PREDICTIVE LAND VALUE MODELLING USING A GEOSTATISTICAL APPROACH AND SPACE SYNTAX IN GUATEMALA CITY

MSc Andres Morales
Prof. Dr. A. Stein, Dr. J. Flacke, Prof. Dr. J. Zevenbergen
Guideline

- Background and objectives
- Case study and data
- Methods
- Results
- Conclusions
PROPERTY VALUE (S) = LAND VALUE (S) + CONSTRUCTION VALUE (S)

LAND VALUE $/m^2 (Y,S) = LOCATIONAL FACTORS (S) + PLOT CHARACTERISTICS + STRUCTURAL CHARACTERISTICS
Background

Housing affordability in Global South regions is constrained by land-markets dynamics. In these regions there is also a potential to unlock values to finance social-housing and infrastructure investments.

Spatial information on land-values is fundamental in supporting planners working towards sustainable futures. Individual appraisals for a city are costly and unpractical, requiring land-value predictive modelling.
Background

Geographic Access

- Proximity to Core Business District
- And a number of services
- Impedance: Aerial distance, network-based distance, time (transport mode)

1. Mapping urban accessibility using a geographic and geometric approach
2. Modelling residential land values using a geographic and geometric access, following Bourassa, et al. (2007)
3. Predictive land value modelling using a geostatistical approach and space syntax

Transport modelling tradition

Graph and Network theory applied to urban areas, Hillier, B. (2007)

- Closeness and Betweenness

Geometric

Standard OLS

Spatial Econometrics

Geostatistics
The objective was to predict a land-value ($\hat{y}$) map using geostatistics that considers geographic and geometric access metrics and relevant variables.
Methodology OLS: Backward stepwise regression

\[ Y = x^T \beta \]

\[ y(S_1) = \beta_0 + \sum_{k=1}^{p} \beta_p \cdot X(s_1) + \varepsilon(s_1), \mu = 0 \text{ and } \sigma^2 \]

\[ \hat{y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p} \hat{\beta}_k \cdot X(s_1) \]

- Characteristics of the parcel
- Characteristics of the neighbourhood
- Access metrics to positive or negative externalities
- Network centrality metrics

\[ \hat{y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p-1} \hat{\beta}_p \cdot X(s_1) \]

\[ \hat{y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p-1...n} \hat{\beta}_p \cdot X(s_1) \]

\[ \hat{y}(S_0) = \hat{\beta}_0 + \sum_{k=1}^{p-n} \hat{\beta}_p \cdot X(s_0) \]

\[ AIC = 2k - 2\log(\hat{L}) \]

if \( AIC_{k-1} < AIC_p \) → discharge covariable

if \( AIC_{k-1...n} \geq AIC_p \) → retain covariable
Methodology: Spatial Model, backward stepwise regression

\[ Y = x^T \beta \]

\[ y(S_1) = \beta_0 + \sum_{k=1}^{p} \beta_k \cdot X(s_1) + \sum_{i=1}^{n} \omega_i \cdot \varepsilon(s_1) + \varepsilon, \mu = 0 \text{ and } \sigma^2 \]

\[ \hat{y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p} \hat{\beta}_k \cdot X(s_1) + \sum_{i=1}^{n} \omega_i \cdot \varepsilon(s_1) \]

\[ \hat{\beta}_k \cdot X(S_1) \quad \text{Characteristics of the parcel} \]

\[ \hat{\beta}_k \cdot X(S_1) \quad \text{Characteristics of the neighbourhood} \]

\[ \hat{\beta}_k \cdot X(S_1) \quad \text{Access metrics to positive or negative externalities} \]

\[ \hat{\beta}_k \cdot X(S_1) \quad \text{Network centrality metrics} \]

\[ \hat{y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p-1} \hat{\beta}_p \cdot X(s_1) + \sum_{i=1}^{n} \omega_i \cdot \varepsilon(s_1) \]

\[ \hat{y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p-1} \ldots n \cdot \hat{\beta}_p \cdot X(s_1) + \sum_{i=1}^{n} \omega_i \cdot \varepsilon(s_1) \]

\[ \hat{y}(S_0) = \hat{\beta}_0 + \sum_{k=1}^{p-n} \hat{\beta}_p \cdot X(s_0) + \sum_{i=1}^{n} \omega_i \cdot \varepsilon(s_1) \]

\[ \sigma^2(S) = \sigma^2 \{ \hat{y}(S) \} + \sigma^2 \{ \hat{\varepsilon}(S) \} \]

\[ \text{Variance} \]

\[ \text{Predictor} \]

\[ AIC = 2k - 2 \log(\hat{L}) \]

if \( AIC_{k-1} < AIC_p \) → discharge covariable

if \( AIC_{k-1\ldots n} \geq AIC_p \) → retain covariable
Methodology: Spatial model

\[ \gamma(h) = \frac{1}{2ng} \sum_{i=1}^{n} \sum_{j=i}^{n} \delta_{ij} (y_i - y_j)^2 \]

Empirical semivariogram of residuals

\[ \gamma(h) = \begin{cases} 0, & \text{for } |h| = 0 \\ C_0 + C_1 \left[ 1 - e^{\left(\frac{h}{r}\right)} \right], & \text{for } |h| > 0 \end{cases} \]

Exponential semivariogram function
Methodology: Assessment and cross validation

Goodness of fit over training data

\[ R_{Adj}^2 = R^2 - (1 - R^2) \frac{p}{n - p - 1} \]

Prediction over test data

- Mean Error
- Mean Squared Errors
- Root Mean Squared Error

\[ R_{Adj}^2 = R^2 - (1 - R^2) \frac{p}{n - p - 1} \]
Case study and data

- 1200 Observations of real estate appraisals (2008-2014)
- Hexagonal tessellation (inscribed radius of 150m).

Land value in Q/m²

Min= 325
Max= 7035
Mean= 1655
St. Dev.=920

Log transformed land value

Min= 5.79
Max= 8.86
Mean= 7.28
St. Dev.=0.51
Case study and data

- Centroids were used for accessibility analyses
- Then, interpolated using ordinary kriging (radius 300m)
Case study and data

Geometric Access (Space Syntax)

Geographic Access

Externalities

Geometric via Geographic

Network Closeness reachable via mobility modes

Network Betweenness

Network Closeness

Time based per mobility mode

Submarkets

Parcel level

Neighbourhood characteristics

Parcel characteristics

Prediction Grid

Train & Test

Integration_grav
Global Betweenness
Betweenness_75
Betweenness_50
Betweenness_25
Betweenness_15
Betweenness_08
Global Closeness
Closeness_75
Closeness_50
Closeness_25
Closeness_15
Closeness_08
Markets
Clinics
Neighborhood Groceries
Banks_Restaurants
Schools
Parks
Culture
Hospitals
University
XL_Sport Facilities
XL_Groceries store
Shopping Malls
Jobs_Gravity
Proximity to bus stops
Distance to main Road
Core Business District
Flat_Segments
Condos_Segments
East
West
% of private vehicle users
Density of new condos
Density of new buildings
Socio-economic strata
Population Density
Potential for new building
Block corner
Regular Geometry
Construction area m²
Plot surface m²
Case study and data

Geometric Access
(Space Syntax)

Geographic Access

Neighbourhood characteristics

Parcel characteristics

Network Closeness reachable via mobility modes

Network Betweenness

Network Closeness

Time based per mobility mode

Submarkets

Parcel level

Externalities

Internal

Integration _grav
Global Betweenness
Betweenness_75
Betweenness_50
Betweenness_25
Betweenness_15
Betweenness_08

Global Closeness
Closeness_75
Closeness_50
Closeness_25
Closeness_15
Closeness_08

Markets
Clins
Neighborhood Groceries
Banks_Restaurants
Schools
Parks
Culture
Hospitals
University
XL_Sport Facilities
XL_Groceries store
Shopping Malls
Jobs_Gravity
Proximity to bus stops
Distance to main Road

Core Business District
Flat_Segments
Condos_Segments
East
West
% of private vehicle users
Density of new condos
Density of new buildings
Socio-economic strata
Population Density

Potential for new building
Block corner
Regular Geometry
Construction area m2
Plot surface m2
Prediction Grid
Train & Test
Case study and data

Geometric Access (Space Syntax)
- Geometric via Geographic
- Network Betweenness
  - Global Betweenness
  - Betweenness_75
  - Betweenness_50
  - Betweenness_25
  - Betweenness_15
  - Betweenness_08
- Global Closeness
  - Closeness_75
  - Closeness_50
  - Closeness_25
  - Closeness_15
  - Closeness_08
- Markets
- Clinics
- Neighborhood Groceries
- Banks_Restaurants
- Schools
- Parks
- Culture
- Hospitals
- University
- XL_Sport Facilities
- XL_Groceries store
- Shopping Malls
- Jobs_Gravity
- Proximity to bus stops
- Distance to main Road

Geographic Access
- Time based per mobility mode
- Core Business District
- Flat_Segments
- Condos_Segments
- East
- West
- % of private vehicle users
- Density of new condos
- Density of new buildings
- Socio-economic strata
- Population Density
- Prediction Grid
- Train & Test

Externalities
- Internal

Parcel characteristics
- Submarkets

Parcel level

Network Closeness
- Network Closeness reachable via mobility modes

Neighbourhood characteristics
Case study and data

Geometric via Geographic

Geometric Access (Space Syntax)

Geographic Access

Neighbourhood characteristics

Parcel characteristics

Externalities

Network Closeness reachable via mobility modes

Network Betweenness

Network Closeness

Submarkets

Parcel level

\[ Clos_\theta(x) = \left( \sum_{i=1}^{n} D_\theta(x, i) \right)^{-1} \]

Prediction Grid

Train & Test

\[ \text{Externalities} \]

\[ \text{Internal} \]

\[ \mathcal{C}l_{\theta} \]

\[ x = i \]

\[ n \]

\[ D_{\theta}(x, i) \]

\[ -1 \]
LAND VALUE MODELLING IN GUATEMALA CITY

Case study and data

Geometric via Geographic Access

Geometric Access (Space Syntax)

Geographic Access

Neighbourhood characteristics

Parcel characteristics

Parcel level

Submarkets

Network Betweenness

Network Closeness reachable via mobility modes

Network Closeness

Externalities

Internal

Integration_grav

Global Betweenness

Betweenness_75

Betweenness_50

Betweenness_25

Betweenness_15

Betweenness_08

Global Closeness

Closeness_75

Closeness_50

Closeness_25

Closeness_15

Closeness_08

Markets

Clinics

Neighborhood Groceries

Core Business District

Flat_Segments

Condo_Segments

East

West

% of private vehicle users

Density of new condos

Density of new buildings

Socio-economic strata

Population Density

Potential for new building

Block corner

Regular Geometry

Construction area m²

Plot surface area m²

Train & Test

Prediction Grid

\[
Bet_\theta(x) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma(i, x, j)}{(n - 1)(n - 2)/2}
\]

\(\sigma\) is the number of shortest paths between the nodes i and x, and j and x.
Case study and data

Geometric via Geographic

Geometric Access (Space Syntax)

Global Betweenness
Betweenness_75
Betweenness_50
Betweenness_25
Betweenness_15
Betweenness_08
Global Closeness
Closeness_75
Closeness_50

Network Closeness reachable via mobility modes

Network Betweenness

\[
\sum \left( \frac{n - 1}{\sum_{i=1}^{n} D_\theta (x, i)} \right) \text{hex j} \cdot \alpha \exp(-\beta \cdot t_{\text{hex i-hex j}})
\]

Geographic Access

Neighbourhood characteristics

Parcel characteristics

Parcel level

Externalities

Parks
Culture
Hospitals
University
XL_Sport Facilities
XL_Groceries store
Shopping Malls
Jobs_Gravity
Proximity to bus stops
Distance to main Road

Core Business District

Flat_Segments
Conds_Segments
East
West
% of private vehicle users
Density of new condos
Density of new buildings
Socio-economic strata
Population Density
Potential for new building
Block corner
Regular Geometry
Construction area m²
Plot surface area m²

Submarkets

Time based per mobility mode

Prediction Grid

Train & Test
Variable selection process

- Backward stepwise regression using a spatial model
- Exponential semivariance function MLE
\( \hat{Y}(S_1) = \hat{\beta}_0 + \sum_{k=1}^{p-1} \hat{\beta}_k \cdot X(s_1) + \sum_{i=1}^{n} \omega_i \cdot \epsilon(s_1) \)

- Ranking of the models: AIC statistic
Variable selection process

Discharged variables
Results and Model Comparison

Standard OLS step-wise regression
- 30 variables
- Adj. $R^2=0.73$
  - Mean error = -0.03
  - Mean squared error = 0.34
  - Root mean squared error = 0.58
  - Variability explained = 63%

Plug the variables selected to a Regression Kriging formulation
- 30 variables
- Adj. $R^2=0.79$
  - Mean error = -0.003
  - Mean squared error = 0.004
  - Root mean squared error = 0.06
  - Variability explained = 65%

Spatialized step-wise regression kriging
- 21 variables
- Adj. $R^2=0.81$
  - Mean error = -0.003
  - Mean squared error = 0.002
  - Root mean squared error = 0.05
  - Variability explained = 70%
Inspection of model outputs

Mapping of residuals (train)

Residuals

-0.89

Predictions

0.91
Further data collection
Land value map prediction

Legend
- Test obs.
  - 333 - 950
  - 951 - 1,666
  - 1,687 - 2,386
  - 2,397 - 3,414
  - 3,415 - 4,819
  - 4,820 - 6,224
  - 6,225 - 7,496
  - 7,497 - 8,965
- Kriged centroids

Value Q/m²
- 227 - 950
- 951 - 1,666
- 1,687 - 2,386
- 2,397 - 3,414
- 3,415 - 4,819
- 4,820 - 6,224
- 6,225 - 7,496
- 7,497 - 8,965

Kilometers
Conclusions

• Based on our cross validation statistics, we were able to predict a land value map for Guatemala City using a geostatistical approach and Space Syntax.

• The spatialized variable selection process by means of a spatialized modelling approach yielded an improved parsimonious model and lead to a more accurate prediction at non observed locations (70%).

• The approach provides evidence that validates how Space Syntax network centrality metrics (geometric access) adds spatialized model information to explain residential land values in Guatemala City.

• Further work:
  • On the ranking of the variables
  • On the improvement of prediction at (S0)