Estimating quality of life dimensions from urban spatial pattern metrics

Marta Sapena [a,b], Michael Wurm [b], Hannes Taubenböck [b,c], Devis Tuia [d,e,*], Luis A. Ruiz [a]

[a] *Geo-Environmental Cartography and Remote Sensing Group, Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València, Camí de Vera, s/n, 46022 Valencia, Spain*
[b] *German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, 82234 Wessling, Germany*
[c] *Institute for Geography and Geology, Julius-Maximilians-Universität Würzburg, 97074 Würzburg, Germany*
[d] *Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, 6700, PB, Wageningen, the Netherlands*
[e] *Environmental Computational Science and Earth Observation Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1950 Sion, Switzerland*

A B S T R A C T

The spatial structure of urban areas plays a major role in the daily life of dwellers. The current policy framework to ensure the quality of life of inhabitants leaving no one behind, leads decision-makers to seek better-informed choices for the sustainable planning of urban areas. Thus, a better understanding between the spatial structure of cities and their socio-economic level is of crucial relevance. Accordingly, the purpose of this paper is to quantify this two-way relationship. Therefore, we measured spatial patterns of 31 cities in North Rhine-Westphalia, Germany. We rely on spatial pattern metrics derived from a Local Climate Zone classification obtained by fusing remote sensing and open GIS data with a machine learning approach. Based upon the data, we quantified the relationship between spatial pattern metrics and socio-economic variables related to 'education', 'health', 'living conditions', 'labor', and 'transport' by means of multiple linear regression models, explaining the variability of the socio-economic variables from 43% up to 82%. Additionally, we grouped cities according to their level of 'quality of life' using the socio-economic variables, and found that the spatial pattern of low-dense built-up types was different among socio-economic groups. The proposed methodology described in this paper is transferable to other datasets, levels, and regions. This is of great potential, due to the growing availability of open statistical and satellite data and derived products. Moreover, we discuss the limitations and needed considerations when conducting such studies.

1. Introduction

How we organize space in urban areas has a decisive influence on how we live and what effects this has on our closest environment: what kind of mobility we choose, how large our ecological footprint is, how close we are to utilities, or what access we have to jobs or leisure facilities. These are just a few of the many exemplary factors that influence the quality of life and the sustainability by its spatial design.

This is especially relevant in cities, as the world population is becoming urban. The share of people living in urban areas has been growing in the last decades and this trend is expected to continue (United Nations, 2018). Heretofore, population growth has been accompanied by a significant increase of the urban layout, triggering environmental and socio-economic consequences (e.g. Haase, Kabisch, & Haase, 2013; Ribeiro-Barranco, Batista e Silva, Marin-Herrera, &

Lavalle, 2014; Taubenböck et al., 2012). The *quality of life* and *sustainable development* of urban and peri-urban areas depend on the successful management of their growth. Both are common goals in cities around the world. They are described in multiple dimensions: 'quality of life' is a broad concept assessed on various factors ranging from living conditions and employment to experience of life. It is usually represented by a multiple set of indicators such as income, deprivation rate, education attainment, employment rate, life expectancy, air quality, etc. (Eurostat, 2017; OECD, 2017). Also 'sustainable development' is addressed by the United Nations in their 2030 Agenda for Sustainable Development. It collects seventeen Sustainable Development Goals (SDGs), which aim at ending poverty by means of promoting economic growth, addressing social needs, while protecting the environment and fighting climate change.

Urban form is constituted by spatial and socio-economic processes

developed over time and space (Abrantes et al., 2019; Salat, 2011). It is accepted in scientific literature to wield a powerful influence on shaping societies (Oliveira, 2016; Salat, 2011; Tonkiss, 2013). Urban form is a key element for understanding urban systems as it drives where people live and work and how the interaction is spatially structured (e.g. Grimm, Cook, Hale, & Iwaniec, 2015; Taubenböck, 2019). However, it is not self-evident to establish a universal link between the urban spatial structure, here considered as the organization of urban areas in terms of the distribution of physical structures and human activities (Krehl & Siedentop, 2019), and quality of life. Accordingly, in this study we want to explicitly investigate this relation between urban structural features and socio-economic parameters, and whether quality of life can be interpreted based on spatial and statistical methods.

Urban areas with similar physical appearance tend to feature similar social, economic, and environmental characteristics (Patino & Duque, 2013; Taubenböck et al., 2009; Wurm & Taubenböck, 2018). Consequently, several authors have described qualitatively and quantitatively these influences. Concerning social factors, many relevant concerns such as crime, public safety, gentrification, health, and poverty, have been linked to diversity and configuration of land uses, road network patterns, or remote sensing derived variables (e.g. Hankey & Marshall, 2017; Jacobs, 1961; Lehrer & Wieditz, 2009; Patino, Duque, Pardo-Pascual, & Ruiz, 2014; Sandborn & Engstrom, 2016; Wurm et al., 2019). In terms of economic issues, wealth indicators were positively related to the diversity of land uses, and productivity and innovation were influenced by density, centricity, and urban size (e.g. Tapiador, Avelar, Tavares-Corrêa, & Zah, 2011; UN-Habitat, 2015). For the environmental dimension, the identification of land cover and urban structural types allowed for instance determining urban heat islands or green area facilities, which contributed to climate change studies (Bechtel et al., 2019; Stewart & Oke, 2012), while pollution, energy use, and transport means have also been related to different properties of urban form, such as density, diversity or centrality of land use (e.g. Anderson, Kanaroglou, & Miller, 1996; Hankey & Marshall, 2017). However, there are few studies that measure this widely agreed linkage between the spatial structure of cities and their socio-economic status in a quantitative manner. These studies mostly rely on earth observation data to extract the physical information, such as buildings, roads, land-use/land-cover (LULC) and their spatial distribution, or structure and texture features. This approach has been applied so far to model neighborhood deprivation (Venerandi, Quattrone, & Capra, 2018), poverty (Duque, Patino, Ruiz, & Pardo-Pascual, 2015; Jean et al., 2016; Wurm & Taubenböck, 2018), income and property value (Taubenböck et al., 2009), and demographic, living conditions, labor and transport factors (Sapena, Ruiz, & Goerlich, 2016). These examples present previous attempts to identify links between urban spatial structure and socio-economic parameters.

Insofar, the investigation of these relations has been possible due to the increasing accessibility of open databases and earth observation products. On the one hand, satellite images allow for increasing capabilities to provide high-resolution geoinformation. In this context, LULC data have been an important source of information for urban studies; however, it lacks three-dimensional information of urban structures, considered a fundamental aspect in such studies (Wentz et al., 2018). Therefore, the characterization of cities into urban structural types and land cover, with Local Climate Zones (LCZ) (Stewart & Oke, 2012) as one concept, has great potential in its relation with socio-economic functions (Bechtel et al., 2015). LCZ have additional inherent information on the physical composition of cities compared to other LULC legends by their density, building types, heights, greenness and their land cover that are worth to explore. Besides, it is a conceptually consistent, generic, and culturally-neutral description and thus a replicable classification system. On the other hand, global, national and local institutes provide more and more statistical data for different dates and spatial levels. Notwithstanding all the urban theories relating these two components, and the growing availability of both, spatial and socio-

economic databases, studies aiming to quantify the relations between the spatial structure of urbanized areas and the quality of life of inhabitants or the well-known SDGs are still scarce. Methods based on the quantification of spatial patterns by means of spatial metrics and the clustering of urban areas based on their socio-economic performance (e. g.: Abrantes et al., 2019; Sapena, Ruiz, & Goerlich, 2016; Schwarz, 2010) have shown to be suitable for the combined analysis of spatial and socio-economic variables in urban areas as well as for their relation.

In this framework, the general objective of this work is to understand better the relationship between the spatial structure of cities and the socio-economic level of city dwellers. For this reason, we explore the value of the LCZ, as urban structural types, in relation to quality of life indicators at the city level. First, we quantify the relationships between socio-economic variables and the spatial distribution of LCZ. Then, we group cities according to their similar levels of quality of life and describe their spatial structure.

2. Data

2.1. Study area

We selected North Rhine-Westphalia (NRW) as a study case for its socio-economic relevance in Europe, reinforced by the availability of statistical data. The historical and political background of many cities located in this Federal State is similar, which diminishes external influences in our analysis. We base our study on a sample of 31 cities in NRW as consistent spatial and socio-economic databases are available there. The location and identification of cities is presented below in Fig. 2.

Regarding the Federal State of NRW, it is the most populous of the sixteen German states, accounting for 21.7% of the total population in Germany (Eurostat, 2019). The Ruhr industrial region, in NRW, is a competitive industrial region of Germany. NRW is an economic centre in Europe, with a regional GDP of € 672 billion in 2016 (21.4% of the German GDP). However, the per capita level is slightly below the national level. Nowadays, the economy of NRW is based on small and medium-sized enterprises, hosting more than 20% of companies in Germany, and providing work to near 80% of the active population (European Commission, 2019).

2.2. Socio-economic variables

For the socio-economic analysis we used the City statistics database (https://ec.europa.eu/eurostat/web/cities/data/database). This database was originally created with the purpose to provide information that supports more evidence-based decisions in planning and managing tasks (Eurostat, 2016). The City statistics project covers several aspects of quality of life – i.e., demography, housing, health, economic activity, labour market, income disparity, educational qualifications, environment, climate, travel patterns, tourism, and cultural infrastructure - for cities and their commuting zones in Europe (Eurostat, 2018). At the city level, it contains 171 variables and 62 indicators for more than one thousand cities that have an urban core of at least 50,000 inhabitants. The data are available at different dates from 1990 onwards. In this study the city level is the basic spatial unit. At this level a rich source of data for comparative studies in Europe is provided.

For the purpose of this study, we selected a set of socio-economic variables and indicators for 31 cities in NRW for the year 2009 (Table 1), to coincide with the date of the satellite images used for LCZ classification. When data from 2009 were not available, the previous or subsequent year was used instead. Subject to the availability of data, we selected indicators of five dimensions of 'quality of life' covered in the database – education, health, living condition, labor and transport. We linked the dimensions to the SDGs policy commitments, as was previously done by the OECD (2017) to evidence the global efforts that are being made to reduce inequalities in the socio-economic level of citizens

Table 1

Description of the selected socio-economic variables (dependent variables in the models from Table 5) representing five dimensions of quality of life and their link to the Sustainable Development Goals (SDGs).

Dimension	Name	Description	SDGs
Education	education	The proportion of population (aged 25–64) with lower secondary as the highest level of education	SDG 4 (education)
Health	health	Crude death rate per 1000 inhabitants	SDG 3 (health)
Living conditions	housing	Average price for buying an apartment in euros	SDG 1 and 11 (poverty and sustainable cities)
	income	Median disposable annual household income in euros	
	affordability	Ratio reflecting the ability of a city to pay for housing. Housing price compared to income.	
Labor	employment	Number of employments per 1000 inhabitants (work place-based)	SDG 8 (decent work and economy)
Transport	transport	The share of journeys to work by car or motor cycle (%)	SDG 9 and 11 (Infrastructure and sustainable cities)
	commuting	People commuting out of the city per 1000 residents	

(Table 1).

2.3. Earth observation and ancillary data

For classification of the physical structures describing the cities' spatial structure we rely on remotely sensed and geospatial data extracted from three data sources:

- High-resolution remote sensing imaging: a Rapid-Eye mosaic for the year 2009 was constructed for the whole area. This satellite provides images at 6.5 m resolution (orthorectified and resampled to 5 m) with five spectral bands (red, green, blue, near infrared and red edge).
- 3D model: A normalized digital surface model (nDSM) was derived from 135 individual Cartosat-1 stereo images (collected between 2009 and 2013) and processed according to Wurm, d'Angelo, Reinartz, & Taubenböck (2014) to retrieve above ground heights.
- GIS layers from OpenStreetMap: the amenities and road layers from the open repository of geospatial data was used (downloaded in 2014, openstreetmap.org).

3. Methods

3.1. Patterns describing the spatial structure of cities

For the derivation of the spatial patterns describing the spatial structure of cities, we applied the LCZ framework that allows characterizing the morphologic appearance of cities in a conceptually consistent manner. It comprises several urban structural and land covers types with uniform surface cover, structure, material and use (Stewart & Oke, 2012). Out of the 17 original LCZ classes, 12 were present in the region (Fig. 1). The spatial pattern describes the distribution of phenomena across space, e.g., concentration, dispersion, clustered patterns, etc. (Getis & Paelinck, 2004). In particular, we refer to the arrangement of urban structural types and land covers within cities.

For the classification of LCZ, we followed the protocol presented in Tuia, Moser, Wurm, & Taubenböck (2017). We modelled LCZs on a grid composed of cells of size 200 × 200 m. In total, 89 variables were

Fig. 1. Summary of the LCZ classes present in NRW (from Stewart & Oke, 2012). LCZs 2 to 10 are built-up classes, LCZs A to G land cover types.

extracted for each cell (Table 2) to train the classifier. A ground truth of 2658 cells was defined by photointerpretation, where the cognitive perception of an interpreter was used to define the predominant LCZ.

The classifier was based on random forests, a method building several decision trees with heavy randomization of features (Breiman, 2001). The initial result was then further improved by making the model aware of two spatial relationships between cells using a Markovian Random Field formulation (see Tuia, Moser, Wurm, & Taubenböck (2017)): (1) By predicting with higher probability co-occurrence of neighboring LCZ that attract or repel each other spatially; (2) By favoring a map respecting a rank-size distribution of urban settlements, according to Zipf's law.

For testing the relationships between socio-economic and spatial structure of cities, we extracted a large set of spatial pattern metrics, henceforth referred to as spatial metrics, related to attributes such as density, aggregation, shape, etc., using the spatial module of the software tool IndiFrag (Sapena & Ruiz, 2015). This tool computes spatio-temporal metrics that quantify spatial patterns and their changes from thematic maps. We used the LCZ classification as a base to characterize the spatial structure of cities. The level of analysis to extract the spatial metrics was the city level that corresponds to the level in which the socio-economic variables are provided. Therefore, we computed all spatial metrics included in IndiFrag, obtaining one set of metrics per LCZ class for the spatial level of the city, and another set of metrics for the city, regardless of the LCZ classes. Then, we standardized the values of

Table 2

Description of the geo-spatial variables for each cell from remote sensing and GIS data.

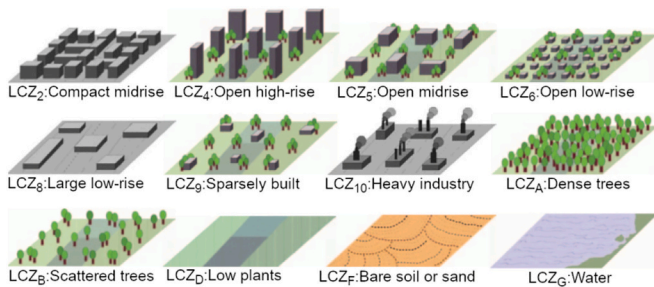| Source	Type	Description	No. of variables
Remote sensing	Bands	Mean and standard deviation of the pixel values. Data were atmospherically corrected and haze removed.	10
	Texture	Co- and occurrence features (local standard deviation, average, homogeneity, entropy, dissimilarity, correlation, contrast and angular moments).	60
	3D	Mean and standard deviation for both nDSM and buildings only. The number of buildings was also added as a feature.	5
	Land cover	Percentage of the area occupied by buildings, trees, grassland and impervious surfaces. The land cover is issued from an object-based classification on a Rapid-Eye mosaic (Montanges, Moser, Taubenböck, Wurm, & Tuia, 2015).	4
GIS	Roads	Total line length for *highway*, *primary*, *secondary*, *tertiary roads*, *residential streets*, *streams* and *rivers*, smoothed with a Gaussian kernel at the cell level.	7
	POIs	Counts for POIs *cafes*, *restaurants* and *rail stations*, smoothed with a Gaussian kernel at the cell level.	3

the metrics as the mean divided by the standard deviation, in order to obtain comparable regression coefficients and avoid influence of measurement units.

When working with spatial metrics and a great diversity of structural types, it is common to find redundant information since metrics depend on similar variables, cover similar spatial patterns, or can be complementary to each other (Reis, Silva, & Pinho, 2015; Schwarz, 2010). Therefore, feature selection is a fundamental step (Genuer, Poggi, & Tuleau-Malot, 2015), as the high correlation of features may introduce noise in the process and affect the accuracy of results. In particular, regression models need the independence of predictors to minimize the multicollinearity, which makes the model unstable. We followed three consecutive approaches for the objective selection of metrics. First, we discarded the non-discriminative spatial metrics, those with a coefficient of variation lower than 5%. Second, we conducted a correlation analysis to identify redundancies in the spatial information. We omitted those metrics showing strong correlations to others (Pearson correlation coefficient > 0.8), keeping one metric per group of correlated metrics. Third, we applied a recent method proposed by Genuer, Poggi, & Tuleau-Malot (2015), called Variable Selection Using Random Forests (VSURF) that selects a specific subset of metrics adapted to each socioeconomic variable. This is based on measuring the relevance of every metric in relation to each socio-economic variable using a random forest regression. We kept one subset of metrics for each socio-economic variable (the different subsets of selected metrics are reported in Table 4).

3.2. Estimating socio-economic and spatial pattern links

A model was obtained for each socio-economic variable from Table 1 applying stepwise multiple linear regression analysis, using the subset of spatial pattern metrics previously selected as independent variables. We applied a min-max normalization transforming the socio-economic variables in a range between zero and one as follows: $z_i = (x_i - min(x)) / (max(x) - min(x))$, where $x = (x_1, ..., x_n)$, x_i is the ith original value and z_i is the normalized value. For education, health, housing, affordability, transport, and commuting the normalization was inversed and thus, higher values mean better conditions for all variables. The number of independent variables was restricted to a maximum of four spatial metrics to avoid overfitting, considering the limited number of observations (cities) in our dataset. The residuals were tested for normality using the Shapiro-Wilk test (Shapiro & Wilk, 1965), and for statistical significance by requiring *p-values* to be lower than 0.05. Leave-one-out cross-validation was employed to evaluate the models. We estimated the root mean squared error (RMSE) and the coefficient of determination (R^2) to summarize the proportion of variance explained by the model, and thus the goodness-of-fit.

To verify whether the level of 'quality of life' in cities is reflected in their urban spatial structure we conducted a two-step analysis: (1) we used the *k*-Means clustering method to group cities according to their values of socio-economic variables, representing variables from five dimensions of quality of life considered in our study (Table 1), out of nine (Eurostat, 2017). Using the Elbow method (Ketchen & Shook, 1996) we found an appropriate number of groups. Consequently, we created and described four clusters that group cities based on their socioeconomic similarities. Moreover, we represented the 'quality of life' for each city and group using star plots, as well as the average of the region, which facilitates the interpretation of the different groups of cities; (2) we applied a stepwise discriminant analysis for selecting a relevant and reduced set of spatial metrics - based on their significance - that better separates the cities into these groups. Afterwards, the values of the spatial metrics, and thus the spatial structure of cities, were interpreted for each group.

4. Results

4.1. Spatial analysis of cities

In Table 3 we present the composition of the training/test sets and the per-class and overall accuracies obtained for the LCZ classification. Per-class accuracy is given by the user's and producer's accuracy, where the number of correct classified cells in a class divided by the total number of cells classified as that class is the user's accuracy (commission error), and if divided by total number of cells of a class in the ground truth is the producer's accuracy (omission error) (Congalton, 1991). The region was split in two parts (North and South) and training was performed on the Northern region, while testing was performed on the Southern to avoid positive biases related to spatial co-location of cells. We obtained an overall accuracy of 83%, which is slightly lower than in Tuia, Moser, Wurm, & Taubenböck (2017), most probably due to the larger amount of testing samples used in this study. With the exception of the LCZ$_2$ "Compact midrise" class and the LCZ$_5$ "Open midrise" all the other classes are classified with more than 70% accuracy. Moreover, the average accuracies of 76.5% and 82.7% also show that the errors are not systematic on the small classes. In Fig. 2 we illustrate the LCZ classification for the 31 sample cities in NRW. The detailed example of the city of Münster reveals how the structural variety of the built and natural landscape is captured by the LCZ classification.

Concerning the spatial metrics, in total 22 global metrics per city and 24 class metrics per LCZ and city were calculated. Since our classification map had 12 LCZ classes, 310 metrics were obtained at city level. After the correlation analysis, a reduced subset of 72 uncorrelated metrics remained. This subset was the input in VSURF for each socioeconomic variable, obtaining one group of metrics per socio-economic variable with sizes between 19 and 31 metrics (Table 4). This output was part of the input in the following section as explained below.
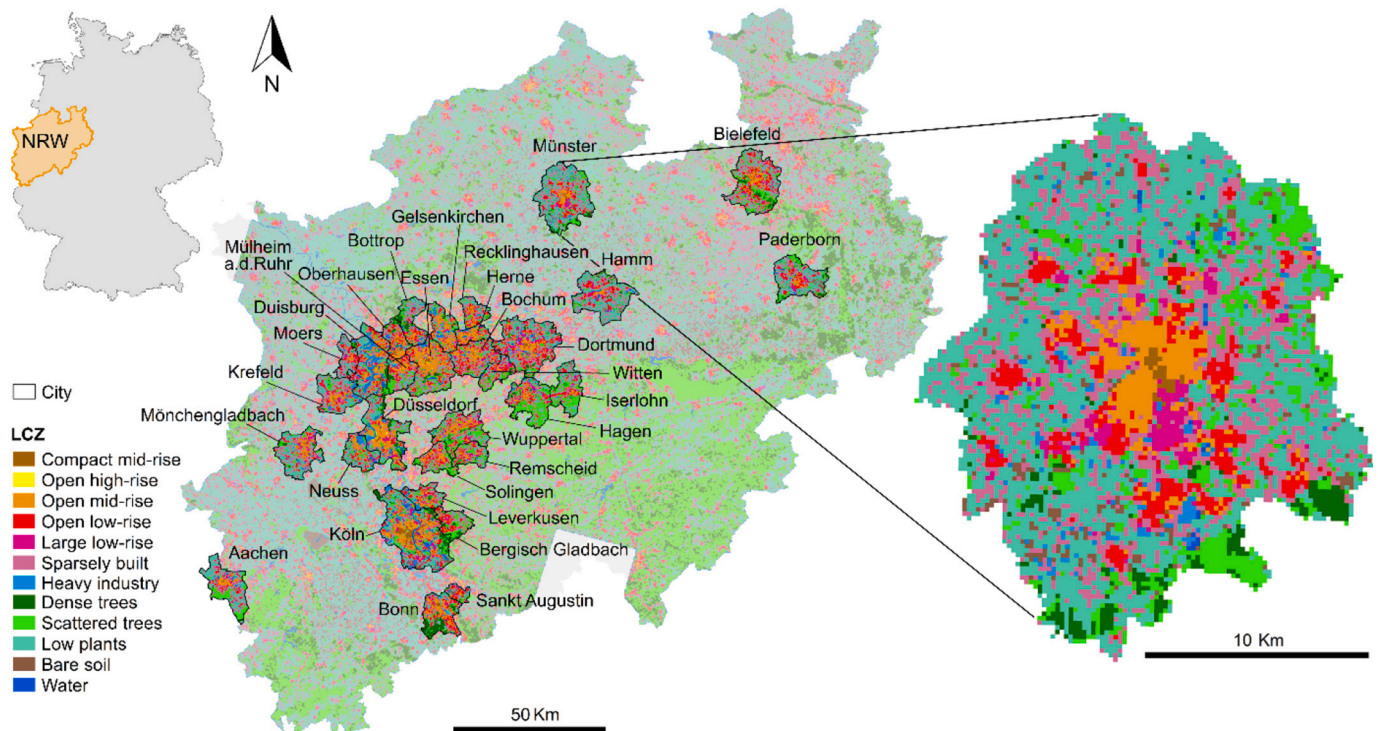
4.2. Models of socio-economic variables

In Table 5 we show the results of the eight fitted models, one for each socio-economic variable. The numerical goodness-of-fit indicators show that the models are statistically significant (*p-value* < 0.05) and explain from 43% to 82% of the variability (R^2) of the socio-economic variables by means of the spatial structure of cities, with RMSEs ranging from 0.10 to 0.17. The values of the model of *housing* were normalized with a logarithmic transformation to obtain a normal distribution of the residuals and improve the adjustment (Table 5). The spatial metrics included in each model and their associated coefficients allow interpreting which and to what extent spatial patterns explain the modelled variable (Table 5). As they are all standardized to z-scores prior to the analysis, their direct contribution is represented by the regression coefficients.

The relationships we found between the spatial structure of the cities in this region and the socio-economic variables are as follows: cities with a better level of *education* have less open, and thus more continuous, built-up (PU), however, the distribution of open midrise is more scattered (DEMp$_5$), dense tree patches are furthest away from the city center (DimR$_A$), and there is a higher density of open high-rise buildings (DC$_4$). In terms of *health*, the model relates a lower death rate in cities with a fragmented and distant distribution of sparsely built (DEM$_9$) and scattered tree (DEM$_B$) patches. Conversely, larger and less fragmented areas of heavy industry (IS$_{10}$) are usually present in cities with higher levels of death rates. On the one hand, a compact shape of open midrise (C$_5$), scattered from city centers towards the suburban areas (DEP$_5$) with a compact midrise core (TM$_2$) is related to lower prices of *housing*. On the other hand, *income* is higher in cities with bigger extensions of open low-rise (TEM$_6$), clustered dense trees (IS$_A$), and contiguous areas of sparsely built with very few open areas (P$_9$). Regarding the ability to pay for housing based on income, the model is similar to housing model (TM$_2$ and C$_5$), except that the affordability is inversely proportional to the

Table 3

Numerical results of the LCZ classification and number of samples used for train/test steps. User's and producer's accuracy and global statistics.

	Code	LCZ	No. samples (train/test)	User's accuracy	Producer's accuracy
Land cover	LCZ A	Dense trees	186 / 114	76.99%	76.32%
	LCZ B	Scattered trees	118 / 107	74.77%	74.77%
	LCZ D	Low plants	169 / 131	81.41%	96.95%
	LCZ F	Bare soil or sand	108 / 192	97.33%	94.79%
	LCZ G	Water	194 / 106	98.85%	91.13%
Built-up	LCZ 2	Compact midrise	82 / 39	45.59%	79.49%
	LCZ 4	Open high-rise	13 / 26	100%	3.85%
	LCZ 5	Open midrise	124 / 72	67.95%	73.61%
	LCZ 6	Open low-rise	152 / 48	88.10%	77.08%
	LCZ 8	Large low-rise	100 / 101	91.01%	80.20%
	LCZ 9	Sparsely built	153 / 55	81.67%	89.10%
	LCZ 10	Heavy industry	148 / 192	88.62%	90.83%
		Average accuracy	–	82.69%	76.51%
		Overall accuracy	–	83.08%	
		kappa	–	0.81	

Fig. 2. Location of NRW in Germany (left). Result of the classification for the NRW region highlighting the classification in the analyzed cities (middle). Detailed example of the classification for Münster (right).

fragmentation of open low-rise (IS_6). Therefore, the ability to pay is lower in bigger cities with a compact midrise core surrounded by fragmented clusters of open low-rise structures.

In relation to the economic aspects (*employment*), open low-rise located towards the periphery of the city ($DimR_6$) and a fragmented and distant distribution of sparsely built (DEM_9) are characteristic of cities with higher employment rate, moreover, LCZ patches are bigger, which means more continuous LCZ classes and, in general, less isolated small patches (TEM). Concerning *transport*, fragmented cities (TEM) in small continuous clusters (GC), with higher proportion of sparsely built areas (DC_9) and a lower number of large low-rise areas (DO_8) commute more by car or motor cycle. Meanwhile, citizens living in cities associated with more compact areas of sparsely built structural type (C_9) commute more out of the city (*commuting*). The way in which open low-rise is allocated affects commuting patterns. The higher the number of compact clusters (LPF_6 and IS_6) the more the commuting proportion.

4.3. Categorization of cities

Fig. 3 shows the clustering of cities according to their socio-economic similarities using the normalized values of six socio-economic variables (we excluded the share of journeys to work by car or motor cycle since statistics were available only from 29 cities and the ability to pay for housing since housing and income were included instead). The individual plots show the location of cities by means of the bi-dimensional spaces defined by each pair of socio-economic variables. Cities are identified by means of a number and color. The map depicts how cities and groups are distributed in the region. It can be seen that the first group (green) is easily identified by means of income and commuting levels (income and commuting plots in Fig. 3). While education discerns the second group (blue, education plots in Fig. 3), the identification of the third group (orange) is not straightforward. However, the fourth (red) can be identified by means of death rate, price of buying an apartment and employment rates (health, housing, and employment

Table 4

Description of the selected spatial metrics (independent variables in the models from Table 5). The significant relations between metrics and socio-economic variables according to VSURF, are shown in the intersection of the rows and columns. The characters show whether metrics computed at the class level (with their LCZ short codes, see Table 3), at the city level (X), or lack of relation (−). Formulas can be consulted in Sapena and Ruiz (2015, 2019). Patch means a group of contiguous pixels with the same LCZ class.

Spatial metric	Description	Education	Health	Housing	Income	Affordab.	Employ.	Transport	Commut.
Compactness (C)	Measures the shape complexity of the LCZ class.	−	F	A,F,5,9	−	A,F,5,9	A,F,5,9	A,5,9	A,F,5,9
Class density (DC)	The ratio between the LCZ class area and city area.	2,4,10	D, F,5,9	A,F, G,2,5,6,9	6	A,F, G,2,5,6	9	6,9	6
Density-diversity (DD)	Sum of the ratios between the areas of every LCZ class and the largest LCZ class. Informs about the richness and heterogeneity.	−	−	X	−	−	X	X	X
Patch nearest neighbor (DEM)	Mean Euclidean distance between the nearest patches from the same LCZ class (km).	B,5	A, B,8,9	10	9	10	5,9	5,9	9
Pixel nearest neighbor (DEMp)	Mean Euclidean distance between the nearest pixels from the same LCZ class (km).	A,5,6	D,6	8	A	A	D	−	A
Wei. Standard distance (DEP)	Area-weighted mean distance of patches from the same LCZ to their centroid (km). Informs about the concentration degree.	−	A,5	−	5	A,6	A	D,6	8
Object density (DO)	The number of patches of the same LCZ divided by the area of the city.	6,9	B,5,6	D,5	−	D	D,5	D,8	D,5
Urban density (DU)	The ratio between the built-up area (LCZ2-10) and the city area.	−	X	−	−	−	X	X	−
Radius dimension (DimR)	Measures the centrality of the LCZ classes with respect to the city center given.	A,6,9	A,B,G	A,6,9	G,8,9	6,9	6,9	A,6,9	6,9
Coherence deg. (GC)	The probability that two random points are in the same patch in a city.	X	−	X	−	X	X	X	X
Shape index (IF)	A normalized ratio of patch perimeter-area in which the complexity of patch shape is compared to a square of the same size.	2	−	2	−	2	−	2	2
Splitting index (IS)	The number of patches when dividing the LCZ class into equal size parts with the same division.	X,G	F,10	X,6	A,6	X,6	X,6	X,6	X,6,10
Leapfrog (LPF)	The proportion of isolated pixels with respect to the entire LCZ class.	6	5,8	A,B	A	5,6	−	5,6	5,6
Urban−/porosity (PU, P)	The ratio of open space (area of holes within the built-up area or LZC class) compared to the city or LCZ area (Reis, Silva, & Pinho, 2015).	X	−	5,9	X,9	5	−	5	5
Contrast (RCB)	The sum of the segment lengths of pixels adjacent to different LCZ, divided by the perimeter.	D	D	A	X	−	X,D	D	−
Effective mesh size (TEM)	Measures the connectivity. Low values mean fragmentation (ha).	2	6,8	2,8	2,6,8	2,6,8	X,2,8	X,2,8	X,6
Object mean size (TM)	Average size of the patches from a LCZ class (ha).	2,4,9	F,G,9	2	G,2,4	F,2	−	2,9	2

Table 5

Multiple linear regression models for the normalized socio-economic variables, where higher values mean better conditions for all variables (dependent variables, DV), using the spatial metrics (independent variables (IV) in bold, with the LCZ class in the subscript). The intercept, coefficients of IV, leave-one-out cross-validation coefficient of determination (R^2), the root mean square error (RMSE), the *p*-value of the model, and the number of observations or cities (Ob) are shown. The acronyms of LCZ and spatial metrics can be found in Tables 3 and 4, respectively.

DV	Intercept	IV 1	IV 2	IV 3	IV 4	R^2	RMSE	*p-value*	Ob
Education	0.407	$0.159 \cdot \mathbf{DEMp_5}$	$0.119 \cdot \mathbf{DimR_A}$	$-0.110 \cdot \mathbf{PU}$	$0.077 \cdot \mathbf{DC_4}$	54.87	0.174	$7.7 \cdot 10^{-6}$	31
Health	0.389	$0.130 \cdot \mathbf{DC_F}$	$0.102 \cdot \mathbf{DEM_B}$	$0.081 \cdot \mathbf{DEM_9}$	$0.081 \cdot \mathbf{IS_{10}}$	49.78	0.175	$9.9 \cdot 10^{-6}$	31
Log(housing)	0.536	$0.138 \cdot \mathbf{C_5}$	$0.093 \cdot \mathbf{P_9}$	$-0.060 \cdot \mathbf{TM_2}$	$0.057 \cdot \mathbf{DEP_5}$	50.61	0.151	$1.2 \cdot 10^{-5}$	31
Income	0.560	$0.112 \cdot \mathbf{TEM_6}$	$-0.092 \cdot \mathbf{IS_A}$	$-0.057 \cdot \mathbf{P_9}$		43.4	0.165	$2.7 \cdot 10^{-5}$	31
Affordability	0.580	$-0.088 \cdot \mathbf{TM_2}$	$0.076 \cdot \mathbf{C_5}$	$-0.067 \cdot \mathbf{IS_6}$		53.98	0.153	$3.9 \cdot 10^{-6}$	31
Employment	0.281	$0.108 \cdot \mathbf{DimR_6}$	$0.108 \cdot \mathbf{DEM_9}$	$0.093 \cdot \mathbf{TEM}$		56.84	0.157	$2.0 \cdot 10^{-7}$	31
Transport	0.457	$0.236 \cdot \mathbf{TEM}$	$-0.149 \cdot \mathbf{GC}$	$0.114 \cdot \mathbf{DO_8}$	$-0.057 \cdot \mathbf{DC_9}$	51.22	0.166	$5.1 \cdot 10^{-6}$	29
Commuting	0.637	$-0.222 \cdot \mathbf{C_9}$	$0.100 \cdot \mathbf{LPF_6}$	$0.078 \cdot \mathbf{DEM_9}$	$-0.059 \cdot \mathbf{IS_6}$	82.29	0.101	$7.4 \cdot 10^{-11}$	31

plots in Fig. 3).

The interpretation of groups by their mean values (i.e.: group centroids using the non-scaled socio-economic variables, Table 6), shows that *group 1* is formed by four cities with medium and low rates of mortality and low-education, the prices of buying an apartment are the lowest in contrast to the highest income levels (i.e., the capacity to pay for housing is higher), however, the low employment is balanced by the highest commuting level to work out of the city. *Group 2* accounts for the majority of cities (15 out of 31). This group is characterized for having lower education and employment together with higher death rates, the prices for buying an apartment and the income are medium-low in

comparison with the rest of the groups, and a close to 15% commute out of the city. *Group 3*, which most closely approximates to the mean values of the region (Table 6), clusters seven cities; in this group the education level and health are medium, the price of buying an apartment is high in contrast with the lower income levels (i.e., low capacity to pay for housing), however, the employment rate is medium-high and there are low commuting rates. Finally, *group 4* gathers five cities with the lowest proportion of the low-educated population, lower rates of mortality, and the highest prices for buying an apartment accompanied by high-income values; however, the huge discrepancy suggests housing prices are less affordable, the employment rate is the highest of the region and the level
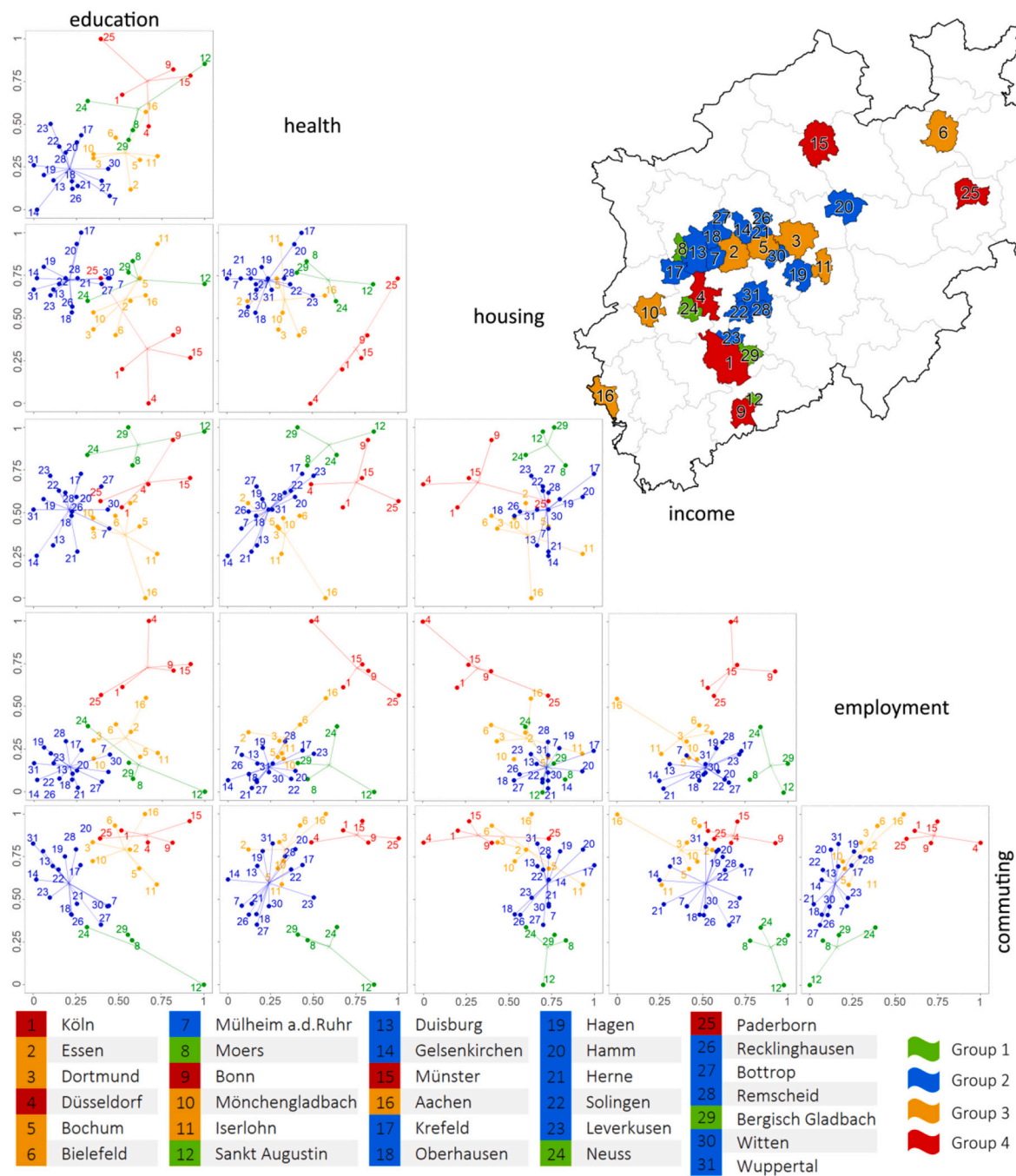
Fig. 3. Clustering of cities into four groups using the scaled socio-economic variables. The individual scatter plots show: the location of the cities according to each pair of socio-economic variable (row and column, e.g.: the top-left plot corresponds to 'education' and 'health'), the centroid of each group, and the distance of cities to their centroid. The map locates spatially the clusters and combined with the table identifies the cities (identification number, group and name of the city). It compares cities relatively based on to their socio-economic performance and groups them according to their similarities.

Table 6
Mean of non-scaled socio-economic variables (centroids) and number of cities per group. The last row shows the mean values of the NRW region.

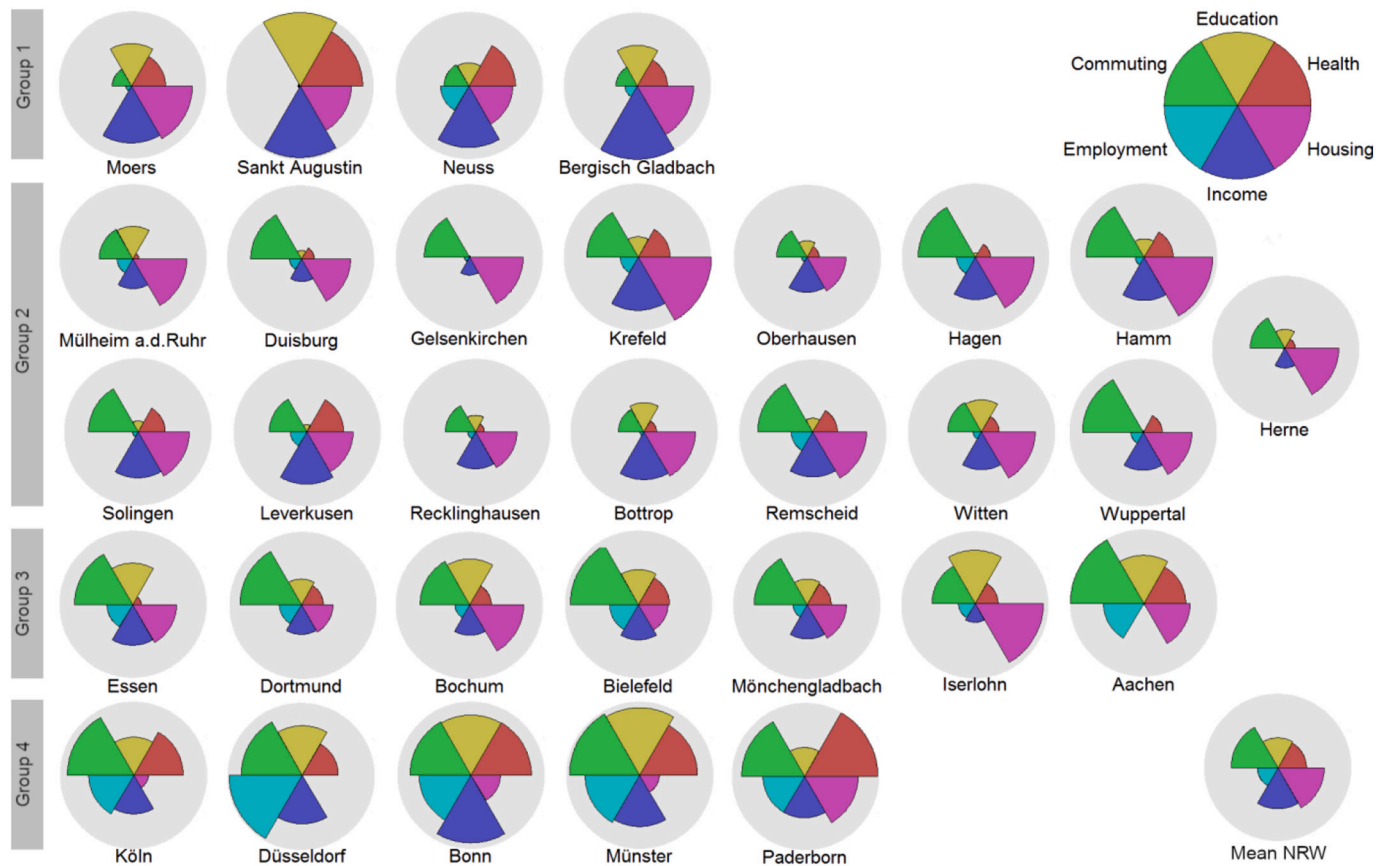| Group	Cities	Education (%)	Health (n°/1000)	Housing (€)	Income (€)	Employment (n°/1000)	Commuting (n°/1000)
1	4	27.59	9.92	91,875	23,975	426.01	215.27
2	15	37.24	11.89	91,933	20,900	421.49	148.49
3	7	29.44	11.36	104,000	19,700	498.79	112.94
4	5	26.37	9.01	134,400	22,200	686.05	97.82
NRW	31	32.48	11.05	101,500	21,235	482.20	140.91

Fig. 4. Multi-dimensional quality of life star plots of cities by group. Values are relative, as the socio-economic variables were min-max normalized between zero and one. For education, health, housing, and commuting the normalization was inversed and thus, higher values mean better conditions. The legend (top right) shows the maximum value of each socio-economic variable, equal to one, and its name related to the position and color. The mean values of the NRW region are represented in the bottom right. The gray background shows the maximum reachable value.

of commuting out of the city is the lowest.

By representing the cities multi-dimensionally using the socio-economic values by means of star plots (Fig. 4), the shape of each city becomes an indicator of its 'quality of life' (here based on five dimensions), the more complete (i.e., the area of the gray circle is covered) the better. *Group 1* shows high levels of commuting out of the city, education, health, house affordability, and income, but very few employments (work place-based). This shape can be related to satellite cities with good quality of life (regarding education, health and living conditions) but a less desirable situation in terms of sustainability due to the high commuting shares, to balance against the low employment rate. In *group 2* we find the lowest values of education and health in the region, commuting is medium-high and housing is affordable compared to income levels, however employment is quite low. There are similarities with the first group in the values of employment and housing, however, the analysis of the rest of variables suggests that this group has the lowest quality of life relative to the entire region. *Group 3* presents the lowest values of income in the region, and health is slightly lower than that of the mean NRW value. However, the remaining socio-economic variables are quite close to the mean values, which may suggest a quality of life close to the NRW average. Finally, *group 4* has the lowest values of commuting out of the city, while education, health, employment and income are considerably high and, as a counterpart, housing is less affordable. Additionally, this can be considered the most sustainable group in terms of commuting shares. Thus, according to the analyzed dimensions, it could be objectively said that it shows the highest quality of life in the region.

The spatial structure of urban spaces, as mentioned in the introduction, is related to this measured 'quality of life'. To explore such

relationships, we selected the spatial metrics that best identify these groups. We started from the subset of metrics selected with the VSURF method. Five spatial metrics for three structural types were the most influential in terms of grouping cities into different levels of quality of life. Those metrics were: the distance between sparsely built and open midrise structures patches within the city (DEM_9 and DEM_5), the number of open areas within the sparsely built patches (P_9), the connectivity and size of open low-rise patches (TEM_6), the compactness of open midrise (C_5), and the centrality (proximity to the city center) of sparsely built and open low-rise ($DimR_9$ and $DimR_6$). The spatial patterns that better differentiate between the derived levels of quality of life can be analyzed by representing the values of these metrics for each socio-economic group in box-and-whiskers plots (Fig. 5). The spatial patterns that better represent the cities in *group 1* are the presence of the biggest continuous areas of open low-rise, the highest compact shapes of open midrise patches but spatially scattered, and the compact distribution of sparsely built close to the city centers. For *group 2*, the metrics portray an even distribution of the sparsely built areas through the city, with fragmented and centralized open low-rise. *Group 3* shows open midrise structures scattered across the city, plus high values of open areas in the sparsely built environment, close to each other but farther from the urban cores, pointing that these urban structures are located in the surrounding areas of the city centers that are mainly occupied by high and medium rise types. Finally, cities in *group 4* are especially characterized by a compact nucleus of open midrise structures, with irregular shape, combined with fragmented distribution of sparsely built far from the urban cores, probably as they are located in the outskirts of the city, as well as the fragmented and decentralized distribution of open low-rise (Fig. 5). That is, cities in *group 4* have a compact urban core
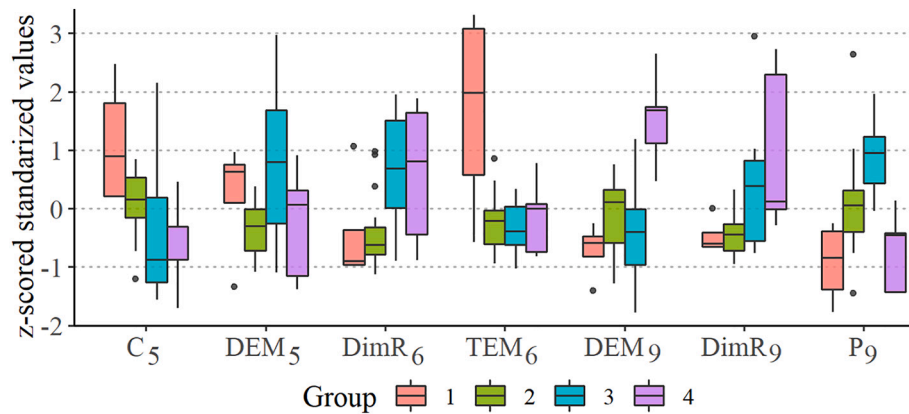
Fig. 5. Box-and-whiskers plot illustrating the standardized values of the spatial metrics for each socio-economic group of cities. Where: C = compactness, DEM = patch nearest neighbor, DimR = radius dimension, TEM = effective mesh size, and P = porosity. The subscript shows the LCZ: 5 = open midrise, 6 = open low-rise and 9 = sparsely built.

becoming gradually less compact as the distance to the core increases, eventually with low-dense structures located in the outskirts. For example, Münster (detailed example from Fig. 2) present this spatial pattern, with a compact midrise core (orange), with decentralized fragmented clusters of open low-rise (red) intermixed with a scattered and isolated distribution of sparsely built (pink).

5. Discussion

Our study in the cities of North Rhine-Westphalia in Germany shows the interrelation of urban spatial structure with quality of life dimensions. Our findings show that the education, mortality, income, employment, and other quality of life indicators can be partially explained by urban spatial pattern metrics extracted from urban structural types and land covers. No more than four metrics were needed to explain more than 40% of the variability of the socio-economic levels in cities with a similar economic and historical background for a given time. For example, the level of education tended to be better in more compact cities but also in cities with low-dense structures (i.e., open low-rise and sparsely built), which correspond to major cities and their satellite cities in NRW, respectively. This link can be related to higher-educated people moving to bigger cities seeking better job opportunities, and eventually moving to satellite cities. This seems to differ with a study where higher education levels were found in low-dense urban areas against high-dense areas in North America (Batchis, 2010). Cities with distant agglomerations of sparsely built areas and vegetation, combined with fewer and more scattered industry areas showed fewer death rates. In this sense, Oliveira (2016) compiled case studies that related walkability, diversity of land uses, and urban form with an improvement in health habits. The positive relation between death rate and bigger areas of heavy industry, besides higher shares of death in cities from groups 2 and 3, could be related to the fact that most of these cities are located in the highly industrialized Ruhr region, where death rates are high (Kibele, 2012). Apartments in cities of NRW with midrise structures (compact in the core and open towards the suburbs) and patches of dense tree are prone to be more expensive. We also found that income is measured higher in cities with a larger share of continuous and homogeneous areas of very low-dense built-up areas (i.e. sparsely built and open low-rise structural types), a spatial pattern that is especially seen in the satellite cities in NRW. Similarly, we found that commuting out of the city is higher in cities with more clusters and more compact areas of these low-dense built-up structural types, and the share of people choosing to commute by car or motor cycle is higher in less diverse and low-dense cities, which could be related to more mono-functional and dispersed cities. This tendency is widely discussed in the literature, for example, Travisi, Camagni, & Nijkamp (2010) found

higher automobile dependency in low-dense Italian cities. The positive relation of low-dense cities with higher incomes and commuting shares, especially by car or motor cycle, is likely to be linked to preferences of high-income households to live in less dense areas despite the higher travel cost. Additionally, the proportion of employment showed a positive relation to the homogeneity of structural types, they seem to be more organized, that may suggest that cities with more jobs are planned in a more uniform spatial distribution, with the exception of the sparsely built that tends to be more fragmented in these cities. Other authors also found relationships between spatial metrics and percentages of land uses with employment sector statistics (Ghafouri, Amiri, Shabani, & Songer, 2016).

The socio-economic variables used in this study cover several dimensions of quality of life (Eurostat, 2017). Therefore, grouping cities according to the socio-economic variables allowed us identifying various levels of quality of life within the analyzed cities. One group presented the lowest level in the region, but it does not necessarily mean that the quality of life is poor because we are comparing relative values. On the contrary, two groups stood out for having better levels of quality of life. These groups differ in commuting patterns, housing affordability, and employment rates, and coincide with major cities and satellite cities. Despite having a good quality of life, satellite cities here identified with low-dense built structures, are unsustainable in terms of commuting and transport choices, besides low-dense cities are more inefficient in the use of land, energy and resources (Bhatta, 2010). We also found common spatial patterns related to the built-up structural types in cities that had similar levels of quality of life, which again suggests the two-sided impact of spatial structure of cities on their socio-economic levels. We should note here that this specific morphology found for cities in NRW for a given date do not necessarily have the same relations in other areas. Context is - as Tonkiss (2013) argues - all in this debate. However, similar correlations between urban spatial structures and economical functions have been previously discussed in the literature. For instance, Mouratidis (2018) found a positive relation between social well-being and high density, short distances to the center, and land use diversity. Venerandi, Quattrone, & Capra (2018) modelled deprivation to population density, higher proportion of bare soil and regular street patterns, while other authors predicted indicators of wealth, poverty and crime through satellite data (Irvine, Wood, & McBee, 2017). Studies relating urban spatial structures with socio-economic values are usually focused in single dimensions, such as education, poverty, transport, air pollution, health, energy consumption, etc. (e.g.: Batchis, 2010; Duque et al., 2015; Hankey & Marshall, 2017; Sandborn & Engstrom, 2016; Wurm et al., 2019). Only a few of them analyze various indicators (e.g.: Irvine, Wood, & McBee, 2017; Sapena, Ruiz, & Goerlich, 2016) or combine them (Tapiador, Avelar, Tavares-Corrêa, & Zah, 2011). In contrast, we

tackled several aspects individually related to the quality of life and also in a combined way. We found that cities with a higher socio-economic status in NRW have a core with spatially compact midrise structures, while on the periphery there are small groups of low-rise structures and sparsely structures disaggregated with a high proportion of open green spaces. However, we are fully aware that the urban spatial structure of cities does not define or fully explain their success, since there are many other factors that play an important role. However, in the region analyzed, cities with similar spatial appearance also had similar socio-economic levels; this is a strong indication that the spatial structure of the cities do influence socio-economic performance. It should also be recalled that correlation does not imply causation, and thus the variation of a spatial pattern does not necessarily improve the socio-economic level of a certain area, although, it will certainly alter its state (Lehrer & Wieditz, 2009; Williams, 2014).

We also faced some limitations. We based the relations on a large and consistent set of variables. However, a comprehensive set of variables is inexistent which means that we are not able to depict the manifold interrelationships holistically. Moreover, it is worth noting that we did not include external influences in the models, such as policies, individual historical background, etc. While of course, every city is unique, the overall urban spatial structure of the analyzed cities is comparable to a certain extent, thus reducing these influences. However, when conducting a global analysis, externalities should be considered, as well as measuring the spatial stratified heterogeneity (Wang, Zhang, & Fu, 2016) to test whether the variables are distributed unevenly across different parts of the study area, in which case it would be convenient to perform different models. Besides, it was not possible to model some socio-economic factors such as the 'proportion of economically active population' or the 'share of persons at risk of poverty after social transfers' with the spatial distribution of LCZs, and thus they were not included in the analysis. Regarding the data used in this study, it is important to mention the following: on the one hand, the urban spatial patterns of cities were extracted by means of spatial pattern metrics. When working with spatial metrics, diminishing redundancies by removing duplicated information and the selection of the most significant variables is of great importance (Sapena & Ruiz, 2019; Schwarz, 2010). Another consideration is that the accuracy and spatial resolution of the image classification (here measured with 83% overall accuracy) affects to the spatial metrics, this fact needs to be considered when extracting conclusions of such studies. On the other hand, the socio-economic data used in this study have a great potential for comparative studies in Europe, but as mentioned, the average values at city level disregard internal socio-economic variations assuming that cities are homogeneous. Another limitation is the use of only five aspects of quality of life based on eight indicators, instead of a larger subset of socio-economic variables to enrich the analysis. Although these variables were able to represent a significant part of the different levels of quality of life in the region, we were subject to the availability of data. Moreover, data quality and spatial units should be considered (Schwarz, 2010; Venerandi, Quattrone, & Capra, 2018). Whereas remote sensing and GIS derived products, such as the LCZs classification, have no boundary or time-scale limitations, socio-economic statistics are usually restricted by administrative boundaries and census dates. Nevertheless, there is a recent tendency to provide these data in a different format, such as gridded datasets that swap irregularly shaped census boundaries to a regular surface (EFGS, 2019). These new datasets will serve as an opportunity to conduct studies that are not restricted by administrative boundaries.

The quality of life of the population and the sustainable development of urban areas are in the spotlight (OECD, 2017). Improving the understanding of the spatial structure of urban areas, the demographic, social and economic levels of these areas and their interrelations contributes to planning the development of cities with a view to meeting the global policy objectives set out in the New Urban Agenda (UN-Habitat, 2016). In order to unravel the interactions between the spatial structure

of cities and their socio-economic levels, in this paper we quantified their relationships by means of statistical models. This supports the hypotheses that assume that the spatial structure of cities reflects social and economic indicators of their inhabitants, and eventually influence their quality of life. The applied methodology can be used as a tool to obtain empirical evidences as well as learning from past trends and understanding the present to design a better future.

6. Conclusions

The spatial structure of urban spaces is related to the quality of live and sustainability of our cities. This is clearly confirmed by this analysis of cities in NRW. We extracted the spatial structure of cities using spatial pattern metrics from a LCZ classification based on machine learning algorithms applied to multimodal geospatial data. These attributes explained the variability of quality of life related indicator, which are linked to six out of seventeen SDGs. Moreover, grouping cities into different levels of quality of life showed common spatial patterns within the groups. We ascertained that the spatial structure of cities has a strong influence on their socio-economical functions, but does not fully determine them.

In times of increasing availability of socio-economic and spatial data (e.g. from remote sensing) in ever-increasing spatial resolutions, there is a huge demand for systematic research in this direction. Of particular interest is research that systemizes these relations in dependence of context, that is policies, culture, demography, etc., for a more general and quantifiable knowledge of the influence of the urban spatial structure on socio-economic parameters of cities and their people – this paper testifies to this.

Although this study accounts for cities in NRW in a specific period, and thus is not globally representative, results show a trend that is worth investigating further. This is feasible due to the growing availability of data for both local and global levels. Moreover, the methods applied in this study are directly transferable to other regions and datasets, which would broaden the analysis and derived conclusions. Additionally, the increase of the temporal scale would allow gaining knowledge of how urban growth affects cities and urban areas spatially and socio-economically, and inversely the influence of socio-economic policies and evolution on the urban growth patterns, as well as quantifying their interrelationships. The analysis of the relationships of urban developing processes, e.g. the effect of city design choices on urban fabric, is of high interest for future research and will provide a valuable source of information for supporting policy makers and city planners to ensure the quality of life and sustainable development in urban areas.

Declaration of Competing Interest

None.

References

Abrantes, P., Rocha, J., Marques da Costa, E., Gomes, E., Morgado, P., & Costa, N. (2019). Modelling urban form: A multidimensional typology of urban occupation for spatial analysis. *Environment and Planning B: Urban Analytics and City Science, 46*(1), 47–65.

Anderson, W. P., Kanaroglou, P. S., & Miller, E. J. (1996). Urban form, energy and the environment: A review of issues, evidence and policy. *Urban Studies, 33*(1), 7–35.

Batchis, W. (2010). Urban sprawl and the constitution: Educational inequality as an impetus to low density living. *The Urban Lawyer, 42*, 95–133.

Bechtel, B., Alexander, P. J., Böhner, J., Ching, J., Conrad, O., Feddema, J., … Stewart, I. (2015). Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information, 4*(1), 199–219.

Bechtel, B., Demuzere, M., Mills, G., Zhan, W., Sismanidis, P., Small, C., & Voogt, J. (2019). SUHI analysis using local climate zones - a comparison of 50 cities. *Urban Climate, 28*, 100451.

Bhatta, B. (2010). Causes and consequences of urban growth and sprawl. In *Analysis of urban growth and sprawl from remote sensing data* (pp. 17–37). Heidelberg, Germany: Springer.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*, 35–46.

Duque, J. C., Patino, J. E., Ruiz, L. A., & Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landscape and Urban Planning, 135*, 11–21.

European Commission. Growth, internal market, industry, entrepreneurship and SMEs. https://ec.europa.eu/growth/tools-databases/regional-innovation-monitor/base-profile/north-rhine-westphalia, 2019 (accessed 22/01/19).

EFGS. European and Global Forum for Geography and Statistics. https://www.efgs.info/, 2019. (accessed 10/03/19).

Eurostat. (2016). Cities (Urban Audit). https://ec.europa.eu/eurostat/web/cities/background (accessed 01/10/18).

Eurostat. (2017). *Final report of the expert group on quality of life indicators*. Luxembourg: Publications Office of the European Union.

Eurostat. (2018). Statistics on European cities. http://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_European_cities (accessed 09/10/18).

Eurostat. (2019). Regional statistics by NUTS classification. https://ec.europa.eu/eurostat/web/regions/data/database (accessed 15/01/19).

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2015). VSURF: An R package for variable selection using random forests. *The R Journal, 7*(2), 19–33.

Getis, A., & Paelinck, J. H. (2004). An analytical description of spatial patterns. *L'Espace geographique, 33*(1), 61–68.

Ghafouri, B., Amiri, B. J., Shabani, A. A., & Songer, M. (2016). Examining relationships between socioeconomic factors and landscape metrics in the southern basin of the Caspian Sea. *Environmental Modeling and Assessment, 21*(5), 669–680.

Grimm, N. B., Cook, E. M., Hale, R. L., & Iwaniec, D. M. (2015). A broader framing of ecosystem services in cities. In *Benefits and challenges of built, natural or hybrid system function*. Routledge Handbooks Online.

Haase, D., Kabisch, N., & Haase, A. (2013). Endless urban growth? On the mismatch of population, household and urban land area growth and its effects on the urban debate. *PLoS One, 8*(6), 1–8.

Hankey, S., & Marshall, J. D. (2017). Urban form, air pollution, and health. *Current Environmental Health Reports, 4*(4), 491–503.

Irvine, J. M., Wood, R. J., & McBee, P. (2017). *Remote sensing to analyze wealth, poverty, and crime* (pp. 1–9). Washington, DC: IEEE Applied Imagery Pattern Recognition Workshop.

Jacobs, J. (1961). *The death and life of great American cities*. New York: Random House.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794.

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–458.

Kibele, E. U. (2012). *Regional mortality differences in Germany*. Springer Science & Business Media.

Krehl, A., & Siedentop, S. (2019). Towards a typology of urban centers and subcenters–evidence from German city regions. *Urban Geography, 40*(1), 58–82.

Lehrer, U., & Wieditz, T. (2009). Condominium development and gentrification: The relationship between policies, building activities and socio-economic development in Toronto. *Canadian Journal of Urban Research, 18*(1), 140–161.

Montanges, A. P., Moser, G., Taubenböck, H., Wurm, M., & Tuia, D. (2015). Classification of urban structural types with multisource data and structured models. *Joint Urban Remote Sensing Event*.

Mouratidis, K. (2018). Built environment and social well-being: How does urban form affect social life and personal relationships? *Cities, 74*, 7–20.

OECD. (2017). *How's life? Measuring well-being*. Paris: OECD Publishing.

Oliveira, V. (2016). *Urban morphology. An introduction to the study of the physical form of cities*. Springer.

Patino, J. E., & Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. *Computers, Environment and Urban Systems, 37*(1), 1–17.

Patino, J. E., Duque, J. C., Pardo-Pascual, J. E., & Ruiz, L. A. (2014). Using remote sensing to assess the relationship between crime and the urban layout. *Applied Geography, 55*, 48–60.

Reis, J. P., Silva, E. A., & Pinho, P. (2015). Spatial metrics to study urban patterns in growing and shrinking cities. *Urban Geography, 3638*, 1–26.

Ribeiro-Barranco, R., Batista e Silva, F., Marin-Herrera, M., & Lavalle, C. (2014). Integrating the MOLAND and the urban atlas geo-databases to analyze urban growth in European cities. *Journal of Map & Geography Libraries, 10*(3), 305–328.

Salat, S. (2011). *Cities and forms: On sustainable urbanism*. Hermann.

Sandborn, A., & Engstrom, R. N. (2016). Determining the relationship between census data and spatial features derived from high-resolution imagery in Accra, Ghana. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(5), 1970–1977.

Sapena, M., & Ruiz, L. A. (2015). Description and extraction of urban fragmentation indices: The Indifrag tool. *Revista de Teledetección, 43*, 77–89.

Sapena, M., & Ruiz, L. A. (2019). Analysis of land use/land cover spatio-temporal metrics and population dynamics for urban growth characterization. *Computers, Environment and Urban Systems, 73*, 27–39.

Sapena, M., Ruiz, L. A., & Goerlich, F. J. (2016). Analysing relationships between urban land use fragmentation metrics and socio-economic variables. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1029–1036*. XLI-B8.

Schwarz, N. (2010). Urban form revisited: Selecting indicators for characterising European cities. *Landscape and Urban Planning, 96*, 29–47.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3–4), 591–611.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *American Meteorological Society, 93*, 1879–1900.

Tapiador, F. J., Avelar, S., Tavares-Corrêa, C., & Zah, R. (2011). Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. *International Journal of Remote Sensing, 32*, 6437–6456.

Taubenböck, H. (2019). *Remote sensing for the analysis of global urbanization*. DLR-Research Report 2019-10. Habilitation Thesis, University Würzburg. 10, 600.

Taubenböck, H., Wurm, M., Setiadi, N., Gebert, N., Roth, A., Strunz, G., … Dech, S. (2009). Integrating remote sensing and social science. *Joint Urban Remote Sensing Event*, 1–7.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., & Dech, S. (2012). Monitoring of mega cities from space. *Remote Sensing of Environment, 117*, 162–176.

Tonkiss, F. (2013). *Cities by design: The social life of urban form*. Cambridge, UK: Polity Press.

Travisi, C. M., Camagni, R., & Nijkamp, P. (2010). Impacts of urban sprawl and commuting: A modelling study for Italy. *Journal of Transport Geography, 18*(3), 382–392.

Tuia, D., Moser, G., Wurm, M., & Taubenböck, H. (2017). Land use modeling in North Rhine-Westphalia with interaction and scaling laws. *Joint Urban Remote Sensing Event*.

UN-Habitat. (2015). The economics of urban form: A literature review. In *Nairobi*.

UN-Habitat. (2016). In *World cities report 2016 - urbanization and development: emerging futures*. Nairobi: United Nations Human Settlements Programme.

United Nations. (2018). *World urbanization prospects: The 2018 revision, key facts*. Department of Economic and Social Affairs, Population Division.

Venerandi, A., Quattrone, G., & Capra, L. (2018). A scalable method to quantify the relationship between urban form and socio-economic indexes. *EPJ Data Science, 7*(4), 1–21.

Wang, J. F., Zhang, T. L., & Fu, B. J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators, 67*, 250–256.

Wentz, E. A., York, A. M., Alberti, M., Conrow, L., Fischer, H., Inostroza, L., … Taubenböck, H. (2018). Six fundamental aspects for conceptualizing multidimensional urban form: A spatial mapping perspective. *Landscape and Urban Planning, 179*, 55–62.

Williams, K. (2014). Urban form and infrastructure: A morphological review. In *Future of cities: Working paper*. Foresight, Government Office for Science.

Wurm, M., d'Angelo, P., Reinartz, P., & Taubenböck, H. (2014). Investigating the Applicability of Cartosat-1 DEMs and Topographic Maps to Localize Large-Area Urban Mass Concentrations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Wurm, M., & Taubenböck, H. (2018). Detecting social groups from space – Assessment of remote sensing-based mapped morphological slums using income data. *Remote Sensing Letters, 9*(1), 41–50.

Wurm, M., Weigand, M., Stark, T., Goebel, J., Wagner, G. G., & Taubenböck, H. (2019). Modelling the impact of the urban spatial structure on the choice of residential location using "big earth data" and machine learning. *Joint Urban Remote Sensing Event*, 1–4.