

Synthese von Haushaltsdaten für den Kanton Zürich

Simulationsprojekt "Erreichbarkeit, Infrastruktur und
Raumentwicklung"

Working Paper

Author(s):

Bürgle, Michaela

Publication date:

2007-01

Permanent link:

<https://doi.org/10.3929/ethz-a-005371344>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften 9

**Simulationsprojekt
„Erreichbarkeit, Infrastruktur und Raumentwicklung“**

Synthese von Haushaltsdaten für den Kanton Zürich

Michaela Bürgle

Arbeitsberichte Polyprojekt "Zukunft urbane Kulturlandschaften"

9

Januar 2007

**DARCH
DBAUG NSL**

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Dank

Dank geht an David Simmons und Olga Feldman, die mit ihren Ausführungen zur Haushaltssynthese im britischen Kontext die Anwendung von Simulated Annealing als Optimierungsmethode ins Gespräch brachten.

Inhaltsverzeichnis

1	Einleitung	3
2	Datengrundlagen	4
2.1	Verwendete Datensätze	4
2.2	Datenaufbereitung im Vorfeld der Synthese	7
3	Algorithmus zur Haushaltssynthese	9
3.1	Anforderungen an die synthetisierten Haushalte	9
3.2	Simulated Annealing.....	9
3.3	Simulated Annealing – Algorithmus zur Haushaltssynthese	10
3.4	Parameterwahl.....	11
3.5	Problematische Hektare	13
3.6	Imputation fehlender Attribute	15
3.7	Übertragung auf andere Regionen	15
4	Literatur.....	16

Tabellenverzeichnis

Tabelle 1	Volkszählung 1990: Für die Haushaltssynthese relevante Felder.....	4
Tabelle 2	PUS 1990: Für die Haushaltssynthese relevante Felder.....	5
Tabelle 3	Mikrozensus Verkehr 2000: Für die Haushaltssynthese relevante Felder ...	6
Tabelle 4	Für die Haushaltssynthese relevante räumliche Information.....	6
Tabelle 4	Parameterbelegungen zum Test des Optimierungsverfahrens	12

Abbildungsverzeichnis

Abbildung 1	Optimierungsalgorithmus zur Haushaltssynthese	11
Abbildung 2	Wert der Fehlerfunktion im Lauf der Iterationen.....	13

Arbeitsberichte Polyprojekt "Zukunft urbane Kulturlandschaften"

9

Synthese von Haushaltsdaten für den Kanton Zürich

Michaela Bürgle
IRL/IVT/NSL
ETH Hönggerberg
CH-8093 Zürich

Telefon: 044 633 45 43
Telefax: 044 633 11 02
buergle@nsl.ethz.ch

Januar 2007

Kurzfassung

Der vorliegende Arbeitsbericht beschreibt die Datengrundlagen und Schritte zur Erstellung einer Menge von synthetisierten Haushaltsdatensätzen für den Grossraum Zürich im Jahre 1990. Die erzeugten Daten finden im Rahmen des Projekts „Infrastruktur, Erreichbarkeit und Raumentwicklung“ Verwendung, das 2004 bis 2006 im Forschungsfeld „Zukunft urbane Kulturlandschaften“ am Netzwerk Stadt und Landschaft der ETH Zürich durchgeführt wird. Sie stellen eine der Inputgrössen für eine integrierte Landnutzungs-Verkehrssimulation im Grossraum Zürich dar.

Das Syntheseverfahren geht von auf Hektare aggregierten Haushaltsattributen aus, um Aussagen über Eigenschaften und Verortung einzelner Haushalte abzuleiten. Dabei wird das Optimierungsverfahren „Simulated Annealing“ eingesetzt. Die entwickelte Methode kann ohne Anpassungen für andere Regionen der Schweiz durchgeführt werden.

Schlagworte

Haushaltssynthese; Simulated Annealing; UrbanSim

Zitierungsvertrag

Bürgle, M. (2007) Synthese von Haushaltsdaten für den Kanton Zürich, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, 9, NSL, ETH Zürich, Zürich.

Title of the work

Michaela Bürgle
IRL/IVT/NSL
ETH Hönggerberg
CH-8093 Zürich

Telefon: 044 633 45 43
Telefax: 044 633 11 02
buergle@nsl.ethz.ch

January 2007

Abstract

This working paper describes data used and steps taken to create a set of synthetic households for the Greater Zurich area in the year 1990. The dataset thus created is used in the context of the project „Infrastructure, Accessibility and Spatial Development“ which is carried out as part of the research focus “Future of Urbanised Landscapes” of ETH Zurich’s Network City and Landscape from 2004 to 2006. The household data constitute one of the inputs for an integrated land use – transport simulation of the Greater Zurich area.

The synthesis method utilises household attributes aggregated on hectares and derives from them properties and locations of single households. This is accomplished using Simulated Annealing as an optimization technique. The method thus designed can be applied to other regions in Switzerland without any customisation.

Keywords

Household synthesis; Simulated Annealing; UrbanSim

Preferred citation style

Bürgle, M. (2007) Synthese von Haushaltsdaten für den Kanton Zürich, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, 9, NSL, ETH Zürich, Zürich.

1 Einleitung

Im Rahmen des Projekts „Infrastruktur, Erreichbarkeit und Raumentwicklung“ wird auf Hektarebene die Entwicklung von Landnutzung und Verkehr im Grossraum Zürich über mehrere Jahre hinweg simuliert (vgl. Bürgle, Löchl und Waldner, 2005). Eine der für das Basisjahr der Simulation vorzugebenden Eingabegrössen ist eine Sammlung detaillierter Informationen über jeden einzelnen Haushalt im Simulationsgebiet. Diese Informationen umfassen Haushaltsgrösse, Alter des Haushaltsvorstands, Haushaltseinkommen und das Vorhandensein von Berufstätigen oder Kindern, die Verfügbarkeit von Personenwagen und ausserdem die Verortung im Hektarraster.

Da zum Zeitpunkt der im vorliegenden Bericht beschriebenen Synthesearbeiten (2005) reale Datensätze nur für hektarweise aggregierte Einzelmerkmale von Haushalten verfügbar waren, nicht aber in der korrelierten Form, die für die Simulation vorausgesetzt wird, wurden die benötigten Haushaltsdatensätze aus den zur Verfügung stehenden Datensätzen synthetisiert. Dazu werden zunächst alle Haushalte im betrachteten Simulationsgebiet aus der Volkszählung 1990 und dem Public Use Sample (PUS) 1990 mit Hilfe des Optimierungsverfahrens „Simulated Annealing“ erzeugt. Da sich aus der Volkszählung nicht alle für die Simulation benötigten Attribute entnehmen lassen, werden in einem zweiten Schritt die fehlenden Attribute Einkommen und Autobesitz imputiert.

Im vorliegenden Arbeitsbericht ist das Vorgehen zur Erzeugung von Haushaltsdatensätzen für das Jahr 1990 für den Teil des Simulationsgebiets beschrieben, der im Kanton Zürich liegt. Die Übertragung auf andere Kantone erfordert keine Anpassung des Verfahrens.

Als Basisjahr für die Simulation sollte zunächst 1990 verwendet werden. Da sich später aufgrund der Zeitstände anderer Datensätze das Basisjahr auf 1996 änderte und zusätzlich personenscharfe Daten aus der Volkszählung 2000 verfügbar wurden, kam am Ende ein einfacheres Verfahren zur Erzeugung von Haushalten zum Einsatz, in dem die für 1990 erzeugten Haushalte kaum eine Rolle spielen. Der Java-Quellcode für die Durchführung der hier beschriebenen Haushaltssynthese ist im CVS des IVT archiviert (ivt/src/hhGenerator). Alle in diesem Programm verwendeten Datenbankabfragen sind in der Quellcodedatei `SQLdefinitions.java` gesammelt. Für den Einsatz der TestShell muss zusätzlich JfreeChart (Object Refinery Limited, 2007) installiert werden.

2 Datengrundlagen

2.1 Verwendete Datensätze

Zur Synthese von Haushaltsdaten wurden vor allem drei Datensätze verwendet: die Volkszählung 1990 des Schweizer Bundesamts für Statistik (BfS), das zugehörige Public Use Sample (PUS) 1990 des BfS und der Mikrozensus zum Verkehrsverhalten 2000 (Bundesamt für Raumentwicklung - ARE und BfS). Zusätzlich wird die Verfügbarkeit einer Zuordnung von Hektarzellen zu Gemeinden vorausgesetzt.

2.1.1 Volkszählung 1990

In Tabelle 1 sind die für die Haushaltssynthese relevanten Attribute der Volkszählung 1990 aufgelistet und erläutert. Bezugsgrösse ist jeweils ein Hektar.

Die Volkszählungsdaten 1990 lagen für die Haushaltssynthese in der nicht klassierten Version vor (absolute Werte von 1 bis 3 sind getrennt ausgewiesen, nicht wie in der allgemein verfügbaren Version zu einer Klasse mit Wert 3 zusammengefasst).

Tabelle 1 Volkszählung 1990: Für die Haushaltssynthese relevante Felder

Feldname	Beschreibung
x	x-Koordinate des Hektars
y	y-Koordinate des Hektars
p9btot	Einwohnerzahl
p9bm[01-19]	Zahl männlicher Einwohner der Altersgruppe [1-19]*
p9bw [01-19]	Zahl weiblicher Einwohner der Altersgruppe [1-19]*
p9emt	Zahl der männlichen Erwerbstätigen
p9ewt	Zahl der weiblichen Erwerbstätigen

h9tot	Gesamtzahl der Haushalte
h9ptot	Zahl der Privathaushalte
h9p0[1-6]	Zahl der Privathaushalte mit [1,... 6 oder mehr] Haushaltsmitgliedern

* Altersgruppen: 1 = 0-4 jährig, 2 = 5-9-jährig, 3 = 10-14 jährig, 4 = 15-19 jährig, 5 = 20-24 jährig, 6 = 25-29 jährig, 7 = 30-34 jährig, 8 = 35-39 jährig, 9 = 40-44 jährig, 10 = 45-49 jährig, 11=50-54 jährig, 12 = 55-59 jährig, 13 = 60-64 jährig, 14 = 65-69 jährig, 15 = 70-74 jährig, 16 = 75-79 jährig, 17 = 80-84 jährig, 18 = 85-89 jährig, 19 = 90 jährig oder älter

2.1.2 PUS 1990

Ergänzend zur Volkszählung 1990 hat das Schweizer BfS eine 5%-Stichprobe von Personendatensätzen zu Einzelpersonen mit einer Reihe von Attributen veröffentlicht. Aus Datenschutzgründen fehlt hier jedoch der genaue Wohnort, die Verortung ist nur kantonsscharf. Tabelle 2 zeigt, welche Felder des PUS in der Haushaltssynthese Eingang fanden. Die Referenzperson der PUS-Datensätze wurde mangels genauerer Information jeweils als Haushaltsvorstand aufgefasst.

Tabelle 2 PUS 1990: Für die Haushaltssynthese relevante Felder

Feldname	Beschreibung
kant	Kanton (nur Datensätze mit kant=ZH wurden verwendet)
aper	Anzahl Personen im Haushalt (1, 2, ...,6 für 1-6 Personen, 7 für 7-25 Personen, 8 für 26 Personen oder mehr)
raltj	Alter der Referenzperson im Haushalt*
sth2	Anzahl (Stief-)Kinder im Haushalt (bis 19 Jahre)
etto	Zahl der Erwerbstätigen im Haushalt

* Altersgruppen: 1 = 15-29 jährig, 2 = 30-44 jährig, 3 = 45-64 jährig, 4 = 65 jährig oder älter

2.1.3 Mikrozensus zum Verkehrsverhalten 2000

Der Mikrozensus 2000 wurde als Informationsquelle zur Verteilung von Haushaltseinkommen und Personenwagenbesitz verwendet. Da das Haushaltseinkommen im Mikrozensus 1989 noch nicht enthalten war, musste die Daten für das Jahr 2000 verwendet werden, unter der Annahme, dass der statistische Zusammenhang zwischen Haushaltsgrösse, Einkommen und Personenwagenbesitz unverändert übernommen werden können. Tabelle 4 listet die verwendeten Datenfelder auf:

Tabelle 3 Mikrozensus Verkehr 2000: Für die Haushaltssynthese relevante Felder

Feldname	Beschreibung
hhanzper	Haushaltsgrösse (1 – 15 Personen, für die Synthese aggregiert zu sechs Klassen wie in Volkszählung 1990)
ein	Monatliches Haushaltseinkommen in CHF (acht Klassen in 2000er Schritten von < 2000 CHF bis \geq 14.000 CHF)
f30100	Anzahl Autos im Haushalt (aggregiert zu zwei Klassen 0 und 1)

2.1.4 Hektarraster und Gemeindenummern

Die Haushalte werden gemeindeweise erzeugt, um den Syntheseablauf besser kontrollieren zu können und ein besseres Laufzeitverhalten zu erzielen. Tabelle 4 zeigt die dafür benötigten Informationen zum Hektarraster. Es wird im folgenden davon ausgegangen, dass die hier beschriebene Tabelle nur Hektare beinhaltet, für die Haushalte erzeugt werden sollen.

Tabelle 4 Für die Haushaltssynthese relevante räumliche Information

Feldname	Beschreibung
XY	Eindeutiger Bezeichner für die Hektarzelle (Wert entspricht $VZ90.X * 100 + VZ90.Y / 100$)
City_ID	Eindeutiger Bezeichner für die Gemeinde, in der die Zelle liegt (BfS-Gemeindenummer)

2.2 Datenaufbereitung im Vorfeld der Synthese

Das Programm zur Haushaltssynthese wurde in Java erstellt. Die unter 2.1 beschriebenen, für die Synthese verwendeten Daten wurden in eine MS-ACCESS-Datenbank importiert, um über JDBC (Sun microsystems, 2007) und eine ODBC-Datenquelle auf sie zugreifen zu können. Beim Export wurden die Feldnamen wie in 2.1 beschrieben vergeben. Da Datenbankzugriffe den Engpass des Syntheseprogramms darstellen, wurden den aus BfS-Daten erzeugten Tabellen einige zusätzliche Felder hinzugefügt. Die zusätzlichen Werte werden aus vorhandenen Attributen berechnet, um dies nicht zur Laufzeit des Programms durchführen zu müssen. Die Schritte zur Berechnung der zusätzlichen Werte werden in den folgenden Abschnitten beschrieben.

Ausserdem wurden Datensätze gelöscht, deren Koordinaten nicht im Simulationsgebiet liegen, auch dies eine Massnahme zur Beschleunigung der Arbeit mit der Datenbank.

2.2.1 Aufbereitung der Volkszählungsdaten

Zunächst wird die gesamte Volkszählungstabelle (hier benannt mit „vz90“) auf die Hektardatensätze reduziert, deren XY-Koordinaten im Simulationsgebiet liegen. Die Hektarinformationen liegen in einer Tabelle „grid“ vor. Als Ergebnis entsteht die Tabelle „vz90_simarea“:

```
SELECT vz90.* INTO vz90_simarea
FROM grid INNER JOIN vz90 ON grid.XY=vz90.xy;
```

Anschliessend werden alle Datensätze mit P9BTOT=0 gelöscht, da sie unbewohnte Hektare beschreiben:

```
DELETE * FROM vz90_simarea WHERE p9btot=0;
```

Für die Zahl der Berufstätigen und die Zahl der Kinder pro Hektar wird jeweils ein zusätzliches Feld eingeführt (p9etot bzw. p9ktot), dessen Wert sich als Summe der vorhandenen Einzelangaben ergibt. Als Kinder zählen Personen bis 19 Jahre, entsprechend der Einteilung im PUS.

```
UPDATE vz90_simarea SET p9etot = p9emt0t + p9ewt0t;
UPDATE vz90_simarea SET p9ktot = p9bm01 + p9bm02 + p9bm03 + p9bm04 +
p9bw01 + p9bw02 + p9bw03+ p9bw04;
```

2.2.2 Aufbereitung des PUS

Aus dem PUS 1990 wurden nur die Datensätze für den Kanton Zürich verwendet, entsprechend der Abgrenzung des Simulationsgebiets. Es wurden keine Umkodierungen vorgenommen. Die im Vergleich zur Volkszählung feinere Klassierung der Haushaltsgrössen wird im Verlauf der Synthese ausgeglichen.

Datensätze, bei denen Angaben zur Zahl der Kinder und/oder Berufstätigen im Haushalt, dem Alter der Referenzperson oder der Haushaltsgrösse fehlen, werden nicht bei der Haushaltssynthese berücksichtigt. Eine Überprüfung ergab, dass diese Auslassungen nicht zu systematischen Verzerrungen führen.

3 Algorithmus zur Haushaltsynthese

3.1 Anforderungen an die synthetisierten Haushalte

Ziel der Haushaltssynthese ist die Erzeugung von Haushalten mit den Attributen

- Haushaltsgrösse; Klassen 1, 2, 3 (1, 2 bzw. 3 oder mehr Personen)
- Alter des Haushaltsvorstands; Klassen 1, 2, 3, 4 (Altersgruppen 15-29, 30-44, 45-64 bzw. 65 oder älter)
- Zahl der Berufstätigen im Haushalt; Klassen 0, 1 (keine Berufstätigen bzw. Berufstätige vorhanden)
- Zahl der Kinder im Haushalt; Klassen 0, 1 (keine Kinder bzw. Kinder vorhanden)
- Personenwagen-Verfügbarkeit; Klassen 0, 1 (kein PW bzw. PW vorhanden)
- Haushaltseinkommen (in 2000-Franken-Schritten)
- Verortung des Haushalts im Hektarraster (eindeutige Hektar ID gemäss „gridcells“-Tabelle der Basisjahrdaten für UrbanSim)

Die Auswahl der Attribute ergibt sich aus der Spezifikation von UrbanSim (Waddell et al. 2007).

3.2 Simulated Annealing

Da es sich bei der Erzeugung von Haushalten unter Beachtung vorgegebener Constraints in Bezug auf die Haushaltsattribute um ein komplexes kombinatorisches Problem handelt, für das angesichts der zu erzeugenden Menge von Haushalten (ca. 500.000 für das Simulationsgebiet) eine erschöpfende Lösungssuche nicht in Frage kommt, wurden heuristische Herangehensweisen herangezogen. Eine vergleichende Betrachtung von Hill-Climbing, Simulated Annealing und Genetischen Algorithmen zur Lösung einer ähnlichen Aufgabenstellung im britischen Kontext (Williamson, Birkin und Rees, 1998) ergab, dass Simulated Annealing von diesen drei Herangehensweisen die besten Resultate lieferte. Zur Erzeugung von Haushalten im Grossraum Zürich wurde deshalb ein Simulated Annealing Verfahren gewählt.

Simulated Annealing ist ein Optimierungsverfahren, das in Analogie zu metallurgischen Verfahren des kontrollierten Abkühlens von Materialien entwickelt wurde. Der Algorithmus basiert auf dem Prinzip des Hill-Climbing, indem er eine zufällig gewählte Startlösung schrittweise optimiert. Er enthält aber eine probabilistische Komponente, die unter bestimmten Umständen schlechtere Lösungen akzeptiert, um lokale Optima verlassen zu können. Diese probabilistische Komponente wird durch den sogenannten „Temperaturparameter“ bestimmt. Der Temperaturparameter wird zu Beginn der Optimierung relativ hoch gewählt (Analogie zu höherer Bewegungsenergie von Atomen bei hohen Temperaturen) und mit dem Fortschreiten der Berechnungen schrittweise reduziert („Abkühlung“). Das Akzeptieren von schlechteren als der aktuellen Lösung wird so mit dem Fortschreiten der Lösungssuche immer unwahrscheinlicher.

Die Verfügbarkeit von Personenwagen und das Haushaltseinkommen können für die einzelnen erzeugten Haushalte nicht in diesem Syntheseschritt bestimmt werden, da die für das Optimierungsverfahren einsetzbaren Daten (Volkszählung 1990, PUS 1990) keine entsprechenden Angaben enthalten.

3.3 Simulated Annealing – Algorithmus zur Haushaltssynthese

Das hier angewendete Simulated Annealing-Verfahren zur Erzeugung von Haushalten geht von einer Stichprobe von Haushaltsdatensätzen als Grundgesamtheit aus, aus der als Startlösung eine Zufallsauswahl von Haushalten für die betrachtete räumliche Einheit gezogen wird. Diese Startlösung wird so gewählt, dass die Vorgaben der Volkszählung in Bezug auf die Zahl von Privathaushalten einer bestimmten Grösse erfüllt sind. Bis zum Erreichen eines geforderten Übereinstimmungsgrads der zufälligen Haushaltsauswahl mit den vorgegebenen statistischen Werten werden immer wieder einzelne Haushalte der Auswahl durch andere aus der Grundgesamtheit gezogene mit gleicher Haushaltsgrösse ersetzt. Jeweils vor und nach dieser Ersetzung wird die Qualität der Auswahl durch Vergleich der Zahl Kinder und Erwerbstätigen mit den Vorgaben der Volkszählung ermittelt. Ergibt sich ein kleinerer Fehler, wird die neue Lösung beibehalten. Vergrössert sich der Fehler durch den Tausch, wird anhand von Formel (1) (vgl. Williamson, Birkin und Rees, 1998, S. 797) über die Akzeptanz dieser schlechteren Lösung entschieden.

$$p(\Delta E) = \exp(-\Delta E/T) \quad (1)$$

ΔE stellt dabei die Verschlechterung durch Übernahme der neuen Lösung dar, T ist der oben vorgestellte Temperaturparameter. Abbildung 1 zeigt die Grundidee des im Programm verwendeten Algorithmus.

Abbildung 1 Optimierungsalgorithmus zur Haushaltssynthese

```

reduce := number of swaps at each temperature
anneal := rate of temperature decline
tl := lower temperature bound:
for each hectare xy do
- temperature parameter  $T := t_{xy}$ , swap-counter swaps := 0
- get a random initial household selection hsample from PUS that contains as many
  households of each size as stated in census for xy
- add up all workers and children of hsample and calculate the squared difference from
  census information for xy as  $e_1$ 
- do
  · if  $e_1$  is lower than a set quality measure  $q$ : break
  · else if  $T < t_l$ : break
  · else if maximum number of loops has been exceeded: break
  · else:
     $e_1 :=$  error of hsample
    hsample' := hsample with one household replaced by a randomly drawn one
     $e_2 :=$  error of hsample'
    if  $e_1 > e_2$ : hsample := hsample', swaps := swaps + 1
    else:
      - rd := random number  $\geq 0.0$  and  $< 1.0$  drawn from a uniform distribution
      - if  $rd < \exp((e_1 - e_2) / t)$ : hsample := hsample', swaps := swaps + 1
    fi
    if swaps  $\geq$  reduce:  $T := T - \textit{anneal} * T$ , swaps = 0 fi
  · fi
- while (true)
- add the households in hsample to database
end for each

```

3.4 Parameterwahl

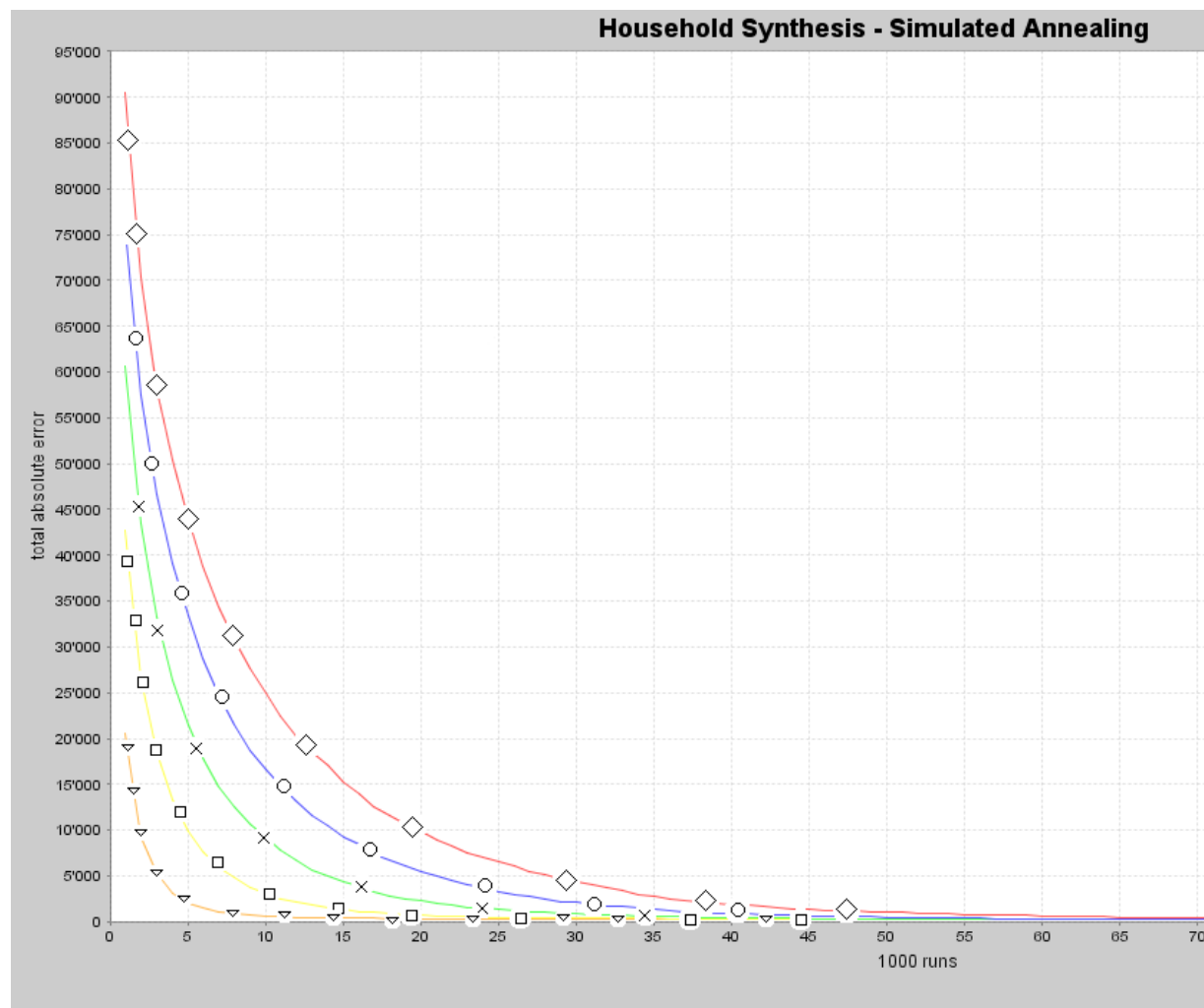
Die Wahl der Parameter Starttemperatur, Temperaturreduktion und Dauer des Verbleibs auf einer Temperaturstufe hat Einfluss auf die Qualität der Ergebnisse. Dieser Einfluss wurde in Testläufen systematisch untersucht. Analog zu den in Williamson, Birkin und Rees (1998) beschriebenen Versuchen wurden jeweils fünf Synthesedurchläufe für verschiedene Parameterbelegungen durchgeführt. Die getesteten Einstellungen zeigt Tabelle 5.

Tabelle 5 Parameterbelegungen zum Test des Optimierungsverfahrens

Konfiguration	Starttemperatur	Temperaturreduktion	Temperaturwechsel
1	50	5	5
2	40	5	4
3	30	5	3
4	20	5	2
5	10	5	1

Abbildung 2 zeigt den Verlauf der Fehlerkurven in Abhängigkeit von der Zahl der Optimierungsiterationen. Für jede Parameterbelegung wurden Mittelwerte der 5 Testläufe gebildet, um dem zufallsgesteuerten Charakter des Verfahrens Rechnung zu tragen. Wie die Abbildung zeigt, ist für das vorliegende Problem nach ca. 100.000 Iterationen kaum mehr eine Verbesserung der Lösung zu erwarten. In diesem Bereich unterscheiden sich die verschiedenen Varianten auch nicht mehr stark. Will man mit wenig Iterationen eine relativ gute Lösung erhalten, bieten sich die durch niedrige Starttemperatur und schnelle Temperatursenkung gekennzeichneten Parameterbelegungen an, ein Ergebnis, zu dem auch Williamson, Birkin und Rees kamen. Dass Varianten, die nahe an reinen Hill-Climbing-Verfahren liegen, die besten Ergebnisse produzieren, kann vielleicht darauf zurückgeführt werden, dass die Fehlerfunktion sehr einfach ist und nur von wenigen Variablen abhängt.

Abbildung 2 Wert der Fehlerfunktion im Lauf der Iterationen



Konfiguration 1: —◇—

Konfiguration 2: —○—

Konfiguration 3: —×—

Konfiguration 4: —□—

Konfiguration 5: —▽—

3.5 Problematische Hektare

Bestimmte Konstellationen erschweren das Erzielen eines guten Ergebnisses für einzelne Hektare. In diesen Fällen muss mit zusätzlichen Annahmen gearbeitet werden, was wahrscheinlich zu einer Verzerrung der Ergebnisse in jenen Hektaren führt.

3.5.1 Kollektivhaushalte

Eine schwer zu bestimmende Grösse ist die Belegung eines Hektars mit Kollektivhaushalten (= alle nicht privaten Haushalte). Der Volkszählung kann zwar die Zahl der Kollektivhaushalte je Hektar entnommen werden, nicht aber deren jeweilige Mitgliederzahl oder die Zahl der in diesen Haushalten lebenden Erwerbstätigen und Kinder. Da es sich bei Kollektivhaushalten meist um relativ grosse Gebilde handelt (z.B. Anstalten, Heime), ist für die betroffenen Hektare die Zuordnung von Kindern und Berufstätigen zu den Privathaushalten mit grossen Unsicherheiten belastet, die sich nicht anhand der Volkszählungsdaten auflösen lassen. Die initiale (Privat-)Haushaltsauswahl für solche Hektare wird fast zwangsläufig ein grosses Defizit an Kindern und/ oder Berufstätigen aufweisen. Um Verzerrungen zu vermeiden, wird beim Vorhandensein eines Kollektivhaushalts in einer Hektarzelle ein gewisser Prozentsatz (aktuell 80%) der in der initialen Haushaltsauswahl fehlenden Berufstätigen und Kinder dem Kollektivhaushalt zugerechnet. Nur die dann noch fehlenden 20% gehen als Zielwert in die Optimierung ein. Wird dieser Zielwert mit Abschluss des Optimierungsverfahrens nicht erreicht, rechnet man auch die dann noch „fehlenden“ Berufstätigen oder Kinder dem Kollektivhaushalt zu.

Gibt es auf einem Hektar mehrere Kollektivhaushalte, bleibt das Verfahren unverändert. Da Kollektivhaushalte in der angestrebten Landnutzungssimulation zur Zeit nicht berücksichtigt werden stellt diese Vorgehensweise kein Problem dar. Sollten zukünftig auch Modelle für Kollektivhaushalte geschätzt werden, muss die Verteilung von Personen auf Kollektivhaushalte neu überdacht werden.

3.5.2 Grosse Privathaushalte

Ein weniger gewichtiger Problemfall sind grosse Privathaushalte, die in der Volkszählung unter der Rubrik „6 Personen oder mehr“ zusammengefasst sind. Für Hektare, die insgesamt wenig Einwohner aber trotzdem einen hohen Fehler haben, kann man teilweise rechnerisch bestimmen, dass tatsächlich ein oder mehrere Haushalte mit deutlich mehr als sechs Mitgliedern vorhanden sind (z.B. in einem Fall ein 24-Personen-Haushalt). Datensätze mit solchen Haushaltgrössen gibt es aber im PUS nicht.

3.6 Imputation fehlender Attribute

Die Attribute Autoverfügbarkeit und Einkommen lassen sich nicht mit dem oben beschriebenen Verfahren erzeugen, da keine derartigen Angaben in zur Verfügung stehenden Datensätzen enthalten sind. Diese fehlenden Informationen wurden wie folgt imputiert: Der Mikrozensus Verkehr Schweiz 2000 enthält Angaben zur Haushaltsgrösse, zum Bruttoeinkommen des Haushalts (in 2000 Franken Schritten) und zur Zahl der Autos im Haushalt. Daraus wurde zunächst Anteile der acht Einkommensklassen für die verschiedenen Grössenklassen ermittelt, dann Anteile von Haushalten mit Personenwagen nach Haushaltsgrösse und Einkommen. Diese Verteilungen wurden als Wahrscheinlichkeiten für Konstellationen aus Haushaltsgrösse, -einkommen und Autobesitz auf die synthetisch erzeugten Haushalte angewendet. Dieses Verfahren ist nicht unkritisch zu sehen, da die Fallzahlen im Mikrozensus für manche Kombinationen sehr gering sind. Eine bessere Datenquelle stand jedoch nicht zur Verfügung.

3.7 Übertragung auf andere Regionen

Eine Übertragung des bisher beschriebenen Verfahrens auf andere Regionen ist problemlos möglich. Die Synthese muss kantonsweise erfolgen, um die räumlichen Angaben des PUS, das Datensätze kantonsgenau zuordnet, möglichst gut auszunutzen. Es wird jeweils eine Auflistung der mit Haushalten zu füllenden Hektare und deren Gemeindezugehörigkeit benötigt wie unter 2.1.3 beschrieben.

Es empfiehlt sich bei der Anwendung auf andere Kantone eventuell, die unter 3.4 vorgestellte Parameterwahl noch einmal einem Vergleichstest zu unterziehen. Diese Auswahl wurde bisher nur für den Kanton Zürich validiert. Da die Rechenzeiten der Haushaltssynthese für Gebiete der Grössenordnung von Gemeinden sehr kurz sind und die Testläufe automatisiert werden können, stellt dies keinen grossen Zusatzaufwand dar.

4 Literatur

Bürgle, M., M. Löchl und U. Waldner (2005) Entwicklung eines Simulationsmodells – Infrastruktur, Erreichbarkeit und Raumentwicklung, *DISP*, **160**, 94—95.

Object Refinery Limited (2007) JfreeChart. <http://www.jfree.org/jfreechart/> 17. Jan 2007

Sun microsystems (2007) JDBC. <http://java.sun.com/products/jdbc/>. 17. Jan 2007

Waddell et.al. (2007) UrbanSim: Dokumentation der „households“-Tabelle
http://www.urbansim.org/docs/data/database_tables/household_tables.xml#Households.
17. Jan 2007

Williamson P., M Birkin und P. H. Rees (1998) The estimation of population microdata by using data from small area statistics and samples of anonymised records, *Environment and Planning, Part A*, **30**, 785—816.