What data tells us: Applications and Value Potentials of 3D Spatial Data

Zur Verabschiedung Prof. Dr. Dr. h.c. Günter Müller

Prof. Dr. Dirk Neumann
Chair for Information Systems Research
University of Freiburg, Germany

Freiburg 20.4.2018
Within the last decade, several complex data sources (e.g. text, images, videos) have been incorporated into decision-making.

Progress in airborne laser scanning (e.g. via drones) and photogrammetry has made 3D building geometries available in map services such as GoogleMaps.

- So far, mainly visualization purposes
- Little usage for managerial decision-making with analytics

Real buildings\(^2\) and corresponding 3D models

1. What is the value of 3D spatial data for decision-making?
2. How can 3D data be incorporated into analytical modeling?
3D spatial data consists of an **object representation** and a **representation of the environment**. At least one of the representations is in 3D:

Showcase 1: What do building geometries tell us about real estate valuations?

Showcase 2: What does the 3D environment tell us about the adoption of solar photovoltaic systems?
1. **Goal:** Predict future *rent prices of flats* based on observed prices of the past

2. **Study design:**
   - **Berlin** in order to obtain a sufficiently large sample of historic rent prices
   - Object representation: *Simple level-of-detail 1* (LoD1) to demonstrate added value even for low-resolution 3D data
     - Environment representation: *Simple 3D environment* (location and elevation)
       - No interdependence between nearby objects
Showcase 1: What do building geometries tell us about real estate valuations?

3. Variables and descriptive statistics:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Origin</th>
<th>Mean</th>
<th>Median</th>
<th>Stdev</th>
<th>Min</th>
<th>Max</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>height [m]</td>
<td>3D buildings</td>
<td>21.74</td>
<td>22.49</td>
<td>7.66</td>
<td>1.92</td>
<td>67.01</td>
<td>Building height</td>
</tr>
<tr>
<td>footprint [m²]</td>
<td>3D buildings</td>
<td>3369.65</td>
<td>1823.58</td>
<td>4659.01</td>
<td>23.37</td>
<td>33322.40</td>
<td>Buildings footprint</td>
</tr>
<tr>
<td>elevation [m]</td>
<td>3D environment</td>
<td>39.39</td>
<td>35.61</td>
<td>7.39</td>
<td>30.96</td>
<td>61.54</td>
<td>Elevation above sea level</td>
</tr>
<tr>
<td>district</td>
<td>3D environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Urban district</td>
</tr>
<tr>
<td>square_meters</td>
<td>flat offerings</td>
<td>83</td>
<td>74</td>
<td>39</td>
<td>20</td>
<td>482</td>
<td>Size of the flat</td>
</tr>
<tr>
<td>rooms</td>
<td>flat offerings</td>
<td>2.52</td>
<td>2</td>
<td>1.01</td>
<td>1</td>
<td>7</td>
<td>Number of rooms</td>
</tr>
<tr>
<td>rent [EUR/month]</td>
<td>flat offerings</td>
<td>1068.31</td>
<td>924</td>
<td>599.08</td>
<td>220</td>
<td>5000</td>
<td>Outcome variable</td>
</tr>
</tbody>
</table>

Observations: 984
Showcase 1: What do building geometries tell us about real estate valuations?

4. Choice of potential prediction methods

- **Conventional methods**: Random forest, support vector machine, OLS regression, Elastic net regression

- **Spatial error model**:

  \[ Rent_i \sim \text{Normal}(\mu_i, \varepsilon) \]

  \[ \mu_i = \alpha + \beta_X X_i + \omega(s) \]

  \[ \omega(s) \sim \text{MVNormal}(0, K(\theta)) \]

  \[ C(s, t) = \sigma^2 \rho(s, t; \phi) \]

  \[ \rho(s, t; \phi) = \exp(-\phi \|s - t\|) \]

  \( \leftarrow \) Spatially correlated errors

  \( \leftarrow \) Correlation depends on distance between observations s & t

  \( \leftarrow \) Exponential decay of correlation
Showcase 1: What do building geometries tell us about real estate valuations?

5. Results (RMSE): Rent $\sim X$

<table>
<thead>
<tr>
<th>X</th>
<th>rooms</th>
<th>square_meters</th>
<th>district</th>
<th>footprint</th>
<th>elevation</th>
<th>height</th>
<th>SEM</th>
<th>RF</th>
<th>SVM</th>
<th>OLS</th>
<th>OLS &amp; Elastic Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>645.15</td>
<td>645.15</td>
<td>645.15</td>
<td>645.15</td>
<td>645.15</td>
</tr>
<tr>
<td>1D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>261.15</td>
<td>259.07</td>
<td>239.20*</td>
<td>261.80</td>
<td>263.28</td>
</tr>
<tr>
<td>2D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>260.35</td>
<td>248.01</td>
<td>239.43*</td>
<td>259.71</td>
<td>260.78</td>
</tr>
<tr>
<td>3D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>259.26</td>
<td>222.47</td>
<td>220.24*</td>
<td>254.52</td>
<td>258.28</td>
</tr>
</tbody>
</table>

- **3D model outperforms** more naïve 1D and 2D models
- **Conventional prediction methods** perform better than spatial model

Conclusion: 3D building data can improve mass valuations of real estate, e.g. in web portals.
Showcase 2: What does the 3D environment tell us about the adoption of photovoltaic systems?

Background: roof-mounted photovoltaic (PV) systems

- Provide electricity
- Increasing markets worldwide
- Electricity yield depends on size and roof characteristics
  - Orientation
  - Shading

A roof-mounted PV system

1 Adrian Pingstone, Public Domain, https://commons.wikimedia.org/w/index.php?curid=33115816
Showcase 2: What does the 3D environment tell us about the adoption of photovoltaic systems?

1. **Goal:** Create application to assist **PV system vendors** in identifying target customers

2. **Study design:**
   - 2 municipalities in Southern Germany comprising **suburban**, as well as **rural and hilly** areas
   - Learn adoption patterns from several municipal districts and **predict adoption of PV systems** in an unseen district
   - **Object representation:** LoD2  
     3D Environment + interactions
Showcase 2: What does the 3D environment tell us about the adoption of photovoltaic systems?

3. Data preparation and variable choice

Irradiance Simulation:
- **Input**: 3D spatial data, historic weather and sun position
- **Simulation**: Solar irradiance (and shading), temperature losses
- **Output**: Potential yearly electricity production of PV systems for each roof
Showcase 2: What does the 3D environment tell us about the adoption of photovoltaic systems?

4. Results (AUC): PV ~ X

| X | municipality | neighborhood | building_function | building_density | roof_surface | roof_type | roof_orientation | roof_inclination | PV_power
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEM</td>
<td>RF</td>
<td>RF ROSE</td>
<td>SVM</td>
<td>Logit</td>
<td>Logit &amp; Elastic Net</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.5513</td>
</tr>
<tr>
<td>2D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.7091*</td>
</tr>
<tr>
<td>3D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- Models based on **3D data outperform** more naïve 1D and 2D models
- **Spatial prediction method** outperforms conventional methods
Showcase 2: What does the 3D environment tell us about the adoption of photovoltaic systems?

5. Model use

![Graph showing the comparison of detected PV systems and targeted homeowners according to P (PV = yes) for 1D, 2D, 3D models, and a random guess.]

Conclusion: 3D building data can improve targeted marketing of PV systems.
Insights: What else can 3D spatial data tell us?

Utilizing 3D spatial data for decision making can tell us much more:

- **Marketing:**
  - Population estimations to improve spatial demand analyses (e.g. for retail store locations, transportation demand)
  - Visibility analyses for pricing of facade advertisements

- **Security and Safety:**
  - Visibility analyses for crime prediction
  - Detection of sniper hazards
  - Emergency response planning (flooding, terror attacks)

- **Sustainability and Health:**
  - Simulate noise and pollution propagation to explain population health
  - Estimate heating, cooling and electricity demand
## Insights: Key modeling questions

<table>
<thead>
<tr>
<th>Goal definition</th>
<th>Study design &amp; data collection</th>
<th>Data preparation</th>
<th>Exploratory data analysis</th>
<th>Variable choice</th>
<th>Choice of potential methods</th>
<th>Method and model selection</th>
<th>Model use and reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the relevant real-world relations?</td>
<td>What is the adequate model representation?</td>
<td>Disjoint hold-out sample?</td>
<td>Data quality?</td>
<td>How to extract variables from the environment?</td>
<td>What is the value of spatial prediction methods for 3D data?</td>
<td>How does 3D data impact model training and evaluation?</td>
<td></td>
</tr>
<tr>
<td>Location and extent of study area?</td>
<td>Level of detail of 3D data?</td>
<td>How to handle missing values?</td>
<td>Is the data aligned?</td>
<td>Do spatial patterns signal dependence?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- D. Neumann | What data tells us | 20.4.2018
Insights: Modeling guidelines

1. What are the **relevant real-world relations**?
   → Rely on **domain knowledge**, consider socio-economic environment

2. What is the adequate **model representation**?
   → Evaluate **trade-off** between modeling complexity and information loss

3. How to **extract variables** from the environment?
   → Summarize or simulate

4. What is the **value of spatial prediction methods** for 3D data?
   → **No silver bullet**. Implicit measurement of dependence is an advantage of spatial methods. Assumption of spatially-smoothed effects is a disadvantage

5. How does 3D data impact model training and evaluation?
   → **Spatial dependence** and **data imbalance** are challenges that require special data resampling methods