

Relationship between land cover type and Body Mass Index in Geneva

Estelle Rochat*[¶], Solange Duruz*[¶], Ivo Widmer*, Alain Clémence[†], Olivier Desrichard^{||}, Daniel Rappo[‡], Olivier Ertz[‡], Jens Ingensand[‡], Jean-Marc Theler[§], Idris Guessous[§] and Stéphane Joost*

*LASIG, EPFL, Lausanne, Switzerland; Email: stephane.joost@epfl.ch

[†]Laboratoire de Psychologie Sociale, UNIL, Lausanne, Switzerland

^{||}Groupe de recherche en psychologie de la santé, UNIGE, Genève, Switzerland

[‡]COMEM& G2C, HEIG-VD, Yverdon-les-bains, Switzerland

[§]Unit of Population Epidemiology, HUG, Geneva, Switzerland

[¶]These authors contributed equally to this work

Abstract—Past studies conducted in urban areas analyzed the impact of the presence of green spaces on public health, and highlighted in particular the psychological benefits of interacting with nature. To investigate a supposed relationship between overweight and dense built environment, we focused on the State of Geneva, Switzerland, and calculated the correlation between Body Mass Index (BMI) in a representative sample of 6663 adults and the percentage of natural areas at the locations where these individuals were living. To this end, we used population-based health data from the “Bus Santé” study (Geneva University Hospitals) and multi-scale land cover maps obtained by means of satellite images and LiDAR data classification. We found little correlation between BMI (as a proxy for health) and land cover data and were not able to verify the working hypothesis at local and regional scales. However, an important phenomenon highlighted here is the difference in the results obtained between the city center and the whole State.

I. INTRODUCTION

During the last few decades, the increasing concentration of the population in cities led to a progressive destruction of green spaces [1]. This gradual urbanization implies a strong fragmentation of semi-natural environments, what clearly endangers biodiversity and causes decline in the quality of life of the population. Indeed, the presence of natural environment in cities favors the good health of the residents, since the latter are encouraged to spend time outside and to carry out physical activities [2] [3] [4]. However, this trend has to be balanced by other results showing no significant correlation between obesity in children and the presence of green space [5]. Here, the results obtained are explained by different factors, including socio-cultural aspects (car dependent cities and communities), climate (long winters in Calgary [5]), and scale-related issues (analysis at the city level in [3] and [5], and at the national level in [2]). It is therefore difficult to know *a priori* if the presence of green spaces will affect the weights of the inhabitants of a given area.

In this paper, we used classifications based on a SPOT satellite image and on LiDAR data (also useful for classification, see [6]) to identify green/natural spaces versus concrete/asphalt surfaces in the urban State of Geneva, Switzerland. This permitted to calculate the percentage of the classes in regular grids showing cells of different sizes, as well as in

statistical sectors subdividing the territory under study (multi-scale approach). To test for the existence of a potential positive effect of the natural environment on the health of residents, we calculated the correlation between the percentage of natural classes around the resident’s home mentioned above and Body Mass Index (BMI) data measured by the Unit of Population Epidemiology of the Geneva University Hospitals (HUG).

II. DATA

A. Satellite image

The multi-spectral SPOT-5 image from May, 24, 2004, was used. This image contains four reflectance bands (green, red, near infrared and mid infrared) and has a spatial resolution of 10m [7]. The spatial extent of the picture was clipped to the the State of Geneva (“SPOT area” in the rest of the paper).

B. LiDAR and DEM

For the classification based on LiDAR, we used a point cloud acquired in 2009, with a mean point density of 6 pts/m². On this basis, the “Système d’Information du Territoire à Genève” (SITG) created a Digital Terrain Model (DTM) and a Digital Surface Model (DSM) [8] in raster format with a spatial resolution of 1m (applying an MDL algorithm). We used these two models as well as a 2D-model of buildings in a vector format (also provided by the SITG) [9]. The spatial extent covered by LiDAR classification is smaller than the SPOT area mentioned above. It is limited by the following coordinates in the Swiss Coordinate System CH1903: $X_{\min} = 492000$; $Y_{\min} = 116000$; $X_{\max} = 501000$; $Y_{\max} = 122000$. This area will be referred to as “LiDAR area” and approximately corresponds to Geneva downtown. The classification obtained with the LiDAR data thus shows a smaller extent but a finer spatial resolution.

C. Health data

We used health data provided by the “Bus-santé” study of the Population Epidemiology Unit of the Geneva University Hospitals (HUG). Between 2001 and 2009, every year a representative stratified sample of 500 men and 500 women from the Geneva States population (~ 500 000 inhabitants) is

recruited and studied [10]. The sample used here is composed of 6663 adults, and describes the age, the postal address and the BMI of each individual. This index is calculated as follows: the weight (in kg) of an individual divided by the square of his height (in m).

As income level is thought to be an obesity risk factor to be considered when conducting BMI spatial dependence analyses [10] (here corr coeff $r=-0.08$), BMI was adjusted for areas income level by means of median regression. The area used here is a statistical subsector named “GIREC” characterized with income level obtained from the 2009 Geneva Census (Office Cantonal de la Statistique, www.ge.ch/statistique). We will refer to “adjusted BMI” for the rest of the paper.

III. METHODS

A. Image classification

The method chosen is a supervised classification in five classes: i) Water (*water*); ii) Low vegetation (*vege*); iii) Forest (*forest*); iv) Bare soil (*soil_bare*); v) Urban (buildings and asphalt; *urban*). When used, the combination of low vegetation and forest is abbreviated *vf*, and in turn the combination of *vf* with bare soil is abbreviated *vsf*.

To maximize κ [11] in the cross-validation test, we used the quadratic method available in the Matlab function [12] “classify”, which performs an adjustment of a Gaussian curve with different variances for each class.

B. Classification based on LiDAR and DEM

We used an object-based approach to classify land cover using the LiDAR point cloud, the Digital Elevation Models (DSM and DTM) and the 2D-model of buildings [13]. Using the height (DSM minus DTM) and LiDAR intensity (the ratio of the strength of the light reflected from an object related to the light emitted [6]), five classes were obtained using the following criteria: i) Water (*water*) with height = 0 and intensity = 0; ii) Permeable soil (*soil_per*) with height < 1.5m and intensity > 30; iii) Forest (*forest*) with height > 1m, not matching the 2D built zone; iv) Impermeable soil (*soil_imp*) with height < 1.5m and intensity < 30; v) Buildings (*building*) with height > 1m matching the 2D built zone.

Note that the categories “Buildings” and “Impermeable soil” correspond to the “Urban” class obtained with the supervised classification (*urban*), while the class “Permeable soil” corresponds either to “Bare soil” or “Low vegetation” of the supervised classification. The combination of permeable soil and forest is called *sp_f*.

C. Interaction with heath data

Two approaches were used to assess the presence and to quantify the footprint of natural environment in the neighborhood of the individuals sampled: i) regular grids centered on the individuals with cells showing 3 different spatial resolutions (200m, 800m, 3200m), and ii) statistical subsectors (GIREC).

As regards regular grids, we computed the percentage of each land cover class in the cell encompassing each individuals postal address. On this basis a table was created to contain the

following variables for each individual: i) unadjusted BMI; ii) adjusted BMI; iii) percentage of the different land cover classes. Then we calculated the correlation between the two BMI variables and the percentages of land cover for the 3 regular grids.

The approach was different for GIRECs: instead of taking into account each individual separately, we computed statistics for the two BMI variables: average (*bmi_avg*), median (*bmi_med*), percentage of people showing overweight (BMI>25 as defined by the World Health organization [14], *perc_overw*) and percentage of obese people (BMI>30, *perc_obese*). Then for each GIREC, we also computed the percentage of each land cover class. Correlations were then computed between the BMI statistics listed above and land cover percentages. GIRECs with less than 5 sampled individuals were not taken into account in the analysis, as proposed by Guessous *et al.* [10].

IV. RESULTS

A. Classification

The level of accuracy as revealed by the κ coefficient [11] is of 0.93 for the supervised classification on SPOT images (SCSI) and therefore indicates a high degree of inter-rater agreement (Tab. I). The same coefficient was used to compare accuracy in SCSI results versus the ones obtained on the basis of the object-based LiDAR land cover classification (OBL). Tab. II takes into account pixels that are located within the LiDAR area only. The goal here is not to evaluate the power of the LiDAR technology, but rather to assess if the two approaches produce convergent and coherent results. In this case, the κ coefficient is of 0.59 and the relatively low accuracy values shown in Tab. II are mainly due to the difference in pixel resolution (1m versus 10m).

B. Correlation

Fig. 1 shows the correlation between BMI variables (un-adjusted and income-adjusted) and the classification based on the SPOT image. Fig. 2 shows the correlation matrix with the classification based on the LiDAR data. The results in both cases look quite similar: the BMI variables are not significantly correlated with any percentage of land cover class (correlation between 0.1 and -0.1), while the median income is inversely correlated with built surfaces (asphalt and concrete). In both cases, the correlation between the median income of the corresponding GIREC with the percentage of land cover type is also provided as a landmark. Indeed, there is a clear positive relationship between the different classes constituting a natural environment and high income values.

TABLE I. ACCURACY OF ALL CLASSES FOR THE SUPERVISED CLASSIFICATION OF SPOT DATA

	Vegetation	Water	Forest	Bare soil	Urban
User accuracy	0.97	1.00	0.95	0.88	0.92
Producer accuracy	0.95	0.97	0.96	0.95	0.87

TABLE II. COMPARISON OF THE ACCURACY BETWEEN THE SUPERVISED CLASSIFICATION OF SPOT AND LIDAR DATA

	Vegetation + Bare soil	Water	Forest	Urban
User accuracy	0.68	0.93	0.80	0.69
Producer accuracy	0.78	0.96	0.53	0.78

At the level of the GIREC sectors (Figs. 3 and 4), the correlation coefficients are weak (below $|0.3|$). Nevertheless, considering the SPOT area (whole State of Geneva canton, Fig. 3) one observes a weak negative relationship ($r \sim -0.2$) between green spaces (forest, low vegetation, bare soil) and unadjusted BMI-variables (average, median, percentage of obese people). On the opposite, urban land cover classes are weakly positively correlated ($r \sim 0.2$) with the same BMI-variables. These relationships however almost completely disappear when considering adjusted BMI values.

In general, the different responses obtained according to the SPOT-based and to the LiDAR-based classifications and perceptible in Fig. 3 are mainly due to the size of the area considered: considering the restricted LiDAR area and thus focusing on downtown Geneva (Fig. 4) reinforces the correlation coefficients mainly between the SPOT-based classification and BMI variables suggesting a disappearance of this signal as we move towards the urban periphery. In the same downtown area, a weak positive correlation exists between green spaces and BMI adjusted variables ($r \sim 0.1-0.2$), while a weak negative correlation ($r \sim -0.1$) appears between BMI variables and the urban environment; both constitute a switch as compared with the type and intensity of relationship observed in the SPOT area.

V. DISCUSSION

At the scale of the GIREC subsectors (3), the sign of the correlation calculated between the unadjusted BMI variables and land cover classes based on the SPOT classification are in accordance with information provided by previous studies [2], [3], [4], [15], [16]. This type of relationship would imply that individuals living in a dense urban environment are the

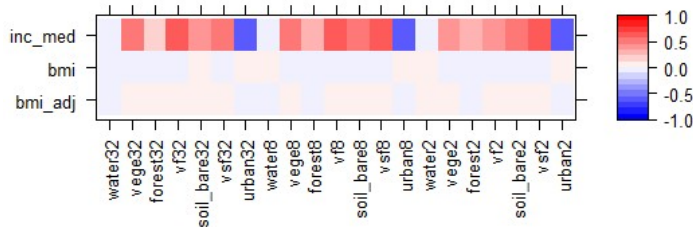


Fig. 1. Correlation values between BMI variables and percentage of land cover classes based on SPOT data. The numbers 2, 8 and 32 correspond to the size of the cell in meters divided by 100. The correlation with the median income (inc_med) is calculated for the GIREC including the cell of interest.

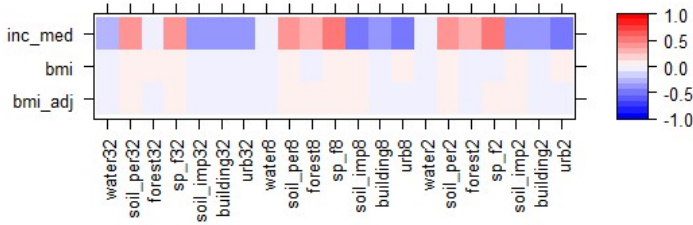


Fig. 2. Correlation values between BMI variables and percentage of land cover classes based on LiDAR data. The numbers 2, 8 and 32 correspond to the size of the cell in meters divided by 100. The correlation with the median income (inc_med) is calculated for the GIREC including the cell of interest.

most likely to show overweight and obesity problems. However the correlation values detected remain very low. Now, when observing the relationship with adjusted BMI variables, this correlation disappears and even becomes slightly negative clearly showing that there is no global association between the natural component of the environment and the health indicator characterizing individuals sampled in this study.

A positive correlation between the percentage of green spaces and income is clearly shown in Figs. 1 and 2). As BMI and income are slightly inversely correlated ($r = -0.08$, see section II, Health data), this implies the existence of a correlation between low BMI values and a high proportion of green spaces (as seen in Fig. 3 for SPOT data and unadjusted BMI values). Then, if an individual shows a high income and a low BMI, it is difficult to distinguish between a situation with a location of residence of better quality (greener) and favoring physical activity, and a situation for which the higher income makes it possible to eat healthily. The inclusion of other parameters like age and sex [17] would be necessary to better explain the possible influence of the natural environment on the health of residents.

About the influence of the area considered, in the LiDAR area we observe a weak positive correlation between adjusted

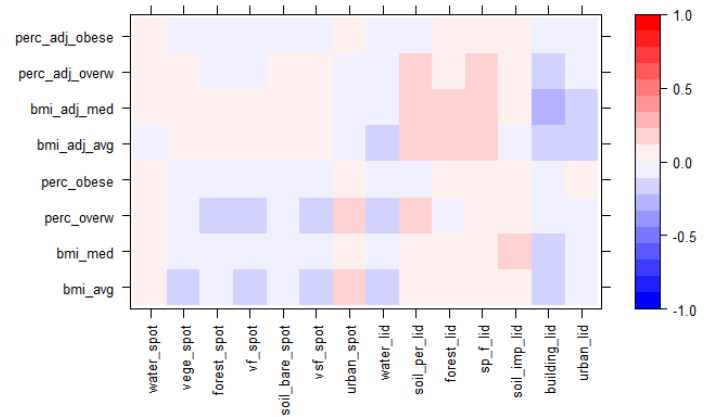


Fig. 3. Correlations values between BMI/weight-related variables and SPOT/LiDAR classifications in all GIRECs for SPOT and LiDAR zone for LiDAR

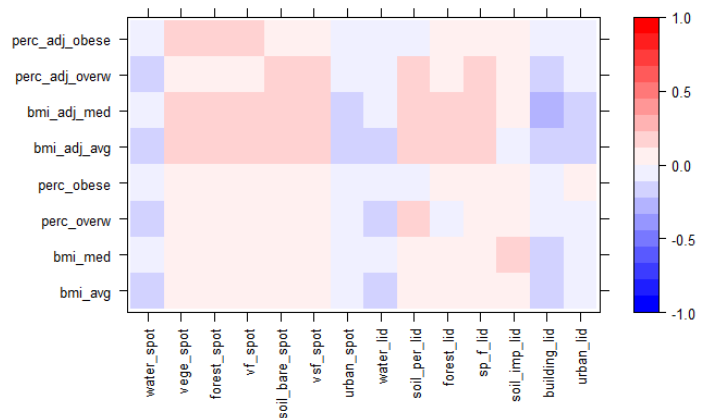


Fig. 4. Correlations values between BMI/weight-related variables and SPOT/LiDAR classifications in GIRECs included within the LiDAR area only

BMI variables and green spaces, which contrasts with the results discussed above but also with most of the results found in the literature (see introduction). The correlations calculated being very low, the result as such is not very meaningful. However this highlights the importance of the degree of centrality of the area considered rather than the role of the scale (see Tab. III, where + corresponds to a positive correlation [0 – 0.1], ++ to [0.1 – 0.2] and +++ to >0.2. The same is applied with inverse correlations (–, ––, – – –). A negative relationship is expected to support the working hypothesis according which a high percentage of natural environment would favor lower BMI values). As described in section IV-B, if we focus on the individuals located in the center of the city (LiDAR area) one gets a stronger positive correlation between BMI and natural environment classes, thus invalidating the potential beneficial effects of the latter on public health, and showing by the way a result congruent with what is observed in all cases with adjusted BMI values (see second column of Tab. III). Two explanations are possible and may merge with each other: i) the density of individuals sampled gradually decreases from downtown towards periphery in all directions causing a decrease in statistical power, and ii) the percentage of green spaces increases from downtown towards periphery in all directions progressively injecting an overestimation of green spaces for a decreasing number of individuals considered.

VI. CONCLUSION

To verify the hypothesis according to which green spaces have a positive effect on the health of residents in urban environments, we analyzed the relationship between the Body Mass Index (adjusted according to income) of 6663 individuals living in the State of Geneva and the percentage of natural land cover classes - determined with the help of SPOT images and LiDAR data - at their place of residence.

The correlations obtained at 3 different scales are very low but positive, while an inverse relationship was expected. However, while the sign and the intensity of the correlations is consistent through scales, an important phenomenon highlighted here is the difference in the results obtained between the central zone of the city and the suburbs. The difference of population density between downtown and the periphery makes it very difficult to prevent the number of individuals sampled from decreasing towards the outside of the city and causing a decrease in statistical power. Moreover, this process is parallel to the gradual increase of the percentage of natural environment towards the outside.

To improve the results, complementary analyses could be implemented at the national level. Moreover, further investigations need to rely on an adapted sampling design adding constraints e.g. weighting depending on population density on top of the stratified random sampling strategy implemented

TABLE III. BMI VERSUS NATURAL ENVIRONMENT RELATIONSHIP BASED ON SPOT DATA

	BMI vs Env	BMI_ adj vs Env
GIREC Downtown	+	+++
GIREC State (downtown + suburbs)	--	+
200m cell	-	+
800m cell	-	+
3200m cell	-	+

[10]. And other parameters known to influence BMI like the age and the sex of the residents will have to be included in the analysis.

ACKNOWLEDGMENT

We are grateful to Gisèle Ethal from the University of Lilles' Team for Random Analysis of Space and the GIRAPH group for having provided us with the necessary inspiration.

REFERENCES

- [1] S. Campbell, "Green cities, growing cities, just cities?: Urban planning and the contradictions of sustainable development," *Journal of the American Planning Association*, vol. 62, no. 3, pp. 296–312, 1996.
- [2] S. De Vries, R. A. Verheij, P. P. Groenewegen, and P. Spreeuwenberg, "Natural environments-healthy environments? an exploratory analysis of the relationship between greenspace and health," *Environment and planning A*, vol. 35, no. 10, pp. 1717–1732, 2003.
- [3] J. Feng, T. A. Glass, F. C. Curriero, W. F. Stewart, and B. S. Schwartz, "The built environment and obesity: a systematic review of the epidemiologic evidence," *Health & place*, vol. 16, no. 2, pp. 175–190, 2010.
- [4] A. Lee and R. Maheswaran, "The health benefits of urban green spaces: a review of the evidence," *Journal of Public Health*, vol. 33, no. 2, pp. 212–222, 2011.
- [5] M. L. Potestio, A. B. Patel, C. D. Powell, D. A. McNeil, R. D. Jacobson, and L. McLaren, "Is there an association between spatial access to parks/green space and childhood overweight/obesity in calgary, canada?" *International Journal of Behavioral Nutrition and Physical Activity*, vol. 6, p. 77, 2009.
- [6] J. H. Song, S. H. Han, K. Y. Yu, and Y. I. Kim, "A study on using lidar intensity data for land cover classification," in *ISPRS Commission III, Symposium 2002 September 9-13, 2002, Graz, Austria*, 2002.
- [7] S. I. Corporation. (2002) Spot-5 satellite sensor. [Online]. Available: <http://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/spot-5/>
- [8] SITG. (2009) Mns-modele numerique de surface/ mna hauteur. [Online]. Available: http://ge.ch/sitg/sitg_catalog/sitg_donnees?keyword=&distribution=tous&datatype=tous&topic=elevation&service=tous
- [9] —. (2009) Batiments et objets classes. [Online]. Available: http://ge.ch/sitg/sitg_catalog/sitg_donnees?keyword=&distribution=tous&datatype=tous&topic=structure&service=tous
- [10] I. Guessous, S. Joost, E. Jeannot, J. Theler, P. Mahler, J. Gaspoz, N. Cantoreggi, J. Chételat, and J. Simos, "A comparison of the spatial dependence of body mass index among adults and children in a swiss general population," *Nutrition & diabetes*, vol. 4, no. 3, p. e111, 2014.
- [11] D. Gomez and J. Montero, "Determining the accuracy in image super-vised classification problems," in *Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology*. Atlantis Press, 2011, pp. 342–349.
- [12] MATLAB. (2014) Discriminant analysis - matlab classify - mathworks schweiz. [Online]. Available: <http://www.mathworks.ch/help/stats/classify.html>
- [13] A. Antonarakis, K. S. Richards, and J. Brasington, "Object-based land cover classification using airborne lidar," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2988–2998, 2008.
- [14] WHO. (2004) Bmi classification. [Online]. Available: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html
- [15] G. F. Dunton, J. Kaplan, J. Wolch, M. Jerrett, and K. D. Reynolds, "Physical environmental correlates of childhood obesity: a systematic review," *Obesity Reviews*, vol. 10, no. 4, pp. 393–402, 2009.
- [16] E. Coombes, A. P. Jones, and M. Hillsdon, "The relationship of physical activity and overweight to objectively measured green space accessibility and use," *Social science & medicine*, vol. 70, no. 6, pp. 816–822, 2010.
- [17] L. R. Mobley, E. D. Root, E. A. Finkelstein, O. Khavjou, R. P. Farris, and J. C. Will, "Environment, obesity, and cardiovascular disease risk in low-income women," *American journal of preventive medicine*, vol. 30, no. 4, pp. 327–332, 2006.