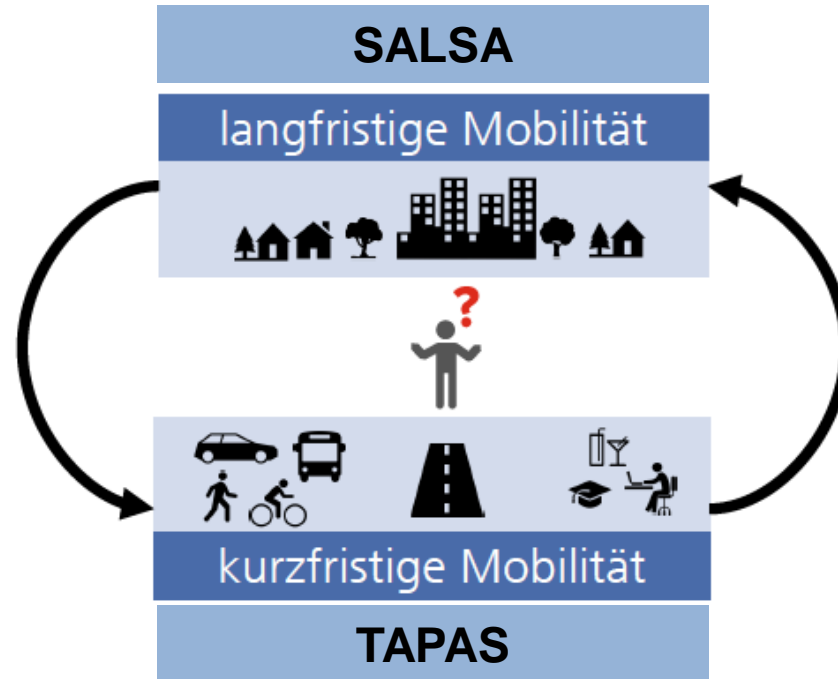


Generierung synthetischer Bevölkerungen für Berlin - Möglichkeiten und Grenzen

Rita Cyganski, Antje von Schmidt, Benjamin Heldt
DLR-Institut für Verkehrsforschung



Synthetische Bevölkerung – warum die Mühe?



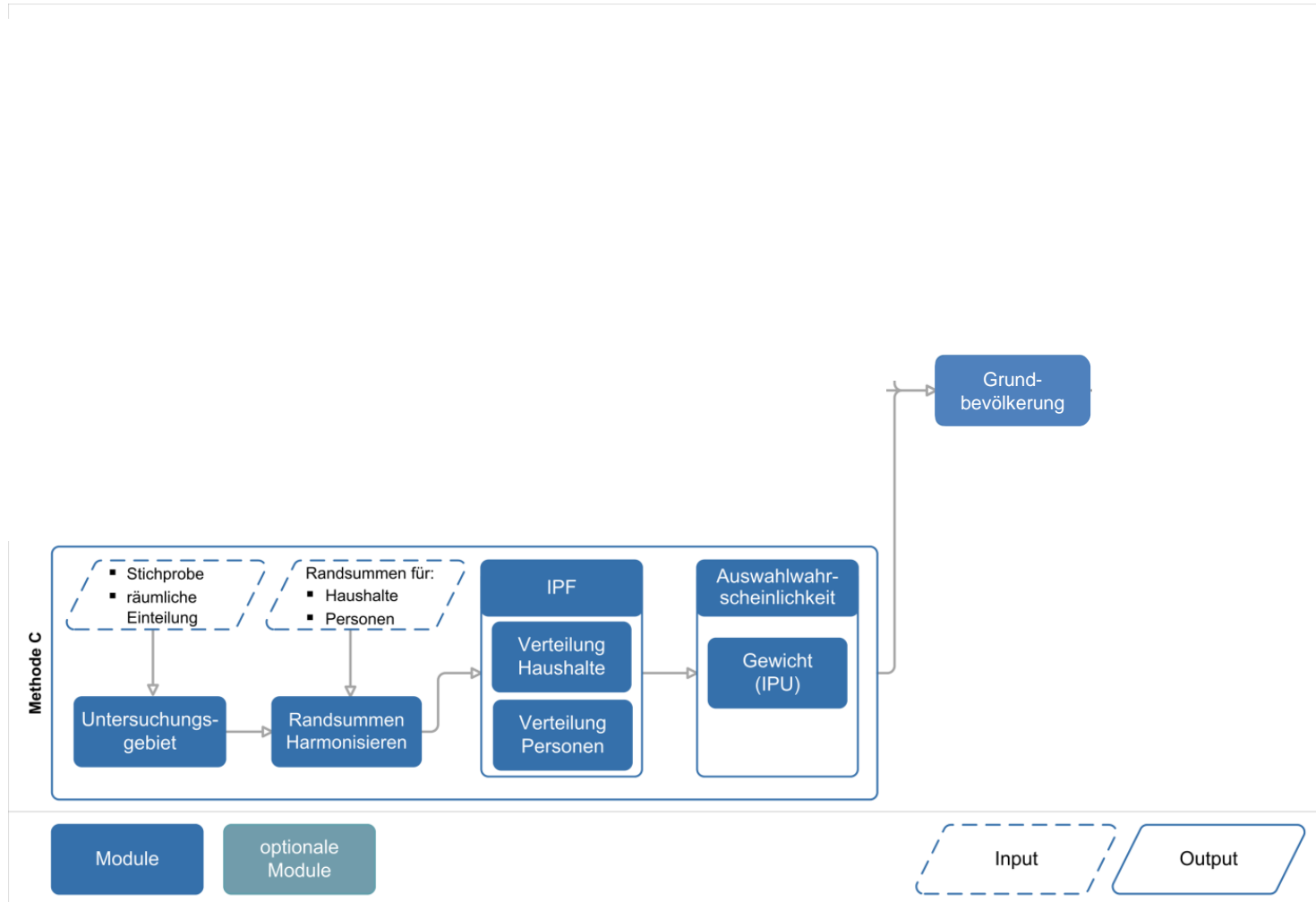
Perspektivische Nutzung

- Möglichkeit der aggregierten Nutzung in makroskopischen Modellen (PTV VISUM)
- Unternehmensstandortwahl (SALSA)
- Fortschreibung von Strukturgrößen, z.B. Standorte des Lebensmittelhandels nach Größe und Angebotstyp



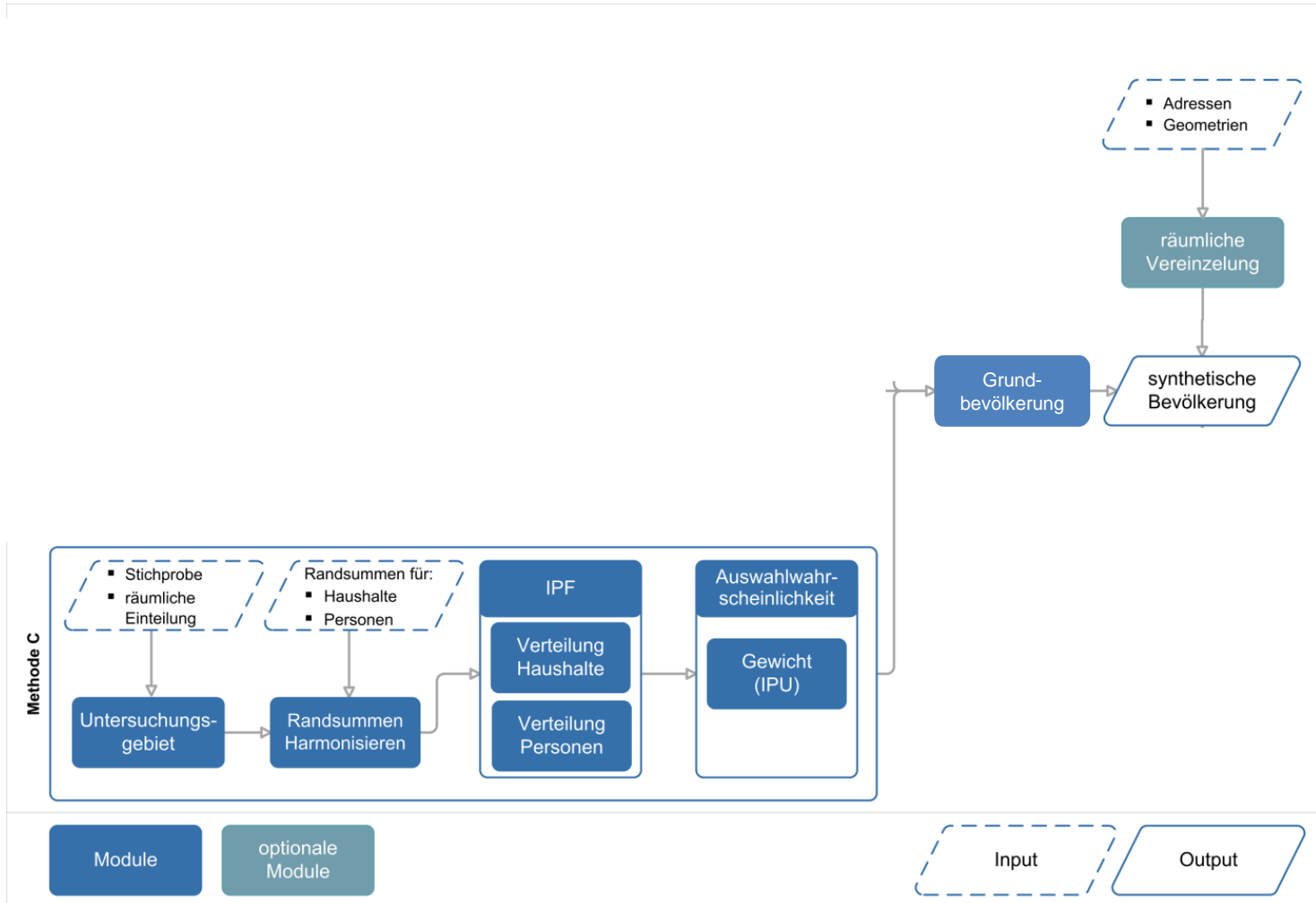
Erstellung synthetischer Bevölkerungen mit dem SYNTHESIZER

1



Erstellung synthetischer Bevölkerungen mit dem SYNTHESIZER

1



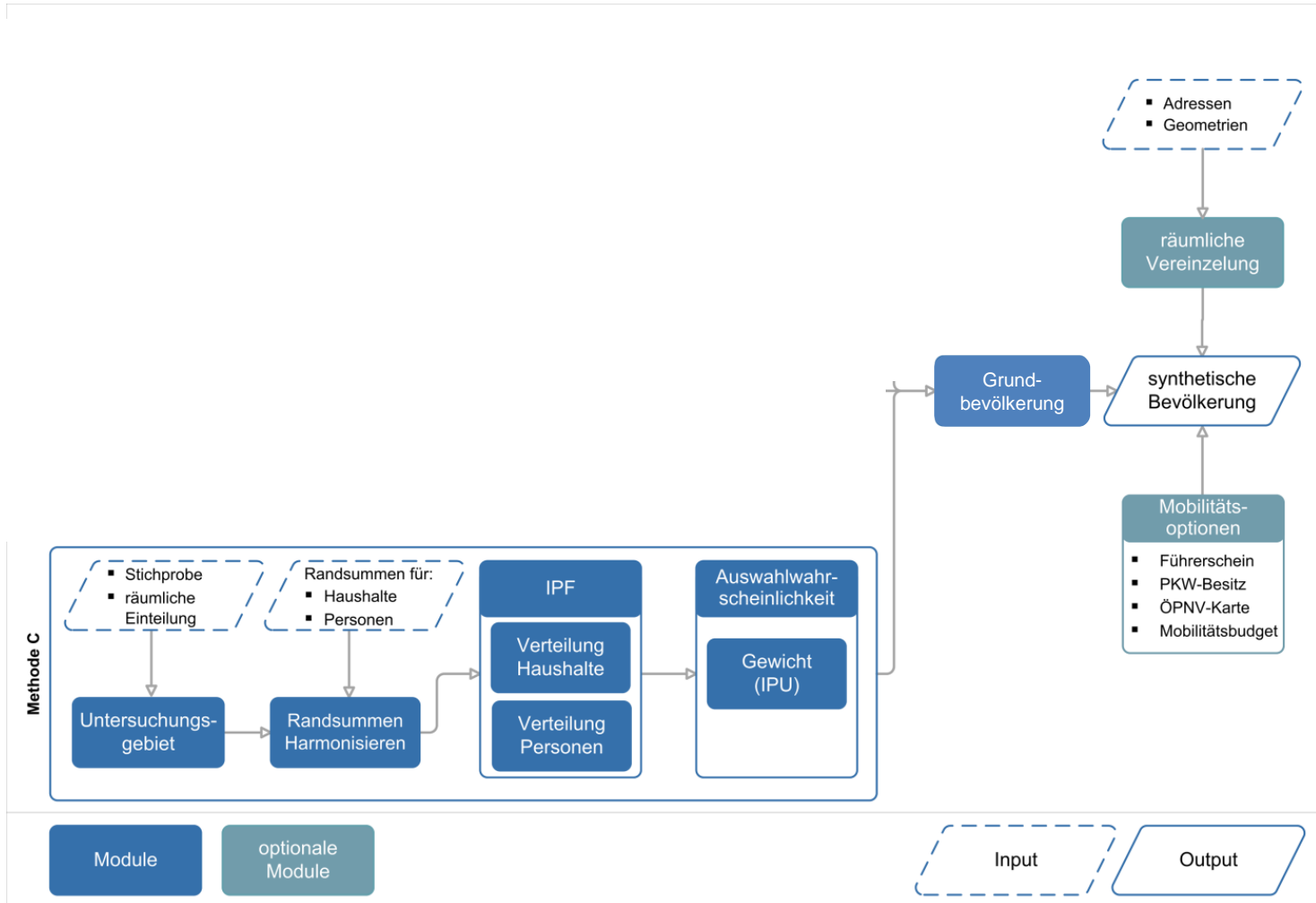
2

3



Erstellung synthetischer Bevölkerungen mit dem SYNTHESIZER

1

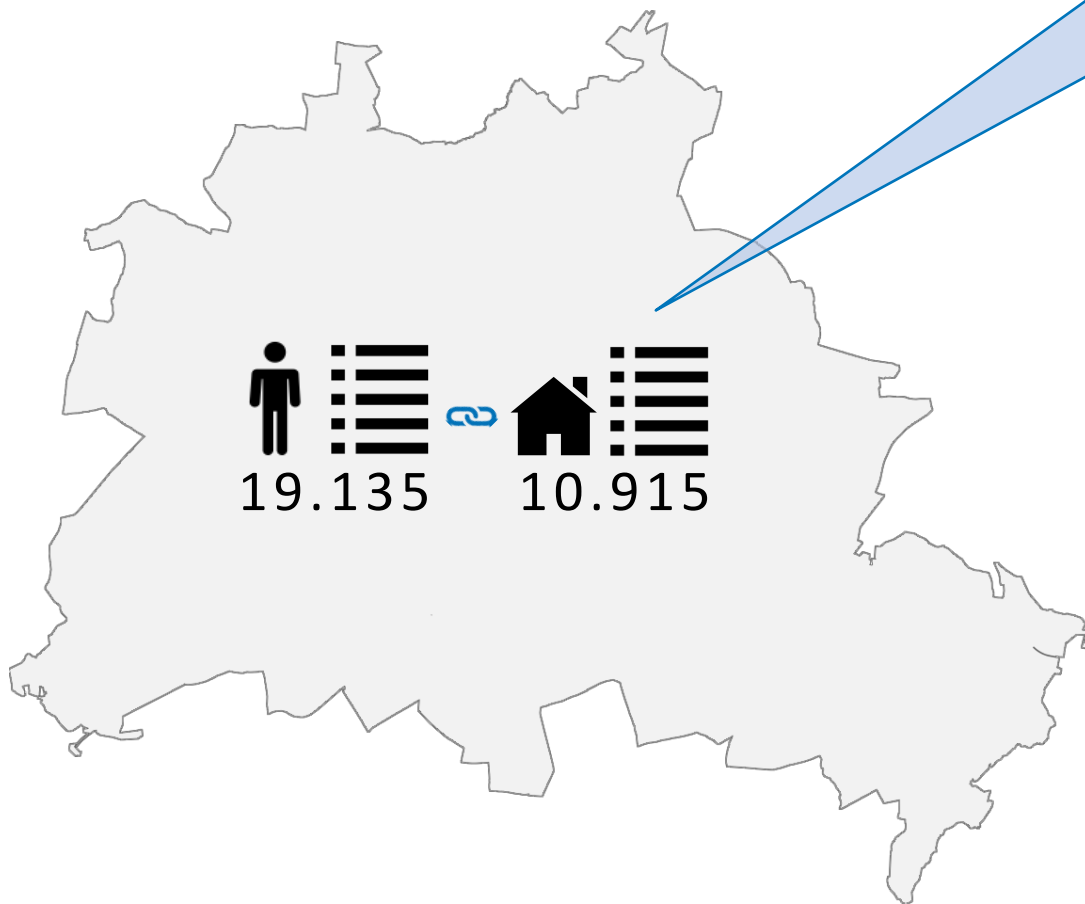


2

3



Unsere Ausgangsbasis: Stichprobendaten



Stichprobe
disaggregierte soziodemographische
Haushalts- und Personendaten
(Berlin gesamt) Quelle:
Mikrozensus 2010

HH-ID	HH-Größe	Einkommen	...
10001	2 Personen	5000 €	...

...

Haushalte

P-ID	HH-ID	Geschlecht	Alter	...
1000101	10001	männlich	52	...
1000102	10001	weiblich	48	...

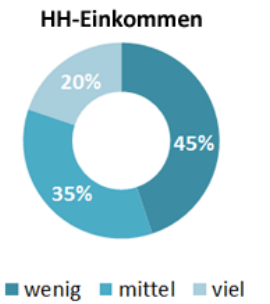
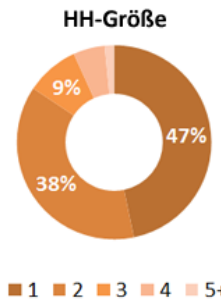
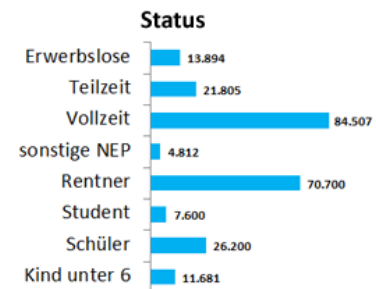
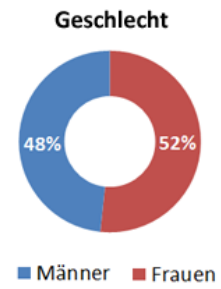
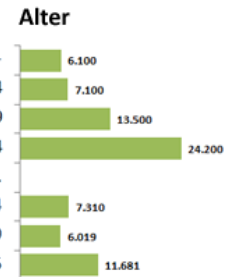
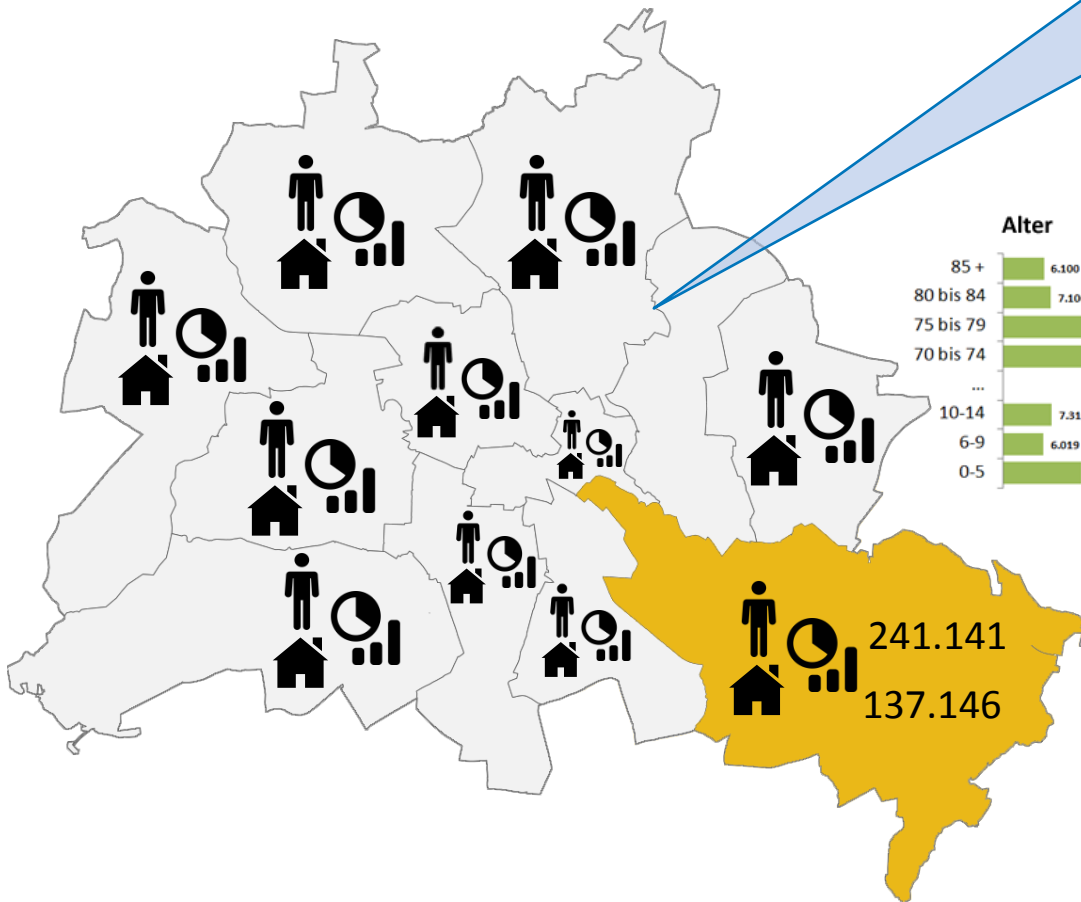
...

Personen



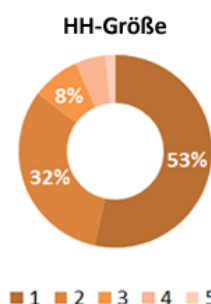
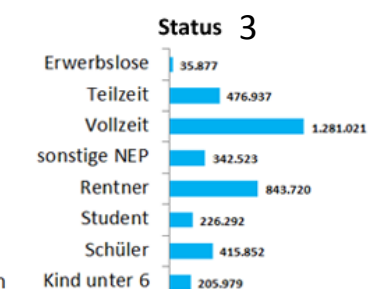
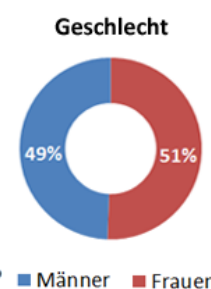
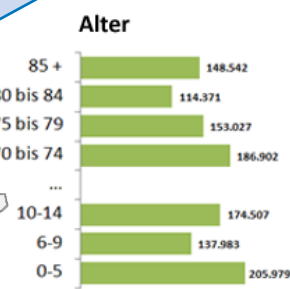
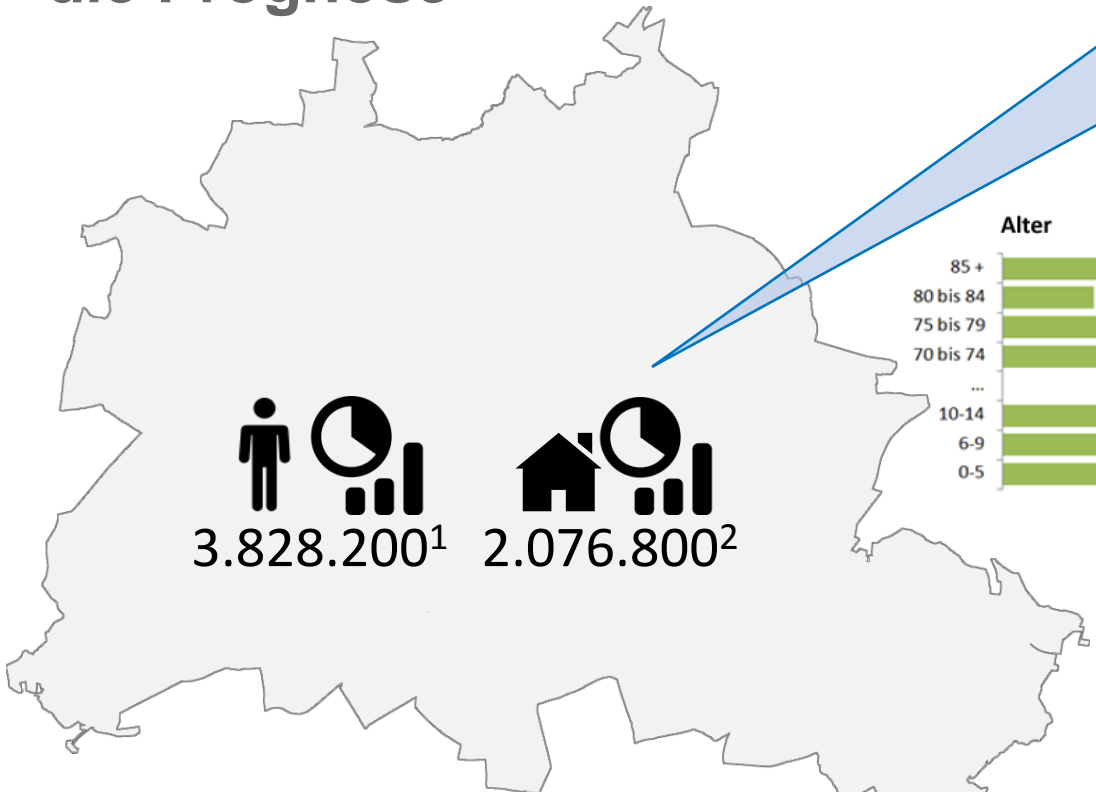
Festlegung der Zielverteilung für die Referenz

Randsummen (Referenzjahr)
 aggregierte soziodemographische
 Haushalts- und Personendaten
 (Berliner Bezirke) Quelle:
Mikrozensus 2010



Festlegung der Zielverteilung für die Prognose

Randsummen (Prognosejahr)
 aggregierte soziodemographische
 Haushalts- und Personendaten
 (Berlin gesamt) Quelle:
Bevölkerungsprognose 2030 (Senat)
Bevölkerungsprognose 2035 (BBSR)



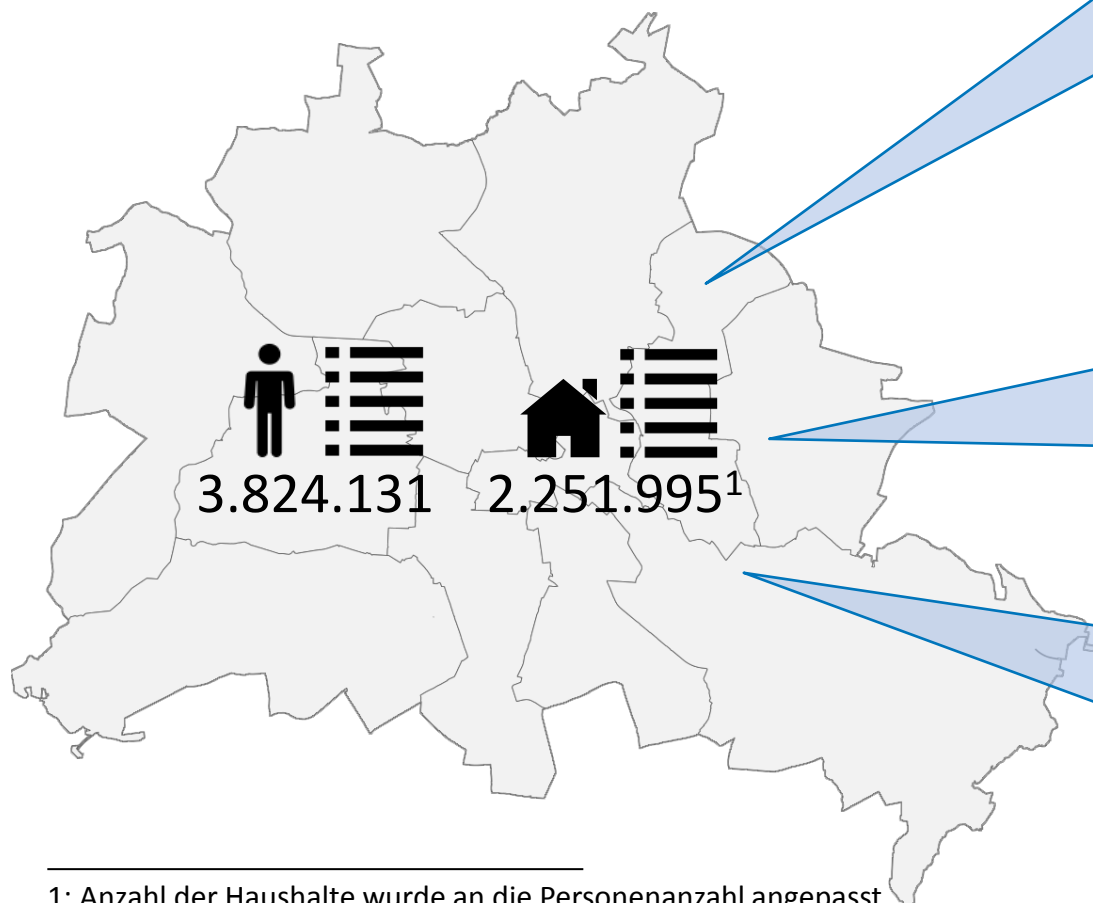
HH-Einkommen

% jährliche Wachstumsrate

- 1: Berliner Senatsverwaltung
- 2: BBSR 2030; Anzahl Haushalte basiert auf 3.534.255 Personen
- 3: Kinder, Schüler und Rentner gemäß Alter; rel. Änderung Stud.zahl (Kultusminister Konferenz); Erwerbspersonen (VEU2); 98% Beschäftigung (VP2030); Anteil Vollzeit/Teilzeit (Mikrozensus 2010)



Erweiterung der Grundbevölkerung



Grundbevölkerung
beinhaltet disaggregierte
soziodemographische Daten

Räumliche Auflösung
basiert auf den verwendeten
Randsummen (**Bezirksebene**)
deshalb **Haushalte auf
Adresskoordinaten verteilen**

Mobilitätsoptionen hinzufügen

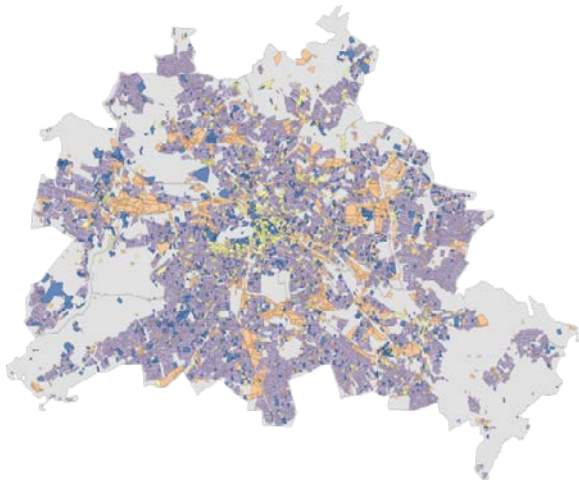
Quelle:
Verkehrserhebungen²

1: Anzahl der Haushalte wurde an die Personenanzahl angepasst.

2: Mobilität in Deutschland (MID), Mobilität in Städten (SrV)



2. Adresskoordinaten-feine Verteilung der Haushalte



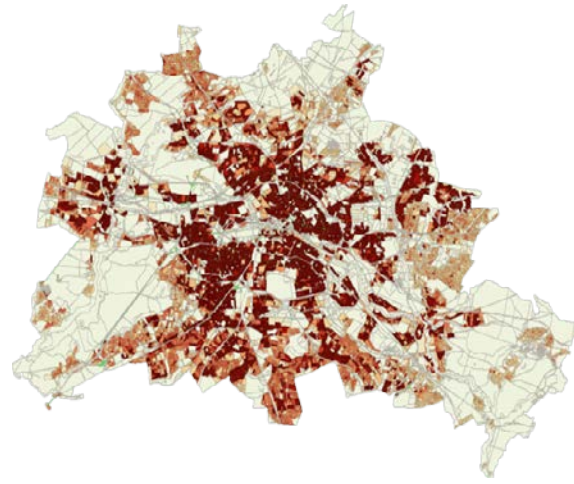
Flächennutzungstypen

(Wohnbaufläche, Industrie- und Gewerbefläche, usw.)

Quelle:

Digitales Landschaftsmodell (DLM)

Bundesamts für Kartographie und Geodäsie (BKG)

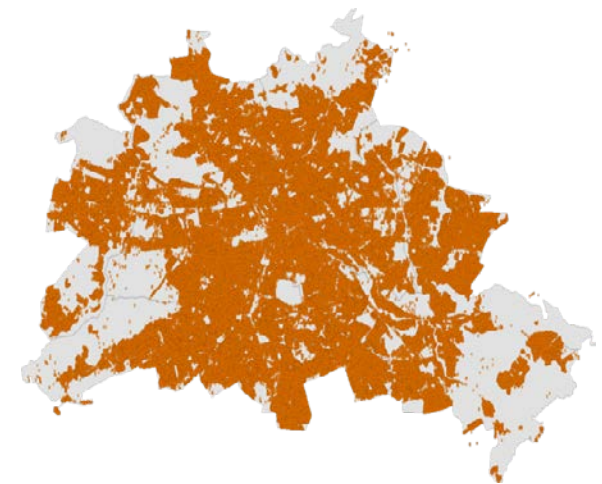


Gebäudedaten

(Wohnblöcke und Anzahl Bewohnern)

Quelle:

Amtliches Liegenschafts- katasterinformationssystem (ALKIS)



Adresdaten

(Koordinaten)






Quelle:

Adresdatensatz

Bundesamts für Kartographie und Geodäsie (BKG)



3. Ergänzung von Mobilitätsoptionen

	ca. 20 Minuten	ca. 40 Minuten	ca. 25 Minuten	ca. 10 Minuten	ca. 85 Minuten
Attribute					
Alter	✓		✓	✓	✓
Geschlecht	✓		✓		
Status		✓ ¹	✓ ^{2,3}		
HH-Größe	✓	✓		✓	
HH-Einkommen	✓	✓	✓	✓	✓
Anzahl Führerschein im HH	→	✓	✓	✓	
Anzahl Erwachsene im HH		✓	✓		
Anzahl Männer im HH (18+)		✓			
Alter älteste Person im HH		✓			
Anzahl Autos im HH			→	✓	✓
TAPAS HH-Typ					✓

Quelle

MiD 2008 MiD/SrV 2008

MiD 2008

EVS 2003

Status: 1)Anzahl Vollzeit Beschäftigte 2)Anzahl Teilzeit Beschäftigte 3)Anzahl Erwerbstätige



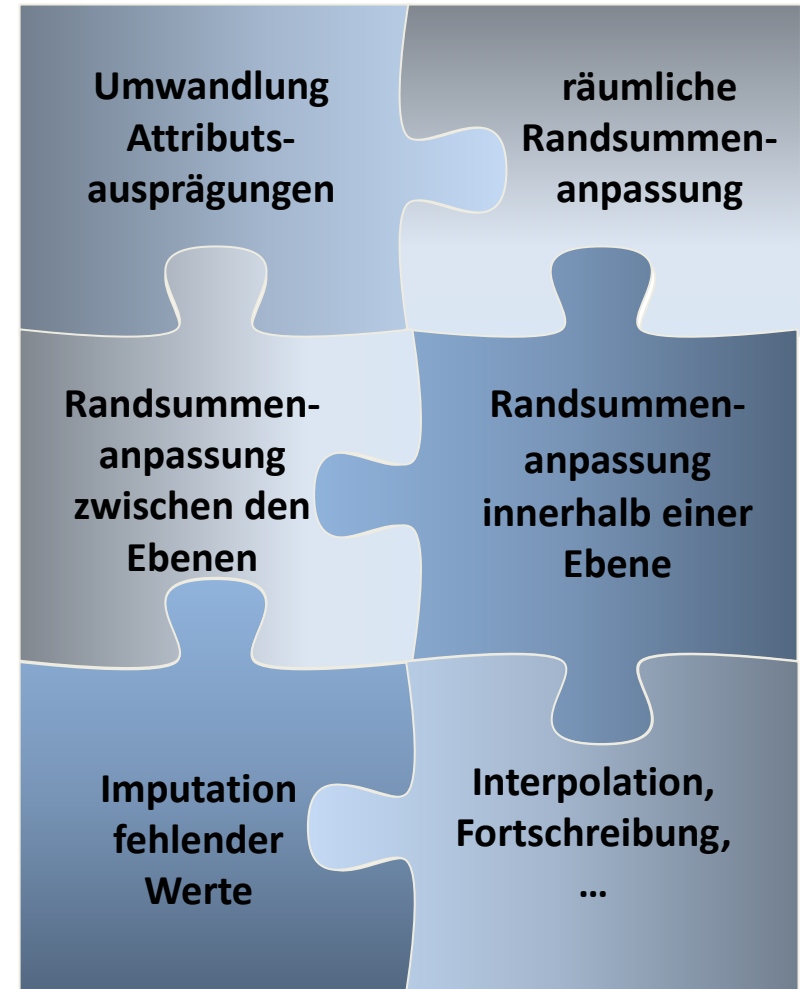
Attribute der erstellten synthetischen Bevölkerung

Grundbevölkerung				Erweiterte Bevölkerung	
Randsummen-kontrolliert		Nicht kontrolliert		Regressionsmodelle	
Alter	21	Kinder im Haushalt	0/1	Führerschein	0/1
Geschlecht	2	HH-Typ	18	ÖPNV-Abo	0/1
HH-Größe	5	HH-EK diskret		Anzahl Autos	0/1/2
HH-EK, gruppiert	6	<i>Bildungsstand</i>		Art der Autos	
Status	8	<i>Arbeitsplatzdetails</i>		Fahrrad	0/1
		<i>Entfernung Arbeitsstätte</i>		Verkehrszelle	
		<i>Renten und Transferl.</i>		Adresskoordinate	
		<i>Nationalität</i>		MB var	
		<i>Migr.hintergrund</i>		MB fix	
		<i>Wohnungswechsel</i>		MB ÖPNV	
		<i>Zweitwohnsitz</i>			
		...			



Herausforderung heterogene Datenbasis

- Synthese benötigter Attribute aus unterschiedlichen Datenquellen
 - Basisbevölkerung, Fortschreibungsinformationen, Mobilitätsoptionen, finanz. Situation
 - Datenschutzerfordernungen
 - Geringe Fallzahlen für spezifische Untersuchungsräume
 - Geringe räumliche Auflösung, insb. für Prognose
 - Heterogene Attributsausprägungen, Bezugsjahre, räumliche Bezüge und Auflösungen, Angaben / Werte
 - Langandauernde Datenakquisen, teilweise häufige Aktualisierungen
- **Notwendigkeit umfangreicher Harmonisierungsarbeiten !**



Qualitätsprüfung während und nach der Erstellung

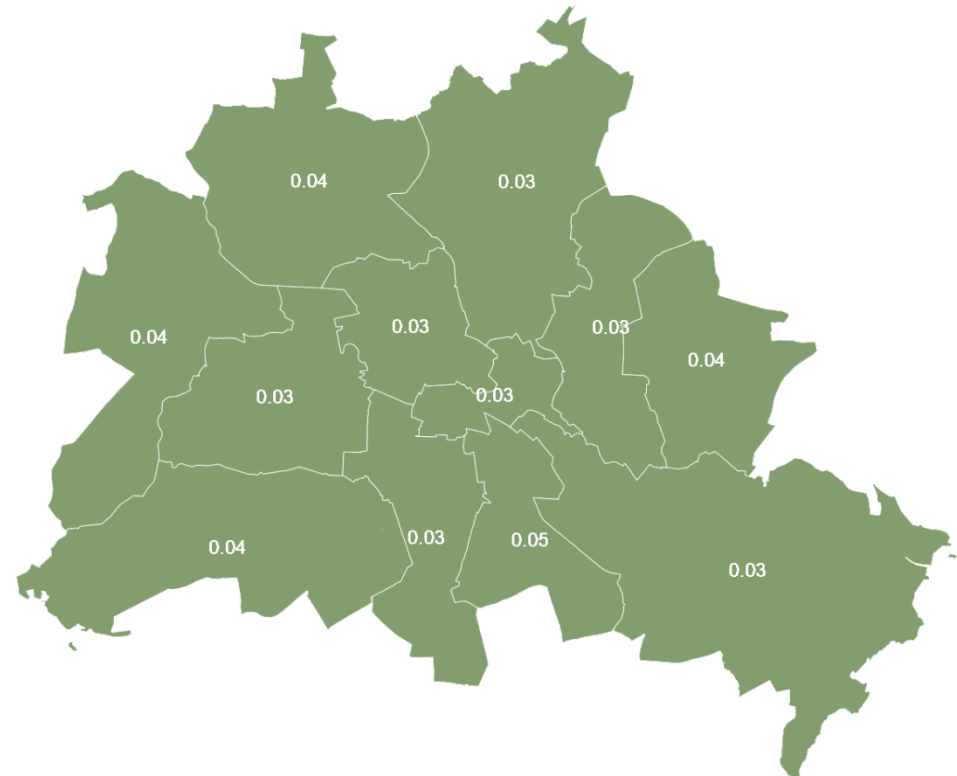
- Basisbevölkerung:
 - Randsummenprüfung während der Erstellung mit dem Synthesizer: Standardized Root Mean Square Error (SRMSE)¹
 - Visuelle Prüfung relativer und absoluter Abweichungen (Karten, Verteilungen,...)
- Mobilitätsoptionen
 - Visuelle Prüfung relativer und absoluter Abweichungen (Karten, Verteilungen,...)

Personenverteilung
Alter (21) x Geschlecht (2) x Status (8)

$$\text{SRMSE} = \sqrt{\frac{\sum_{ijk} (\text{Observed}_{ijk} - \text{Expected}_{ijk})^2}{ijk}}{\frac{\sum_{ijk} \text{Expected}_{ijk}}{ijk}}$$

SRMSE

{ min 0 (stimmt exakt überein)
max 1 (stimmt nicht überein)



Aktualität und Beschränkungen

- Separate Erstellung für jede betrachtete Zeitscheibe
- Aktualisierung der synthetischen Bevölkerung je nach Projektanforderungen:
 - Datenaktualität
 - Konsistenz mit anderen Eingangsdaten
 - spezifische räumliche Bezüge
 - spezifische Anforderungen an die Flotteninformationen
- Beschränkungen
 - Räumliche Auflösung der Daten relativ grob
 - Teilweise umfangreiche Harmonisierung notwendig
 - Fortschreibung und Prognose: Annahme Verhaltens- und Raumkonsistenz
 - SALSA: zahlreiche Attribute nicht vorhanden (Umzugshistorie, Zuzugsinformationen, Migrationshintergrund, Präferenzstruktur, ...)



Quo vadis?

- Stärkere Nutzung der Methodik zur Fortschreibung von Strukturdaten in den Modellen
 - Unternehmensstandortwahl (SALSA)
 - Zielwahl (TAPAS)
- Überarbeitung und Aktualisierung der Modelle für die Mobilitätsoptionen gemäß neuer Verhaltensdaten (MiD und SrV 2017, EVS)
- Verbesserung der Fortschreibung, eventuell mit Evolutionsmodellen

- Nutzung auch separat von den beiden originären Ziel-Modellen
 - Analysen zur Ladeinfrastruktur und Pkw-Besitz
 - Mikrosimulationsmodell städtischer Güterverkehr
 - ...



Heureka-Beitrag: elib.dlr.de/111905/

Rita Cyganski
Institut für Verkehrsforschung
Rita.Cyganski@dlr.de
+49 (30) 670 55 – 147
www.dlr.de/vf

