



# Leaf Area Index for Assessment of Urban Green Space and Tree Coverage to Support Elder Friendly Cities

Jiri Horak\*

VSB-TU Ostrava, FMG, dept. of Geoinformatics,  
17.listopadu 2172/15, Ostrava, CZ  
jiri.horak@vsb.cz

Lucie Orlikova

VSB-TU Ostrava, FMG, dept. of Geoinformatics,  
17.listopadu 2172/15, Ostrava, CZ  
lucie.orlikova@vsb.cz

Pavel Kukuliac

VSB-TU Ostrava, FMG, dept. of Geoinformatics,  
17.listopadu 2172/15, Ostrava, CZ  
pavel.kukuliac@vsb.cz

Peter Golej

VSB-TU Ostrava, FMG, dept. of Geoinformatics,  
17.listopadu 2172/15, Ostrava, CZ  
peter.golej@vsb.cz

## ABSTRACT

Current walkability and living conditions assessments lack appropriate evaluation of urban green space and tree canopy which improves comfort of walking and living mainly for elderly. The study compares municipal registers of trees and green areas with satellite-derived Leaf Area Index (LAI) using Sentinel-2 imagery processed in SNAP. For evaluation correlation analysis, multiple linear regression (MLR) and XGboost modelling was applied for two types of pilot areas in Ostrava, CZ. Relationships between LAI and green, building and water areas, as well as number and kernel density of registered trees were explored. S-pilot areas with almost completed evidence of tree and green areas provide much better modelling results ( $R^2=0.41$ ) than U-pilot areas. For quantification of urban green space and trees, the LAI performs much better than the municipal evidence. XGboost modelling outperforms MLR and overcomes issues with heteroscedascity and normality of models' residuals.

## CCS CONCEPTS

• **Image processing**; • **Machine learning**; • **Environmental sciences**;

## KEYWORDS

urban green space, Sentinel-2, SNAP, LAI, XGboost, elderly

## ACM Reference Format:

Jiri Horak\*, Pavel Kukuliac, Lucie Orlikova, and Peter Golej. 2022. Leaf Area Index for Assessment of Urban Green Space and Tree Coverage to Support Elder Friendly Cities. In *2022 International Conference on Big data, IoT, and Cloud Computing (ICBICC 2022)*, December 02–04, 2022, Chengdu, China. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3588340.3588364>

## 1 INTRODUCTION

Trees represent an important part of viable urban environments in that they are essential for the well-being of citizens, biodiversity,

healthy air condition and mitigation of urban microclimates. Their occurrence positively influences inhabitants' propensity to walk and relax outside, especially for elderly and other vulnerable groups of people.

An environmental mobility determinant represents one of five fundamental categories of elderly mobility, usually recognised as a part of the "comfort" aspect. A 10-year review of urban walkability [1] discovered only 21% of papers discuss some components of comfort such as the weather, aesthetics of buildings, enclosure ratio, cleanliness, shade, and the presence of trees. Alvez et al. [2] designed Walkability Index for Elderly Health and included in the "urban scene" dimension the existence of trees/vegetation but with a simplified classification of urban condition into three levels (no trees/vegetation, moderate, and strong occurrence). A walkability index which quantifies tree occurrences or even tree shade does not yet exist. The reason for this is simple – a lack of data. To overcome this issue, we evaluated two possible data sources – digital municipal evidence of trees and green space, and several satellite-derived vegetation indices where the LAI seems to match best for quantification of the tree coverage.

Current cities recognize the value of city trees for the urban environment, carefully invest in them, expand green areas, water trees, and protect them against diseases and pests. The required digital evidence of public trees is commonly based on regular updating and monitoring. The problem is that this evidence covers only trees owned by municipalities, usually standalone trees, and is not comprised of all trees in parks and forests within a city, nor does it include private gardens with many trees. Additionally, tree registers usually do not contain important parameters for assessment of tree growth and its canopy such as age, height, and crown diameter.

The Leaf Area Index (LAI) is specified as the one-sided green leaf area per unit surface area. The LAI serves as a measure for the amount of plant canopy and its density. Both direct methods (e.g., leaf traps) and indirect methods (e.g., hemispherical photography, LIDAR or satellite imagery [3-4]) can be used for enumeration of LAI. LAI was utilized e.g., to measure the urban forest impact of decreasing UV and temperature, the effects of surface coverage types on temperature and relative humidity of the urban environment, and to explore different distributions of dense trees to ameliorate urban microclimates. A satellite-derived LAI faces several challenges – it measures 'effective' leaf area which is less than actual LAI due to the expected random distribution of leaves [5], or underestimation



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICBICC 2022, December 02–04, 2022, Chengdu, China

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9954-8/22/12.

<https://doi.org/10.1145/3588340.3588364>

for really high LAI values due to trends of saturation of vegetation indices caused by the limited penetrability of sunlight [6].

Large amounts of data and their asymmetric distribution require advanced processing and utilization of machine learning methods. Among others, the XGBoost (Extreme Gradient Boosting) method offers parallelized tree building, cache awareness and out of core computing, regularization for avoiding overfitting, efficient handling of missing data and in build cross-validation capability [7].

This paper evaluates the possibility of using the LAI instead of municipal evidence of trees and green areas for assessment of urban green/tree coverage, useful e.g., for walkability modelling.

## 2 STUDY AREA AND DATA SOURCES

The city of Ostrava (population 290,000; area 214 km<sup>2</sup>) lies in the NE corner of Czechia close to the Polish and Slovakian borders. Its heterogeneous urbanization consists of an agglomeration of urban blocks separated by crop fields, forests, and industrial parks.

To assess LAI, Sentinel 2 MSI image (9.9.2021, processing level 1C) was processed using SNAP (Sentinel Application Platform). The LAI was calculated with a Biophysical Processor using the following reflectance bands (B3, B4, B5, B6, B7, B8A, B11 and B12), as well as other information such as solar zenith and relative azimuth angles using radiative transfer models. For the Sentinel-2, images have been implemented in two different neural network architectures, the NNET 10m and NNET 20m.

Geographic Information System of Ostrava City (GISMO) comprises various data sets to support urban planning and development, maintenance of municipal property, and solving environmental and social issues in the city. The important part here is a register of trees, currently including 154,874 individuals. Unfortunately, the evidence covers only trees which are municipal property, thus only trees occurring in public spaces such as streets or parks. Private trees (in gardens, industrial, agricultural or forested areas) are not included. Also, some public spaces (parks, riverbanks, etc.) are not fully covered due to the number of trees and their natural reproduction. To address these issues, two types of pilot areas were delimited in Ostrava: S-areas with almost complete evidence of trees (dominating public spaces with many registered trees) where a good correspondence between evidence of trees and real tree coverage is expected (9 suitable pilot areas, sizes 0.9-3.4 km<sup>2</sup>), (fig 1), and U-areas with weak tree evidence due to prevailing family dwellings with private gardens, or natural or seminatural areas (i.e., river banks, parks, forests) (4 unsuitable pilot areas, sizes 0.9-1.6 km<sup>2</sup>) (fig. 2). Other auxiliary datasets include buildings, water bodies (OSM), and green areas (GISMO).

## 3 METHODOLOGY

Sentinel-2 MSI image processing covers atmospheric corrections, resampling, subset and LAI calculation. Atmospheric correction using Sen2Cor (ESA) was applied to the input Sentinel-2 L1C data. Sen2Cor is a standalone application that allows atmospheric, terrain and cirrus correction of TOA level 1C input data. Further, all bands were resampled to 10m pixel size. Finally, the Biophysical Processor S2 was used to compute LAI (fig. 1,2 down). All operations were conducted in free available SNAP software version 9.0.0 (<https://step.esa.int/main/download/snap-download/>). According to



Figure 1: Green space and trees in one S-pilot area (Ostrava-Poruba) (up) and LAI values (down)

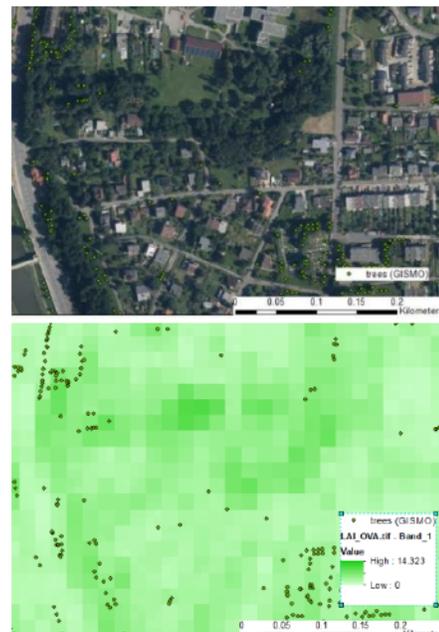


Figure 2: Green space and trees in one U-pilot area (Ostrava-Silesian) (up) and LAI values (down)

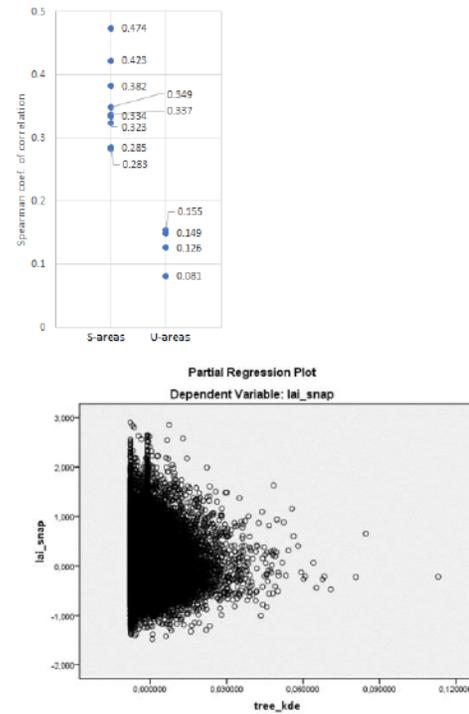
the recommendation of [8], LAI values are satisfactorily evaluated in the range of 0.0-8.0 values; cases with higher values outside tolerance limits are considered invalid. All cases with LAI values outside this range were excluded from the following statistical processing.

For the same spatial scheme, all other factors were enumerated. Point evidence of trees were aggregated to the cell units corresponding to LAI pixels (`tree_count`). Point locations of tree trunks is obviously not sufficient to represent a canopy. Trees create different crown sizes, on Ostrava's streets usually ranging from 0 to 18 m, with an average of 8 m. Accordingly, a kernel density estimation (KDE) of tree coverage was calculated (ArcMap v.10.8, Kernel density function, 10 m bandwidth and 10 m cell size) (`tree_kde`). Other supporting data is green areas registered by GISMO mainly for the purpose of lawn cutting and maintenance. The indicator was calculated as a percentage of pixel size coverage by green areas (`green_area`). Tree coverage is naturally suppressed by water bodies and buildings. Similar to green areas, water and building areas were enumerated with expectations of negative correlations to tree coverage and LAI (`water_area`, `build_area`).

An Explanatory Data Analysis provided basic statistical characteristics of data sets including histograms. Multiple scatter plots enabled evaluation of types of relationships between variables. Bivariate correlation analysis included Pearson and Spearman coefficients of correlation to explore the pairwise correlations between all variables. Further, multiple linear regression (MLR, SPSS version 18) is applied to understand the relationship of LAI (a dependent variable) on this set of independent factors, including multicollinearity evaluation, ANOVA, and assessment of beta standardized coefficients.

Finally, an advanced machine learning method, implemented in Python programming language, XGBoost (<https://xgboost.readthedocs.io/en/stable/>) was applied to improve the model and overcome issues of MLR evaluation. XGBoost [9], is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework. Boosting algorithms combine weak learners into a strong learner in an iterative way. The following parameters were tuned for XGBoost in this study: learning rate, number of gradient boosted trees (`n_estimators`), L1 regularization of leaf weights, minimum loss reduction, maximum depth of the tree (`max_depth`), fraction of features to be evaluated at each split (`colsample_bylevel`), subsampling rate (`subsample`), and random number seed (`random_state`).

In order to interpret the model and its results, we used unified framework SHAP (Shapley Additive Values), a widely used method to explain the predictions of machine learning models [10] where the model's prediction  $f(x)$  can be represented as the sum of its computed SHAP values, plus a fixed base value (Lundberg, Lee, 2017). To explain our model, we used SHAP TreeExplainer() class. The feature importance bar plot displays the important features in descending order of their importance and the magnitude of feature attributions. The summary plot combines feature importance with feature effects. Each point on the summary plot is a Shapley value for a feature and an instance. The features that influence the model's outcome in a positive way are highlighted in red, whereas the features that impact the model's outcome negatively are highlighted in blue [10].



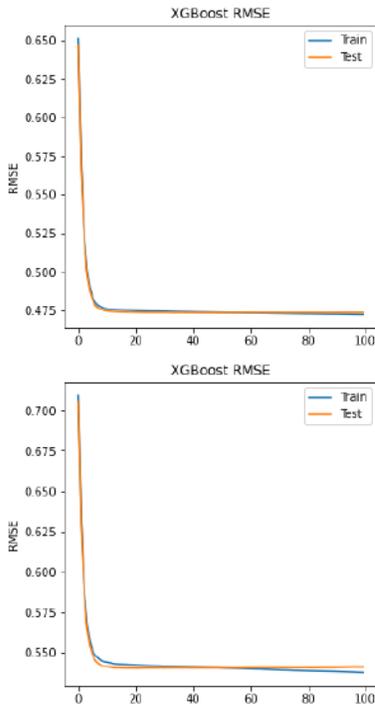
**Figure 3: Spearman  $r$  between LAI and `tree_kde` for both types of pilot areas ( $p=0.001$ ) and partial regression plots (MLR) for S-pilot areas**

## 4 RESULTS

Results of the bivariate correlation analysis for two main interested variables LAI and `tree_kde` show important differences of correlation between suitable and unsuitable areas (fig. 3 up). While the Spearman coefficient of correlation for S-pilot area ranges between 0.283 and 0.474 ( $p=0.001$ ), the correlations for U-pilot areas are significantly lower (0.081-0.155).

MLR for S-pilot areas ( $N=160997$ ) reached adjusted  $R^2=0.354$  ( $p=0.000$ ) and ANOVA discovered that the model explains 35% of data variability. MLR for U-pilot areas ( $N=49789$ ) provides only  $R^2=0.136$  and ANOVA indicates the model explains only 14% of data variability. Both models show satisfactory low multicollinearity (VIF in the range of 1.181-1.355) but residuals are not  $N$  distributed and some of the predictors indicate issues of heteroscedascity in partial regression plots (fig. 3 down).

To overcome issues in MLR models, XGBoost was applied. In order to evaluate the performance of the XGBoost model, we split our dataset into two train and validation subsets (80:20) using scikit-learn library (version 1.1.1) and the method `train_test_split()`. Train data is used for learning while validation data enables an unbiased evaluation of the model while tuning model hyperparameters ( $R^2$  and RMSE metrics). XGBoost model is defined as thus: `model_xgb = xgb.XGBRegressor(n_estimators = 100, learning_rate = 0.3, max_depth = 3, reg_alpha=0.5, gamma=0, subsample = 1, booster = 'gbtree', colsample_bylevel = 1, random_state = 44)`



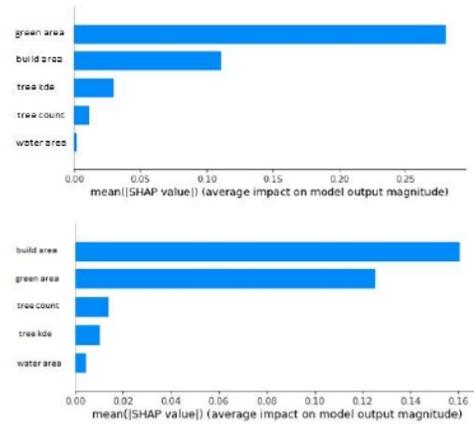
**Figure 4: Training and test losses for S-pilot (up) and U-pilot (down) areas**

The configuration of hyperparameters of XGBoost models is based on the performance model evaluation for each iteration of the algorithm during training and validation with a plot depicting learning curves using the train and test dataset where the x-axis is the number of iterations of the algorithm (or the number of trees added to the ensemble) and the y-axis is the RMSE of the model.

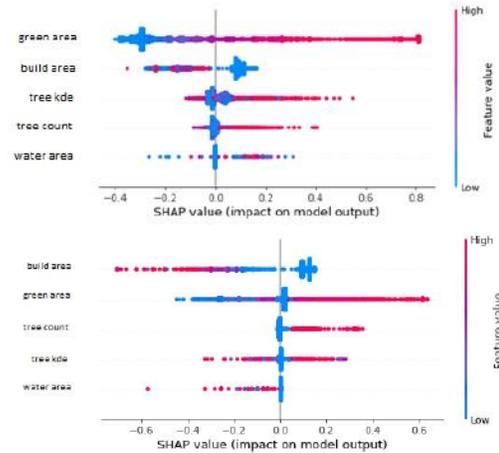
XGBoost was applied to both datasets covering suitable (S-pilot) and unsuitable (U-pilot) areas. The model for S-pilot areas achieved a coefficient of determination  $R^2=0.41$ . The plot of training loss decreases to a point of stability with  $RMSE=0.475$  (fig.4 up). Conversely, the model for U-pilot areas, as expected, achieved only  $R^2=0.23$ . The plot of training loss decreases to a point of stability around 0.54 (RMSE 0.54) (fig.4 down). The plot of test loss decreases to a point of stability and has, compared to the previous case, a bigger gap with the training loss suggesting overfitting.

The feature importance bar plots (fig.5) illustrate a feature importance as defined by SHAP algorithm. For S-pilot areas, the green\_area variable represents the most important feature, changing the predicted LAI value by 27 percentage points. It is followed by the build\_area variable, changing the predicted value by 12 percentage points. The influence of registered trees is about 3 and 1.5 percentage points for KDE and tree counts respectively. Surprisingly, the water\_area variable has very little impact on the model's predictions.

Results for U-pilot areas show changes in the order of feature priorities. As expected, the building area variable had the highest impact (16%), followed by the green\_area variable (13%). While



**Figure 5: Average feature importance (mean SHAP value) for S-pilot (up) and U-pilot (down) areas**



**Figure 6: SHAP values of variables in the XGboost model for S-pilot (up) and U-pilot (down) areas**

tree\_count also had the same influence as in S-pilot areas, the density of trees (tree\_kde) reaches only about half the influence as in U-pilot areas. This is caused by a large number of missing trees in the city's register (GISMO). The very low influence of water areas was confirmed.

A detail distribution of SHAP values can be explored in fig. 6 (the summary plots), where feature importance is combined with feature effects. As expected, large values of green\_area increase the predictions of LAI, while, on the other hand, and large values of build\_area make the model to decrease the predictions of the LAI. Tree\_kde and tree\_count both indicate that with more trees the predictions of the LAI values increase. Their positive effect is demonstrated by asymmetric value distribution in the case of S-pilot areas (fig. 6 up). The indifferent influence of tree\_kde in fig.6 downright shows issues with the model for U-pilot areas. Values of water\_area are distributed both in positive and negative influence

effects (SHAP values), resulting in an overall small average impact of this variable.

## 5 CONCLUSION

All used methods (correlation analysis, MLR and XGBoost modelling) confirm significant differences between S-pilot and U-pilot areas in Ostrava. S-pilot areas provide much higher parametric and non-parametric correlation coefficients between LAI and tree\_kde than U-pilot areas. MLR shows at least double adjusted  $R^2$  for S-areas compared with U-areas. However, results of MLR models failed in several modelling requirements (heteroscedascity, N distribution of residuals). XGBoost models achieved the best results. The model for S-pilot areas reached  $R^2=0.41$  (RMSE=0.48) where green and built-up areas play the main role followed by a cumulative importance of tree\_kde and count. Results for U-pilot areas are much worse. This confirms our hypothesis that the LAI works better than simple evidence of individual trees (or kernel density estimation of tree coverage) for evaluation of urban green space. It opens the possibility of using LAI as a suitable proxy for assessment of this important urban factor for walkability evaluation as well as other evaluations of urban living conditions.

## ACKNOWLEDGMENTS

This research was funded by the Czech Science Foundation, grant number 21-22276S "Accessibility modelling for seniors, perception of accessibility and determinants of their spatial mobility".

## REFERENCES

- [1] J. Arellana, M. Saltarin *et al.*, 'Urban walkability considering pedestrians' perceptions of the built environment: A 10-year review and a case study in a medium-sized city in Latin America', *Transp. Rev.*, 40, 183–203, 2020, doi: 10.1080/01441647.2019.1703842.
- [2] F. Alves, S. Santos Cruz, A. Ribeiro, A. Silva, J. Martins, and I. Cunha, 'Walkability Index for Elderly Health: A Proposal', *Sustainability*, 12/7360, 2020, doi: 10.3390/su12187360.
- [3] S. Vafaei, O. Fathizadeh, N. Puletti, H. Fadaei, S. B. Rasooli, and G. V. Laurin, 'Estimation of forest leaf area index using satellite multispectral and synthetic aperture radar data in Iran', *Iforest-Biogeosciences For.*, vol. 14, pp. 278–284, 2021, doi: 10.3832/ifor3633-014.
- [4] J.-P. George *et al.*, 'Method comparison of indirect assessments of understory leaf area index (LAI<sub>u</sub>): A case study across the extended network of ICOS forest ecosystem sites in Europe', *Ecol. Indic.*, vol. 128, p. 107841, Sep. 2021, doi: 10.1016/j.ecolind.2021.107841.
- [5] S. van der Graaf, R. Kranenburg, *et al.*, 'Satellite-derived leaf area index and roughness length information for surface-atmosphere exchange modelling: a case study for reactive nitrogen deposition in north-western Europe using LOTOS-EUROS v2.0', *Geosci. Model Dev.*, vol. 13, pp. 2451–2474, 2020, doi: 10.5194/gmd-13-2451-2020.
- [6] B. He, X. Li, X. Quan, and S. Qiu, 'Estimating the Aboveground Dry Biomass of Grass by Assimilation of Retrieved LAI Into a Crop Growth Model', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 2, pp. 550–561, 2015, doi: 10.1109/JSTARS.2014.2360676.
- [7] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, 'A Comparative Analysis of XGBoost', 2019. Accessed: 16.8.2022. <https://ui.adsabs.harvard.edu/abs/2019arXiv191101914B>
- [8] M. Weiss and F. Baret, 'S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER', 53, 2016.
- [9] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [10] A. Bhattacharya, *Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more*. Packt Publishing, 2022.