperability between the frequently updated Google Place data and the master-planning data. The semantic representation of the archetypes lays the foundation for a live and automated workflow to update the programme profiles and the archetypes at regular intervals when integrated into a Semantic City Planning System, providing the underlying technology to enable the monitoring of shifts in programmatic use patterns over time.

We want to highlight five research outlooks. First, as mentioned, we aim automate the workflow introduced in this work in our Cities Knowledge Graph system. This would enable us to update the programme profiles and archetypes frequently, whenever Google

Place data are updated. Second, the ontoMixedUseZoning ontology can be extended to include programmes for all other zoning types in the Singapore master plan, not just mixed-use zoning types. Third, we want to apply the methodology introduced in this work to urban contexts beyond Singapore, as mixed-use urban planning initiatives are gaining ground worldwide. Fourth, in addition to this work's use case in urban building energy modelling, we will likely apply the ontoMixedUseZoning ontology in other types of urban modelling, such as transport modelling. Finally, the introduction of these archetypes introduces the need to adapt and extend UBEM use types in simulation software tools such as the City Energy Analyst.

Acknowledgements

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. Markus Kraft gratefully acknowledges the support of the Alexander von Humboldt foundation.

A Appendix – Google Place types and programme types

The 46 Google Place types are beauty_salon, hair_care, spa, pharmacy, liquor_store, laundry, convenience_store, movie_rental, locksmith, florist, hardware_store, restaurant, bar, cafe, gym, bowling_alley, dentist, doctor, physiotherapist, pet_store, veterinary_care, bank, post_office, embassy, home_goods_store, car_repair, car_wash, bicycle_store, jewellery_store, night_club, art_gallery, museum, library, book_store, school, electronics_store, car_dealer, car_rental, furniture_store, clothing_store, shoe_store, supermarket, department_store, lodging, movie_theatre and casino.

Table 4: *36 programme types correspond to one or multiple Google Place types.*

Programme types	Google Place types
beauty_service	beauty_salon, hair_care, spa
pharmacy	pharmacy
liquor_store	liquor_store
laundry	laundry
convenience_store	convenience_store
movie_rental	movie_rental
locksmith	locksmith
florist	florist
hardware_store	hardware_store
restaurant	restaurant
bar/cafe	bar, cafe
doctor	doctor, dentist, doctor, physiotherapist
veterinary_care	pet_store, veterinary_care
bank	bank, post_office
embassy	embassy
home_goods_store	home_goods_store
car_repair	car_repair, car_wash
bicycle_store	bicycle_store
jewellery_store	jewellery_store
gym	gym
bowling_alley	bowling_alley
night_club	night_club
art_gallery	art_gallery
museum	museum
library	library
book_store	book_store
school	school
electronics_store	electronics_store
car_rental	car_rental, car_dealer
furniture_store	furniture_store
apparel_store	clothing_store, shoe_store
supermarket	supermarket
department_store	department_store

Table 4 – continued from previous page

Programme types	Google Place types
lodging	lodging
movie_theatre	movie_theatre
casino	casino

B Appendix – 3,064 formulated programme profiles

Figure 13: *Programme profiles formulated for all the 3,064 plots by zoning types.*

C Appendix – measured programme profiles

This section describes the three malls used for the validation of the Tensorflow model results: Clementi 321, City Square Mall and Ngee Ann City. The rationale for choosing these particular malls for the validation was four-fold: data was easily available for them; their plots had the same (Commercial) zoning type; they were located on a single plot; and they differed in terms of size, geographical location and programme distribution, thereby increasing the generalisability of the validation. Table 5 presents information on the location, zoning properties, floor area and Google Place counts for each mall.

Table 5: *Characteristics of the malls used for validation*

	Takashimaya	**321 Clementi**	**City Square Mall**
plot ID	#10282	#10342	#10186
Location	Orchard (Central)	Clementi (Western)	Farrer Park (Central)
Zoning type	Commercial	Commercial	Commercial
Plot GPR	6.3	3	unknown
Plot size $[m^2]$	26,865	2,420	11,103
Floor area $[m^2]$	111,665	5,572	31,517
Programme count	327	39	239

Table 6: *A quantitative comparison between the formulated (F) and the measured (M) programme profiles for the three selected malls of validation.*

programme type	**programme type ratios**					
	321 Clementi (F)	**321 Clementi (M)**	**Takashimaya (F)**	**Takashimaya (M)**	**City Square Mall (F)**	**City Square Mall (M)**
gym	4.2%	20.1%	0.6%	0%	1.7%	3.9%
bank	0%	0.2%	2%	1.1%	1.1%	2%
apparel_store	0%	0%	15.8%	23%	8.9%	15.5%
car_repair	0%	0%	0.2%	0%	0%	0%
electronics_store	0%	0%	0.012%	0%	9.2%	1.1%
veterinary_care	0%	0%	0%	0%	0.6%	0.2%
restaurant	19.2%	12.9%	8.6%	11.5%	22.4%	23.5%
doctor	2.8%	6.7%	1.7%	0%	0.8%	3.2%
beauty_service	20.1%	4.2%	9.4%	4.1%	8.8%	4.7%
supermarket	0%	0%	8.7%	3.5%	0%	6.4%
school	0%	0%	0.3%	0.9%	9.5%	14.2%
bar/cafe	0%	0%	1.2%	5.5%	4.1%	2.8%
department_store	0%	0%	27.6%	36.8%	6.6%	0%
laundry	0%	0%	0.8%	0.4%	1.1%	0.1%
convenience_store	0%	0%	0.2%	0%	0%	0%
furniture_store	0%	0%	2.3%	0%	3.5%	2.2%
home_goods_store	0%	0%	1.4%	0.6%	3%	9.7%
locksmith	0%	0%	0.7%	0.1%	0%	0%
book_store	0%	0%	1.1%	5.9%	2.1%	1.2%
pharmacy	0%	0%	3.4%	2.2%	4.8%	0.8%
florist	0%	0%	0.3%	0%	1%	0%
jewellery_store	0%	0%	10.2%	4.3%	1.8%	0.5%
lodging	0%	0%	2%	0%	0%	0%
art_gallery	0%	0%	0.3%	0%	0%	0%
liquor_store	0%	0%	0.6%	0.2%	0.8%	0.3%
night_club	4.9%	2.6%	0.2%	0%	0.7%	0%
movie_theatre	48.8%	53.3%	0%	0%	6.6%	6.4%
bowling_alley	0%	0%	0%	0%	0%	1.2%

D Appendix – City Energy Analyst

The building energy demand simulations of CEA use an hourly resistance-capacitance model [22]. The solar heat gain calculations of CEA use DAYSIM [21], a validated software which considers the mutual shading between building geometries. The infiltration airflow simulations of CEA uses a constant envelope leakage rate [9].

The basic inputs for the CEA energy demand forecast are building geometries, building occupancy profile as well as building construction and system properties. In our analysis, we used as inputs the building geometries shown in Figure 12 (a). Occupancy profiles were derived from the buildings' programme profiles, shown in Figure 12 (a), using occupancy schedules adapted from the ASHRAE (the American Society of Heating, Refrigerating and Air-Conditioning Engineers) standards. The building construction and system properties are pre-defined in the CEA database for Singapore. Based on these inputs, the CEA energy demand forecast calculates the hourly electricity demand for lighting and appliances and thermal energy demand for space cooling and water heating.

With CEA, we calculate the hourly PV electricity yield. First, we reuse the solar radiation results in the energy demand forecast. Then, using PV technology properties stored in the CEA database, we calculate the plot's total PV electricity yield. We assume that PV panels are installed on those building envelope surfaces receiving at least 200 $[kWh/m^2]$ of solar radiation, based on a recent study on the carbon emissions of PV electricity and Singapore's city grid electricity [13].

E Appendix – 163 mixed-use zoning archetypes

The .csv file containing 163 mixed-use zoning archetypes' GPRs and ratios of each programme type's GFA is attached below.

References

[1] ASHRAE Project Committee 90.1. Schedules and internal loads for Appendix C, 2019.

[2] I. Bedini. Automatic Ontology Generation : State of the Art, 2007. URL https://www.semanticscholar.org/paper/Automatic-Ontology-Generation-%3A-State-of-the-Art-Bedini/81989605fba59a415a28603ce709566cc6ebc45c.

[3] W. Bowley and R. Evins. Assessing energy and emissions savings for space conditioning, materials and transportation for a high-density mixed-use building. *Journal of Building Engineering*, 31:101386, 2020. doi:10/gmffbh.

[4] A. Chadzynski, N. Krdzavac, F. Farazi, M. Q. Lim, S. Li, A. Grisiute, P. Herthogs, A. von Richthofen, S. Cairns, and M. Kraft. Semantic 3D City Database — An enabler for a dynamic geospatial knowledge graph. *Energy and AI*, 6:100106, 2021. doi:10/gmfgm3.

[5] S. Chang, N. Saha, D. Castro-Lacouture, and P. P.-J. Yang. Multivariate relationships between campus design parameters and energy performance using reinforcement learning and parametric modeling. *Applied Energy*, 249:253–264, 2019. doi:10/gk9ndt.

[6] City of Toronto. Official Plan of Toronto, Aug. 2017. URL https://www.toronto.ca/city-government/planning-development/official-plan-guidelines/official-plan/.

[7] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, and J. Glazer. EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319–331, 2001. doi:10/fb2jvz.

[8] A. Eibeck, A. Chadzynski, M. Q. Lim, K. Aditya, L. Ong, A. Devanand, G. Karmakar, S. Mosbach, R. Lau, I. A. Karimi, E. Y. S. Foo, and M. Kraft. A Parallel World Framework for scenario analysis in knowledge graphs. *Data-Centric Engineering*, 1:e6, 2020. doi:10/ghgpvz.

[9] J. A. Fonseca, T.-A. Nguyen, A. Schlueter, and F. Marechal. City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts. *Energy and Buildings*, 113:202–226, 2016. doi:10/f8dcvk.

[10] Genting Singapore. Resorts World Sentosa, 2021. URL http://www.gentingsingapore.com/#!/en/business/resorts-world-sentosa.

[11] Google. Google Maps Platform - Place Types, Places API, 2021. URL https://developers.google.com/maps/documentation/places/web-service/supported_types.

[12] Greater London Authority. The London Plan 2021, Mar. 2021. URL https://www.london.gov.uk//what-we-do/planning/london-plan/new-london-plan/london-plan-2021.

[13] G. Happle, Z. Shi, S. Hsieh, B. Ong, J. A. Fonseca, and A. Schlueter. Identifying carbon emission reduction potentials of BIPV in high-density cities in Southeast Asia. *Journal of Physics: Conference Series*, 1343:012077, 2019. doi:10/gmv24x.

[14] G. Happle, J. A. Fonseca, and A. Schlueter. Context-specific urban occupancy modeling using location-based services data. *Building and Environment*, 175:106803, 2020. doi:10/gmv24z.

[15] O. Ho. S'pore calling for universal climate deal. *Strait Times*, Nov. 2015.

[16] A. Horni, K. Nagel, and K. W. Axhausen. *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, Aug. 2016. doi:10.5334/baw.

[17] J. Jacobs. *The Death and Life of Great American Cities, Reissue Edition*. Vintage Books, New York, 1992.

[18] Z. Liang, S. Wu, Y. Wang, F. Wei, J. Huang, J. Shen, and S. Li. The relationship between urban form and heat island intensity along the urban development gradients. *Science of The Total Environment*, 708:135011, 2020. doi:10/gk9ndv.

[19] V. Mehta and J. K. Bosson. Revisiting Lively Streets: Social Interactions in Public Space. *Journal of Planning Education and Research*, 41(2):160–172, 2021. doi:10/gjtnsz.

[20] G. W. Milligan. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45(3):325–342, 1980. doi:10/fmn927.

[21] MIT Sustainable Design Lab. Daysim. MIT Sustainable Design Lab, May 2020.

[22] M. Mosteiro-Romero, D. Maiullari, M. Pijpers-van Esch, and A. Schlueter. An Integrated Microclimate-Energy Demand Simulation Method for the Assessment of Urban Districts. *Frontiers in Built Environment*, 6, 2020. doi:10/gmv24v.

[23] V. Oliveira. *Urban Morphology*. The Urban Book Series. Springer International Publishing, Cham, 2016.

[24] J. Parker, A. Hardy, D. Glew, and C. Gorse. A methodology for creating building energy model occupancy schedules using personal location metadata. *Energy and Buildings*, 150:211–223, 2017. doi:10/gbvkph.

[25] Sands Casino. Marina Bay Sands, Singapore, Official Site Sands Casino, 2018. URL https://www.sandscasino.com/singapore/casino-marina-bay-sands.html.

[26] E. Saratsis, T. Dogan, and C. F. Reinhart. Simulation-based daylighting analysis procedure for developing urban zoning rules. *Building Research & Information*, 45 (5):478–491, 2017. doi:10/ghxp3k.

[27] SAS Institute Inc. JMP Pro, 2019.

[28] P. M. Schirmer and K. W. Axhausen. A multiscale classification of urban morphology. *Journal of Transport and Land Use*, 9(1), 2015. doi:10/gfw6h9.

[29] Z. Shi, S. Hsieh, J. A. Fonseca, and A. Schlueter. Street grids for efficient district cooling systems in high-density cities. *Sustainable Cities and Society*, 60:102224, 2020. doi:10/gmfgmz.

[30] Z. Shi, J. A. Fonseca, and A. Schlueter. Floor area density and land uses for efficient district cooling systems in high-density cities. *Sustainable Cities and Society*, 65: 102601, 2021. doi:10/gmfgmw.

[31] Z. Shi, J. A. Fonseca, and A. Schlueter. A parametric method using vernacular urban block typologies for investigating interactions between solar energy use and urban design. *Renewable Energy*, 165:823–841, 2021. doi:10/gmfgmx.

[32] Singapore Government. The planning act - master plan written statement, 2019.

[33] B. R. Sperry, M. W. Burris, and E. Dumbaugh. A Case Study of Induced Trips at Mixed-Use Developments. *Environment and Planning B: Planning and Design*, 39 (4):698–712, 2012. doi:10/f387zm.

[34] Y. Tao, Z. Zhang, W. Ou, J. Guo, and S. G. Pueppke. How does urban form influence PM2.5 concentrations: Insights from 350 different-sized cities in the rapidly urbanizing Yangtze River Delta region of China, 1998–2015. *Cities*, 98:102581, 2020. doi:10/gjnbrj.

[35] The CEA team. City Energy Analyst v3.4.0. Zenodo, 2020.

[36] Urban Redevelopment Authority. Singapore Master Plan 2019, 2019.

[37] Urban Redevelopment Authority. URA SPACE, 2019. URL https://www.ura.gov.sg/maps/?service=MP.

[38] A. van Nes and Z. Shi. Network Typology, Junction Typology and Spatial Configuration and Their Impacts on Street Vitality in Singapore. *ResearchGate*, Jan. 2009.

[39] A. Vialard. *A Typology of Block-Faces*. Doctoral Thesis, Georgia Institute of Technology, Atlanta, 2013.

[40] A. von Richthofen, P. Herthogs, M. Kraft, and S. Cairns. Semantic City Planning Systems (SCPS): A Literature Review, 2021. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/270/.

[41] J. R. Vázquez-Canteli, S. Ulyanin, J. Kämpf, and Z. Nagy. Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities. *Sustainable Cities and Society*, 45:243–257, 2019. doi:10/gksc2h.

[42] Y. Yao, H. Liang, X. Li, J. Zhang, and J. He. Sensing Urban Land-Use Patterns By Integrating Google Tensorflow And Scene-Classification Models. *arXiv:1708.01580 [cs]*, Aug. 2017.

[43] T. Yoshida, Y. Yamagata, and D. Murakami. Energy demand estimation using quasi-real-time people activity data. *Energy Procedia*, 158:4172–4177, 2019. doi:10/gmffbg.