Quantifying Total Imperviousness from Building Footprint Area and Very High Resolution Air Photographs

Dennis M. Fox[1,], Mostafa Banitalebi[1], Richard Fournie[2], Yacine Bouroubi[2]*

[1] Université Côte d'Azur, UMR ESPACE CNRS, Nice, France, dennis.fox@univ-cotedazur.fr, mostafa.banitalebi@univ-cotedazur.fr

[2] Department of Applied Geomatics, CARTEL, Université de Sherbrooke, Sherbrooke Canada, Richard.Fournier@USherbrooke.ca, Yacine.Bouroubi@USherbrooke.ca

* corresponding author

doi: 10.5281/zenodo.11584934

Abstract: Imperviousness is the sealing of the soil surface by artificial materials that inhibit water transfer between the surface and soil. It has become a major environmental indicator of land cover change due to its impacts on hydrological and energy fluxes in the environment. Quantifying imperviousness has improved in the past decades with remote sensing technologies, but several challenges remain due to classification errors. We developed an innovative method based on vector building and road layers to quantify total imperviousness with greater accuracy when compared to current methods. Imperviousness was predicted with an accuracy approaching or surpassing 90% depending on the method (random forest, regression): with building footprint alone, R^2 value is about 0.88 when comparing simulated to observed values, and this increases to 0.94 when a road layer is added.

Keywords: imperviousness; building footprint area; machine learning.

1 Introduction

The conversion of agricultural or natural land covers to urban/suburban uses represents one of the fastest land cover transitions occurring globally. As cities expand onto agricultural and natural soils, the surface is overlain with impervious materials that inhibit water flow into the soil. Imperviousness therefore lowers groundwater recharge and increases flood risk by increasing peak discharge and total storm runoff (Arnold and Gibbons 1996, Jacobson 2011). In addition, imperviousness aggravates urban heat island effects (Hua et al. 2020, Shi et al. 2023) among other impacts. Imperviousness is therefore a key environmental indicator that is evolving rapidly both spatially and temporally throughout the world. The objective of this study is to improve imperviousness estimates by using vector building footprint and road layers in combination with very high resolution (20 cm) ortho-rectified aerial photographs.

2 Materials and methods

Quantifying imperviousness was carried out in 2 steps. Firstly, imperviousness was quantified for a selection of cities in France; secondly, statistical methods were used to relate imperviousness to building footprint and road vector layers.

2.1 Quantifying imperviousness

For 37 main cities located throughout France, 100 x 100 m polygons were created over a range of apparent building density conditions. Cities from all 13 regions were included. The number of polygons per city ranged from 4 to 20 with a total of 230 polygons in all. For each polygon, 100 random points were generated. The points were overlain on high resolution (20 cm) aerial photographs, and the surface of each point was visually classed as impervious (building, road, parking lot or other) or pervious (vegetation, agricultural field). In all, 23,000 points were visually identified and manually inputted into an attribute table. These observations provided the reference or "Observed" imperviousness values.

2.2 Creating the imperviousness model

Observed imperviousness was related to 3 potential predictors: 1 – building area alone, 2 – building area and cumulative road length, 3 – building area and road area. Hence, for each of the 230 polygons, the building area, road length, and road area were extracted automatically from the National Geographic Information (IGN) BD-TOPO database; road area was calculated as the product of road length * road width. To predict imperviousness, several techniques were initially tested to relate imperviousness to the predictors described above. These were progressively reduced to the following methods: random forest, non-linear regression, and linear regression on transformed (square root) variables. For each method, the total sample of 230 observations was divided into 160 calibration polygons (70%) and 70 validation polygons (30%). Metrics used to evaluate model performance were the following: R^2, Mean Absolute Error (MAE), Root Mean Square Error (RMSE). For a selection of 68 polygons located in the central region of France, imperviousness values were categorized as Urban continuous or Urban discontinuous (Suburban).

3 Results

Results are presented in two sub-sections: the first provides basic descriptive statistics of the variables used in the analysis and the second deals with the prediction results.

3.1 Polygon data characteristics

In theory, the random point method should provide representative distributions of building area in the polygons. However, the distributions shown in Figure 1 (left and middle) are not identical. Both distributions show a

first mode in the 0-5% range, but the random numbers method shifts the second mode to greater values than the BD-TOPO footprint area data. Mean and median random number building footprint areas are 31.2% and 30.5%, respectively; corresponding values for the BD-TOPO data are 25.7% and 20.8%, respectively. The random numbers method therefore has slightly greater building area values than the BD-TOPO method. Figure 1 (right) plots random numbers percentage building area versus BD-TOPO percentage building area, and the curve shows that the initial imperviousness values, in the range of about 0-10% fall on the 1:1 line, but that most of the remaining values are slightly greater with the greatest spread in the 10–50% range. Visual examination shows that the building footprint in the BD-TOPO layer is based on ground level contours whereas the air photo random number method includes overhanging rooftops, so values are greater. The Pearson-r correlation coefficient for this plot is 0.94.

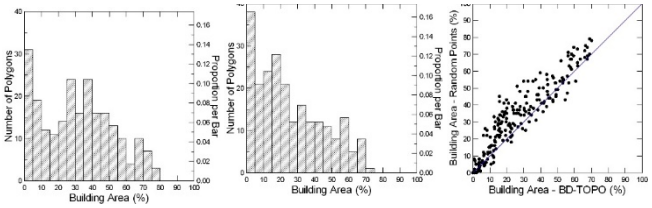

Figure 1. Frequency histogram of building area from random point data (left), Frequency distribution of building area from BD-TOPO (middle), Building area from random points versus building area from polygon layer (right).

Road length unit (m) (Figure 2 left) is different from the other values, so comparing distributions has limited meaning. The frequency histogram is provided nonetheless for other modelers; mean and median road length values are 212.8 m and 202.6 m per 10,000 m^2 polygon, respectively. Road area (middle) values are presented in the same % unit as building area values, and percentage road area rarely surpasses 20% of the total polygon area. Mean and median values are 8.4% and 8.1%, respectively. This corresponds roughly to about 25% to 30% of the percentage building area within a polygon.

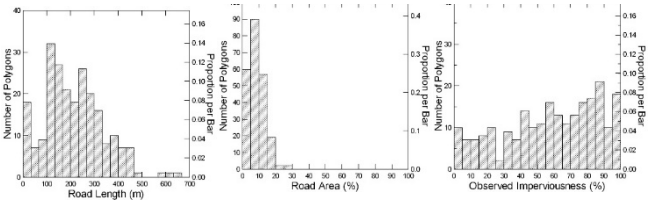

Figure 2. Frequency histogram of cumulative road length (left), Frequency histogram of road area (middle), Frequency Distribution of observed imperviousness values (right).

The distribution of observed imperviousness (right) reflects the cumulative impervious surfaces within a polygon. Where building area never surpassed 80%, imperviousness can reach 100%. For a selection of 68 polygons, mean and median impervious values for Urban were 88.6% and 89.5%, respectively. As expected, corresponding values for the Suburban category were lower at 70.3% and 73.0%, respectively. The range in imperviousness values was greater for Suburban (from 20.0% to 99.0%, std. dev 19.1%) compared to Urban, which ranged from 47.0% to 100.0% (std. dev. 10.3%), so in terms of imperviousness, the Urban category represents a less heterogenous environment than Suburban where imperviousness can be quite low to very high.

3.2 Predicting imperviousness from building and road layers

Prediction results will be presented according to the explanatory variables used, from the simplest (building area alone) to the most elaborate (building area and road area). Accuracy metrics for all the prediction methods are summarized in Table 1.

Figure 3 shows results for building area alone. Although the results are globally less good than when road length is included, the accuracy metrics shown in Table 1 remain high. Non-linear regression gives the best results, ahead of random forest, and linear regression has the lowest accuracy. This order is maintained throughout the results regardless of the explanatory variables used. In addition, residuals are clearly not randomly distributed in the linear regression model used.

Table 1. Accuracy metrics for the different predictive models from validation data (bold numbers represent best results).

Method	R^2	MAE	RMSE
RF Bldg. Area	0.83	8.37	11.69
Non-Lin., Bldg. Area1	0.90	6.25	8.97
Linear, Bldg. Area2	0.84	9.40	12.06
RF Bldg. Area & Road length	0.90	7.25	10.47
Non-Lin., Bldg. Area & Road length3	0.93	5.63	7.59
Linear, Bldg. Area & Road length4	0.87	9.22	11.52
RF Bldg. Area & Road area	0.90	6.32	8.81
Non-Lin., Bldg. Area & Road area5	0.93	5.41	7.39
Linear, Bldg. Area & Road area6	0.88	8.91	11.08

[1] imperv.=(8.661+3.982*x)/(1+0.025*x);
x=percentage building area;
[2] sqrt. imperv.=2.760+0.954*x;
x=sqrt. percentage building area;
[3] imperv.=(-1.431+4.033*x_1+2.322*x_2)/(1+0.030*x_1+0.009*x_2);
x_1=percentage building area,
x_2=road length;
[4] sqrt. Imperv.=1.595+0.820*x_1+0.651*x_2;
[5] imperv.=(-1.148+4.224*x_1+0.082*x_2)/(1+0.033*x_1); x_1=percentage building area, x_2=percentage road area;
[6] sqrt.Imperv.=1.752 + 0.825 *x_1+0.115*x_2;
x_1=sqrt. building area,
x_2=sqrt. road area

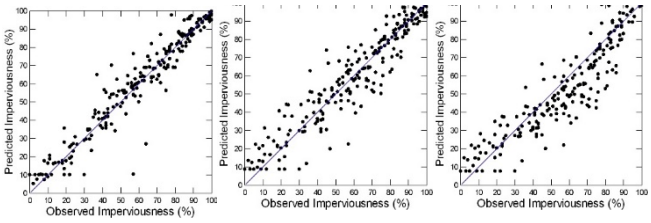

Figure 3. Imperviousness predicted from Random Forest on building area alone (left), non-linear regression on building area alone (middle), linear regression on building area alone (right).

Metrics improve as road length is added to percentage building area as an explanatory variable (Figure 4). The improvement is greatest for random forest which increases by 0.07 units in the R^2 value and the MAE decreases by about 1.1%. Corresponding values for non-linear regression are 0.03 and 0.62%.

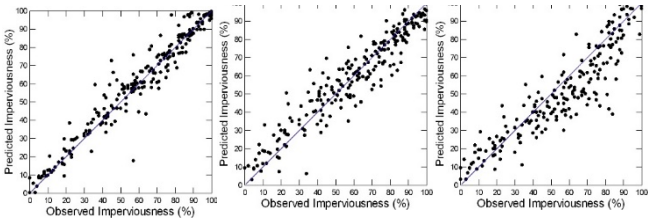

Figure 4. Imperviousness predicted from Random Forest on building area and road length (left), non-linear regression (middle), linear regression (right).

Using road area (Figure 5) instead of road length to predict imperviousness did not change the R^2 values for random forest or non-linear regression, but it improved both the MAE and the RMSE substantially (Table 1). When all three accuracy metrics are taken into account, the best performance is the non-linear regression using percentage building and road areas, non-linear regression using percentage building area and road length, and finally, random forest using percentage building and road areas.

Figure 6. Example of high imperviousness and low building/road areas.

4 Discussion

The methods described above provide a quick and efficient means of estimating imperviousness accurately from building and road vector layers. The most frequent cases of outliers (e.g. the 2 points in the lower right corner of Figure 3) arise in the presence of large parking lots where there are few buildings and roads but a high impervious surface (Figure 6).

When the observed values are compared to the values in the 100-m Copernicus Imperviousness Density 2018, the Pearson R-value is high (0.90), but Figure 7 shows that Copernicus tends to underpredict imperviousness below an imperviousness threshold of about 40% compared to our observed results. Similar results were also observed in Norway (Strand 2022).

5 Conclusions

Imperviousness was estimated accurately from building and road vector layers. Explained variance for the best models

was greater than 90% and MAE values were less than 6%. Where national building layers are available, quantifying the entire national coverage can take only a few hours versus several weeks or months of extensive image classification. Most countries are currently elaborating their own building footprint layers, so we can expect data

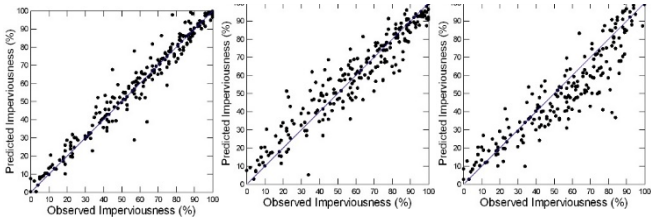

Figure 5. Imperviousness predicted from Random Forest on building area and road area (left), non-linear regression on building area and road area (middle), linear regression on building area and road area (right).

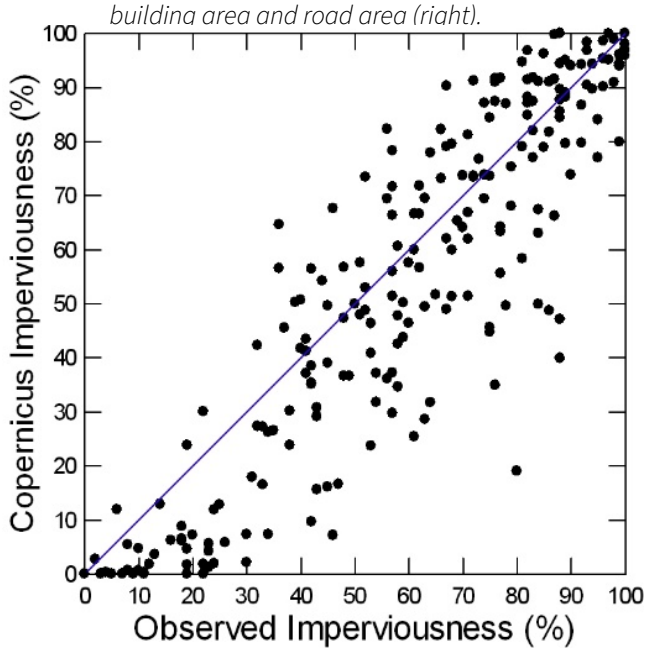

Figure 7. Copernicus imperviousness versus Observed imperviousness.

availability to grow exponentially. We are currently starting to explore the potential of extrapolating our results to other countries with freely available building polygon data.

Acknowledgments: This research was carried out within the framework of the FORESEE Project, supported by the Academy of Excellence 3 (Space, Environment, Risk and Resilience), and has been supported by the French government, through the UCA JEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01. The authors would also like to thank the interns who carried out the random point counting: L. Kitanovska, Y. Moshasha and P. Promduangsri.

6 References

Arnold, C.L., Gibbons, C.J., 1996. Impervious surface coverarge: the emergence of a key environmental indicator. Journal of the American Planning Association 62, 243-258.

Copernicus, n.d. High Resolution Layer Imperviousness, https://land.copernicus.eu/en/products/high-resolution-layer-imperviousness. (Accessed 11 June 2024).

Hua, L., Zhang, X., Nie, Q., Sun, F., Tang, L., 2020. The Impacts of the Expansion of Urban Impervious Surfaces on Urban Heat Islands in a Coastal City in China.

Jacobson, C.R., 2011. Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review. Journal of Environmental Management 92, 1438-1448.

Shi, Z., Li, X., Hu, T., Yuan, B., Yin, P., Jiang, D., 2023. Modelling the intensity of surface urban heat island based on the impervious surface area. Urban Climate 49, 101529.

Strand, G.H., 2022. Accuracy of the Copernicus High-Resolution Layer Imperviousness Density (HRL IMD) Assessed by Point Sampling within Pixels. Remote Sensing 14 (15), 3589.