



Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungfortschreibung und Mobilfunkdaten

Sandra Hadam 

Eingegangen: 8. Mai 2022 / Angenommen: 13. März 2023 / Online publiziert: 24. April 2023
© Der/die Autor(en) 2023

Zusammenfassung Kleinräumige und aktuelle Bevölkerungszahlen sind für politische Entscheidungsfindungen unerlässlich. Die Bevölkerungfortschreibung ermöglicht die Angabe aktueller Einwohnerzahlen auf geografischer Ebene der Gemeinden. Die Einwohnerzahl wird hierbei auf Basis des Zensus 2011 anhand von Angaben der Statistiken zu Geburten und Sterbefällen sowie der Wanderungsstatistik laufend fortgeschrieben. Um den wachsenden Bedarf an kleinräumigeren Bevölkerungszahlen kurzfristig zu decken, wird die Bevölkerungfortschreibung mit einem neuen experimentellen Lösungsansatz ergänzt.

Im Projekt *Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungfortschreibung und Mobilfunkdaten* werden im Rahmen eines Verteilungsverfahrens die Ergebnisse der Bevölkerungfortschreibung von der Gemeindeebene anhand von Mobilfunkdaten bundesweit auf INSPIRE-konforme 1 × 1 km Gitterzellen umverteilt und mittels einer interaktiven Karte frei nutzbar zur Verfügung gestellt. Mobilfunkdaten bieten aufgrund ihres starken Zusammenhangs mit der Bevölkerungsverteilung sowie ihrer hohen zeitlichen und räumlichen Auflösung eine geeignete Datengrundlage für die kleinräumige Verteilung der amtlichen Einwohnerzahl. Auf Basis zusätzlicher Geodaten der deutschen Landesvermessung, wie bspw. dem Landbedeckungsmodell für Deutschland, werden die resultierenden experimentellen georeferenzierten Bevölkerungszahlen auf Plausibilität geprüft und Verbesserungen in der Mobilfunkdatenaufbereitung hergeleitet und umgesetzt. Im Ergebnis resultieren bundesweit experimentelle georeferenzierte Bevölkerungszahlen, die in urbanen, dicht besiedelten Regionen plausibel erscheinen, jedoch im ländlichen, weniger dicht besiedelten Raum zu systematischen Fehlschätzungen neigen.

✉ Sandra Hadam
Statistisches Bundesamt, Wiesbaden, Deutschland
E-Mail: sandra.hadam@destatis.de

Schlüsselwörter Mobilfunkdaten · Bevölkerungsfortschreibung · Georeferenzierung · Experimentell · INSPIRE-konform · Bevölkerungszahl

JEL-Klassifikation C81 · R23 · Y10 · Y91

Experimental georeferenced population figure based on intercensal population updates and mobile network data

Abstract Small-area and up-to-date population figures are essential for policy decision-making. The intercensal population update makes it possible to provide current population figures at the geographic level of municipalities. The number of inhabitants is continuously updated on the basis of the 2011 census using data from statistics on births and deaths and migration statistics. In order to satisfy the demand for small-area population figures in the short term, the population update is supplemented with a new experimental approach.

In the project *Experimental georeferenced population figure based on intercensal population updates and mobile network data*, the results of the intercensal population update are redistributed nationwide from the municipality level to INSPIRE-compliant 1×1 km grid cells using mobile network data and made available for free public use on an interactive map. Mobile network data offer a suitable data basis for the small-area distribution of the official population due to their strong correlation with the distribution of the population as well as their high temporal and spatial resolution. Based on additional geodata from the German Land Survey, such as the land cover model for Germany, the resulting experimental georeferenced population figures are tested for plausibility and improvements in the mobile network data processing are derived and implemented. As result, nationwide experimental georeferenced population figures are obtained, that appear plausible in urban, densely populated regions, but tend to be systematically misallocated in rural, less densely populated areas.

Keywords Mobil network data · Intercensal population update · Georeferencing · Experimental · INSPIRE-compliant · Population figure

1 Motivation

Aktuelle und valide Einwohnerzahlen sind für politische Entscheidungsfindungen unerlässlich und bspw. für den Finanzausgleich zwischen Bund und Ländern, für die Einteilung des Bundesgebietes in Wahlkreise und zur Bestimmung der Größe der Wahlbezirke oder für allgemeine Planungsaufgaben von Bedeutung (Statistisches Bundesamt 2021b). Auch bei Fragestellungen zur Verkehrsnutzung oder zur räumlichen Gesundheitsversorgung können diese von hoher Relevanz sein.

Der Bedarf an möglichst kleinräumigen Bevölkerungszahlen wird aktuell vor allem im Rahmen von georeferenzierten Auswertungen im Bereich des Gesundheitssystems, zum Beispiel hinsichtlich des Zugangs zu Gesundheitseinrichtungen, gesehen. Bspw. untersuchte Information und Technik Nordrhein-Westfalen (IT.NRW

2018) die Verteilung von Arztpraxen in Nordrhein-Westfalen anhand von georeferenzierten Einwohnerzahlen aus dem Zensus 2011. Eine weitere Verwendungsmöglichkeit wird in diesem Zusammenhang im Einbezug kleinräumiger Bevölkerungszahlen beim Krankenhaus-Atlas¹ gesehen, der deutschlandweit die Krankenhäuser unter Angabe ihrer jeweiligen Fachabteilungen, ihrer Erreichbarkeit sowie in Abhängigkeit von der kleinräumigen Bevölkerungsdichte basierend auf dem Zensus 2011 interaktiv darstellt. Hierzu ist anzumerken, dass die zeitliche Diskrepanz in der Bevölkerungsdichte zwischen dem Zensus 2011 sowie den zu einem späteren Zeitpunkt hieran anknüpfenden Forschungsfragen eine mögliche Unsicherheit bei der Interpretation der Ergebnisse darstellt, da keine aktuelleren Ergebnisse der jährlichen Bevölkerungsfortschreibung auf kleinräumiger Ebene vorlagen.

Dieses Problem findet sich auch bei Arbeiten zur Raumforschung und bei der Bewertung von Raumentwicklungen. Bei Fina et al. (2019) werden bspw. anhand kleinräumiger Bevölkerungszahlen aus dem Zensus 2011 Analysen hinsichtlich der innerstädtischen Mobilität durchgeführt, um Zusammenhänge zwischen bspw. Mobilitätsarmut und der soziodemografischen Struktur auf stadtregioener Ebene herzustellen. Sie betonen, dass eine Fortschreibung differenzierter kleinräumiger Bevölkerungsdaten für derartige Analysen aufgrund abnehmender Aktualität der Zensus-Ergebnisse erforderlich sei.

Die Bevölkerungsfortschreibung ermöglicht die Angabe aktueller Einwohnerzahlen auf geografischer Ebene der administrativen Einheiten. Die Einwohnerzahl wird hierbei auf Basis des Zensus 2011 anhand von Angaben der Statistiken zu Geburten und Sterbefällen sowie der Wanderungsstatistik laufend bzw. jährlich fortgeschrieben. Die kleinste administrative Ebene, auf der hierbei veröffentlicht werden kann, ist die Gemeindeebene. Entsprechend ist es nicht möglich Bevölkerungsdaten für nicht-administrative Einheiten zu ermitteln, die sich nicht aus Gemeinden zusammensetzen (Statistisches Bundesamt 2021b). Hierunter fallen unter anderem georeferenzierte Daten in Form von INSPIRE²-konformen Gitterzellen.

Gitterzellen sind gleich große Quadrate, die bei einer flächendeckenden, gleichmäßigen Anordnung bzw. Verteilung, ein Raster bilden. Die sogenannten INSPIRE-konformen Gitterzellen stellen ein europaweit einheitliches geographisches Gitter dar und sind dadurch unabhängig von administrativen Einheiten, wodurch sie zugleich unabhängig von Gebietsstandsänderungen sind (BKG 2020b). Demzufolge sind INSPIRE-konforme Gitterzellen zeitlich stabil und bleiben dauerhaft vergleichbar, selbst bei Zusammenfassung der Gitterzellen auf einer höheren Ebene. Sie erlauben somit flexible Auswertungen nicht nur für administrative oder statistische Gebiete, sondern auch für frei definierbare Gebiete wie innerstädtische Strukturen oder Stadt-Umland-Übergänge (Wonka et al. 2009; BBSR 2021; Kirchner et al. 2014). Ein weiterer Vorteil neben der beliebigen räumlichen Zusammenfassung besteht in der vereinfachten Verschneidung mit weiteren Datenquellen auf dieser räumlich neutralen und – in Europa – länderübergreifend standardisierten Struktur. Flächendeckende Bevölkerungsdaten auf Rasterebene ermöglichen demnach eine differenziertere Be-

¹ Siehe hierzu den Krankenhausatlas der Statistischen Ämter des Bundes und der Länder: <https://krankenhausatlas.statistikportal.de/>.

² INSPIRE= INfrastructure for SPatial InfoRmation in Europe.

trachtung soziodemografischer Entwicklungen in Gemeinden, Ortsteilen etc. Da der Bedarf an kleinräumigeren Bevölkerungszahlen aktuell nicht von der Bevölkerungsfortschreibung gedeckt werden kann, wird ein neuer experimenteller Lösungsansatz verfolgt.

Demzufolge gibt es bereits diverse Herangehensweisen bzw. Ansätze, um kleinräumige und aktuelle Bevölkerungszahlen zu erzeugen. Neben mittlerweile gängigen Methoden wie der sogenannten Small Area Estimation, eine kleinräumige Schätzmethode zur Schätzung von kleinräumigen Bevölkerungszahlen, wie bereits in Simpson et al. (1996) diskutiert, werden zusätzlich insbesondere Fernerkundungsdaten – vorzugsweise Satellitendaten – zur Herleitung kleinräumiger Bevölkerungsverteilungen verwendet. Hierbei werden Bevölkerungszahlen, insbesondere Bevölkerungsdichten, vorrangig durch Kartierungsmethoden oder kleinräumige Schätzverfahren anhand von Fernerkundungsdaten auf Basis der letzten Volkserhebung auf räumlich feine Einheiten prognostiziert (Stevens et al. 2015).³ Lloyd et al. (2017) nutzen hierbei im sogenannten WorldPop Programm⁴ basierend auf diversen Geodaten einen gewichteten dasymetrischen Ansatz, worunter ein Prozess der räumlichen Umverteilung von interessierenden Größen durch eine flächenhafte Interpolation zu verstehen ist, bei dem anhand eines Random Forest Modells Bevölkerungszahlen kleinräumig geschätzt werden.⁵ Schug et al. (2021) kartieren die Bevölkerung in Deutschland unter Verwendung von Gewichtungsschichten, die bspw. von der Gebäudedichte, Gebäudehöhe und den Gebäudetypen aus Satellitendaten (Copernicus Sentinel-1 und Sentinel-2 Daten) hergeleitet werden. Neuere Ansätze wie in Kobe et al. (2022) kombinieren Satellitendaten und das Small Area-Schätzverfahren SPREE⁶, eine Methode der strukturerhaltenden Schätzung, die insbesondere für regional und demografisch differenzierte Bevölkerungsfortschreibungen zwischen den Zensen auf kleinräumiger Ebene verwendet wird. Anhand der Hilfsinformationen aus den Satellitendaten werden folglich kleinräumige Bevölkerungszahlen für den Senegal selbstständig fortgeschrieben.

Um die Qualitätsaspekte der amtlichen Statistik möglichst nicht zu tangieren, werden in dieser Arbeit die amtlich fortgeschriebenen Bevölkerungszahlen anhand neuer digitaler Daten kleinräumig umverteilt. Die Qualität der fortgeschriebenen Bevölkerungszahl ist ab der Gemeindeebene aufwärts unangetastet, da sie den Bevölkerungszahlen der Bevölkerungsfortschreibung entsprechen. Mit diesem Verfahren wird die amtliche Statistik unterstützt – jedoch nicht ersetzt – da keine Bevölkerungszahlen fortgeschrieben werden, sondern diese nur anhand einer zusätzlichen

³ Ferner vergleichen Leyk et al. (2019) anhand verschiedener dasymetrischer Kartierungsmethoden erzeugte und zugängliche Rasterdatensätze zu Bevölkerungszahlen und -dichten mit den Genauigkeiten und Qualitäten der Ergebnisse und geben Hilfestellung zum beabsichtigten Verwendungszweck.

⁴ Siehe hierzu auch: <https://www.worldpop.org/methods/populations>.

⁵ Kartografisch dargestellt wird dabei die Bevölkerungsdichte anhand der geschätzten Anzahl der Wohnbevölkerung pro Gitterzelle, wobei die hier verwendeten Gitterzellen nicht INSPIRE-konform sind. Die geschätzte Gesamtbevölkerung der Länder wird so angepasst, dass sie mit den entsprechenden offiziellen Bevölkerungsschätzungen der Vereinten Nationen und nicht mit den amtlich veröffentlichten Einwohnerzahlen – sofern vorhanden – der einzelnen Länder übereinstimmen.

⁶ SPREE = Structure preserving estimation.

externen Datenquelle – genauer Mobilfunkdaten – kleinräumig unterhalb der Gemeindeebene verteilt werden.

Dass die Verteilung der Bevölkerung mit den vorliegenden Mobilfunkdaten grundsätzlich gut und zeitnah abgebildet werden kann, zeigen Hadam et al. (2020) bereits in den bisherigen Analysen zur Bevölkerungsdarstellung mit Mobilfunkdaten. Der ausschlaggebende Vorteil der Mobilfunkdaten im Vergleich zu anderen Datenquellen oder Hilfsinformationen besteht hierbei in den starken Zusammenhängen der Mobilfunkdaten mit der Bevölkerung sowie in ihrer zeitlich und räumlich hohen Auflösung. Zudem sind Mobilfunkdaten robust gegenüber administrativen Gebietsstrukturänderungen und können für jede gewünschte räumliche Einheit aufbereitet werden und sind dadurch auch im Zeitverlauf vergleichbar. Im Gegensatz zu anderen Datenquellen können Mobilfunkdaten die tatsächlichen Aufenthaltsorte der Bevölkerung somit valide und zeitnah darstellen.⁷

Douglass et al. (2015) haben den Nutzen von Mobiltelefonaten zur Darstellung hochauflösender Bevölkerungsschätzer bereits erkannt und fokussieren sich darauf, die Bevölkerung in Mailand durch ein Random Forest Modell für den Zeitraum zwischen den Zensen zu schätzen, das auf den bekannten Zensusdaten trainiert wird. Sie verwenden hierzu sogenannte individuelle Mobiltelefonaten oder auch Call Detail Records (CDRs). Deville et al. (2014) zeigen ferner, wie Mobiltelefonaten bzw. CDRs die gängigen Ergebnisse der Volkszählung durch kleinräumige Schätzungen oder auch bei der Messung der Bevölkerungsdynamik ergänzen können. Zudem vergleichen sie die geschätzte Bevölkerungsdichte, die auf Basis von CDRs sowie durch Fernerkundungsdaten hergeleitet wird, anhand der amtlichen Bevölkerungszahlen in Portugal und schlussfolgern, dass die Kombination beider Datenquellen und Methoden eine Verbesserung der räumlichen und zeitlichen Auflösung verspricht.⁸

Der Vorteil dieser CDRs liegt mitunter in der sehr individuellen Angabe von Informationen zu Mobiltelefonnutzenden auf einer hohen räumlichen Auflösung, die im Gegensatz zu den Signaldaten – im Folgenden nur noch als Mobilfunkdaten bezeichnet – jedoch ereignisbasiert sind. Bei den Mobilfunkdaten werden alle erzeugten Signale im entsprechenden Mobilfunknetz vom Netzbetreiber erfasst (Hadam 2021). Die CDRs sind daher nur verfügbar, wenn der Telefonnutzende bspw. aktiv einen Anruf tätigt oder eine SMS bzw. mobile Daten sendet. Zudem liegen CDRs nur von Vertragskundinnen und -kunden vor, die im Rechnungssystem des Mobilfunkanbieters hinterlegt sind. Um Aussagen über die Bevölkerungszahlen anhand von Daten mobiler Endgeräte zu tätigen, bieten sich CDRs aufgrund der offensichtlichen Selektivitäten (Vertrag- vs. Prepaid-Kundin/-Kunde) daher nicht an.

Im Projekt *Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten* wird darauf aufbauend erforscht, ob und inwieweit mit Mobilfunkdaten die vorhandene Bevölkerungsfortschreibung un-

⁷ Daher bieten sich Mobilfunkdaten auch als potenzielle Informationsquelle für den Katastrophenschutz an, insbesondere um zu ermitteln, wo sich die Tagesbevölkerung im Zeitverlauf (Tages- und Wochenverlauf) befindet. Der Informationsgehalt ist jedoch abhängig von der erforderlichen räumlichen und zeitlichen Auflösung.

⁸ Die bessere räumliche Auflösung findet sich bei Deville et al. (2014) in den Fernerkundungsdaten, deren Aufnahmen jedoch abhängig von den Wetterverhältnissen sind. Die zeitliche Genauigkeit findet sich hierbei in den Mobiltelefonaten.

ter Verwendung eines Verteilungsverfahrens kleinräumig, von der Gemeindeebene bundesweit auf INSPIRE-konforme 1×1 km Gitterzellen, verteilt und abgebildet werden kann. Bis die erste amtliche georeferenzierte Bevölkerungszahl auf Basis des Zensus 2022 vorliegt, kann die zeitliche Lücke behelfsweise durch die Nutzung von Mobilfunkdaten geschlossen und als experimentelles Ergebnis genutzt werden. Zudem werden die erstellten kleinräumigen Ergebnisse anhand von Geodaten der deutschen Landesvermessung des Bundesamtes für Kartographie und Geodäsie (BKG) validiert und die experimentelle Bevölkerungsverteilung auf Plausibilität geprüft. Fehlzusweisungen, die unplausible Ergebnisse hervorrufen, werden weiterhin durch eine nachfolgende Modifizierung der Mobilfunkdaten bestmöglich korrigiert. Insgesamt stellt der Artikel damit den aktuellen Stand der Arbeit zur experimentellen Georeferenzierung der Bevölkerungszahl mittels Mobilfunkdaten dar.

Dieser Artikel ist wie folgt gegliedert: Im nachfolgenden Abschnitt werden die Datengrundlagen beschrieben, die sich in die amtliche Bevölkerungsfortschreibung sowie die verwendeten Mobilfunkdaten aufteilen. Hierbei wird insbesondere auf die Datenstrategie und -aufbereitung der Mobilfunkdaten eingegangen. In Abschn. 3 wird die Methode – genauer das Verteilungsverfahren – zur Umverteilung der Bevölkerungszahlen anhand der zuvor beschriebenen Daten erläutert. Die resultierenden Ergebnisse werden in Abschn. 4 diskutiert und auf Plausibilität geprüft. Im letzten Abschnitt wird ein Fazit zur hier beschriebenen Erstellung kleinräumiger Bevölkerungszahlen unter Verwendung von Mobilfunkdaten gezogen und es werden weitere Schritte sowie Schlussfolgerungen zur Diskussion gestellt.

2 Datengrundlage: Bevölkerungsfortschreibung und Mobilfunkdaten

Die Bevölkerungsfortschreibung ermöglicht die Angabe aktueller Einwohnerzahlen bis auf Ebene der Gemeinden und gibt die Bevölkerungszahl und die Zusammensetzung der Bevölkerung untergliedert nach Geschlecht, Alter, Familienstand und Staatsangehörigkeit wieder (Statistisches Bundesamt 2021b).

Die Einwohnerzahl wird nach § 5 BevStatG auf Basis der letzten Volkszählung (gegenwärtig Zensus 2011) anhand von Angaben der Statistiken zu Geburten und Sterbefällen, zu Staatsangehörigkeitswechseln und Lösungen von Ehen und Lebenspartnerschaften sowie der Wanderungsstatistik laufend fortgeschrieben (Statistisches Bundesamt 2021b). Die Fortschreibung der Bevölkerungszahlen insgesamt sowie untergliedert nach Alter und Geschlecht resultiert aus den statistischen Ergebnissen der Bevölkerungsbewegungen, worunter Wanderungen, Geburten, Sterbefälle und Eheschließungen zu verstehen sind. Untergliedert werden diese nach den natürlichen Bevölkerungsbewegungen, hierunter fallen Geburten sowie Sterbefälle, und nach den räumlichen Bevölkerungsbewegungen, den Zu- und Abwanderungen über Gemeindegrenzen hinweg, die aus entsprechenden Verwaltungsdaten von Standesämtern und Meldebehörden gezogen werden (Statistisches Bundesamt 2021b).

Die demografischen Merkmale der Zusammensetzung aus der Bevölkerungsfortschreibung liegen zudem in unterschiedlicher regionaler Gliederungstiefe vor, wobei die Merkmale Geschlecht, Alter und Staatsangehörigkeit (deutsch/nicht-deutsch) bis auf Gemeindeebene und der Familienstand nur auf der Kreisebene sowie einzel-

ne Staatsangehörigkeiten auf der Landesebene vorliegen (Statistisches Bundesamt 2021b). Insgesamt werden die Ergebnisse auf Ebene der Gemeinden, Kreise, Bundesländer und das Bundesgebiet nach dem Gemeindeverzeichnis⁹ des Statistischen Bundesamtes ausgewiesen.

Neben der Bevölkerungsfortschreibung stellen Mobilfunkdaten die zweite elementare Datengrundlage in diesem Artikel dar. Aufgrund des Potenzials, die Verteilung der Tages- und Wohnbevölkerung gut und zeitnah abzubilden (Hadam et al. 2020), stellen sie – besonders durch die starken Zusammenhänge mit der Wohnbevölkerung – eine geeignete Grundlage dar, um die Ergebnisse der Bevölkerungsfortschreibung kleinräumig zu verteilen.¹⁰ Seit dem Jahr 2019 besitzen über 97% der privaten Haushalte in Deutschland ein mobiles Endgerät (Statistisches Bundesamt 2021a), weshalb die gezählten Mobilfunkaktivitäten bundesweit flächendeckend zu einer realitätsnahen Darstellung der Tages- und Wohnbevölkerung in Deutschland beitragen können. Im Vergleich zu anderen Datenquellen, insbesondere traditionellen Erhebungsdaten, liegen Mobilfunkdaten damit zeitnah, hochaktuell und kleinräumig zur Verfügung und sind grundsätzlich nicht von äußeren Einflüssen, wie Wetterbedingungen, beeinflussbar. Zudem ist der Aufwand der Datenerfassung und -aufarbeitung bei Mobilfunkdaten tendenziell geringer, weshalb die zeitliche Aktualität bei anderen Datenquellen, so wie bspw. traditionellen Erhebungsdaten, tendenziell abnimmt.

Ziel der hier aufbereiteten Mobilfunkdaten ist eine möglichst perfekte Abbildung der potenziellen Wohnbevölkerung in den Mobilfunkdaten, um anhand dessen die Bevölkerungszahlen der Bevölkerungsfortschreibung kleinräumig zu verteilen. Zu diesem Zweck werden Mobilfunkdaten aus dem Netz der Telefónica Deutschland verwendet, die vom Datenanbieter Teralytics GmbH aufbereitet und zur Verfügung gestellt werden. Die Mobilfunkdaten liegen aus 8 ausgewählten Wochen aus dem Jahr 2019, exkl. Ferien und Feiertage, auf einem flächendeckenden INSPIRE-konformen 1 × 1 km Raster¹¹ vor, um Verzerrungen durch touristische und freizeitliche Aktivitäten zu vermeiden.¹² Da es sich hierbei um Signaldaten handelt, werden alle Signale im entsprechenden Mobilfunknetz vom Netzbetreiber automatisch erfasst, sofern das mobile Endgerät nicht ausgeschaltet ist oder sich im Flugmodus befindet.

⁹ Siehe hierzu: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/_inhalt.html.

¹⁰ Hierbei wird vereinfachend von Wohnbevölkerung – genauer der potenziellen Wohnbevölkerung – in den Mobilfunkdaten gesprochen, die jedoch gleichbedeutend mit der Aufenthaltsbevölkerung in den Mobilfunkdaten ist. Dies ist darin begründet, dass die Mobilfunkdatenaufbereitung zwangsweise anderen Definitionen zur Ermittlung der Wohnbevölkerung unterliegt als die amtliche Statistik. Jedoch wurden die Mobilfunkdaten mit dem Ziel aufbereitet, ausschließlich bzw. bestmöglich die potenzielle Wohnbevölkerung abzubilden. Weitere Informationen und Analysen zur Tagesbevölkerung oder auch Aufenthaltsbevölkerung sind in Statistisches Bundesamt (2019), Hadam et al. (2020) und Hadam (2021) aufgeführt.

¹¹ Entsprechend der INSPIRE-Richtlinie wurde dazu die Lambert Azimuthal Equal Area Projektion verwendet (ETRS89-LAEA Europe – EPSG:3035).

¹² Eine weitere Unterscheidung der potenziellen Wohnbevölkerung in den Mobilfunkaktivitäten zwischen Erst- und Zweitwohnsitz ist zudem nicht möglich, beides ist in den Mobilfunkdaten vorhanden. Da die Mobilfunkdaten jedoch derart aufbereitet sind, dass der hauptsächliche Aufenthaltsort der Aktivitäten im Jahresdurchschnitt ermittelt wird, ist diese Unterscheidung besonders für die weitere Ermittlung der Tagesbevölkerung nicht ausschlaggebend.

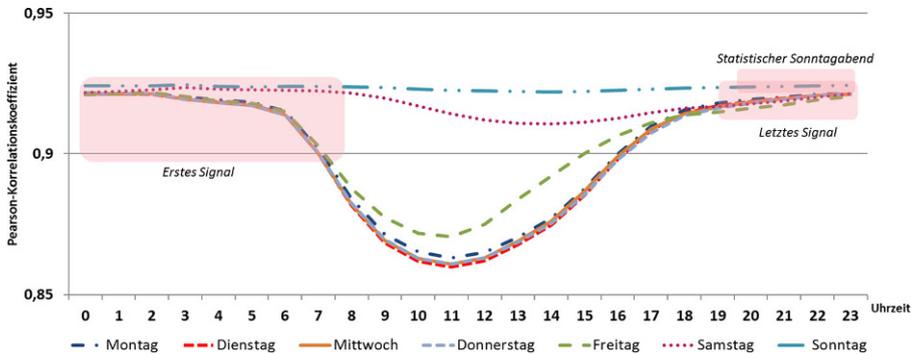


Abb. 1 Pearson-Korrelationskoeffizienten zwischen den Bevölkerungszahlen des Zensus 2011 und den Mobilfunkdaten basierend auf einem bundesweiten Mischraster sowie Darstellung/Hervorhebung der beiden möglichen Mobilfunkdatenstrategien

Dabei wird lediglich die Ortsangabe des Funkmastes registriert, mit dem das mobile Endgerät zu einem bestimmten Zeitpunkt verbunden ist.

Abb. 1 stellt beifolgend die Pearson-Korrelationskoeffizienten zwischen den Bevölkerungszahlen des Zensus 2011 und den aufbereiteten Mobilfunkdaten aus dem Netz der Telefónica Deutschland basierend auf einem Mischraster, wie in Statistisches Bundesamt (2019), nach Wochentag und Uhrzeit erstmalig für ganz Deutschland in einem Liniendiagramm dar. Eine hohe Korrelation in Abb. 1 lässt schlussfolgern, dass zu den entsprechenden Zeitpunkten die mobilen Aktivitäten am Wohnort getätigt wurden, da der Zusammenhang zwischen den Bevölkerungszahlen des Zensus 2011 und den Mobilfunkdaten entsprechend stark positiv ist.

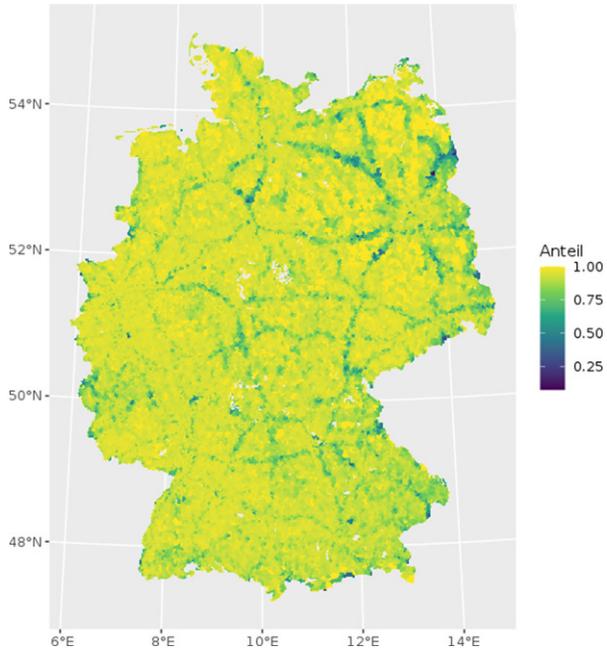
Ferner stehen zwei Strategien zur Zählung mobiler Aktivitäten zur Abbildung der potenziellen Wohnbevölkerung zur Verfügung. Die beiden möglichen Datenstrategien sind visuell in den rötlich hinterlegten Rechtecken in Abb. 1 hervorgehoben.

Die erste bereits bekannte Option nach Hadam et al. (2020) und dem Statistisches Bundesamt (2019) bildet einen Datensatz für einen statistischen Sonntagabend, wie in Abb. 1 dargestellt, welcher den Durchschnittswert aller Mobilfunkaktivitäten von 20 bis 23 Uhr an den ausgewählten Sonntagen mit einer zweistündigen Verweildauer im Untersuchungsgebiet enthält. Aufgrund der in Abb. 1 sichtbaren höchsten Korrelation zwischen den Mobilfunkdaten am statistischen Sonntagabend und den Bevölkerungszahlen des Zensus 2011 wird angenommen, dass dieser Zeitraum einen guten Indikator für die Darstellung der Bevölkerungsverteilung liefert.

Die zweite und damit alternative Datenstrategie zur Ermittlung der Wohnbevölkerung wird als Heimatort-Strategie bezeichnet und ermittelt den Herkunftsort aller erfassten Mobilfunksignale anhand des ersten und des letzten Signals innerhalb von 24 Stunden. Hierbei gibt es verschiedene Möglichkeiten, die die Erfassungen der Signale betreffen, die in Abb. 1 hypothetisch im Korrelationsdiagramm in den rötlichen Rechtecken (*Erstes Signal*, *Letztes Signal*) hervorgehoben sind.¹³

¹³ Um welche Uhrzeit das erste und letzte Signal innerhalb von 24 Stunden getätigt wird, ist nicht definiert. Daher werden die Hervorhebungen in Abb. 1 zur vereinfachten Veranschaulichung der Datenstrategie angenommen.

Abb. 2 Anteil der identischen ersten und letzten Signale an allen verfügbaren Signalen (erstes u./o. letztes) über die Werktage (Mo.–Do.) mit sichtbar höherem Anteil des separaten ersten oder letzten Signals auf dem deutschen Autobahnnetz



Die eingängigste Definition stellt das räumlich identische erste und letzte Signal dar. Eine räumliche Einheit wird als Herkunftsort bzw. potenzieller Wohnort aus den Mobilfunkdaten bestimmt, wenn das erste und letzte Signal des mobilen Endgerätes innerhalb von 24 Stunden in derselben räumlichen Einheit erfasst wurde. Alternativ hierzu können das erste Signal oder auch das letzte Signal innerhalb von 24 Stunden separat verwendet werden, sofern diese nicht räumlich übereinstimmen. Dabei wird der potenzielle Wohnort auf dem ersten Signal innerhalb von 24 Stunden bestimmt, falls das Signal vor 8 Uhr erfasst wurde und das erste und letzte Event nicht übereinstimmen und vice versa bei der Bestimmung nach dem letzten Signal innerhalb von 24 Stunden.

Weiterhin wird bei der Heimatort-Strategie im Gegensatz zum statistischen Sonntagabend ein Werktagsdurchschnitt gebildet, ohne das Wochenende sowie den Freitag einzubinden, was am veränderten (Bewegungs-)Verhalten der Bevölkerung bzw. der Mobilfunknutzenden in diesem Zeitraum liegt (siehe auch Hadam 2021). Zudem muss beachtet werden, dass die Angabe des Herkunftsortes nur über das separate erste oder letzte Signal verzerrt ist, wie Abb. 2 veranschaulicht.

Werden die räumlich identischen ersten und letzten Signale ins Verhältnis zu allen verfügbaren (ersten u./o. letzten) Signalen über die Werktage Montag bis Donnerstag gesetzt, sticht das Autobahnnetz in Deutschland insbesondere im östlichen Teil Deutschlands in Abb. 2 sichtbar hervor. Der Anteil des separaten ersten oder letzten Signals in diesen Bereichen erscheint gering im Vergleich zu den anderen Regionen, jedoch bilden sie an den Bundesgrenzen sowie auf dem Autobahnnetz die Hauptaktivitäten in den verfügbaren Mobilfunkdaten. Insgesamt macht das erste Signal nur 2,2% aller Mobilfunkaktivitäten aus, das letzte Signal nur 3,9% aller

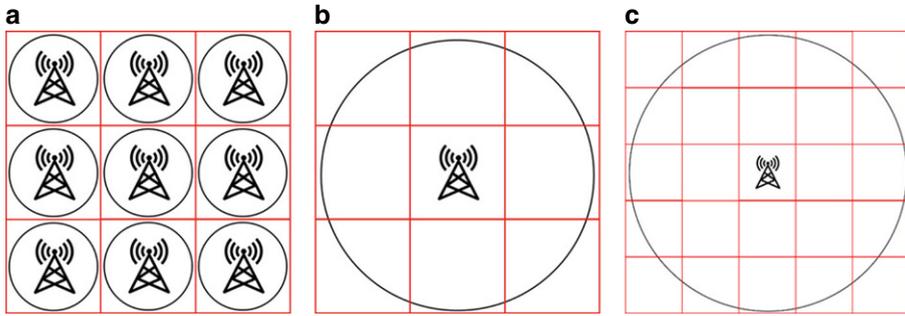


Abb. 3 Räumliche Verteilung/Modellierung der Mobilfunkaktivitäten durch den Datenanbieter nach Modellierungsgrad **a** keine, **b** kleinräumig und **c** großflächig modelliert

Aktivitäten und damit fallen 93,8% aller Aktivitäten der ausgewählten Werkzeuge auf die räumlich identischen ersten und letzten Signale, weshalb durch diese Anforderung kein Informationsverlust in den Mobilfunkdaten entsteht. Schlussendlich stellen diese Herkunftsorte nicht die potenzielle Wohnbevölkerung in Deutschland dar und werden schließlich in den nachfolgenden Analysen nicht weiter einbezogen.

Da nur Mobilfunkdaten eines von insgesamt drei Mobilfunkanbieter auf dem deutschen Markt zur Verfügung stehen, wurden die Mobilfunkdaten vom Datenanbieter extrapoliert, wobei ein konstanter Extrapolationsfaktor auf Landkreisebene basierend auf Einwohnerzahlen der Bevölkerungsfortschreibung berechnet wurde. Dabei wurde die Extrapolation nur für Mobilfunkaktivitäten deutscher SIM-Karten durchgeführt, um Verzerrungen durch ausländische oder touristische Aktivitäten zu vermeiden. Dies erfolgte durch eine sogenannte Roamerkorrektur, wobei die Roamer (nicht-deutsche SIM-Karten) bei der Berechnung der Extrapolationsfaktoren herausgerechnet wurden.

Um flächendeckende Mobilfunkdaten für das 1×1 km Raster zu erhalten, mussten die Mobilfunkaktivitäten durch den Datenanbieter in einem letzten Schritt – sofern notwendig – anhand von weiteren Bevölkerungszahlen¹⁴ modelliert bzw. räumlich verteilt werden. Das bedeutet, dass die Anzahl der mobilen Aktivitäten nicht in jedem Fall eindeutig einer einzelnen Gitterzelle zugewiesen werden kann.

Abb. 3 visualisiert hierbei vereinfachend das Modellierungsprinzip. Im Idealfall wird jede Gitterzelle von mindestens einer Mobilfunkzelle abgedeckt, auf deren Basis die mobilen Aktivitäten initial erfasst werden (siehe Abb. 3a). Bei einem bundesweiten 1×1 km Raster wird diese Bedingung nur in dicht besiedelten Regionen erfüllt. Sofern die Mobilfunkzelle mehr als eine Gitterzelle, wie in Abb. 3b und c, abdeckt, werden die Mobilfunkaktivitäten vom Datenanbieter Teralytics anhand der ihnen zur Verfügung stehenden Bevölkerungszahlen probabilistisch in die Gitterzellen verteilt. Dies ist vor allem in ländlichen oder weniger dicht besiedelten Regionen der Fall. Der Modellierungsgrad des Datenanbieters zeigt auf, dass 5,6% der Mobilfunkdaten den 1×1 km Gitterzellen eindeutig zugewiesen werden konnten

¹⁴ Hierbei handelt es sich um Daten eines Customer Intelligence Unternehmens auf Ebene von Wohnbezirken.

(vergleichbar mit Abb. 3a). 37,5 % wurden kleinräumig modelliert (vergleichbar mit Abb. 3b), was bedeutet, dass die Mobilfunkzelle zwischen zwei und neun Gitterzellen abdeckt. Deckt die Mobilfunkzelle mehr als neun Gitterzellen ab, werden die Mobilfunkaktivitäten großräumig modelliert bzw. mit einfachen Annahmen räumlich verteilt und die Genauigkeit der Zuweisungen mobiler Aktivitäten lässt deutlich nach (vergleichbar mit Abb. 3c). Dies ist bei 56,9 % der Gitterzellen der Fall und der größte Treiber möglicher Unsicherheiten in den resultierenden Ergebnissen.

Zu guter Letzt liegen ebenfalls die soziodemografischen Merkmale Altersgruppe und Geschlecht ausschließlich der Vertragskundinnen und -kunden vor (Datenstand: 2021).

3 Methodik

3.1 Räumliche Zuordnung der Gitterzellen und Mobilfunkaktivitäten

Für eine präzise Umverteilung der Ergebnisse der Bevölkerungsfortschreibung von der Gemeindeebene auf Ebene der 1×1 km Gitterzellen bedarf es einer akkuraten Zuordnung der Gitterzellen zu der überdeckenden Gemeinde. Andernfalls besteht die Gefahr einer fehlerhaften kleinräumigen Verteilung von Bevölkerungszahlen innerhalb ihrer Gemeinde durch die aufbereiteten Mobilfunkdaten.

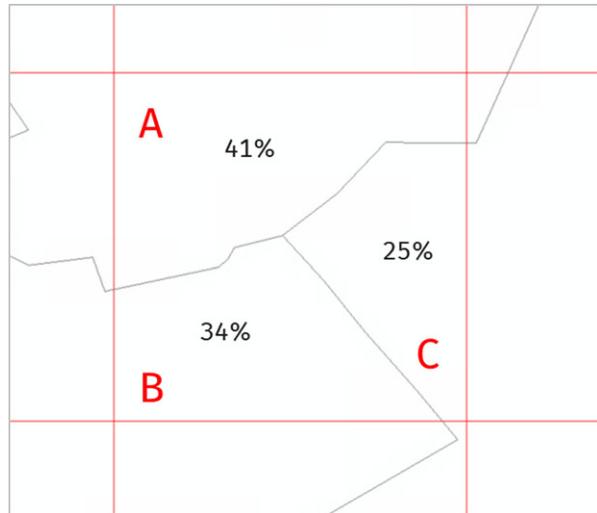
Die einfachste Methode bietet eine geografische Mittelpunktzuordnung, bei der der Mittelpunkt einer Gitterzelle anhand ihrer räumlichen xy-Koordinaten in einer eindeutig überdeckenden Gemeinde verortet und dieser zugeordnet wird. D.h., die Gitterzelle wird derjenigen Gemeinde zugewiesen, deren Mittelpunkt sie überdeckt. Der Vorteil dieser Methode ist die schnelle und einfache Umsetzung in gängigen Geoinformationssystem-Softwareprodukten. Der Nachteil liegt in der teilweise fehlerhaften oder auch nicht möglichen Zuordnung von Gitterzellen zu Gemeinden. Dies ist vorwiegend in Bundesländern mit flächenmäßig kleinen Gemeinden der Fall, bei der eine Zuordnung nicht oder nur stark verzerrt möglich ist. Im vorliegenden Fall resultieren bei einer Mittelpunktzuordnung 46 Gemeinden in Deutschland, die nicht von Gitterzellen überdeckt und zugeordnet werden können (siehe Tab. 1).

Eine zweite und zuverlässigere Möglichkeit bietet die Zuordnung der Gitterzelle zu einer überdeckenden Gemeinde ausschließlich anhand ihrer Fläche, mit der die Gitterzelle die Gemeinde überdeckt. Dies wurde auch in dieser Arbeit umgesetzt. Das Ziel hierbei ist es, die Gitterzelle derjenigen Gemeinde zuzuordnen, die den größten Flächenanteil an einer Gemeinde besitzt, unabhängig von den zugrundeliegenden Bevölkerungsdichten.

Tab. 1 Anzahl fehlender Gemeinden nach Zuordnungsstrategie der Gitterzellen sowie Mobilfunkdatenwahl

	Mittelpunktzuordnung (Sonntagabend)	Flächenzuordnung (Sonntagabend)	Flächenzuordnung (Heimortort- Werktagsdurchschnitt)
Anzahl fehlender Gemeinden	46	4	2

Abb. 4 Flächenzuordnung der Gitterzelle zu den überdeckenden Gemeinden anhand der berechneten Flächenanteile



In einem ersten Schritt werden die Flächenanteile jeder Gitterzelle zur überdeckenden Gemeinde berechnet. Abb. 4 zeigt beispielhaft eine Gitterzelle, die drei Gemeinden überdeckt. Die Berechnung der Fläche ergibt, dass Gemeinde A den größten Flächenanteil mit 41 % an der Gitterzelle besitzt. In einem zweiten Schritt werden die Mobilfunkaktivitäten aus dieser Gitterzelle anhand der jeweiligen Flächenanteile anteilmäßig auf die drei Gemeinden A, B und C verteilt und damit nicht zu 100 % einer Gemeinde zugeteilt. Dadurch erfolgt in diesem Beispiel eine implizite Dreiteilung der Gitterzelle bei der Berechnung der gruppenspezifischen Ziehungswahrscheinlichkeit in Abschn. 3.2. Hiermit werden weniger Verzerrungen bzw. Unsicherheiten in den Ergebnissen durch grenzüberlappende Gitterzellen verursacht, weil die Mobilfunkdaten anteilmäßig den Gemeinden zugeordnet werden, in denen diese flächenmäßig liegen. Im letzten Schritt wird jede Gitterzelle sowie die berechnete experimentelle georeferenzierte Bevölkerungszahl eindeutig der Gemeinde mit dem höchsten Flächenanteil zugeordnet (hier Gemeinde A).

In den hier aufgezeigten Zuordnungsstrategien und im nachfolgend beschriebenen Verteilungsverfahren wird damit implizit eine Strukturgleichheit auf Gemeindeebene sowie dem 1×1 km Raster angenommen. Diese wird gleichsam in sowie zwischen den Mobilfunkdaten und den zu verteilenden Bevölkerungszahlen der Bevölkerungsforschreibung übernommen und geht aus der nachweisbaren Korrelation beider Datenquellen hervor. Letzteres resultiert zudem in der Annahme, dass sich die vorliegende (kleinräumige) Verteilung der Mobilfunkdaten proportional zur Verteilung der amtlichen Bevölkerungszahlen verhält, weshalb diese als Verteilungsgrundlage in dieser Arbeit Bestand haben. Entsprechend ist auch eine triviale Aufteilung der Mobilfunkaktivitäten von der Gitterzelle auf mehrere Gemeinden anhand der Flächenanteile umsetzbar.

In Bezug zu den beiden Mobilfunkdatenstrategien aus Abschn. 2 werden beim letzten Verfahren damit alle Gemeinden – mit Ausnahme von zwei Gemeinden, in denen keine Mobilfunkdaten unter anderem durch fehlende Mobilfunkmasten vor-

liegen – mit Gitterzellen abgedeckt, wie in Tab. 1 gelistet. Durch die Abhängigkeit der Datenverfügbarkeit aufgrund der ausgewählten Mobilfunkdatenstrategie bzw. Datenwahl kann auch die präferierte Flächenzuordnung zu fehlenden Ergebnissen einzelner Gemeinden führen, wie es im Fall des statistischen Sonntagabends zur Abbildung der potenziellen Wohnbevölkerung vorliegt (vgl. Tab. 1). Wegen fehlender Mobilfunkaktivitäten beim statistischen Sonntagabend können zwei weitere Gemeinden nicht kleinräumig durch Gitterzellen dargestellt werden. Dies trägt maßgeblich zur Wahl der Heimatort-Strategie für die Umsetzung in Abschn. 4.1 bei.

3.2 Verteilungs- und Rundungsverfahren

Im Rahmen eines Verteilungsverfahrens werden die Ergebnisse der Bevölkerungsfortschreibung von der Gemeindeebene auf eine kleinräumigere Ebene umverteilt. Für die kleinräumige Verteilung der Bevölkerungszahlen aus der Bevölkerungsfortschreibung werden die den Gemeinden räumlich zugeordneten Gitterzellen benötigt, um aus den zugrundeliegenden Mobilfunkaktivitäten den Verteilungsvorgang herzu-leiten. Hierfür werden gruppenspezifische Ziehungswahrscheinlichkeiten (P_{ID}) für jede Gitterzelle aus den Mobilfunkdaten in Abhängigkeit von der zugrundeliegen- den Gemeinde errechnet:

$$P_{ID} = \frac{n_{ID}}{n_i}, \quad (1)$$

wobei n_{ID} die Anzahl der Mobilfunkaktivitäten pro Gitterzelle ID und n_i die Summe aller Mobilfunkaktivitäten in der zugeordneten Gemeinde i ist. Vereinfachend ausgedrückt, wird der Anteil der Mobilfunkaktivitäten in Gitterzelle ID im Verhältnis zur Gesamtanzahl aller Aktivitäten in der zugeordneten Gemeinde i berechnet, sodass P_{ID} innerhalb der Gemeinde variiert.

Anhand der Ziehungswahrscheinlichkeiten P_{ID} wird die amtliche Bevölkerungszahl im nächsten Schritt kleinräumig verteilt. Die experimentelle georeferenzierte Bevölkerungszahl ($ExpGeoBFS_{ID}$) pro Gitterzelle ID ergibt sich dabei aus der Multiplikation der Bevölkerungszahl der Bevölkerungsfortschreibung in Gemeinde i mit der gruppenspezifischen Ziehungswahrscheinlichkeit P_{ID} aus Gl. 1:

$$ExpGeoBFS_{ID} = BFS_i * P_{ID}, \quad (2)$$

wobei BFS_i die Einwohnerzahl der Bevölkerungsfortschreibung in Gemeinde i darstellt.

Wie in Abb. 5 dargestellt, ermöglicht Gl. 2 eine kleinräumige Verteilung der Bevölkerungszahl von jeder möglichen administrativen Einheit – wie im vorliegenden Fall die Gemeindeebene – auf jede mögliche kleinräumigere Ebene. Notwendig hierfür sind gleichgroße kleinräumige Strukturen, bei denen die gruppenspezifischen Ziehungswahrscheinlichkeiten errechnet werden können. Burgdorf (2010) und Steinocher et al. (2005) führen im Vergleich hierzu eine räumliche Disaggregation von Bevölkerungsdaten mittels Bebauungsinformation bzw. Bebauungsdichten aus dem amtlichen Digitalen Basis-Landschaftsmodell (ATKIS-Basis-DLM) durch. Burgdorf (2010) nimmt dabei für bestimmte Objektarten unterschiedliche Bevölkerungsdich-

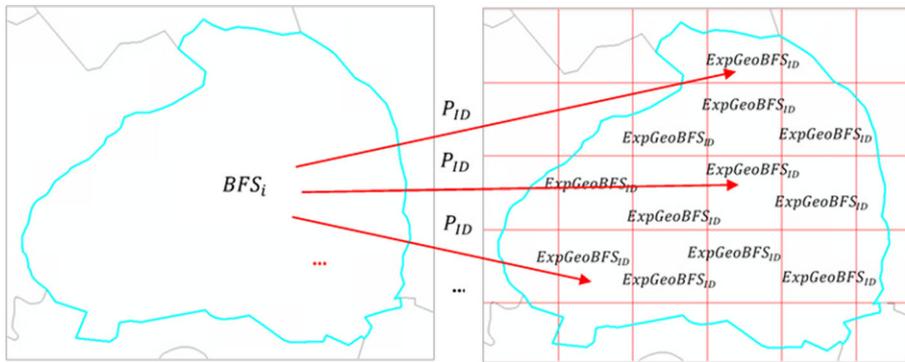


Abb. 5 Visuelle Darstellung des Verteilungsverfahrens

ten an, die zur Vergabe von Gewichten für die Umverteilung der Bevölkerungszahl verwendet werden. Bei Steinnocher et al. (2005) erfolgt die räumliche Aufteilung über eine gewichtete Summenfunktion. Sie leiten bspw. für die untersuchte Region einen spezifischen Faktor ab, der innerhalb der Region konstant ist und vom Verhältnis der Gesamtbevölkerung zur Summe des Flächenanteils und der Bebauungsdichte der entsprechenden Bebauungsklasse abhängt.

Hierbei wird zudem die Relevanz einer bestmöglichen Zuordnung von Gitterzelle zu Gemeinde aus Abschn. 3.1 deutlich. Wird die Gitterzellenzuordnung nicht akkurat durchgeführt, hat dies andere gruppenspezifische Ziehungswahrscheinlichkeiten (P_{ID}) aus Gl. 1 zur Folge. Daraus ergeben sich zwangsläufig veränderte experimentelle georeferenzierte Bevölkerungszahlen ($ExpGeoBFS_{ID}$) in Gl. 2.

Um nun die amtliche Bevölkerungszahl je Gemeinde aus der Bevölkerungsfortschreibung zu erhalten, werden die experimentellen kleinräumigen Bevölkerungszahlen in einem entsprechenden Verfahren gerundet. Norman et al. (2008) oder Rees et al. (2003) verwenden dafür die sogenannte iterative proportionale Anpassung (Iterative Proportional Fitting (IPF)), um Bevölkerungsgruppen kleinräumig zu disaggregieren und gleichzeitig die Randwerte zu erhalten, sodass Zeilen- und Spaltensummen immer der Gesamtzahl der Bevölkerungsgruppe entsprechen. Im Vergleich dazu werden hier vereinfacht die aus Gl. 2 resultierenden $ExpGeoBFS_{ID}$ auf Gemeinde i aufsummiert und anschließend anhand der jeweiligen Bevölkerungszahl pro Gemeinde (BFS_i) gerundet, wobei die Summe der $ExpGeoBFS_{ID}$ pro Gemeinde i ($\sum_i ExpGeoBFS_{ID}$) dem Wert der Einwohnerzahl der Bevölkerungsfortschreibung (BFS_i) entsprechen muss:

$$\sum_i ExpGeoBFS_{ID} = BFS_i. \quad (3)$$

Daraus ergeben sich experimentelle georeferenzierte Bevölkerungszahlen, deren Eckwerte denen der amtlichen Bevölkerungsfortschreibung entsprechen.

Gl. 3 wird dabei wie folgt ausgeführt: Im ersten Schritt werden alle kleinräumig verteilten Bevölkerungszahlen anhand der angegebenen Dezimalstellen abgerundet, entgegen der klassischen Rundungsregel. Basierend darauf wird die Differenz zwi-

schen der amtlichen Bevölkerungszahl pro Gemeinde und der aufsummierten kleinräumig verteilten experimentellen Bevölkerungszahl für die zugrundeliegende Gemeinde gebildet.

Als Beispiel sei der folgende Zahlenvektor für die experimentellen georeferenzierten Bevölkerungszahlen (ExpGeoBFS_{ID}) aufgeführt: (27,27273; 34,34343; 38,38384). Nach der klassischen Rundungsregel, oder auch Fünferndung genannt, würde eine Summe von 99 resultieren.¹⁵ Der wahre Wert beträgt jedoch 100, welcher durch das implizite Rundungsverfahren in Gl. 3 hervorgeht. Weiter ergibt sich nun eine Differenz zwischen dem gerundeten und wahren Wert von 1. Daraufhin werden die zuvor abgerundeten ExpGeoBFS_{ID} nach ihren Dezimalstellen geordnet. Dies ist notwendig, um die abgerundeten Dezimalstellen von denjenigen ExpGeoBFS_{ID} um den Faktor 1 zu erhöhen, die letztlich die größten Nachkommastellen aufweisen. Die eingangs ermittelte Differenz, hier als Beispiel in Höhe von 1, zwischen BFS_i und der aufsummierten kleinräumig verteilten experimentellen Bevölkerungszahl ($\sum_i \text{ExpGeoBFS}_{ID}$) bestimmt hierbei die Anzahl der ExpGeoBFS_{ID} pro Gemeinde i , deren Werte erhöht bzw. aufgerundet werden müssen, um Gl. 3 zu erfüllen. Im angegebenen Beispiel wird von den drei Werten im Zahlenvektor durch die Differenz von 1 nur ein Wert mit der größten Dezimalstelle aufgerundet, hier 38,38384 auf 39. In Summe resultiert nach Gl. 3:

$$27 + 34 + 39 = 100.$$

Das hier beschriebene Verfahren lässt sich grundsätzlich auch auf weitere Angaben zur Bevölkerung, wie bspw. soziodemografische Merkmale umsetzen, sofern geeignete Mobilfunkdaten oder andere räumlich passende Datenquellen vorliegen, die eine kleinräumige Umverteilung anhand gruppenspezifischerziehungswahrscheinlichkeiten zulassen.

4 Diskussion der resultierenden experimentellen georeferenzierten Bevölkerungszahlen

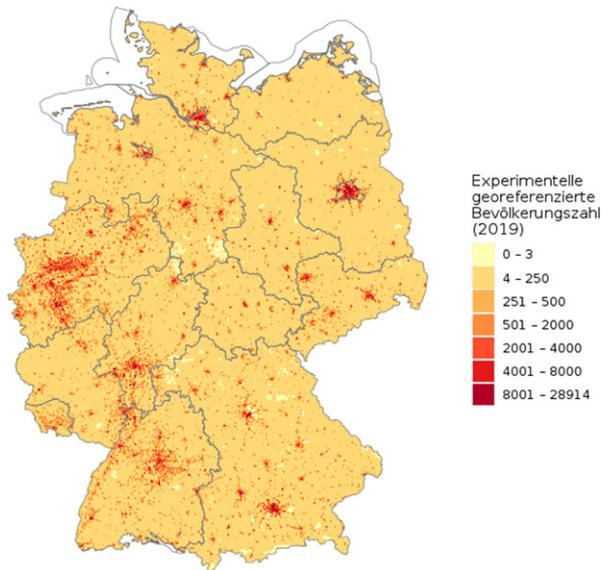
4.1 Die experimentelle georeferenzierte Bevölkerungszahl

Wie in Abschn. 2 bereits beschrieben, stehen zwei mögliche Mobilfunkdatenstrategien für die Bestimmung experimenteller georeferenzierter Bevölkerungszahlen nach Abschn. 3.2 zur Verfügung. Die Wahl der Datenstrategie hängt von zwei grundlegenden Aspekten ab: bundesweit flächendeckende Abdeckung bzw. Datenverfügbarkeit sowie bestmögliche Wiedergabe der Bevölkerungsverteilung.

Aus Tab. 1 wurde bereits sichtbar, dass die Mobilfunkdaten nach dem statistischen Sonntagabend nicht flächendeckend vorliegen bzw. die Datenverfügbarkeit durch die restriktive Annahme der ausschließlich gezählten sonntagabendlichen Mobilfunkaktivitäten abnimmt. Daher findet hier die Heimatort-Strategie Verwendung. Wie bereits Abb. 2 veranschaulichte, werden hierbei zudem nur die räumlich iden-

¹⁵ Klassisch gerundet auf: (27, 34, 38).

Abb. 6 Bundesweit visualisierte experimentelle georeferenzierte Bevölkerungszahl auf Ebene der 1×1 km Gitterzellen



tischen ersten und letzten Signale mobiler Aktivitäten einbezogen. Da das hier beschriebene Verfahren für die folgenden Berichtsjahre 2020 und 2021 umgesetzt wird, ist die flächendeckende Datenverfügbarkeit für die Folgejahre der ausschlaggebende Entscheidungsfaktor. Weitere Informationen zur Mobilfunkdatenwahl und die Auswirkungen der gewählten Mobilfunkdatenstrategie auf die Berechnung der experimentellen georeferenzierten Bevölkerungszahlen sind im Anhang A.1 dargestellt.

Die experimentelle georeferenzierte Bevölkerungszahl nach Abschn. 3.2 wird anhand der Heimatort-Strategie, basierend auf dem Werktagsdurchschnitt von Montag bis Donnerstag, und dem räumlich identischen ersten und letzten Mobilfunksignal der Mobilfunknutzenden innerhalb von 24 Stunden berechnet.¹⁶ Die Ergebnisse sind in Abb. 6 in einer statischen Karte dargestellt.

Abb. 6 stellt die räumliche Verteilung der experimentellen georeferenzierten Bevölkerungszahl anhand einer klassierten Skala dar, wobei die Grenzen der Bundesländer für die Einordnung der Werte hervorgehoben sind. Experimentelle georeferenzierte Bevölkerungszahlen zwischen 0 und 3 werden durch die Angabe eines Intervalls geheim gehalten und in der Karte als solches sowie farblich hell hinterlegt. Hohe Werte der experimentellen georeferenzierten Bevölkerungszahl werden dunkel schattiert (farblich rot) hervorgehoben und niedrige Werte hell schattiert (gelb/orange). Auf den ersten Blick erscheinen die berechneten Werte und die Verteilung plausibel, da besonders dicht besiedelte Regionen bzw. Städte wie Berlin, Hamburg, München, Köln, Bonn oder das Ruhrgebiet entsprechend stark in der Karte hervorstechen. Der eher weniger dicht besiedelte oder auch der ländliche Raum sind entsprechend gelb/orange in Abb. 6 hinterlegt. Der weniger dicht besiedelte Raum stellt visuell den Großteil der Fläche in Deutschland dar. Um die Ergebnisse

¹⁶ Tab. 6 im Anhang hebt die entsprechenden (finalen) Ergebnisse aus der Summenstatistik kräftig hervor.

kleinräumiger und individueller betrachten zu können, wurde zusätzlich eine interaktive Rasterkarte erstellt und auf der Seite *Statistik visualisiert* des Statistischen Bundesamtes veröffentlicht (weitere Informationen siehe hierzu den Anhang A.2).¹⁷

Die Eckwerte der experimentellen georeferenzierten Bevölkerungszahlen können in einigen Gebietsstrukturen oberhalb der Gemeindeebene (Kreis, Bundesland) leicht von den Ergebnissen der amtlichen Bevölkerungsfortschreibung¹⁸ abweichen. Grund hierfür sind nicht verfügbare Mobilfunkdaten in einigen Gitterzellen, die zu fehlenden experimentellen georeferenzierten Bevölkerungszahlen in den Ergebnissen der zugrundeliegenden Gemeinde führen können. Im vorliegenden Fall können zwei Gemeinden in Schleswig-Holstein (Helgoland, Nieby) wegen nicht vorhandener Mobilfunkaktivitäten unter anderem durch fehlende Mobilfunkmasten mit experimentellen georeferenzierten Bevölkerungszahlen ausgewiesen werden (vgl. auch Tab. 1). Insgesamt handelt es sich hier um rund 0,05 % der Gesamtbevölkerung in Schleswig-Holstein, die folglich nicht mit einer experimentellen georeferenzierten Bevölkerungszahl abgebildet werden können. Eine Aggregation der experimentellen georeferenzierten Bevölkerungszahlen von Gemeinde- auf bspw. Kreisebene kann demnach zu einer geringeren Einwohnerzahl führen als amtlich angegeben.

Weiterhin wurde untersucht, inwieweit sich das in Abschn. 3.2 beschriebene Verteilungsverfahren auf die soziodemografischen Angaben, wie Altersgruppe und Geschlecht, anwenden lässt. Wie durch das Statistische Bundesamt (2021c) bereits ausführlich dargestellt wurde, unterliegen die soziodemografischen Angaben der Mobilfunkanbieter starken Verzerrungen, die sich auch in dem hier beschriebenen Verfahren wiederfinden lassen. Um diesen Verzerrungen entgegenzuwirken, wurden Anpassungsfaktoren auf nationaler Ebene durch den Datenanbieter berechnet und die Verteilungen entsprechend angepasst. Besonders ein Fehlen der Nicht-Vertragsmündigen sowie der Prepaid-Kundinnen und -Kunden verhindert dennoch eine plausible Darstellung der experimentellen georeferenzierten Bevölkerungszahl differenziert nach Altersgruppen sowie Geschlecht.

Ferner ergeben sich Herausforderungen bei der Nutzung der soziodemografischen Merkmale für das hier beschriebene Verfahren in der resultierenden Datenverfügbarkeit durch die Kreuzkombinationen von Mobilfunkaktivität pro Gitterzelle und soziodemografischem Merkmal beim Datenanbieter sowie dem umgesetzten Anonymisierungsverfahren. Durch die Unterteilung der Merkmale in mehrere Untergruppen bzw. Klassen erhöht sich das Risiko, dass viele 1 × 1 km Gitterzellen ohne entsprechenden Wert ausgegeben werden und dadurch keine flächendeckenden Ergebnisse berechnet und dargestellt werden können. Zusätzliche Ausführungen zur Verwendung soziodemografischer Merkmale für die experimentelle georeferenzierte Bevölkerungsfortschreibung sind im Anhang A.3 ausgeführt.

Eine Qualitätseinschätzung der Mobilfunkdaten sowie der Ergebnisse (der experimentellen georeferenzierten Bevölkerungszahl) ist nur anhand weiterer vergleichba-

¹⁷ Siehe hierzu die am 14. Feb. 2022 erstmals veröffentlichte Anwendung unter: https://www.destatis.de/DE/Service/Statistik-Visualisiert/Bevoelkerung-Geo/Bevoelkerung_Karten.html.

¹⁸ Siehe hierzu: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/_inhalt.html.

rer Datengrundlagen gegeben, welche in den nachfolgenden Abschnitten besprochen werden.¹⁹

4.2 Validierung der Ergebnisse – Erste Plausibilitätsprüfung anhand des Zensus 2011

In Abschn. 4.1 wurden die ermittelten experimentellen georeferenzierten Bevölkerungszahlen visualisiert und beschrieben. Hierbei wurden erwartbare regionale Differenzen der experimentellen Bevölkerungsdichte sichtbar, die insbesondere urbane Gebiete im Vergleich zum ländlichen Raum stark hervorheben. Aufgrund der räumlich genaueren Aufbereitung der Mobilfunkaktivitäten im urbanen Raum (vgl. Abschn. 2; Abb. 3) war dies zu erwarten und unterstützt die Annahme, dass die Ergebnisse – insbesondere die Verteilung der experimentellen georeferenzierten Bevölkerungszahlen – in urbanen Regionen plausibel sind. Umgekehrt wird angenommen, dass Unsicherheiten insbesondere in ländlichen Gebieten auftreten und die Ergebnisse dort weniger plausibel erscheinen.

Da es sich bei den Ergebnissen um keine Schätzung im eigentlichen Sinne handelt, ist eine gängige Bestimmung von Unsicherheitsmaßen, wie dem Mean Squared Error (MSE) o. ä., nicht gegeben. Auch eine Prüfung der absoluten Werte der experimentellen georeferenzierten Bevölkerungszahl ist aktuell nicht möglich, da es hierfür keine vergleichbare Datengrundlage gibt. Dies wird erst mit den Ergebnissen des Zensus 2022 möglich sein.

Stattdessen wird eine zweistufige Plausibilisierung durchgeführt. Als erster Schritt zur Validierung der Ergebnisse aus Abschn. 4.1 werden diese den georeferenzierten Bevölkerungszahlen des Zensus 2011 gegenübergestellt, um eine erste Einschätzung der Ergebnisse zu erhalten. Im zweiten Plausibilisierungsschritt wird dann in Abschn. 4.3 die räumliche Verteilung der experimentellen georeferenzierten Bevölkerungszahl anhand aktueller amtlicher Geodaten auf Plausibilität geprüft. Grundsätzlich soll anhand dessen im ersten Plausibilisierungsschritt ermittelt werden, inwieweit sich die hier ermittelten experimentellen Bevölkerungszahlen von denen aus dem Zensus 2011 unterscheiden und worin die Unterschiede bzw. mögliche Fehlerquellen bestehen. Aufgrund der zeitlichen Differenz beider Datenquellen stellt dies nur eine grobe Annäherung dar und ermöglicht keine absoluten Aussagen. Entsprechend müssen die Befunde mit Vorsicht interpretiert werden.

Um regionale Unterschiede bei der Gegenüberstellung beider Bevölkerungszahlen feststellen zu können, werden diese anhand des Verstärterungsgrades in Deutschland aufgegliedert. Der Verstärterungsgrad wird nach der Definition von der Europäischen Kommission und Eurostat (2021) in drei Kategorien unterteilt, die anhand des Anteils der regionalen Bevölkerung ermittelt werden. Es wird nach Städten oder dicht besiedelten Gebieten unterschieden, die mindestens 50 % ihrer Bevölkerung in

¹⁹ Anhand eines vorliegenden Pearson-Korrelationskoeffizienten zwischen den hier verwendeten Mobilfunkdaten und den Einwohnerzahlen aus der Bevölkerungsfortschreibung 2019 auf Ebene der Gemeinden in Höhe von 9,999 kann bereits festgehalten werden, dass die Mobilfunkdaten gut auf höherer administrativer Ebene aufbereitet wurden und eine geeignete Datengrundlage für die kleinräumige Verteilung der amtlichen Bevölkerungszahlen aus der Bevölkerungsfortschreibung darstellen.

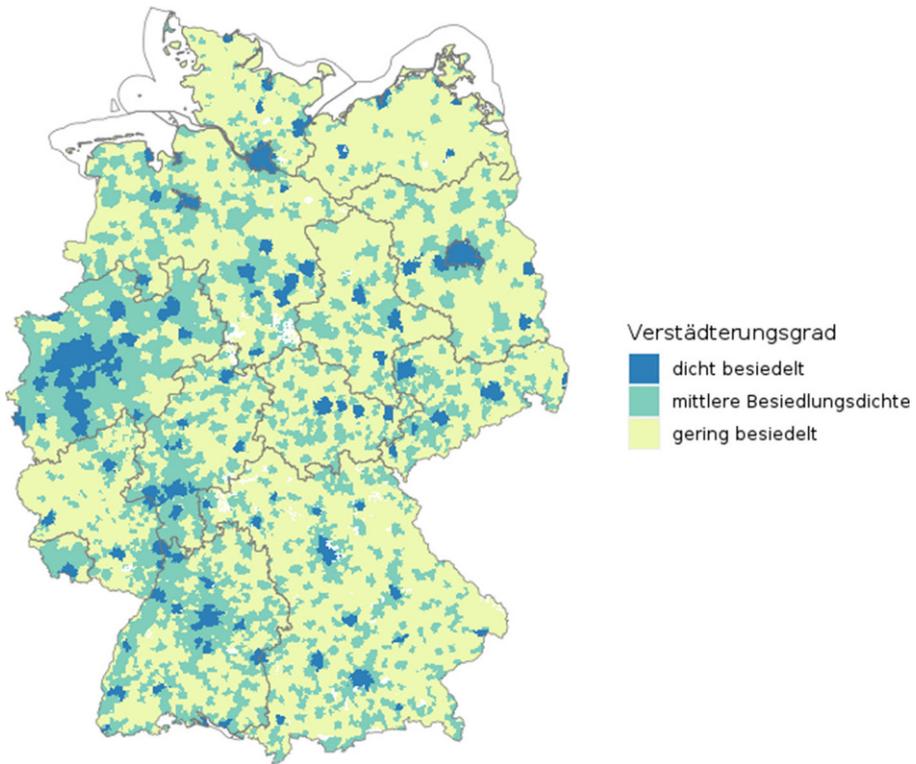


Abb. 7 Verstädterungsgrad in Deutschland

städtischen Zentren nachweisen. Hinzu kommen kleinere Städte und Vororte bzw. Gebiete mit mittlerer Bevölkerungsdichte, die weniger als 50 % ihrer Bevölkerung in städtischen Zentren und nicht mehr als 50 % ihrer Bevölkerung in ländlichen Gebieten vorweisen sowie ländliche Gebiete oder dünn besiedelte Gebiete, die mehr als 50 % ihrer Bevölkerung in ländlichen Gebieten aufweisen. Abb. 7 stellt den Verstädterungsgrad anhand der drei beschriebenen Kategorien kartografisch dar. Hierbei ist weiterhin entscheidend, wie sich der Flächenanteil sowie die deutsche Wohnbevölkerung auf die drei Raumtypen verteilen (siehe Tab. 2).

Aus Abb. 7 wird visuell ersichtlich, dass der urbane Raum den geringsten Flächenanteil in Deutschland ausmacht und der ländliche Raum dagegen den größten. In Deutschland gelten nach Tab. 2 rund ca. 2 % der Fläche als dicht besiedeltes Gebiet, ca. 24,5 % als Gebiet mit mittlerer Besiedlungsdichte und ca. 73,5 % als ländliches

Tab. 2 Flächen- und Bevölkerungsanteile nach Verstädterungsgrad in Deutschland

Grad der Verstädterung	Flächenanteil in %	Anteil der ansässigen Bevölkerung je Kategorie in %
dicht besiedelt	2,0	39,5
mittlere Besiedlungsdichte	24,5	40,3
gering besiedelt	73,5	20,2

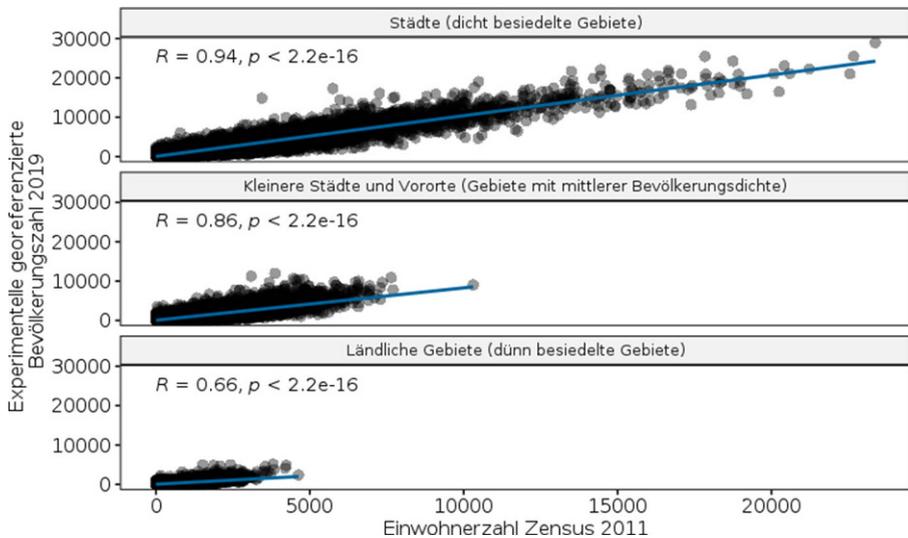


Abb. 8 Korrelationsdiagramm der experimentellen georeferenzierten Bevölkerungszahl 2019 und der Einwohnerzahl basierend auf dem Zensus 2011 nach Verstädterungsgrad

und demnach gering besiedeltes Gebiet. Jedoch wird in Tab. 2 gleichzeitig sichtbar, dass knapp 80 % der deutschen Bevölkerung auf nur rund 27 % der Fläche mit dichter sowie mittlerer Besiedlungsdichte angesiedelt sind. D. h. im Umkehrschluss, dass ein Großteil der Fläche in Deutschland gering besiedelt ist oder nicht bewohnt wird. Weiterhin wird beim Vergleich der beiden Abb. 6 und 7 deutlich, dass der Verstädterungsgrad sowie die räumliche Verteilung bzw. Dichte der experimentellen georeferenzierten Bevölkerungszahl im Bundesgebiet deckungsgleich sind.

Für die Feststellung möglicher regionaler Unterschiede bei der Gegenüberstellung beider Bevölkerungszahlen, werden in Abb. 8 nun die Pearson-Korrelationskoeffizienten der Gegenüberstellung bzw. die Zusammenhänge der experimentellen georeferenzierten Bevölkerungszahl 2019 und der georeferenzierten Einwohnerzahl basierend auf dem Zensus 2011 und den 1×1 km Gitterzellen differenziert nach dem Verstädterungsgrad betrachtet. Auf der x-Achse ist die absolute Einwohnerzahl aus dem Zensus 2011 hinterlegt und auf der y-Achse die experimentelle georeferenzierte Bevölkerungszahl 2019.

Zunächst geht aus den Korrelationskoeffizienten je Verstädterungsgrad in Abb. 8 insgesamt hervor, dass die Zusammenhänge beider Bevölkerungszahlen trotz zeitlicher Differenz von 8 Jahren stark positiv sind und mit einem Koeffizienten von maximal 0,94 in dicht besiedelten Gebieten einhergehen. Weiterhin wird ersichtlich, dass die Korrelationen mit abnehmender Bevölkerungszahl – in Form des Verstädterungsgrades und damit der Bevölkerungsdichte – stetig abnehmen. In ländlichen bzw. dünn besiedelten Gebieten fällt die Korrelation am geringsten aus, womit auch der positive Zusammenhang mit dem Zensus 2011 geringer wird. In den Gitterzellen, die den dünn besiedelten Gemeinden zugeordnet werden, wird außerdem dazu tendiert, die experimentelle georeferenzierte Bevölkerungszahl im Vergleich

Tab. 3 Mögliche Fehlschätzungen in der experimentellen georeferenzierten Bevölkerungszahl im Vergleich zum Zensus 2011 auf Ebene der 1×1 km Gitterzellen

Mögliche Fehlschätzungen im Verteilungsverfahren (laut Zensus 2011)	Fehlschätzungen pro Gitterzelle in %	Verstädterungsgrad in %		
		dicht besiedelt	mittlere Besiedlungsdichte	gering besiedelt
Keine Einwohner	39,0	3,0	27,3	69,7
Keine Geheimhaltung/ bewohnt	0,2	4,0	27,2	68,8

zum Zensus 2011 zu überschätzen, was durch die sichtbare Streuung oberhalb der Diagonalen in Abb. 8 (unterstes Streudiagramm) sichtbar wird. Hier besteht die Möglichkeit, dass durch die Überschätzung im ländlichen Raum eine Unterschätzung im städtischen Raum durch die anzunehmende fehlerhafte Verteilung in einigen Gitterzellen vorliegen könnte.

Tab. 3 schlüsselt weiterhin mögliche Fehlschätzungen in den experimentellen georeferenzierten Bevölkerungszahlen auf, die aus dem Vergleich mit den Bevölkerungszahlen aus dem Zensus 2011 auf Ebene der 1×1 km Gitterzellen resultieren. Hierbei fällt besonders auf, dass in 39 % aller Gitterzellen mit dem hier verwendeten Verteilungsverfahren Bevölkerung kleinräumig verteilt wird, in denen laut dem Zensus 2011 keine Einwohner gemeldet waren. Dies ist vorwiegend in dünn besiedelten Gemeinden mit knapp 70 % der betroffenen Gitterzellen der Fall sowie zu einem merklichen Anteil von 27,3 % in kleineren Städten und Vororten. In 0,2 % der Gitterzellen gibt die experimentelle georeferenzierte Bevölkerungszahl keine Bevölkerung oder einen geheim zuhaltenden Wert aus, in denen dies laut dem Zensus 2011 nicht der Fall ist. Offensichtlich wird aus Tab. 3 und Abb. 8, dass Unsicherheiten und Fehlschätzungen bei den kleinräumigen Zuordnungen der Bevölkerungszahl im ländlichen oder weniger dicht besiedelten Raum in den (vorwiegend modellierten) Gitterzellen angenommen werden müssen.

Die Erkenntnisse aus Tab. 2 relativieren hierbei die weniger guten Ergebnisse in ländlichen Gebieten in Abb. 8 sowie Tab. 3. Zwar wird ein flächenmäßig großer Anteil in Deutschland mit dem hier beschriebenen Verfahren mit tendenziell nicht plausiblen experimentellen georeferenzierten Bevölkerungszahlen ausgegeben. Jedoch wird demgegenüber aufgrund der Verteilung der Bevölkerung auf die drei kategorialen Gebiete angenommen, dass der Großteil der Bevölkerung besonders im urbanen Raum nachvollziehbar und plausibel kleinräumig verteilt wird. Diese Annahmen werden im folgenden zweiten Plausibilitätsschritt eingehender geprüft.

4.3 Zweite Plausibilitätsprüfung anhand amtlicher Geodaten

In einem zweiten Plausibilisierungsschritt werden die Erkenntnisse aus der Gegenüberstellung mit dem georeferenzierten Zensus 2011 aus Abschn. 4.2 aufgegriffen und die Ergebnisse anhand weiterer Datenquellen, genauer Geodaten aus amtlichen Vermessungsdaten, untersucht, die Informationen zu räumlichen Gegebenheiten in Bezug auf Landnutzung und Wohnflächen liefern. Sie werden verwendet, um die räumliche Verteilung der Ergebnisse in den Gitterzellen auf Plausibilität zu prüfen.

Der Vorteil bei der Nutzung von Vermessungsdaten bzw. Geodaten im Vergleich zu originären Fernerkundungsdaten liegt in den dort bereits aufbereiteten Geometrien. Sie enthalten belastbare Informationen zur Landnutzung und sind vergleichsweise einfach zu verarbeiten, können jedoch aufgrund der aufwändigen Datenaufbereitung seitens des BKG und der Vermessungsämter der Länder meist nur in einem Turnus von mehreren Jahren aktualisiert werden. Somit ist eine gewisse zeitliche Differenz zwischen den zugrundeliegenden Daten vorhanden.

Wie von Deville et al. (2014) bereits vorgeschlagen, wird eine Kombination aus Mobilfunkdaten sowie Geodaten, welche teilweise aus Fernerkundungs- oder genauer Satellitendaten hergeleitet oder mit diesen aktualisiert werden, umgesetzt. Jedoch wird anhand dessen keine zusätzliche Schätzung der experimentellen georeferenzierten Bevölkerungszahl wie in Schug et al. (2021) erfolgen, stattdessen wird anhand dieser Geodaten eine zusätzliche Plausibilitätsprüfung durchgeführt und bei Bedarf werden Verbesserungsmaßnahmen in der Aufbereitung der Mobilfunkdaten definiert.

Zu diesem Zweck werden aktuelle Geodaten der deutschen Landesvermessung des BKG verwendet, um die Verteilung der experimentellen georeferenzierten Bevölkerungszahl auf Plausibilität zu prüfen. Hierfür werden die *amtlichen Hausumringe Deutschland* (HU-DE) sowie die Geodaten aus dem Datensatz *Haushalte Einwohner Bund* (HH-EW-Bund) verwendet. Dabei wird analysiert, welche Gitterzellen Wohnflächen bzw. eine Wohnnutzung aufweisen und ob folglich eine Wohnbevölkerung in dieser zu erwarten ist oder ausgeschlossen werden kann.

Allein die Fläche zu betrachten, ist für die Plausibilisierung der Ergebnisse nicht ausreichend, da bspw. in ländlichen Gebieten Einfamilienhäuser eine relativ große Fläche für vergleichsweise wenige Einwohnerinnen und Einwohner beanspruchen können. Durch die Kombination der HU-DE mit den Daten des HH-EW-Bund lassen sich genauere Aussagen zur absoluten Verteilung der experimentellen georeferenzierten Bevölkerungszahl treffen. Hierzu werden die georeferenzierten Umringspolygone der Gebäudegrundrisse in Deutschland nach Wohngebäuden anhand der Gebäudefunktionskennung gefiltert und die Anzahl der Hausumrisse in der entsprechenden Gitterzelle bestimmt (BKG 2021b). Anhand dessen wird primär die räumliche Verteilung der experimentellen georeferenzierten Bevölkerungszahl auf Plausibilität geprüft. Zusätzlich werden die georeferenzierten Daten des HH-EW-Bund verwendet, um die Wohngebiete des HU-DE mit georeferenzierten Adressdaten (GA)²⁰ sowie die Anzahl der Haushalte pro Adresse und Gitterzelle zu ergänzen. Letztere werden dabei von der infas 360 GmbH anhand diverser Quellen sowie einer stichprobenhaften Erhebung hergeleitet (BKG 2021a). Der Stand des HH-EW-Bund Datensatzes ist auf Januar 2019 datiert und der des HU-DE auf das Jahr 2021, weshalb die Kombination beider Geodatenquellen aufgrund der zeitlichen Differenz in einigen Fällen zu nicht konsistenten Angaben führen kann.²¹

²⁰ Bei den verwendeten Adressen handelt es sich um die Haus-/Postanschrift aus den Quelldaten (BKG 2021a).

²¹ Grund für diese Differenz ist, dass die genutzte Gebäudefunktionskennung erst im HU-DE Datensatz von 2021 vorliegt.

Eine grundlegende Annahme, die bei der Plausibilitätsprüfung mit diesen Geodaten getroffen wird, ist, dass die Höhe der experimentellen georeferenzierten Bevölkerungszahl mit der Wohnfläche bzw. der Wohnbebauung, den Adresspunkten sowie der damit verbundenen Anzahl an Haushalten einhergeht. Konkret bedeutet dies, dass je mehr Adressdaten pro Gitterzelle vorliegen und je höher der Anteil der Wohnfläche bzw. der Hausumringe sowie der Anzahl der Haushalte pro Gitterzelle ist, desto höher sollte die experimentelle georeferenzierte Bevölkerungszahl tendenziell liegen. Im gleichen Maße geht damit einher, dass keine experimentelle georeferenzierte Bevölkerungszahl in einer Gitterzelle vorliegen darf, in der keine Wohnbebauung bzw. -fläche, Adressdaten etc. vorliegen.

Daraus ergeben sich folgende Kennzahlen im Sinne eines hier definierten Ampel-Konzepts, das die experimentelle georeferenzierte Bevölkerungszahl in drei Plausibilisierungskategorien einteilt: plausibel (grün), teils plausibel (gelb) und unplausibel (rot).

Als *unplausibel* gelten hierbei alle Gitterzellen mit experimentellen georeferenzierten Bevölkerungszahlen, die einem unbewohnten Gebiet bzw. den dazugehörigen Gitterzellen (keine Wohnfläche bzw. Wohngebäude, Adressen oder Haushalte) zugeordnet wurden oder keine in einer bewohnten Gitterzelle. Als *plausibel* gelten alle Bevölkerungszahlen, die in Gitterzellen verteilt wurden, die mit einem entsprechend hohen Wohnflächenanteil, einer hohen Anzahl an Adresspunkten und an Haushalten einhergehen sowie Gitterzellen, die eine experimentelle Bevölkerungszahl von Null angeben, in der auch keine Wohnflächen etc. ausgewiesen werden. Als *teils plausibel* werden Gitterzellen bewertet, deren zugeordnete experimentelle georeferenzierte Bevölkerungszahl zu hoch oder zu niedrig in Zusammenhang mit der vorhandenen Wohnfläche bzw. der Anzahl an Hausumringen, den Adresspunkten und den geführten Haushalten erscheinen. Hierbei wird zudem als Schwellenwert zur Bewertung der Höhe der experimentellen georeferenzierten Bevölkerungszahl die Anzahl der

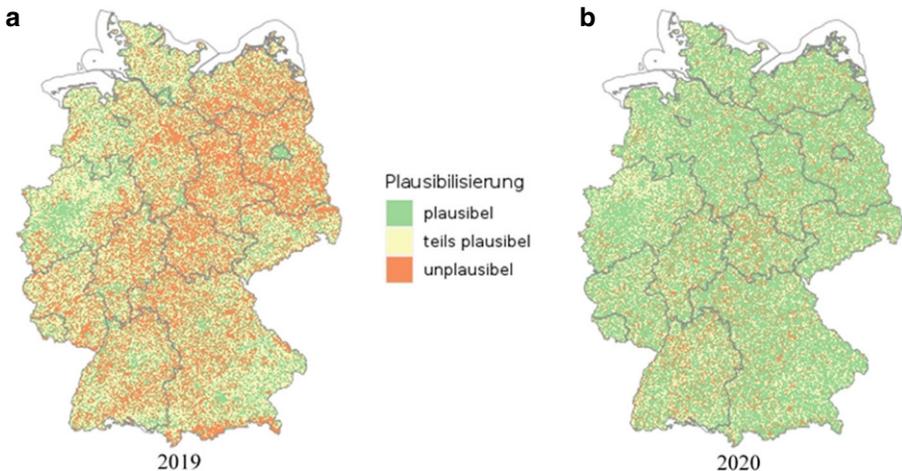


Abb. 9 Kartografische Visualisierung der Kennzahlen (Ampel-Konzept) der Plausibilitätsprüfung anhand der regionalen Ergebnisse für die Berichtsjahre 2019 (a) und 2020 (b) (mit räumlicher Korrektur der Mobilfunkdaten)

Tab. 4 Auflistung der Ergebnisse der berechneten Kennzahlen (Ampel-Konzept) nach dem Verstärterungsgrad

Grad der Verstärterung	Plausibilität der Ergebnisse in %		
	plausibel (grün)	teils plausibel (gelb)	unplausibel (rot)
dicht besiedelt	53,0	31,4	15,6
mittlere Besiedlungsdichte	30,6	39,1	30,3
gering besiedelt	23,2	36,9	39,9

Haushalte pro Gitterzelle sowie ihre durchschnittliche Anzahl an Personen in Höhe von zwei angeführt (Statistisches Bundesamt 2020).²²

Die in Abb. 9a visualisierten Ergebnisse dieser zweiten Plausibilitätsprüfung unterstützen die ersten Annahmen aus Abschn. 4.2, dass insbesondere in urbanen Räumen plausible experimentelle georeferenzierte Bevölkerungszahlen ermittelt bzw. zugewiesen werden. Umgekehrt bestätigt sich, dass ländliche Räume tendenziell häufiger mit unplausiblen Ergebnissen, aufgrund der hier verwendeten Mobilfunkdaten und des verwendeten Vertriebsverfahrens, einhergehen. Insgesamt geben die Kennzahlen rund 27,5 % der Gitterzellen und die ihnen zugeordneten experimentellen Bevölkerungszahlen als plausibel, 37,2 % als teils plausibel und rund 35,3 % als unplausibel an (vgl. Tab. 5). Hierbei wird erneut in Abb. 9a visuell deutlich, dass die weniger dicht besiedelten bzw. ländlichen Gebiete (vgl. mit Abb. 7) mit überdurchschnittlich vielen unplausiblen Werten einhergehen. Der Unterschied wird bspw. im Raum Berlin-Brandenburg ersichtlich.

Um weitere Rückschlüsse der Plausibilität der Ergebnisse durch regionale Unterschiede herzuleiten, werden die Ergebnisse aus Abb. 9a zusätzlich nach dem Verstärterungsgrad (vgl. Abschn. 4.2; Abb. 7) aufgeschlüsselt. Tab. 4 listet den gerundeten prozentualen Anteil der Kennzahlen nach Plausibilität – sowie farblich nach Abb. 9a – und nach dem Grad der Verstärterung auf.

Offensichtliche Unsicherheiten ergeben sich bei Zuordnungen der Bevölkerungszahlen in ländlichen, weniger dicht besiedelten Gebieten vorrangig durch die kleinräumige Aufbereitung der mobilen Aktivitäten beim Datenanbieter.²³ 53,0 % der Werte je Gitterzelle werden in dicht besiedelten Regionen als plausibel gekennzeichnet, während es in gering besiedelten Gebieten nur 23,2 % sind (vgl. Tab. 4). Dagegen sind nur 15,6 % der Gitterzellen in Städten mit nicht plausiblen experimentellen Bevölkerungszahlen versehen, während es 39,9 % der Gitterzellen im

²² Als Beispiel für ein teils plausibles Ergebnis wird eine Gitterzelle mit einer experimentellen georeferenzierten Bevölkerungszahl von 20 angenommen, in der jedoch über 200 Gebäude sowie Adresspunkte und doppelt so viele Haushalte verortet werden. Es kann hierbei davon ausgegangen werden, dass die tatsächliche Bevölkerungszahl aufgrund der hergeleiteten Wohnkapazitäten deutlich höher sein wird, als mit der experimentellen georeferenzierten Bevölkerungszahl berechnet und daher nur als teilweise plausibel eingestuft werden kann.

²³ Die plausiblen Werte verteilen sich weiterhin zu 65,6 % auf die nicht modellierten Mobilfunkaktivitäten bzw. die zugrundeliegenden Gitterzellen, zu 29,5 % auf die kleinräumig modellierten und nur zu 20,9 % auf die modellierten Gitterzellen. Dagegen sind über 43,5 % der unplausiblen Ergebnisse in den modellierten, 28,2 % in kleinräumig modellierten und nur 8,0 % in nicht modellierten Gitterzellen zu finden. Dies bestärkt weiterhin die Problematik der Mobilfunkdatenaufbereitung aus Abschn. 2.

ländlichen Gebiet sind. Die teils plausiblen Ergebnisse verteilen sich zu relativ gleichen Teilen auf alle Raumeinheiten bzw. etwas verstärkter in Gebieten mit mittlerer Besiedlungsdichte. Die teils plausiblen Werte sollten zudem mit Vorsicht interpretiert werden. Da es sich hierbei – aufgrund der grundsätzlich nachvollziehbaren räumlichen Verteilung aber der gleichzeitig nicht validierbaren Höhe der zugewiesenen Bevölkerung – um schwer einzuschätzende experimentelle georeferenzierte Bevölkerungszahlen handelt, wird von einer weiteren Analyse der absoluten Werte abgesehen. Insgesamt muss hierbei daher beachtet werden, dass die absoluten Werte der einzelnen experimentellen georeferenzierten Bevölkerungszahlen mit den Geodaten nicht hinreichend bzw. nicht absolut verglichen werden können. Eine Einschätzung hinsichtlich der Plausibilität der Verteilung und der Werte kann mit dem hier beschriebenen Verfahren aber grundsätzlich abgegeben werden.

4.4 Schlussfolgerungen der Plausibilitätsprüfung – Räumliche Korrektur der Mobilfunkdaten

Insgesamt bleibt damit die Schlussfolgerung bestehen, dass das hier durchgeführte Verfahren insbesondere im urbanen Raum zu nachvollziehbaren, plausiblen experimentellen georeferenzierten Bevölkerungszahlen führt. Ungefähr ein Drittel der Gitterzellen in gering besiedelten Gebieten und Gebieten mit mittlerer Besiedlungsdichte werden dagegen mit unplausiblen experimentellen georeferenzierten Bevölkerungszahlen ausgewiesen. Hierbei handelt es sich um unbewohnte Gitterzellen, denen durch die hier verwendeten Mobilfunkdaten und dem Verteilungsverfahren fälschlicherweise Bevölkerungszahlen zugewiesen werden, was auch in Abschn. 4.2 durch den Vergleich mit dem Zensus 2011 bereits angedeutet wurde. Diese offensichtlichen Fehlzusweisungen können allerdings nicht im Nachgang korrigiert werden. Hierfür müssen die Mobilfunkdaten, also die Datengrundlage anhand derer die Bevölkerungszahlen der Bevölkerungsfortschreibung kleinräumig verteilt werden, entsprechend modifiziert werden.

Gründe für die unplausible Verteilung speziell in gering besiedelten Gebieten sind vor allem in der Aufbereitung der Mobilfunkaktivitäten zu finden, die aus der Modellierung der Gitterzellen herrühren. Durch die probabilistische Verteilung der Aktivitäten durch den Datenanbieter ohne Einbezug weiterer Datenquellen als Hilfsinformationen (siehe Abschn. 2; Abb. 3) werden nicht bewohnte Regionen wie Naturschutzgebiete, Waldgebiete oder Industriegebiete bei der Verteilung der Aktivitäten stets zu gleichen Anteilen mitberücksichtigt.

Liegt nun wie in Deutschland der Fall vor, dass sich die Dichte des Mobilfunknetzes an der regionalen Bevölkerungsdichte orientiert, wird die Netzabdeckung in urbanen Räumen flächendeckender und engmaschiger und in ländlichen Räumen grobmaschiger und möglicherweise lückenhafter.²⁴ Dadurch sind Stadtzentren bzw. Innenstädte entsprechend mit kleinen Mobilfunkzellen und städtische Randbezirke und generell weniger dicht besiedelte Gebiete mit größeren ausgelegt. Daraus re-

²⁴ Siehe hierzu die Netzabdeckung in Deutschland von der Bundesnetzagentur im Mobilfunk-Monitoring nach Mobilfunknetzbetreiber und Technologie: <https://www.breitband-monitor.de/mobilfunkmonitoring/karte>.

sultieren in urbanen Räumen kleinräumige und räumlich genauere Verortungen von gezählten Mobilfunkaktivitäten als im weniger dicht besiedelten Raum.

Im Rahmen der Mobilfunkdatenaufbereitung konnte diese Netzabdeckung bei der Verteilung der Mobilfunkaktivitäten auf das hier verwendete 1×1 km Raster nicht einbezogen werden. Stattdessen wurden die Aktivitäten in weniger dicht besiedelten Gebieten mit einfachen Annahmen räumlich modelliert bzw. gleichmäßig verteilt. Folglich resultieren unplausible Werte aus den räumlich ungenauen Verteilungen und damit ergibt sich eine Unsicherheit in den Ausgangsdaten sowie den resultierenden Ergebnissen.

Die hieraus resultierende Konsequenz besteht in einer Verbesserung der Mobilfunkdatenaufbereitung beim Datenanbieter für das Berichtsjahr 2020, um anhand einer Modifizierung die Bevölkerungszahlen der Bevölkerungsfortschreibung auf kleinräumiger Ebene genauer zu verteilen. Dies wurde in Form einer Optimierung der Modellierung, genauer einer Steigerung der räumlichen Genauigkeit, von Mobilfunkaktivitäten auf das 1×1 km Raster anhand zusätzlicher realitätsbasierter Annahmen erzielt und durch Verwendung der Geodaten des *Landbedeckungsmodells für Deutschland* (LBM-DE) umgesetzt.²⁵

Eine präferierte Lösung, um die unbebaute Fläche strikt aus dem Modellierungsprozess der Mobilfunkdaten zu entfernen, gleichzeitig jedoch nicht Wohnflächen herauszufiltern bzw. auszuschließen, bietet eine Filterung der Gesamtfläche des Bundesgebietes. Hierfür wurden räumliche Gebiete nach Landbedeckungs- bzw. Landnutzungskategorien gefiltert, die nicht bebaut sind oder der Landnutzung *Wohnen*²⁶ nicht zugehörig sind, wie bspw. Industrieanlagen oder bebaute Flächen für den Verkehr, und wo prinzipiell keine Wohnbevölkerung verortet werden darf (siehe hierzu BKG 2020a). Somit werden alle nicht bebauten Flächen sowie Flächen ohne Siedlungsfunktion exkludiert, was in Deutschland ca. 90 % der Gesamtfläche ausmacht, die daraufhin für Verteilungsmaßnahmen nicht mehr beachtet werden.²⁷

Die Auswirkungen dieser räumlichen Korrektur auf die Ergebnisse werden in Abb. 9 präsentiert. Sie stellt die resultierenden Plausibilitätsprüfungen für das Berichtsjahr 2019 dem Berichtsjahr 2020 gegenüber.²⁸ Letzteres basiert dabei auf den experimentellen georeferenzierten Bevölkerungszahlen unter Verwendung der räumlich korrigierten Mobilfunkdaten. Hierbei wird bereits visuell deutlich, dass im Ergebnis deutlich plausiblere Ergebnisse durch die Verwendung von Landnutzungs-

²⁵ Der aktuelle Datenstand des LBM-DE bezieht sich auf das Referenzjahr 2018 und wird in einem Dreijahreszyklus aktualisiert. Dabei werden die amtlichen Vermessungsdaten des ATKIS-Basis-DLM für die Landnutzung sowie Bilddaten der RapidEye und Sentinel-2 Satelliten als auch digitale Orthophotos aus Überfliegungen der Landesvermessungsämter für die Landbedeckung genutzt (BKG 2020a). Insofern basiert die Abgrenzung von bspw. Wohngebieten oder Produktionsstandorten auf den Daten der Vermessungsbehörden der Länder und weisen eine hohe Genauigkeit auf.

²⁶ Darunter fallen auch Gebiete wie *Fußgängerzonen* oder *Grasland mit Bäumen zur Wohnnutzung gehörig* (BKG 2020a).

²⁷ Die Bodenfläche nach Nutzungsarten in Deutschland zum Stichtag 31.12.2020 kann auch nachvollzogen werden unter: <https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Landwirtschaft-Forstwirtschaft-Fischerei/Flaechennutzung/Tabellen/bodenflaeche-ingesamt.html>.

²⁸ Siehe hierzu auch die Veröffentlichung der Ergebnisse für das Berichtsjahr 2020 vom 6. Juli 2022 des Statistischen Bundesamtes (2022).

Tab. 5 Ergebnisse der räumlichen Anpassung für die experimentelle georeferenzierte Bevölkerungszahl

Plausibilität	Plausibilität der Ergebnisse in %	
	2019	2020
plausibel (grün)	27,5	67,8
teils plausibel (gelb)	37,2	22,1
unplausibel (rot)	35,3	10,1

informationen im Rahmen der Mobilfunkdatenaufbereitung erzeugt werden (vgl. Abb. 9b). Insgesamt nehmen damit die als plausibel eingestuften zugeordneten experimentellen Bevölkerungszahlen in Tab. 5 um über 40 Prozentpunkte zu, 22,1 % gelten als teils plausibel und nur noch 10,1 % als unplausibel.

Ferner werden nun 69,5 % der Werte je Gitterzelle in dicht besiedelten Regionen als plausibel gekennzeichnet sowie 69,4 % in gering besiedelten Gebieten. Dagegen sind nur noch 10,3 % der Gitterzellen im ländlichen Gebiet mit nicht plausiblen experimentellen Bevölkerungszahlen versehen. Die teils plausiblen Ergebnisse verteilen sich weiterhin verstärkter in Gebieten mit mittlerer Besiedlungsdichte.²⁹

Die zusätzliche Einbindung von Landnutzungsdaten bzw. Bebauungsinformation bei der Datenaufbereitung des Mobilfunkdatenanbieters bewirkt eine plausiblere Umverteilung der Mobilfunkdaten mit deutlich weniger Unsicherheit und eine höhere Qualität der Daten sowie der resultierenden Ergebnisse. Dadurch werden insgesamt nicht die Intensitäten der Mobilfunkaktivitäten beeinflusst, aber die Verteilungen deutlich verbessert, weshalb auch bereits plausibel geschätzte experimentelle georeferenzierte Gitterzellen Änderungen an ihren absoluten Werten erfahren haben. Ein weiterer positiver Nebeneffekt dieses Vorgehens liegt in einer leicht umsetzbaren und nachvollziehbaren Änderung der Methodik zur Datenaufbereitung beim Datenanbieter sowie dem damit einhergehenden Einfluss und der Mitgestaltung bei der Mobilfunkdatenaufbereitung.

5 Fazit und Schlussfolgerungen

Im Projekt *Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten* werden im Rahmen eines Verteilungsverfahrens die Ergebnisse der Bevölkerungsfortschreibung von der Gemeindeebene anhand von Mobilfunkdaten bundesweit auf INSPIRE-konforme 1 × 1 km Gitterzellen kleinräumig umverteilt. Durch den starken nachweisbaren Zusammenhang zwischen Mobilfunkdaten und amtlichen Bevölkerungszahlen bieten Mobilfunkdaten eine geeignete Datenquelle, um die Wohnbevölkerung in Deutschland kleinräumig zu verteilen. Mit diesem Vorgehen wird die vorhandene amtliche Bevölkerungsfortschreibung um ein experimentelles kleinräumiges Ergebnis ergänzt und kann damit zur Schließung der Lücke einer fehlenden aktuellen und georeferenzierten Bevölkerungszahl beitragen, bis die ersten georeferenzierten Bevölkerungszahlen auf Basis

²⁹ Die unplausiblen Ergebnisse verteilen sich zudem nur noch zu 14,8 % in den modellierten, 13,0 % in kleinräumig modellierten und nur zu 5,0 % in nicht modellierten Gitterzellen.

des Zensus 2022 vorliegen. Durch die Erstellung und Veröffentlichung eines frei zugänglichen experimentellen Produktes stehen die Ergebnisse zudem uneingeschränkt zur Verfügung. Weiterhin ist das Verfahren zur Erstellung experimenteller kleinräumiger Bevölkerungszahlen grundsätzlich bei allen weiteren statistischen Ämtern umsetzbar, die eine laufende Bevölkerungsfortschreibung erstellen sowie Zugang zu anonymisierten und aggregierten Mobilfunkdaten haben.

Neben den üblichen Nutzungszwecken der amtlichen Bevölkerungsfortschreibung für politische Entscheidungsfindungen können weitere Anwendungsfälle wie bei Fina et al. (2019) oder im Zusammenhang mit dem Krankenhaus-Atlas der Statistischen Ämter des Bundes und der Länder aufgegriffen werden. Grundsätzlich werden die experimentellen georeferenzierten Bevölkerungszahlen sämtlichen Forschungsvorhaben zugutekommen, die aktuelle und kleinräumige Bevölkerungszahlen benötigen und für die bislang nur die georeferenzierten Bevölkerungszahlen des Zensus 2011 zur Verfügung standen. Zwar erfüllen die Ergebnisse nicht den Qualitätsanspruch der amtlichen Statistik, jedoch wird deren Qualität mit den Erkenntnissen aus der Plausibilitätsprüfung verbessert.

Durch das hier verwendete Verteilungsverfahren resultiert ein regional differenziertes Bild der Bevölkerung, das im Schnitt ein plausibles Ergebnis insbesondere im urbanen Raum darstellt. Fehlzusweisungen kleinräumiger Bevölkerungszahlen werden durch die Unsicherheiten im ländlichen Raum durch die Modellierung der dort befindlichen Gitterzellen und ihrer Mobilfunkaktivitäten hervorgerufen, die aufgrund der probabilistischen Zuordnung beim Datenanbieter zu unplausiblen Ergebnissen führen kann. Dies ist verstärkt in den Ergebnissen des Berichtsjahres 2019 der Fall. Die Verwendung von amtlichen Geodaten des HU-DE und des HH-EW-Bund ermöglichen in der Plausibilisierungsprüfung, diese Bereiche zu identifizieren und anhand des Verstädterungsgrades zu kategorisieren. Bei diesem Verfahren ist eine genaue, uneingeschränkte Prüfung der absoluten Zahlen der experimentellen georeferenzierten Bevölkerungszahl allerdings nicht gegeben. Jedoch können anhand der verwendeten Geodaten Tendenzen zur Höhe der Werte angegeben werden. Die ultimative Validierungsgrundlage stellen daher die georeferenzierten Ergebnisse des Zensus 2022 dar, welche voraussichtlich im Jahr 2023 zur Verfügung stehen.

Aufgrund der Resultate der hier durchgeführten Plausibilitätsprüfung in den zu ermittelnden experimentellen georeferenzierten Bevölkerungszahlen für das Berichtsjahr 2019 kam es folglich zu einer Modifizierung der Mobilfunkdatenaufbereitung anhand des LBM-DE, indem die unbewohnten Flächen in der kleinräumigen Verteilung der mobilen Aktivitäten vom Datenanbieter nicht wieder einbezogen wurden und damit zu einer sichtlichen Verbesserung der Ergebnisse für das Berichtsjahr 2020 führte. Wie Deville et al. (2014) bereits vorgeschlagen haben oder auch Schug et al. (2021) in anderer Form umsetzen, kommt es dadurch für die nachfolgenden Berichtsjahre von einer indirekten (in Form der Plausibilitätsprüfung) zu einer direkten Kombination (in Form der Modifizierung der Mobilfunkdatenaufbereitung) aus Geodaten aus dem LBM-DE sowie Mobilfunkdaten, was zur Qualitätssteigerung der Ergebnisse führt.

Dennoch wird durch die weiterhin bestehende Abhängigkeit vom Datenanbieter hinsichtlich der Mobilfunkdatenaufbereitung die Qualität der Ergebnisse tangiert bzw. eine Aussage hierzu deutlich erschwert, so dass diese nicht die Qualität der

amtlichen Statistik vorweisen können. Durch die Verwendung von Mobilfunkaktivitäten nur eines Netzanbieters in Deutschland werden dadurch entstehende Abweichungen und Unsicherheiten in den Ergebnissen sowie in den soziodemografischen Merkmalen durch die jeweiligen regionalen Marktanteile und die verwendete und nicht im Detail offengelegte Methodik des Datenanbieters bei der Datenaufbereitung bedingt (siehe hierzu auch Hadam 2021). Verzerrungen in unterschiedlichen Regionen können durch ein besseres Extrapolationsverfahren bzw. Gewichtungsverfahren entgegengewirkt werden. Hierzu müssen Stärke und Lage der Verzerrung im Raum bekannt sein, wie es ansatzweise in Statistisches Bundesamt (2019, 2021c) aufgeführt wird. Für ein zielorientierteres Gewichtungsverfahren sind weitere Informationen zur Mobilfunknutzung in der Bevölkerung erforderlich. Diese könnten bei zukünftigen Haushaltserhebungen, wie dem Mikrozensus, durch zusätzliche Fragen, zum genutzten Mobilfunkanbieter sowie zur Anzahl und Nutzung von Mobilfunkgeräten ermittelt werden, ohne die Auskunftgebenden hierbei zu stark zu belasten.

Neben der schwierigen Ableitung weiterer Merkmale bleibt grundsätzlich auch die Positionsschätzung der mobilen Aktivitäten, genauer der aktiven SIM-Karten, unterschiedlich präzise, wie in Saidani et al. (2022) aufgeführt. Sie zeigen auf, dass die Ableitung von Positionen der SIM-Karten aufgrund des ungleichförmigen Mobilfunknetzes bspw. anhand von Voronoi-Polygonen keine genaue Schätzung der mobilen Aktivitäten ermöglicht. Fehlzureisungen in dem hier beschriebenen Verteilungsverfahren resultieren daher aus Beeinträchtigungen in der Datenverfügbarkeit durch das Mobilfunknetz des Anbieters, sofern Regionen nicht durch das entsprechende Mobilfunknetz abgedeckt werden oder eine genaue Verortung der Mobilfunkaktivitäten aufgrund zu großer Mobilfunkzellen nicht gegeben ist. Ausfälle von Mobilfunkmasten spielen hierbei keine ausschlaggebende Rolle, da für die Erstellung der experimentellen georeferenzierten Bevölkerungszahl ein Jahresdurchschnitt der mobilen Aktivitäten gebildet wird. Auch mögliche Doppelzählungen von SIM-Karten in den Mobilfunkdaten, wie bspw. durch Zweitverträge oder SIM-Karten aus anderen nicht personenbezogenen Geräten provoziert, die nicht durch eine Deduplizierung ausgeschlossen werden konnten, können ebenfalls Einschränkungen in den Ergebnissen hervorrufen. Da weiterhin keine modellbasierte Schätzung vorliegt, ist die Umsetzung einer klassischen Varianzschätzung nicht gegeben, die besonders vor dem Hintergrund der Qualitätsstandards der amtlichen Statistik notwendig ist. Hierbei eignen sich insbesondere Small-Area-Methoden, um die Genauigkeit von kleinräumig geschätzten Bevölkerungsschätzungen zu evaluieren oder auch eine Varianzreduzierung zu bewirken (Simpson et al. 1996; Rao und Molina 2015). Insgesamt bleiben daher Einschränkungen in der Qualitätseinschätzung der Ergebnisse bestehen.

Die grundsätzlich angestrebte Nutzung von Mobilfunkdaten für die Produktion amtlicher Statistiken kann letztlich nur anhand von Daten aller Mobilfunkanbieter in Deutschland erfolgen, um die bundesweite Repräsentativität und Qualität der Daten einschätzen und verbessern zu können. Hierfür bedarf es der Schaffung einer Rechtsgrundlage, um den Zugang zu privat gehaltenen Daten zu ermöglichen und dauerhaft zu sichern und diese langfristig in die amtliche Statistikproduktion integrieren zu können. Erst wenn diese Datenquelle für die amtliche Statistik dauerhaft

und vollständig zugänglich ist, kann diese dem Qualitätsanspruch der amtlichen Statistik entsprechend aufbereitet und dauerhaft genutzt werden.

6 Anhang

6.1 A.1. Zusätzliche Information zur Mobilfunkdatenwahl für die kleinräumige Verteilung der Bevölkerungszahlen

Die Wahl der Mobilfunkdatenstrategie hat Auswirkungen auf die Berechnung der experimentellen georeferenzierten Bevölkerungszahl, auf welche nachfolgend eingegangen wird.

Tab. 6 listet die Ergebnisse der nach Abschn. 3.2 berechneten experimentellen georeferenzierten Bevölkerungszahlen nach den drei möglichen Optionen der Datenwahl (statistischer Sonntagabend, Heimatort Werktagsdurchschnitt ,erstes u./o. letztes Signal‘ und ,nur identisches erstes u. letztes Signal‘) anhand einer Summenstatistik. Wie bereits genannt, wird bei der Heimatort-Strategie aufgrund der Art

Tab. 6 Summenstatistik der berechneten experimentellen georeferenzierten Bevölkerungszahlen nach Mobilfunkdatenwahl

Mobilfunkdatenwahl/-Strategie	Min.	1st. Qu.	Median	Mean	3rd Qu.	Max.
Statistischer Sonntagabend	0	26	55	230,2	128	29.008
Heimatort Werktagsdurchschnitt (erstes u./o. letztes Signal)	0	26	56	230,2	130	28.846
Heimatort Werktagsdurchschnitt (nur identisches erstes u. letztes Signal)	0	26	55	230,2	128	28.914

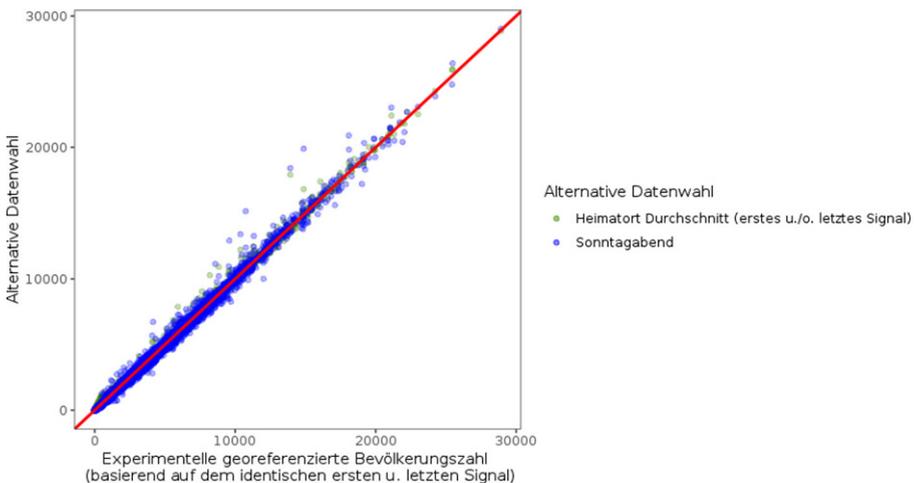


Abb. 10 Vergleich der Verteilungen der experimentellen georeferenzierten Bevölkerungszahlen nach Mobilfunkdaten-Strategie

der Mobilfunkdatenaufbereitung durch das erste und letzte Signal innerhalb von 24 Stunden jeweils ein Werktagsdurchschnitt von Montag bis Donnerstag gebildet, um veränderte Gewohnheiten an Wochenenden durch die Mobilfunknutzenden zu exkludieren. Offensichtlich wird, dass sich die Ergebnisse im Durchschnitt nicht voneinander unterscheiden und sich ausschließlich Unterschiede ab dem 3. Quantil zeigen. Im Maximum findet sich bei allen drei Optionen in einer 1×1 km Gitterzelle eine experimentelle georeferenzierte Bevölkerungszahl von fast 30.000. Die final ausgewählten und diskutierten Ergebnisse sind in Tab. 6 kräftig hervorgehoben.

Abb. 10 visualisiert die Verteilung der berechneten experimentellen georeferenzierten Bevölkerungszahlen nach der jeweiligen Mobilfunkdaten-Strategie aus Tab. 6 weiterhin in einem Streudiagramm. Auf der x-Achse sind die Ergebnisse des räumlich identischen ersten und letzten Signals laut der Heimatort-Strategie abgebildet. Auf der y-Achse werden die Ergebnisse nach den alternativen Optionen abgetragen – der statistische Sonntagabend sowie der Werktagsdurchschnitt ‚erstes u./o. letztes Signal‘. Hier wird noch einmal sichtbar, dass sich die Ergebnisse nicht signifikant voneinander unterscheiden. Jedoch fällt in Abb. 10 eine merkbare Streuung um die rote Diagonale (entspricht einer identischen Verteilung der Ergebnisse) auf, die insbesondere durch die unterschiedlichen Ergebnisse der Sonntagabend-Strategie entsteht (blaue Punkte). Diese schätzen die experimentelle georeferenzierte Bevölkerungszahl in Gitterzellen ab einer geschätzten Bevölkerungszahl von 10.000 tendenziell höher ein als die Heimatort-Strategie ‚nur identisches erstes u. letztes Signal‘.

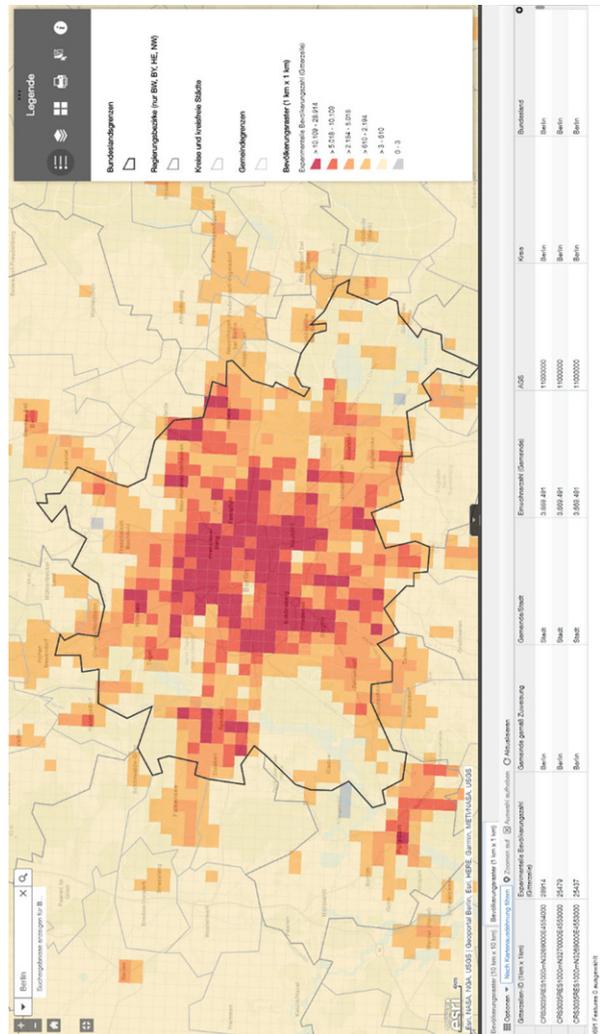
Weiterhin gibt es wenig sichtbare Unterschiede bei den beiden Heimatort-Strategien. Da anhand der Abb. 2 jedoch bereits der negative Einfluss der separaten ersten und letzten Signale zum Vorschein kam, werden im Artikel (vgl. Abschn. 4.1) die experimentellen georeferenzierten Bevölkerungszahlen anhand der räumlich identischen ersten und letzten Mobilfunksignale innerhalb von 24 Stunden erstellt. Der statistische Sonntagabend wurde vor allem durch die schlechtere Flächenabdeckung und Datenverfügbarkeit (vgl. Tab. 1) nicht ausgewählt.

6.2 A.2. Die experimentelle georeferenzierte Bevölkerungszahl als interaktive Karte

Die Ergebnisse der experimentellen georeferenzierten Bevölkerungszahl können anhand einer ArcGIS Online-Anwendung auf der Seite *Statistik visualisiert* des Statistischen Bundesamtes individuell und nutzerfreundlich gefiltert, heruntergeladen und für weitere Forschungsfragen ausgewertet werden. Die erstmals am 14. Feb. 2022 veröffentlichte Anwendung ist unter https://www.destatis.de/DE/Service/Statistik-Visualisiert/Bevoelkerung-Geo/Bevoelkerung_Karten.html zu finden.

Der Mehrwert der experimentellen kleinräumigen Bevölkerungszahlen wird insbesondere auf der 1×1 km Ebene unter Verwendung zusätzlicher Geodaten-Layer sichtbar. Die experimentelle georeferenzierte Bevölkerungszahl liegt auf Basis von 1×1 km sowie zusätzlich von 10×10 km Gitterzellen vor. Dabei wird für jede Gitterzelle die ermittelte experimentelle georeferenzierte Bevölkerungszahl ausgewiesen, sofern die Zellen zum aktuellen Zeitpunkt mit mobilen Aktivitäten gefüllt sind und

Abb. 11 Ausschnitt der interaktiven Karte im Internetauftritt des Statistischen Bundesamtes (https://www.destatis.de/DE/Service/Statistik-Visualisiert/Bevoelkerung-Geo/Bevoelkerung_Karten.html) zur interaktiven Darstellung der experimentellen georeferenzierten Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten



zusätzlich nicht der Anonymisierung sowie der anschließenden Geheimhaltung unterliegen.

Abb. 11 stellt einen Ausschnitt der interaktiven Anwendung dar. Beispielhaft werden die Ergebnisse für den Raum Berlin dargestellt. Abgebildet werden die experimentellen georeferenzierten Bevölkerungszahlen mit ansteigender Anzahl: In hell eingefärbten Zellen fällt die experimentelle georeferenzierte Bevölkerungszahl gering aus, in dunklen Zellen ist sie höher. Dies ermöglicht zusätzlich einen regionalen Vergleich der aktuellen Bevölkerungsverteilung. Erwartungsgemäß zeigen sich deutliche Unterschiede in der regionalen Verteilung der experimentellen georeferenzierten Bevölkerungszahl zwischen städtischen und ländlichen Gebieten.

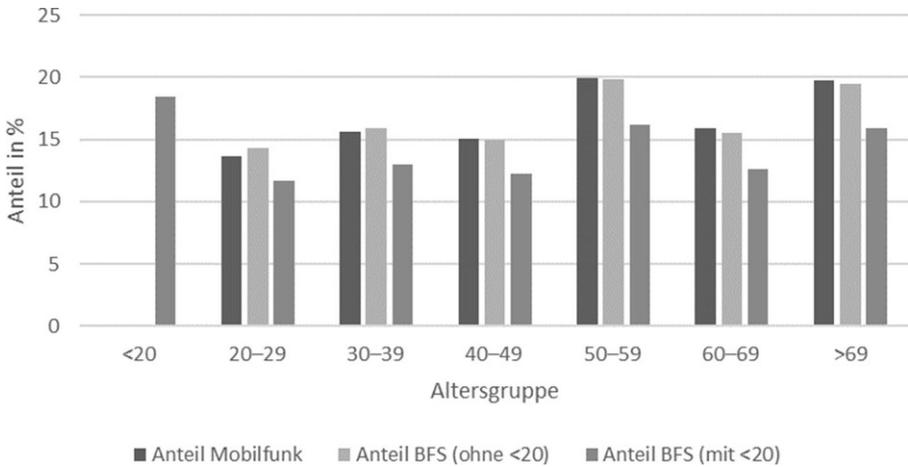


Abb. 12 Gegenüberstellung der Verteilungen der Altersgruppen nach der Bevölkerungsfortschreibung (BFS) 2019 sowie nach den vorliegenden Mobilfunkdaten aus dem Netz der Telefónica Deutschland

6.3 A.3. Verwendung soziodemografischer Merkmale für die experimentelle georeferenzierte Bevölkerungsfortschreibung

Wie bereits in Abschn. 4.1 erläutert, wurde die Anwendung des Verteilungsverfahrens auf die soziodemografischen Angaben, wie Altersgruppe und Geschlecht, geprüft und aufgrund der starken Verzerrungen in den soziodemografischen Angaben der Mobilfunkanbieter als nicht umsetzbar erachtet.

Abb. 12 stellt zur Veranschaulichung der vorliegenden Verzerrung beispielhaft die Verteilungen der Altersklassen nach der Bevölkerungsfortschreibung 2019 sowie nach den vorliegenden Mobilfunkdaten aus dem Netz der Telefónica Deutschland in einem Balkendiagramm dar. Es werden die jeweiligen Anteile nach Altersgruppe angegeben, die in den Daten vorliegen. Abb. 12 unterscheidet zusätzlich die Verteilung der Altersklassen in der Bevölkerungsfortschreibung mit und ohne die minderjährige Bevölkerung, wobei Letzteres durch die Verteilung der Altersgruppe in den Mobilfunkdaten bedingt wird. Hierbei wird vereinfachend mit der Altersgruppe der unter 20-Jährigen der Anteil der Minderjährigen abgedeckt. Die Altersgruppen 18 und 19 Jahre werden für die Vergleichbarkeit in die Altersgruppe 20–30 gezählt, da dies in dieser Form einfachheitshalber durch den Datenanbieter umgesetzt wurde.

Durch die ausschließliche Nutzung der Merkmale der Vertragskundinnen und -kunden wird bereits aus Abb. 12 ersichtlich, dass anhand dieser Datenquelle keine Angaben zu den Minderjährigen gemacht werden können und diese ausgehend von der Bevölkerungsfortschreibung die größte Gruppe aller dargestellten Altersgruppen ausmachen. Entsprechend ist keine plausible Darstellung der Bevölkerungsverteilung nach Altersgruppen möglich. Demzufolge ist auch eine Berechnung der Anteile der verschiedenen Altersgruppen für die 1×1 km Gitterzellen nicht gegeben, da bereits die Grundgesamtheit fehlerhaft ist. Bei der Nutzung des Merkmals Geschlecht aus den Mobilfunkdaten sind die Folgen weniger drastisch als bei den Altersgruppen, jedoch sind die ungleichen Verteilungen und sichtbaren Verzerrungen zwischen

Mobilfunkdaten und Bevölkerungsfortschreibung wie in vorherigen Untersuchungen des Statistischen Bundesamtes (2021c) auch hier weiterhin der Fall, zumal ebenfalls beim Merkmal Geschlecht alle Minderjährigen sowie Prepaid-Kundinnen und -Kunden nicht einbezogen werden.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt S. Hadam gibt an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- BBSR – Bundesinstitut für Bau-, Stadt- und Raumforschung (2021) Der demografische Wandel. Ein wichtiger Faktor für die Entwicklung regionaler Teilmärkte. Dezembertagung des DGD-Arbeitskreises „Städte und Regionen“ in Kooperation mit dem BBSR Bonn am 5. und 6. Dezember 2019 in Berlin. BBSR-Online-Publikation 01, S 4–14, Bonn
- BKG – Bundesamt für Kartographie und Geodäsie (2020a) Dokumentation: Digitales Landbedeckungsmodell für Deutschland – LBM-DE2018. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/lbm-de2018.pdf. Zugegriffen: 26. Febr. 2022
- BKG – Bundesamt für Kartographie und Geodäsie (2020b) Dokumentation: Geographische Gitter für Deutschland – GeoGitter. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/geogitter.pdf. Zugegriffen: 20. Jan. 2022
- BKG – Bundesamt für Kartographie und Geodäsie (2021a) Dokumentation: Haushalte Einwohner Bund – HH-EW-Bund. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/hh-ew-bund.pdf. Zugegriffen: 26. Febr. 2022
- BKG – Bundesamt für Kartographie und Geodäsie (2021b) Dokumentation: Hausumringe Deutschland – HU-DE. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/hu-de.pdf. Zugegriffen: 10. März 2022
- Burgdorf M (2010) Disaggregation von Bevölkerungsdaten mittels ATKIS Basis DLM. In: Strobl J et al (Hrsg) *Angewandte Geoinformatik 2010*. Wichmann, Heidelberg, S 474–483
- Deville P, Linard C, Martin S, Gilbert M, Stevens FR et al (2014) Dynamic population mapping using mobile phone data. *Proc Natl Acad Sci USA* 111(45):15888–15893
- Dougllass RW, Meyer DA, Ram M, Rideout D, Song D (2015) High resolution population estimates from telecommunications data. *EPJ Data Sci* 4(1):4. <https://doi.org/10.1140/epjds/s13688-015-0040-6>
- Europäische Kommission, Eurostat (2021) *Applying the Degree of Urbanisation: a methodological manual to define cities, towns and rural areas for international comparisons—2021 edition*. Publications Office of the European Union, Luxembourg <https://doi.org/10.2785/706535>
- Fina S, Gerten C, Gehring-Fitting K, Rönsch J (2019) Geomonitoring und die große Transformation – Methoden zur kritischen Bewertung nachhaltiger Raumentwicklung. ILS-TRENDS [extra]. <https://www.ils-forschung.de/files/publikationen/pdfs/trends-dez-19.pdf>. Zugegriffen: 7. Nov. 2021
- Hadam S (2021) Pendler Mobil: Die Verwendung von Mobilfunkdaten zur Unterstützung der amtlichen Pendlerstatistik. *ASTa Wirtsch Sozialstat Arch* 15:197–235. <https://doi.org/10.1007/s11943-021-00294-z>

- Hadam S, Schmid T, Simm J (2020) Kleinräumige Prädiktion von Bevölkerungszahlen basierend auf Mobilfunkdaten aus Deutschland. In: Klumpe B, Schröder J, Zwick M (Hrsg) Qualität bei zusammengeführten Daten. Springer, Wiesbaden, S 27–44 https://doi.org/10.1007/978-3-658-31009-7_3
- IT NRW (2018) Raum neu gefasst: Wie verteilen sich Arztpraxen in Nordrhein-Westfalen? Potenziale georeferenzierter Auswertungen des statistischen Unternehmensregisters. Statistik kompakt 09/2018. <https://webshop.it.nrw.de/gratis/Z259%20201859.pdf>. Zugegriffen: 7. Nov. 2021
- Kirchner T, Pflanz F, Techen A, Wagenknecht L (2014) Kleinräumige Gliederung, Georeferenzierung und Rasterdarstellung im Zensus. *Z Amtliche Stat Berlin Brandenburg* 3:28–32
- Koebe T, Arias-Salazar A, Rojas-Perilla N, Schmid T (2022) Intercensal updating using structure-preserving methods and satellite imagery. *J R Stat Soc Ser A Stat Soc* 185(Suppl. 2):170–196. <https://doi.org/10.1111/rssa.12802>
- Leyk S, Gaughan AE, Adamo SB, de Sherbinin A, Balk D et al (2019) The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth Syst Sci Data* 11:1385–1409. <https://doi.org/10.5194/essd-11-1385-2019>
- Lloyd CT, Sorichetta A, Tatem AJ (2017) High resolution global gridded data for use in population studies. *Sci Data* 4(1):10017. <https://doi.org/10.1038/sdata.2017.1>
- Norman P, Simpson L, Sabater A (2008) “Estimating with confidence” and hindsight: New UK small area population estimates for 1991. *Popul Space Place* 14(5):449–472
- Rao JNK, Molina I (2015) Small area estimation, 2. Aufl. Wiley, New York
- Rees P, Brown D, Norman P, Dorling D (2003) Are socioeconomic inequalities in mortality decreasing or increasing within some British regions? An observational study, 1990–98. *J Public Health Med* 25(3):208–214
- Saidani Y, Bohnensteffen S, Hadam S (2022) Qualität von Mobilfunkdaten – Projekterfahrungen und Anwendungsfälle aus der amtlichen Statistik. *WISTA Wirtsch Stat* 74(5):55–67
- Schug F, Frantz D, van der Linden S, Hostert P (2021) Gridded population mapping for Germany based on building density, height and type from Earth Observation data using census disaggregation and bottom-up estimates. *PLoS ONE* 16(3):e249044. <https://doi.org/10.1371/journal.pone.0249044>
- Simpson S, Diamond I, Tonkin P, Tye R (1996) Updating small area population estimates in England and Wales. *J R Stat Soc Ser A Stat Soc* 159(2):235–247
- Statistisches Bundesamt (2019) Bevölkerungsdarstellung mit Mobilfunkdaten. <https://www.destatis.de/DE/Service/EXDAT/Datensaeetze/mobilfunkdaten.html>. Zugegriffen: 18. Nov. 2021
- Statistisches Bundesamt (2020) 2040 wird voraussichtlich jeder vierte Mensch in Deutschland alleine wohnen – Pressemitteilung Nr. 069 vom 2. März 2020. https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/03/PD20_069_122.html. Zugegriffen: 24. Nov. 2021
- Statistisches Bundesamt (2021a) Ausstattung mit Gebrauchsgütern – Daten aus den Laufenden Wirtschaftsrechnungen (LWR) zur Ausstattung privater Haushalte mit Informationstechnik. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Ausstattung-Gebrauchsgueter/Tabellen/a-infotechnik-d-lwr.html>. Zugegriffen: 19. Febr. 2022
- Statistisches Bundesamt (2021b) Fortschreibung des Bevölkerungsstandes (Bevölkerungsfortschreibung 2020). Qualitätsbericht. https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/bevoelkerungsfortschreibung-2020.pdf?__blob=publicationFile. Zugegriffen: 4. Nov. 2021
- Statistisches Bundesamt (2021c) Strukturvergleich von Mobilfunkdaten zweier Mobilfunkanbieter. <https://www.destatis.de/DE/Service/EXDAT/Datensaeetze/mobilfunkanbieter-strukturvergleich.html>. Zugegriffen: 21. Nov. 2021
- Statistisches Bundesamt (2022) Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten. <https://www.destatis.de/DE/Service/EXDAT/Datensaeetze/bevoelkerung-geo-mobilfunkdaten.html>. Zugegriffen: 2. Nov. 2022
- Steinnocher K, Petri F, Tötzer T, Weichselbaum J (2005) Räumliche Disaggregation von sozio-ökonomischen Daten. In: Strobl J et al (Hrsg) Angewandte Geoinformatik 2005 – 17. AGIT-Symposium. Wichmann, Heidelberg, S 702–707
- Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* 10(2):e107042. <https://doi.org/10.1371/journal.pone.0107042>
- Wonka E, Kaminger I, Katzlberger G (2009) Regionalstatistische Auswertungen mit geographischen Rastern in der Raumplanung. *Inf Raumentwickl* 10/11:661–675

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.