

Workshop-Bericht: Synthetische Populationen für die Politikberatung in der Schweiz

Bern, 8. Dezember 2017



Bundesamt für Raumentwicklung ARE, Sektion Grundlagen

Nicole Mathys

Andreas Justen

EBP

Peter de Haan

Adrian Stetter

30. Januar 2018

2.4 Generierung synthetischer Bevölkerungen für Berlin - Möglichkeiten und Grenzen (Cyganski, DLR)

Name:	Antworten
Ersteller	Institut für Verkehrsforschung, Deutsches Zentrum für Luft- und Raumfahrt e.V.
Partner	-
Software	SYNTHESIZER (Freigabe open source geplant)
Themengebiete	Bisher v.a. Verkehrsnachfragemodellierung; (Wohn-) Standortwahlmodellierung
Zeitliche & Geographische Auflösung	Referenzjahr 2010, Grossregion Berlin, Haushalte mit Adresskoordinaten für jede Jahresscheibe
Nächste Schritte	Nutzung für weitere Szenarienrechnungen der Verkehrsnachfrage und Standortwahl; Fortschreibung Strukturgrößen über ein Evolutionsmodell; Neuberechnung Regressionsmodelle für die Mobilitätswerkzeuge mit aktualisierten Datenquellen
Prognosefähigkeit	Ja, separate Zeitscheiben für 2020, 2030, 2040
Grösse der grundlegenden Datenmenge	Stichprobengröße variierend je nach Untersuchungsraum; in der Regel aufbauend auf Mikrozensus (0.1% aller Haushalte)

Anforderungen an die synthetische Population

Künstlich generierte Bevölkerungen gehören zu den wichtigsten Eingangsdaten von Verkehrs- und Standortwahlmodellen. Güte und Genauigkeit der mit den Simulationen erzeugten Ergebnisse hängen dabei direkt mit der Qualität der genutzten Eingangsdaten zusammen. Die konkreten Anforderungen an die Genauigkeit und die Attribute der synthetischen Bevölkerungsdaten sind dabei abhängig von der Art des Modelleinsatzes und der avisierten Fragestellung. Am Institut für Verkehrsforschung des Deutschen Zentrums für Luft- und Raumfahrt e.V. (IVF) finden entsprechende Eingangsdaten bisher ihre Anwendung für die Berechnung der Verkehrsnachfrage im mikroskopischen Nachfragemodell TAPAS (Heinrichs et al. 2016) sowie dem auf dem Simulationspaket Cube Land basierenden Standortwahlmodell SALSA (Martinez et al 2010; Heldt et al. 2017). Im Fokus liegt dabei in der Regel der Untersuchungsraum Berlin, für den je nach Projektanforderung verschiedene Zeitscheiben sowie Prognosen der Bevölkerungsentwicklung genutzt werden.

Für eine automatisierte Erstellung der synthetischen Bevölkerung wurde eine modulare, Java-basierte Applikation entwickelt, die die Nutzung verschiedener gängiger Verfahren zu Erstellung einer Bevölkerung auf Basis einer vorliegenden Stichprobe ermöglicht. Der sogenannte SYNTHESIZER (von Schmidt et al. 2017) ermöglicht a) die einfache Hochrechnung einer Stichprobe anhand eines gegebenen Hochrechnungsfaktors, b) die Anpassung einer Stichprobe an

Haushalts- oder Personenrandsummen oder c) die gleichzeitige Einhaltung gegebener Haushalts- und Personenrandsummen (vgl. Abbildung 6).

Variante A stellt die einfachste Art dar, auf Basis einer gegebenen, repräsentativen Stichprobe aus der Gesamtbevölkerung sowie eines bekannten Hochrechnungsfaktors eine synthetische Bevölkerung des Untersuchungsgebietes zu erzeugen. Hierbei wird jeder Haushalt samt der dazugehörigen Personen so oft aus der Stichprobe kopiert, wie er nach dem jeweiligen Hochrechnungsfaktor vorkommt (vgl. Moekel 2016).

Auch *Variante B* startet mit einer disaggregierten Bevölkerungsstichprobe. Darüber hinaus setzt sie das Vorhandensein von Randsummen all derjenigen Attribute voraus, hinsichtlich derer die Zielbevölkerung eine korrekte Verteilung aufweisen soll. Dabei kann es sich entweder um personen- oder haushaltsbezogene Randsummen handeln. Mit Hilfe mathematischer Anpassungsverfahren, in der Regel des sogenannten Iterative Proportional Fitting (IPF) (siehe Beckman et al. 1996; Farooq et al. 2013), werden im Laufe der Erstellung der synthetischen Bevölkerungen die Wahrscheinlichkeiten für die Auswahl eines Haushaltes oder einer Person derart iterativ angepasst, dass die Randsummen getroffen werden. Das Verfahren läuft solange, bis eine vorab definierte Qualitätsschwelle oder maximale Iterationszahl erreicht wurde.

Variante C erweitert das Vorgehen aus Variante B dahingehend, dass gleichzeitig vorliegende Randsummen für die Attribute der Personen sowie der Haushalte eingehalten werden. Neben dem IPF-Verfahren kommt hier in der Regel das sogenannte Iterative Proportional Updating (IPU; siehe von Schmidt et al. 2017, Müller und Axhausen 2011 sowie Ye et al. 2009) Verfahren zur Anpassung der Haushaltsgewichte zum Einsatz.

Detaillierte Informationen zum Vorgehen sowie den Vor- und Nachteilen der einzelnen Varianten finden sich in (von Schmidt et al. 2017); Müller und Axhausen (2011) gibt einen Überblick über unterschiedliche Varianten und Erweiterungen des IPF. Eine Vorstellung des IPU findet sich bei Ye et al. (2009). In der Regel findet dabei bei der Erstellung von synthetischen Bevölkerungen für die Simulationsmodelle am Institut für Verkehrsforschung die Variante C Verwendung, bei der die Einhaltung der vorgegebenen Randsummen mit Hilfe iterativer IPF- und IPU-Anpassungen erfolgt.

Wie anhand der Abbildung 6 ersichtlich, erfolgt die Erstellung der synthetischen Bevölkerung in drei Arbeitsschritten, die sich im Vorgehen sowie der zugrundeliegenden Datenbasis stark unterscheiden. Zunächst wird die Grundbevölkerung mit Hilfe eines der oben genannten Verfahren erstellt. Die nachfolgend näher beschriebenen Datengrundlagen weisen dabei zwei Schwachpunkte auf, die in den beiden anschließenden Bearbeitungsschritten kompensiert werden sollen. Zunächst liegen die Randsummen meist mit einer geringen räumlichen Auflösung vor, sodass anschließend eine räumliche Vereinzelung mit der Zuweisung einer konkreten X-Y-Koordinate erfolgt. Zudem weisen die verwendeten Daten keine mobilitätsbezogenen Informationen auf. Angaben zum Besitz eines Führerscheins, eines ÖV-Abos und eines Pkws werden daher mit Hilfe spezifischer Regressionsmodelle in einem weiteren Arbeitsschritt angespielt. Je nach Projektkontext kann hier auch eine Differenzierung der Pkw-Flotte vorgenommen werden (vgl. z.B. Heinrichs et al. 2016). Für alle drei Bearbeitungsschritte sollen die verwendeten Datenquellen, Attribute und Einschränkungen nachfolgend kurz adressiert werden.

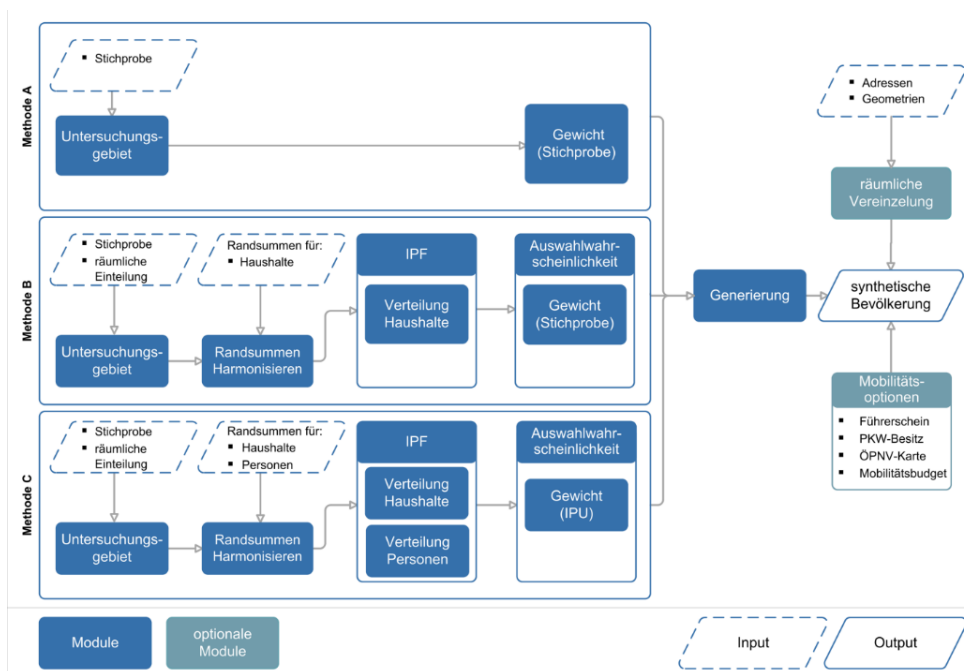


Abbildung 6: Methoden der Erstellung synthetischer Bevölkerungen mit dem SYNTHESIZER.

Für die Erstellung der Basisbevölkerung wird auf den deutschen Mikrozensus (Statistisches Bundesamt 2018) als Stichprobenbasis zurückgegriffen. Er enthält umfassende Informationen zu den einzelnen Personen, ihrem Haushaltskontext und ihrer Wohnsituation. Insgesamt beinhaltet er über 100 Attribute, von denen jedoch nur eine Auswahl bei der Bevölkerungsgenerierung hinsichtlich ihrer Verteilung berücksichtigt wird. Detaillierte Angaben sind hier nur auf Bundeslandebene zu erhalten, Randsummen auch auf Bezirksebene. Für die Erstellung der Randsummen, hinsichtlich derer die Bevölkerung im Erstellungsverfahren geprüft wird, wird für die Basisbevölkerung in der Regel ebenfalls auf den Mikrozensus zurückgegriffen. So vorhanden können hier jedoch auch räumlich spezifischere Daten Verwendung finden: im Falle des Untersuchungsraumes Berlin liegen so teilweise Daten der Berliner Senatsverwaltung auf Ebene der Berliner Bezirke vor. Insgesamt lässt sich konstatieren, dass die benötigten Daten zur Bestimmung der Randsummen bereits für die Abbildung des Basiszustandes nur teilweise in der benötigten Art und räumlich differenzierten Auflösung zur Verfügung stehen und oftmals aus unterschiedlichen Quellen stammen. Dies gilt insbesondere für die Informationen zum Berufsstatus der Personen sowie zum Haushaltseinkommen. Dies führt in der Regel zu aufwändigen Arbeiten für die Harmonisierung der Randsummen sowie zur Notwendigkeit der Festlegung der wichtigsten Referenzquelle.

Zu den häufig notwendigen Harmonisierungsschritten 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).

Datengrundlagen

Als wichtigste Datenquellen für die Basissituation sind der Mikrozensus zu nennen, für die Prognosezeitpunkte wird in der Regel auf die Bevölkerungsfortschreibung des Bundesamtes für Bau-, Stadt- und Raumforschung (BBSR 2012) zurückgegriffen.

Für die präzise räumliche Verortung der so generierten Grundbevölkerung auf Ebene der Bezirke werden Daten des Digitalen Landschaftsmodells (DLM) des Bundesamts für Kartographie und Geodäsie (BKG), des amtlichen Liegenschaftskatasterinformationssystems (ALKIS) sowie der Adressdatensatz des Bundesamts für Kartographie und Geodäsie (BKG) verschnitten. Auf diese Weise wird ermittelt, welche Adresskoordinaten für die Zuweisung eines Wohnstandortes genutzt werden sollten. Die Grundbevölkerung wird dann entsprechend in den einzelnen Häusern positioniert, und die Koordinateninformationen können für die Generierung einer hochauflösenden Nachfrage eingesetzt werden.

Angaben zu den Mobilitätsoptionen liegen in der Regel nicht an den für die Generierung der Grundbevölkerung genutzten Daten vor. Auf Basis von spezifischen Verkehrserhebungen, für Berlin dem Datensatz der Erhebung Mobilität in Städten (SrV) (TU Dresden 2018), werden separate Regressionsmodelle zur Ausstattung der Personen und Haushalte mit Führerscheinen, ÖV-Karten sowie Pkws geschätzt. Anhand der Attribute der synthetischen Bevölkerung werden anschließend die jeweiligen Ausprägungen an die Basisbevölkerung angespielt.

Tabelle 1 zeigt eine Übersicht der Attribute und ihrer Ausprägungen der fertigen synthetischen Bevölkerung. Fett dargestellt sind dabei diejenigen Attribute, die bei der Erstellung der Bevölkerung mit Hilfe der Methode Variante C hinsichtlich ihrer Randsummen geprüft werden.

Attribut	Attributausprägung
HH-ID	...
HH-Größe	1, 2, 3, 4 und 5+ Personen
HH-Einkommensgruppe (Euro)	0 - 899, 900 - 1499, 1500 - 1999, 2000 - 2599, 2600 - 3199, 3200+
HH-Einkommen (Euro)	Für jede Einkommensgruppe wird eine gleichverteilte Zufallszahl erzeugt, um somit diskrete Einkommenswerte zu ermitteln. Bei Prognosejahren wird zusätzlich noch eine jährliche Wachstumsrate hinzuaddiert.
Kinder im HH	Ja/Nein
HH-Typ	TAPAS-HH-Typ z.B. Zweipersonenhaushalt (2 Erwachsene), Zweipersonenhaushalt (1 Erwachsener mit Kind)
Anzahl Autos im HH	0/1/2
Welches bzw. welche Autos?	z.B. Größe/Antrieb/Automatisierung
Verkehrszelle-ID	z.B. Bezirk/TVZ
Koordinaten	Adresse
Personen-ID	...
HH-ID	...
Geschlecht	männlich, weiblich
Alter	0, 1, 2, ..., 100
Status	<u>Nichterwerbspersonen (NEP)</u> Kind unter 6, Schüler, Student, Rentner, sonstige NEP <u>Erwerbspersonen (EP)</u> Vollzeit, Teilzeit, Erwerbslos
Personen-Typ	TAPAS-Personen-Typ z.B. erwerbstätig, Mann oder Frau, kein Pkw im HH, bis 24 Jahre
Führerschein	JA/NEIN
ÖPNV-Ticket	JA/NEIN
Fahrrad	JA/NEIN
MB-MIV-Variable	Mobilitäts-Budget: MIV-Variable (z.B. für Kraftstoff, Wartung)
MB-MIV-Fix	Mobilitäts-Budget: MIV-Fix (z.B. für Steuern, Versicherung, Kfz-Erwerb)
MB- ÖPNV	Mobilitäts-Budget: ÖPNV (z.B. für Monatsticket)

Tabelle 1: Übersicht der Bevölkerungsattribute und ihrer Ausprägungen; bei der Erstellung nach Variante C mit Hilfe der Randsummen kontrollierte Attributverteilungen sind fett markiert

Weiterführende Arbeiten: Prognose, Qualitätskontrolle und Ausdehnung des Untersuchungsraumes

Die Erstellung der synthetischen Bevölkerung erfolgt momentan für jede betrachtete Zeitscheibe separat; eine Fortschreibung über die Jahre mit Hilfe von Evolutionsmodelle etc. ist bisher noch nicht umgesetzt. Eine Aktualisierung der Bevölkerung wird in Abhängigkeit von den jeweiligen Anforderungen der Projekte vorgenommen, die oftmals unterschiedliche Anforderungen hinsichtlich der Datenaktualität, der Konsistenz mit anderen Eingangsdaten oder spezifischen räumlichen Bezügen aufweisen.

Aufgrund der unterschiedlichen Datenquellen kommen für die Prüfung der Qualität der erstellten Bevölkerung unterschiedliche Verfahren zum Einsatz. Für die erstellte Grundbevölkerung findet eine Prüfung der Abweichung der Attributverteilungen in der erstellten Bevölkerung von den bekannten Randsummen statt. Als Indikator wird hier der oft verwendete Standardized *Root Mean Square Error* (SRMSE) berechnet (Prichard und Miller 2012). Da die Randsummen oftmals nur auf hoher Aggregationsebene vorliegen, finden darüber hinaus kartenbasierte visuelle Prüfungen der absoluten und relativen Bevölkerungsanteile und ihre Veränderung zwischen verschiedenen Zeitscheiben Anwendung. Diese werden auch genutzt, um die feinräumliche Verteilung der Bevölkerung nach der Vereinzelung zu prüfen. Die Prüfung der mobilitätsbezogenen Eigenschaften der Bevölkerung findet im Abgleich mit den für die Regressionsmodelle verwendeten Eingangsdaten, in der Regel der SrV (TU Dresden 2018) oder der MiD (BMVI 2018), statt. Hier werden die jeweiligen Anteile für die vorliegenden Bezugseinheiten berechnet sowie eine kleinräumigere Prüfung anhand von Karten vorgenommen. Eine Prüfung der Verteilungen der im Mikrozensus vorhandenen zusätzlichen Attribute, die nicht bei der Erstellung der Bevölkerung als Randsummen Eingang finden, wird nicht vorgenommen.

Der SYNTHESIZER wurde bisher vorrangig zur Erstellung der Bevölkerungen für Berlin sowie den Untersuchungsraum Braunschweig eingesetzt. Der flexible, modulare Aufbau ermöglicht jedoch eine schnelle Anpassung auch für andere Einsatzzwecke. Neben einer potentiellen Nutzung für die Generierung passender Eingangsdaten für das deutschlandweite makroskopische Personenverkehrsnachfragemodell des IVF wird momentan der Einsatz für die Fortschreibung von Strukturdaten für die Nachfragemodellierung geprüft. Auf Basis einer vorliegenden Vollerhebung von Standorten des Lebensmittel-einzelhandels im Jahr 2015 erfolgte eine nach Art und Größenklasse differenzierte Fortschreibung der Einkaufsorte 2030 mit Hilfe der Methode Variante B.

6. Literaturverzeichnis

- Beckman R.J., Baggerly K.A., McKay M.D. (1996): Creating synthetic baseline populations. *Transportation Research Part A* 30 (6), 415–429.
- BBSR (2012): Raumordnungsprognose 2030. Bevölkerung, private Haushalte, Erwerbspersonen. *Analysen Bau.Stadt.Raum Bd.9*, Bonn.
- BMVI (2018): Bundesministerium für Verkehr und digitale Infrastruktur: Mobilität in Deutschland (MiD), bundesweite Befragung von Haushalten zu ihrem alltäglichen Verkehrsverhalten, <http://www.mobilitaet-in-deutschland.de/>
- Bundesamt für Raumentwicklung ARE (2014): Entwicklung eines Flächen-nutzungsmodells für die Schweiz <http://www.are.admin.ch/flnm>
- Bundesamt für Raumentwicklung ARE (2017): Weiterentwicklung Flächen-nutzungsmodellierung: Wohnstandortwahl: Erweiterung des Modells FaLC: Verhaltensmodelle und synthetische Population; <http://www.are.admin.ch/flnm>
- Farooq B., Bierlaire M., Hurubia R., Flötteröd G. (2013): Simulation based population synthesis. *Transportation Research Part B, Volume 58*, December 2013, Pages 243-263.
- Heinrichs, M., Krajzewicz, D., Cyganski, R. & von Schmidt, A. (2016): Introduction of car sharing into existing car fleets in microscopic travel demand modelling, in: *Personal and Ubiquitous Computing*, S. 1-11, Springer. DOI: <https://doi.org/10.1007/s00779-017-1031-3>, ISSN 1617-4909, 2017.
- Heinrichs, M., Krajzewicz, D., Cyganski, R. & von Schmidt, A. (2016): Disaggregated Car Fleets in Microscopic Travel Demand Modelling, In *Procedia Computer Science*, Volume 83, S. 155-162, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.04.111>.
- Heldt, B., Donoso, P., Bahamonde-Birke, F., & Heinrichs, D. (2017): Estimating bid-auction models of residential location using census data with imputed household income. Accepted for *Journal of Transport and Land Use*.
- Horni, A., K. Nagel and K.W. Axhausen (eds.) (2016): *The Multi-Agent Transport Simulation MATSim*, Ubiquity, London. DOI: <https://doi.org/10.5334/baw>
- Martínez, F., & Donoso, P. (2010): The MUSSA II land use auction equilibrium model. In F. Pagliara, J. Preston, & D. Simmonds (Eds.), *Residential Location Choice* (S. 99-113): Springer.
- Moekel, R. (2016): Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. *Journal of Transport and Land Use*, Vol. 10(2), S. 1-18.
- 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.
- Prichard, D. & Miller, E. J. (2012): Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. *Transportation*, 39 (3), S. 685-704.
- Statistisches Bundesamt 2018. Mikrozensus Verkehr. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Mikrozensus.html>
- TU Dresden (2018): Forschungsprojekt Mobilität in Städten – SrV, https://tu-dresden.de/die_tu_dresden/fakultaeten/vkw/ivs/srv
- von Schmidt, A., Cyganski, R. & Krajzewicz, D. (2017): Generierung synthetischer Bevölkerungen für Verkehrsnachfragemodelle - Ein Methodenvergleich am

Beispiel von Berlin, in: HEUREKA'17 - Optimierung in Verkehr und Transport, S. 193-210, FGSV-Verlag, ISBN 978-3-86446-177-4, 2017.

Ye, X., Konduri, K., Pendyala, R. M., Sana, B., Waddel, 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.