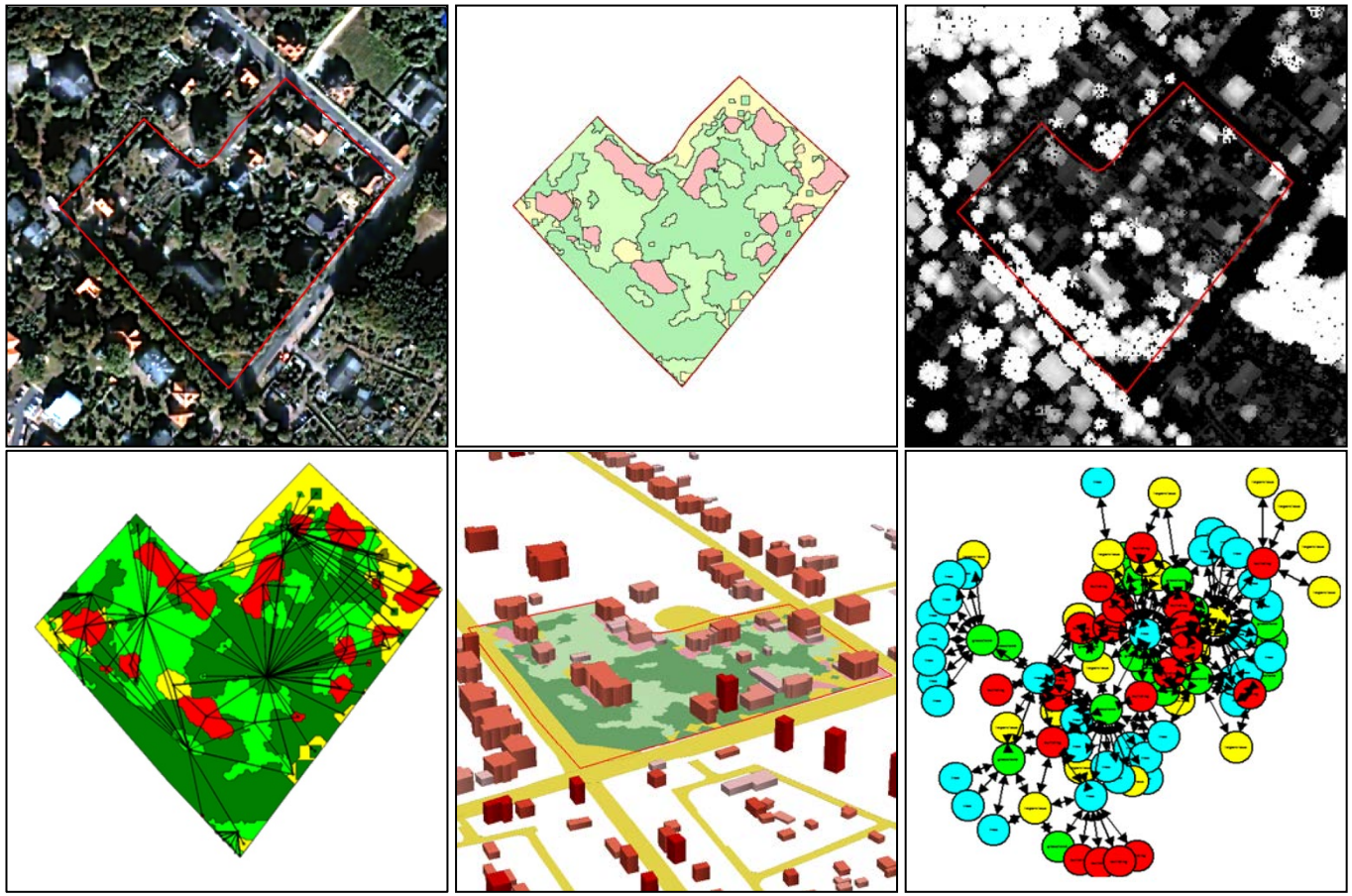
Vom Satellitenbild zu Stadtstrukturtypen: Wie Graphen die Flächennutzung charakterisieren



Übersicht

1. Datenbeschreibung

2. Methoden
   a) Urbane Strukturtypen
   b) Polygon Nachbarschaft und Graphen
   c) Graphen-Indikatoren
   d) Random Forest Analyse

3. Ergebnisse
   a) Graphen-Indikatoren
   b) Random Forest Analyse
   c) Validierung

4. Zusammenfassung

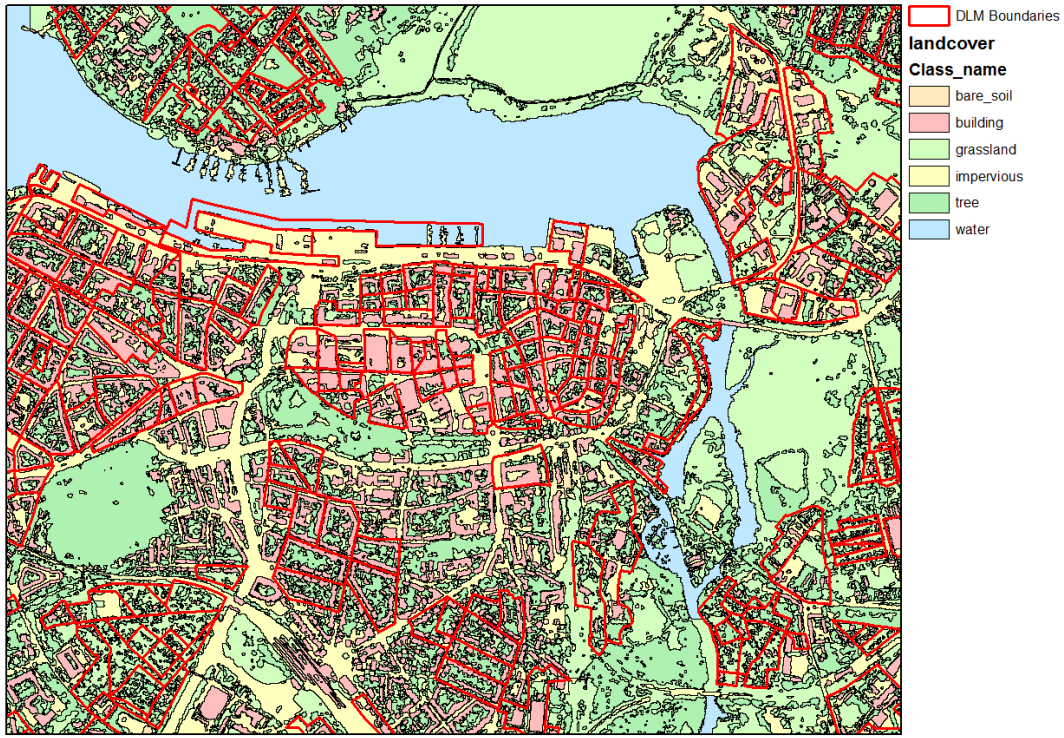
1. Datenbeschreibung

Quickbird Satellitenbild
(© DigitalGlobe, Inc., 2011)

September 2009

Pan 0,6m
MS 2,4m

LiDAR
Normalisiertes Höhenmodell



OBIA-Landbedeckungs-Ableitung beschrieben in: Berger, C.; Voltersen, M.; Hese, S.; Walde, I.; Schmullius, C. (accepted)

2. Methoden

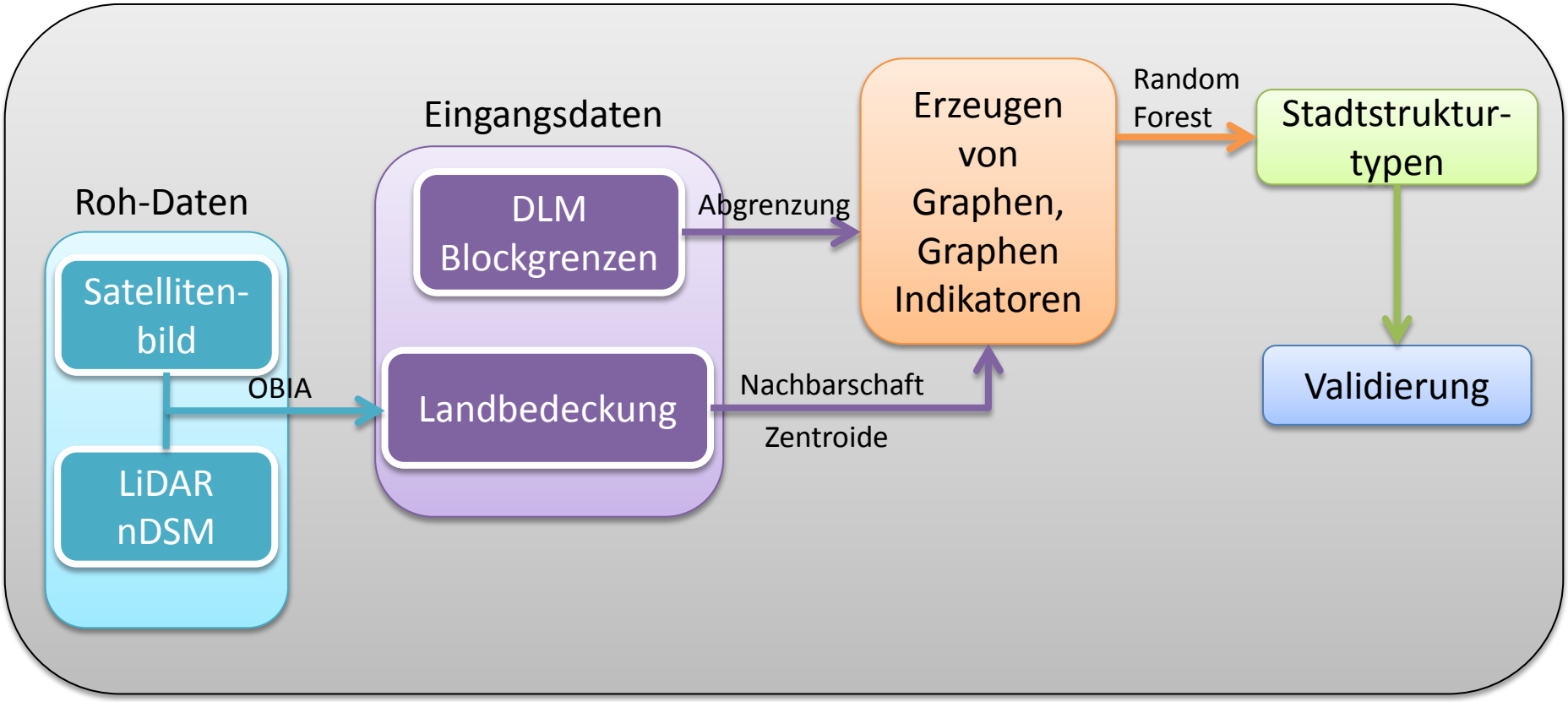
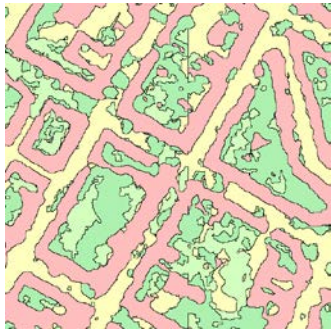


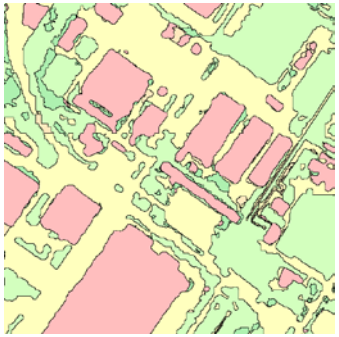
Abb.: Workflow

2. Methoden -Urbane Strukturtypen

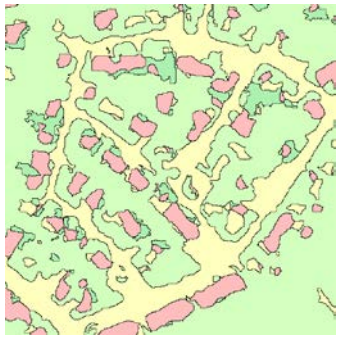
Stadtzentrum:



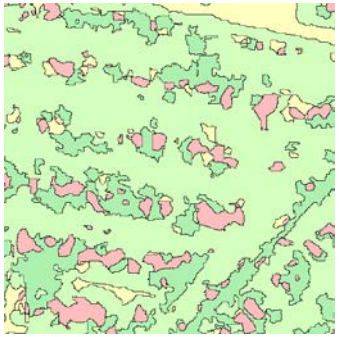
Industriegebiet



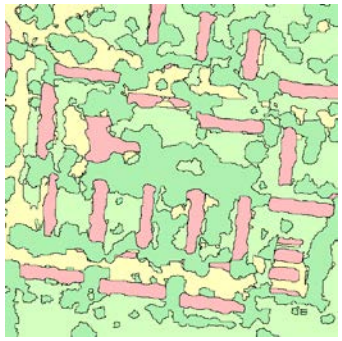
Einfamilien haus - Siedlung:



Gartenanlage

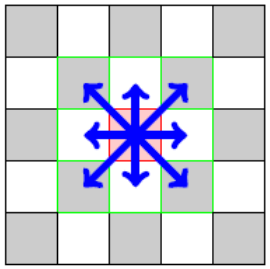


Block Gebäude



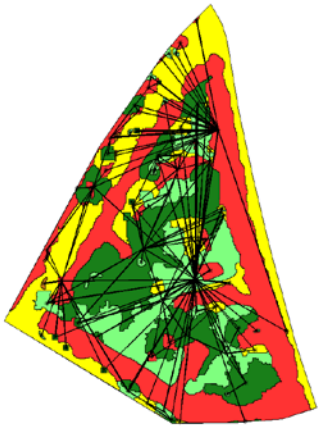
2. Methoden - Polygon Nachbarschaft und Graphen

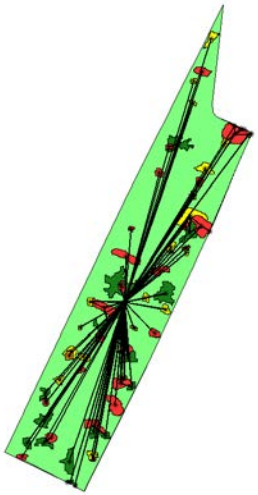
Queen's Nachbarschaft



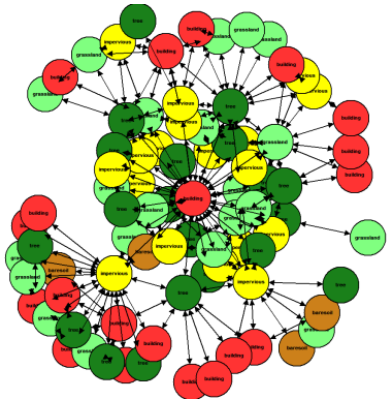
Stadtzentrum | Gartenanlage

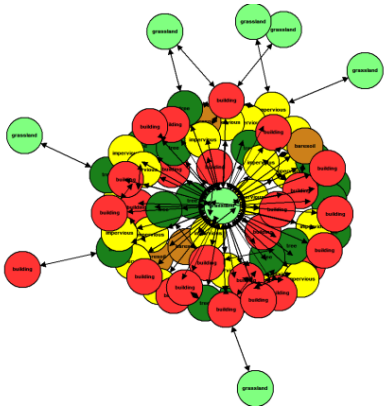
mit geographischer Referenz





ohne geographischer Referenz





building grassland tree impervious bare soil

2. Methoden – Graphen-Indikatoren

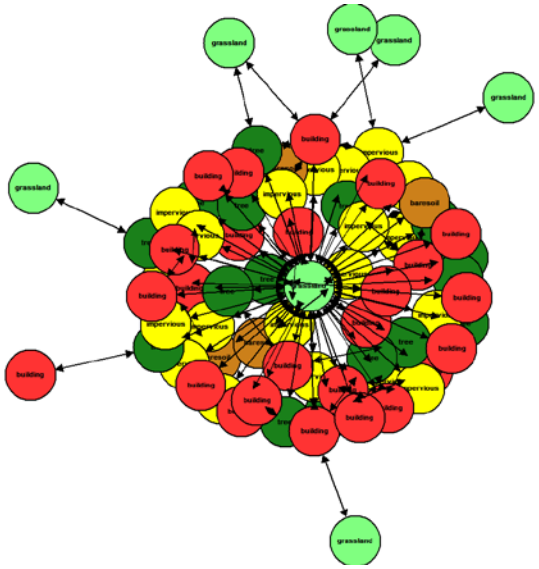
- Konnektivität (2)
- Zentralität (3)
- Adjazenz-Event (13)
- Zusätzliche Maße (4)

2. Methoden – Graphen-Indikatoren

- Konnektivitäts-Indikatoren

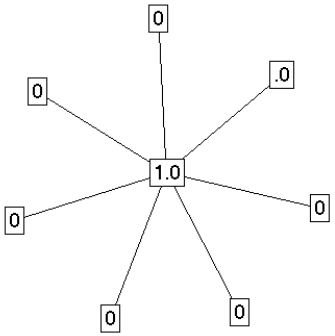
  Beta Index= Kanten/Knoten (siehe Walde, I.; Hese, S.; Berger, C.; Schmullius, C. (accepted))

- Zentralitäts-Indikatoren
  - Degree Centrality (Borgatti and Halgin(2011), Barr and Barnsley (1998), Gunduz et al. (2004))



  - Betweenness Centrality (Borgatti and Halgin (2011), Newman et al. (2001))

    Anzahl der kürzesten Wege eines Graphen, die durch einen bestimmten Knoten laufen

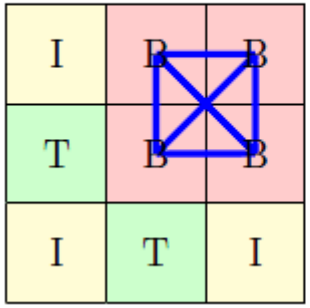


  - Mittlerer Knotengrad von Gebäuden

2. Methoden - Graphen-Indikatoren

- Adjazenz-Event Maße

Quelle: Barnsley, M. J.; Barr, S. L. (1996)



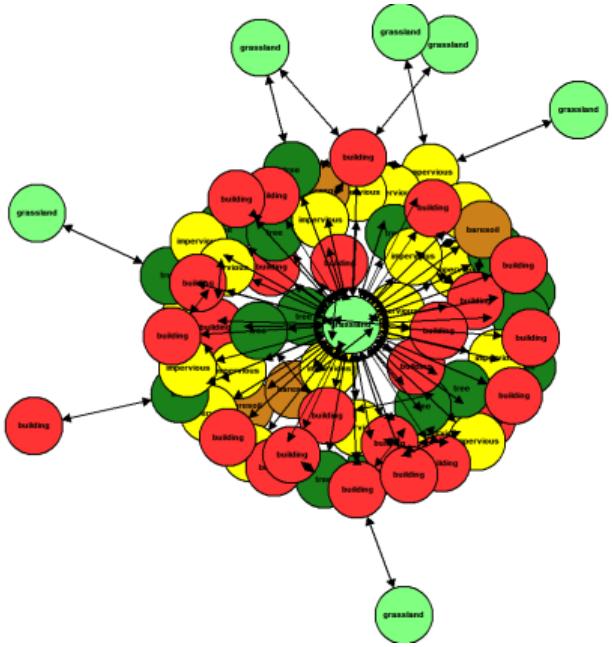
	B	T	I
B	6	4	5
T	–	1	4
I	–	–	0

- Berücksichtigen Häufigkeit und räumliche Anordnung von Landbedeckungsklassen

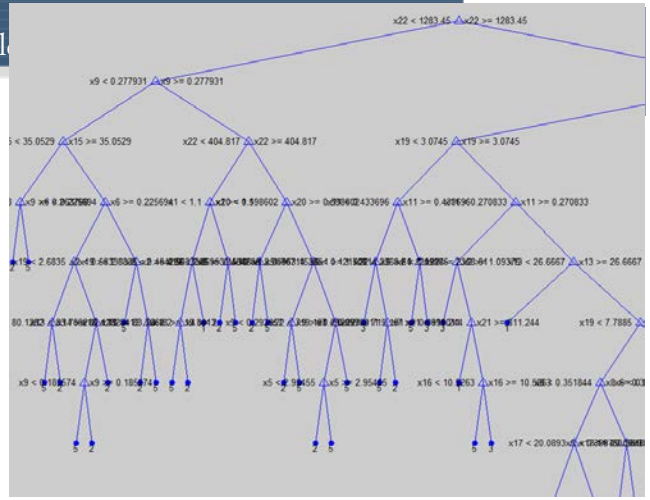
- Ähnlich dem Contagion Index in Landschaftsökologischen Untersuchungen

2. Methoden - Graphen-Indikatoren

- Adjazenz-Event Maße
  - Norm. Kantenanzahl zu Versiegelt, Baum, Gras
  - Norm. Kantenanzahl von Gebäude zu Versiegelt, Baum, Gras
  - % an Gebäuden mit mindestens 1 Kante zu Versiegelt, Baum, Gras
  - Diversität (Fraction) um Gebäude
    (wie viele Landbedeckungsklassen sind Gebäude-Nachbarn)
  - % an Gebäuden umschlossen von Versiegelt, Baum, Gras
    (Knotengrad=1)



- Zusätzliche Graphen-Maße
  - Gebäude mit höchstem Knotengrad im Graph
    - Gebäudehöhe
    - Gebäudefläche
    - Gebäudevolumen
    - Gebäude Shape Index (Form)



2. Methoden - Random Forest Analyse

- Klassifikator: Machine learning
- Nicht-parametrische Methode
- Generiert ein Ensemble von Entscheidungsbäumen (bagged decision trees)
- Bagging = *b*ootstrap *agg*regat*ing*
  - unterschiedl. Auszug an Trainingsdaten werden genutzt um die einzelnen Bäume zu erzeugen
  - 2/3 Trainingsdaten, 1/3 out-of-bag
- Split Kriterium: nicht alle Variablen werden benutzt
  - $\sqrt[2]{n_{\mathrm{var}}}$ - beste Variable wird für den Split verwendet
- Klassen-Zuweisung: Mehrheit der Stimmenvorhersagen der Bäume

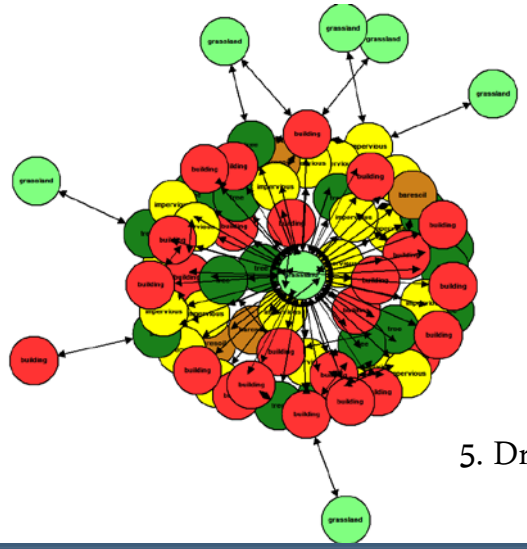
- Implementiert z.B. in R und Matlab (TreeBagger)

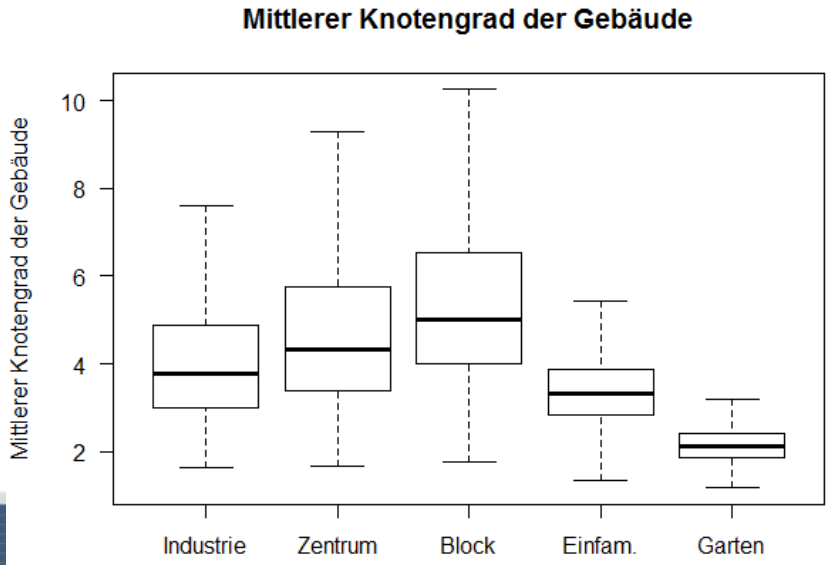
Quelle: Breiman, L. (2001), Breiman, L. (2002) , Liaw, A. and Wiener, M.(2002), Strobl, C.; Boulesteix, A.; Zeileis, A.; Hothorn, T. (2007)

3. Ergebnisse - Graphen-Indikatoren – Zentralitäts-Indikatoren

Tab.: Landbedeckung vom Knoten mit höchster Betweeness und Degree Centrality von 1588 Baublöcken in Rostock in Prozent

Betweenness / Degree Centrality	Industrie	Zentrum	Block	Einfamilienhäuser	Gartenanlage
Wasser	1,1/0,5	0,0/0,0	0,0/0,0	0,0/0,0	0,0/0,0
Boden	0,5/1,1	0,0/0,0	0,0/0,0	0,0/0,0	0,0/0,0
Baum	3,8/2,7	2,4/4,1	12,1/8,9	11,3/8,5	6,5/5,7
Gras	24,3/21,1	3,3/4,1	46,8/46,8	69,5/70,5	92,8/93,6
Gebäude	7,0/7,6	64,2/61,8	15,8/14,7	1,5/1,3	0,3/0,3
Versiegelt	63,2/67,0	30,1/30,1	25,3/29,5	17,8/19,8	0,4/0,4
Summe Baublöcke	185	123	190	400	690





3. Ergebnisse - Graphen-Indikatoren – Adjazenz-Event Maße

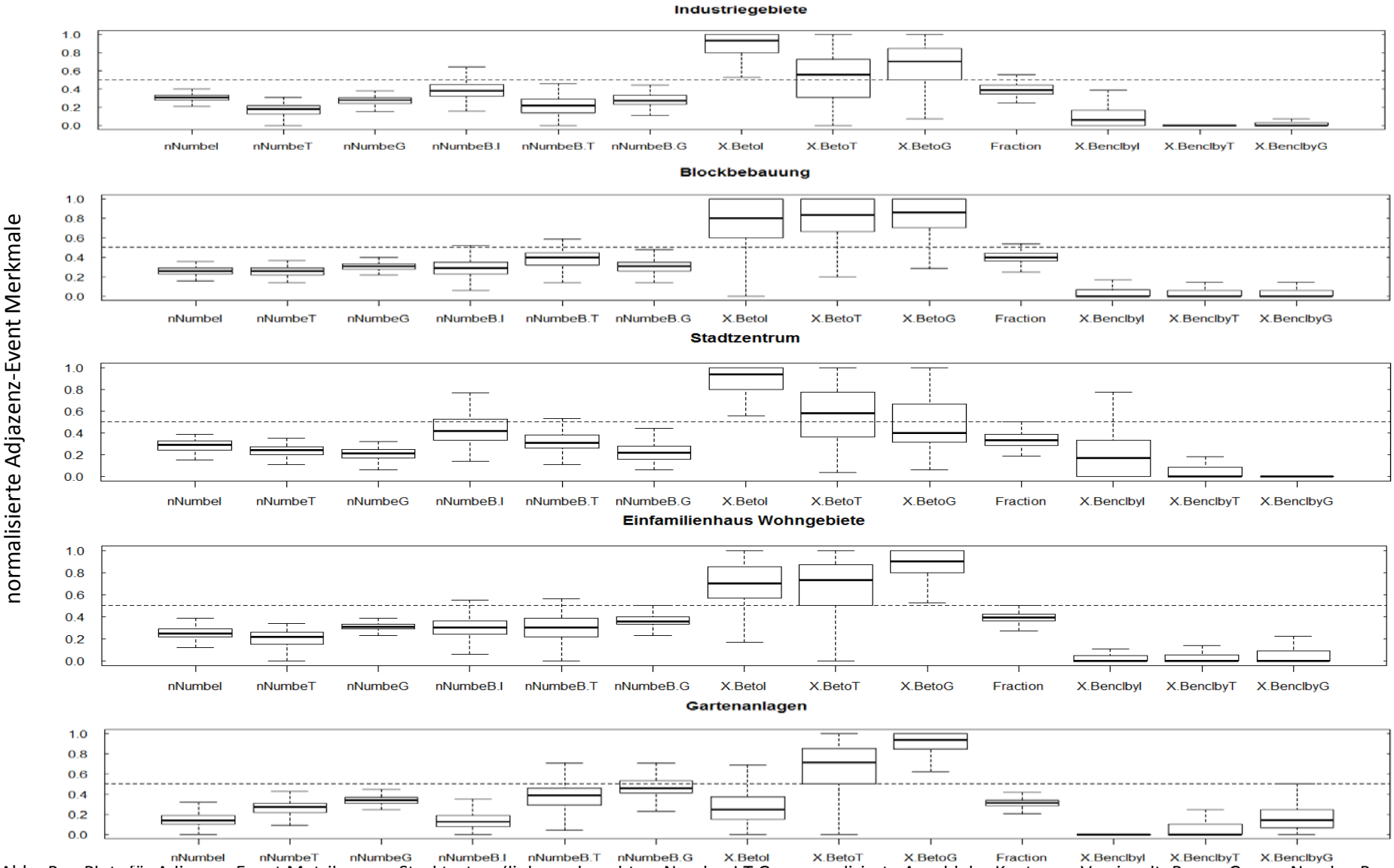
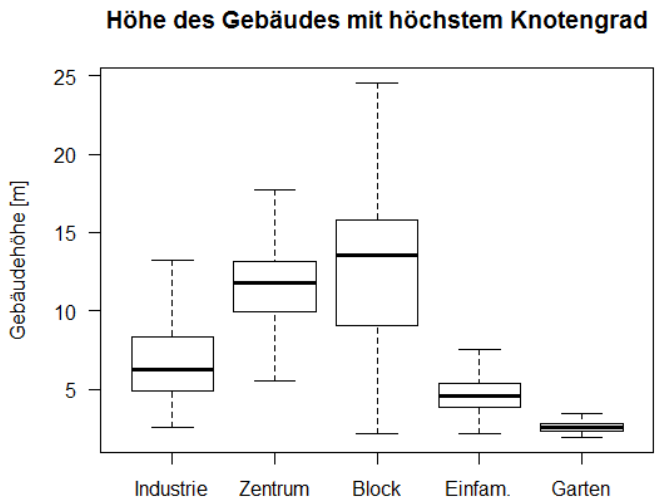
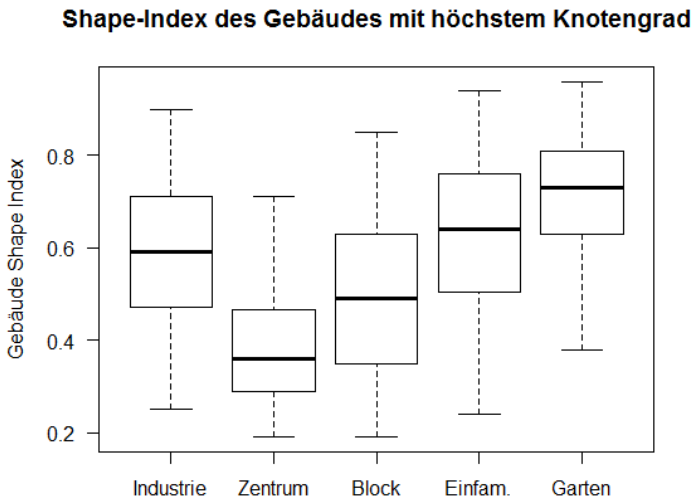
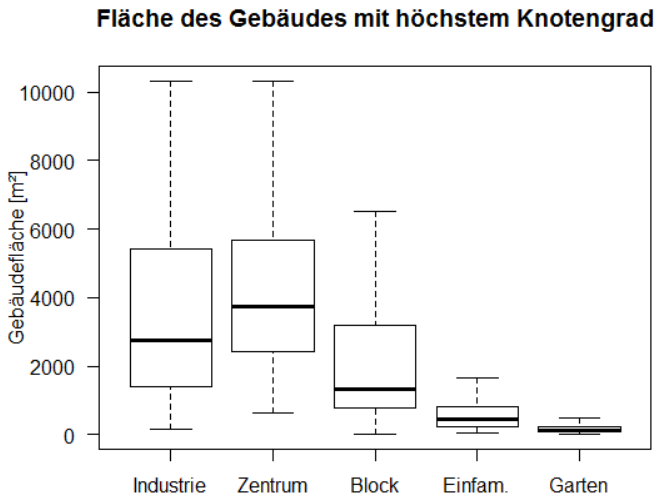


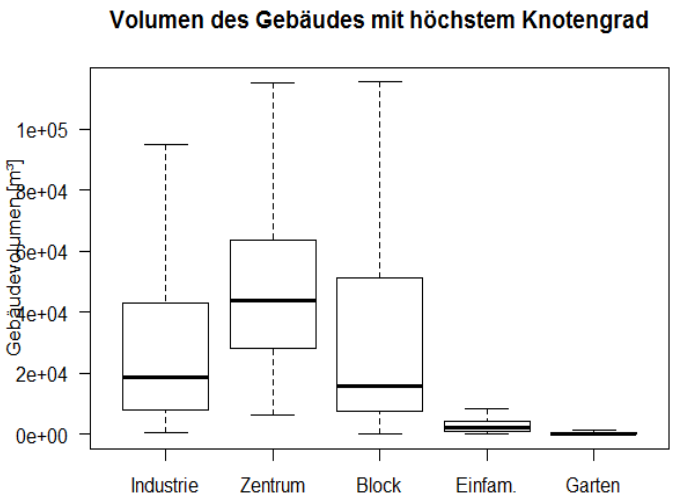
Abb.: Box-Plots für Adjazenz-Event Metriken pro Strukturtyp. (links nach rechts: nNumb e I,T,G... normalisierte Anzahl der Kanten zu Versiegelt, Baum, Gras; nNumb e B-I,T,G... normalisierte Anzahl der Kanten von Gebäude zu Versiegelt, Baum, Gras; % B e to I,T,G... % von Gebäuden mit mindestens einer Kante zu Versiegelt, Baum, Gras; FractionAround B... Landbedeckungs-Diversität um Gebäude; % encl by I,T,G... % umschlossener Gebäude von Versiegelt, Baum, Gras

3. Ergebnisse - Graphen-Indikatoren – Zusätzliche Graphen-Maße





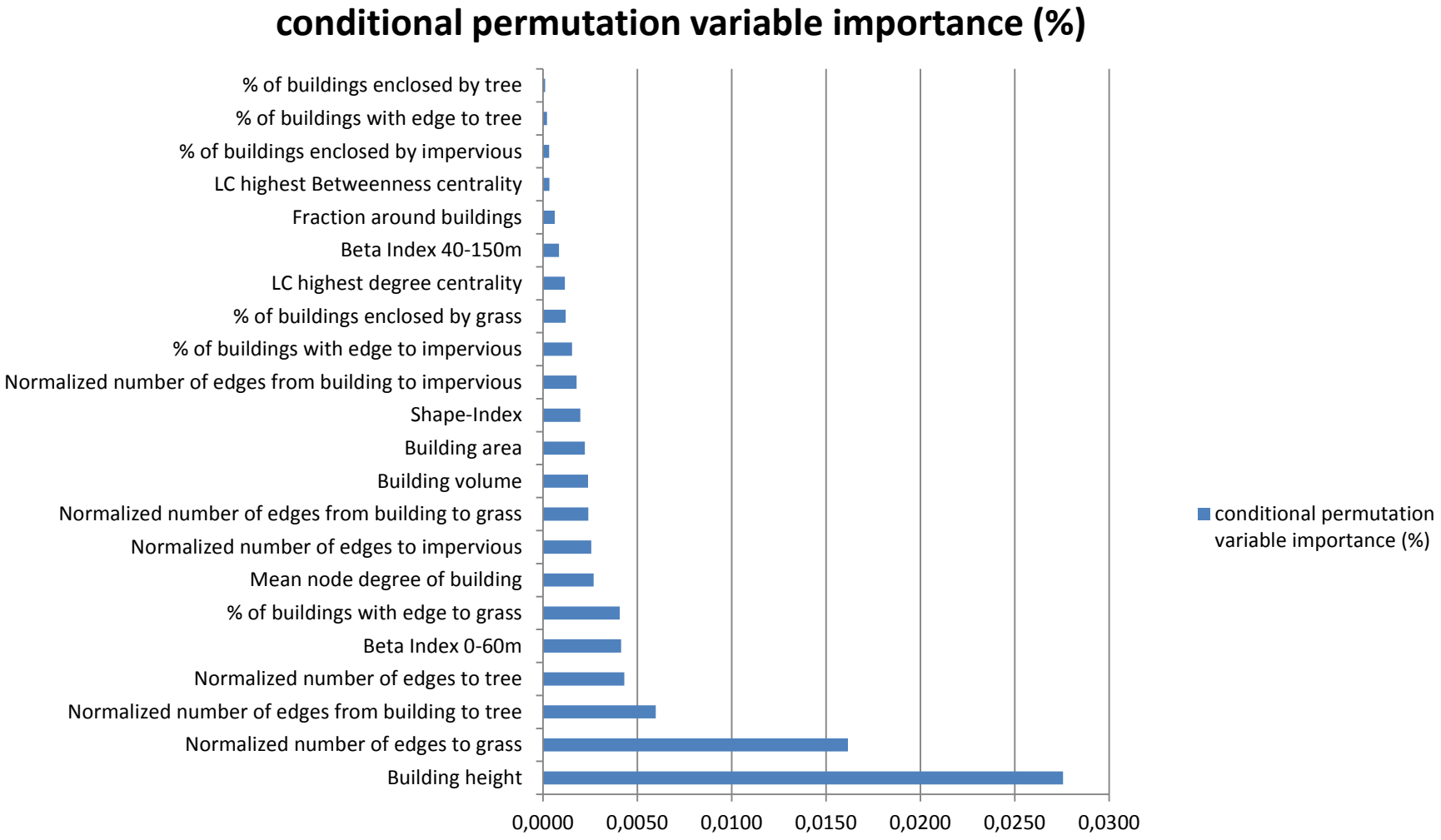




3. Ergebnisse - Random Forest Analyse

- 1588 Graphen (urbane Blöcke), 22 Variablen (Graphen Indikatoren)
- Verwendete Software: R, cforest (party-package)
- 60 zufällig gewählte Trainingsgebiete pro Stadtstrukturtyp um Bäume für jede Klasse gleichmäßig zu trainieren
- Anzahl der Bäume: 500
- Anzahl der Variablen für Split: $\sqrt[2]{n_{\text{var}}} = 5$ (2 und 10 getestet aber nicht besser)

3. Ergebnisse - Random Forest Analyse: Variablen Wichtigkeit



3. Ergebnisse - Random Forest Analyse: Variablen Wichtigkeit

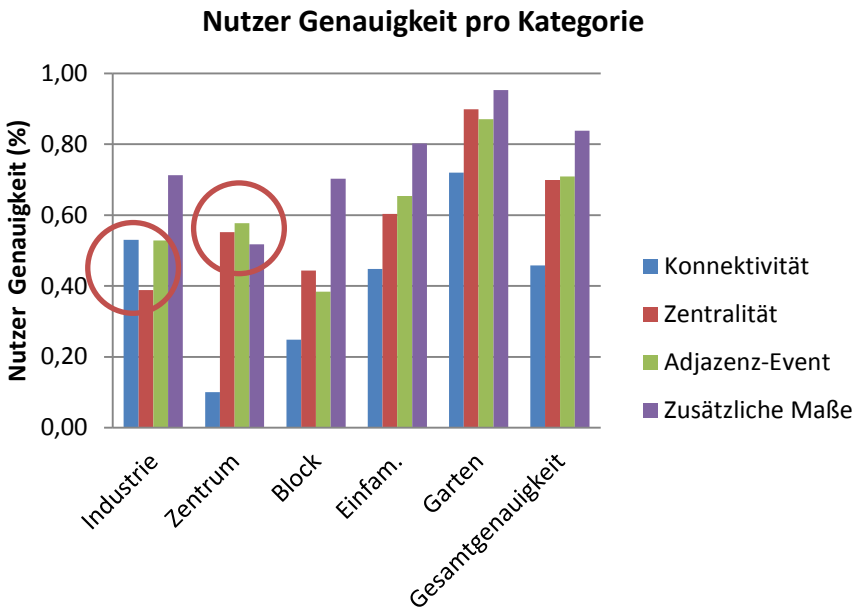
Tab.: Variablen Reduktion

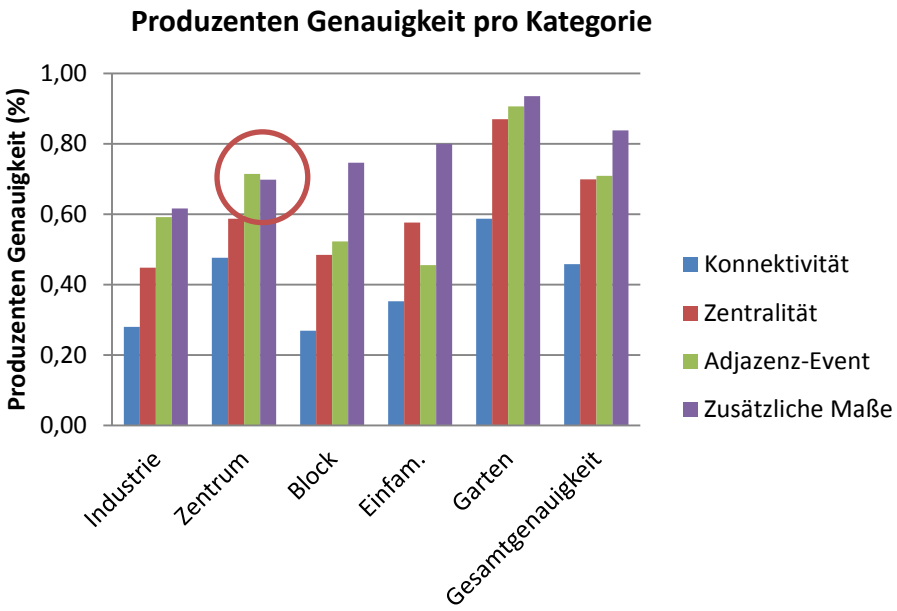
Anzahl der benutzten Variabeln	entfernte Variable	Testset - Gesamtgenauigkeit (%)
22	alle Variablen verwendet	0,864
21	% of buildings enclosed by tree	0,864
20	% of buildings enclosed by impervious	0,865
19	% of buildings with edge to tree	0,861
18	LC highest Betweenness centrality	0,865
17	LC highest degree centrality	0,868
16	Fraction around buildings	0,862
15	Beta Index 40-150m	0,867
14	% of buildings enclosed by grass	0,865
13	Normalized number of edges from building to impervious	0,866
12	Normalized number of edges to impervious	0,865
11	Building area	0,859
10	Shape-Index	0,865
9	% of buildings with edge to impervious	0,855
8	Normalized number of edges from building to grass	0,846
7	Building volume	0,823
6	% of buildings with edge to grass	0,828
5	Normalized number of edges to tree	0,825
4	Mean node degree of building	0,807
3	Beta Index 0-60m	0,788
2	Normalized number of edges from building to tree	0,773
1	Normalized number of edges to grass	0,725

Entfernen der ersten 5 Variablen steigert die Klassifikationsgenauigkeit!

Genauigkeit von 81% wird erreicht wenn nur die Gebäudehöhe entfernt wird

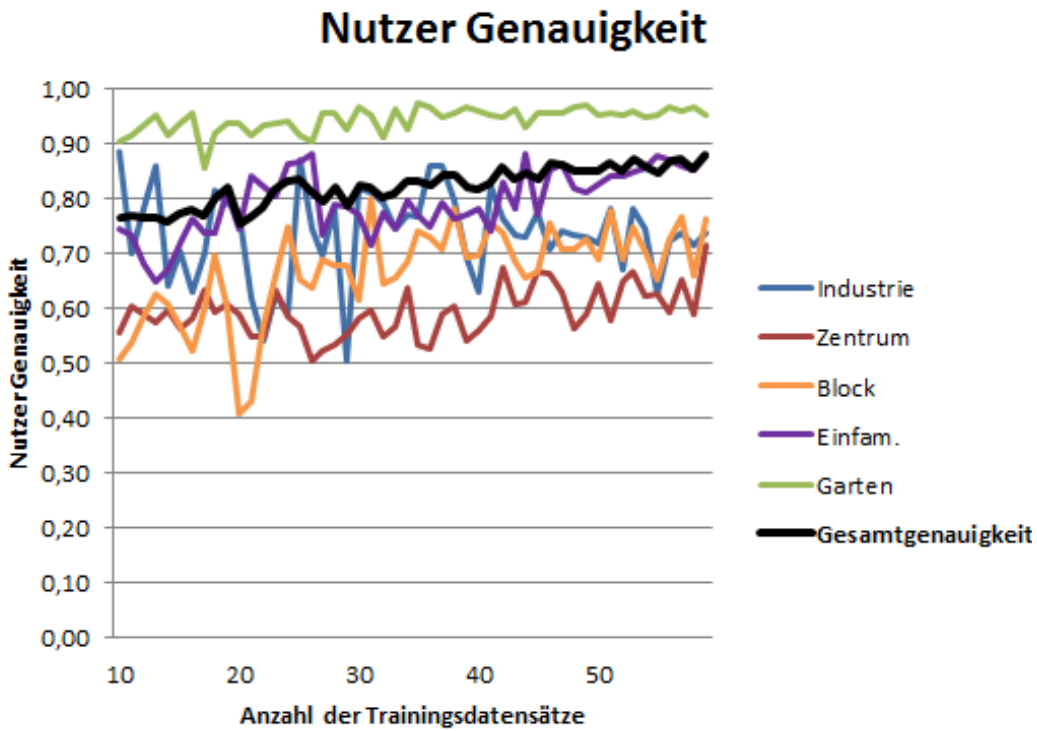
3. Ergebnisse - Random Forest Analyse: Variablen Wichtigkeit

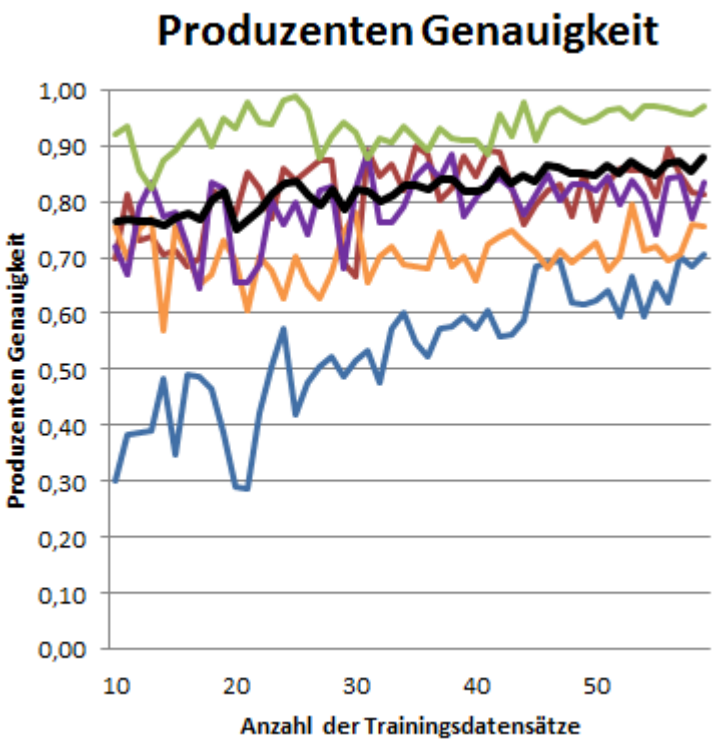




- Konnektivitäts-Maße (2) -> geringste Genauigkeit
- Zentralitäts-Maße (3) und Adjazenz-Event Maße (13) wechseln sich im Ranking ab
- Zusätzliche Maße (4) -> höchste Genauigkeit (maßgeblich Gebäudehöhe)

3. Ergebnisse - Random Forest Analyse: Reduktion der Trainingsdaten





- Inkrementelle Reduktion der Trainingsdaten um 1 von 60 auf 10
- Mehr Trainingsdaten – bessere Genauigkeit
- Zufriedenstellende Gesamtgenauigkeiten über 80% werden mit mindestens 33 Trainingsdaten pro Klasse erreicht
- Gibt Auskunft über Variabilität innerhalb einer Klasse (Industrie: hohe Variabilität, Gartenanlagen: niedrige Variabilität)

3. Ergebnisse - Validierung

- n_{tree}= 500
- 60 Trainingssamples pro Strukturtyp
- Insgesamt 1288 Testdatensätze
- Nur die 17 besten Variablen verwendet
- Kappa Koeffizient: 0.80

Tab.: Konfusionsmatrix der klassifizierten Test - Datensätze

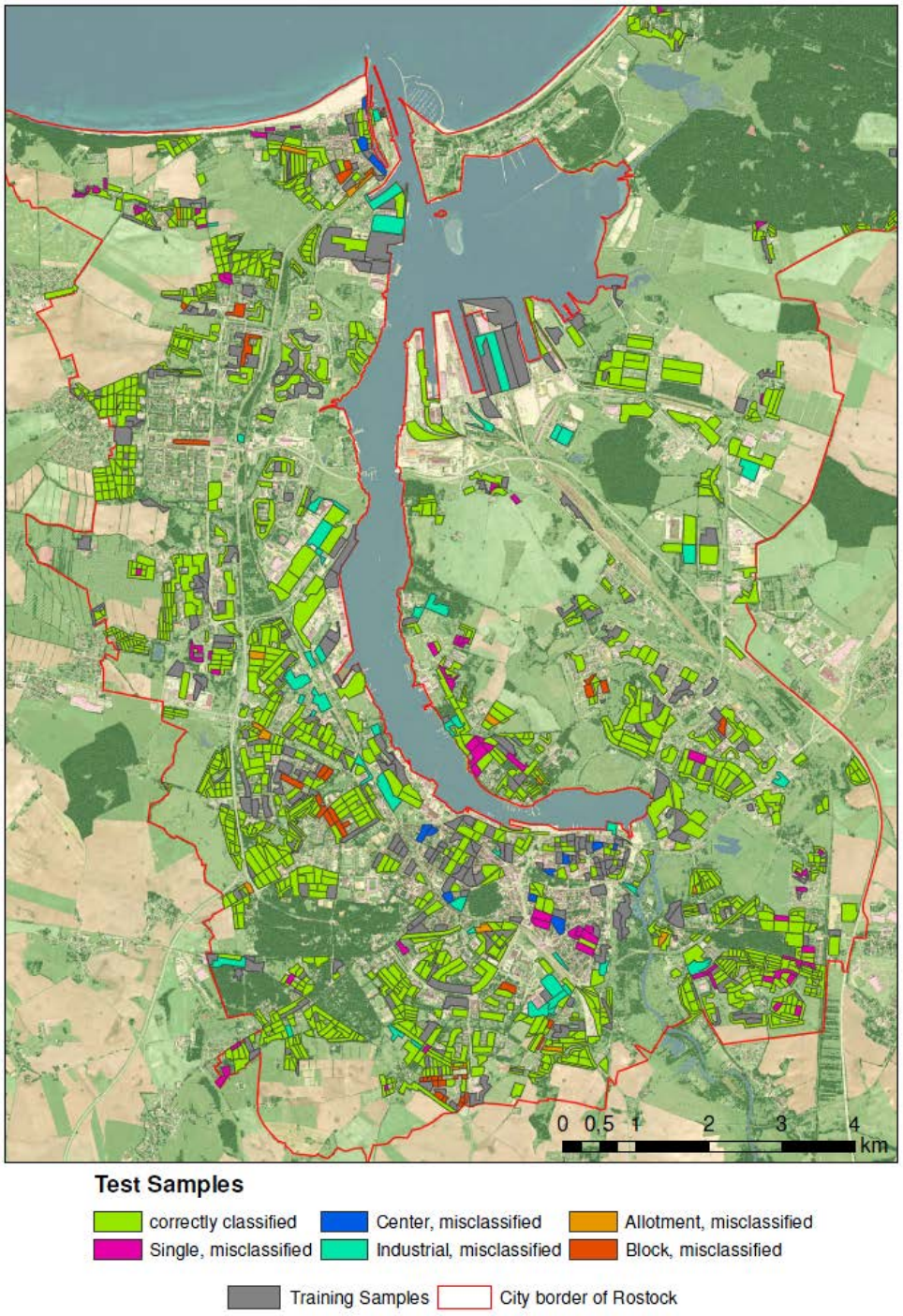
UST	Industrie	Zentrum	Block	Einfam.	Garten	Summe	User Acc. (%)
Industrie	82	20	12	10	1	125	0,66
Zentrum	4	51	8	0	0	63	0,81
Block	10	5	95	18	2	130	0,73
Einfam.	20	4	13	276	27	340	0,81
Garten	0	0	2	14	614	630	0,97
Summe	116	80	130	318	644	1288	
Prod. Acc. (%)	0,71	0,64	0,73	0,87	0,95		0,87

3. Ergebnisse - Validierung

60 Trainingssamples pro Strukturtyp (300)

1118 von 1288 korrekt klassifiziert

170 Falsch-Klassifikationen



4. Zusammenfassung

- Nachbarschafts-Graphen aus 6 Landbedeckungsklassen wurden erzeugt

- Klassifikation der Graphen in 5 Stadtstrukturtypen

- Nachbarschafts-Graphen sinnvoll für die Stadtstrukturtypen Klassifikation

- Random Forest: praktische Methode zur überwachten Klassifikation zur gleichzeitigen Analyse von mehreren Variablen

- Gebäudehöhen: signifikant zur Trennung urbaner Strukturen
- Adjazenz-Event-Maße: großen Einfluss auf das Klassifikationsergebnis

- Hohe Intra-Klassen-Variabilität bei Industrial und Center
- Geringe Intra-Klassen-Variabilität bei Gartenanlagen

- To Do: Übertragbarkeitstests auf andere Städte
    - Landbedeckungs-Klassen müssen in vergleichbarer Genauigkeit vorliegen!
    - Ähnlichkeit der Strukturtypen

Vielen Dank für die Aufmerksamkeit!

Mail: irene.walde@uni-jena.de

Referenzen:

Barnsley, M. J.; Barr, S. L. (1996): Inferring Urban Land Use from Satellite Sensor Images Using Kernel-Based Spatial Reclassification. In *Photogrammetric Engineering & Remote Sensing* 62 (8), pp. 949–958.

Barr, Stuart; Barnsley, Michael (1998): Application of structural pattern-recognition techniques to infer urban land use from Ordnance Survey digital map data. In *Geocomputation*. Available online at http://www.geocomputation.org/1998/43/gc_43.htm.

Berger, C.; Voltersen, M.; Hese, S.; Walde, I.; Schmullius, C. (accepted): Robust Extraction of Urban Land Cover Information from HSR Multi-Spectral and LiDAR Data. In *IEEE Journal of selected topics in Applied Earth Observations & Remote Sensing*.

Berger, Christian; Voltersen, Michael; Hese, Sören; Walde, Irene; Schmullius, Christiane (2012): Using Geographic Object-Based Image Analysis (GEOBIA) for Urban Land Cover Mapping and Settlement Density Assessment. In : Proceedings of the 4th GEOBIA. GEOBIA. Rio de Janeiro, Brasil, May 7-9, 2012, pp. 503–508.

Borgatti, Stephen P.; Halgin, Daniel S. (2011, c2011): Analyzing Affiliation Networks. In John Scott, Peter J. Carrington (Eds.): The SAGE handbook of social network analysis. London ;, Thousand Oaks, Calif: SAGE.

Breiman, Leo (2001): Random Forests. In *Machine Learning* 45 (1), pp. 5–32.

Gunduz, C.; Yener, B.; Gultekin, S. H. (2004): The cell graphs of cancer. In *Bioinformatics* 20 (Suppl 1), pp. i145–i151.

Liaw, Andy; Wiener, Matthias (2002): Classification and regression by random forest. In *R News* 2 (3), pp. 18–22. Available online at http://www.webchem.science.ru.nl:8080/PRiNS/rF.pdf.

Newman, M. E. J. (2001): Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. In *Phys. Rev. E* 64 (1).

Strobl, Carolin; Boulesteix, Anne-Laure; Zeileis, Achim; Hothorn, Torsten (2007): Bias in random forest variable importance measures: Illustrations, sources and a solution. In *BMC Bioinformatics* 8:25.

Walde, I.; Hese, S.; Berger, C.; Schmullius, C. (accepted): Graph based Mapping of Urban Structure Types from High Resolution Satellite Image Objects - A case study of the German cities Rostock and Erfurt. In *IEEE Geoscience and Remote Sensing Letters*.

Walde, Irene; Hese, Sören; Schmullius, Christiane (2012): Graph Based Mapping of Urban Structure Types from High Resolution Satellite Image Objects. In : Proceedings of the 4th GEOBIA. GEOBIA. Rio de Janeiro, Brasil, May 7-9, 2012, pp. 355–360.

Walde, Irene; Hese, Sören; Berger, Christian; Schmullius, Christiane (2012): Graph-based Urban Land Use Mapping from High Resolution Satellite Images. In *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* I-4, pp. 119–124.