

This is an excerpt from the thesis “*Cost-Sensitive Active Learning mit Fernerkundungsdaten zur Abschätzung der Gebäudevulnerabilität bei Hochwasser und Erdbeben*”.

Please contact Matthias Thoma for a full version of the thesis.

Extended works on this topic are documented in:

Geiß, C., Thoma, M., Pittore M., Wieland, M., Dech, S., and Taubenböck, H. (2017): Multitask Active Learning for Characterization of Built Environments with Multisensor Earth Observation Data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 10, 12, 5583–5597.

Geiß, C., Thoma, M., and Taubenböck, H. (2018): Cost-sensitive Multitask Active Learning for Characterization of Urban Environments with Remote Sensing. *IEEE Geoscience and Remote Sensing Letters*, 15, 6, 922–926.

Abschlussarbeit
im Studiengang Master of Science in Praktischer Informatik

**Cost-Sensitive Active Learning mit Fernerkundungsdaten
zur Abschätzung der Gebäudevulnerabilität
bei Hochwasser und Erdbeben**

Matthias Thoma
Januar 2017

Gutachten

Erstgutachterin: Prof. Dr. Gabriele Peters
Zweitgutachter: Prof. Dr. Christoph Beierle

FernUniversität in Hagen
Fakultät für Mathematik und Informatik
Lehrgebiet Mensch-Computer-Interaktion

Inhaltsverzeichnis

1	Einführung	1
1.1	Motivation	1
1.2	Zielsetzung	2
1.3	Gliederung	3
2	Grundlagen und aktuelle Verfahren	5
3	Datengrundlagen und Merkmalsvektor	7
3.1	Untersuchungsgebiet	7
3.2	Datengrundlagen	9
3.3	Merkmalsvektor	12
3.3.1	Geometrische Merkmale	13
3.3.2	Spektrale Features	17
3.3.3	Relative Maße auf Blocklevel	18
3.4	Klassifikationsschema der Ziel-Klassifikation	19
4	Referenzklassifikation mit Supervised Learning	23
4.1	Support Vector Machines	23
4.2	Feature-Selektion	27
4.3	Single-Task Supervised-Klassifikation	33
5	Klassifikation mit Active Learning	37
5.1	Active Learning	38
5.1.1	Unsicherheitskriterium: Multiclass-Level Uncertainty	40
5.1.2	Verschiedenheitskriterium: Angle Based Diversity	41
5.1.3	Multi-Task AL: Klassifikation mehrerer Variablen gleichzeitig	44
5.1.4	Berücksichtigung der Kosten bei der Auswahl der Samples	46
5.1.5	Retuning-Strategie für SVM-Hyperparameter	49
5.2	Klassifikation mit AL im Untersuchungsgebiet	51
5.2.1	Single-Task AL-Klassifikation	51
5.2.2	1. Szenario: Multi-Task AL-Klassifikation	55

5.2.3	2. Szenario: Cost-Sensitive Multi-Task AL-Klassifikation auf geomer-Daten	61
6	Zusammenfassung, Bewertung und Ausblick	65
A	Quelltexte	67
A.1	Hilfsfunktionen	67
A.2	Datenvorbereitung	72
A.3	Feature-Selektion	84
A.4	Lernen / Query-Funktionen	94
A.5	Ausgabe / Aufbereitung der Ergebnisse	138
	Literaturverzeichnis	161

Zusammenfassung

In dieser Arbeit wird der Einsatz von Active Learning (AL) zur effizienten Charakterisierung von Gebäuden mit Hilfe von aus Fernerkundungsdaten gewonnenen Merkmalen untersucht. Bei AL wird basierend auf bisherigem Wissen (in Form von gelabelten Samples) ein vorläufiges Modell zur Klassifikation einer Zielvariablen gebildet; eine Query-Funktion fragt in einem iterativen Prozess gezielt bei einem Experten diejenigen Samples als zusätzliche Trainingsdaten an, die den größten Informationsgewinn versprechen. Im Gegensatz zu vielen Aufgabenstellungen mit nur einer Zielvariable (Single-Task AL), bei denen AL eingesetzt wird, liegen im betrachteten Anwendungskontext „Abschätzung der Gebäudevulnerabilität bei Hochwasser und Erdbeben“ zwei Besonderheiten vor: Erstens ist eine Vielzahl von gleichzeitig zu klassifizierenden Zielvariablen interessant (Multi-Task AL); zweitens ist die Erfassung mancher Eigenschaften zeit- und kostenintensiv, so dass bei der Auswahl der anzufragenden Samples der erwartete Aufwand berücksichtigt werden sollte (Cost-Sensitive AL).

Entsprechend wird eine leistungsfähige Single-Task AL-Technik, die Samples nach Unsicherheit und Vielfalt anfragt, zunächst für Multi-Task AL erweitert. Hierbei werden verschiedene Strategien, die auf abwechselnder Auswahl bzw. auf Rang-Kombination basieren, evaluiert. Schließlich wird als weiteres Auswahlkriterium eine Kostenkomponente hinzugefügt, wobei der Zeitaufwand anhand der Entfernungen auf dem Straßennetz unter Berücksichtigung unterschiedlicher Reiseschwindigkeiten (mit dem Auto / zu Fuß) realitätsnah abgeschätzt und die Tourplanung als Problem des Handlungsreisenden optimiert wird.

Die experimentelle Untersuchung anhand zweier Anwendungsszenarios im Kölner Stadtgebiet im Hinblick auf insgesamt fünf Zielvariablen zeigt, dass die entwickelten und eingesetzten Multi-Task-Strategien bessere Klassifikationsergebnisse erzielen als zufällige Sample-Auswahl und dass durch Cost-Sensitive Multi-Task AL vergleichbare Genauigkeit mit geringerem Zeit- und Kostenaufwand bzw. in gegebenem Zeitbudget bessere Klassifikationsergebnisse erreicht werden können als mit Query-Funktionen, die bei der Auswahl die erwarteten Kosten nicht berücksichtigen.

Weltweit sind viele Siedlungen in von Naturkatastrophen bedrohten Gebieten entstanden. Das Bauen entlang Flüssen und am Meer ist seit jeher aus Versorgungs- und Infrastrukturgründen attraktiv, bringt jedoch Hochwassergefahren mit sich. Einige Megacities mit mehr als zehn Millionen Einwohnern sowie eine Vielzahl kleinerer Groß- und Millionenstädte liegen in erdbebengefährdeten Gebieten – gerade in diesen bevölkerungsreichen Ballungsräumen findet jedoch häufig unkontrolliertes Stadtwachstum statt. Viele Gebäude wurden oder werden mit unzulänglichen Qualitätsstandards in gefährdeten Gebieten errichtet, Hochwasser- und Erdbebensicherheit finden dabei häufig keine oder nur unzureichende Berücksichtigung. Über das umfangreiche Gebäudeinventar und das Gefährdungsrisiko bei Katastrophen gibt es in vielen Gegenden keine belastbaren Aussagen, die jedoch für gezielte Stadtplanungsprozesse, für lebensrettende präventive Verbesserungsmaßnahmen oder auch für die Einsatzplanung im Katastrophenfall nötig wären.

1.1 Motivation

Das individuelle Risiko jedes Gebäudes bei Naturkatastrophen kann aus einer Kombination verschiedener Attribute abgeschätzt werden. Zur Ermittlung der Verwundbarkeit bei Erdbeben wird hierfür beispielsweise die GEM Building Taxonomy 2.0 [BSC⁺13] eingesetzt. Umfangreiche Aufnahmen des Gebäudezustands vor Ort durch Bauingenieure sind jedoch zeit- und kostenintensiv, die gewonnenen Daten veralten bei rapider Stadtentwicklung relativ schnell. Aus Fernerkundungsdaten können hingegen für viele Gebäude gleichzeitig mit vergleichsweise geringem Aufwand Merkmale zu ihrer Charakterisierung gewonnen werden, wobei allerdings nur wenige dieser Merkmale direkte Relevanz in Bezug auf die Verwundbarkeit bei Naturkatastrophen haben.

Mit Techniken des Maschinellen Lernens kann anhand weniger vor Ort bewerteter Gebäudesamples ein Modell trainiert werden, mit dem für viele andere Gebäude die gesuchten Eigenschaften indirekt aus den Fernerkundungsdaten abgeleitet werden können. Um eine ausreichende Vorhersagegüte zu erhalten, sind bei überwachten Lernverfahren (SL, supervised learning – es werden dabei alle vorab gelabelten Samples zur Modellbildung herangezogen) vergleichsweise viele Samples zum Trainieren des Modells notwendig;

die Auswahl geeigneter Samples muss durch Experten erfolgen, die eine passende Samplingstrategie entwickeln.

In diesem Anwendungskontext sind Verfahren gesucht, bei denen eine hohe Vorhersagegüte mit möglichst wenig manuellem Erfassungsaufwand vor Ort erreicht werden kann. Die zu erfassenden Gebäude sollen so ausgewählt werden, dass sie möglichst schnell und kostengünstig bewertet werden können. Lernverfahren, die nicht mehr ausschließlich auf vorab gelabelten Samples basieren, versprechen hier Vorteile.

1.2 Zielsetzung

Mit dieser Arbeit soll untersucht werden, wie sich durch Aktives Lernen (AL, active learning) der Aufwand für die Auswahl geeigneter Samples reduzieren lässt. Bei AL werden zunächst nur wenige Samples zur Modellbildung herangezogen. Der Algorithmus entscheidet in jedem Schritt eines iterativen Prozesses, welche Gebäude besonders interessant erscheinen und einen hohen Informationsgewinn versprechen. Für diese Instanzen werden die Attribute angefragt, durch einen Experten gelabelt und ein verbessertes Modell errechnet. Bei der Auswahl der Samples kann neben der erwarteten Verbesserung des Modells auch der Aufwand für das (manuelle) Erfassen der angefragten Samples berücksichtigt werden, hierfür sollen entsprechende Kostenfunktionen implementiert werden.

Zur experimentellen Untersuchung der Verfahren stehen Fernerkundungsdaten zur Verfügung, aus denen für einen Großteil der Gebäude im Kölner Stadtgebiet charakteristische Merkmale auf drei Hierarchieebenen (Einzelgebäude, Gebäudegruppen, Blöcke) errechnet wurden. Vom Geoforschungszentrum Potsdam wurde ein Datensatz mit einigen in Köln vor Ort aufgenommenen erdbeben- und hochwasserrelevanten Attributen auf etwa 800 Gebäuden zur Verfügung gestellt – mit drei Zielvariablen wird ein 1. Szenario durchgespielt: Vor Ort wurden bereits in einem Gebiet Aufnahmen von Bauten gemacht, die für dieses Gebiet typisch sind. Jetzt müssen diese Aufnahmen „im Labor“ ausgewertet und die in-situ-Informationen typisiert werden. Durch AL wird dieser Labeling-Prozess beschleunigt, es müssen nur noch spürbar weniger Gebäude bewertet werden. Der Active Learner soll bei der Auswahl der zu labelnden Gebäude die Relevanz für alle drei Variablen berücksichtigen (= Multi-Task AL).

Zusätzlich steht ein von der Firma geomer bereitgestellter Datensatz zur Verfügung, der für ein Projekt der Stadtentwässerungsbetriebe Köln verwendet wurde. Hieraus werden zwei Zielvariablen für einen weiteren Testlauf mit Multi-Task AL genutzt und entsprechend das 1. Szenario in einem wesentlich größeren Betrachtungsgebiet simuliert.

Zudem werden die geomer-Daten für ein 2. Szenario verwendet: In-situ-Daten sind noch nicht gegeben und müssen vor Ort gesammelt werden; AL wird eingesetzt, um den Routing-/Labeling-Prozess zu beschleunigen. Neben dem zu erwartenden Informationsgewinn durch das nächste Label fließen auch die Kosten (z.B. Wegekosten / Zeiteinsatz) in die Auswahl ein (= Cost-Sensitive Multi-Task AL).

1.3 Gliederung

Diese Arbeit ist in folgende Teile gegliedert: In Kapitel 2 wird zunächst ein Überblick über Konzepte und aktuelle Verfahren gegeben. Kapitel 3 beschreibt das Untersuchungsgebiet in Köln, die verwendeten Datengrundlagen, die berechneten Merkmale sowie die für die experimentelle Untersuchung ausgewählten Zielvariablen und ihre jeweiligen Klassen.

In Kapitel 4 wird nach einer Einführung zu Support Vector Machines (SVM), die als Klassifikator dienen, eine Feature-Selektion durchgeführt, um für die Zielvariablen jeweils passende Prediktorvariablen auszuwählen. Es folgt als Referenz oder Benchmark für die späteren Aktiven Lernverfahren eine Klassifikation mit Überwachtem Lernen (SL mit SVM). Die eingesetzten AL-Verfahren und entsprechende Experimente werden in Kapitel 5 beschrieben. In Kapitel 6 folgen Bewertung und Zusammenfassung der Ergebnisse.

This is an excerpt from the thesis “*Cost-Sensitive Active Learning mit Fernerkundungsdaten zur Abschätzung der Gebäudevulnerabilität bei Hochwasser und Erdbeben*”.

Please contact Matthias Thoma for a full version of the thesis.

Extended works on this topic are documented in:

Geiß, C., Thoma, M., Pittore M., Wieland, M., Dech, S., and Taubenböck, H. (2017): Multitask Active Learning for Characterization of Built Environments with Multisensor Earth Observation Data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 10, 12, 5583–5597.

Geiß, C., Thoma, M., and Taubenböck, H. (2018): Cost-sensitive Multitask Active Learning for Characterization of Urban Environments with Remote Sensing. *IEEE Geoscience and Remote Sensing Letters*, 15, 6, 922–926.

- [APC10] ANGEL, Shlomo ; PARENT, Jason ; CIVCO, Daniel L.: Ten compactness properties of circles: measuring shape in geography. In: *Canadian Geographer / Le Géographe canadien* 54 (2010), Nr. 4, 441–461. <http://dx.doi.org/10.1111/j.1541-0064.2009.00304.x>. – DOI 10.1111/j.1541-0064.2009.00304.x. – ISSN 1541-0064
- [BKR16] BIVAND, Roger ; KEITT, Tim ; ROWLINGSON, Barry: *rgdal: Bindings for the Geospatial Data Abstraction Library*, 2016. <https://CRAN.R-project.org/package=rgdal>. – R package version 1.1-10
- [Bou16] BOUNEFFOUF, Djallel: Exponentiated Gradient Exploration for Active Learning. In: *Computers* 5 (2016), Nr. 1. <http://dx.doi.org/10.3390/computers5010001>. – DOI 10.3390/computers5010001
- [Bre01] BREIMAN, Leo: Random Forests. In: *Machine Learning* 45 (2001), Nr. 1, S. 5–32. <http://dx.doi.org/10.1023/A:1010933404324>. – DOI 10.1023/A:1010933404324. – ISSN 08856125
- [Bri03] BRINKER, Klaus: Incorporating Diversity in Active Learning with Support Vector Machine. In: *Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003)*. Menlo Park, California : AAAI Press, 2003, S. 59–66
- [BSC⁺13] BRZEV, Svetlana ; SCAWTHORN, Charles ; CHARLESON, Andrew W. ; ALLEN, Luke ; GREENE, Marjorie ; JAISWAL, Kishor ; SILVA, Vitor: *GEM Building Taxonomy (Version 2.0)*. <http://www.globalquakemodel.org/resources/publications/technical-reports/gem-building-taxonomy-report/>. Version: 2013, Abruf: 14.01.2017
- [BTL⁺14] BELGIU, Mariana ; TOMLJENOVIC, Ivan ; LAMPOLTSHAMMER, Thomas ; BLASCHKE, Thomas ; HÖFLE, Bernhard: Ontology-Based Classification of Building Types Detected from Airborne Laser Scanning Data. In: *Remote Sensing* 6 (2014), Nr. 2, S. 1347–1366. <http://dx.doi.org/10.3390/rs6021347>. – DOI 10.3390/rs6021347. – ISSN 2072-4292

- [BVE⁺13] BERGER, Christian ; VOLTERSEN, Michael ; ECKARDT, Robert ; EBERLE, Jonas ; HEYER, Thomas ; SALEPCI, Nesrin ; HESE, Soren ; SCHMULLIUS, Christiane ; TAO, Junyi ; AUER, Stefan ; BAMLER, Richard ; EWALD, Ken ; GARTLEY, Michael ; JACOBSON, John ; BUSWELL, Alan ; DU, Qian ; PACIFICI, Fabio: Multi-Modal and Multi-Temporal Data Fusion: Outcome of the 2012 GRSS Data Fusion Contest. In: *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 6 (2013), June, Nr. 3, S. 1324–1340. <http://dx.doi.org/10.1109/JSTARS.2013.2245860>. – DOI 10.1109/JSTARS.2013.2245860. – ISSN 1939–1404
- [CGJ96] COHN, David A. ; GHAHRAMANI, Zoubin ; JORDAN, Michael I.: Active Learning with Statistical Models. In: *J. Artif. Int. Res.* 4 (1996), März, Nr. 1, 129–145. <http://dl.acm.org/citation.cfm?id=1622737.1622744>. – ISSN 1076–9757
- [CL11] CHANG, Chih-Chung ; LIN, Chih-Jen: LIBSVM: A library for support vector machines. In: *ACM Transactions on Intelligent Systems and Technology* 2 (2011), S. 27:1–27:27. – Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [CN06] CSARDI, Gabor ; NEPUSZ, Tamas: The igraph software package for complex network research. In: *InterJournal Complex Systems* (2006), 1695. <http://igraph.org>
- [Coh60] COHEN, Jacob: A Coefficient of Agreement for Nominal Scales. In: *Educational and Psychological Measurement* 20 (1960), Nr. 1, S. 37–46. <http://dx.doi.org/10.1177/001316446002000104>. – DOI 10.1177/001316446002000104. – ISSN 0013–1644
- [CST00] CRISTIANINI, Nello ; SHAWE-TAYLOR, John: *An introduction to support vector machines: And other kernel-based learning methods*. Cambridge and New York : Cambridge University Press, 2000. – ISBN 978–0521780193
- [CSW⁺12] CARTER, Jamie ; SCHMID, Keil ; WATERS, Kirk ; BETZOLD, Lindy ; HADLEY, Brian ; MATAOSKY, Rebecca ; HALLERAN, Jennifer ; NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA) COASTAL SERVICES CENTER (Hrsg.): *Lidar 101: An Introduction to Lidar Technology, Data, and Applications*. <https://coast.noaa.gov/data/digitalcoast/pdf/lidar-101.pdf>. Version: 2012, Abruf: 14.01.2017
- [Dij59] DIJKSTRA, Edsger W.: A Note on Two Problems in Connexion with Graphs. In: *Numerische Mathematik* 1 (1959), Nr. 1, S. 269–271

- [DIN01] *DIN EN 1998-1/NA:2011-01: Nationaler Anhang - National festgelegte Parameter - Eurocode 8: Auslegung von Bauwerken gegen Erdbeben - Teil 1: Grundlagen, Erdbebeneinwirkungen und Regeln für Hochbau.* Norm. Berlin, 2011-01
- [dlr15a] *Höhendaten für Gebäude im Untersuchungsgebiet.* 2015. – bereitgestellt von Christian Geiß, DLR, 01.10.2015
- [dlr15b] *Landcover-Klassifikation im Untersuchungsgebiet.* 2015. – bereitgestellt von Christian Geiß, DLR, 22.12.2015
- [dlr15c] *Spektrale Merkmale für Gebäude im Untersuchungsgebiet.* 2015. – bereitgestellt von Christian Geiß, DLR, 01.12.2015
- [DMB14] DEMIR, Begüm ; MINELLO, Luca ; BRUZZONE, Lorenzo: Definition of Effective Training Sets for Supervised Classification of Remote Sensing Images by a Novel Cost-Sensitive Active Learning Method. In: *IEEE Transactions on Geoscience and Remote Sensing* 52 (2014), Feb, Nr. 2, S. 1272–1284. <http://dx.doi.org/10.1109/TGRS.2013.2249522>. – DOI 10.1109/TGRS.2013.2249522. – ISSN 0196–2892
- [DPB11] DEMIR, Begüm ; PERSELLO, Claudio ; BRUZZONE, Lorenzo: Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images. In: *IEEE Transactions on Geoscience and Remote Sensing* 49 (2011), Nr. 3, S. 1014–1031. <http://dx.doi.org/10.1109/TGRS.2010.2072929>. – DOI 10.1109/TGRS.2010.2072929. – ISSN 0196–2892
- [DZZ16] DU, S. de ; ZHANG, F. ; ZHANG, X.: Semantic classification of urban buildings combining VHR images and GIS data, 2016
- [eco15] TRIMBLE GERMANY GMBH (Hrsg.): *eCognition Developer Reference Book 9.1.2.* München, 2015
- [GCV⁺12] GRAESSER, Jordan ; CHERIYADAT, Anil ; VATSAVAI, Ranga R. ; CHANDOLA, Varun ; LONG, Jordan ; BRIGHT, Eddie: Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (2012), Nr. 4, S. 1164–1176. <http://dx.doi.org/10.1109/JSTARS.2012.2190383>. – DOI 10.1109/JSTARS.2012.2190383. – ISSN 1939–1404
- [GE03] GUYON, Isabelle ; ELISSEEFF, André: An Introduction to Variable and Feature Selection. In: *J. Mach. Learn. Res.* 3 (2003), März, 1157–1182. <http://dl.acm.org/citation.cfm?id=944919.944968>. – ISSN 1532–4435

- [Geo06] GEOLOGISCHER DIENST NRW ; GEOLOGISCHER DIENST NORDRHEIN-WESTFALEN IN ZUSAMMENARBEIT MIT DEM MINISTERIUM FÜR BAUEN UND VERKEHR DES LANDES NORDRHEIN-WESTFALEN (Hrsg.): *Karte der Erdbebenzonen und geologischen Untergrundklassen der Bundesrepublik Deutschland: Bundesland Nordrhein-Westfalen 1:350 000, Karte zu DIN 4149*. Krefeld, 2006
- [geo16] *In-situ-Daten mit Gebäude-Taxonomie*. 2016. – bereitgestellt von Stefan Jäger, geomer, 03.05.2016
- [gfz16] *In-situ-Daten mit Gebäude-Taxonomie*. 2016. – bereitgestellt von Massimiliano Pittore, GFZ, 10.03.2016
- [GPM⁺15] GEISS, Christian ; PELIZARI, Patrick A. ; MARCONCINI, Mattia ; SENGARRA, Wayan ; EDWARDS, Mark ; LAKES, Tobia ; TAUBENBÖCK, Hannes: Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 104 (2015), 175–188. <http://dx.doi.org/10.1016/j.isprsjprs.2014.07.016>. – DOI 10.1016/j.isprsjprs.2014.07.016. – ISSN 0924–2716
- [Grü98] GRÜNTHAL, Gottfried: *Cahiers du Centre européen de géodynamique et de séismologie*. Bd. v. 15: *European macroseismic scale 1998: EMS-98*. [2nd ed.]. Luxembourg : European Seismological Commission, Subcommittee on Engineering Seismology, Working Group Macroseismic scales, 1998. – ISBN 2–87977–008–4
- [GT13] GEISS, Christian ; TAUBENBÖCK, Hannes: Remote sensing contributing to assess earthquake risk: From a literature review towards a roadmap. In: *Natural Hazards* 68 (2013), Nr. 1, S. 7–48. <http://dx.doi.org/10.1007/s11069-012-0322-2>. – DOI 10.1007/s11069-012-0322-2. – ISSN 0921–030X
- [GTT⁺14] GEISS, Christian ; TAUBENBÖCK, Hannes ; TYAGUNOV, Sergey ; TISCH, Anita ; POST, Joachim ; LAKES, Tobia: Assessment of Seismic Building Vulnerability from Space. In: *Earthquake Spectra* 30 (2014), Nr. 4, S. 1553–1583. <http://dx.doi.org/10.1193/121812EQS350M>. – DOI 10.1193/121812EQS350M. – ISSN 8755–2930
- [GTW⁺11] GEISS, Christian ; TAUBENBÖCK, Hannes ; WURM, Michael ; ESCH, Thomas ; NAST, Michael ; SCHILLINGS, Christoph ; BLASCHKE, Thomas: Remote Sensing-Based Characterization of Settlement Structures for Assessing Local Potential of District Heat. In: *Remote Sensing* 3 (2011),

- Nr. 12, S. 1447–1471. <http://dx.doi.org/10.3390/rs3071447>. – DOI 10.3390/rs3071447. – ISSN 2072–4292
- [HCL16] HSU, Chih-Wei ; CHANG, Chih-Chung ; LIN, Chih-Jen: A Practical Guide to Support Vector Classification / Department of Computer Science, National Taiwan University. Version: 2016. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, Abruf: 14.01.2017. 2016. – Forschungsbericht
- [HH16] HAHLER, Michael ; HORNIK, Kurt: *TSP: Traveling Salesperson Problem (TSP)*. <https://CRAN.R-project.org/package=TSP>. Version: 2016. – R package version 1.1-4
- [HL02] HSU, Chih-Wei ; LIN, Chih-Jen: A comparison of methods for multiclass support vector machines. In: *IEEE transactions on neural networks* 13 (2002), Nr. 2, S. 415–425. <http://dx.doi.org/10.1109/72.991427>. – DOI 10.1109/72.991427. – ISSN 1045–9227
- [HSD73] HARALICK, Robert M. ; SHANMUGAM, K. ; DINSTEN, Its'hak: Textural Features for Image Classification. In: *IEEE Transactions on Systems, Man and Cybernetics SMC-3* (1973), Nov, Nr. 6, S. 610–621. <http://dx.doi.org/10.1109/TSMC.1973.4309314>. – DOI 10.1109/TSMC.1973.4309314. – ISSN 0018–9472
- [HTF09] HASTIE, Trevor ; TIBSHIRANI, Robert ; FRIEDMAN, J. H.: *The elements of statistical learning: Data mining, inference, and prediction*. Second edition, corrected 7th printing. New York : Springer, 2009 (Springer series in statistics). <http://statweb.stanford.edu/~tibs/ElemStatLearn/>. – ISBN 978–0387848570
- [JWHT13] JAMES, Gareth ; WITTEN, Daniela ; HASTIE, Trevor ; TIBSHIRANI, Robert: *An Introduction to Statistical Learning*. Bd. 103. New York, NY : Springer New York, 2013. <http://dx.doi.org/10.1007/978-1-4614-7138-7>. <http://dx.doi.org/10.1007/978-1-4614-7138-7>. – ISBN 978–1–4614–7137–0
- [KWW⁺16] KUHN, Max ; WING, Jed ; WESTON, Steve ; WILLIAMS, Andre ; KEEFER, Chris ; ENGELHARDT, Allan ; COOPER, Tony ; MAYER, Zachary ; KENKEL, Brenton ; THE R CORE TEAM ; BENESTY, Michael ; LESCARBEAU, Reynald ; ZIEM, Andrew ; SCRUCICA, Luca ; TANG, Yuan ; CANDAN, Can.: *caret: Classification and Regression Training*, 2016. <https://CRAN.R-project.org/package=caret>. – R package version 6.0-71

- [Ley11] LEYDECKER, Günter: *Geologisches Jahrbuch Reihe E, Geophysik*. Bd. 59: *Erdbebenkatalog für Deutschland mit Randgebieten für die Jahre 800 bis 2008: Mit 5 Tabellen, 9 Anhängen*. Stuttgart : Schweizerbart, 2011. – ISBN 978–3510959891
- [LMGCCV13] LEIVA-MURILLO, Jose M. ; GOMEZ-CHOVA, Luis ; CAMPS-VALLS, Gustavo: Multitask Remote Sensing Data Classification. In: *IEEE Transactions on Geoscience and Remote Sensing* 51 (2013), Nr. 1, S. 151–161. <http://dx.doi.org/10.1109/TGRS.2012.2200043>. – DOI 10.1109/TGRS.2012.2200043. – ISSN 0196–2892
- [MCE15] MCGARIGAL, Kevin ; CUSHMAN, Sam ; ENE, Eduard: *FRAGSTATS: Spatial Pattern Analysis Program for Categorical and Continuous Maps (Fragstats Help)*. <http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf>. Version: 2015, Abruf: 28.03.2016. – Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: <http://www.umass.edu/landeco/research/fragstats/fragstats.html>
- [MDH⁺15] MEYER, David ; DIMITRIADOU, Evgenia ; HORNIK, Kurt ; WEINGESSEL, Andreas ; LEISCH, Friedrich: *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2015. <https://CRAN.R-project.org/package=e1071>. – R package version 1.6-7
- [Mit82] MITCHELL, Tom M.: Generalization as search. In: *Artificial Intelligence* 18 (1982), Nr. 2, S. 203–226. [http://dx.doi.org/10.1016/0004-3702\(82\)90040-6](http://dx.doi.org/10.1016/0004-3702(82)90040-6). – DOI 10.1016/0004-3702(82)90040-6. – ISSN 00043702
- [MUP04] MITRA, Pabitra ; UMA SHANKAR, B. ; PAL, Sankar K.: Segmentation of multispectral remote sensing images using active support vector machines. In: *Pattern Recognition Letters* 25 (2004), Nr. 9, S. 1067–1074. <http://dx.doi.org/10.1016/j.patrec.2004.03.004>. – DOI 10.1016/j.patrec.2004.03.004. – ISSN 01678655
- [OKS05] OSUGI, Thomas ; KUN, Deng ; SCOTT, Stephen: Balancing Exploration and Exploitation: A New Algorithm for Active Machine Learning. In: *Proceedings of the Fifth IEEE International Conference on Data Mining*. Washington, DC, USA : IEEE Computer Society, 2005 (ICDM '05). – ISBN 0–7695–2278–5, 330–337
- [osm15] OPENSTREETMAP (Hrsg.) ; GEOFABRIK (Hrsg.): *Download OpenStreetMap data for this region: Regierungsbezirk Köln*. <http://download.geofabrik.de/europe/germany/nordrhein-westfalen/koeln-regbez.html>. Version: 2015, Abruf: 28.09.2015

- [PBD⁺14] PERSELLO, Claudio ; BOULARIAS, Abdeslam ; DALPONTE, Michele ; GOBAKKEN, Terje ; NÆSSET, Erik ; SCHÖLKOPF, Bernhard: Cost-Sensitive Active Learning With Lookahead: Optimizing Field Surveys for Remote Sensing Data Classification. In: *IEEE Transactions on Geoscience and Remote Sensing* 52 (2014), Nr. 10, S. 6652–6664. <http://dx.doi.org/10.1109/TGRS.2014.2300189>. – DOI 10.1109/TGRS.2014.2300189. – ISSN 0196–2892
- [peg17] WASSER- UND SCHIFFFAHRTSVERWALTUNG DES BUNDES (Hrsg.): *Elektronischer Wasserstraßen-Informationsservice (ELWIS): Amtliche Messwerte des Pegels Köln*. https://www.elwis.de/gewaesserkunde/Wasserstaende/Wasserstaende_start.php?target=2&pegelId=a6ee8177-107b-47dd-bcfd-30960ccc6e9c. Version: 2017, Abruf: 14.01.2017
- [Pel08] PELZING, Rolf: *Erdbeben in Nordrhein-Westfalen*. Krefeld : Geologischer Dienst Nordrhein-Westfalen, 2008. – ISBN 3–86029–971–9
- [PW13] PITTORE, Massimiliano ; WIELAND, Marc: Toward a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing. In: *Natural Hazards* 68 (2013), Nr. 1, S. 115–145. <http://dx.doi.org/10.1007/s11069-012-0475-z>. – DOI 10.1007/s11069–012–0475–z. – ISSN 0921–030X
- [R C16] R CORE TEAM ; R FOUNDATION FOR STATISTICAL COMPUTING (Hrsg.): *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2016. <https://www.R-project.org/>
- [reg12] INITIATIVKREIS EUROPÄISCHE METROPOLREGIONEN IN DEUTSCHLAND (IKM) ; BUNDESINSTITUT FÜR BAU-, STADT- UND RAUMFORSCHUNG (BBSR) IM BUNDESAMT FÜR BAUWESEN UND RAUMORDNUNG (BBR) (Hrsg.): Regionales Monitoring 2012. Daten und Karten zu den Europäischen Metropolregionen in Deutschland. Version: 2012. http://www.deutsche-metropolregionen.org/fileadmin/ikm/01_monitoring/Regionales_Monitoring_IKM_BBR_2013.pdf, Abruf: 14.01.2017. 2012. – Forschungsbericht
- [Ric96] RICHTER, Rudolf: A spatially adaptive fast atmospheric correction algorithm. In: *International Journal of Remote Sensing* 17 (1996), Nr. 6, S. 1201–1214. <http://dx.doi.org/10.1080/01431169608949077>. – DOI 10.1080/01431169608949077. – ISSN 0143–1161

- [RN12] RUSSELL, Stuart J. ; NORVIG, Peter: *Künstliche Intelligenz: Ein moderner Ansatz*. 3., aktualisierte Aufl. München [u.a.] : Pearson, 2012 (Pearson-Studium). – ISBN 978–3868940985
- [RS16] RICHTER, Rudolf ; SCHLÄPFER, Daniel ; DLR (Hrsg.): *Atmospheric / Topographic Correction for Satellite Imagery*. https://www.rese-apps.com/pdf/atcor3_manual.pdf. Version: 2016, Abruf: 01.05.2016. – DLR report DLR-IB 565-01/15
- [RSL77] ROSENKRANTZ, Daniel J. ; STEARNS, Richard E. ; LEWIS, II, PHILIP M.: An Analysis of Several Heuristics for the Traveling Salesman Problem. In: *SIAM Journal on Computing* 6 (1977), Nr. 3, S. 563–581. <http://dx.doi.org/10.1137/0206041>. – DOI 10.1137/0206041. – ISSN 0097–5397
- [RSt16] RSTUDIO TEAM ; RSTUDIO, INC. (Hrsg.): *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc., 2016. <http://www.rstudio.com/>
- [RTH08] REICHART, Roi ; TOMANEK, Katrin ; HAHN, Udo: Multi-task active learning for linguistic annotations. In: *In ACL*, 2008
- [RTH⁺11] ROBIN, Xavier ; TURCK, Natacha ; HAINARD, Alexandre ; TIBERTI, Natalia ; LISACEK, Frédérique ; SANCHEZ, Jean-Charles ; MÜLLER, Markus: pROC: an open-source package for R and S+ to analyze and compare ROC curves. In: *BMC Bioinformatics* 12 (2011), S. 77
- [RW15] REVOLUTION ANALYTICS ; WESTON, Steve: *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*, 2015. <https://CRAN.R-project.org/package=doParallel>. – R package version 1.0.10
- [SC08] STEINWART, Ingo ; CHRISTMANN, Andreas: *Support vector machines*. 1st ed. New York : Springer, 2008 (Information science and statistics). – ISBN 978–0387772417
- [Sch08] SCHLEPÜTZ, Elke: *Mit Sicherheit für Köln : ein Meilenstein für den Hochwasserschutz*. <http://www.steb-koeln.de/Redaktionell/Downloads/Hochwasserschutz/Mit-Sicherheit-f%C3%BCr-K%C3%B6ln-Ein-Meilenstein-f%C3%BCr-den-Hochwasserschutz.pdf>. Version: 2008, Abruf: 14.01.2017
- [Scr11] SCRUGA, Luca: Model-based SIR for dimension reduction. In: *Computational Statistics & Data Analysis* 5 (2011), Nr. 11, S. 3010–3026
- [Set12] SETTLES, Burr: *Synthesis lectures on artificial intelligence and machine learning*. Bd. 18: *Active learning*. San Rafael, Calif. (1537 Fourth Street,

- San Rafael, CA 94901 USA) : Morgan & Claypool, 2012. – ISBN 978–1608457250
- [SLBW08] STEINIGER, Stefan ; LANGE, Tilman ; BURGHARDT, Dirk ; WEIBEL, Robert: An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques. In: *Transactions in GIS* 12 (2008), Nr. 1, 31–59. <http://dx.doi.org/10.1111/j.1467-9671.2008.01085.x>. – DOI 10.1111/j.1467-9671.2008.01085.x. – ISSN 1467-9671
- [SLM⁺14] STUMPF, André ; LACHICHE, Nicolas ; MALET, Jean-Philippe ; KERLE, Norman ; PUISSANT, Anne: Active Learning in the Spatial Domain for Remote Sensing Image Classification. In: *IEEE Transactions on Geoscience and Remote Sensing* 52 (2014), Nr. 5, S. 2492–2507. <http://dx.doi.org/10.1109/TGRS.2013.2262052>. – DOI 10.1109/TGRS.2013.2262052. – ISSN 0196-2892
- [SS02] SCHÖLKOPF, Bernhard ; SMOLA, Alexander J.: *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, Mass. : MIT Press, 2002 (Adaptive computation and machine learning). – ISBN 9780262194754
- [sta15] KÖLN. AMT FÜR STADTENTWICKLUNG UND STATISTIK (Hrsg.): Statistisches Jahrbuch Köln 2015. Version: 2015. <http://www.stadt-koeln.de/politik-und-verwaltung/statistik/jahrbuecher/>. 2015. – Forschungsbericht. – ISSN 0933-632X
- [sta10] STADT KÖLN (Hrsg.): *Kommunale Gebietsgliederung: Stadtteil (Shape)*. <http://offenedaten-koeln.de/dataset/stadtteile>. Version: 2014-07-10, Abruf: 09.01.2016
- [TEF⁺12] TAUBENBÖCK, Hannes ; ESCH, Thomas ; FELBIER, Andreas ; WIESNER, Michael ; ROTH, Achim ; DECH, Stefan: Monitoring urbanization in mega cities from space. In: *Remote Sensing of Environment* 117 (2012), S. 162–176. <http://dx.doi.org/10.1016/j.rse.2011.09.015>. – DOI 10.1016/j.rse.2011.09.015. – ISSN 00344257
- [TK02] TONG, Simon ; KOLLER, Daphne: Support Vector Machine Active Learning with Applications to Text Classification. In: *J. Mach. Learn. Res.* 2 (2002), März, 45–66. <http://dx.doi.org/10.1162/153244302760185243>. – DOI 10.1162/153244302760185243. – ISSN 1532-4435
- [TKW⁺13] TAUBENBÖCK, Hannes ; KLOTZ, Martin ; WURM, Michael ; SCHMIEDER, J. ; WAGNER, B. ; WOOSTER, M. ; ESCH, Thomas ; DECH, Stefan: Delineation of Central Business Districts in mega city regions using

- remotely sensed data. In: *Remote Sensing of Environment* 136 (2013), S. 386–401. <http://dx.doi.org/10.1016/j.rse.2013.05.019>. – DOI 10.1016/j.rse.2013.05.019. – ISSN 00344257
- [TRP⁺09] TUIA, Devis ; RATLE, Frédéric ; PACIFICI, Fabio ; KANEVSKI, Mikhail F. ; EMERY, William J.: Active Learning Methods for Remote Sensing Image Classification. In: *IEEE Transactions on Geoscience and Remote Sensing* 47 (2009), Nr. 7, S. 2218–2232. <http://dx.doi.org/10.1109/TGRS.2008.2010404>. – DOI 10.1109/TGRS.2008.2010404. – ISSN 0196–2892
- [TVC⁺11] TUIA, Devis ; VOLPI, Michele ; COPA, Loris ; KANEVSKI, Mikhail ; MUNOZ-MARI, Jordi: A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. In: *IEEE Journal of Selected Topics in Signal Processing* 5 (2011), Nr. 3, S. 606–617. <http://dx.doi.org/10.1109/JSTSP.2011.2139193>. – DOI 10.1109/JSTSP.2011.2139193. – ISSN 1932–4553
- [TW12] TOMPALSKI, Piotr ; WEZYK, P.: LiDAR and VHRS data for assessing living quality in cities - an approach based on 3D spatial indices. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B6* (2012), S. 173–176. <http://dx.doi.org/10.5194/isprsarchives-XXXIX-B6-173-2012>. – DOI 10.5194/isprsarchives-XXXIX-B6-173-2012. – <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B6/173/2012/>
- [TWF⁺14] TAUBENBÖCK, Hannes ; WIESNER, Michael ; FELBIER, Andreas ; MARCONCINI, Mattia ; ESCH, Thomas ; DECH, Stefan: New dimensions of urban landscapes: The spatio-temporal evolution from a polynuclei area to a mega-region based on remote sensing data. In: *Applied Geography* 47 (2014), S. 137–153. <http://dx.doi.org/10.1016/j.apgeog.2013.12.002>. – DOI 10.1016/j.apgeog.2013.12.002. – ISSN 01436228
- [WBB⁺16] WARNES, Gregory R. ; BOLKER, Ben ; BONEBAKKER, Lodewijk ; GENTLEMAN, Robert ; HUBER, Wolfgang ; LIAW, Andy ; LUMLEY, Thomas ; MAECHLER, Martin ; MAGNUSSON, Arni ; MOELLER, Steffen ; SCHWARTZ, Marc ; VENABLES, Bill: *ggplots: Various R Programming Tools for Plotting Data*, 2016. <https://CRAN.R-project.org/package=ggplots>. – R package version 3.0.1
- [Wic09] WICKHAM, Hadley: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009 <http://ggplot2.org>. – ISBN 978–0–387–98140–6

- [Wor13] DIGITAL GLOBE (Hrsg.): *WorldView-2 Data Sheet*. http://global.digitalglobe.com/sites/default/files/DG_WorldView2_DS_PROD.pdf. Version: 2013, Abruf: 14.01.2017
- [WST16] WURM, Michael ; SCHMITT, Andreas ; TAUBENBÖCK, Hannes: Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions. In: *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 9 (2016), May, Nr. 5, S. 1901–1912. <http://dx.doi.org/10.1109/JSTARS.2015.2465131>. – DOI 10.1109/JSTARS.2015.2465131. – ISSN 1939–1404
- [WTS⁺11] WURM, Michael ; TAUBENBÖCK, Hannes ; SCHARDT, Mathias ; ESCH, Thomas ; DECH, Stefan: Object-based image information fusion using multi-sensor earth observation data over urban areas. In: *International Journal of Image and Data Fusion* 2 (2011), Nr. 2, S. 121–147. <http://dx.doi.org/10.1080/19479832.2010.543934>. – DOI 10.1080/19479832.2010.543934. – ISSN 1947–9832