Generierung synthetischer Bevölkerungen für Verkehrsnachfragemodelle, ein Methodenvergleich am Beispiel von Berlin

Dipl. Inf. (FH) Antje von Schmidt

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Verkehrsforschung, Rutherfordstraße 2, 12489 Berlin, Tel. +49-30-67055295, Fax. +49-30-67055283, E-Mail: antje.vonschmidt@dlr.de

Dipl. Geogr. Rita Cyganski

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Verkehrsforschung, Rutherfordstraße 2, 12489 Berlin, Tel. +49-30-67055147, Fax. +49-30-67055283, E-Mail: rita.cyganski@dlr.de

Dipl. Inf. Daniel Krajzewicz

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Verkehrsforschung, Rutherfordstraße 2, 12489 Berlin, Tel. +49-30-67055273, Fax. +49-30-67055283, E-Mail: daniel.krajzewicz@dlr.de

Zusammenfassung

Verkehrsnachfragemodelle benötigen detaillierte Informationen zur Bevölkerung im Untersuchungsgebiet. In dem vorliegenden Artikel werden verschiedene Methoden zur Erzeugung synthetischer Bevölkerungen vorgestellt und hinsichtlich ihrer Datenanforderungen sowie ihrer Ergebnisqualität und ihrer Einsetzbarkeit besprochen. Das Hauptaugenmerk liegt dabei auf stichprobenbasierten Verfahren. Für drei verschiedene Methoden der Bevölkerungsgenerierung werden anhand des Untersuchungsgebietes Berlin 2010 die resultierenden synthetischen Bevölkerungen untereinander sowie mit realen Bevölkerungsdaten verglichen. Es zeigt sich, dass durch eine zusätzliche Nutzung von Randsummen die Qualität der Ergebnisse verbessert wird.

1 Einleitung

Nicht nur für Verkehrsnachfragemodelle gilt, dass die Qualität der erzeugten Modellergebnisse direkt von der Güte und der Genauigkeit der genutzten Eingangsdaten abhängt. Zu den wichtigsten Strukturdaten der Verkehrsmodellierung zählen hierbei Bevölkerungsdaten, stehen sie doch in direktem Zusammenhang mit dem Volumen aber auch der Art und der räumlichen Verteilung der generierten Verkehrsnachfrage (vgl. u.a. Friedrich 2011 [1]). Attributumfang und räumlicher Detaillierungsgrad der notwendigen Eingangsdaten zur Beschreibung der Bevölkerung sind dabei in starkem Maße abhängig vom gewählten Modellansatz.

Makroskopische Nachfragemodelle arbeiten in der Regel mit einer vergleichsweise geringen Anzahl von Attributen zur Beschreibung der Bevölkerung im Untersuchungsgebiet, die sich vorrangig auf homogene Gruppen von Personen beziehen und auf einer gröberen

räumlichen Auflösung vorliegen müssen. Übliche Unterscheidungsmerkmale umfassen Alter, Geschlecht, Erwerbsstatus oder Lebensphase sowie Führerscheinbesitz und Pkw-Verfügbarkeit auf Personenebene (vgl. u.a. Schnabel und Lohse 2011 [2], Steierwald und Künne 1994 [3]). Seltener finden sich haushaltsbezogene Informationen wie das verfügbare Einkommen oder die Haushaltsgröße, nicht zuletzt weil hier vermehrt mit Problemen der Fortschreibung zu rechnen ist (vgl. Ortúzar und Willumsen 2006 [4]).

Mikroskopische Verkehrsnachfragemodelle bilden das individuelle Mobilitätsverhalten einzelner Personen ab und weisen in der Regel eine höhere räumliche Auflösung sowie eine detailliertere Beschreibung der einzelnen Individuen und ihres jeweiligen Haushaltskontextes auf. So benötigt beispielsweise das agentenbasierte Verkehrsnachfragemodell TAPAS (Heinrichs et al. 2016 [5]) je betrachtetem Untersuchungsgebiet und Zeitpunkt hausnummerngenau verortete und hoch differenzierte Bevölkerungsdaten als Eingabe. Diese umfassen für jede Person neben Angaben zur Soziodemographie (Alter, Geschlecht, Status) auch Informationen zum assoziierten Haushalt (Haushaltsgröße und -einkommen) sowie die Verfügbarkeit verschiedener Mobilitätsoptionen (Führerschein, Anzahl und Art der Pkw im Haushalt, ÖV-Zeitkarte, Mobilitätsbudget). Haushaltinformationen spielen innerhalb mikroskopischer Verkehrsnachfragemodelle auch eine wesentliche Rolle bei der Ziel- und Verkehrsmittelwahl.

Derartige Bevölkerungsdaten stehen für die Untersuchungsgebiete in der Regel nicht zur Verfügung, sondern müssen auf Grundlage empirischer Bevölkerungsdaten und durch geeignete mathematische Verfahren erzeugt werden. Die Zusammenführung aller Informationen erweist sich oftmals als schwierig, da vielfältige und heterogene Datenquellen benutzt werden müssen. Problematisch ist auch die Verfügbarkeit von Daten zur Berechnung zukünftiger Bevölkerungen. Diese stehen für Prognosejahre oftmals nicht zur Verfügung bzw. müssen manuell mittels Trendfortschreibung erzeugt werden. Bei einer zukünftigen Bevölkerung muss berücksichtigt werden, dass nicht nur die Personen älter werden und eventuell ihren Status ändern, sondern sich auch die Haushaltsgrößen sowie Haushaltszusammensetzungen verändern können.

Zur Erzeugung synthetischer Bevölkerungen hat sich in den letzten Jahrzehnten eine Vielzahl verschiedener Verfahren etabliert. Die meisten dieser Methoden sind stichprobenbasiert (siehe Beckman et al. 1996 [6], Guo und Bhat 2007 [7], Bar-Gera et al. 2009 [8], Ye et al. 2009 [9], Auld und Mohammadian 2010 [10], Müller und Axhausen 2011 [11], Pritchard und Miller 2012 [12]). Aufgrund eingeschränkter Datenverfügbarkeit wurden in letzter Zeit Verfahren entwickelt, die ohne eine Stichprobe auskommen, siehe Gargiulo et al. 2010 [13], Barthelemy und Toint 2013 [14] sowie Farooq et al. 2015 [15]. Bevölkerungen für Prognosejahre werden entweder mittels Fortschreibung einer bereits erzeugten synthetischen Bevölkerung (siehe Namazi-Rad et al. 2014 [16]) erstellt oder ebenfalls neu generiert.

Ziel dieses Artikels ist es, verschiedene Methoden zur Erstellung synthetischer Bevölkerungen vorzustellen und sie hinsichtlich ihrer Datenanforderungen, ihrer Ergebnisqualität und ihrer Einsetzbarkeit zu vergleichen. Das Hauptaugenmerk liegt dabei auf stichprobenbasierten Verfahren, bei denen drei verschiedene Methoden unterschieden werden können: A) Hochrechnung einer Stichprobe, B) Anpassung einer Stichprobe an Haushalts- oder Personenrandsummen sowie C) Anpassung einer Stichprobe an Haushalts- und Personenrandsummen.

Anhand der mit den unterschiedlichen Verfahren, aber gleicher Datenbasis generierten synthetischen Bevölkerungen für das Untersuchungsgebiet Berlin 2010 werden die Ergebnisse dieser Methoden mit realen Bevölkerungsdaten verglichen. Abschließend werden Empfehlungen bezüglich der Einsetzbarkeit der vorgestellten Methoden sowohl für makroskopische als auch mikroskopische Verkehrsnachfragemodelle gegeben.

2 Methoden zur Generierung synthetischer Bevölkerungen

Nachfolgend werden die Verfahren zur Erzeugung synthetischer Bevölkerungen vorgestellt. Zunächst erfolgt eine Erläuterung der häufiger verwendeten stichprobenbasierten Methoden; anschließend wird eine stichprobenfreie Alternative skizziert.

2.1 Hochrechnung einer Stichprobe (Methode A)

Eines der einfachsten Verfahren, jedoch in der Literatur zur Erzeugung von synthetischen Bevölkerungen bisher wenig beachtet, ist die Hochrechnung einer Bevölkerungsstichprobe. Solch eine Stichprobe repräsentiert einen Anteil der Gesamtbevölkerung und beinhaltet disaggregierte Haushalts- und Personendaten. In der Regel beinhaltet eine Stichprobe ein Gewicht bzw. einen Hochrechnungsfaktor, mit dessen Hilfe die Bevölkerung auf 100 % hochgerechnet werden kann. Bei der Generierung der synthetischen Bevölkerung wird jeder Haushalt und die dazugehörigen Personen so oft aus der Stichprobe kopiert, wie sie nach dem jeweiligen Hochrechnungsfaktor vorkommen (vgl. Moeckel 2016 [17]).

Abbildung 1 illustriert dieses Vorgehen. Die Farbkodierung in der Stichprobe sowie der generierten Bevölkerung symbolisiert Haushalte mit fünf unterschiedlichen Haushaltstypen. Ersichtlich ist hierbei auch, dass Personenattribute der auf Basis des Haushaltsgewichtes erzeugten Bevölkerung nicht separat kontrolliert werden; sie entsprechen stets denen der Stichprobe. Die Darstellung der Personen weist daher ausschließlich eine Farbkodierung als Merkmal der Zugehörigkeit zu einem Haushalt auf.

Für die Qualität der erzeugten Bevölkerung ist somit die Berechnungsbasis des vorliegenden Hochrechnungsfaktors von substantieller Bedeutung. Idealerweise stimmen die hier verwendeten Kriterien stark mit den im Model verwendeten Bevölkerungsattributen überein.

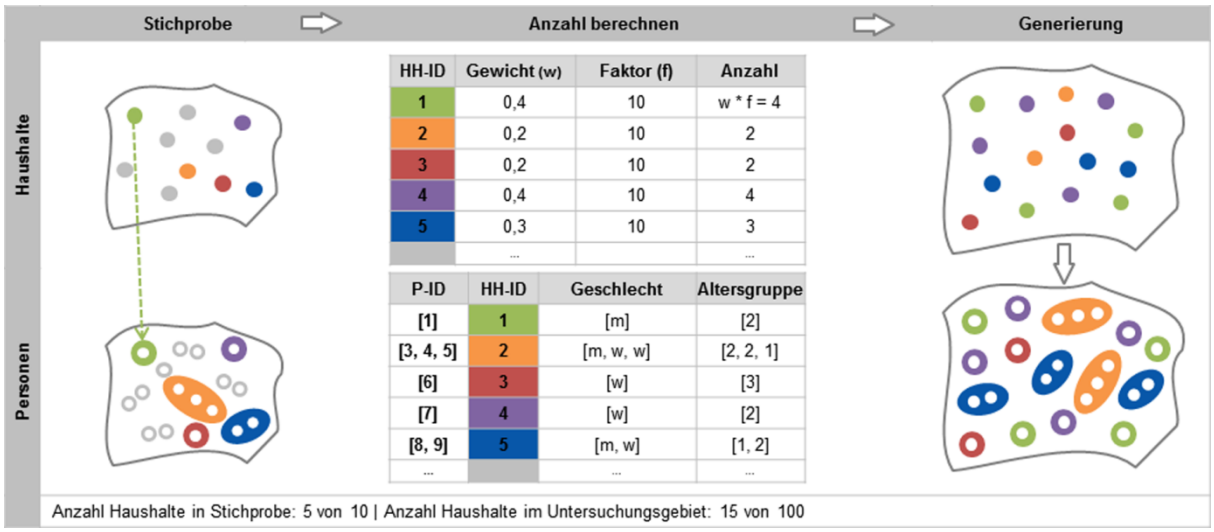

Abbildung 1: Hochrechnung einer Stichprobe

2.2 Anpassung einer Stichprobe an Haushalts- oder Personenrandsummen (Methode B)

Im Gegensatz zur Hochrechnung einer Stichprobe kann diese auch in Verbindung mit empirischen Randsummen verwendet werden. Einen allgemeinen Überblick dieser auf Beckman et al. 1996 [6] zurückgehenden Vorgehensweise findet sich in Bowman 2008 [18], eine Zusammenfassung der zahlreichen Verbesserungs- und Erweiterungsmöglichkeiten in Müller und Axhausen 2011 [19]. Ziel dieses Ansatzes ist es, dass die zu erstellende synthetische Bevölkerung einer gewünschten Haushalts- oder Personenverteilung entspricht.

Zunächst werden die in der Stichprobe enthaltenen Haushalte bzw. Personen entsprechend der modellseitig erforderlichen Attribute und deren Ausprägungen gruppiert und somit entsprechende Haushalts- bzw. Personentypen erzeugt. Anschließend können die so erzeugten Häufigkeiten je Typ aus der Stichprobe als initiale Startwerte für den Iterativ Proportional Fitting (IPF) Algorithmus verwendet werden. Der IPF-Algorithmus, auch bekannt als Biproportional Fitting oder Matrix Scaling, geht zurück auf Deming und Stephan 1940 [20]. Das Verfahren dient der Anpassung der initialen Zellwerte an die gegebenen Zeilen- sowie Spaltenrandsummen bei zwei- bzw. mehrdimensionalen Tabellen.

Eine synthetische Bevölkerung kann sowohl von einer Haushalts- als auch von einer Personenverteilung ausgehend erzeugt werden. Der textlichen Einfachheit halber wird das Verfahren nachfolgend am Beispiel einer gewünschten Einhaltung der Haushaltsrandsummen erläutert.

Für jeden Stichprobenhaushalt wird zunächst eine Auswahlwahrscheinlichkeit ermittelt, welche die Wahrscheinlichkeit seines Vorkommens in der gewünschten synthetischen Bevölkerung widerspiegelt. Die Auswahlwahrscheinlichkeit entspricht dem Gewicht eines Haushaltes aus der Stichprobe geteilt durch die Summe aller Haushaltsgewichte des jeweiligen Haushaltstyps (vgl. Beckman et al. 1996 [6]).

Bei der Generierung nach Haushaltsverteilung werden per Zufall für jeden Haushaltstyp, unter Berücksichtigung der Auswahlwahrscheinlichkeit, die entsprechenden Haushalte sowie die jeweils zugehörigen Personen vervielfältig. Dieser Vorgang wird solange wiederholt bis die Anzahl der Haushalte für jeden Haushaltstyp mit der jeweils gewünschten Menge der zu erzeugenden Haushalte übereinstimmt.

Abbildung 2 skizziert das Verfahren hinsichtlich der Einhaltung der Haushaltsrandsummen (HH-Größe x HH-Einkommen) und des Haushaltstyps „Zweipersonenhaushalt / mittleres Einkommen“. Auf Haushaltsebene werden insgesamt 100 Haushalte erzeugt, von denen 30 Prozent auf Zweipersonenhaushalte sowie 40 Prozent auf die mittlere Einkommenskategorie entfallen. Das blaue Farbschema symbolisiert Haushalte des zu betrachtenden Haushaltstyps. Es unterscheidet sich jedoch die jeweilige Haushaltszusammensetzung. Die Stichprobe enthält fünf solcher Haushalte. Per IPF wird beispielsweise ermittelt, dass insgesamt 15 der 100 Haushalte diesem Typ entsprechen. Unter Einhaltung dieser Gesamtzahl und der Berücksichtigung der Auswahlwahrscheinlichkeiten werden die Haushalte und die zugehörigen Personen vervielfältigt. Die Darstellung der Personen in den erzeugten Haushalten verdeutlicht, dass Personenattribute keine Berücksichtigung finden und auch in diesem Beispiel Alters- und Geschlechtsverteilung ein Zufallsprodukt der Stichprobenstruktur sind.

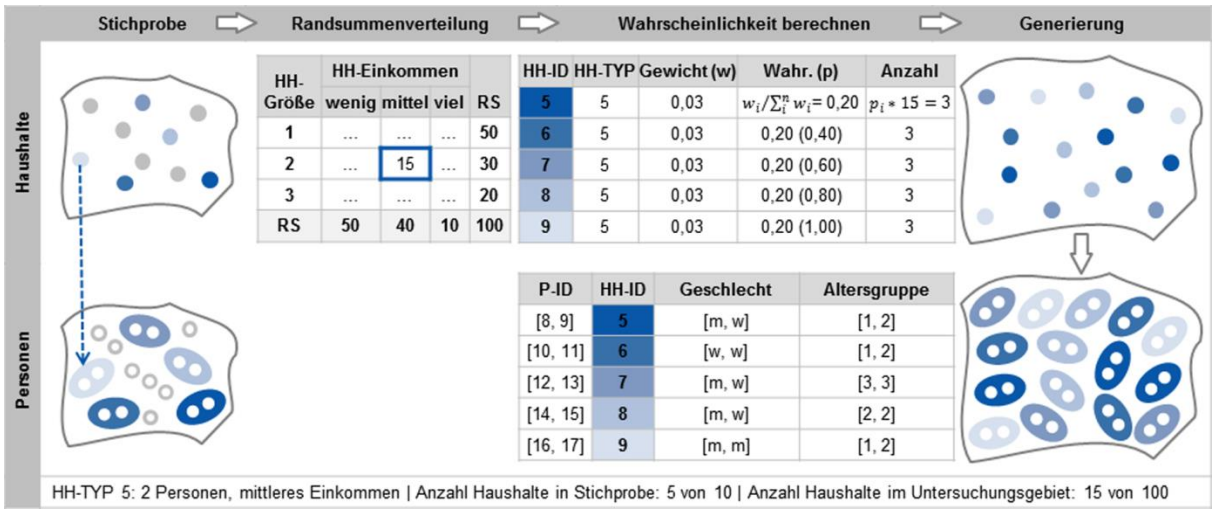

Abbildung 2: Anpassung einer Stichprobe an Haushaltsrandsummen

2.3 Anpassung einer Stichprobe an Haushalts- und Personenrandsummen (Methode C)

Bei diesem Ansatz wird genau wie bei der Methode B eine Bevölkerungsstichprobe in Verbindung mit empirischen Randsummen verwendet. Ziel dieser Erweiterung der Methode B ist jedoch, dass die zu erstellende synthetische Bevölkerung sowohl einer gewünschten Haushalts- als auch einer entsprechenden Personenverteilung entspricht.

Um den modellseitig erforderlichen Attributen und deren Ausprägungen auf Haushalts- sowie Personenebene zu entsprechen wird ein mehrstufiges Verfahren angewandt. Die Anpassung kann entweder separat mittels IPF oder aufeinander aufbauend mit Hilfe des so genannten Hierarchical IPF (HIPF) Algorithmus (Müller und Axhausen 2011 [11]) erfolgen. Die erzeugten Verteilung werden anschließend verwendet, um entweder das Haushaltsgewicht inklusive der Auswahlwahrscheinlichkeit neu zu berechnen oder nur die Wahrscheinlichkeit neu zu bestimmen. Zur Neuberechnung des Haushaltsgewichtes gibt es verschiedene Verfahren, von denen das von Ye et al. 2009 [9] entwickelte Iterative Proportional Updating (IPU) das bekannteste sein dürfte. Neben den Haushaltseigenschaften fließen hier auch explizit die Charakteristika der einzelnen Haushaltsmitglieder ein. Dafür werden die zuvor mittels IPF-Algorithmus berechneten Haushalts- und Personenverteilungen verwendet. Im Gegensatz dazu beschreiben Guo und Bhat 2007 [7] sowie Auld und Mohammadian 2010 [10] alternative Verfahren zur Neubestimmung der Auswahlwahrscheinlichkeit eines Haushaltes basierend auf dem Haushaltsgewicht aus der Stichprobe.

Da bei dieser Methode die Erzeugung der Bevölkerung wahrscheinlichkeitsbasiert erfolgt, wird der Vorgang der Generierung solange wiederholt bis die synthetische Bevölkerung der mittels IPF bestimmten Personenverteilung entspricht bzw. eine maximale Anzahl an Iterationen erreicht ist.

Die Abbildung 3 zeigt ähnlich wie in der Abbildung 2 die Erstellung von 15 Zweipersonenhaushalten mit mittlerem Einkommen. Neben den Haushaltsrandsummen (HH-Größe x HH-Einkommen) werden hier auch Randsummen auf Personenebene (Alter x Geschlecht) berücksichtigt. In den 15 Haushalten leben 30 Personen. Mittels IPF wird beispielsweise folgende Personenverteilung ermittelt: 2 Mädchen, 4 Jungs, 10 Frauen, 6

Männer, 4 Rentnerinnen sowie 4 Rentner. Grün dargestellt sind die sich ändernden Haushaltsgewichte und -wahrscheinlichkeiten. Die Replikationsanzahl unterscheidet sich nun für jeden Ausgangshaushalt im Gegensatz zu Abbildung 2 und resultiert in dem heterogenen Bild der erzeugten Haushalte und Personen. Die Darstellung der Personen symbolisiert, dass hierbei die Haushaltszusammensetzung berücksichtigt wird.

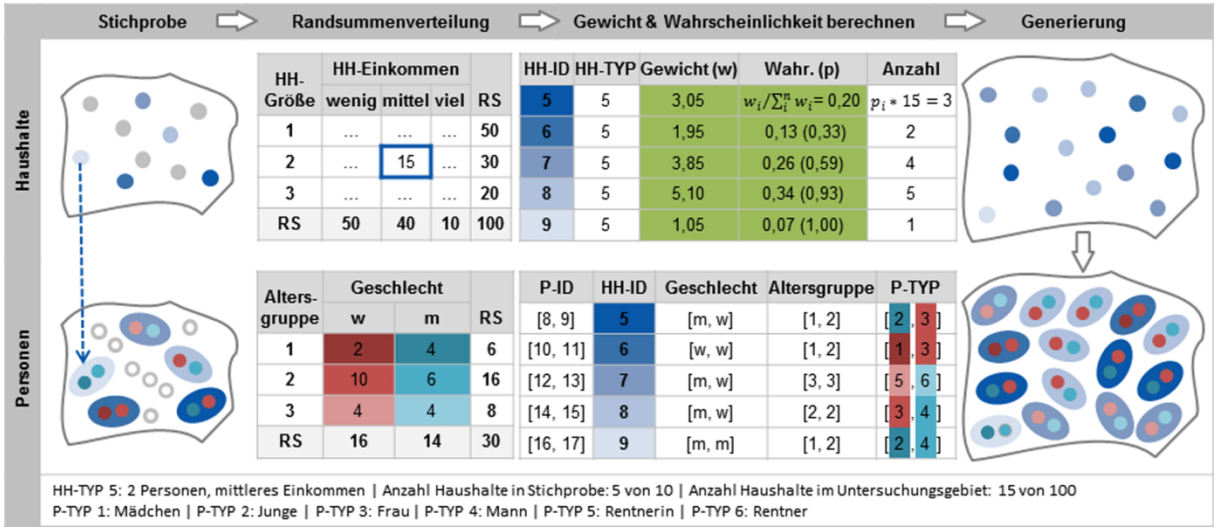

Abbildung 3: Anpassung einer Stichprobe an Haushalts- und Personenrandsummen

Damit eine zu generierende synthetische Bevölkerung sowohl einer gewünschten Haushalts- und Personenverteilung entspricht könnten auch alternative Verfahren wie beispielsweise Bayesian Networks (siehe Sun und Erath 2015 [21]) statt IPF und den darauf aufbauenden Algorithmen verwendet werden.

2.4 Stichprobenfreie Alternative

Nicht immer steht für ein Untersuchungsgebiet eine Bevölkerungsstichprobe zur Verfügung. Als Beispiel eines alternativen Verfahrens sei hier der dreistufige Ansatz von Barthelemy und Toint 2013 [14] skizziert, der ausschließlich auf der Verwendung empirischer Randsummenverteilungen beruht. Anhand bekannter bzw. prognostizierter Personenverteilungen (z.B. Alter x Geschlecht) wird zunächst ein Pool vollkommen synthetischer Personen erstellt. In den nächsten Schritten wird anhand verschiedener mathematischer Verfahren die Haushaltsverteilung berechnet und anschließend die Haushalte erzeugt. Bei der Haushaltserzeugung werden zuerst der Haushaltsvorstand und anschließend mögliche weitere Haushaltsmitglieder aus dem Personenpool den Haushalten zugewiesen. Mit dieser Vorgehensweise haben Barthelemy und Toint 2013 [14] eine synthetische Bevölkerung für Belgien mit stark reduzierten Attributausprägungen erstellt.

3 Empirische Datengrundlagen und Generierung synthetischer Bevölkerungen

Nachdem im letzten Abschnitt die gängigen Verfahren der Erzeugung synthetischer Bevölkerungen vorgestellt wurden, sollen nun die drei stichprobenbasierten Methoden zunächst hinsichtlich ihrer allgemeinen Datenanforderungen verglichen werden.

Anschließend wird auf die speziell für das Untersuchungsgebiet Berlin 2010 verwendeten Datenquellen eingegangen. Schließlich wird das zur Generierung der synthetischen Bevölkerungen verwendete Programm „SYNTHESIZER“ vorgestellt.

3.1 Datenanforderungen bei stichprobenbasierten Methoden

Bei den drei beschriebenen Methoden wird als Datenbasis eine Bevölkerungsstichprobe benötigt. Häufig werden hierfür Zensusdaten verwendet. In Deutschland gibt es den Zensus [22] und den Mikrozensus [23]. Neben diesen kommen auch Erhebungen zum Verkehrsverhalten wie beispielsweise „Mobilität in Deutschland“ sowie „Mobilität in Städten“ (siehe MiD [24] und SrV [25]) oder spezifische Erhebungen für den konkreten Untersuchungsraum in Frage.

Die Methode A „Hochrechnung einer Stichprobe“ hat den Vorteil, dass für die Erzeugung einer synthetischen Bevölkerung lediglich eine Bevölkerungsstichprobe benötigt wird. Die Bevölkerung lässt sich sehr schnell und relativ einfach erstellen. Zu berücksichtigen ist hierbei, dass der Hochrechnungsfaktor bzw. das Gewicht ausschlaggebend für die Güte der erstellten synthetischen Bevölkerung ist. Beispielsweise beinhaltet der Hochrechnungsfaktor im Mikrozensus drei Altersklassen und vier Staatsangehörigkeitsgruppen, jeweils differenziert nach dem Geschlecht (siehe Qualitätsbericht zum Mikrozensus 2010 [26]). Dies entspricht einer Personenverteilung, da bei dem Hochrechnungsfaktor nur personenbezogene Daten verwendet werden. Da die Gewichtungsfaktoren nur für den Erhebungszeitpunkt Gültigkeit haben und keine weiteren Informationen zur Generierung herangezogen werden, verlangt die Methode A im Falle einer gewünschten Bevölkerungsfortschreibung den Einsatz zusätzlicher Verfahren.

Für die Methoden B und C werden zusätzlich modellrelevante Randsummen benötigt. Diese stehen nicht immer in der benötigten Art und räumlich differenzierten Auflösung zur Verfügung. Bei der zu erzielenden Haushalt- bzw. Personenverteilung ist zu berücksichtigen, dass die Wahrscheinlichkeit fehlender Daten in einer Bevölkerungsstichprobe steigt je mehr Attributausprägungen verwendet werden. Zudem müssen die oftmals aus heterogenen Datenquellen stammenden Randsummen harmonisiert werden. Hierzu können folgende Harmonisierungsschritte zählen:

- Umwandlung von Attributausprägungen (z.B. Zusammenfassen bzw. Aufteilen von Altersgruppen)
- räumliche Randsummenanpassung (z.B. ältere Verteilung auf Teilverkehrszellenebene an neue Daten auf Bezirksebene anpassen)
- Randsummenanpassung innerhalb einer Ebene (z.B. Gesamtanzahl der Haushalte nach HH-Einkommen anteilmäßig an die Gesamtanzahl der Haushalte nach HH-Größe anpassen)
- Randsummenanpassung zwischen den Ebenen (z.B. Gesamtanzahl der Haushalte auf HH-Größe anteilmäßig an die Gesamtanzahl der Personen anpassen)

Beide Methoden können jedoch sowohl für das Basisjahr als auch für mögliche Prognosejahre verwendet werden. Problematisch ist hierbei, dass Randsummen für Folgejahre zumeist nur für eine Auswahl der benötigten Attribute bzw. ihrer Ausprägung vorliegen.

Oftmals stehen die Daten einer Bevölkerungsstichprobe disaggregiert nur auf einer groben räumlichen Auflösung zu Verfügung. Somit können Unterschiede innerhalb einzelner Gebiete nicht ausreichend abgebildet werden. Da mikroskopische Verkehrsnachfragemodelle meist hausnummerngenau differenzierte Bevölkerungsdaten benötigen, müssen die erzeugten Haushalte anschließend feiner räumlich verteilt werden. Für die Verteilung der Haushalte auf Adresskoordinaten können beispielsweise Flächennutzungstypen sowie Gebäudedaten verwendet werden.

Je nach verwendetem Verkehrsnachfragemodell werden zusätzlich zu den Grundinformationen (Alter, Geschlecht, HH-Größe) auch Mobilitätsoptionen (Führerschein, ÖPNV-Karte, Pkw, Mobilitätsbudget) benötigt. Diese Informationen stehen nicht unbedingt in einer Bevölkerungsstichprobe zur Verfügung, sondern müssen durch zusätzliche Daten mithilfe geeigneter Modelle an eine synthetische Grundbevölkerung hinzugefügt werden.

3.2 Datenquellen für das Untersuchungsgebiet Berlin

In der Tabelle 1 sind zu den erwähnten Datenquellen Mikrozensus, MiD sowie SrV der jeweilige Stichprobenumfang sowie die räumliche Auflösung für das Untersuchungsgebiet Berlin zu entnehmen. Der Zensus wird als Bevölkerungsstichprobe hier nicht weiter betrachtet, da die anonymisiert disaggregierten Daten lediglich für Analysezwecke auf Antrag in den jeweiligen statistischen Landesämtern zur Verfügung stehen.

	Mikrozensus 2010	MiD 2008	SrV 2008
Haushalte	10.915	1.300	19.354
Personen	19.135	2.582	41.050
Stichprobenumfang (%)	0,6	0,08	1,23
Räumliche Auflösung	Berlin gesamt	Berlin gesamt	Bezirke

Tabelle 1: Datenquellen für Berlin 2010 - Bevölkerungsstichprobe

Sowohl der Mikrozensus 2010 als auch die Daten der SrV 2008 können als Stichprobe eingesetzt werden. Die Daten der MiD 2008 sind bezüglich des geringen Stichprobenumfangs dafür nicht geeignet. Die SrV-Daten bieten einen größeren Stichprobenumfang und eine höhere räumliche Auflösung als der Mikrozensus. Sie wären deshalb prinzipiell besser geeignet und einer bundeslandweiten Stichprobe vorzuziehen. Die relevanten SrV-Daten stammen jedoch aus dem Jahr 2008 und nicht aus 2010. Damit jedoch alle drei stichprobenbasierten Methoden verglichen werden können, ist eine einheitliche Datenbasis notwendig. Aus diesem Grund wird für alle drei Methoden der disaggregierte Mikrozensus 2010 als Stichprobe verwendet. In diesem sind für Berlin 10.915 Haushalte sowie 19.135 Personen enthalten, welche etwa 0,6 % der Berliner Gesamtbevölkerung entsprechen.

Für die Methoden B und C werden zusätzlich zur Stichprobe auch aggregierte Randsummen benötigt. Diese müssen nicht zwangsläufig von der gleichen Bezugsquelle wie der Stichprobe stammen. Im vorliegenden Fall wurden sie jedoch ebenfalls dem Mikrozensus entnommen. Aus diesem geht hervor, dass 2010 in Berlin 3.444.800 Personen in 1.988.600 Haushalten lebten. Die Berliner Bevölkerung verteilt sich auf 1193 Teilverkehrszellen, welche wiederum 12 Bezirken zugeordnet sind. Statistische Randsummen des Mikrozensus stehen

für Berlin auf Bezirksebene zur Verfügung. In der Tabelle 2 sind die jeweiligen Attribute mit ihren Ausprägungen für die Haushalts- als auch Personenebene enthalten, welche als Randsummen verwendet werden.

Die erzeugten Bevölkerungen werden in Abschnitt 4 hinsichtlich der Haushaltsverteilung „HH-Größe x HH-Einkommen“ sowie der Personenverteilung „Alter x Geschlecht x Status“ miteinander verglichen.

Attribut	Attributausprägung
HH-Größe	1, 2, 3, 4 und 5+ Personenhaushalte
HH-Einkommen (Euro)	0 - 899, 900 - 1499, 1500 - 1999, 2000 - 2599, 2600 - 3199, 3200+
Geschlecht	männlich, weiblich
Alter	0 - 5, 6 - 9, 10 - 14, 15 - 17, 18 - 19, 20 - 24, 25 - 29, 30 - 34, 35 - 39, 40 - 44, 45 - 49, 50 - 54, 55 - 59, 60 - 64, 65 - 74, 75 - 79, 80 - 84, 85+
Status	Nichterwerbspersonen (NEP) Kind unter 6, Schüler, Student, Rentner, sonstige NEP Erwerbspersonen (EP) Vollzeit, Teilzeit, Erwerbslos

Tabelle 2: Attribute und deren Ausprägung

3.3 SYNTHESIZER: Programm zur Bevölkerungsgenerierung

Das agentenbasierte Verkehrsnachfragemodell TAPAS (Heinrichs et al. 2016 [5]) benötigt je betrachtetem Untersuchungsgebiet und Zeitpunkt hausnummerngenau differenzierte Bevölkerungsdaten als Eingabe. Für die Erstellung dieser synthetischen Bevölkerungen wurde am DLR-Institut für Verkehrsforschung das Programm SYNTHESIZER entwickelt.

Der SYNTHESIZER ist in der Programmiersprache Java implementiert und modular aufgebaut. Über eine entsprechende Benutzeroberfläche können synthetische Bevölkerungen erstellt und verwaltet werden. Die Daten werden in einer PostgreSQL/Postgis Datenbank gespeichert. Mit dem SYNTHESIZER können, für jede vorgestellte stichprobenbasierte Methode, synthetische Bevölkerungen für beliebige Bezugsräume erstellt werden. In der Abbildung 4 ist der Programmablauf, inklusive der für jede Methode benötigten Module, dargestellt.

Über entsprechende Parameter wird das Untersuchungsgebiet mit seiner räumlichen Auflösung und Abdeckung definiert. Die Methode A benötigt zur Vervielfältigung der Haushalte lediglich das Haushaltsgewicht bzw. den Hochrechnungsfaktor der Stichprobe. Die beiden anderen Methoden benötigen, wie in Abschnitt 2 beschrieben, zusätzliche Berechnungsschritte.

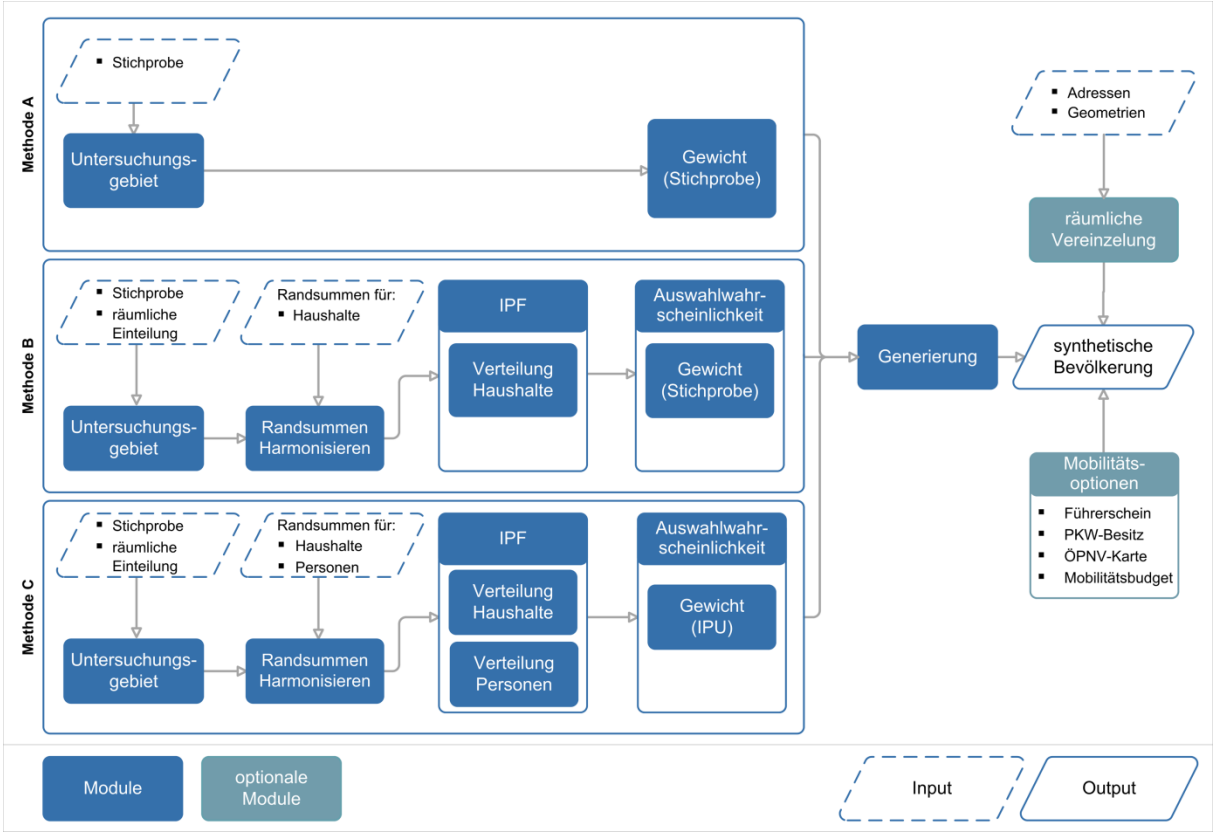

Abbildung 4: SYNTHESIZER - Programmablauf für stichprobenbasierte Methoden

Für die Methoden B und C wird die in Abschnitt 3.2 definierte Haushaltsverteilung zunächst formal beschrieben, siehe Tabellen 3 und 4. Die Festlegung der internen Referenz-randsumme ist notwendig, wenn beispielsweise die Randsummen innerhalb der Haushaltsebene in ihrer Gesamtheit nicht übereinstimmen. Dies ist möglich da die Randsummen oftmals aus verschieden Datenquellen stammen. Eine Übereinstimmung ist jedoch eine Grundvoraussetzung für den IPF-Algorithmus. Wie der Tabelle 3 zu entnehmen ist, werden daher im Beispielfall die Randsummen für HH-Einkommen anteilmäßig an die Randsummen nach HH-Größe angepasst. Ähnlich verhält es sich mit der externen Referenzrandsumme, wobei definiert wird, welche Randsumme verwendet werden soll, um die Randsummen zwischen der Haushalts- und Personenebene anzupassen. Diese Anpassung kann innerhalb der Methode C optional durchgeführt werden. Die in Abschnitt 3.2 erwähnten notwendigen Schritte zur Harmonisierung der Randsummen sind mit Ausnahme der Umwandlung von Attributausprägungen im SYNTHESIZER implementiert. Diese wird im Vorhinein bei der Datenaufbereitung manuell durchgeführt, da hier eine Formalisierung schwer umsetzbar ist.

Attribut-ID	Attribut	Referenzrandsumme (intern)	Referenzrandsumme (extern)
1	**HH-Größe**	TRUE	TRUE
2	**HH-Einkommen**	FALSE	FALSE

Tabelle 3: Haushaltsebene - Attribute

Attribut-ID	Attributausprägung-ID	Attributausprägung
1	**1**	1 Personenhaushalt
1	**…**	…
1	**5**	5+ Personenhaushalt
2	**1**	0 - 899 €
2	**…**	…
2	**6**	3200+ €

Tabelle 4: Haushaltsebene - Attributausprägungen

Basierend auf den Eintragungen in den Tabellen 3 und 4 werden die Haushaltstypen im SYNTHESIZER automatisch erzeugt (siehe Tabelle 5).

	Attributausprägung-ID	
HH-Typ-ID	**HH-Größe**	**HH-Einkommen**
1	1	1
…	**…**	…
30	5	6

Tabelle 5: Haushaltstypen (HH-Größe x HH-Einkommen)

Bei der Methode C wird zusätzlich zur Haushaltsverteilung auch die in Abschnitt 3.2 festgelegte Personenverteilung benötigt. Dafür wird die Personenebene analog zur Haushaltsebene definiert um die entsprechenden Personentypen automatisch zu erzeugen. Durch eine solche parametrisierte Festlegung der Zielverteilungen ist der SYNTHESIZER flexibel einsetzbar.

Die Randsummenverteilungen und die Bestimmung der Auswahlwahrscheinlichkeit eines Stichprobenhaushalts für die Methode C werden im SYNTHESIZER mittels IPF- respektive IPU-Algorithmus umgesetzt.

Bei allen drei Methoden wird zunächst eine synthetische Grundbevölkerung erstellt. Bei Methode A resultiert die Bevölkerung auf räumlicher Auflösung der zugrundeliegenden Stichprobe. Im Falle des Mikrozensus wäre das flächendeckend für ein gesamtes Bundesland. Das Untersuchungsgebiet selber wird nicht weiter differenziert. Bei den Methoden B und C resultieren die Bevölkerungen räumlich der Ebene, auf welcher die Randsummen vorliegen, beispielsweise der Bezirksebene. Makroskopischen Verkehrs-nachfragemodellen genügt eine solche gröbere räumliche Auflösung. Mikroskopische Verkehrsnachfragemodelle benötigten jedoch räumlich hochaufgelöst differenzierte Bevölkerungsdaten. Da diese Daten auf so hoher räumlicher Ebene nicht zur Verfügung stehen, ist im SYNTHESIZER ein Algorithmus zum räumlichen Verteilen der Haushalte auf Adresskoordinaten integriert.

Eine so erzeugte synthetische Bevölkerung könnte auch von anderen mikro- bzw. makroskopischen Modellen verwendet werden. Verkehrsnachfragemodelle benötigen zusätzlich zu der Grundbevölkerung auch Informationen über die Verfügbarkeit verschiedener Mobilitätsoptionen wie beispielsweise der Besitz eines Führerscheins oder

einer ÖPNV-Zeitkarte. Da einige Mobilitätsoptionen einander bedingen, ist beim Hinzufügen auf eine Reihenfolge zu achten. So ist beispielsweise die Wahrscheinlichkeit des Autobesitzes im Haushalt unter anderem von der Anzahl der Führerscheine im Haushalt abhängig. Für das Verkehrsnachfragemodell TAPAS sind im SYNTHESIZER die Mobilitätsoptionen Führerschein, Anzahl und Art der Pkw im Haushalt, ÖPNV-Zeitkarte sowie das Mobilitätsbudget je Person enthalten. Die modulare Implementierung des SYNTHESIZERS lässt flexible Erweiterungen zu. So können beispielsweise für die Abbildung neuer Mobilitätstrends, z.B. CarSharing, entsprechend neue Module hinzugefügt werden.

4 Vergleich und Bewertung der synthetischen Bevölkerungen

Für die in Abschnitt 2 besprochenen stichprobenbasierten Methoden werden nachfolgend die resultierenden synthetischen Bevölkerungen des Jahres 2010 für das Untersuchungsgebiet Berlin untereinander sowie mit realen Bevölkerungsdaten des Mikrozensus 2010 verglichen.

Tabelle 6 zeigt die Summen der jeweils erzeugten Haushalte sowie Personen und gibt die jeweilige Abweichung zu den realen Werten an. Ersichtlich ist hierbei, dass die jeweilige Gesamtanzahl bei allen drei Methoden absolut und relativ kaum abweicht, teilweise den Zielvorgaben sogar exakt entspricht.

	Quelle	**Methode**		
	Mikrozensus 2010	**A**	**B**	**C**
Haushalte	1.988.600	1.993.730	1.988.600	1.988.600
Differenz (%)		0,26	0,00	0,00
Personen	3.444.800	3.470.312	3.445.707	3.443.885
Differenz (%)		0,74	0,026	-0,03

Tabelle 6: Synthetische Bevölkerungen für Berlin 2010

Zur Bestimmung der Qualität von synthetischen Bevölkerungen wird oftmals der Standardized Root Mean Square Error (SRMSE) verwendet. Er ermöglicht die Bestimmung der Unterschiede zwischen zwei Verteilungen. Bei einem SRMSE Wert von 0 stimmen die erzeugten Bevölkerungen exakt mit der Vorgabe überein. Als Maximalwert gilt in der Regel 1 (siehe Pritchard und Miller 2012 [12]). Für alle drei Methoden wurden die Haushalts- und Personenverteilungen für Berlin mit empirischen Werten verglichen. Wie in der Tabelle 7 ersichtlich liefert die Methode C mit einem SRMSE von 0,0 bzw. 0,03 die besten Ergebnisse. Insbesondere bei den Verteilungen der personenbezogenen Attribute zeigen sich im Falle der Methoden A und B mit einem SRMSE-Wert von 0,41 deutliche Abweichungen.

	Methode		
	A	**B**	**C**
Haushaltsverteilung (HH-Größe x HH-Einkommen)			
SRMSE	0,16	0,00	0,00
Personenverteilung (Alter x Geschlecht x Status)			
SRMSE	0,41	0,41	0,03

Tabelle 7: SRMSE für Haushalts- und Personenverteilung

Die unterschiedliche Genauigkeit der Methoden wird auch in der Abbildung 5 deutlich. Hierfür wurden wieder die Haushalts- und Personenverteilungen betrachtet, diesmal jedoch auf Bezirksebene. Für die dargestellten Boxplots wurden die absoluten relativen Abweichungen zwischen den realen und erzeugten Haushaltsverteilungen (30 Ausprägungen je Bezirk) bzw. Personenverteilungen (288 Ausprägungen je Bezirk) ermittelt. Deutlich erkennbar weist ein substantieller Anteil der erzeugten Verteilungen eine deutliche Abweichung von den realen Werten auf: So liegt der Median der Abweichungen der mit Methode A generierten Haushaltsverteilungen bei rund 24 %. Ähnliche Werte gelten auch für die Personenverteilungen bei den Methoden A und B. Gleichzeitig fällt, insbesondere bei der Personenverteilung für Methode C, die deutlich geringere mittlere Abweichung auf. Das in der Abbildung gezeigte genaue Treffen der Haushaltsverteilung in den Methoden B und C ist auf die explizite Anpassung an die Haushaltsrandsummen zurückzuführen. Bei der Personenverteilung zeigt sich, dass durch die zusätzliche Anpassung an Personenrandsummen innerhalb der Methode C auch eine Verbesserung dieser im Vergleich zur Methode B erreicht werden kann.

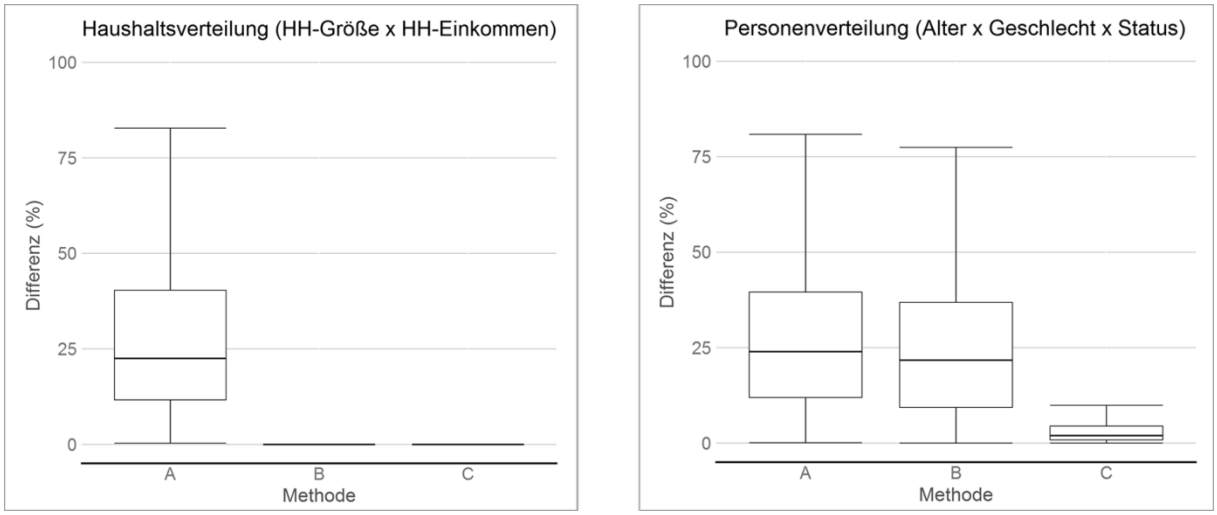

Abbildung 5: Boxplots der Abweichungen synthetischer Bevölkerungsverteilungen

Um die Unterschiede der erzeugten Bevölkerungen zu verdeutlichen werden nachfolgend Auswertungen sowohl auf Personen- als auch Haushaltsebene betrachtet.

In der Abbildung 6 ist auf der linken Seite eine Auswertung nach Altersgruppen und Geschlecht in Form einer Bevölkerungspyramide dargestellt. Sie zeigt, dass alle Methoden gute Ergebnisse liefern, wobei die mittels Methode C erzeugte Bevölkerung am besten mit

den Referenzwerten übereinstimmt. Die Auswertung nach Personenstatus, rechts in der Abbildung, zeigt bei den Kategorien „Schüler“, „Student“, „sonstige Nichterwerbspersonen“ sowie „Teilzeit“ zum Teil starke Unterschiede bei den Methoden A und B. Methode C liefert im Gegensatz dazu wieder ein sehr gutes Ergebnis. Dies ist zurückzuführen auf die explizite Anpassung an Personenrandsummen. Bezogen auf die Gesamtsumme der Nichterwerbspersonen sowie Erwerbspersonen gibt es bei der Methode A eine relative Abweichung von plus 8,93 % respektive minus 6,69 % zu den realen Werten. Bei der Methode B liegt die relative Abweichung der Nichterwerbspersonen bei plus 7,76 % und minus 6,99 % bei den Erwerbspersonen.

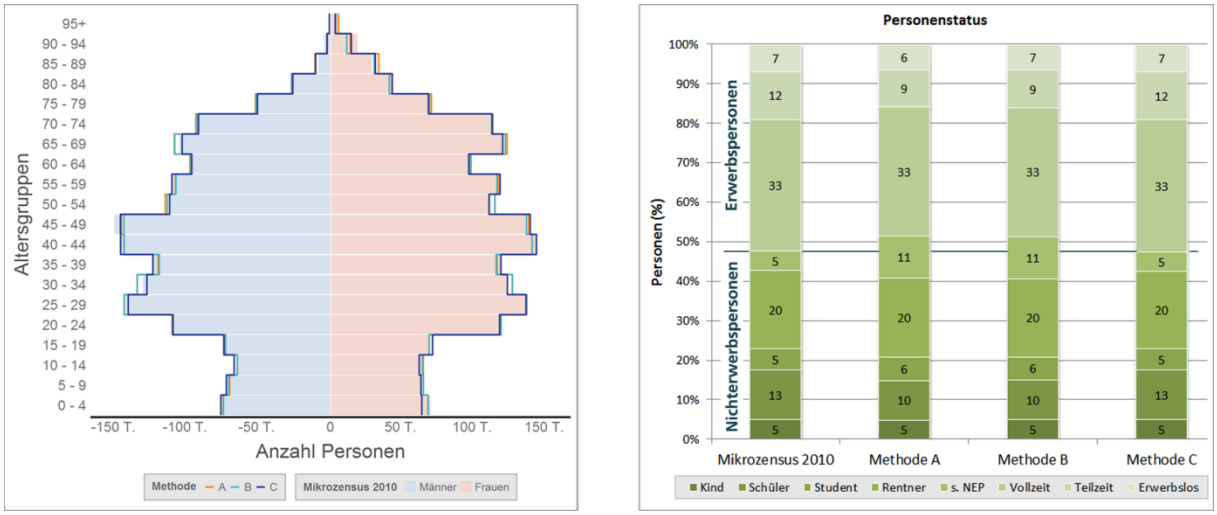

Abbildung 6: Bevölkerungspyramide sowie prozentuale Anteile für Personenstatus

In der Abbildung 7 ist auf der linken Seite ein Vergleich zwischen synthetischen und empirischen Werten bezüglich der Haushaltsgrößen dargestellt. Hierbei stimmen bei allen drei Methoden die Werte mit den Referenzzahlen überein. Beim Betrachten des Haushaltseinkommens (Abbildung 7 rechts) weichen die Werte der Methode A insbesondere in den höheren Einkommensklassen ab. Die Methode B und C liefert perfekte Werte. Dies ist wiederum auf die explizite Anpassung an Haushaltsrandsummen zurückzuführen.

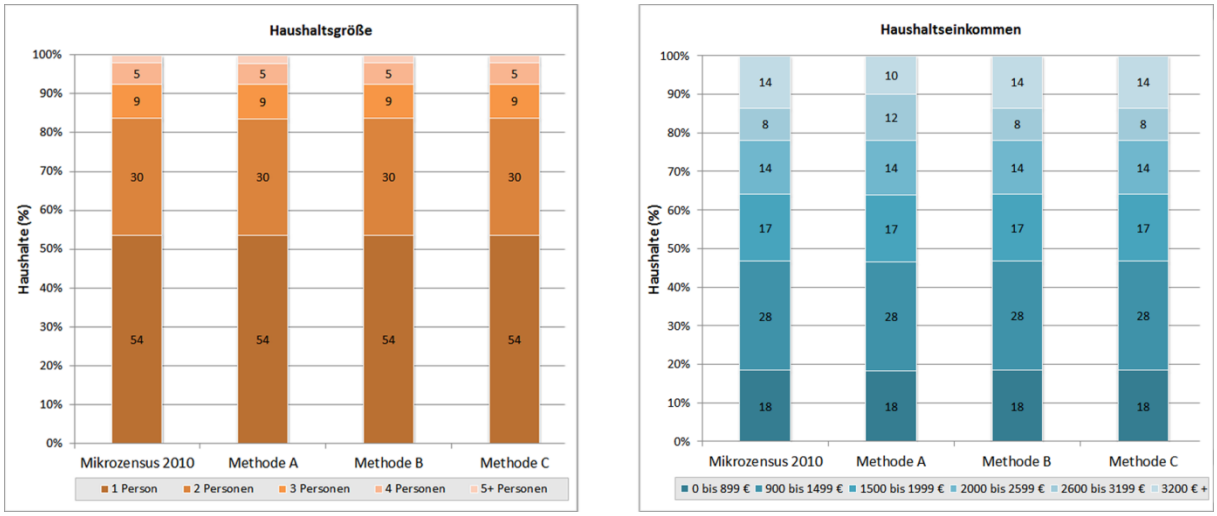

Abbildung 7: Auswertung nach Haushaltsgröße sowie Haushaltseinkommen

Die räumliche Verteilung der Differenz zwischen der jeweils erzeugten Bevölkerungsdichte und den realen Werten auf Ebene der Berliner Bezirke wird in Abbildung 8 dargestellt. Wie schon aus den vorhergehenden Darstellungen zu entnehmen, nimmt die Qualität der Ergebnisse entlang der vorgestellten Methoden A, B und C zu. Während bei Methode A bei einem Viertel der Bezirke Abweichungen von absolut 4 und mehr Prozent auftreten, ist dies bei Methode C in keinem der Bezirke der Fall.

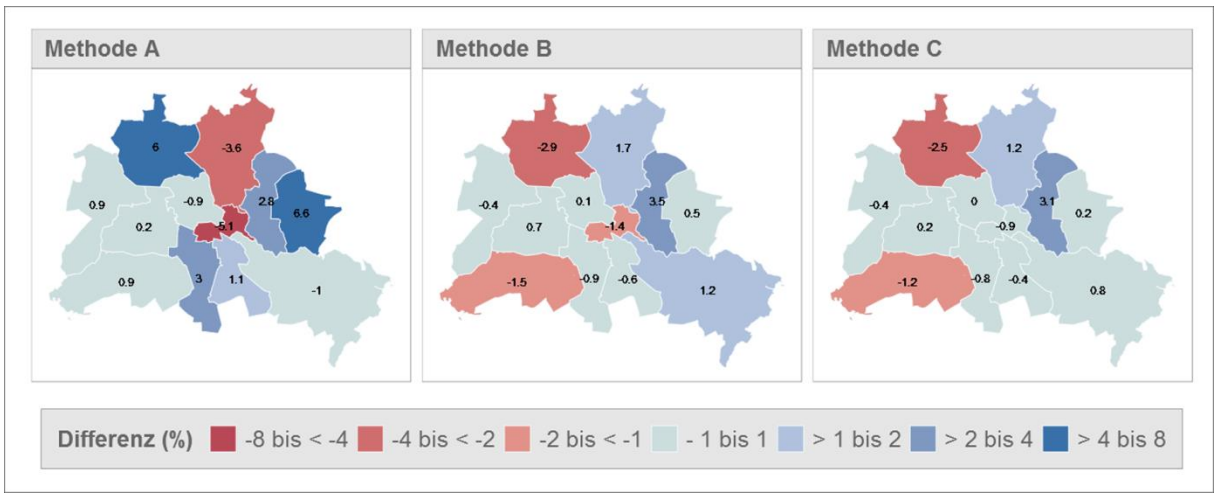

Abbildung 8: Prozentuale Differenz der Bevölkerungsdichte auf Bezirksebene

5 Fazit

Bevölkerungsdaten zählen zu den wichtigsten Strukturdaten der Verkehrsmodellierung. Der benötigte Attributumfang und räumliche Detaillierungsgrad sind dabei stark abhängig vom gewählten Modellansatz. In dem vorliegenden Artikel wurden drei stichprobenbasierte Methoden zur Erzeugung synthetischer Bevölkerungen vorgestellt: A) Hochrechnung einer Stichprobe, B) Anpassung einer Stichprobe entweder an Haushalts- oder an Personenrandsummen sowie C) Anpassung einer Stichprobe an Haushalts- und Personenrandsummen. Für das Untersuchungsgebiet Berlin 2010 wurden mit Hilfe dieser Methoden synthetische Bevölkerungen generiert und miteinander sowie mit realen Bevölkerungsdaten verglichen.

Alle drei beschriebenen stichprobenbasierten Methoden verwenden zur Generierung einer synthetischen Bevölkerung eine disaggregierte Bevölkerungsstichprobe. Die Methoden unterscheiden sich im Hinblick auf die Vervielfältigung der Haushalte sowie der dazugehörigen Personen, die teilweise zusätzlich benötigten Eingangsdaten und die generierten Ergebnisse. In der Tabelle 8 ist eine Übersicht zu den einzelnen Methoden enthalten.

Methode A sollte zur Berechnung von synthetischen Bevölkerungen für ein Basisjahr nur dann verwendet werden, wenn die zur Verfügung stehende Stichprobe aus dem zu erzeugenden Jahr stammt. In der Regel beziehen sich verfügbare Hochrechnungsfaktoren entweder auf Personen- oder Haushaltsattribute. Diese Methode ist insbesondere dann geeignet, wenn die bei der Berechnung des Hochrechnungsfaktors berücksichtigten Attribute stark mit den im Modell verwendeten Bevölkerungsattributen übereinstimmen. Idealer Weise sollte dabei die Übereinstimmung der erzeugten Bevölkerung mit nur einer Verteilung,

beispielsweise der Personenverteilung, angestrebt werden. Dies ist insbesondere bei makroskopischen Verkehrsmodellen der Fall, da hier der Haushaltskontext in der Regel geringe oder keine Beachtung findet. Insgesamt ist der Aufwand zur Erstellung einer synthetischen Bevölkerung für das Basisjahr mittels Methode A als sehr gering einzustufen. Zur Berechnung einer Bevölkerung für Prognosejahre müssen jedoch weitere Verfahren herangezogen werden, um den fehlenden Hochrechenfaktor zu kompensieren.

	Methode A		Methode B		Methode C	
Randsummen						
• **Haushaltsebene**	NEIN		JA	NEIN	JA	
• **Personenebene**	NEIN		NEIN	JA	JA	
• **Harmonisierung**	NEIN		JA		JA	
Mathematische Verfahren						
• **Hochrechnung**	JA		NEIN		NEIN	
• **Berechnung Zielverteilung (z.B. IPF)**	NEIN		JA		JA	
• **Berechnung Haushaltsgewicht (z.B. IPU)**	NEIN		NEIN		JA	NEIN
• **Berechnung Auswahlwahrscheinlichkeit**	NEIN		JA		JA	JA
Verteilung (synthetische Bevölkerung entspricht)						
• **Haushaltsverteilung**	JA	NEIN	JA	NEIN	JA	
• **Personenverteilung**	NEIN	JA	NEIN	JA	JA	
Zeithorizont						
• **Basisjahr**	JA		JA		JA	
• **Prognosejahr**	NEIN		JA		JA	

Tabelle 8: Übersicht der vorgestellten stichprobenbasierten Methoden und ihrer Eigenschaften (farbige Blöcke zeigen Kombinationen von Alternativen)

Für die Methoden B und C werden zusätzlich zur Stichprobe modellrelevante Randsummen benötigt. Oftmals müssen diese aufwändig harmonisiert werden, da sie nicht immer in der benötigten Art und räumlichen Auflösung zur Verfügung stehen (siehe Abschnitt 3.1.)

Prinzipiell könnten die von allen drei Methoden erzeugten disaggregierten Bevölkerungen in aggregierte Bevölkerungsgruppen überführt werden. Auch Methode B eignet sich, in Verbindung mit Randsummen auf Personenebene, für makroskopische Verkehrsnachfragemodelle. Im Vergleich zu Methode A bietet Methode B eine größere Flexibilität bei der Erstellung von Bevölkerungen mit gewünschten Haushalts- oder Personenverteilung. Zudem gewährleistet sie die Unabhängigkeit vom Vorliegen einer Stichprobe für das angestrebte Basisjahr. Eine Berechnung der Personenverteilung mittels IPF-Algorithmus sollte bei einer Nutzung in makroskopischen Modellen genügen, da hierfür lediglich aggregierte Personengruppen als Bevölkerungsinput benötigt werden.

Den größten Aufwand bei der Datenaufbereitung wie auch bei den nachfolgenden Berechnungen bringt Methode C mit sich. Wie die Auswertungen in Abschnitt 4 zeigen, ist die Qualität der erzeugten Bevölkerung bei Methode C aber auch am höchsten. Dies ist auf eine explizite Anpassung an Haushalts- sowie Personenrandsummen zurückzuführen. Methode C eignet sich insbesondere für mikroskopische Verkehrsnachfragemodelle, weil

diese Informationen über einzelne Personen sowie den jeweiligen Haushaltskontext benötigen.

Sowohl Methode B als auch Methode C können für die Erstellung von synthetischen Bevölkerungen für Prognosejahre verwendet werden. Problematisch ist hierbei jedoch die Beschaffung der notwendigen Randsummen, die oftmals nur für eine geringe Anzahl von Attributen auf hohem Aggregationsniveau vorliegen. Dynamische Bevölkerungsfortschreibungen, beispielweise mit Hilfe von Evolutionsmodellen, können hier Abhilfe schaffen.

6 Literatur

[1] FRIEDRICH, M. (2011): Wie viele? Wohin? Womit? Was können uns Verkehrsmodelle wirklich sagen? In: Tagungsbericht Heureka 11, Verlag: FGSV.

[2] SCHNABEL, W.; LOHSE, D. (2011): Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung. Band 2: Verkehrsplanung, 3. Auflage, Verlag: Beuth.

[3] STEIERWALD, G.; KÜNNE, H.-D.; VOGT, W. (1994): Stadtverkehrsplanung. Grundlagen - Methoden - Ziele. Verlag: Springer.

[4] ORTÚZAR, J.; WILLUMSEN, L. G. (2001): Modelling Transport. 3. Auflage, Verlag: Wiley & Chichester.

[5] HEINRICHS, M.; KRAJZEWICZ, D.; CYGANSKI, R.; VON SCHMIDT, A. (2016): Disaggregated Car Fleets in Microscopic Travel Demand Modelling. In: The 7th International Conference on Ambient Systems, Networks and Technologies. http://dx.doi.org/10.1016/j.procs.2016.04.111

[6] BECKMAN, R. J.; BAGGERLY, K. A.; MCKAY, M. D. (1996): Creating Synthetic Baseline Populations. Transportation Research Part A: Policy and Practice, 30 (6), 1996, pp. 415-429.

[7] GUO, J. Y.; BHAT, C. R. (2007): Population synthesis for microsimulating travel behavior, Transportation Research Record, 2014 (12), pp. 92-101.

[8] BAR-GERA, H.; KONDURI, K.; SANA, B; Ye, X.; PENDYALA, R. M. (2009): Estimating survey weights with multiple constraints using entropy optimization methods. Paper presented at the 88th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2009.

[9] YE, X.; KONDURI, K; PENDYALA, R. M.; SANA, B.; WADDELL, P. (2009): A methodology to match distributions of both household and person attributes in the generation of synthetic populations. Paper presented at the 88th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2009.

[10] AULD, J.; MOHAMMADIAN, A. (2010): Efficient Methodology for Generating Synthetic Populations with Multiple Control Levels. Transportation Research Record: Journal of the Transportation Research Board, 2175, pp. 138-147.

[11] MÜLLER, K.; AXHAUSEN, K. W. (2011): Hierarchical IPF: Generating a synthetic population for Switzerland. Paper presented at the 51st Congress of the European Regional Science Association, Barcelona, September 2011.

[12] PRITCHARD, D. R.; MILLER E. J. (2012): Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. Transportation, 39 (3), pp. 685-704.

[13] GARGIULO, F.; TERNES, S.; HUET, S.; DEFFUANT, G. (2010): An iterative approach for generating statistically realistic populations of households. PLoS ONE 5 (1): e8828, http://dx.doi.org/10.1371/journal.pone.0008828

[14] BARTHELEMY, J.; TOINT, P. L. (2013): Synthetic Population Generation without a Sample. Transportation Science 47, no. 2 (2013): pp. 266-279.

[15] FAROOQ, B.; HURTUBIA, R.; BIERLAIRE, M. (2015): Simulation based generation of a synthetic population for Brussels. Integrated Transport and Land Use Modeling for Sustainable Cities, ISBN: 978-2-940222-72-8, pp. 95 - 112

[16] NAMAZI-RAD, M.-R.; HUYNH, N.; BARTHELEMY, J.; PEREZ, P. (2014): Synthetic Population Initialization and Evolution-Agent-Based Modelling of Population Aging and Household Transitions. PRIMA 2014: Principles and Practice of Multi-Agent Systems, Volume 8861, pp. 182-189, http://dx.doi.org/10.1007/978-3-319-13191-7

[17] MOECKEL, R. 2016: Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. In JTLU Volume 10 No. 2 [2017] pp. 1-18.

[18] BOWMAN, J. L. (2008): Population synthesizers, Traffic Engineering and Control, 49 (9) p. 342.

[19] MÜLLER, K.; AXHAUSEN, K. W. (2011): Population Synthesis for Microsimulation: State of the Art. Papers presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C, January 2011.

[20] DEMING, W. E.; STEPHAN, F. F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics, 11 (4), pp. 427-444.

[21] SUN, L.; ERATH. A. (2015): A Bayesian network approach for population synthesis. Transportation Research Part C: Emerging Technologies, 61, pp. 49-62

[22] STATISTISCHES BUNDESAMT: Zensus 2011 https://www.zensus2011.de (Stand: 16.01.2017)

[23] STATISTISCHES BUNDESAMT: Der Mikrozensus stellt sich vor. https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Mikrozensus.html (Stand: 16.01.2017)

[24] MOBILITÄT IN DEUTSCHLAND (MiD): Hrsg. BMVBS http://www.mobilitaet-in-deutschland.de (Stand: 16.01.2017)

[25] MOBILITÄT IN STÄDTEN (SrV): Hrsg. TU Dresden https://tu-dresden.de/die_tu_dresden/fakultaeten/vkw/ivs/srv (Stand: 16.01.2017)

[26] STATISTISCHES BUNDESAMT (2011): Qualitätsbericht - Mikrozensus 2010. https://www.destatis.de/DE/Publikationen/Qualitaetsberichte/Bevoelkerung/Mikrozensus2010.pdf (Stand: 16.01.2017)