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

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

Albert-Ludwigs-Universität Freiburg

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

Freiburg 20.4.2018

UNI FREIBURG

Motivation

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

[2] http://www.pfarrverband-miesbach.de/seite/182439/unsere-kirchen.html

Research Questions

1. What is the **value** of 3D spatial data **for decision-making**?
2. How can 3D data be incorporated into **analytical modeling**?

Approach

3D spatial data consists of an **object representation** and a **representation of the environment**. At least one of the representations is in 3D:

3D spatial data

Object representation

1D | 2D | 3D

LoD1 LoD2 LoD3 LoD4

Environment representation

None | 2D | 3D

Location = (latitude, longitude)

Location = (latitude, longitude, elevation)

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

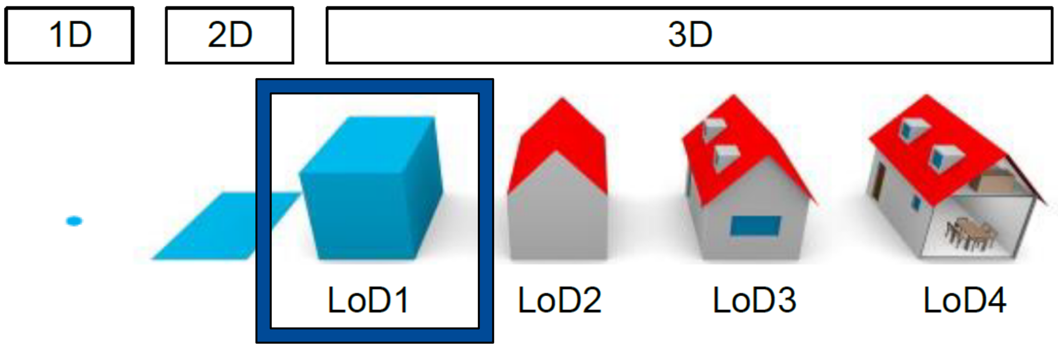
Showcase 2: What does the 3D environment tell us about the adoption of solar photovoltaic systems?

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

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:

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

Observations: 984

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

UNI FREIBURG

4. Choice of potential prediction methods

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

$$Rent_i \sim Normal(\mu_i, \varepsilon)$$

$$\mu_i = \alpha + \beta_X X_i + \omega(s)$$

$$\omega(s) \sim MVNormal(0, K(\theta))$$ ← Spatially correlated errors

$$C(s,t) = \sigma^2 \rho(s,t;\phi)$$ ← Correlation depends on distance between observations s & t

$$\rho(s,t;\phi) = \exp(-\phi \|s-t\|).$$ ← Exponential decay of correlation

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

UNI FREIBURG

5. Results (RMSE): Rent ~ X

X	*rooms*	*square_meters*	*district*	*footprint*	*elevation*	*height*	SEM	RF	SVM	OLS	OLS & Elastic Net
Intercept							645.15	645.15	645.15	645.15	645.15
1D	✓	✓	✓				261.15	259.07	239.20*	261.80	263.28
2D	✓	✓	✓	✓			260.35	248.01	239.43*	259.71	260.78
3D	✓	✓	✓	✓	✓	✓	**259.26**	**222.47**	**220.24***	**254.52**	**258.28**

* best performing method

→ **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

A roof-mounted PV system[1]

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

[1] Adrian Pingstone, Public Domain, https://commons.wikimedia.org/w/index.php?curid=33115816

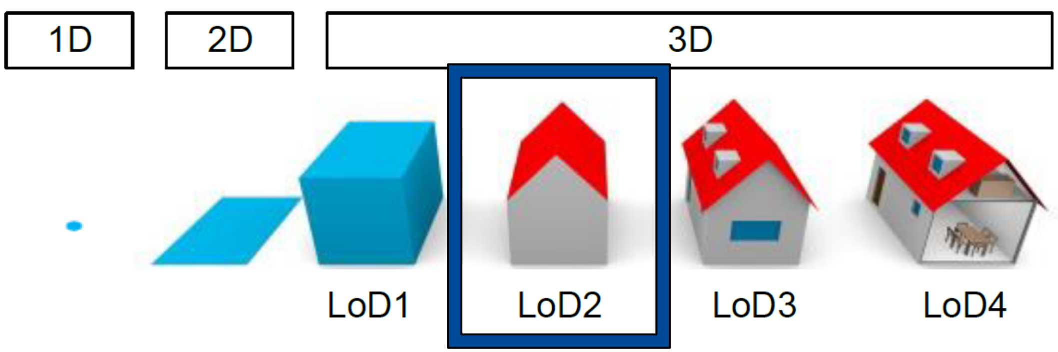
Showcase 2: What does the 3D environment tell us about the adoption of photovoltaic systems?

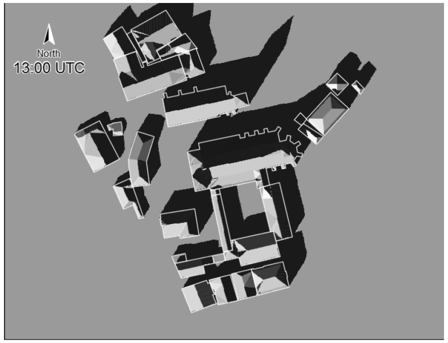
1. Goal: Create application to assist **PV system vendors** in **identifying target customers**

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

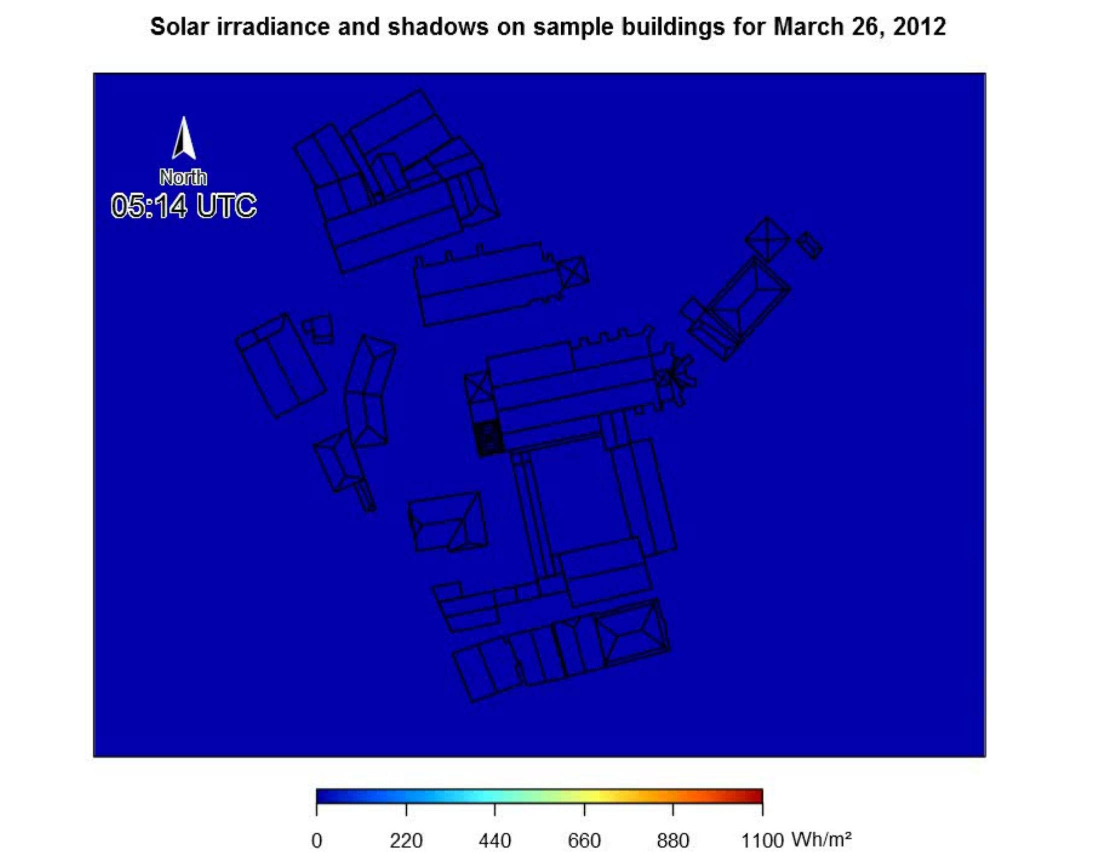
3D Environment + interactions





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?

UNI FREIBURG

4. Results (AUC): PV ~ X

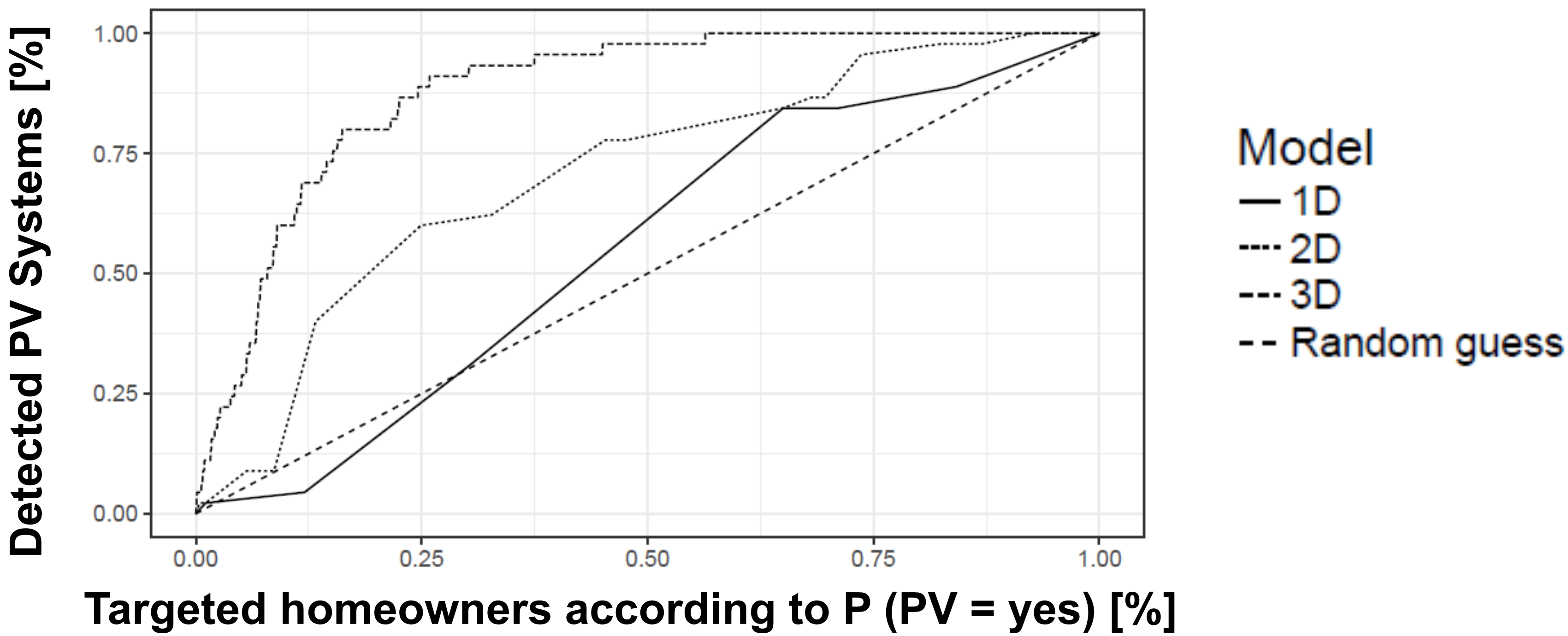
X	*municipality*	*neighborhood*	*building_function*	*building_density*	*roof_surface*	*roof_type*	*roof_orientation*	*roof_inclination*	*PV_power*	SEM	RF	RF ROSE	SVM	Logit	Logit & Elastic Net
Intercept										0.5	0.5	0.5	0.5	0.5	0.5
1D	✓	✓	✓							0.5513	0.5	0.6516	0.5	0.6691*	0.6379
2D	✓	✓	✓	✓						0.7091*	0.5	0.6977	0.5	0.7074	0.7065
3D	✓	✓	✓	✓	✓	✓	✓	✓	✓	**0.8902***	**0.8654**	**0.7872**	**0.7355**	**0.7893**	**0.8591**

→ 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?

UNI FREIBURG

5. Model use



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

Backup

Insights: Key modeling questions

Goal definition	Study design & data collection		Data preparation	Exploratory data analysis	Variable choice	Choice of potential methods	Method and model selection	Model use and reporting
Prediction within or accross areas?	**What are the relevant real-world relations?**	**What is the adequate model re-presentation?**	Disjoint hold-out sample?	Data quality?	**How to extract variables from the environment?**	**What is the value of spatial prediction methods for 3D data?**	**How does 3D data impact model training and evaluation?**	
	Location and extent of study area?	Level of detail of 3D data?	How to handle missing values?	Is the data aligned?				
				Do spatial patterns signal dependence?				

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