JSTARS-2024-00657

Urban morphologic structures retrieved by satellite imagery correlate with socio-economic household data − insights from the City of Kigali, Rwanda

A. Braun, C. Khouri, O. Assmann, G. Warth, M. Schultz and V. Hochschild

Abstract—**A substantial body of research exists on the use of remote sensing in urban contexts. However, only a limited number of studies have contributed to our understanding of the socio-economic conditions of different urban areas. This research aims to demonstrate the potential of very high-resolution images and geospatial data by examining the interrelations between socio-economic data retrieved from household surveys in the city of Kigali and spatial data on urban morphology retrieved by satellite imagery. As the surveys yielded large amounts of data of varying levels of measurement (categorical and numeric), we present different methods of statistical correlation, data mining, and machine learning to highlight socio-economic patterns within the spatial data. The results demonstrate a significant correlation between the share of different building types, building density, average building heights and distances to public infrastructure with a range of surveyed data, including building properties, household members, financial resources and overall lifestyle habits. This highlights the potential of remote sensing and geospatial data to provide valuable insights into the socio-economic conditions of urban areas. It also underscores the importance of using advanced statistical methods, data mining, and machine learning to enhance our understanding of urban morphology and its socio-economic implications. However, it is important to acknowledge the limitations of such approaches, including the lack of information on ownership, potential for false inference and the direction of causation, which require further investigation.**

***Index Terms*—urban areas, social factors, calibration, demography, scientific visualization, data mining**

I. Introduction

The objective of urban planning is to ensure that all residential groups in urban areas have access to essential services and amenities, including infrastructure for the supply and disposal of resources. Additionally, urban planning strives to guarantee that residents have access to vital urban functions, such as education, healthcare, administration, arts, recreation, and security. This is a fundamental responsibility of urban planning, and especially challenging in rapidly growing urban agglomerations. Their dynamic growth leads to both spatial and social segregation and to partially limited access to public services and infrastructures and, therefore, to socio-economic inequalities [1]. These problems are furthermore increased by environmental issues, such as water and air pollution and health-related threats arising from heat and flooding as direct consequences of climate change [2], [3], [4]. Especially in megacities of the Global South, but also in large parts of Asia, unclear settings and conflicting rationalities underline the need for evidence-based planning to tackle these challenges on a political and administrative level [5]. However, reliable and up to date data on the city's infrastructure, its residents, their needs and consumption patterns are needed for this process to be effective, but barely available to decision makers [6]. Their collection, for instance by public census, surveys, or questionnaires, can deliver precise and comprehensive insights into selected households' situations, which help to understand how lifestyles and demands for goods and services differ within the city [7], [8], [9]. However, such surveys are time-consuming and expensive, and can only cover small fractions of the complex environment of the city which are to be determined by systematic sampling in advance [10], [11].

At the same time, many studies have indicated strong interrelations between urban structure and human behavior and consumption patterns over the entire globe: for instance, Salama et al. (2017) found patterns between emerging building types and lifestyle trends and among several Gulf cities [12], Kalbar et al. (2018) statistically identified impacts of the urban environment on food consumption patterns and transportation preferences in Denmark [13], and Mattson (2020) proved the impact of urban density and household expenditures on transit across multiple small and medium US American cities [14]. However, data availability is strongly sector-dependent: Sensor networks and mobile phones emerged as valuable sources of information for studies on traffic and mobility [15], but other methods have to be applied to get insights into environmental challenges, public safety, and socio-economic inequalities within the complex structure of cites [16], [17]

Techniques of Earth Observation can contribute to tackling this complexity and to make rapidly developing cities more understandable by delivering spatially explicit, objective and temporally replicable information which are consistent over large areas [18], [19], [20]. Their applications cover the mapping of land use and land cover [21], the derivation of auxiliary variables which are related to specific phenomena (e.g., gentrification [22]) or groups of residents (e.g. people affected by poverty) [23]), and ultimately, the modelling of human and physical processes [24]. In this context, the mapping of urban structures from space a promising concept for the description of complex urban agglomerations, as well as for comparative studies between different cities, and has proven most effective in terms of replicability, transferability, and

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

consistency when it comes to the quantitative description of a cities morphology [25], [26]. The fundamental premise of this theory is that morphology can be expressed by a range of continuous variables, each providing complementary information content [27]. However, despite the unprecedented technical capabilities and increasing availability of satellite images, most studies are currently limited to the pure quantification of physical characteristics of a city while their interlinkages to socio-economical phenomena is declared the subject of future studies, and the actual delivery of actionable knowledge to planners, and its integration into actual socio-economic reasoning are missing or still to be proven [18].

Dividing urban areas into units of similar morphological characteristics has already been proven successful for climatological studies where areas of similar building density and height were aggregated to estimate their albedo and heat fluxes and finally their contribution to urban heat islands [28]. Similar findings based on techniques of remote sensing have been presented for the interlinkages between urban form and public transportation [29] and sustainable infrastructures [30], or between spatial patterns and the degree of quality of life [31], for instance. On the other hand, despite the many successful studies on urban remote sensing and its benefits with respect to data consistency and temporal integrity, others argue that physical characteristics can only be coupled with social science data to a limited degree, and that social data doesn't follow spatial grids or administrative units [24]. Another shortcoming addresses aspects which cannot be identified by remote sensing alone and could potentially lead to false conclusions: Taubenböck et al. (2009) name the ownership status of a building as an example that shows that a "high class" building is not necessarily inhabited by wealthy residents if they only live there on rent, and vice versa [32]. Lastly, Xing et al. (2024) argue that information retrieved by remote sensing and collected data on residents are of rather complementary nature and that the intended transfer of knowledge from satellite images into spatially-explicit socio-economic information is only one possible approach among many to achieve better understanding of a city and its residents [20].

To address the research gaps, questions, and shortcomings identified in the previous paragraphs, we want to extend existing studies by focusing on the needs and habits of the residents to put the hypothesis to the test that there is a statistical relationship between urban morphology as quantified by satellites and the underlying socio-economic conditions of the residents at certain spatial scales. More precisely, we investigate

- which are the aspects of the life of residents that have the strongest relations to urban morphology,
- which morphological urban structures are most relevant for this purpose,
- which statistical measures are suitable to describe the different types of relations of the above with respect to their level of measurement (nominal, ordinal, interval, ratio), and
- how significant the identified patterns are in terms of urban planning.

For this purpose, variables retrieved by satellites and geospatial data serve as predictive variables to explain socio-economic conditions as retrieved by household surveys.

As a case region, we chose Kigali, the capital of Rwanda, because of its dynamic development since the 1990s and annual growth rates of up to 10 %, underlining its need for information to assist decision making in urban planning [33].

This study combines the ideas of Warth et al. (2020) who demonstrated how attributes of dwellings, such as size, height, form and location, correlate with the socio-economic status of their residents [34], and the findings of Baffoe et al. (2022) who state that neighborhoods in Kigali are only partially defined by administrative boundaries but rather by morphology, accessibility and living conditions [35]. In chapter II, we outline our data collection, preparation and processing, followed by the statistical analysis of correlations between results of household surveys and spatial variables which are presented in chapter III. The main findings are then discussed and summarized at the end of the study.

In essence, our objective is to identify those specific parameters related to urban planning that are represented by remote sensing and that are therefore relevant and worthy of further investigation in future studies. By doing so, we aim to establish the foundation for a more comprehensive approach that links urban structures to socio-economic conditions. This approach would allow to reduce the complexity of urban agglomerations by means of remote sensing.

II. Data and Methods

A. Study design

The overall concept of this study is to analyze the information retrieved by remote sensing and geospatial analysis on urban morphology (explanatory variables) with respect to the responses of household surveys undertaken in the city of Kigali (target variables). After filtering out unusable data, these survey results were compared to the geospatial data based on different types of statistical analysis as a result of multiple levels of measurement. The entire process is illustrated as a graphical workflow in Figure 1.

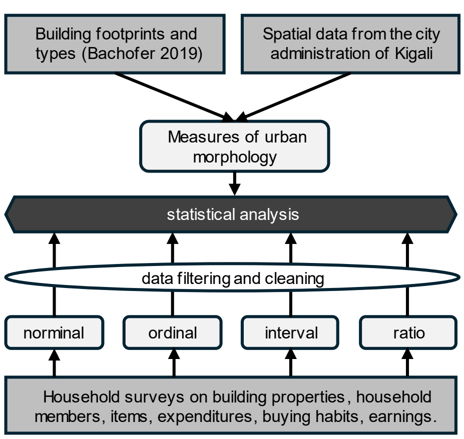

Fig. 1. Data and information flow of this study.

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

B. Household surveys

To gain a detailed picture of the socio-economic situation of the residents of Kigali, household surveys were conducted in September and October 2016 in all parts of the city based on a stratified random sampling design, more precisely in the neighborhoods of Agatare / Rwampara, Biryogo, Gacuriro, Kanserege, Kibagabaga, Kibagabaga, Kimihurura, Kiyovu, Kiyovu, Nyabikenke, Nyamabuye, and Rwezamenyo. Digital questionnaires implemented in the open data kit (ODK, [36]) helped to retrieve 609 geocoded records of households including 254 questions on housing and infrastructure, specific information about the household members, questions on the availability of specific items, expenditures, eating and buying habits, and earnings of the household. The responses were dichotomous (i.e., yes/no), predetermined with single or multiple options, or numeric in nature and treated as the target variables in subsequent statistical analysis. A full list containing all target variables is given in Appendix B.

C. Satellite image and spatial data processing

A stereoscopic image product of the Pléiades mission acquired on 09.08.2015 was used to extract building footprints using an object-based image analysis (OBIA) approach. In a second step, the building footprints were classified into nine predefined classes based on their size, shape, height, and roof color. These classes include *Basic*, *Block*, *Bungalow*, *Villa*, *Mid-rise*, *High-rise*, *Hall*, *Special* and *Construction*, of which the first six are residential and the latter three are uninhabited. A total number of 211,458 buildings were identified at an accuracy of 95.4 % which was determined based on 900 reference buildings which were visited during field campaigns (see next section). Details on the generation of this building dataset are documented and published for open use by Bachofer et al. (2019) [37]. Additional data on civil infrastructures (e.g., sewer network, water kiosks, electric power network) were provided by the administration of the City of Kigali and summarized in Appendix A.

In a second step, morphological variables representing spatial variables representing the morphological nature of the city were retrieved and summarized at the building block level which was provided as outlines by the city administration. These spatial variables included the number of all buildings per block (and shares per building type), the share built-up areas, green spaces and agricultural areas per block (example given in Figure 2), the mean and standard deviation distance between all buildings within a block, the mean slope, the average distances to public infrastructures were computed, as well as the density of buildings and roads and the average building height within a radius of 50 m, 100 m, 250 m and 500 m. Subsequently, additional variables were derived at the building level, using morphological indices describing the shape of the building polygons, such as perimeter, inner edge ratio, sphericity and the number of vertices. The final list of all explanatory variables at the block level and the building levels is shown in Appendix A. In a previous study, these spatial variables on morphology were employed to delineate a map of classified urban structure types in Kigali [25].

In a final step, several filters were applied to ensure the quality and reliability of the statistical results. Initially, variables that exhibited a strong correlation with one another (correlation coefficients > 0.8 or < -0.8) were excluded from the analysis to avoid the potential issues associated with multicollinearity. In addition, variables with low variability were excluded: This concerned variables that predominantly had zero values (> 10%) or no variance (standard deviation = 0). Variables with an insufficient number of entries (N/A > 25%) were also excluded. Morphological and socio-economic variables with a medium correlation (correlation coefficient greater than 0.4 or less than -0.4) were analyzed. In addition, we examined whether data of different groups (e.g. building types) differed significantly from each other to ensure that all relevant differences were considered in the data.

In order to assess the strength and direction of relationships between linear and non-linear associations, both Pearson and Spearman correlation tests were employed. Additionally, a binary decision tree (Tree Classifier) was used to understand how explanatory variables impact the outcomes, providing a clear, interpretable model of the variable interactions.

To confirm the statistical significance of the differences observed, an Analysis of Variance (ANOVA) was performed, followed by a t-Test to validate the significance of the differences identified in the ANOVA. The Chi-Square Test was used to analyze categorical variables and test the independence of variables. For non-normally distributed data, the Kruskal-Wallis Test was employed along with the Post-hoc Bonferroni Test to detect significant differences between multiple groups.

The relationships between spatially derived data at the building and block level retrieved by satellite imagery (independent predictive variables) and the information from the households as punctually collected in surveys (dependent target variables) were then assessed employing the ReliefF score [38] because it takes into account partial correlation between independent variables and other methods of univariate and multivariate statistics, machine learning as described in the following chapter.

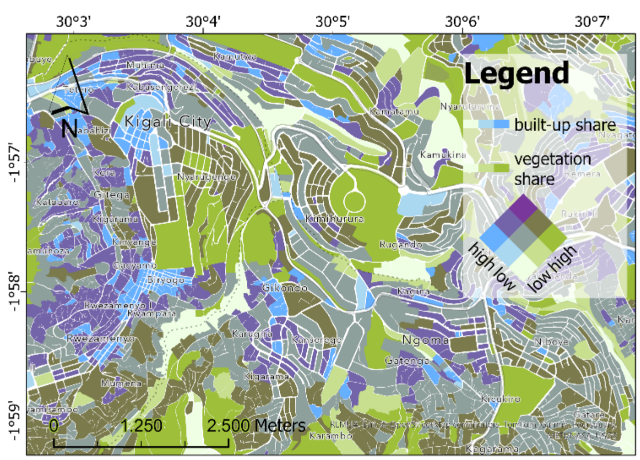

Fig. 2. Share of built-up area (blue) and vegetation (green) per building block in the city of Kigali.

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

III. Results

This chapter presents the relationships identified between morphological variables and socio-economic data obtained from the surveys. It first presents a one-by-one comparison of all variables followed by the main findings of each target category based on different analytical methods.

A. Overall results

All variable pairs which revealed strong positive or negative correlations (r > 0.4 < -0.4) are shown in Figure 3 (Pearson correlation of numerical target features) and Figure 4 (Spearman rank correlation of categorical target features), where the colors indicate the direction of correlation (red for negative and blue for positive correlations) and the dot size and color intensity reflect correlation strength, with larger and more intense dots indicating stronger correlations.

As shown in Figure 3, there are a number of interesting insights to be gained from the correlations. Firstly, among all initial numerical target variables (n=95, Appendix B), only 18 demonstrated at least a moderate positive or negative linear correlation with morphologic variables. This indicates that only a small proportion of approximately 20 % of the surveyed socio-economic data can be directly linked to spatial variables, with the majority falling within category 3 (household items and lifestyle). Moreover, in addition to the initial three variables pertaining to the characteristics of the building, all target variables exhibit comparable patterns of positive and negative correlation. This indicates that they are significantly intercorrelated and present redundancies with respect to spatial variables. Of particular note is the total number of household items (3.6.1), which has the strongest correlation. This is indicative of the fact that this is a significant factor in its own right, and therefore merits assessment. Furthermore, this similarity exists even between categories (household items and expenditures), which lends further weight to this conclusion.

From the perspective of spatial variables, the share and density of high-rise buildings stand out as a distinct form of correlation that is not contained in the other categories. This underscores the peculiarity of high-rise buildings within Kigali as an indicator of the highest urban development, which contrasts with the situation in large parts of the city [39].

To also include categorical target variables, Spearman rank correlations are shown in Figure 4. They all correspond in terms of strength and direction of correlation which can be attributed to the fact that most of them are from category 5 (food and buying habits) and therefore are strongly intercorrelated. Contrary to Figure 3, the density and share of high-rise buildings does not stand out in this analysis, but there is a clear distinction between *basic*-type buildings and *bungalow* or *villa*-type buildings, suggesting that the quality of the building is reflected in opposing consumption or behavior patterns in the data. But again, only a small percentage of all available categorical variables (n=134, Appendix B) is actually reflected by linear correlation. Accordingly, these relationships are investigated at more detail in the following based on further statistics.

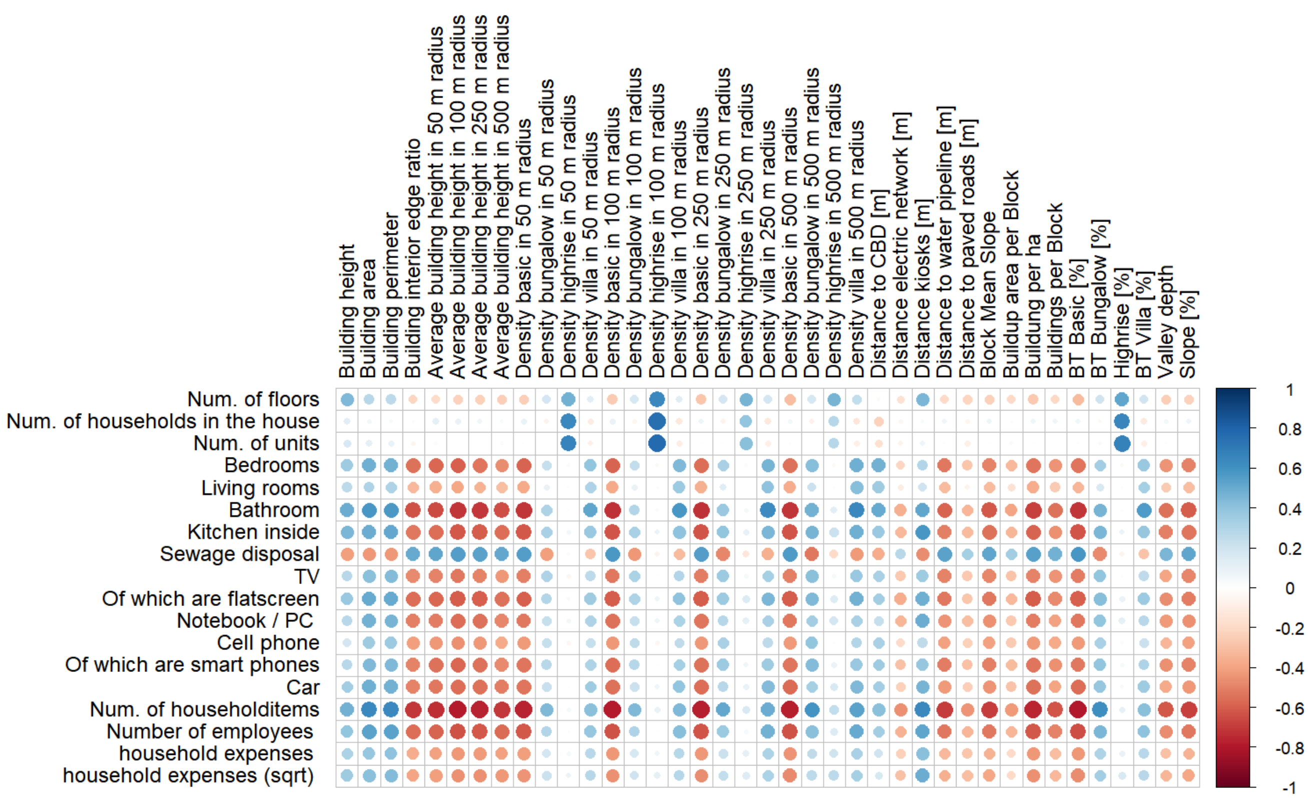

Fig. 3. Correlation matrix of socio-economic and morphological variables (Pearson correlation > 0.4 or <-0.4).

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

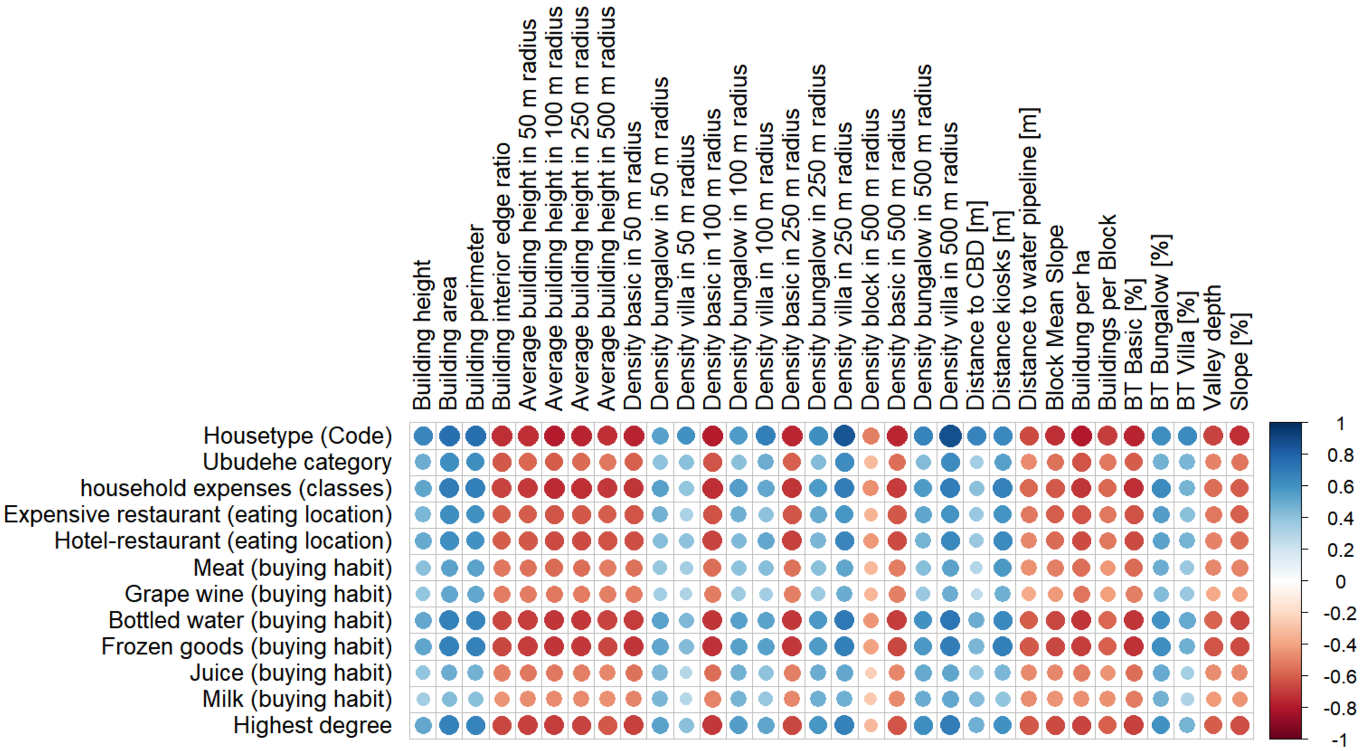

Fig. 4. Correlation matrix of socio-economic and morphological variables (Spearman correlation > 0.4 or <-0.4).

B. Building-related information

As most of the dependent survey variables of this category have binary classified answers (e.g., yes or no) and the independent spatial variables are continuous, we conducted an analysis of variance (ANOVA [40]) to identify variable pairs with very high significance ($p < 0.001$) as proven by a t-test [41] which are presented in the following.

Firstly, there is a significant difference between privately owned and rented homes with respect to their average distance to the central business district (CBD): While rented homes have an average distance to the CBD of 3591 m (± 1702 m), privately-owned homes are located at 4175 m (± 1888 m), as shown in Figure 5A. Although this might be specifically valid for Kigali only, it still shows the socio-economic patterns which can underlie spatial data.

Secondly, the availability of private security service is higher for larger buildings, as shown in Figure 5B: Buildings without security service have an average size of 151 m^2 (± 97 m^2), while those with security service are significantly larger (275 m^2 ± 115 m^2).

Thirdly, Figure 5C shows that the availability of flush toilets in a household is strongly associated with the density of bungalows in its neighborhood. Buildings with flush toilets are, at average, surrounded by 110.9 (± 41.4) bungalows while their presence in neighborhoods without flush toilets is significantly lower (26.5 ± 26.8).

As for the overall sanitation infrastructure, our data shows that households which have toilets are more frequent in blocks with larger shares of building types of higher quality (*bungalow*, *villa*) while they are less frequent in neighborhoods mainly consisting of *basic* or *block* buildings. While this can be clearly attributed to the construction costs and general living conditions of different building types, it was surprising to observe that the households with no sewer system have the smallest average distance to the sewer network (313 m) while those with septic tank (320 m) and an official connection to the sewer system (529 m) are significantly farer away ($p<0.01$). This means that, contrary to our expectations, a smaller distance to the sewer network does not bring direct benefits to the surrounding households. In turn, the relationship is inverted with those buildings closest to the network having the least sanitary facilities. This highlights the potential to improve living conditions in areas close to the network with comparably low efforts, but also indicates that proximity to the sewer network brings potential threats of contamination to drinking water through broken pipes [42].

C. Household members

Besides basic demographics, the Ubudehe category was assessed during the survey, which is a classification system in Rwanda to categorize households based on their income and assets, facilitating targeted social protection and poverty alleviation programs. This system divides households into categories ranging from the most vulnerable and economically disadvantaged (A) to the wealthiest (E), serving as a tool for policy-making and resource allocation in various social services

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

and development initiatives [43]. As shown in Figure 6, these categories are well represented by the building types retrieved by remote sensing, with all buildings of class *basic* belonging to the most disadvantaged class. In turn, buildings of type *bungalow* or *villa* are stronger represented in the categories C and D, respectively.

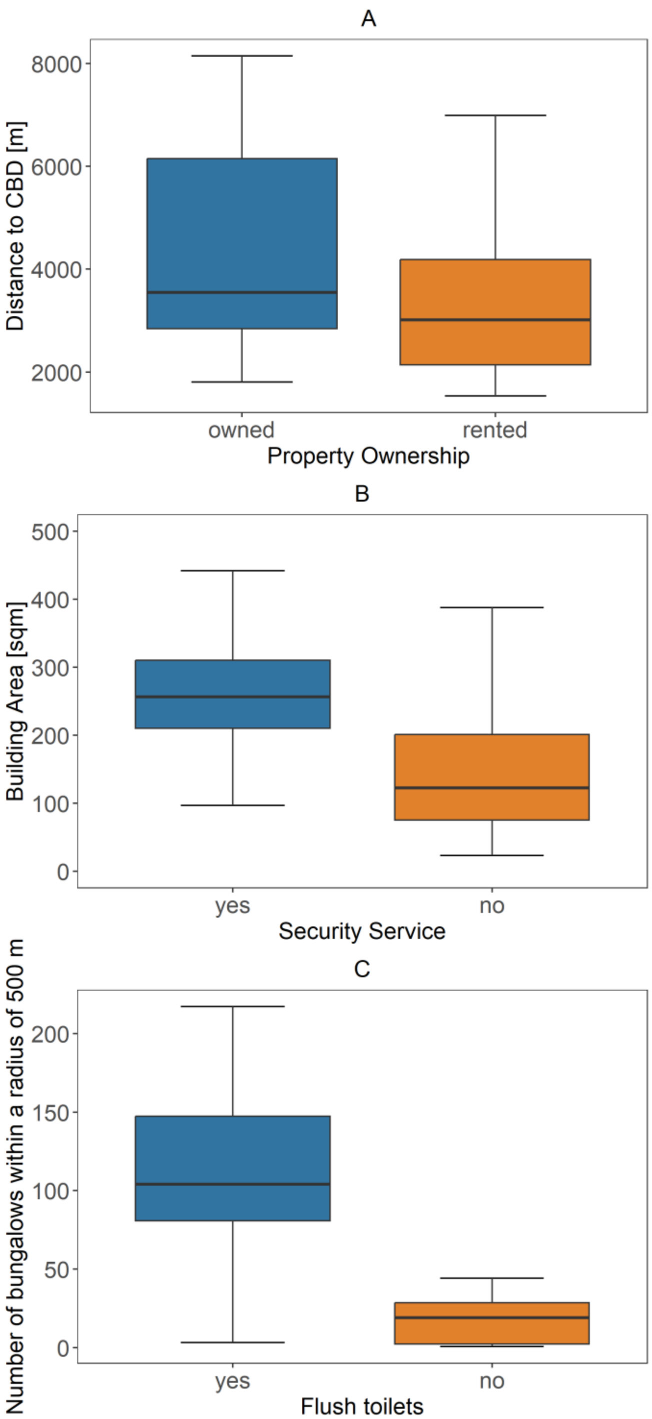

Fig. 5. A: Distance to the central business district by property ownership. B: Building area by Security Service. C: Bungalow density in 500 m radius by flush toilets.

Furthermore, the Ubudehe category allows to statistically analyze social disadvantages at the spatial level: Accordingly, the wealthiest class (D) lives in largely flat areas (average slope of 6.3 %), followed by class C (9.6 %), class B (14.0 %), and class A (18.0 %) confirming highly significant differences ($p<0.001$) regarding the steepness of the neighborhood in these classes which confirms the observations of Kuffer et al. (2017) [44] and many others who identified a larger exposure and vulnerability of the urban poor towards environmental hazards, in this case flooding and landslides in neighborhoods at Kigali's many slopes [45], [46].

Looking at the zoning of the city highlights a spatial pattern regarding the share of children (< 12 years) within the households: At average, it lies around 16.6 % ± (22.1 %) for areas defined as *commercial* according to the official master plan [47] and at 21.1 % (±21.2 %) for the *urban* parts while it is significantly higher for the *rural* parts (32.7 ±18.7 %; $p<0.001$). The opposite is the case for the share of working members of the household which is the lowest for blocks in rural zones (22.6 %), and significantly higher for *urban* (32.3 %) and *commercial* (36.7 %; $p<0.01$) zones.

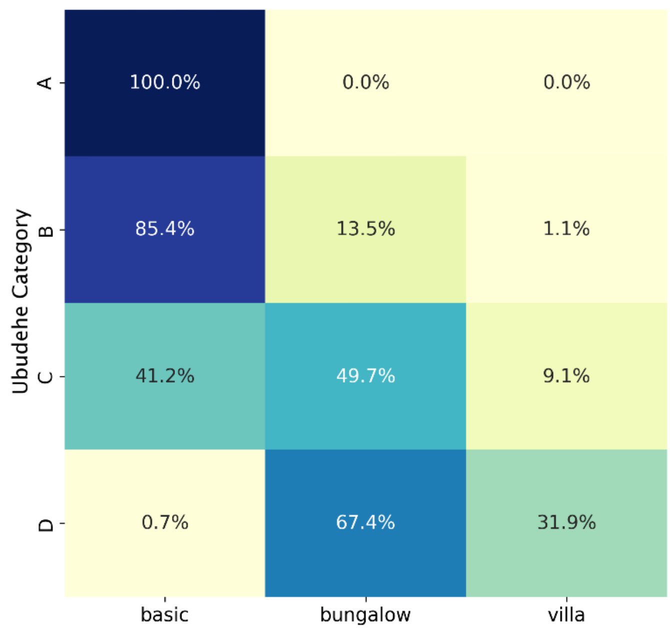

Fig. 6. Contingency map of selected building types (x-axis) and Ubudehe category A-D (y-axis). Note: Uninhabited building types (special and hall) and those of too small survey samples (mid-rise and high-rise) were omitted. Class E was not present among the surveyed households.

Lastly, we delineated the socio-economic class of the surveyed households (as responded by the residents) by a binary tree classifier [48] composed of four levels to an accuracy of 91.4 %: As shown in Figure 7, the root of the tree is split by the percentage of basic building types, indicating that a higher presence of basic structures is strongly associated with lower socio-economic status (100% low). For areas with less than 87.71% BT basic, the distance to markets becomes the next

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

significant predictor, with those closer to markets more likely to belong to higher socio-economic classes (36.9% high). Further splits reveal that proximity to the central business district (CBD) and the average building height in a 100-meter radius are additional key determinants, with closer distances and greater building heights correlating with higher socio-economic status. This hierarchical structure underscores the importance of urban form and accessibility in shaping socio-economic outcomes.

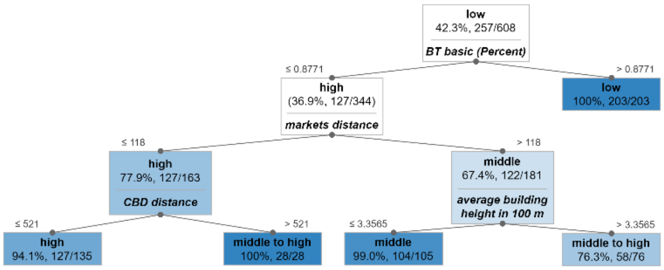

Fig. 7. Tree classifier for the socio-economic class of the households.

D. Household items and lifestyle

The third survey category refers to the items which are available in the interviewed households, among others, these were television, computer or notebook, cell phones, refrigerator, electric stove, microwave, dishwasher, boiler, washing machine, electric fan, air conditioner, bicycle, motorbike, car, bank account and credit card. The number of 'yes' answers to these questions was summarized and ranged between 0 and 24 throughout all households with an average number of 10 owned items. It has a strong negative correlation of r=-0.792 with the share of buildings of the type *basic*. Accordingly, the more *basic* buildings within a block, the lower is the average number of household items. As shown in Figure 8, this relationship is characterized by two clusters: Blocks with a low to medium share (0-50%) of *basic* buildings and blocks which predominantly consist of *basic* buildings (>75%), but the negative correlation was proven strongly significant (p<0.001) by a Chi square test.

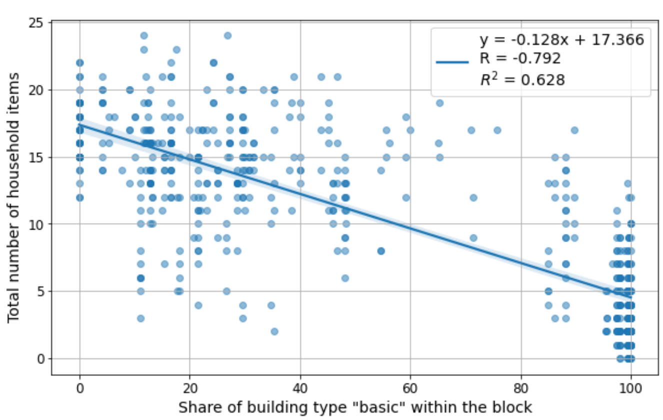

Fig. 8. Scatter plot between the share of buildings of type "basic" within the block (x-axis) and the total number of household items (y-axis).

A similar pattern was observed for the number of household items and the building density as expressed by the number of buildings within a radius of 100 m (r=-0.78), and also the share of buildings of the categories *bungalow* (r=0.58) and *villa* (r=0.42), with respective positive correlations.

E. Household expenses

A Kruskal-Wallis test [49] indicated that the total calculated household expenses differ with respect to the building types identified from the satellite image (X^2=292.17, p>0.05). A post-hoc Bonferroni test [50] proves the significance of these differences which are shown in Figure 9.

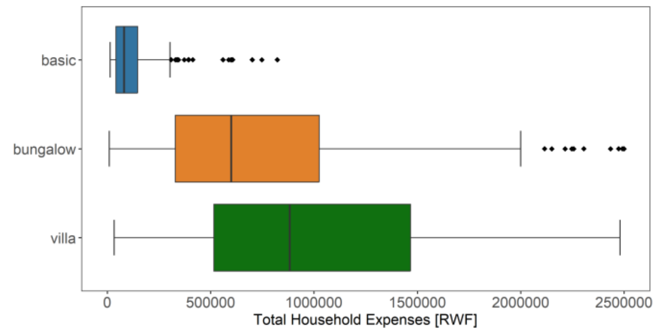

Fig. 9. Total household expenses in Rwanda Franc (x axis) by selected building types (y axis).

All three building types show an increase in total household expenses with increasing number of household items. A Pearson correlation test (r=0.705, p<0.05, n=224) indicates a very strong positive correlation for the building type *basic* and total household expenses (Figure 10, blue markers). The scatter plot confirms the general expectation that household expenditure will increase with a higher-quality lifestyle for this type of building. However, for buildings of type *bungalow* there is only a medium correlation (r=0.458, p<0.1, n=254, orange markers) and no significant correlation can be determined for the building type *villa* (green markers).

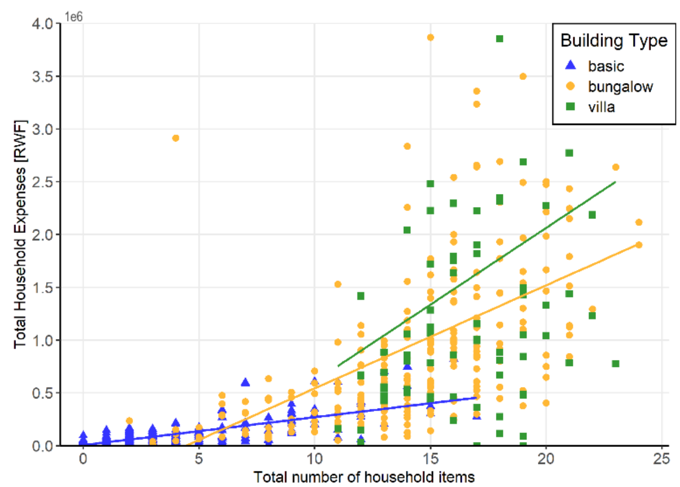

Fig. 10. Total household expenses in Rwanda Franc (y axis) vs. number of household items (x axis).

We furthermore identified a significant link between the average building height within a radius of 100 m per block and the type of health insurance: Persons with private insurance live in blocks with significantly taller houses (3.9 m) as compared to persons with public (3.1 m) or community (2.8 m; $p<0.001$).

F. Food and buying habits

Our data shows that the frequency of buying and eating meat is closely connected to the size of buildings. More precisely, there are highly significant differences ($p<0.001$) in building sizes for people which consume meat 'very rarely' (120 m² ± 78 m²), '1-2 times a week' (207 m² ± 104 m²), and 'almost daily' (272 m² ± 131 m²). The group of households with reported never eating meat has the largest range of building sizes (159 m² ± 103 m²) which indicates that a vegetarian diet is rather a personal choice and observed throughout all socio-economic groups, in contrast to the other three responses which are clearly attributed with increasing purchasing power, which is in line with the findings of Weatherspoon et al. (2019) [51].

A similar pattern is found for the place to shop for vegetables which is strongly dependent on the share of buildings classified as *basic* within a building block. As shown in Table 1, respondents who reported to shop for vegetables in supermarkets and mini markets predominantly live in blocks with low shares of *basic* buildings (both around 25 % to 30 %), while their share is significantly higher for respondents which prefer to retrieve their vegetables from street markets (nearly 80 %, all with $p<0.001$). This can be partially explained by the higher prices of vegetables in supermarkets and mini markets, which are small but official places of trade while the street market is widely considered informal and of lesser quality and therefore cheaper [52]. Only the 'local markets' which are a larger form of organized shopping do not show significant differences regarding the share of basic buildings ($p=0.165$).

We did not identify other patterns regarding buying habits which correlate with morphologic or spatial information in our data.

Table 1. Share of *basic* buildings within a block by points of purchase for vegetables as responded in the household survey.

	Yes	No
Supermarket	26.1 % (± 23.3 %)	78.6 % (± 32.9 %)
Street market	78.1 % (± 32.3 %)	36.4 % (± 33.7 %)
Mini market	29.3 % (± 28.4 %)	65.1 % (± 34.9 %)
Local market	51.9 % (± 38.2 %)	57.1 % (± 40.4 %)

G. Earnings of the household

Similar to the findings on household sanitation, the building type composition within the blocks strongly reflects the educational level of the residents. The share of *bungalow* buildings lies at around 50 % for households where the highest degree is bachelor or master, but rapidly decreases for households with secondary school graduation (22 %) and takes up nearly 0% for households where the highest degree is primary school or with no school graduation at all ($p<0.001$).

As for the total job income of the households (*inc*), which ranged between several spatial variables turned out to have an impact, notably in a complementary way: The highest feature importance score was achieved by the average density of buildings within a radius of 100 m per block (*d100*), the average building height within a radius of 100 m per block (*h100*), and the average distance to the electric network (*dE*). Although these show correlations of different strengths and directions, (-0.785, +0.395, -0.248), they allow to form a multiple regression model (1) with a coefficient of determination of $R^2=0.464$ and a root mean square error (RMSE) of 648,000 RWF (around 600 USD) which means that these three variables together can be used to explain 46% of the total observed variance of the total job income of the surveyed households. Its application to all residential blocks in Kigali is shown in Figure 11 to give an example on how such information can be used to spatially predict socio-economic variables for the entire city based on morphologic variables. The map illustrates how the estimated income follows clear and explainable patterns: Highest values are achieved in the central business district (CBD) of Nyarugenge in the western part of the city, in the upper-class areas of Vision City in Kinyiya, as well as in the residential areas of Kigarama in the south. In turn, lowest incomes spatially match the presence of informal, unplanned or basic buildings, such as in Nyakabanda, Rwezamenyo, Kimisagara and Gitega in the West, but also in southern Gisozi and the densely built areas around the airport in the southwest.

$$inc = -720*dE + 379{,}737*d_{100} - 22{,}115*h_{100} + 293{,}755 \quad (1)$$

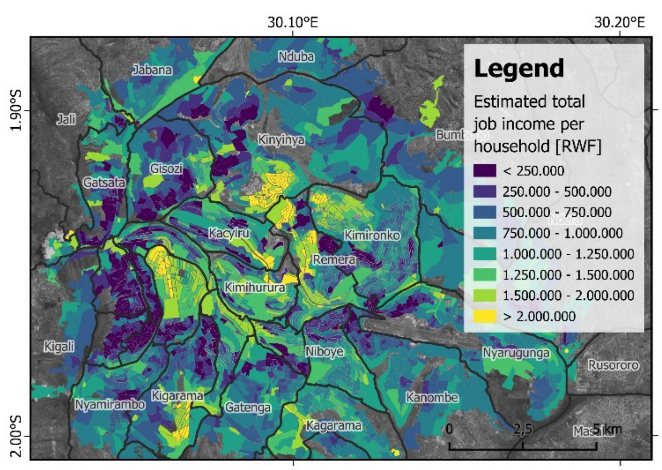

Fig. 11. Estimated total job income of households in Kigali based on multiple regression ($R^2=0.464$).

IV. Discussion and Outlook

A. Major findings

This study examined how well socio-economic information retrieved from household surveys can be attributed by spatial data. These findings can be grouped into three main categories which will be discussed in the following:

(a) Physical properties: Firstly, buildings with a private security service had significantly larger sizes as compared to those without security service. The same applies for meat

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

consumption which is the highest in large buildings and the lowest in small buildings (III.E). Also, our data showed that private health insurance is more frequent in houses of greater height (III.E).

(b) Building type: At the building level, the building type was closely related to the availability of flush toilets (III.B.), the Ubudehe category of its residents (III.C.), or the total amount of household expenses (III.E.). Moreover, the number of household items and expenses were stronger correlated for buildings of the category basic than for villa-type buildings. This indicates that more of the expenses are used for household equipment in basic buildings as compared to higher-class buildings where a higher share of money is spent on leisure or other luxury [53].

It was furthermore shown that the composition of building types within a block can be an indicator for socio-economic information, for example for the availability of sanitation infrastructure which is more available in buildings within blocks of high amounts of bungalow or villa buildings and less frequent in neighborhoods mainly consisting of basic or block buildings (III.B). The share of basic buildings in a block also strongly negatively correlated with the number of household items (III.D.) and showed clear patterns regarding the place to shop for vegetables with the highest share for street markets as compared to other shopping facilities (III.F.). In turn, the share of bungalow buildings of a block was the highest in households of with high education (III.G.)

(c) Spatial patterns: Distances were used to identify patterns, as demonstrated by the relationship between the availability of building sewer systems and their distance to the sewer network, which was a negative correlation for the case of Kigali (III.B.). In turn, buildings closer to the CBD were predominantly rented while the share of privately-owned homes significantly increased with higher distances from the CBD (III.B.).

The density of buildings is negatively correlated with the number of household items (III.D.), which clearly reflects the less-equipped households in poorer or informal parts of the city where buildings are closely built together.

Also, administrative zoning was reflected in the survey with higher shares of children and lower shares of working household members in rural parts in comparison to areas classified as commercial or urban. (III.C.).

Even topography plays a role in Kigali, with more wealthy people living in flat areas while residents of lower Ubudehe classes live in buildings on significantly steeper slopes (III.C.).

Lastly, there are multi-variate connections between the household surveys and the geospatial data as proven by the decision tree to predict the socio-economic class of the household (III.C.) which was largely determined by the share of basic buildings in the block, the distance to markets, and the building height. Also, the estimated total job income can be expressed as a function of building density, building height, and distance to the electric network (III.G.), again combining information of all three mentioned categories.

B. Critical evaluation and outlook

Incorporating remote sensing data into the analysis of cities allows to enlarge the understanding of urban morphology and its socioeconomic implications. Our findings not only align with previous studies which utilize geospatial data to better understand cities and the spatial distribution of their residents [54], but also but also advances the ideas of those which analyze aspects of urban demographics, poverty or wealth, and lifestyle or consumption patterns which were proven to be reflected by building shapes [26] and measures of distance and access [22]. However, to proof these relationships, extensive data on households and their citizens are required which are often outdated, not existing, or only available at coarse spatial units, for instance at the level of city districts [55]. Unlike most other studies which focus on the validation of their modeled results by stratified samples [56] or cross-validation based on training data [57], we were ably to directly link spatial data to a large amount of empirically collected target variables and thus achieve insights of unrivalled quality. Even though the household surveys in this study were already conducted in 2016, this does not detract from the quality of the results, or the insights gained from them, as our study does not focus on temporal change, but on the general relationships between socio-economic and urban structure types, and these have been proven throughout all groups of target variables.

It is important to note that despite the interesting findings presented in this study, many target variables collected during the survey did not yield any significant results. As shown in chapter III.A, only for a small proportion of the target variables a linear correlation with the explanatory variables could be established. This can partially be explained by the points raised in the following, but also underlines the need for multi-variate statistics which allow to analyze the contribution of multiple variables to one target feature, such as suggested by the classification and regression tree to predict the socio-economic status in chapter III.C or the multiple regression to model the total job income of households in chapter III.G. Further aspects to be considered are to follow in the next paragraphs.

Firstly, some questions in the interview addressed aspects of the respondents' lives that are not connected to the location of their residence or their neighborhood. This indicates that a considerable number of demographic characteristics, behavioral patterns and personal attributes cannot be adequately represented by geospatial proxies, at least not within the confines of our dataset. This also includes the aspect of ownership addressed in the introduction which is not visible from space and may indicate wrong connections between the morphology of a single house and the wealth of its residents.

Another aspect to mention is that this study has mainly used the information content of building blocks as provided by the city administration. However, these strongly differ regarding their shape and size and therefore underlie the modifiable areal unit problem [58] which can be avoided by strictly using moving averages or a consistent spatial form of aggregation (e.g., regular hexagon grids) to calculate zonal statistics for the independent variables.

From the perspective of data collection, it turned out that the spatial sampling of the household surveys did not equally include the different building types and morphological structures of Kigali. Accordingly, some of the statistics could not be generated due to small sample sizes. For instance, the composition of building types and their shares within the blocks was mainly dominated by *basic* buildings while *villas*, *mid-rise* and *high-rise* residential buildings were underrepresented in the surveyed areas. Future studies should therefore place more emphasis on strategic sampling to cover the entire magnitude of urban structures, for instance based on a pre-classification of urban structure types [25] as a base for a stratified allocation of surveyed households. Lastly, as a general risk of survey research, the reliability of the retrieved answers is not fully known and could be biased because of confusion, misconception, or misrepresentation [59] which then lead to incorrect data and less significant statistics.

Lastly, all statistics should be interpreted with care and knowledge of the analyzed city. It is not advisable to purely rely on numbers which can in some cases be simply the result of chance. Any conclusions should therefore be drawn with caution and logical thinking. Accordingly, is important to note that statistical correlation does not necessarily imply causality. In the context of this study, this implies that patterns identified between Earth Observation data and socio-economic variables may correspond without a direct linkage. This could be due to the influence of a third, unknown variable. Furthermore, mutual feedback effects between urban structure and human behavior exist, making it challenging to determine which aspect was the primary driver of the other [60].

For future studies, it is the responsibility of interdisciplinary teams consisting of geographers and socio-scientists to identify the most suitable predictor variables for the city of interest and how they can be aggregated or combined to highlight and exploit patterns [61].

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3466298

APPENDIX

A. Morphologic variables

Table 2. Spatial data serving as explanatory variables in this study

	Category	**Variable**	**Source**	**Year**	**Unit**	**Calculation**
Building	Building Type	Building Type	Bachofer et al. (2019)	2015		
	Dimension and Geometry	Height	Bachofer et al. (2019)	2015	m	
		Area (A)			m²	
		Perimeter (P)			m	
		Vertices			count	
	Form and Structure	Interior edge ratio	Bachofer et al. (2019)	2015	ratio	P/A
		Sphericity			ratio	P/(2*sqrt(A*pi)))
	Density in 50 / 100 / 250 / 500 m radius	building height	Bachofer et al. (2019)	2015	m	Kernel density
		building type: basic, block, bungalow, hall, high-rise, midrise, special, villa			count	
		roads	MinInfra	2015	m	
	Distance to infrastructure	CBD	Own	2023	m	
		markets	OSM	2023	m	
		electric network, sewer net, water kiosks, water pipeline, paved roads, unpaved roads	MinInfra	2015	m	
Block	Area and Topography	Block Mean Slope	CoK	2015	percent	calculated on a DEM with 10x10m spatial resolution
		Block Area	MinInfra	2015	m²	
	Landuse	Total veg. area	CoK	2015	m²	
		Total veg. area (non-agricultural)			m²	
		Total agricultural area			m²	
		Share of vegetation			%	(Total veg. area (non-agricultural)/Block Area) * 100
		Share of agricultural area			%	(Total agricultural area/Block Area) * 100
	Building Development	Buildings per Block	Bachofer et al. (2019)	2015	count	
		Building per ha	Bachofer et al. (2019), MinInfra	2015	buildings/ha	Total num of buildings / (Block Area / 10000)
		Buildup area per Block			m²	
		Block BCR			%	(Buildup area per Block / Block Area) * 100
		Block NN_index			count	
		Building Type Distribution: basic, block, bungalow, hall, high-rise, midrise, special, villa	Bachofer et al. (2019)	2015	%	

Bachofer et al (2019 [37]
MinInfra: Ministry of Infrastructure Rwanda, KG 1 Roundabout, Kigali, Rwanda
CoK: City of Kigali, One Stop Centre KN 3 Ave, Kigali, Rwanda

B. Household survey variables

Table 3. Socio-economic variables from the household surveys serving as target variables in this study

Category	Variable	Type of answer / unit
1. Building-related information	Number of floors, households, units, different uses and different rooms	count
	Overall size of the plot, Overall size of the flat, Footprint of the house	m^2
	Access to the structure	type (e.g., road, pathway)
	Property of the house or flat	type (owned, rented)
	Kind of toilet	type (e.g., flush, pit latrine)
	Sewage disposal	type (e.g., septic, sewer)
	Flush toilets, Private water tab	boolean (yes/no)
	security service	boolean (yes/no)
2. Household members	Number of family members	count
	Age of family members	category (< 12y, 12-17y, 18-60y, >60y)
	Number of working family members	count
	Ubudehe category	category (A-E)
	Total financial resources	RWF
3. Household items and lifestyle	TV, flat screen, satellite dish, pay TV, notebook, internet connection, cell phone, smartphone, household appliances (refrigerator, electric stove, microwave, ...), air conditioning, bicycle, motorcycle, car, bank account, credit card	boolean (yes/no) count
	Total items of the household	count
4. Household expenditures	total expenditures, food and beverages, mobility, leisure, rent, electricity, water supply, waste disposal, charcoal, firewood, gas, leisure (gym, tennis, golf, clubs), journeys (family visits, holiday travel), education (pre-school, public school, private or international school, college / university), medical care, health insurances, further insurances (car, house), rent, support for relatives, employees, clothing	RWF per month
5. Food and buying habits	frequency of street food, local restaurants, fast food, resort restaurant	category (almost daily, 1-2 times a week, very rarly, never)
	frequency of eating meat, vegetables, ubugali, wine, beer, bottled water, frozen goods, cereal, milk, juice	category (almost daily, 1-2 times a week, very rarly, never)
	self-produced food (meat, vegetables, crops)	boolean (yes/no)
	Type of buying locations	type (street market, local market, mini market, supermarket, farms, online)
	Estimated distance to different buying locations	m
6. Earnings of the household	Highest degree	type (no school, primary school, secondary school, bachelor/master)
	Profession	type
	Number of current jobs	count
	Characteristics of main job	type (none, part time, fully)
	total income of all jobs, rental income, pension payment, support from governmental institutions, support from relatives, earnings from agriculture	RWF/month
7. Assessment by the interviewer	Personal estimation of the socio-economic class	class (low - high)
	House/ flat in relation to the surrounding	type (worse, typicall, better)
	Quality/reliability of the answers	boolean (realistic/ unrealistic)

Acknowledgment

The authors would like to thank the administration of the City of Kigali for supporting the work with permissions and the provision of geospatial data on blocks and public infrastructures. We furthermore thank all collectors involved in the socio-economic household surveys organized by IUWA (Institute for Eco-Industrial Analyses) Heidelberg.

Parts of this work were funded by the German Ministry for Education and research (BMBF) within the project Rapid Planning under grant identifier 01LG1301K.

We acknowledge the support by the Open Access Publishing Fund of the University of Tübingen.

References

[1] M. A. Kuddus, E. Tynan, and E. McBryde, “Urbanization: a problem for the rich and the poor?,” *Public Health Rev.*, vol. 41, p. 1, Jan. 2020, doi: 10.1186/s40985-019-0116-0.

[2] L. P. Singh, C. J. van Westen, P. C. Ray, and P. Pasquali, “Accuracy assessment of InSAR derived input maps for landslide susceptibility analysis: a case study from the Swiss Alps,” *Landslides*, vol. 2, no. 3, pp. 221–228, Jan. 2005.

[3] M. Araos, S. E. Austin, L. Berrang-Ford, and J. D. Ford, “Public Health Adaptation to Climate Change in Large Cities,” *Int. J. Health Serv.*, vol. 46, pp. 53–78, 2016.

[4] S. Galea and D. Vlahov, “Urban health: evidence, challenges, and directions,” *Annu. Rev. Public Health*, vol. 26, pp. 341–65, Jan. 2005, doi: 10.1146/annurev.publhealth.26.021304.144708.

[5] R. Satgé, *Urban Planning in the Global South: Conflicting Rationalities in Contested Urban Space*. Cham: Palgrave Macmillan US, 2018.

[6] K. Mouratidis, “Urban planning and quality of life: A review of pathways linking the built environment to subjective well-being,” *Cities*, vol. 115, p. 103229, Jan. 2021, doi: 10.1016/j.cities.2021.103229.

[7] J. D. Kressner and L. A. Garrow, “Using Third-Party Data for Travel Demand Modeling: Comparison of targeted Marketing, Census, and household travel survey Data,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2442, no. 1, pp. 8–19, Jan. 2014, doi: 10.3141/2442-02.

[8] S. Zhou and F. Teng, “Estimation of urban residential electricity demand in China using household survey data,” *Energy Policy*, vol. 61, pp. 394–402, Oct. 2013, doi: 10.1016/j.enpol.2013.06.092.

[9] G. M. Oliveira, D. G. Vidal, and M. P. Ferraz, “Urban Lifestyles and Consumption Patterns,” in *Sustainable Cities and Communities*, W. Leal Filho, A. Marisa Azul, L. Brandli, P. Gökçin Özuyar, and T. Wall, Eds., Cham: Springer International Publishing, 2020, pp. 851–860. doi: 10.1007/978-3-319-95717-3_54.

[10] R. F. Grais, A. M. Rose, and J.-P. Guthmann, “Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys,” *Emerg. Themes Epidemiol.*, vol. 4, no. 1, p. 8, Dec. 2007, doi: 10.1186/1742-7622-4-8.

[11] K. Beegle, J. De Weerdt, J. Friedman, and J. Gibson, “Methods of household consumption measurement through surveys: Experimental results from Tanzania,” *J. Dev. Econ.*, vol. 98, no. 1, pp. 3–18, May 2012, doi: 10.1016/j.jdeveco.2011.11.001.

[12] A. M. Salama, F. Wiedmann, and H. G. Ibrahim, “Lifestyle trends and housing typologies in emerging multicultural cities,” *J. Archit. Urban.*, vol. 41, no. 4, pp. 316–327, Dec. 2017, doi: 10.3846/20297955.2017.1415773.

[13] P. P. Kalbar, M. Birkved, M. Hauschild, S. Kabins, and S. E. Nygaard, “Environmental impact of urban consumption patterns: Drivers and focus points,” *Resour. Conserv. Recycl.*, vol. 137, pp. 260–269, Oct. 2018, doi: 10.1016/j.resconrec.2018.06.019.

[14] J. Mattson, “Relationships between density, transit, and household expenditures in small urban areas,” *Transp. Res. Interdiscip. Perspect.*, vol. 8, p. 100260, Nov. 2020, doi: 10.1016/j.trip.2020.100260.

[15] M. M. Rathore, A. Paul, W.-H. Hong, H. Seo, I. Awan, and S. Saeed, “Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data,” *Sustain. Cities Soc.*, vol. 40, pp. 600–610, Jan. 2018, doi: 10.1016/j.scs.2017.12.022.

[16] H. Kaur and P. Garg, “Urban sustainability assessment tools: A review,” *J. Clean. Prod.*, vol. 210, pp. 146–158, Jan. 2019, doi: 10.1016/j.jclepro.2018.11.009.

[17] M. Luca, G. M. Campedelli, S. Centellegher, M. Tizzoni, and B. Lepri, “Crime, inequality and public health: a survey of emerging trends in urban data science,” *Front. Big Data*, vol. 6, 1124526 vols., Jan. 2023, doi: 10.3389/fdata.2023.1124526.

[18] T. Wellmann *et al.*, “Remote sensing in urban planning: Contributions towards ecologically sound policies?,” *Landsc. Urban Plan.*, vol. 204, p. 103921, Jan. 2020, doi: 10.1016/j.landurbplan.2020.103921.

[19] J. E. Patino and J. C. Duque, “A review of regional science applications of satellite remote sensing in urban settings,” *Comput. Environ. Urban Syst.*, vol. 37, pp. 1–17, Jan. 2013, doi: 10.1016/j.compenvurbsys.2012.06.003.

[20] X. Xing, B. Yu, C. Kang, B. Huang, J. Gong, and Y. Liu, “The Synergy Between Remote Sensing and Social Sensing in Urban Studies: Review and perspectives,” *IEEE Geosci. Remote Sens. Mag.*, vol. 12, no. 1, pp. 108–137, Mar. 2024, doi: 10.1109/MGRS.2023.3343968.

[21] J. Yin *et al.*, “Integrating remote sensing and geospatial big data for urban land use mapping: A review,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 103, p. 102514, Dec. 2021, doi: 10.1016/j.jag.2021.102514.

[22] L. Lin, L. Di, C. Zhang, L. Guo, and Y. Di, “Remote Sensing of Urban Poverty and Gentrification,” *Remote Sens.*, vol. 13, no. 20, p. 4022, Oct. 2021, doi: 10.3390/rs13204022.

[23] M. Wurm and H. Taubenböck, “Detecting social groups from space – Assessment of remote sensing-based

mapped morphological slums using income data," *Remote Sens. Lett.*, vol. 9, no. 1, pp. 41–50, Jan. 2018, doi: 10.1080/2150704X.2017.1384586.

[24] E. Wentz *et al.*, "Supporting Global Environmental Change Research: A Review of Trends and Knowledge Gaps in Urban Remote Sensing," *Remote Sens.*, vol. 6, no. 5, pp. 3879–3905, Apr. 2014, doi: 10.3390/rs6053879.

[25] A. Braun, G. Warth, F. Bachofer, M. Schultz, and V. Hochschild, "Mapping Urban Structure Types Based on Remote Sensing Data—A Universal and Adaptable Framework for Spatial Analyses of Cities," *Land*, vol. 12, no. 10, p. 1885, Oct. 2023, doi: 10.3390/land12101885.

[26] J. Wang, S. Georganos, M. Kuffer, A. Abascal, and S. Vanhuysse, "On the knowledge gain of urban morphology from space," *Comput. Environ. Urban Syst.*, vol. 95, p. 101831, Jul. 2022, doi: 10.1016/j.compenvurbsys.2022.101831.

[27] A. Wendnagel-Beck *et al.*, "Characterizing Physical and Social Compositions of Cities to Inform Climate Adaptation: Case Studies in Germany," *Urban Plan.*, vol. 6, no. 4, pp. 321–337, Dec. 2021, doi: 10.17645/up.v6i4.4515.

[28] I. D. Stewart and T. R. Oke, "Local Climate Zones for Urban Temperature Studies," *Bull. Am. Meteorol. Soc.*, vol. 93, no. 12, pp. 1879–1900, Jan. 2012, doi: 10.1175/BAMS-D-11-00019.1.

[29] H. Zhou and H. Gao, "The impact of urban morphology on urban transportation mode: A case study of Tokyo," *Case Stud. Transp. Policy*, vol. 8, no. 1, pp. 197–205, Jan. 2020, doi: 10.1016/j.cstp.2018.07.005.

[30] N. K. Downes, "Linking Dynamic Urban Settlement Patterns to Environmental Infrastructure Needs: The Case of Da Nang City, Vietnam," in *ICSCEA 2021*, J. N. Reddy, C. M. Wang, V. H. Luong, and A. T. Le, Eds., in Lecture Notes in Civil Engineering. , Singapore: Springer Nature Singapore, 2022, pp. 133–139. doi: 10.1007/978-981-19-3303-5_10.

[31] B. Jiang, K. Shen, W. C. Sullivan, Y. Yang, X. Liu, and Y. Lu, "A natural experiment reveals impacts of built environment on suicide rate: Developing an environmental theory of suicide," *Sci. Total Environ.*, vol. 776, p. 145750, Jan. 2021, doi: 10.1016/j.scitotenv.2021.145750.

[32] H. Taubenböck *et al.*, "Integrating Remote Sensing and Social Science - The correlation of urban morphology with socioeconomic parameters," presented at the Joint Urban Remote Sensing Event, Shanghai: IEEE, 2009, pp. 1–7. [Online]. Available: https://api.semanticscholar.org/CorpusID:59060505

[33] G. Nduwayezu, V. Manirakiza, L. Mugabe, and J. M. Malonza, "Urban Growth and Land Use/Land Cover Changes in the Post-Genocide Period, Kigali, Rwanda," *Environ. Urban. ASIA*, vol. 12, no. 1_suppl, pp. S127–S146, Mar. 2021, doi: 10.1177/0975425321997971.

[34] G. Warth, A. Braun, O. Assmann, K. Fleckenstein, and V. Hochschild, "Prediction of Socio-Economic Indicators for Urban Planning Using VHR Satellite Imagery and Spatial Analysis," *Remote Sens.*, vol. 12, no. 11, p. 1730, Jan. 2020, doi: 10.3390/rs12111730.

[35] G. Baffoe, J. Malonza, V. Manirakiza, and L. Mugabe, "Understanding the Concept of Neighbourhood in Kigali City, Rwanda," *Sustainability*, vol. 12, no. 4, p. 1555, Jan. 2020, doi: 10.3390/su12041555.

[36] C. Hartung, A. Lerer, Y. Anokwa, C. Tseng, W. Brunette, and G. Borriello, "Open data kit: tools to build information services for developing regions," in *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, London United Kingdom: ACM, Dec. 2010, pp. 1–12. doi: 10.1145/2369220.2369236.

[37] F. Bachofer *et al.*, "Building Stock and Building Typology of Kigali, Rwanda," *Data*, vol. 4, no. 3, p. 105, Jan. 2019, doi: 10.3390/data4030105.

[38] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Machine learning: Proceedings of the fourteenth international conference*, 1997, pp. 296–304.

[39] V. Manirakiza and A. Ansoms, "Modernizing Kigali: The Struggle for Space in the Rwandan Urban Context," in *Losing your Land*, 1st ed., A. Ansoms and T. Hilhorst, Eds., Boydell and Brewer Limited, 2014, pp. 186–203. doi: 10.1017/9781782043416.010.

[40] L. Ståhle and S. Wold, "Analysis of variance (ANOVA)," *Chemom. Intell. Lab. Syst.*, vol. 6, no. 4, pp. 259–272, 1989, doi: 10.1016/0169-7439(89)80095-4.

[41] Student, "The Probable Error of a Mean," *Biometrika*, vol. 6, no. 1, p. 1, Mar. 1908, doi: 10.2307/2331554.

[42] I. Ross and J. Pinford, *Kigali Urban Sanitation Study: Final report - Volume 1: synthesis report*. Oxford UK, 2017.

[43] A. Niringiye and C. Ayebale, "Impact evaluation of the Ubudehe Programme in Rwanda: an examination of the sustainability of the Ubudehe Programme," *J. Sustain. Dev. Afr.*, vol. 14, no. 3, Jan. 2012.

[44] M. Kuffer, F. Orina, R. Sliuzas, and H. Taubenbock, "Spatial patterns of slums: Comparing African and Asian cities," in *2017 Joint Urban Remote Sensing Event*, IEEE, 2017, pp. 1–4. doi: 10.1109/JURSE.2017.7924587.

[45] L. Nahayo *et al.*, "Estimating landslides vulnerability in Rwanda using analytic hierarchy process and geographic information system," *Integr. Environ. Assess. Manag.*, vol. 15, no. 3, pp. 364–373, 2019.

[46] V. Kabalisa and M. F. Kagambira, "Disaster Risk Management of Higher Risk Areas on Floods and Landslides in Rwanda," 2021.

[47] M. J. Nyiransabimana, I. Rwabudandi, W. T. Vries, J. P. Bizimana, and G. G. Benineza, "Impact of Kigali City master plan implementation on living conditions of urban dwellers: case of Nyarugenge District in Rwanda," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 389, no. 1, p. 012018, Jan. 2019, doi: 10.1088/1755-1315/389/1/012018.

[48] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Jan. 1996, doi: 10.1007/BF00058655.

[49] A. Vargha and H. D. Delaney, “The Kruskal-Wallis Test and Stochastic Homogeneity,” *J. Educ. Behav. Stat.*, vol. 23, no. 2, pp. 170–192, Jun. 1998, doi: 10.3102/10769986023002170.

[50] R. J. Simes, “An improved Bonferroni procedure for multiple tests of significance,” *Biometrika*, vol. 73, no. 3, pp. 751–754, 1986, doi: 10.1093/biomet/73.3.751.

[51] D. D. Weatherspoon, S. Miller, J. C. Ngabitsinze, L. J. Weatherspoon, and J. F. Oehmke, “Stunting, food security, markets and food policy in Rwanda,” *BMC Public Health*, vol. 19, no. 1, p. 882, Dec. 2019, doi: 10.1186/s12889-019-7208-0.

[52] J. B. Ndikubwimana, “Analysis of the Factors Causing the Persistence of Street Vending in the City of Kigali, Rwanda,” *East Afr. J. Sci. Technol.*, vol. 10, no. 2, 2020, doi: doi:10.62103/eajst.v10i10.133.

[53] E. Nzeyimana, M. M. Sakwa, and G. Namusonge, “Urban poor families access to assets and implication on household’s food security, the situation of low-income households in City of Kigali, Rwanda,” *Int. J. Sci. Res. Publ. IJSRP*, vol. 10, no. 2, p. p9802, Feb. 2020, doi: 10.29322/IJSRP.10.02.2020.p9802.

[54] X. Liu and M. Herold, “Population Estimation and Interpolation Using Remote Sensing,” in *Urban Remote sensing*, Boca Raton, FL: CRC Press/Taylor & Francis, 2018, pp. 269–290.

[55] P. Merodio Gómez *et al.*, “Earth Observations and Statistics: Unlocking Sociodemographic Knowledge through the Power of Satellite Images,” *Sustainability*, vol. 13, no. 22, p. 12640, Nov. 2021, doi: 10.3390/su132212640.

[56] A. Abascal *et al.*, “AI perceives like a local: predicting citizen deprivation perception using satellite imagery,” *Npj Urban Sustain.*, vol. 4, no. 1, pp. 1–14, Mar. 2024, doi: 10.1038/s42949-024-00156-x.

[57] S. Georganos *et al.*, “Modelling the Wealth Index of Demographic and Health Surveys within Cities Using Very High-Resolution Remotely Sensed Information,” *Remote Sens.*, vol. 11, no. 21, p. 2543, Oct. 2019, doi: 10.3390/rs11212543.

[58] A. S. Fotheringham and D. W. S. Wong, “The Modifiable Areal Unit Problem in Multivariate Statistical Analysis,” *Environ. Plan. Econ. Space*, vol. 23, no. 7, pp. 1025–1044, Jul. 1991, doi: 10.1068/a231025.

[59] J. Hewitt, “Ethical Components of Researcher—Researched Relationships in Qualitative Interviewing,” *Qual. Health Res.*, vol. 17, no. 8, pp. 1149–1159, Oct. 2007, doi: 10.1177/1049732307308305.

[60] J. A. Priess, N. Schwarz, and S. Lautenbach, “Feedbacks in socio-environmental land systems,” in *International Congress on Environmental Modelling and Software*, Ottawa, Canada, Jul. 2010. [Online]. Available: https://scholarsarchive.byu.edu/iemssconference/2010/all/279

[61] G. Vigderhous, “The Level of Measurement and ‘Permissible’ Statistical Analysis in Social Research,” *Pac. Sociol. Rev.*, vol. 20, no. 1, pp. 61–72, Jan. 1977, doi: 10.2307/1388904.

Andreas Braun (*21.07.1986) works on spatial analysis and geoinformatics and received his B.Sc. (2010) and M.Sc. (2012) in the field of geography at the Eberhard Karls University in Tübingen, Germany. He received his PhD in geography in the year 2019 for his thesis on the contribution of spaceborne radar imagery to humanitarian action at the University of Tübingen.

He engaged in various international research projects on urban remote sensing and landscape monitoring and was a former member of the Christian Doppler Laboratory for geospatial and EO-based humanitarian technologies (GEOHUM) at the Paris Lodron University of Salzburg. Since 2022, he is a senior lecturer and researcher at the institute of geography within the department of geosciences at the faculty of science of the University of Tübingen.

Dr. Braun was awarded the best dissertation (Förderpreis Geoinformatik) by the Runder Tisch GIS e.V. at the Technical University of München in 2020 and was granted the water monitoring price by the Dr. Erich Ritter-Stiftung in cooperation with the Water Science Allegiance e.V. in 2023 for his project on the impact of sand dams in semi-arid ecosystems.