German Remote Sensing Data Center (DFD)
Land Surface Dynamics

Machine Learning-based Property Valuation using Remote Sensing and Geospatial Data in Kigali, Rwanda

Felix Bachofer[1], Jonathan Bower[2] , Andreas Braun[3] , Paul Brimble[4] , Patrick McSharry[5,6,7]

Introduction

Valuation of land in developing countries can contribute to the achievement of various important global, regional and national development goals such as: sustainable urban development, promoting responsible investments and addressing conflicts often associated with large-scale land investments. Taxes on land and property can serve as an important tool to finance these development goals. The revenue is critical for municipalities and other local authorities to upgrade informal settlements, support the resettlement of displaced people and for infrastructure investment and development projects. In addition, it is vital for improving transparency in opaque land markets.

Context in Rwanda

Rwanda has historically had relative low land taxes but these were recently increased. A new property tax law will come into force on 1st January 2019. Along with Rwanda's good land governance, the new tax law may represent an opportunity to generate municipal revenue. Land value can be determined by a list of factors such as public investment in infrastructure, land use regulations, demographic change and economic development, private investment in land, and locational properties. Earth observation (EO) and geospatial analysis provide data, tools and methods that help to explain the relationship between land values and such explanatory variables.

The study is conducted with the full knowledge of Rwandan government institutions, most notably the Ministry of Finance and Economic Planning (MINECOFIN) and Rwanda Revenue Authority (RRA). We hope this study will provide a tool to help improve property tax valuations and hence, revenues. The main method of property valuation in Rwanda is currently through taxpayer self-assessments. Our model could help in two possible ways. The model could assess the validity of these self assessments and trigger additional on-the-ground valuations in cases in which under-valuation is suspected. Alternatively, the property values that the model generates could result in properties being assigned to bands with different tax levels.

Study Design

This study trials a property valuation methodology for Rwanda's capital, Kigali by applying machine learning techniques to parcel transactions data from 2013 to 2015 and remote sensing data from 2009 and 2015 on building footprints. The building information (building footprint and building type) is derived from aerial images (2009) and a Pleiades scene (2015). Additionally, variables describing the spatial interrelations of properties within the city are derived from a multi-source dataset and include distance-based values (distance to central business district, schools, public transport, etc.), accessibility, location parameters (topographic position) and neighborhood statistics (green area, urban structure type, density values, etc.). This approach can help to understand the key determinants of land and building values and be used to create a database with accurate estimates of property values for each parcel in Kigali.

Applications & Outlook

We used Maximum Relevance and Minimum Redundancy (MRMR) approach to predict property values in Kigali Province and got R^2 of 0.71.

Policy applications

- Evaluation of revenue potential of property tax (Ministry of Finance and Economic Planning)
- Evaluating self-assessments (Rwanda Revenue Authority)
- Potential basis for future CAMA (Rwanda Revenue Authority)
- Property price indices (National Bank of Rwanda)

Extensions

- Updated parcel data
- Panel regression analysis

A. Remote sensing analysis

Pléiades imagery for the densely built-up central part of the City of Kigali in order to derive building footprint and building archetypes. Having this information, aerial images of the study area of the years 2008/09 were used to create a change monitoring dataset on single building basis. A semi-automated approach was chosen with an object based image analysis with an expert-based revision.

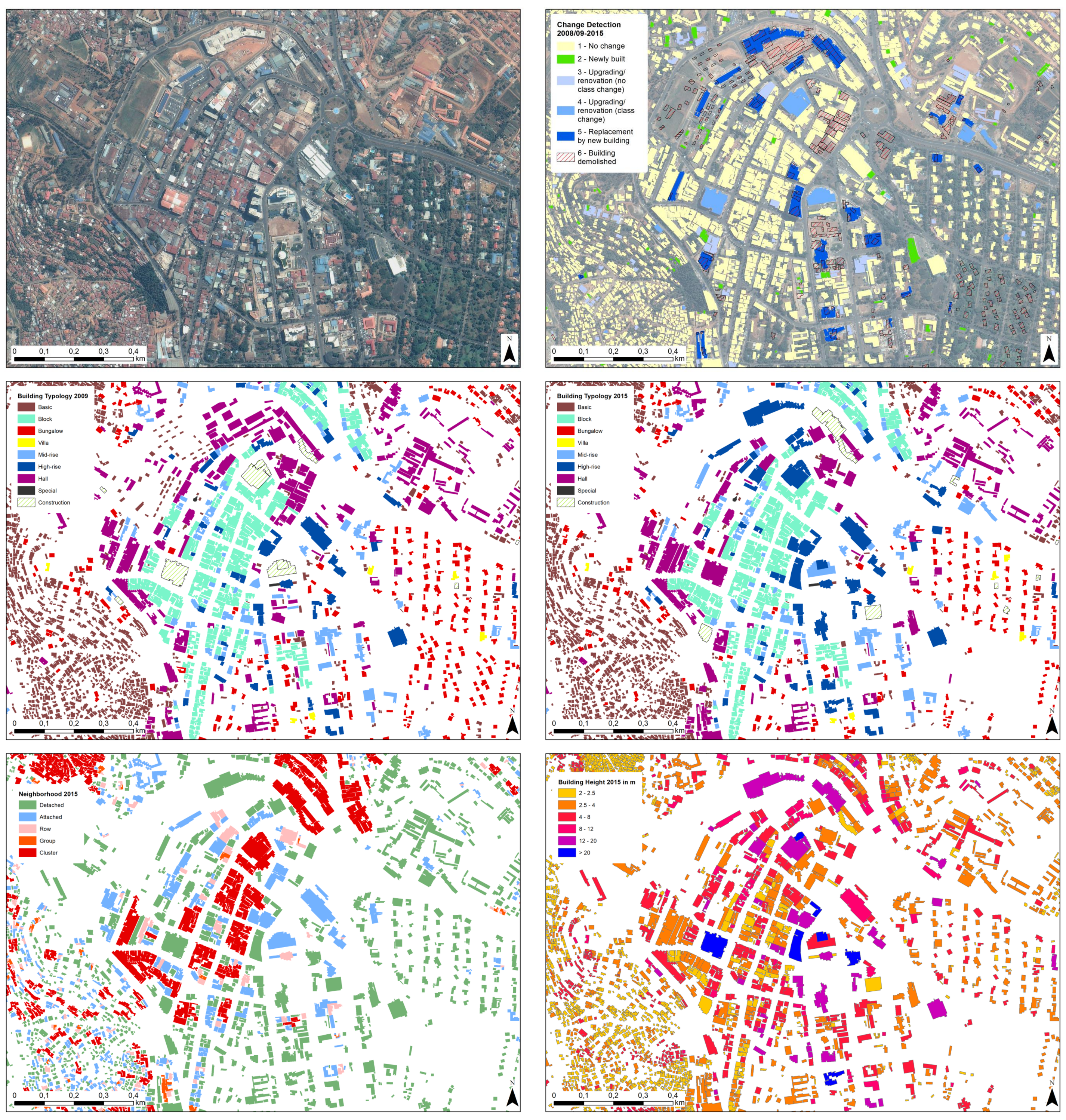

Fig. 1. Upper left: Pleiades scene 2015 CBD Kigali; upper right: change monitoring 2008/09 – 2015; center left: building types 2008/09; center right: building types 2015; lower left: adjacency building objects; lower right: building heights 2015.

Building Type	No. of buildings 2008/09	No. of buildings 2015	Change number of buildings	Change in %	Area Change in %
Basic	135,577	170,805	35,228	26.0%	20.5%
Block	1,609	1,587	-22	-1.4%	0.3%
Bungalow	20,313	27,610	7,297	35.9%	34.4%
Villa	1,542	4,781	3,239	210.1%	190.9%
Mid-rise	503	806	303	60.2%	62.2%
High-rise	211	371	160	75.8%	95.9%
Hall	3,396	4,306	910	26.8%	29.5%
Special	211	253	42	19.9%	27.2%
Construction	2,263	939	-1,324	-58.5%	-49.0%
SUM	165,625	211,458	45,833	27.7%	27.8%

Table 1. Result of remote sensing analysis.

B. Data

Dependent variable: Sales Data (Rwanda's Land Administration and Information System)

Year	2015	2016	2017	2018
Parcel Transactions (Kigali)	**10,246**	13,991	16,352	15,155

Independent variables: Parcel Data (available for all parcels for 2015)

- Structural Data (land area, slope, perimeter, location)
- Building Data (floor area, heights, typography, adjacency)
- Distances to Amenities (roads, bus stops, CBD etc.)
- Counts of Amenities (schools, hospitals, markets)
- Zoning (vegetation, agriculture, nature)
- Economic Factors (number of firms by size, employed labour)

C. Feature Selection

- Dependent variable: log of the sales value per m^2
- Independent variables: 511 parcel attributes (including log and squared)
- Estimation sample: 7,445 filtered sales from 2015
- Maximum Relevance and Minimum Redundancy (MRMR) machine learning approach to develop best OLS model
- We used 10-fold cross-validation to avoid overfitting (90% of data used for estimation and 10% used for testing)
- The model jointly predicts values for both vacant land (37% of estimation sample) and improved parcels (63% of estimation sample)

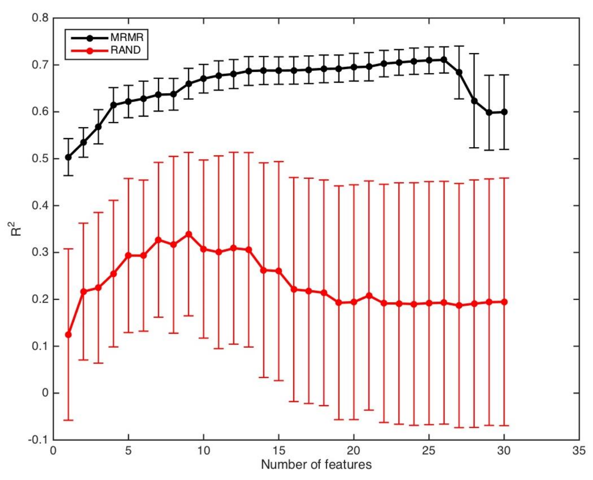

Fig. 2. Feature selection graph.

D. Benchmark model

Features (23):

- *Land* (3): Perimeter (Contribution to R^2 0.265)
- *Building* (4): Building Types, Floor Area & Volume (0.409)
- *Location - Distance* (7): Roads, Bus Stops, Routes (0.564)
- *Location - Zoning* (9): Vegetation, Agriculture, Forest, Single Family, Vacant (0.578)

Model Diagnostics (cross-validated):

- R^2: **0.708**
- MAE: **0.625**
- RMSE: **0.825**
- ±20%: 22.2%

E. Forecasting

Given static parcel data in 2015, the forecast accuracy of the model decreases over time.

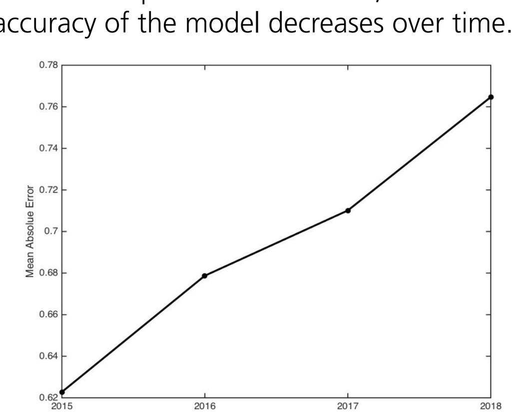

Fig. 3 Forecasting accuracy.

urban tep EO-derived data will be made available on the Urban Thematic Exploitation Platform https://urban-tep.eu/

Contact: **Dr. Felix Bachofer**, felix.bachofer@dlr.de

[1] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD-LAX), Oberpfaffenhofen, Germany
[2] International Growth Center (IGC), Kigali, Rwanda
[3] University of Tuebingen, Faculty of Science, Tuebingen, Germany
[4] Ministry of Finance and Economic Planning, Kigali, Rwanda
[5] Carnegie Mellon University - Africa, Kigali, Rwanda
[6] University of Rwanda, Rwanda, Kigali, Rwanda
[7] Oxford University, Oxford, United Kingdom

Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center