



Flächennutzungsmonitoring VII Boden – Flächenmanagement – Analysen und Szenarien

IÖR Schriften Band 67 · 2015

ISBN: 978-3-944101-67-5

Techniken zur Dateninspektion am Beispiel der Tagbevölkerungsdichte

Martin Behnisch, Odette Kretschmer, Alfred Ultsch

Behnisch, Martin; Kretschmer, Odette; Ultsch, Alfred (2015):
Techniken zur Dateninspektion am Beispiel der Tagbevölkerungsdichte. In: Gotthard Meinel, Ulrich Schumacher, Martin Behnisch, Tobias Krüger (Hrsg.): Flächennutzungsmonitoring VII. Boden – Flächenmanagement – Analysen und Szenarien. Berlin: Rhombos-Verlag, 2015, (IÖR-Schriften; 67), S. 245-256

Techniken zur Dateninspektion am Beispiel der Tagbevölkerungsdichte

Martin Behnisch, Odette Kretschmer, Alfred Ultsch

Zusammenfassung

Die Reduktion der Flächeninanspruchnahme als auch die standörtliche Lenkung von Bautätigkeiten sind wesentliche Ziele einer nachhaltigen Flächenpolitik. Ein Erkennen und Beurteilen räumlicher Unterschiede in Ausmaß und Lokalität der in Anspruch genommenen Siedlungs- und Verkehrsflächen bedarf der Berücksichtigung standörtlicher Gegebenheiten, welche in einem engen Zusammenhang mit Flächenausweisungen stehen. In diesem Beitrag werden Techniken des Data-Mining vorgestellt, die sich zur Dateninspektion von Messgrößen der Siedlungs- und Freiraumentwicklung auf Ebene der Gemeinden in Deutschland ($n = 11\,441$) eignen. Auf diese Weise lassen sich statistische Eigenschaften der Messgrößen präzisieren und darauf aufbauend räumliche Charakteristika darstellen. Die Ergebnisse der Dateninspektion eignen sich als Grundlage für weiterführende mehrdimensionale Raumbeschreibungen. Perspektivisch können diese Techniken wichtige analytische Beiträge im Zuge eines raumbezogenen Monitoringsystems leisten, welches Messgrößen der Siedlungs- und Freiraumentwicklung beobachtet und bewertet.

1 Einführung

Das Spannungsfeld zwischen der anhaltend hohen Umwidmung von bislang nicht baulich genutzten Flächen in Siedlungs- und Verkehrsflächen (= Flächeninanspruchnahme) und dem Leitbild einer nachhaltigen Flächenentwicklung ist in der nationalen und internationalen raumwissenschaftlichen Forschung (u. a. REFINA 2006-2012; FONA 2009-2015) sowie der raumentwicklungspolitischen Diskussion von hoher Relevanz (Nationale Strategie zur biologischen Vielfalt, Deutsche Anpassungsstrategie an den Klimawandel, Nationale Nachhaltigkeitsstrategie, DB 2011; MKRO 2010). „Im Gegensatz zu Wasser oder Biomasse sind Böden endliche, nicht erneuerbare, im Wesentlichen nicht vermehrbare, ökologisch sensible und unter großem technischen und finanziellen Aufwand wiederherstellbare Naturgüter“ (Lexer 2005). Aufgrund der zunehmenden Verfügbarkeit georeferenzierter Daten sowie Sachdaten der amtlichen Statistiken haben sich die Möglichkeiten zur Untersuchung von Eigenschaften der Siedlungs- und Freiraumentwicklung und ihrer Einflussgrößen in den vergangenen Jahren wesentlich verbessert. Data-Mining-Methoden aus den Bereichen Statistik, Maschinelles Lernen, Datenbanken und Visualisierungen eignen sich, um den Prozess des Auffindens verborgener und interessanter Informationen in großen Datenbeständen adäquat zu unter-

stützen. Trotz allem befinden sich derartige Analysen noch in einer initialen Phase. Der Beitrag verfolgt das Ziel, Techniken der Dateninspektion in ihrer Anwendung am Beispiel der Messgröße Tagbevölkerungsdichte vorzustellen.

2 Daten: Messgröße Tagbevölkerungsdichte

Die Messgröße Tagbevölkerungsdichte wird für die vorgestellten Analysen aus dem Quotienten der Tagbevölkerung und der Summe der Gebäude und Freifläche in einer Gebietseinheit berechnet. Gemessen wird ein Dichtewert in Personen pro Hektar. Die Tagbevölkerungsdichte wird in diesem Beitrag auf Ebene der Gemeinden berechnet. Die Tagbevölkerung entspricht der Summe der Einwohner und der vor Ort Beschäftigten. Die vor Ort Beschäftigten (Pendlersaldo) werden aus der Differenz von sozialversicherungspflichtig Beschäftigten am Arbeitsort und sozialversicherungspflichtig Beschäftigten am Wohnort ermittelt. Die Daten zur Gebäude- und Freifläche entstammen der amtlichen Flächenerhebung (FETN). Die Datenqualität dieser statistischen Daten wird in zahlreichen Publikationen differenziert diskutiert (siehe u. a. Deggau 2008; Krüger et al. 2013). Das weitere statistische Datenmaterial nimmt Bezug auf die amtliche Statistik der sozialversicherungspflichtig Beschäftigten (siehe Beschäftigte am Wohnort und Beschäftigte am Arbeitsort) sowie die amtliche Bevölkerungsstatistik. Es sei erwähnt, dass es sich um gebietsstandsbereinigte Gemeindedaten ($n = 11\,441$, Stand: 31.12.2010) handelt. Gemeindefreie Gebiete wurden aus dem Datensatz eliminiert ($n = 228$). Gedankt sei an dieser Stelle dem Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR), welches die umgeschätzten Datenreihen für die quantitative Analyse ausgewählter Messgrößen der Flächeninanspruchnahme zur Verfügung gestellt hat.

3 Methodische Schritte und ausgewählte Ergebnisse

Die Vorgehensweise zur Entdeckung von räumlichen Mustern auf Ebene der deutschen Gemeinden folgt der Methodik des Data-Minings bzw. Knowledge Discovery, welche bereits anderweitig ausführlicher vorgestellt wurde (Ultsch et al. 2015; Behnisch, Ultsch 2015; Ultsch, Behnisch 2009). Dieser Beitrag fokussiert auf Techniken der Dateninspektion, die nachfolgend in aufeinander aufbauenden Schritten vorgestellt werden.

3.1 Untersuchung einer einzelnen Messgröße

Zu Beginn steht die Analyse einer jeden Messgröße für sich, mit dem Ziel, den Typus der Verteilung zu ermitteln, dem die Messgröße grundsätzlich folgt (De Gruijter et al. 2006). Die häufigste Annahme ist eine Normalverteilung der Daten als Konsequenz des zentralen Grenzwertsatzes. In Bezug auf die Verteilungsuntersuchung einzelner Messgrößen verwendet die Mehrzahl der quantitativen Studien Histogramme, die z. B. auch als

Standardmethode in der Mehrzahl der Geografischen Informationssysteme implementiert sind. Zu beachten ist dabei, dass die Darstellung eines Histogramms je nach Wahl der Intervallgrenzen bzw. Anzahl der Bins unterschiedlich ausfällt (Scott 1979).

Eine systematischere Untersuchung des Verteilungsverlaufes wird durch nachfolgend beschriebene Visualisierungswerkzeuge unterstützt: dem Quantile-/Quantile-Plot (QQplot) und der Kerndichteschätzung der Wahrscheinlichkeitsdichte (probability density function, pdf). In diesem Beitrag wird hierfür die Pareto Density Estimation (PDE-plot) verwendet. Dieser Kerndichteschätzer ist insbesondere dazu entwickelt worden, um eventuell vorhandene Teilkomponenten einer Verteilung entdecken zu können (Ultsch 2005). Ein QQplot vergleicht zwei Verteilungen, indem die jeweiligen Quantile gegeneinander aufgetragen werden. Falls die zwei Verteilungen vom gleichen Typ sind, zeigt der QQplot eine Gerade. Falls die Verteilung nicht normalverteilt ist und auf eine schiefe Verteilung hindeutet, so ist es hilfreich, sogenannte Standardtransformationen aus Tukey's „ladder of power“ (Tukey 1977) anzuwenden. In Bezug auf die Wertausprägungen der Tagbevölkerungsdichte wurde erwartet, dass eine kleine Anzahl von Gemeinden über hohe Tagbevölkerungsdichten verfügt und die Mehrzahl der Gemeinden durch eher niedrige bis moderate Dichten gekennzeichnet ist. Erwartet wurde eine schiefe Verteilung. Mithilfe der genannten Techniken konnte diese Vermutung bestätigt werden. Wird die Messgröße z. B. logarithmiert $\text{LogTagbevölkerungsdichte} = \log_{10}(\text{Tagbevölkerungsdichte}_{2010})$, so entsteht in erster Näherung eine Normalverteilung. Der QQplot in Abbildung 1 vergleicht die Verteilung von $\text{LogTagbevölkerungsdichte}$ mit der Normalverteilung. Daraus kann man ablesen, dass $\text{LogTagbevölkerungsdichte}$ näherungsweise normalverteilt ist. Lognormalverteilungen sind häufig in den Natur- und Umweltwissenschaften anzutreffen. Sie weisen in der Regel auf Wachstumsprozesse oder Konzentrationsvorgänge hin. Die Lognormalverteilung der Daten wird auch durch den PDEplot mit einer eingezeichneten Gaussverteilung (schwarz dargestellt in Abb. 2) bestätigt. Ist eine Einteilung in Klassen für die Messgröße angedacht, so lassen sich aufgrund der erkannten Normalverteilung folgende Klassengrenzen unter Berücksichtigung von Mittelwert (m) und Standardabweichung (s) rechtfertigen: $m-2s$, $m-s$, m , $m+s$, $m+2s$. Für die Tagbevölkerungsdichte ergeben sich folgende Klassengrenzen: 6,11, 18, 29 und 47 (Personen/ha).

Abweichungen von der Normalverteilung (grau in Abb. 2) zeigen sich in den Randbereichen und deuten auf weitere Komponenten. Diese Komponenten lassen sich über sogenannte Mischungsmodelle in ihrem Verteilungsverlauf genauer abschätzen, z. B. als Summe von zusammengesetzten Verteilungen. Ergebnisse von Mischungsmodellen werden detaillierter im Zuge der Untersuchungen von Veränderungen der Messgrößen über die Zeit beschrieben (Abschnitt 3.3). Die räumliche Verteilung der Klassifizierung zur Tagbevölkerungsdichte (Abb. 3) zeigt deutliche Überlagerungen mit bereits bekannten Mustern der Siedlungsdichte. Hohe Dichten sind zum einen in Agglomerationsräu-

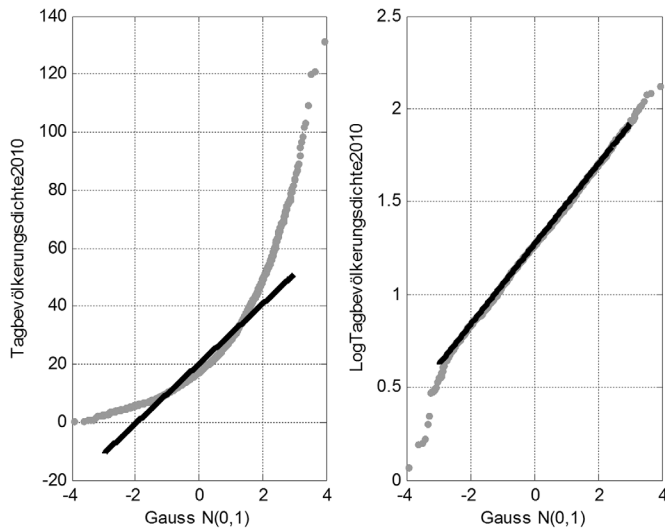


Abb. 1: QQplot. Tagbevölkerungsdichte (links) und LogTagbevölkerungsdichte (rechts) vs. Gauß-Verteilung (Quelle: eigene Berechnung)

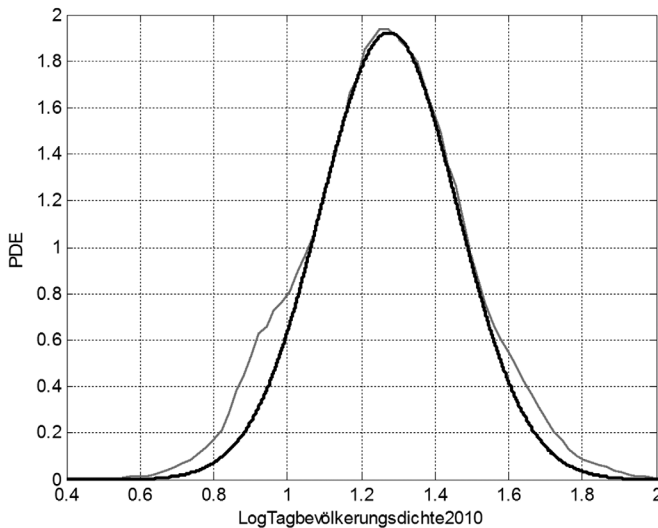


Abb. 2: PDEplot: LogTagbevölkerungsdichte 2010 (grau) mit zentralen Gauß (schwarz) (Quelle: eigene Berechnung)

men zu finden. Zum anderen weisen auch einige ländliche Regionen (z. B. Schwarzwald, Alpenvorland, Vogtland) hohe Tagbevölkerungsdichten auf. Niedrige Dichten treten vor allem in den nordöstlichen Regionen auf, wobei dort auch die Gegensätze zwischen sehr hoher und sehr niedriger Dichte enger beieinander liegen als im Südwesten.

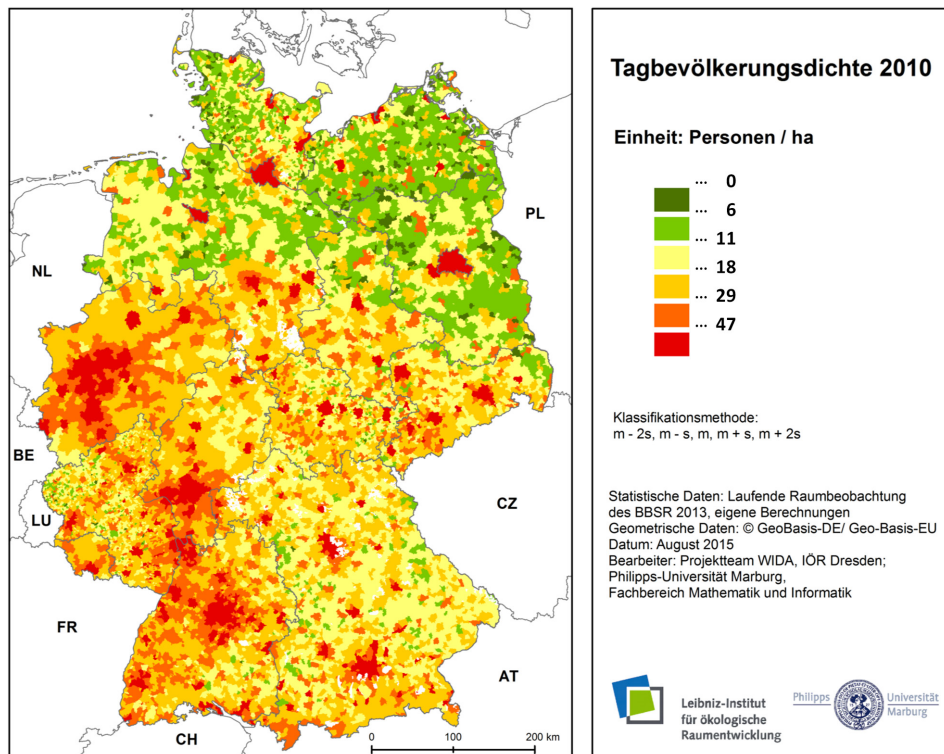


Abb. 3: Klassifizierung zur Tagbevölkerungsdichte (Quelle: eigene Berechnung)

3.2 Messen der Veränderung

Im Zuge der Dateninspektion lassen sich aufbauend auf den Erkenntnissen aus der Untersuchung einer einzelnen Messgröße zu einem festen Zeitpunkt auch Ausprägungen von Messgrößen über die Zeit betrachten.

Die Pareto Density Estimation (PDEplot) eignet sich ebenfalls, um Verteilungen einer Messgröße m zu verschiedenen Zeitpunkten im Vergleich darzustellen. Abbildung 4 zeigt die PDEplots für die Messgröße Tagbevölkerungsdichte der Jahre 2000 und 2010. Der Mittelwert der Messgrößen ist als Lagemaß mit ausgewiesen. Es zeigt sich, dass die Tagbevölkerungsdichte in Deutschland zwischen 2000 und 2010 insgesamt abgenommen hat ($m_{10} = 17,6$ und $m_{00} = 21,7$).

Für den Zweck der Untersuchung der zeitlichen Veränderung von Messgrößen der Siedlungs- und Freiraumentwicklung wird in nationalen (u. a. Dech, Klein 2009; BBSR 2012) wie auch internationalen Projekten (u. a. Shrestha et al. 2012) häufig die relative Veränderung in Prozent (RPV) verwendet. Die Verwendung von RPV ist aus verschiedenen Gründen problematisch: Der Wertebereich ist nicht symmetrisch (i), RPV reicht von

-100 % hin zu unendlich, die zeitliche Kombination von RPV ist problematisch (ii) und RPV verfügt über einige weitere numerische Probleme (iii) (Ultsch 2009).

In diesem Beitrag wird die relative Differenz (RelDiff) zur Messung der Veränderung verwendet. Hier wird die Veränderung der Messgröße als Delta m (Δm) in Relation zum Durchschnitt der Messwerte m_0 (zum Zeitpunkt t_0) und m_1 (zu einem späteren Zeitpunkt t_1) gesetzt. $\text{RelDiff}(m) = 100 \cdot \Delta m / \text{mean}(m)$, mit Mittelwert $(m) = 0.5 \cdot (m_0 + m_1)$. $\text{RelDiff}(m)$ hat einen symmetrischen Wertebereich zwischen -200 % und +200 %. Im Bereich von -25 % und +25 % sind RelDiffs praktisch (mit einer kaum messbaren Fehlerquote) äquivalent zu RPV (Ultsch 2009).

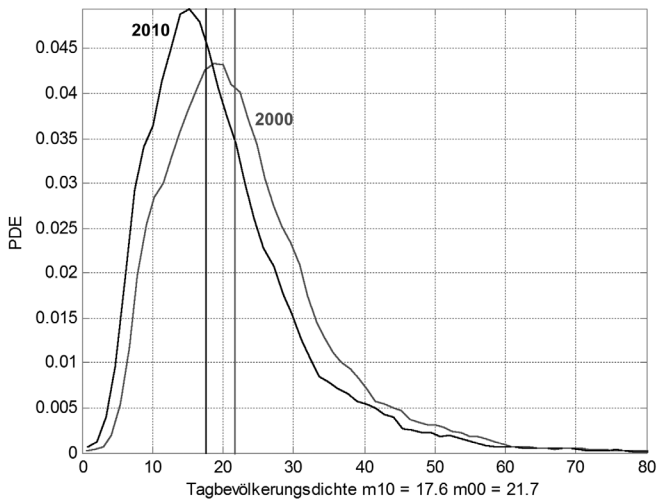


Abb. 4: PDEplot für die Veränderung der Tagbevölkerungsdichte, in Schwarz dargestellt ist das Jahr 2010 und in Grau das Jahr 2000 (Quelle: eigene Berechnung)

Eine wichtige Überlegung in Bezug auf die Beobachtungen zur Veränderung einer Messgröße m über die Zeit ist die Abhängigkeit von $\text{RelDiff}(m)$ vom Ausmaß bzw. der Größenordnung von m . Um auf diesen Aspekt einzugehen, wird ein sogenannter RAplot verwendet (siehe dazu auch in Ultsch et al. 2015). Der RAplot stellt die RelativeDifferen-ces R ($\text{RelDiff}(m)$) mit dem Logarithmus der durchschnittlichen Wertausprägungen der Messgröße über die Zeit $D(\log(\text{mean}(m)))$ in Beziehung. Abbildung 5 zeigt einen solchen RAplot für die Dynamik der Tagbevölkerungsdichte. Die lineare Interpolation (schwarz in Abb. 2) zeigt für die Tagbevölkerungsdichte, dass die Veränderung leicht abhängig vom Ausmaß bzw. der Größenordnung der Tagbevölkerungsdichte ist. Bei einem Wert der Tagbevölkerungsdichte von 10 ($A = 1$) liegt beispielsweise die Veränderung RelDiff bei -20 % und bei einem Wert der Tagbevölkerungsdichte von 100 ($A = 2$) liegt die Veränderung RelDiff nahezu bei Null. Die Varianz von RelDiff Tagbevölkerungsdichte nimmt mit höheren Werten der Tagbevölkerungsdichte ab.

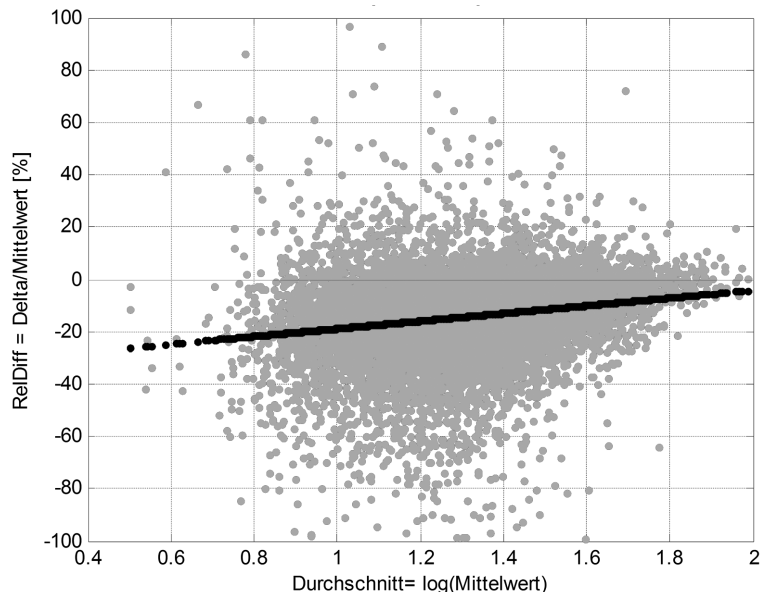


Abb. 5: RAplot für die Veränderung der Tagbevölkerungsdichte, schwarz dargestellt ist eine lineare Interpolation (Quelle: eigene Berechnung)

3.3 Klassifizierung der Veränderung

Eine qualitative Untersuchung von Messgrößen der Veränderung stützt sich auf eine genauere Inspektion der Verteilungen mittels Pareto Density Estimation (PDE) (siehe Abschnitt 3.1). Falls eine Messgröße nicht einem einfachen Gauß folgt, so wird die Verteilung als eine Summe von Komponenten modelliert. Als Komponenten werden für die Dynamik der Messgröße Tagbevölkerungsdichte (RelDiffs) eine zentrale Gaußverteilung (N) und zwei an den Außenseiten befindliche LogNormal (LN) Verteilungen verwendet.

Die relative Differenz folgt einem zentralen Gauß mit einem Mittelwert von -12.69 %. Daraus ergibt sich der Erwartungswert der Tagbevölkerungsdichte in ihrer Dynamik von 2000 bis 2010. Die Ausläufer der Verteilung lassen sich geeignet als LogNormal (LN) Verteilung modellieren. Das Modell kann validiert werden mithilfe von einem statistischen Test, zum Beispiel, dem Chi-Quadrat Test. Ein QQplot eignet sich zusätzlich, um die Eignung des Modelles zu beurteilen (siehe Abschnitt 3.1). Ein solches Modell ermöglicht unter Berücksichtigung des Bayes' theorem (e. g. Lee 2012) die Berechnung von posterior Wahrscheinlichkeiten und die Bestimmung von Entscheidungsgrenzen für die Klassifikation der Messgröße. Die Messgröße RelDiffTagbevölkerungsdichte kann z. B. in drei aussagekräftige Klassen unterteilt werden. Die Klassifizierungsgrenzen (-30 und 12) entsprechen den Schnittpunkten der Wahrscheinlichkeitsdichtefunktionen des Modells (vertikale Linien in Abb. 6).

Die Bayes Posteriors können dazu verwendet werden, um eine qualitative Skalierung der Messgrößen vorzunehmen: Die Differenz (DP) zwischen den Posterior der rechten Komponente abgezogen von der linken Komponente skaliert die Werte in das Intervall -100 bis 100, so dass DP -100 (100) beträgt, wenn die Messgröße eindeutig zur linken oder (rechten) Komponente gehört. DP beträgt 0, wenn die Wertausprägung der Messgröße eindeutig zur zentralen Komponente des Modells zuzuordnen ist.

Abbildung 6 zeigt die qualitative Skalierung für die Messgröße Tagbevölkerungsdichte und ihre Veränderungswerte. Die Klassenzugehörigkeit ist farblich gekennzeichnet.

Das Model erlaubt eine klare Interpretation: die Mehrheit – ca. 83 % der Gemeinden (i. e. die prior Wahrscheinlichkeit der zentralen Komponente) – zeigen eine Veränderung der Tagbevölkerungsdichte, die „wie erwartet“ ist (gelb in Abb. 6). „Wie erwartet“ entspricht einem Anstieg der Tagbevölkerungsdichte von $m = -12.69$ % mit einer Standardabweichung von $s = 10.17$. Diese Werte beziehen sich auf den zentralen Gauß (-12.69, 10.17). Eine Teilmenge von 3 % der Gemeinden zeigt eine Veränderung der Tagbevölkerungsdichte, die „mehr als erwartet“ ist (rot in Abb. 6). „Mehr als erwartet“ bedeutet größer als 12 % und folgt einer LogNormal Verteilung. Eine Teilmenge von 15 % der Gemeinden zeigt eine Veränderung der Tagbevölkerungsdichte, die „geringer als erwartet“ ist (grün in Abb. 6). D. h. weniger als -30 % und die Verteilung folgt einer LogNormal Verteilung.

Abbildung 7 zeigt eine Karte aller Gemeinden bezüglich deren Zugehörigkeit zu den drei Klassen. Erkennbar ist ein sehr heterogenes räumliches Muster, in dem Gemeinden aller drei Klassen teilweise sehr dicht beieinander liegen. Dennoch fällt beispielsweise auf, dass Gemeinden, deren Veränderung in der Tagbevölkerungsdichte zwischen 2000 und 2010 geringer (Prozesse der Entdichtung) ist, als es erwartet wurde, gehäuft in den neuen Bundesländern sowie in den ländlichen Gemeinden von Rheinland-Pfalz, Bayern und Schleswig-Holstein vorkommen.

Hervorzuheben ist, dass die Lokalisierung von Ausprägungen der Messgrößen nicht auf subjektiv festgelegten Klassengrenzen oder reinem Expertenwissen basiert. Vielmehr bilden die Techniken der Dateninspektion eine Charakteristik in den Daten ab und liefern wichtige Entscheidungsgrundlagen für die Wahl der Klassengrenzen.

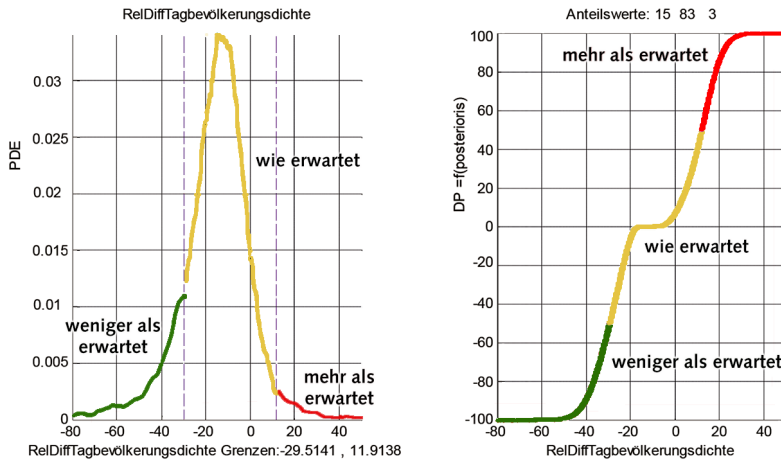


Abb. 6: Qualitative Skalierung für die Messgröße Tagbevölkerungsdichte auf Grundlage eines Gauß Mixtur Modells (Quelle: eigene Berechnung)

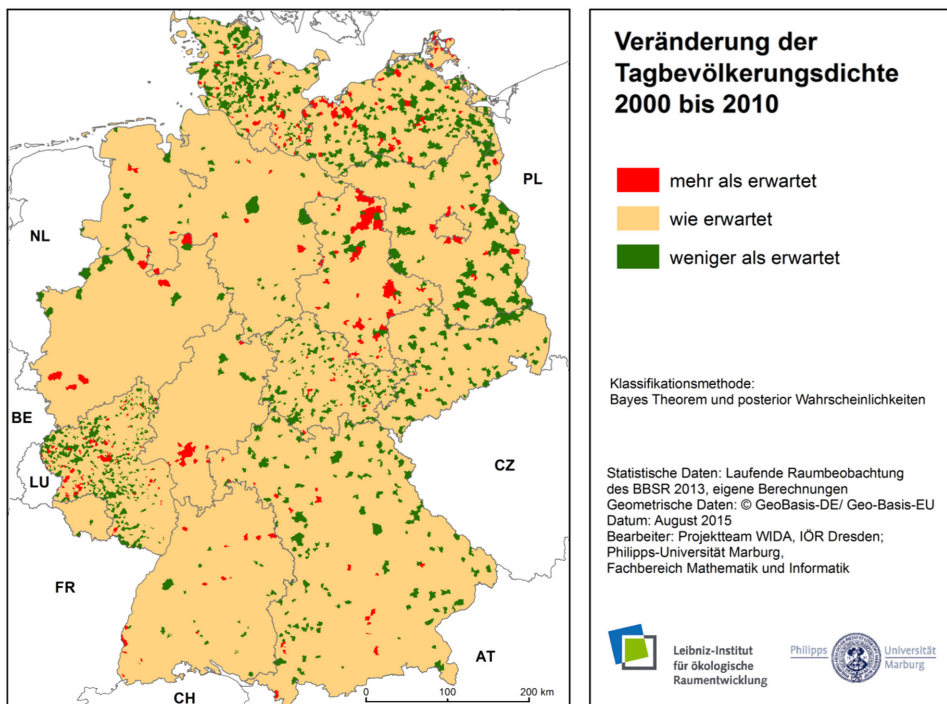


Abb. 7: Lokalisierung der Klassifizierung zur Dynamik der Tagbevölkerungsdichte 2010: 3 Klassen (weniger als erwartet, wie erwartet, mehr als erwartet) (Quelle: eigene Berechnung)

4 Schlussbetrachtung

Ausgangspunkt der Überlegungen zu diesem Beitrag ist das Ziel der Raumbearbeitung, Strukturen und Entwicklungen zu (er-)kennen, zu bewerten und absehbare zukünftige Entwicklungstrends in Strategien und Entscheidungen zu berücksichtigen. Mit Blick auf die stetig wachsenden Geodatenbestände in der jüngeren Vergangenheit werden bereits in der Literatur vielseitig Alternativen zur klassischen Raumanalyse diskutiert. Miller & Han (2009) betonen beispielsweise die Wichtigkeit von adäquaten Analysetechniken, die es künftig noch zu erkunden und zu erforschen gilt (Miller, Han 2009, 37-38).

Der vorliegende Beitrag stellt Data-Mining-Techniken zur Dateninspektion vor und zeigt Anwendungsbeispiele für Messgrößen der Siedlungs- und Freiraumentwicklung auf Ebene der 11 441 Gemeinden in Deutschland (Stand 31.12.2010). Die Vielzahl der zur Verfügung stehenden georeferenzierten Flächendaten sowie Sachdaten kann nur exemplarisch in Betracht gezogen werden.

Im Fokus dieses Beitrages steht die Messgröße Tagbevölkerungsdichte, gemessen für die Zeitpunkte 2000 und 2010. Soweit den Autoren bekannt, wurde die Tagbevölkerungsdichte mit derartigen Techniken auf Ebene der Gemeinden noch nicht analysiert. Es konnte im Zuge der Untersuchung zu einem festen Zeitpunkt (2010) gezeigt werden, dass die Messgröße in erster Annäherung einer Lognormalverteilung folgt. Darauf aufbauend konnte mithilfe von Mittelwert und Standardabweichung eine Klassifizierung gerechtfertigt werden. Im Zuge der Untersuchungen zur Veränderung der Messgröße über die Zeit (2000 bis 2010) konnte eine weitere Klassifizierung statistisch begründet abgeleitet werden. Mithilfe eines Gauß-Mixturmodells mit drei Komponenten (log-normal, normal, log-normal) wurden Klassengrenzen identifiziert. Die Veränderung der Tagbevölkerungsdichte folgt für 83 % der Gemeinden einer Gaußverteilung mit einem Mittelwert von -13 % und einer Standardabweichung von 10 %. Weiterhin ist eine beträchtliche Gruppe der Gemeinden durch eine Veränderung gekennzeichnet, die kleiner als -30 % ist. Denkbar ist künftig die Verwendung von Dreijahresmittelwerten zum Zeitpunkt 2000 und 2010, um den zeitlichen Vergleich über eine Dekade weiter zu präzisieren bzw. um die kurzfristigen Schwankungen von Messgrößen (z. B. Einwohner, Arbeitsplätze usw.) abzufangen.

Zusammenfassend ist festzustellen, dass eine umfassende und detaillierte Dateninspektion die Weichen für die Entdeckung neuer Zusammenhänge in Datensammlungen stellt. Dies auch in Bezug auf weitere Arbeitsschritte im Zuge eines umfassenden Knowledge Discovery in Databases Prozesses (Ultsch et al. 2015; Behnisch, Ultsch 2015; Ultsch, Behnisch 2009). Aus Sicht der Autoren dieses Beitrags eignen sich die hier vorgestellten Techniken, um die Aussagekapazität von bestehenden Instrumenten der Raumbearbeitung und des Flächennutzungsmonitoring zu stärken. Dies sowohl in Bezug auf die statistischen als auch die räumlichen Ausprägungen von Messgrößen der Siedlungs- und Freiraumentwicklung.

5 Literatur

- BBSR – Bundesinstitut für Bau-, Stadt- und Raumforschung (Hrsg.) (2012): Raumordnungsbericht 2011. Bonn: Bundesinstitut für Bau-, Stadt- und Raumforschung (self-publishing).
- Behnisch, M.; Ultsch, A. (2015): Knowledge Discovery in Spatial Planning Data – A Concept for Cluster Understanding. In: Helbich, M.; Arsanjani, J. J.; Leitner, M. (Hrsg.): Computational Approaches for Urban Environments. Berlin: Springer (Geotechnologies and the Environment).
- Behnisch, M.; Ultsch, A. (2009): Urban data mining: spatiotemporal exploration of multidimensional data: In: BuildingResearch & Information, 37(5-6)/2009.
- Dech, S.; Klein, R. (2009): Entwicklung und Evaluierung eines fernerkundungsbasierten Flächenbarometers als Grundlage für ein nachhaltiges Flächenmanagement. Schlussbericht zum Verbundvorhaben (self-publishing).
- Deggau, M. (2008): Die amtliche Flächenstatistik – Grundlage, Methode, Zukunft. In: Meinel, G.; Schumacher, U. (Hrsg.), Flächennutzungsmonitoring – Grundlagen, Statistik, Indikatoren, Konzepte. Aachen: Shaker Verlag.
- De Gruijter, J.; Brus, D.; Bierkens, M.; Knotters, M. (2006): Sampling for Natural Resource Monitoring. Berlin, Heidelberg: Springer-Verlag.
- Dudoit, S.; Yang, Y. H.; Callow, M. J.; Speed, T. P. (2002): Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat. Sin, 12(1)/2002, 111-139.
- Krüger, T.; Meinel, G.; Schumacher, U. (2013): Land-use monitoring by topographic data analysis. Cartography and Geographic Information Science. doi: 10.1080/15230406.2013.809232.
- Lee, P. M. (2012): Bayesian Statistics: An Introduction. Chichester: Wiley.
- Lexer, W. (2004): Zerschnitten, versiegelt, verbaut? – Flächenverbrauch und Zersiedelung versus nachhaltige Siedlungsentwicklung. In: BMLFUW (Hrsg.): Tagungsband „grünstadtgrau“. Fachtagung zur Stadtökologie. 21-22. Oktober 2004, Wien, 35-45.
- Miller, H. J.; Han, J. (2009): Geographic Data Mining and Knowledge Discovery. 2nd ed. London: Chapman & Hall.
- Scott, D. W. (1979): On optimal and data-based histograms. Biometrika 66(3)/1979, 605-610.
- Shrestha, M. K.; York, A. B.; Boone, C. G.; Zhang, S. (2012): Land fragmentation due to rapid urbanization in the Phoenix Metropolitan Area: Analyzing the spatiotemporal patterns and drivers, Applied Geography, 32(2)/2012, 522-531.
- Tukey, J. W. (1977): Exploratory Data Analysis. Reading, MA: Addison-Wesley.

- Ultsch, A. (2005): Pareto Density Estimation: A Density Estimation for Knowledge Discovery, In: Baier, D.; Wernecke, K. D. (Hrsg.), *Innovations in Classification, Data Science, and Information Systems (Proceedings of the 27th Annual Conference of the Gesellschaft für Klassifikation e. V. Brandenburg University of Technology, Cottbus, March 12-14, 2003)*.
- Ultsch, A. (2009): Is log ratio a good value for measuring return in stock investments? In: Fink, A.; Lausen, B.; Seidel, W.; Ultsch, A. (Hrsg.): *Advances in Data Analysis, Data Handling and Business Intelligence (Proceedings 32nd Annual Conference of the German Classification Society, Helmut-Schmidt-University, Hamburg, July 16-18, 2008)*.
- Ultsch, A.; Kretschmer, O.; Behnisch, M. (2015): Systematic Data-Mining into Land Consumption in Germany. In: *Proceedings of 14th International Conference on Computers in Urban Planning and Urban Management, July 7-10, 2015, Cambridge, MA USA, 301-1, 301-19*.