



Flächennutzungsmonitoring V Methodik – Analyseergebnisse – Flächenmanagement

IÖR Schriften Band 61 · 2013

ISBN: 978-3-944101-18-7

Zur Erzeugung hochauflösender datenschutzkonformer Mischrasterkarten

Markus Dießelmann, Gotthard Meinel

Dießelmann, M.; Meinel, G. (2013): Zur Erzeugung hochauflösender datenschutzkonformer Mischrasterkarten. In: Meinel, G.; Schumacher, U.; Behnisch, M. (Hrsg.): Flächennutzungsmonitoring V. Methodik – Analyseergebnisse – Flächenmanagement. Berlin: Rhombos, IÖR Schriften 61, S. 189-197.

Zur Erzeugung hochauflösender datenschutzkonformer Mischrasterkarten

Markus Dießelmann, Gotthard Meinel

Zusammenfassung

Die zunehmende Verfügbarkeit adressbezogener Daten im Zusammenhang mit der Nutzung geometrischer Raster zur Raumuntergliederung haben die Voraussetzungen für kleinräumige Analysen deutlich verbessert. Bei der Verwendung personenbezogener Daten müssen datenschutzrechtliche Vorgaben eingehalten werden, falls die Rasterzellen zu wenig Fallzahlen enthalten. Vielfach werden diese Rasterzellen ausgeblendet, wodurch Informationen in der Karte verloren gehen.

Eine datenschutzkonforme Alternative stellt die Aggregation von Rasterzellen dar, bis die Fallzahlen einen vorgegebenen Grenzwert überschreiten. In diesem Beitrag werden Möglichkeiten vorgestellt und bewertet, nach denen sich datenschutzkonforme Mischrasterkarten erzeugen lassen. Besonderes Augenmerk wird auf die Auflösungsverluste der erzeugten Mischrasterkarten gelegt, um geeignete Datengrundlagen für kleinräumige Analysen zu schaffen.

1 Einführung

Hinsichtlich Darstellung und Analyse raumbezogener Daten eignen sich Raumgliederungen, die auf administrativen Gebietseinheiten basieren, nur bedingt (Szbalski 2006). Zwar sind sie aufgrund der Gebietsgröße bei der Nutzung personenbezogener Daten datenschutzrechtlich in der Regel unbedenklich, jedoch lassen sie keine kleinräumigen Analysen zu. Besonders in mittleren und größeren Maßstabsbereichen wie bspw. in der Stadt- und Regionalplanung, aber auch in privatwirtschaftlichen Anwendungen (Geomarketing usw.) sind kleinräumige Auswertungen von zunehmender Bedeutung.

Kleinräumige Auswertungen setzen entsprechend hochaufgelöste Daten voraus. In der Regel sind dies adressbezogene Daten, die von Meldebehörden (Einwohnermeldeamt, Bundesagentur für Arbeit) erhoben und fortgeschrieben werden. Andererseits haben in den letzten Jahren auch privatwirtschaftliche Unternehmen eine Vielzahl u. a. auch soziodemographischer Daten in hoher Auflösung erhoben. Besonders im Rahmen des Geomarketing geschieht dies mit einer Vielzahl von Individualdaten mit Standortbezug (Dickmann, Sohst 2008). Um diese Daten möglichst kleinräumig abzubilden, sind Alternativen zu administrativen Raumgliederungen nötig. Hierbei sind insbesondere datenschutzrechtliche Vorgaben einzuhalten.

2 Geometrische Raster zur Raumuntergliederung

Auf die Nachteile von administrativen Raumuntergliederungen in der Kartographie wurde bereits frühzeitig hingewiesen. So stellten Witt (1970) und Arnberger (1977) fest, dass räumliche Verteilungen und damit regionalstatistische Unterschiede innerhalb administrativer Gebietseinheiten nicht erkennbar sind und Vergleiche aufgrund der unterschiedlichen Flächenausdehnung nur bedingt möglich sind.

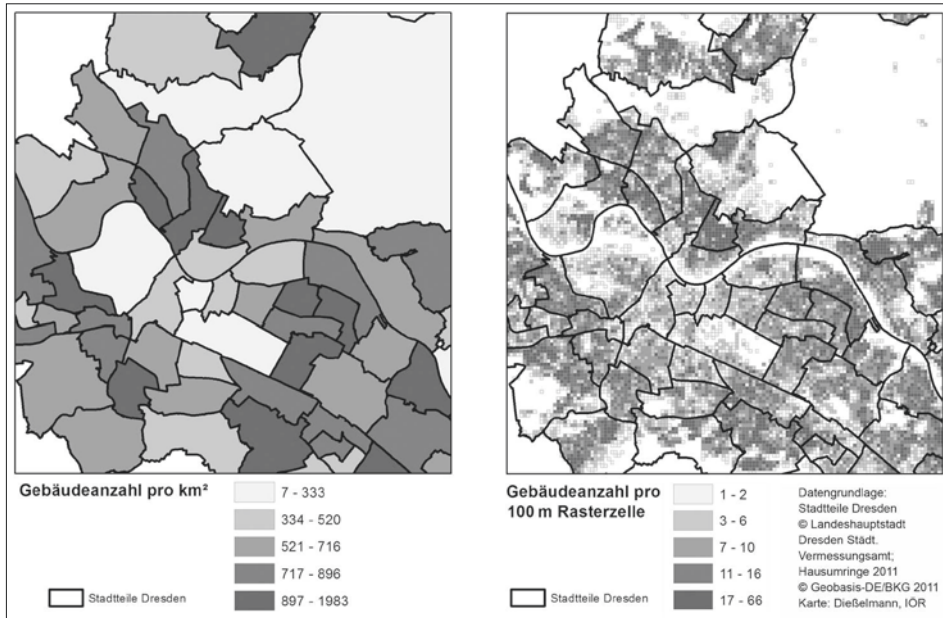


Abb. 1: Vergleich verschiedener Raumgliederungen und der Verteilung des abgebildeten Merkmales. Links: Gebäudedichte pro Stadtteil. Rechts: Gebäudeanzahl pro 100-m-Rasterzelle (Quelle: eigene Darstellung)

Besonders Karten, die Ergebnisse auf Basis administrativer Gebietseinheiten mit einer starken Inhomogenität der Verteilung des zu untersuchenden Merkmales darstellen, können zu Interpretationsfehlern führen, da eine gleichmäßige Verteilung suggeriert wird, die häufig nicht vorhanden ist (Dickmann, Sohst 2008) (s. Abb. 1).

Da die Grenzen administrativer Gebietseinheiten nach anderen Kriterien und nicht nach dem Gesichtspunkt des zu untersuchenden statistischen Sachverhaltes festgelegt wurden (Arnberger 1977; Wonka 2010), wird eine Raumgliederung gesucht, die neutral ist und für viele Zwecke genutzt werden kann. Diese Anforderungen sind mit geometrischen Rastern umsetzbar (Arnberger 1977; Witt 1970). Aufgrund der dem Maßstab anpassbaren Rasterzellengröße lassen sich auch kleinräumige Darstellungen und Analysen umsetzen. Adressbezogene- und andere Daten lassen sich so auf die Rasterzellen des geometrischen Rasters aggregieren, speichern und abbilden. Ist die Lage des Rasters ortsfest und die Rasterzellgröße genau definiert, wie bspw. bei dem

von der europäischen Richtlinie INSPIRE definierten Rastern, sind zeitliche und räumliche Vergleiche möglich. Enthält die Rasterkarte verschieden große Rasterzellen, z. B. durch die Aggregation von Rasterzellen, spricht man von Mischrasterkarten (Meyer 2008).

3 Datenschutzerfordernungen

Handelt es sich bei den abzubildenden Daten um personenbezogene, ist § 3 Abs. 1 des Bundesdatenschutzgesetz (BDSG) zu berücksichtigen. Personenbezogene Daten werden danach als „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlichen Person (Betroffener)“ definiert. Ein Personenbezug lässt sich aber nach § 3 Abs. 6 BDSG ausschließen, wenn die Daten anonymisiert werden. Anonymisieren ist „...das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbarer natürlichen Person zugeordnet werden können“. Die datenverarbeitende Stelle darf nicht in der Lage sein, die Daten zu reidentifizieren.

Wenn Daten beispielweise in einem geometrischen Raster auf Rasterzellen aggregiert wurden stellt sich die Frage, wann eine Aggregation im Rahmen der Anonymisierung nach § 3 Abs. 6 BDSG ausreichend ist. Dafür existieren verschiedene Empfehlungen wie bspw. im 3. Geo-Fortschrittsbericht der Bundesregierung (2012), vom RatSWD (2012) oder in einer Studie des ULD aus dem Jahr 2008. Demnach werden als Grenzwerte drei bis vier Fallzahlen für Personen, Haushalte oder Grundstücke vorgeschlagen. An diesen Empfehlungen orientiert sich auch der Grenzwert, der in diesem Beitrag Berücksichtigung findet.

4 Aggregationsmethoden zur Erzeugung von Mischrasterkarten

Die Aggregation von unterbesetzten Rasterzellen sollte nach dem Kriterium „so detailliert wie möglich, so stark aggregiert wie (datenschutzrechtlich) nötig“ geschehen (Strobl 2005), da eine zu starke Aggregation keine kleinräumigen Analysen mehr zulässt. Die Aggregationsansätze sollen darüber hinaus sowohl für dicht als auch für weniger dicht besiedelte Gebiete anwendbar sein. Die in der Praxis häufig angewandte Ausblendung unterbesetzter Rasterzellen ist problematisch, da diese mit einem Informationsverlust verbunden ist und die Summe der Fallzahlen in der Mischrasterkarte nicht der Summe der Ausgangsrasterkarte entspricht. Nachfolgende Verarbeitungsprozesse können dadurch zu Fehlern bzw. Ungenauigkeiten führen.

Beispielhaft wird als Ausgangsrasterkarte Dresden und sein Umland mit 256 x 256 Rasterzellen (Zellgröße: 250 m x 250 m) gewählt. Aufgrund fehlender Einwohnerdaten

werden die Aggregationsmethoden in diesem Beitrag mithilfe von Gebäudedaten vorgestellt, obwohl diese eigentlich datenschutzrechtlich – ohne eine zusätzliche Attributierung wie z. B. Immobilienwert oder Energieverbrauch – unbedenklich sind. Der Grenzwert entspricht der statistischen Mindestfallzahl, die erreicht werden muss. Diese wurde hier auf > 3 Gebäude pro Rasterzelle festgelegt. Die im Folgenden beschriebenen Berechnungen werden alle im Vektormodell durchgeführt, da es gegenüber dem Rastermodell ein flexibleres Arbeiten bei der Erzeugung der Mischrasterkarten ermöglicht.

4.1 Quadtree basierte Aggregation

Strobl (2005) stellt einen Ansatz zur Erzeugung von Mischrasterkarten vor, der das Prinzip der hierarchischen Datenstrukturen von Quadtrees aufgreift. Mittels der Quadtrees lässt sich der Raum geometrisch rekursiv unterteilen. Der Ansatz orientiert sich an dem Region-Quadtree, welcher den Raum nach dem top-down-Prinzip in vier gleich große Quadranten unterteilt (Samet 1988). Allerdings wird bei der Quadtree-Aggregation nach dem bottom-up-Ansatz vorgegangen. Dies bedeutet, dass die Rasterzellen nicht weiter unterteilt, sondern iterativ in Gruppen von vier mit der Form eines Quadrates

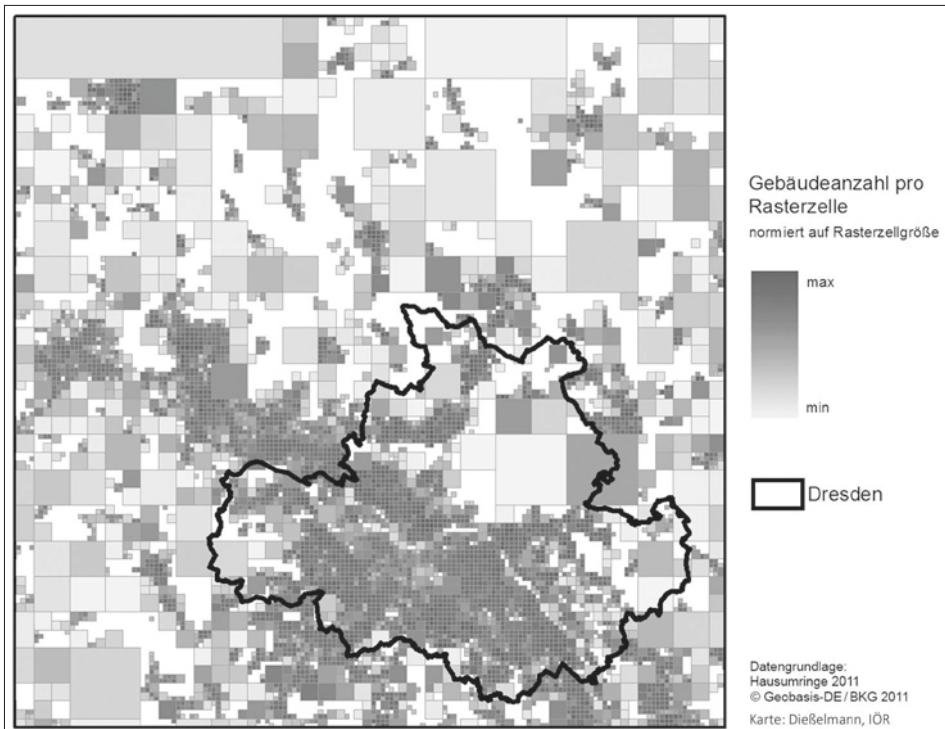


Abb. 2: Ausschnitt der Quadtree-Mischrasterkarte für das Umland von Dresden. Rasterzellen ohne Gebäude wurden ausgeblendet (Quelle: eigene Darstellung)

zusammengefasst werden, solange sie den Grenzwert unterschreiten. Damit wird eine hierarchische Rasterstruktur aufgebaut. Eine umfassende Anwendung der Quadtree-Aggregation zur Visualisierung des Gebäudebestandes findet sich u. a. in Behnisch et al. (2013).

Abbildung 2 zeigt beispielhaft die nach dem Quadtree-Prinzip erzeugte Mischrasterkarte der Gebäudeanzahl. In dicht bebauten Gebieten sind keine oder nur wenige Aggregationen notwendig, um den Grenzwert zu überschreiten, wie die hohe geometrische Auflösung im Zentrum Dresdens zeigt. So sind kleinräumige Analysen in diesen baulich hochverdichteten Bereichen weiterhin möglich. Allerdings sind in den weniger dicht besiedelten Gebieten auch sehr große Rasterzellen gebildet worden. Damit wird der Ortsbezug unscharf und kleinräumige Analysen nur bedingt durchführbar. Tabelle 1 zeigt die im Mittel starke Aggregation an Hand verschiedener Kennzahlen. Die Anzahl an Rasterzellen hat sich gegenüber der ursprünglichen Rasterkarte um über die Hälfte reduziert, die mittlere Gebäudeanzahl pro Rasterzelle mehr als verdoppelt und die Rasterzellengröße im Mittel um nahezu 20 % erhöht. Je höher der Grenzwert ist, umso mehr verstärken sich diese Effekte. Der Quadtree-Ansatz führt also zu einer lokalen Überaggregation in weniger dicht besiedelten Gebieten.

Tab. 1: Maßzahlen zur Bewertung der Quadtree-Mischrasterkarte in Abhängigkeit vom Grenzwert. Der Verhältniswert entspricht dem Quotienten aus den Werten der Mischrasterkarte und der Rasterkarte (Quelle: eigene Bearbeitung)

	Rasterkarte	Dresden Grenzwert > 3		Dresden Grenzwert > 15	
		Mischrasterkarte	Verhältniswert	Mischrasterkarte	Verhältniswert
Anzahl Rasterzellen	65536	27751	0,42	10597	0,16
Durchschnittliche Gebäudeanzahl pro Rasterzelle	7,04	16,62	2,36	43,51	6,18
Kantenlänge der Rasterzellen in m (Mittelwert)	250	294,40	1,18	393,98	1,58

4.2 Formflexible Aggregation

Die lokale Überaggregation des Quadtree-Ansatzes entsteht durch das zwingende Zusammenfassen von vier benachbarten Rasterzellen, solange der Grenzwert nicht erreicht wird. Um nun eine Überaggregation verbunden mit Auflösungsverlust zu vermeiden, wurde ein Ansatz entwickelt, der die Anzahl zu aggregierender Rasterzellen für das Erreichen des Grenzwertes reduziert. Aufgrund der sich ergebenden flexiblen Aggregationsformen, wird der Ansatz im Folgenden Formflexibler Ansatz genannt.

Ausgehend von unterbesetzten Rasterzellen wird zuerst iterativ in einer 4er- und danach in einer 8er-Nachbarschaft nach Rasterzellen gesucht, mit deren Aggregation der Grenzwert überschritten wird. Finden sich in der angrenzenden Nachbarschaft keine Rasterzellen, mit denen der Grenzwert überschritten werden kann, wird unter Zuhilfenahme einer Brückenzelle nach weiter entfernten, besetzten Rasterzellen für die Aggregation gesucht. Rasterzellen, die dann noch immer unterbesetzt sind, besitzen in der vorgegebenen Nachbarschaft keine passenden Aggregationspartner (Abb. 3a) und der Suchradius wird weiter ausgedehnt.

Dies wird mit der Berechnung eines „Relative Neighborhood Graphs“ (Abb. 3b), dessen Knoten die unterbesetzten Rasterzellen sind und der Bildung von Clustern umgesetzt. Die Kanten des Graphen werden hinsichtlich ihrer Lage und Länge ausgedünnt, bis die verbliebenen Kanten die Grundlagen der Cluster bilden (Abb. 3c). Auf dieser Grundlage werden die konvexen Hüllen berechnet (Abb. 3d), die mit dem Raster verschnitten werden, wodurch sich die Rasterzellen ergeben, die die Cluster und damit die finalen Aggregate ergeben. Diese Aggregate enthalten die Gebäudeanzahl aller in ihnen liegenden Rasterzellen. Wird der Grenzwert noch nicht überschritten, wird die nächstgelegene ausreichend besetzte Rasterzelle oder das nächstgelegene Aggregat mit in das Cluster einbezogen. Dies erfolgt auch mit unterbesetzten Einzelrasterzellen, die nicht in einen Cluster aufgenommen wurden.

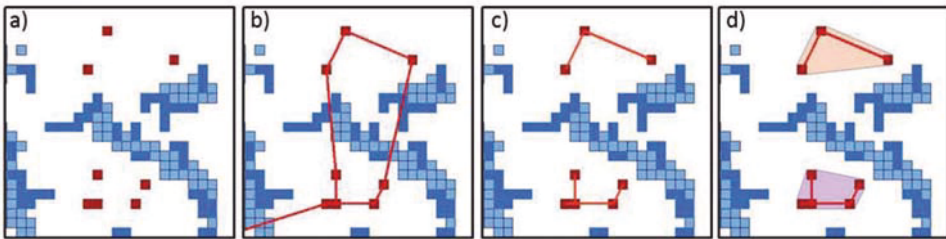


Abb. 3: Graphen-basierte Clusterbildung unterbesetzter Rasterzellen (rot). Aggregate, die im Rahmen der 4er und 8er Nachbarschaft gebildet wurden, sind dunkelblau abgebildet. Rasterzellen ohne Gebäude wurden ausgeblendet (Quelle: eigene Darstellung)

Abbildung 4 zeigt die berechnete Mischrasterkarte auf Basis des Formflexiblen Ansatzes. Im Vergleich zur Quadtree-Mischrasterkarte (s. Abb. 2) ist erkennbar, dass weniger große Aggregate gebildet wurden. Somit wird das Optimierungskriterium „so detailliert wie möglich, so stark aggregiert wie (datenschutzrechtlich) nötig“ besser umgesetzt. Dies zeigt sich auch an Hand der Maßzahlen (s. Tab. 2). Die Anzahl an Rasterzellen gegenüber der originären Rasterstruktur wurde nur um ca. 10 % verringert, gegenüber einer 60%igen beim Quadtree-Ansatz.

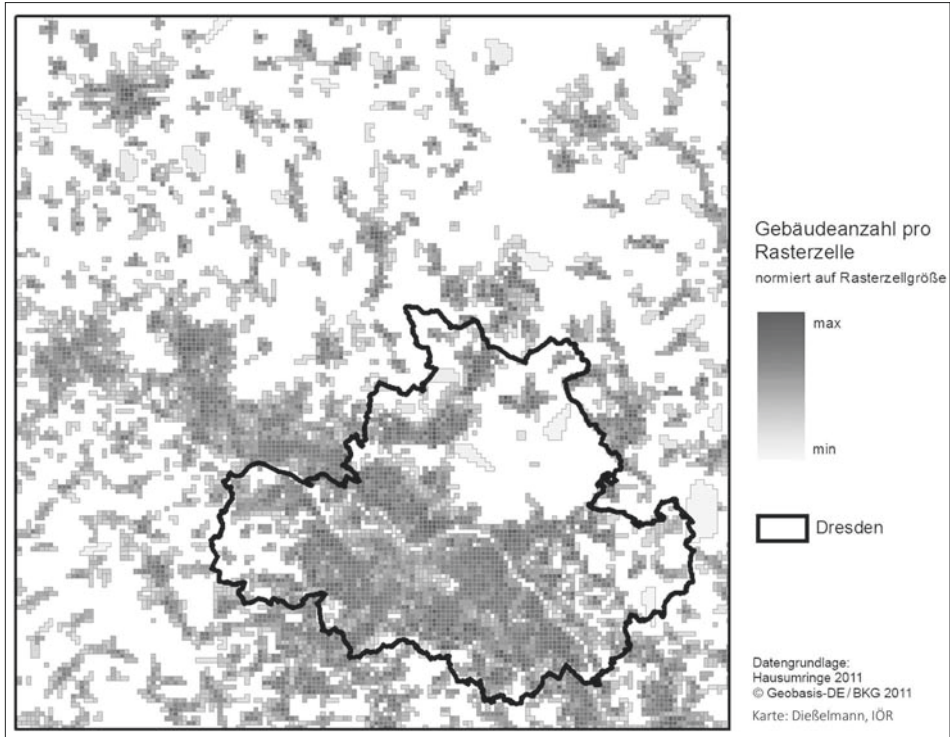


Abb. 4: Ausschnitt der Mischrasterkarte erzeugt nach dem Formflexiblen Ansatz für das Umland von Dresden. Rasterzellen ohne Gebäude wurden ausgeblendet (Quelle: eigene Darstellung)

In verdichteten Gebieten wurde der Grenzwert vielfach schon durch Aggregationen von 2 Nachbarzellen in der 4er Nachbarschaft überschritten. Auch in weniger verdichteten Gebieten konnte mit dem Formflexiblen Aggregationsansatz die Anzahl der aggregierten Rasterzellen verringert werden.

Tab. 2: Maßzahlen zur Bewertung der Aggregation des Formflexiblen Ansatzes im Vergleich zum Quadtree-Ansatz für den Grenzwert > 3. Der Verhältniswert entspricht dem Quotienten aus den Werten der Mischrasterkarte und der Rasterkarte (Quelle: eigene Bearbeitung)

	Rasterkarte	Quadtree Ansatz		Formflexibler Ansatz	
		Mischrasterkarte	Verhältniswert	Mischrasterkarte	Verhältniswert
Anzahl Rasterzellen	65536	27751	0,42	59441	0,91
Durchschnittliche Gebäudeanzahl pro Rasterzelle	7,04	16,62	2,36	7,76	1,10
Kantenlänge der Rasterzellen in m (Mittelwert)	250	294,40	1,18	260,59	1,04

5 Fazit

Das Potenzial punktbezogener bzw. kleinräumiger Geodaten lässt sich auf Basis administrativer Gebietseinheiten nicht differenziert genug nutzen. Verstärkt kommen deshalb geometrische Raster zur Raumgliederung zur Anwendung. Aufgrund des regelmäßigen Aufbaus bieten sie sich für effiziente Aufarbeitungsmöglichkeiten an. Im Rahmen dieses Beitrages konnte dies an Hand der Aggregation von Rasterzellen unter dem Kriterium des Datenschutzes zur Erstellung von Mischrasterkarten gezeigt werden. Damit kommt es zu keinem Ausschluss bzw. Verlust von Fallzahlen der Ausgangsrasterkarte in der Mischrasterkarte.

Die beiden vorgestellten Ansätze eignen sich für die datenschutzkonforme Aggregation von Rasterzellen. Die Überaggregation des Quadtree-Ansatzes konnte durch einen Formflexiblen Ansatz zur Erzeugung von Mischrasterkarten weitestgehend behoben und der Auflösungsverlust deutlich verringert werden. Zeitvergleiche auf Mischrasterkarten sind allerdings nur eingeschränkt möglich, da Größe und Form der Cluster von den jeweiligen Inhalten abhängt.

Derartige Aggregationsansätze lassen sich auch nutzen, um die Fallzahlen von gering besetzten Rasterzellen zu erhöhen und damit die statistische Sicherheit zu verbessern.

6 Danksagung

Die Autoren danken für die Unterstützung der Forschungsarbeiten zur Erstellung von Mischrasterkarten durch die Deutsche Forschungsgemeinschaft (DFG) im Projekt „Quantitative Siedlungsstrukturanalyse – Erhebung, Beschreibung, Bewertung, Visualisierung“ (ME 1592/3-1).

7 Literatur

- Arnberger, E. (1977): Thematische Kartographie. Das Geographische Seminar.
- Behnisch, M.; Meinel, G.; Tramsen, S.; Dießelmann, M. (2013): Using quadtree representation in building stock visualization and analysis. In: *Erdkunde*, 67(2)/2013.
- Bundesministerium des Innern (Hrsg.) (2012): 3. Geo-Fortschrittsbericht der Bundesregierung.
- Dickmann F.; Sohst, A. (2008): Vom Baublock zum Grid – Die Visualisierung künstlicher Raumeinheiten im Geomarketing. In: *Kartographische Nachrichten*, 5/2008.
- Meyer, W. (2008): Globalisierung und lokaler Raum in Mischrasterkarten. In: Strobl, J.; Blaschke, T.; Griesebner, G. (Hrsg.): *Angewandte Geoinformatik 2008, Beiträge zum 20. AGIT-Symposium Salzburg*.
- RatSWD – Rat für Sozial- und Wirtschaftsdaten (Hrsg.) (2012): *Georeferenzierung von Daten. Situation und Zukunft der Geodatenlandschaft in Deutschland*. Berlin.

- Samet, H. (1988): An overview of quadtree, octrees and related hierarchical data structures. In: *Theoretical Foundations of Computer Graphics and CAD*, Volume 40 Issue 1, 51-68.
- Strobl, J. (2005): Hierarchische Aggregation: Detailinformation versus Datenschutz am Beispiel adressbezogen georeferenzierter Datensätze. In: *Salzburger Geographische Arbeiten*. Band 38, Salzburg, 163-171.
- Szibalski, M. (2006): Karten in der amtlichen Statistik. In: *Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik 3/2006*.
- ULD – Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (2008): *Datenschutzrechtliche Rahmenbedingungen für die Bereitstellung von Geodaten für die Wirtschaft – Gutachten im Auftrag der GIW-Kommission*.
- Witt, W. (1970): *Thematische Kartographie, Methoden und Probleme, Tendenzen und Aufgaben*, Veröffentlichung der Akademie für Raumforschung und Landesplanung, Abhandlungen, Band 49.
- Wonka, E. (2010): *Regionalstatistik in Österreich auf der räumlichen Bezugsbasis von regionalstatistischen Rastereinheiten*. Hrsg.: Josef Strobl, Institut für Geographic Information Science, Österreichische Akademie der Wissenschaften gemeinsam mit STATISTIK AUSTRIA.