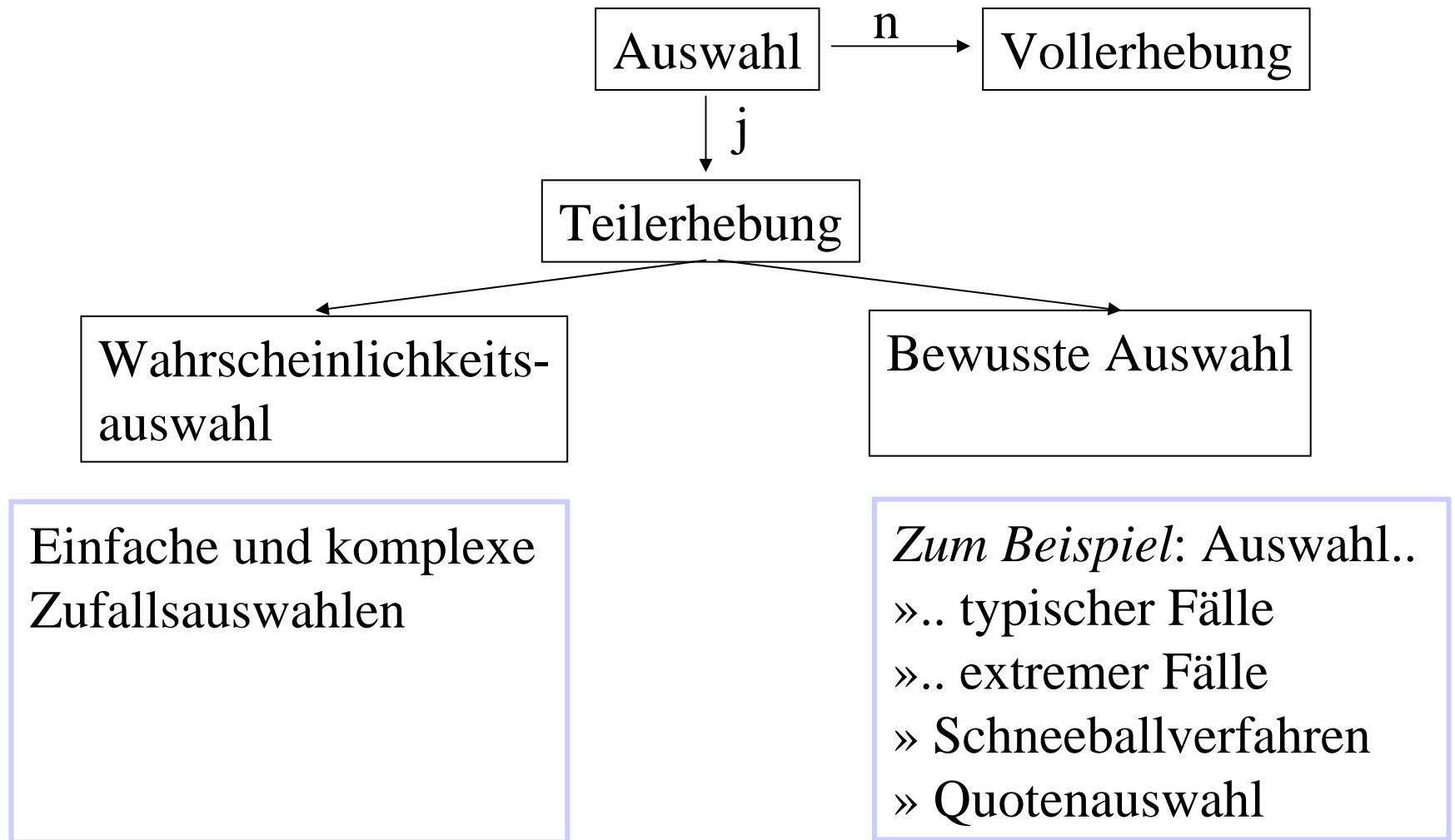
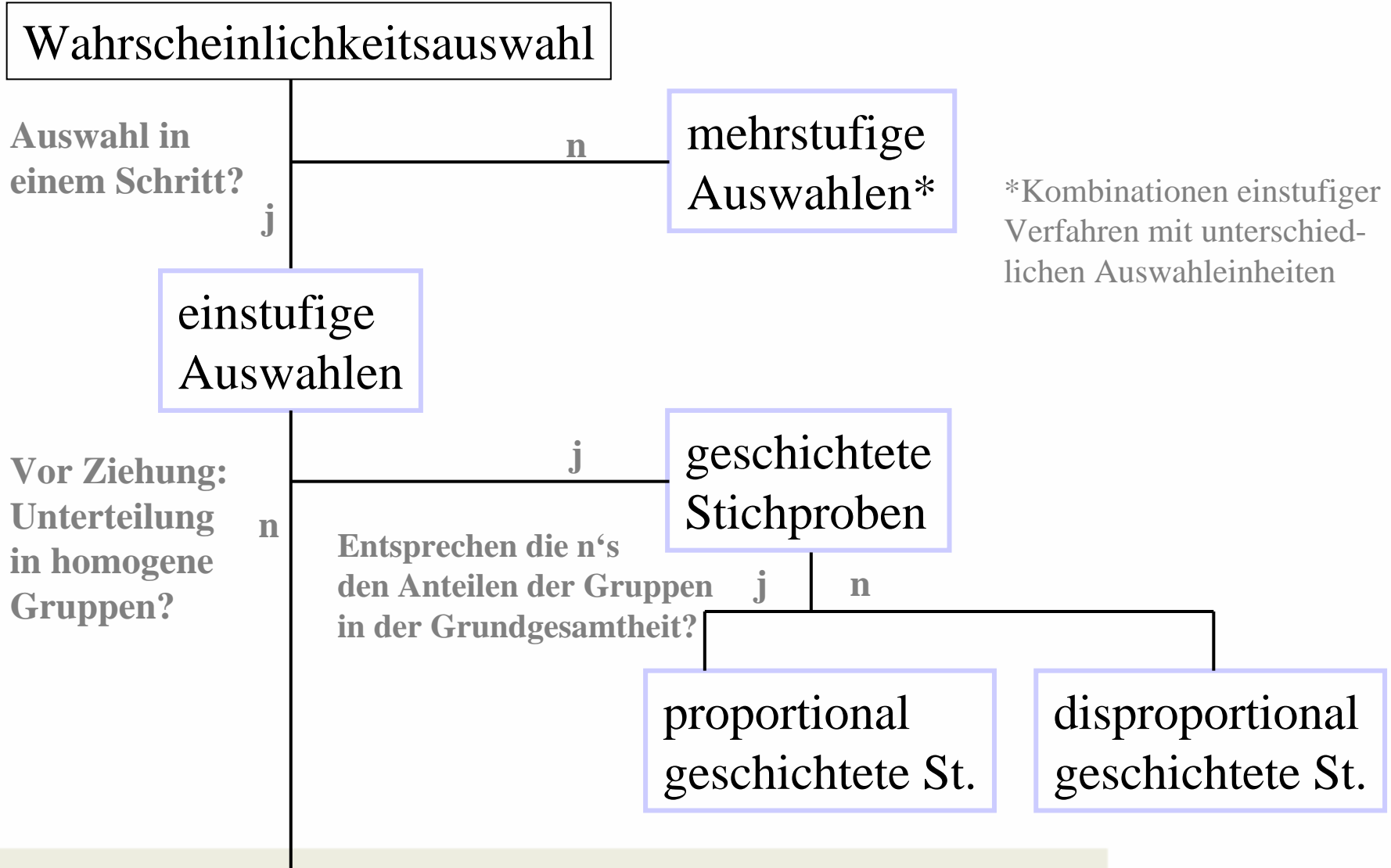


Komponenten des Survey Fehlers

<i>Sampling</i>
<i>Non-Sampling</i>
Frame
Nonresponse
Messung
Fragebogendesign
Modus Datenerhebung
Antwortprozess
Interviewereffekte

Data Processing
Eingangsprüfung
Dateneingabe
Editierung
Kodierung
File Preparation
Gewichtung
Imputation





[Fortsetzung]

Unterteilung
in räumliche
Einheiten und
jeweils
Auswahl aller
Elemente der
gezogenen
Einheiten?

j

Klumpen-
stichprobe

n

Einfache
Zufallsstichprobe

Beispiel (Klumpenauswahl)*

Grundgesamtheit: **$N=39.800$** Haushalte
mit einem Zeitungsabonnement

Jeweils 10 auf der Auslieferungsroute hintereinander liegende Haushalte werden einem Klumpen zugerechnet:

$M=3.980$ Klumpen; **$N_j=10$** für alle Klumpen

Es werden **$m = 40$** Klumpen à 10 Haushalte gezogen, so dass **$n=400$** Haushalte

Einfache Zufallsauswahl (SRS)

Anzahl T möglicher Samples von **n** aus **N** Elementen

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

(Ohne Berücksichtigung der
Ziehungsreihenfolge,
Ohne Zurücklegen)

$$P(\text{Sample}) = \frac{1}{\binom{N}{n}}$$

Beispiel: Population $N=25$; Sample $n=5$

$$T = \frac{25!}{5!(25-5)!} = \frac{25 \times 24 \times \dots \times 1}{5 \times 4 \times 3 \times 2 \times 1 \times 20 \times 19 \times \dots \times 1} = 53130$$

- Lotterieprinzip
- Zufallszahlentafel; Zufallszahlengenerator

Ausschnitt aus einer Zufallszahlentafel*

	(1)	(2)	(3)	(4)	(5)	(6)	...	(14)
1	10480	05011	01536	02011	81647	91646	...	02338
2	22368	46573	25595	85393	30995	89198	...	
3	24130	48390	22527	97265	...			
.								
.								
.								
.								
15	07119	97336	71048	...				
.								
.								
100	38534	01715	94964	...				56947

Systematische Auswahl

- Auswahl jedes k -ten Elements, nach Zufallsstart
- Ermöglicht Sampling auf laufender Basis
ohne vorheriges Listing der Populationselemente
- Kann der echten Zufallsauswahl gleichgesetzt werden:
Wenn eine „zufällige angeordnete (durchmischte)
Auswahlgrundlage vor(liegt), dann sind keine Abweichungen
von der echten Zufallsauswahl zu erwarten“*

*Sturm, Manfred; Thomas Vajna (1974) Planung und Durchführung von Zufallsstichproben. S. 40-80 in Techniken der empirischen Sozialforschung. Band 6 Statistische Forschungsstrategien. München: Oldenbourg, S. 65

ADM – Stichproben – System

- » Stufe 1: Auswahl von Wahlbezirken
- » Stufe 2: Auswahl der Privathaushalte
 - „Address-Random“
(Begehung + Befragung getrennt)
 - „Random-Route“ bzw. „Random-Walk“
(keine Trennung)
- » Stufe 3: Auswahl der Zielpersonen
 - „Schwedenschlüssel“,
 - „Next-“ oder „Last-Birthday-Methode“

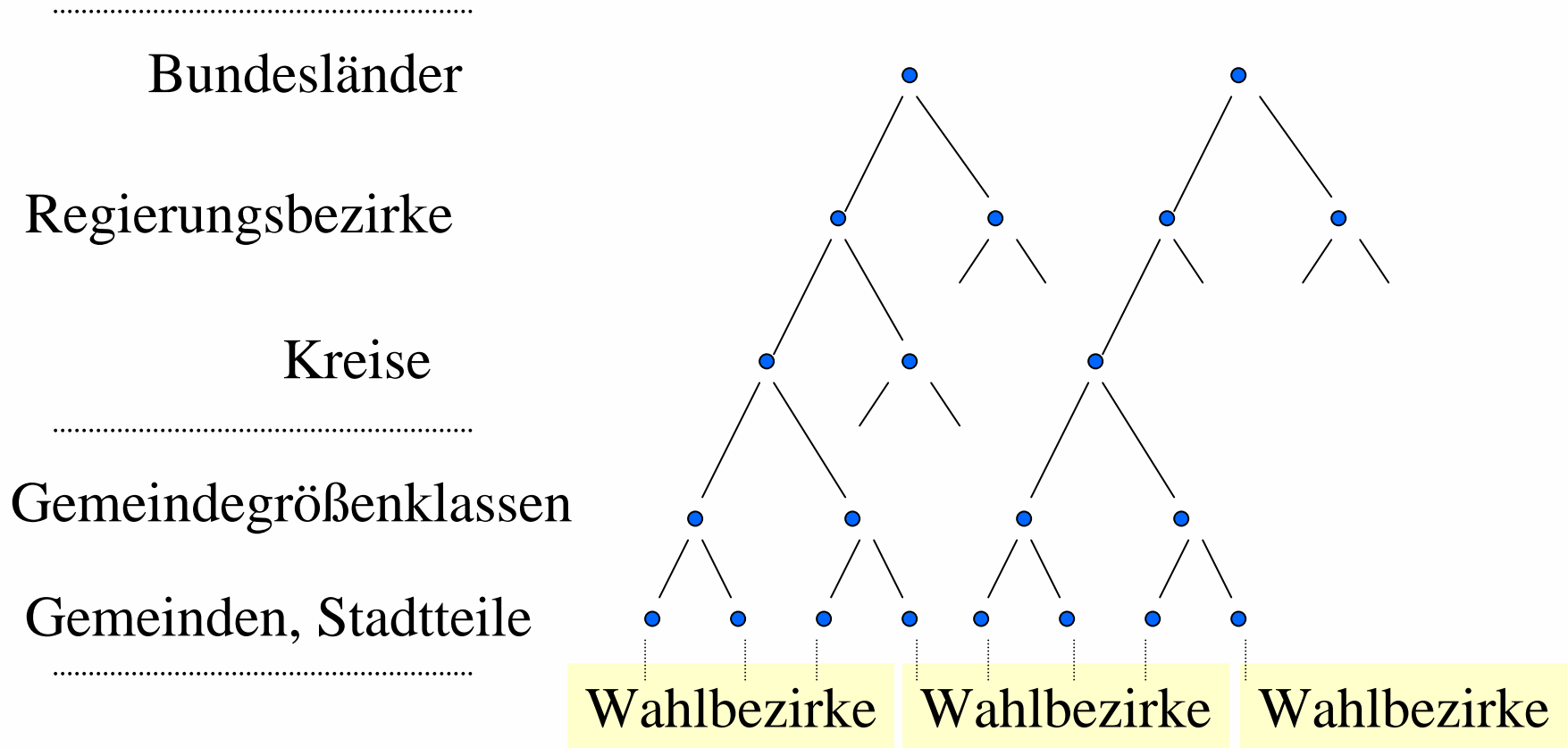
Geschichtete Auswahl der Wahlbezirke nach

Gemeindetyp innerhalb eines **Regionaltyps**

Regionaltyp:

Kreise innerhalb von Regierungsbezirken
innerhalb von Bundesländern

Strukturierte Anordnung zur Ziehung der Wahlbezirke (Sample Points)



ADM – Design

Primäreinheiten (PSUs primary sampling units; sample points)	Westen Ca.	Osten Ca.
Wahlbezirke (n=80.024)	60.000	20.000
zusammengefasst zu „synthetischen“ Stimmbezirken mit jeweils mind. 400 Wahlberechtigten (es verblieben 61.904 reale Stimmbezirke, die übrigen wurden zu 6.765 synthetisierten Stimmbezirken zusammengefaßt)	50.000**	14.000**
128 Netze à Sample Points	128 × 210	128 × [48 (+48)]
„ADM-Mastersample“ Anzahl der Sample Points	26.880	12.288

Anteile der Tabs.		Wenn Anzahl der Erwachsenen im Haushalt:					
		1	2	3	4	5	6+
	Tab.	Wähle Erwachsenen mit Nummer					
1/6	A	1	1	1	1	1	1
1/12	B1	1	1	1	1	2	2
1/12	B2	1	1	1	2	2	2
1/6	C	1	1	2	2	3	3
1/6	D	1	2	2	3	4	4
1/12	E1	1	2	3	3	3	5
1/12	E2	1	2	3	4	5	5
1/6	F	1	2	3	4	5	6

„Kish selection grid“ („Schwedenschlüssel“)

Nach: Kish, Leslie (1965) Survey Sampling. New York: Wiley,
p. 399

Stichprobenziehung für **telefonische Erhebungen**

Telefondichte der Haushalte im Jahr 2001:

94%

Anzahl der Festnetzanschlüsse Ende 1999:

47,8 Mio

Ins Telefonbuch eingetragene Anschlüsse:

75%

» **Random Digit Dialing (RDD)**

Einfache Zufallsziffernanwahl

z.B.: Innerhalb eines Ortsnetzbereiches

06321

2001 (kleinste vergebene Nr.)

6790152 (größte vergebene Nr.)

21000 Nummern veröffentlicht

Trefferquote, auf einen eingetragenen

Anschluss zu treffen, unter einem 1%.

» Randomize Last Digits (RLD)

Ziehung der Nummern aus dem Telefonbuch
und Ersetzung ihrer letzten beiden Stellen durch
zufällig erzeugte Ziffern ersetzt

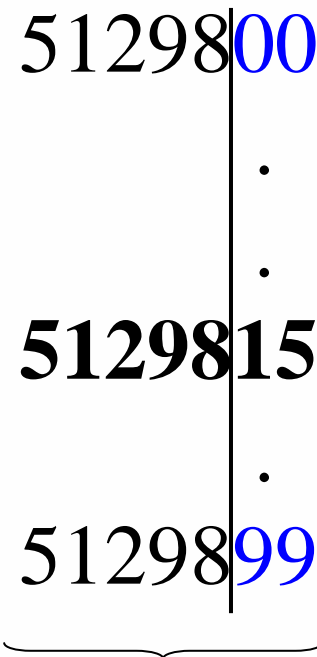
Problem: Inklusionswahrscheinlichkeiten nicht
gleich, sondern davon abhängig, wieviele Nummern
im betreffenden „100er Block“ eingetragen sind

100er Block

Stamm einer Telefonnummer, der nach
Abschneiden der beiden letzten Stellen verbleibt

Block **51298**_{xx}

51298|00
.
.
51298|15
.
51298|99



RLD

Zum Ausgleich ungleicher Inklusions-
wahrscheinlichkeiten:

Ex post Gewichtung der Datensätze mit der
Inversen der Blockbesetzung (Zahl
eingetragener Nummern im 100er Block)

z.B.:

Bei 40 eingetragenen Nr.: $\times 1/40$

Bei 2 eingetragenen Nr.: $\times 1/2$

Gabler/Häder – Design

Bildung der 100er Blöcke auf der Basis der eingetragenen Rufnummern

z.B.:

040 559 61 00	eingetragen
040 559 61 01	generiert
040 559 61 02	generiert
040 559 61 03	eingetragen
...	
040 559 61 99	generiert

Blockbesetzungen (Eintragsdichte):
Schwerpunkt zw. 70 und 80
eingetragenen Anschlüssen pro Block

Durchschnitt: 55 Einträge pro Block

Universum aller Festnetzanschlüsse (im Jahr 2000)

30,7 Mio eingetragene Nummern

41 Mio generierte Nummern

Anteil generierter Nummern: 57%

Regionale Verortung

Eingetragene Nr. – Zuordnung einer Gemeinde-
kennziffer

zwecks

Bestimmung einer Zuordnungswahrscheinlichkeit
bei den generierten Nummern

Regionale Schichtung von Stichproben

Gewichtung nach Angaben der amtl. Statistik

Beispiel für die Bildung der Zuordnungswahrscheinlichkeit

0931 - 665500	eingetragen	Gemeinde A
0931 – 665501	eingetragen	Gemeinde B
0931 – 665502	generiert	50% Gem. A 50% Gem. B

58% der Blöcke verfügen über eine eindeutige Gemeindezuordnung

Anteil korrekter regionaler Verortung:

91% bei den eingetragenen Anschlüssen

84% bei den nicht eingetragenen Anschlüssen