



**Carnegie
Mellon
University**



Using machine learning to value property in Kigali, Rwanda

Felix Bachofer, German Aerospace Center (DLR)

Jonathan Bower, International Growth Centre (IGC)

Andreas Braun, University of Tuebingen

Paul Brimble, Rwanda Ministry of Finance & Economic Planning (MINECOFIN)

Patrick McSharry, Carnegie Mellon University (CMU)

Key messages

- We compiled a rich dataset on property attributes, remotely
- We used Maximum Relevance and Minimum Redundancy (MRMR) approach to predict property values in Kigali Province and got R^2 of 0.71
- We welcome advice on model improvements
- We will follow through with Government of Rwanda on updates and applications

Rwanda's new property tax law necessitates a valuation model

- Rwanda is urbanising from a low base
- Kigali is the primary city, ten times larger than the second largest city
- New property tax law on 1st January 2019
- To help regulate tax collection, **we need a cost-effective, corruption-proof valuation method for the entire city**
- The model must be able to **extrapolate** values for all **367,667** parcels
- The values must also be decomposed into **land** and **building** components

Data

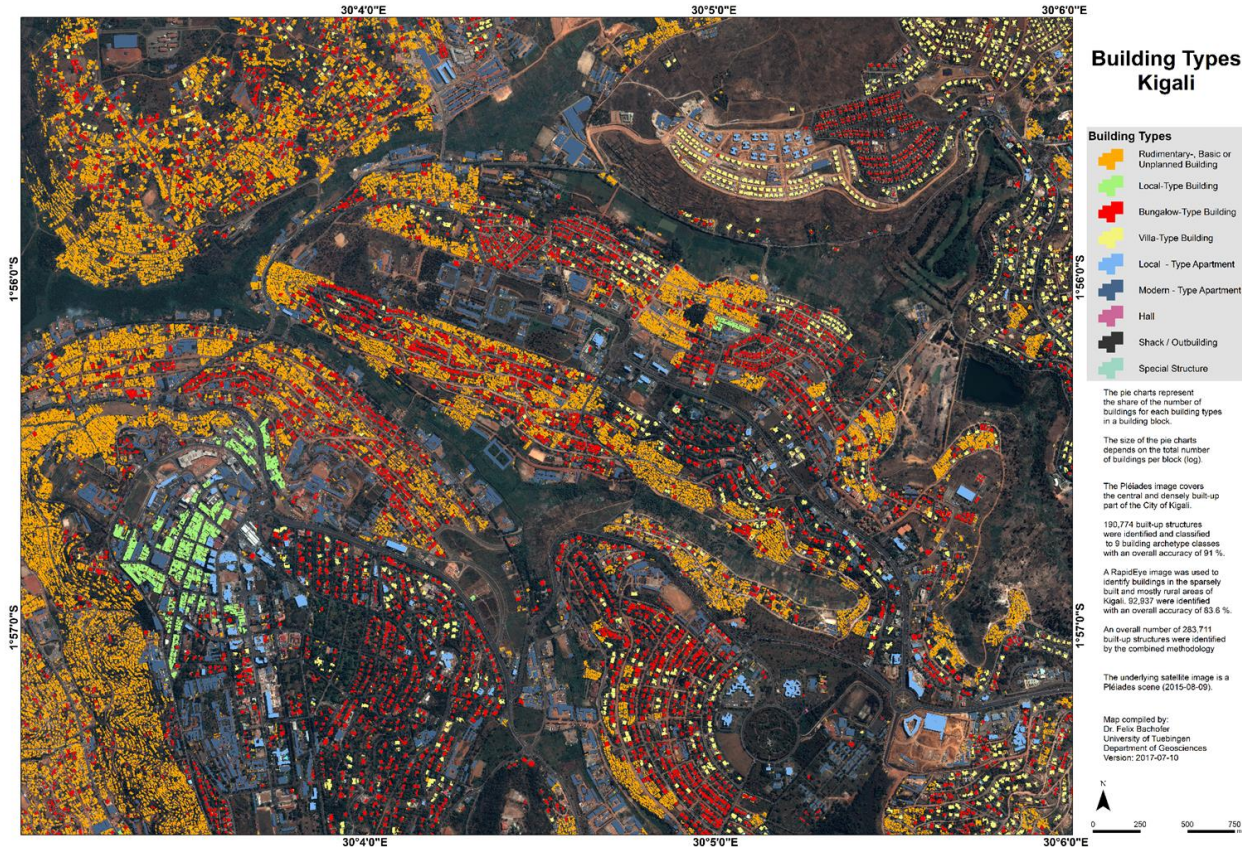
Dependent variable: Sales Data (Rwanda's Land Administration and Information System)

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)

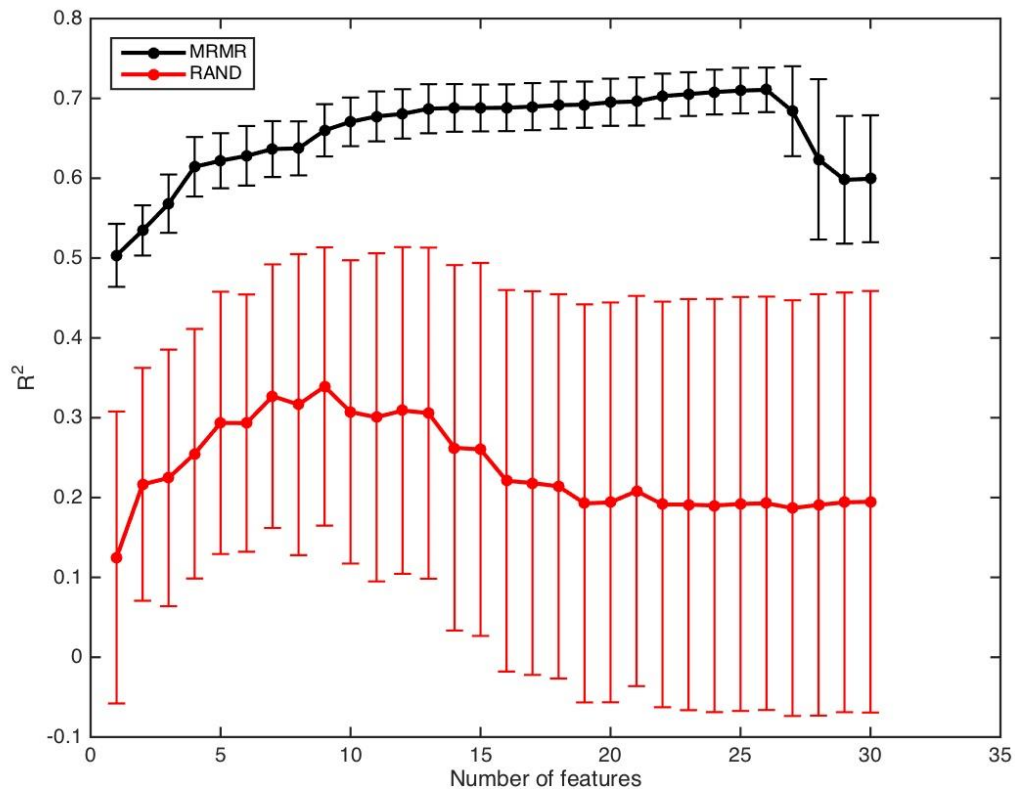
Remote sensing building dataset



Feature selection

- Dependent variable: **log of the sales value per m²**
- 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)

Feature selection visualisation



Benchmark model

Features (23):

- *Structural - Land* (3): Perimeter (Contribution to R^2 0.265)
- *Structural - Building* (4) : Building Count, Floor Area and Volume (0.409)
- *Location - Distance* (7): Roads, Bus Stops, Bus 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**
- $\pm 20\%$: 22.2%

Model adjustments

Location adjustments

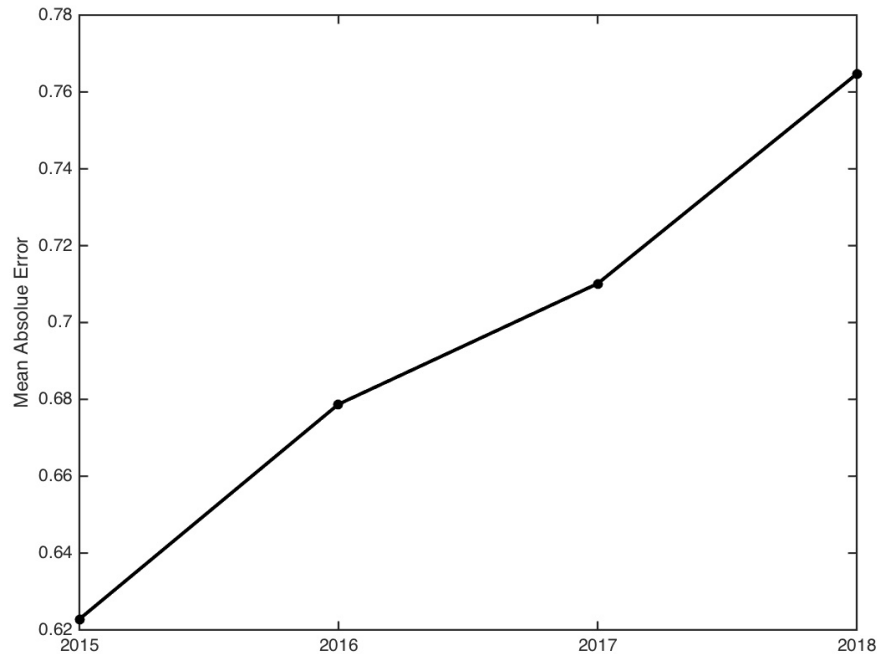
- Locational variables (dummies, coordinates and interactions)
- Spatial autoregressive models (spatial error model)

Housing submarkets

- Building typography (basic, bungalow, villa)
- Building adjacency (detached, attached, clustered)
- Urban structure type (block level classifications)

Forecasting

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



Splitting value into land and buildings

Land values

- The **cross-validated** land value accuracy is evaluated using vacant parcels.
 - R^2 : 0.491
 - MAE: 0.653
 - RMSE: 0.842
 - $\pm 20\%$: 19.6%

Building values

- The building values are determined as the residual of property values and land values, driven by the structural building variables
- Challenge: inability to directly test building value accuracy

Applications

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

Key messages

- We compiled a rich dataset on property attributes, remotely
- We used Maximum Relevance and Minimum Redundancy (MRMR) approach to predict property values in Kigali Province and got R^2 of 0.71
- We welcome advice on model improvements
- We will follow through with Government of Rwanda on updates and applications