Challenges of Data Requirements for Modelling Residential Location Choice - the Case of Berlin, Germany

Benjamin Heldt (presenter), Kay Gade, Dr.-Ing. Dirk Heinrichs

DLR

Agenda

1. Introduction
2. The location choice model
3. Evaluating data suitability
4. Data limitations
5. Conclusions

1. Introduction

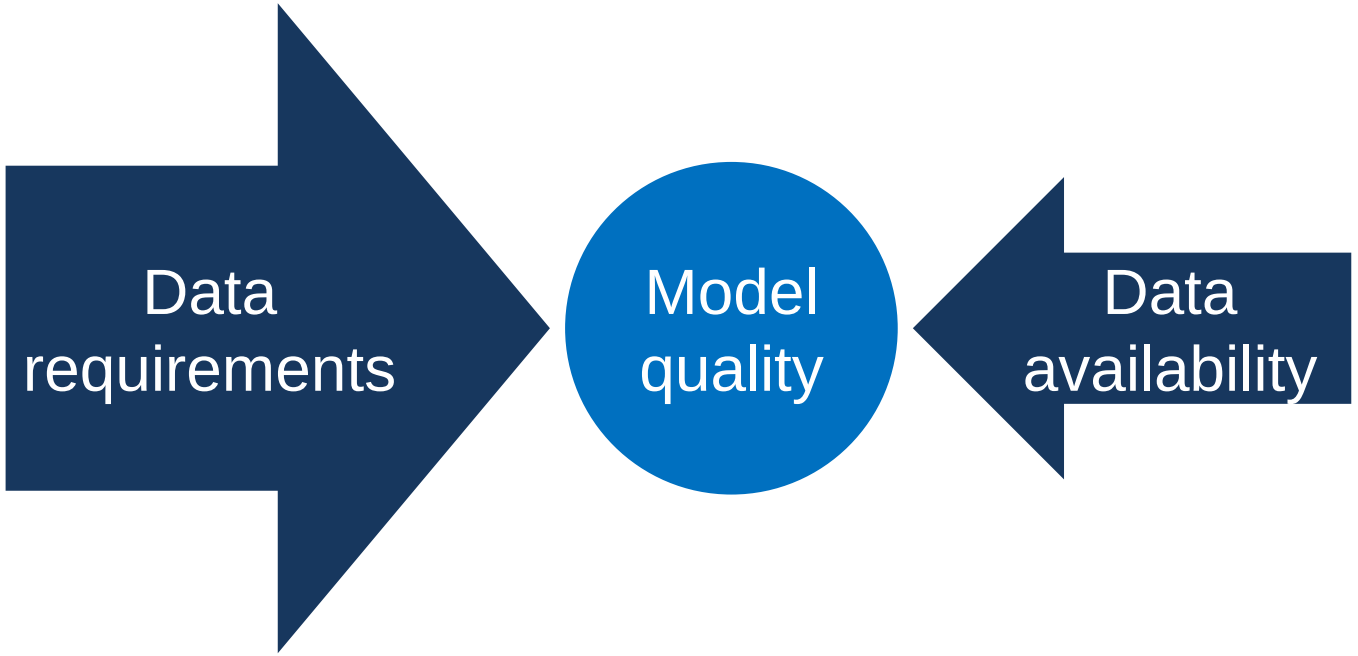
'The data collection for a model of a large metropolis has remained a major effort.'

(Wegener, 1994)

‚A rough approximation is that 75% of the effort and an even higher percentage of the time involved in developing model applications is due to the difficulty of developing the data for the model system.'

(Waddell, 2011)

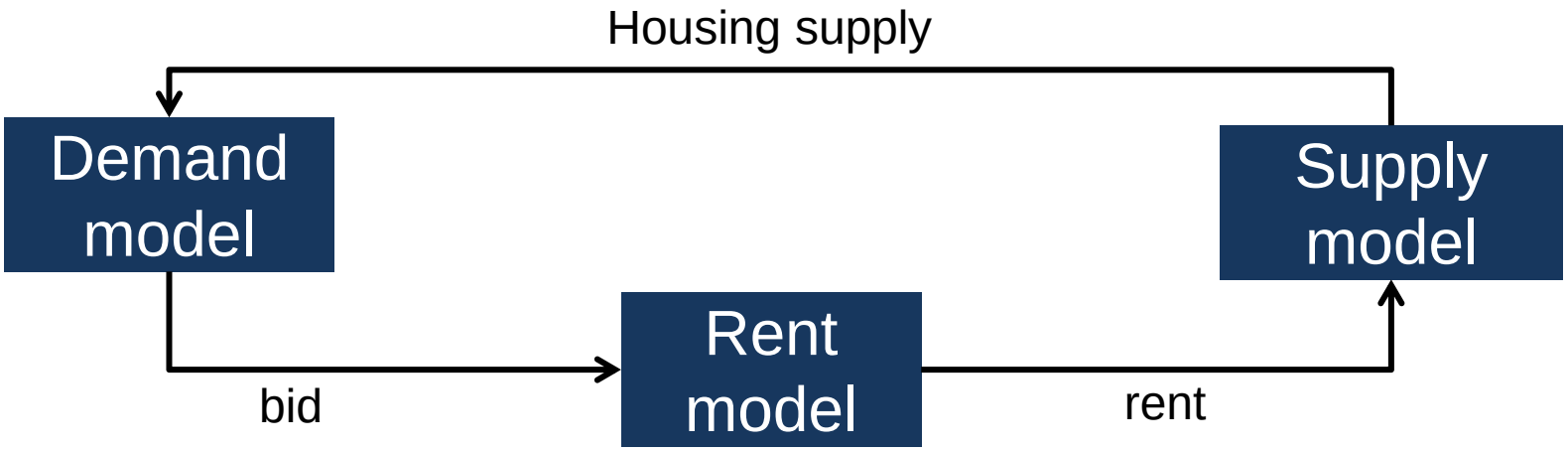
1. Introduction



2. The location model

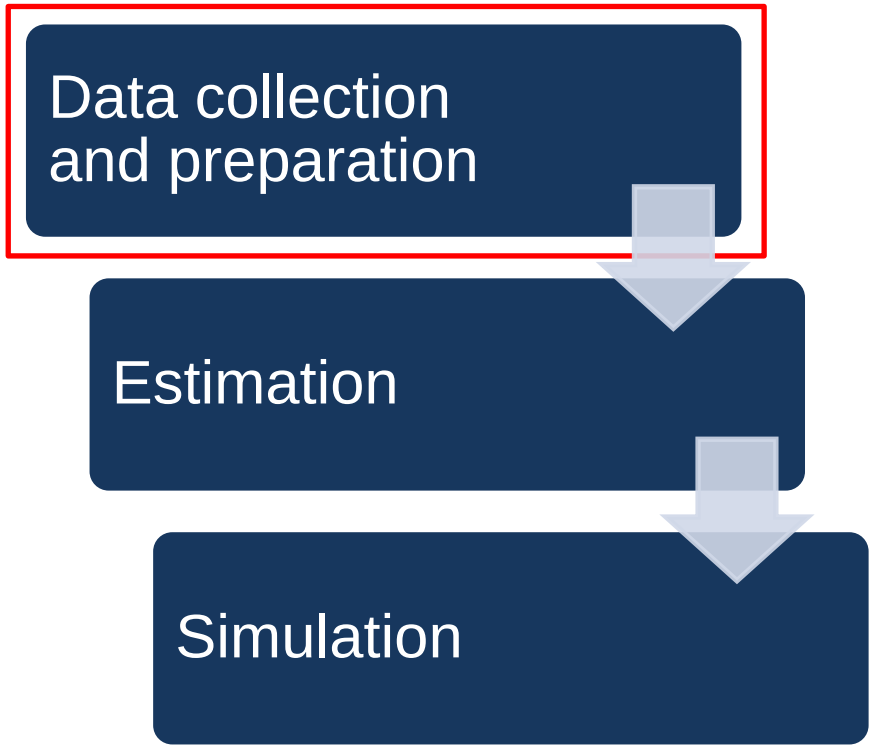
Structure of the land use model *Cube Land*

- Simulates real estate market (residential and non-residential)
- 3 submodels: demand, supply, rent



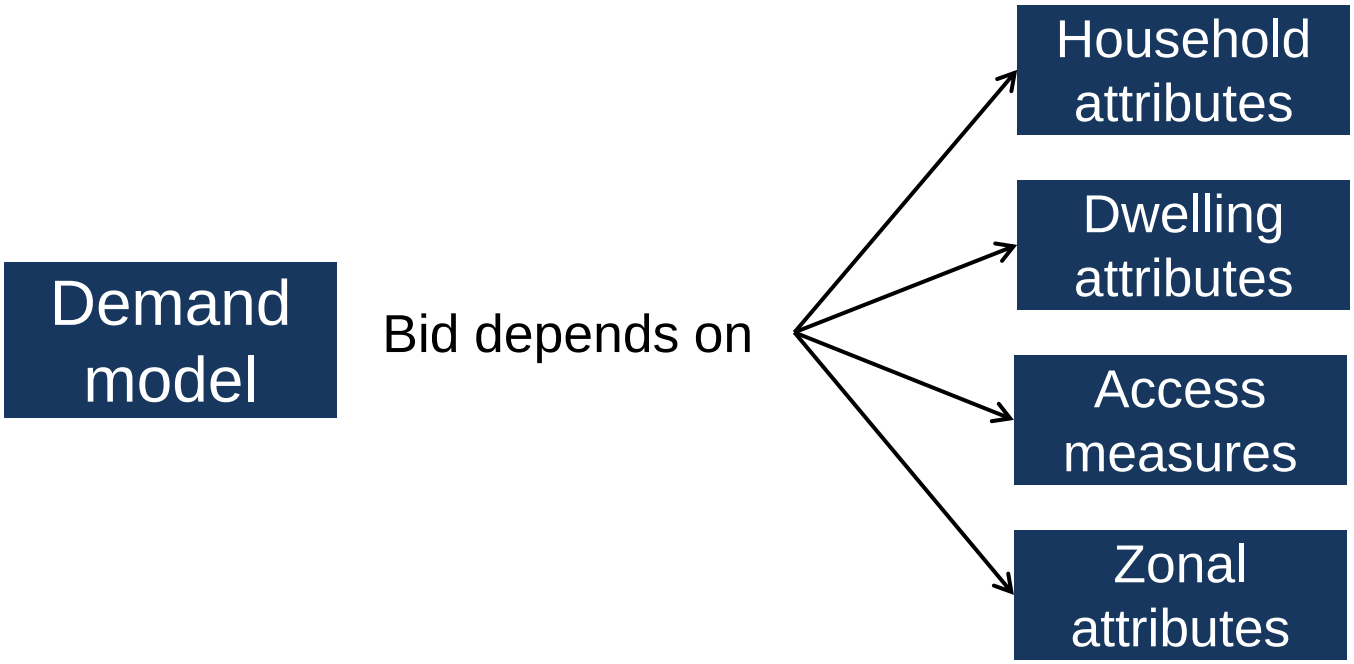
- Bid-choice approach: households issue bids for locations
- Best bidder obtains location

3. Evaluating data suitability



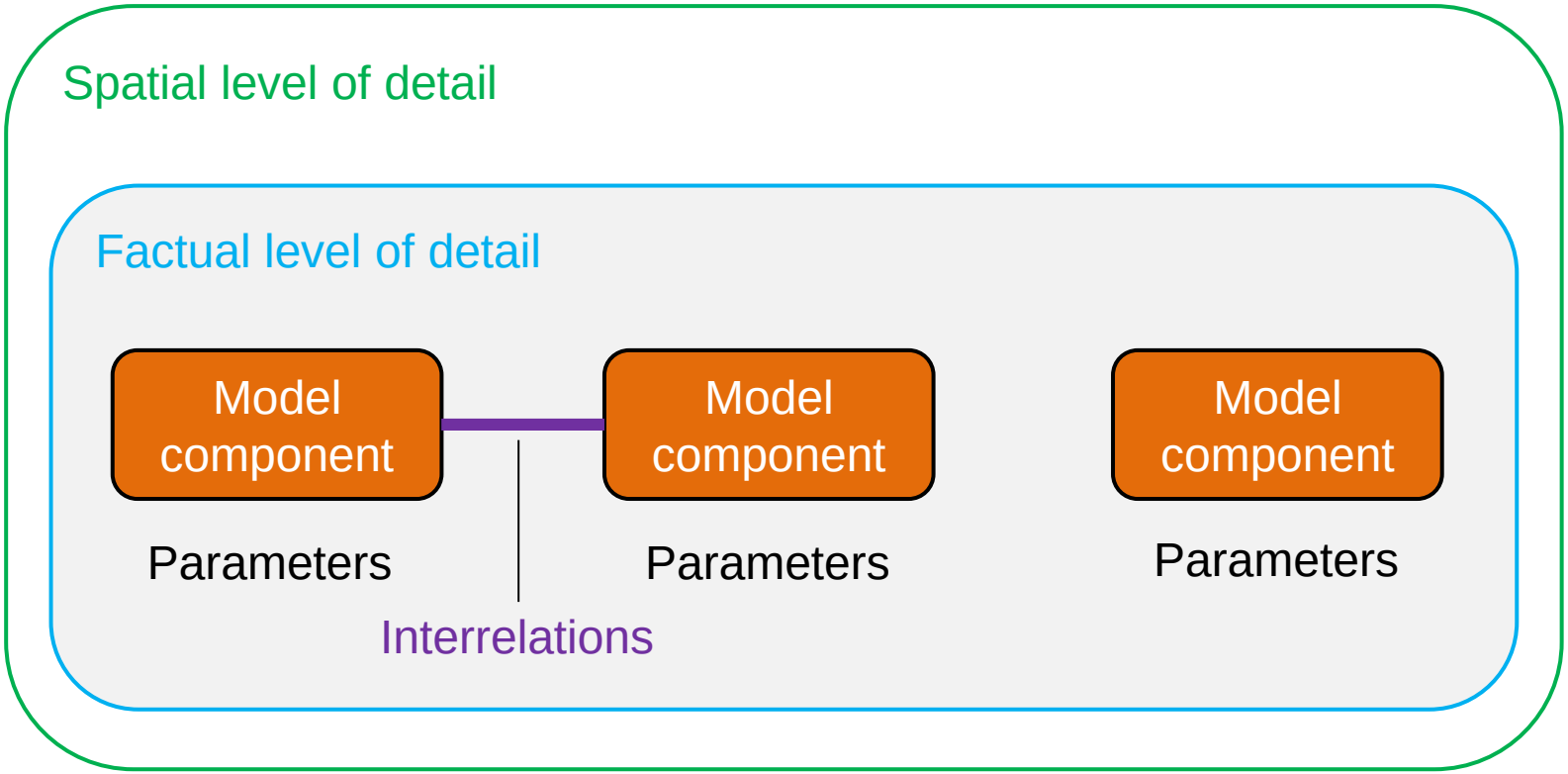
3. Evaluating data suitability

Data requirements of the demand model



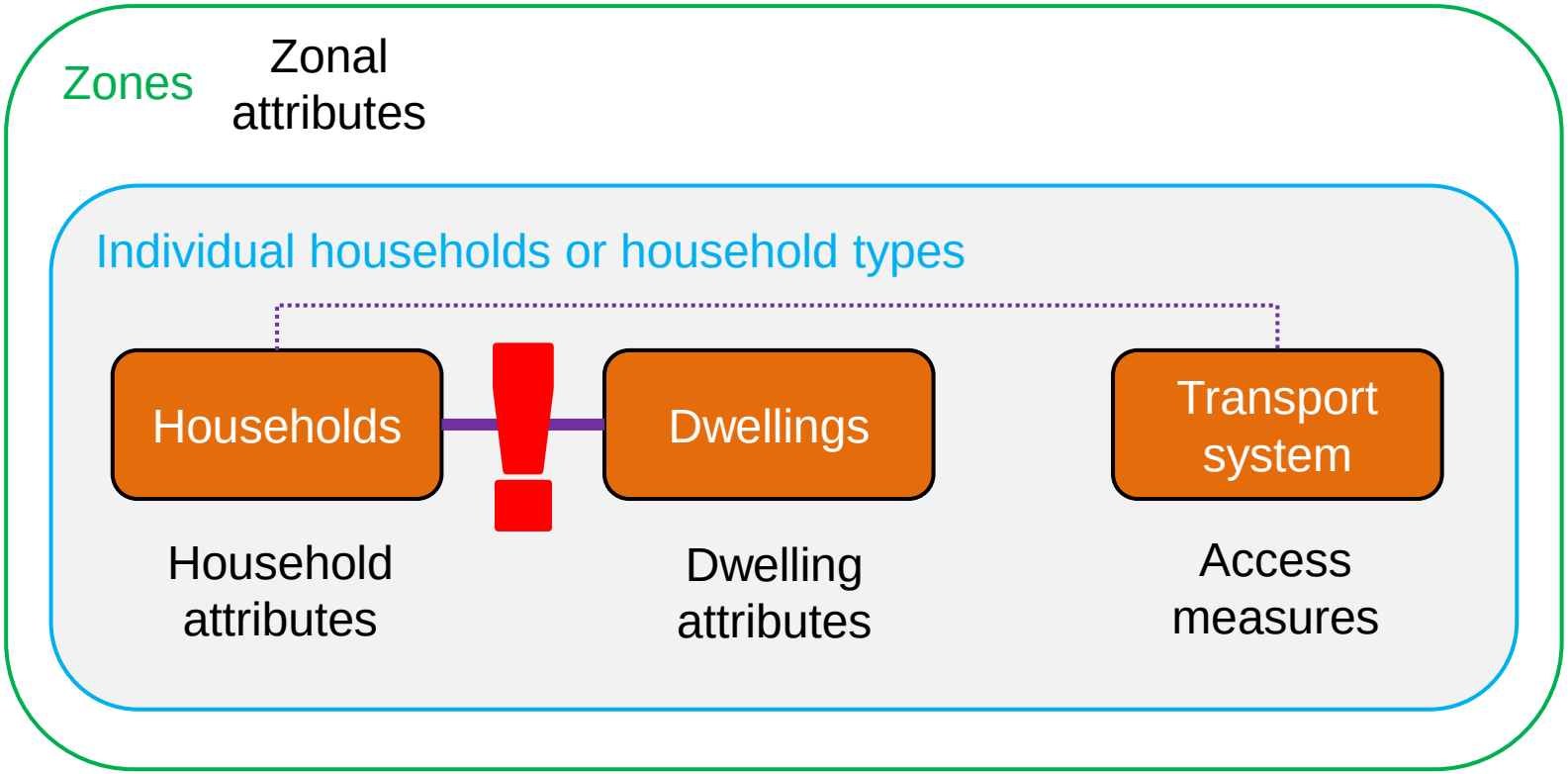
3. Evaluating data suitability

Characteristics of the residential location model



3. Evaluating data suitability

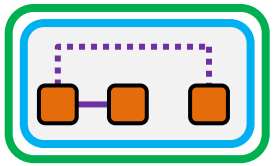
Characteristics of the residential location model



3. Evaluating data suitability

(1) Census of population and buildings	2011	(200.000)
(2) Microcensus of population	2010	15.000
(3) Travel survey (‚SrV‘)	2008	19.000
(4) Commercial data (houses, neighb.)	2010/11	no survey

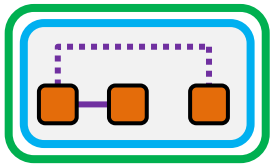
3. Evaluating data suitability



		(1) CENSUS	(2) MICRO-CENSUS	(3) TRAVEL SURVEY	(4)* COMM. DATA
Model components and number of **parameters**	Households	2	3	4	3
	Dwellings	5	3	0	3
	Access measures	-	-	X	-
	Zonal attributes	-	-	-	-
Interrelations	Households and dwellings	X	X	-	-
	Households and zones	X	X	X	-
	Dwellings and zones	X	X	-	-
Factual level of detail	Disaggr. (D) / aggregate. (A)	D/A	D/A	D	A
Spatial level of detail	Number of zones	>12	12	23	3900

* aggregates only (no observations)

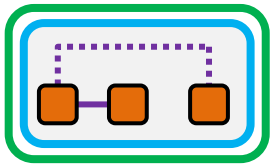
3. Evaluating data suitability



		(1) CENSUS	(2) MICRO-CENSUS	(3) TRAVEL SURVEY	(4)* COMM. DATA
Model components and number of **parameters**	Households	2	3	4	3
	Dwellings	5	3	0	3
	Access measures	-	-	X	-
	Zonal attributes	-	-	-	-
Interrelations	Households and dwellings	X	X	-	-
	Households and zones	X	X	X	-
	Dwellings and zones	X	X	-	-
Factual level of detail	Disaggr. (D) / aggregate. (A)	D/A	D/A	D	A
Spatial level of detail	Number of zones	>12	12	23	3900

* aggregates only (no observations)

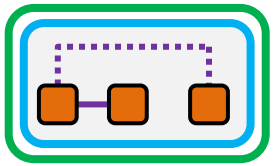
3. Evaluating data suitability



		(1) CENSUS	(2) MICRO-CENSUS	(3) TRAVEL SURVEY	(4)* COMM. DATA
Model components and number of **parameters**	Households	2	3	4	3
	Dwellings	5	3	0	3
	Access measures	-	-	X	-
	Zonal attributes	-	-	-	-
Interrelations	Households and dwellings	X	X	-	-
	Households and zones	X	X	X	-
	Dwellings and zones	X	X	-	-
Factual level of detail	Disaggr. (D) / aggregate. (A)	D/A	D/A	D	A
Spatial level of detail	Number of zones	>12	12	23	3900

* aggregates only (no observations)

3. Evaluating data suitability



		(1) CENSUS	(2) MICRO-CENSUS	(3) TRAVEL SURVEY	(4)* COMM. DATA
Model components and number of **parameters**	Households	2	3	4	3
	Dwellings	5	3	0	3
	Access measures	-	-	X	-
	Zonal attributes	-	-	-	-
Interrelations	Households and dwellings	X	X	-	-
	Households and zones	X	X	X	-
	Dwellings and zones	X	X	-	-
Factual level of detail	Disaggr. (D) / aggregate. (A)	D/A	D/A	D	A
Spatial level of detail	Number of zones	>12	12	23	3900

?

* aggregates only (no observations)

4. Data limitations (1): Incomplete data
Approaches for solutions

Approaches:

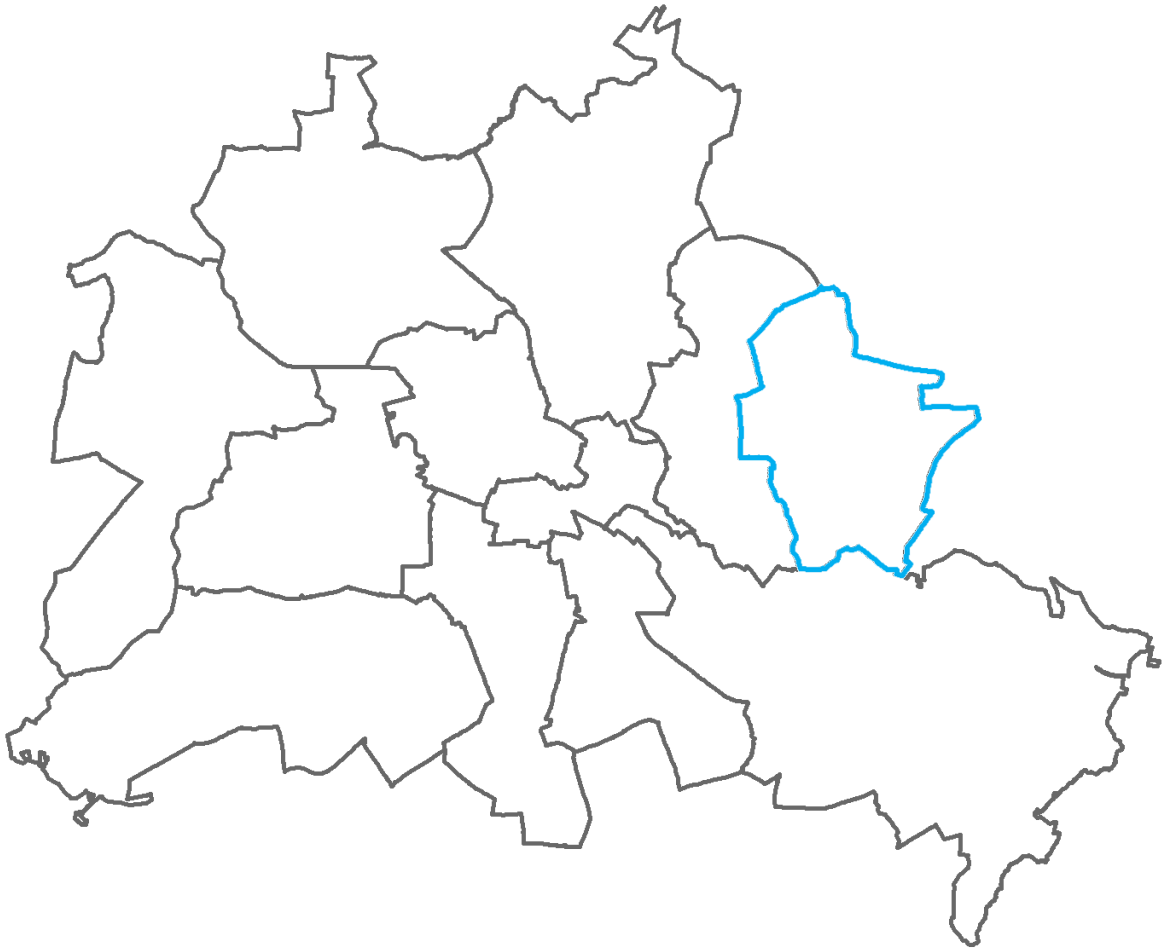
(1) Record linkage

(2) Imputation

(3) Linking aggregate data

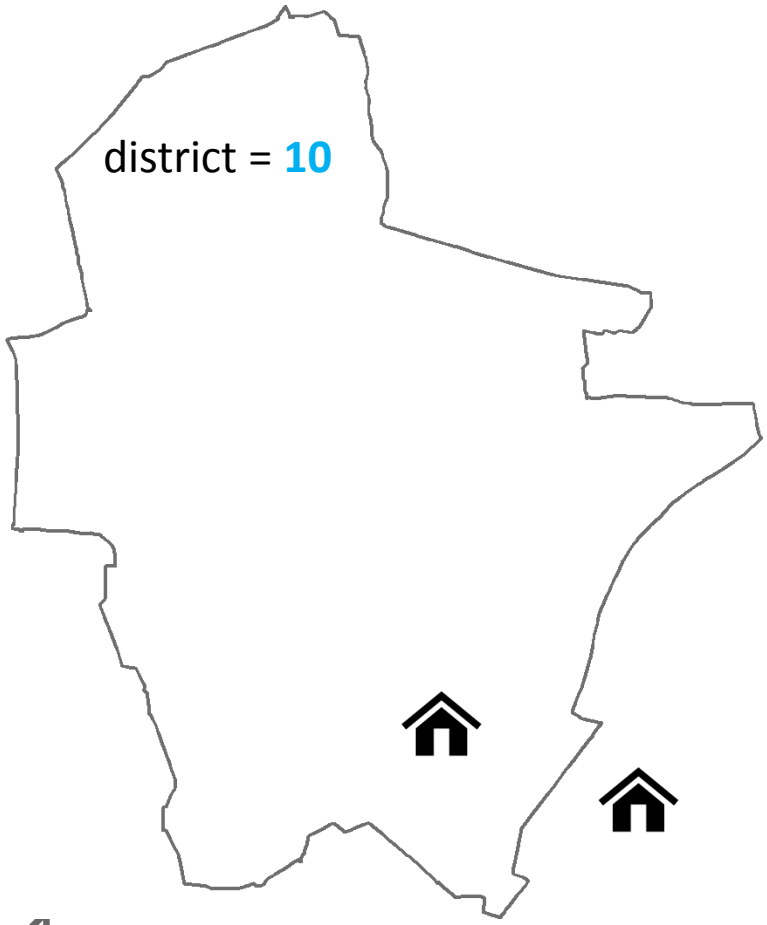
4. Data limitations (1): Incomplete data

Approaches for solutions



4. Data limitations (1): Incomplete data

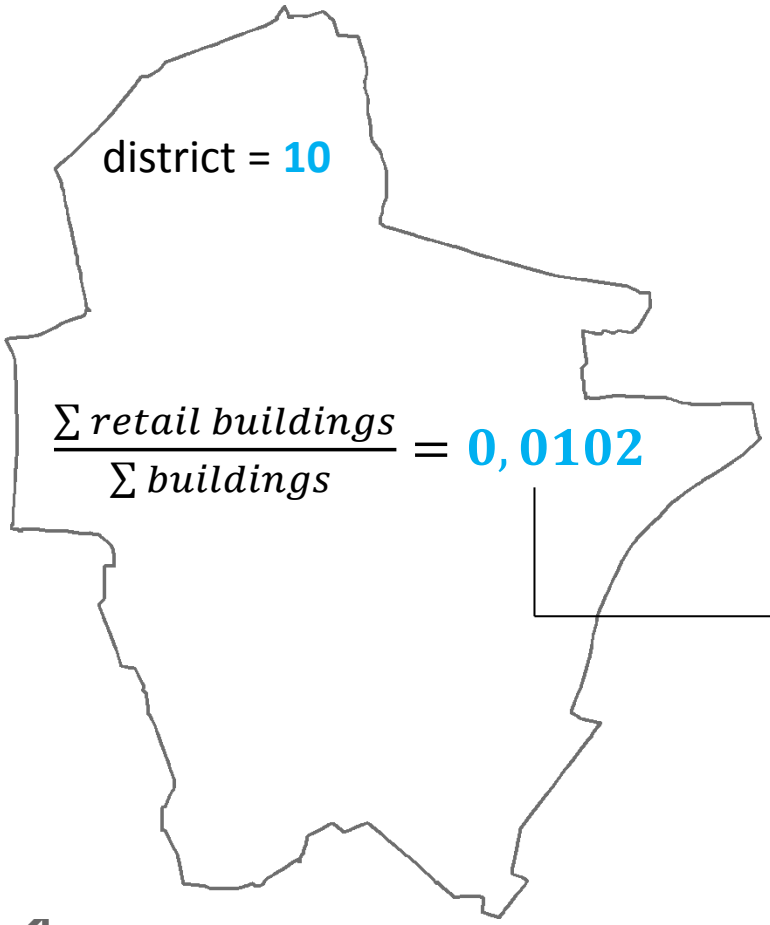
Approaches for solutions



Household	District
1	06
2	09
3	**10**

4. Data limitations (1): Incomplete data

Approaches for solutions



Household	District	Retail_density
1	06	0,0034
2	09	0,0211
3	**10**	**0,0102**

4. Data limitations (1): Incomplete data

Result: Estimation database with zonal attributes

located_type	i	i_run	v	water_area	parc_area	z_in	z_h	total_no	total_area	cw	ret_d	parc_d	res_d	edu_h
1	1	1101	1	1112044,89	5034109,7	1475	198200	20106	39433100	1700,99004	0,01442356	0,12766203	0,57798667	0,00182139
1	1	1101	2	1112044,89	5034109,7	1475	198200	20106	39433100	1700,99004	0,01442356	0,12766203	0,57798667	0,00182139
1	1	1101	3	1112044,89	5034109,7	1475	198200	20106	39433100	1700,99004	0,01442356	0,12766203	0,57798667	0,00182139
1	2	1102	1	612740,49	1128477,83	1400	166900	12691	20136500	1915,4042	0,01481365	0,05604141	0,62043968	0,00109646
1	2	1102	2	612740,49	1128477,83	1400	166900	12691	20136500	1915,4042	0,01481365	0,05604141	0,62043968	0,00109646
1	2	1102	3	612740,49	1128477,83	1400	166900	12691	20136500	1915,4042	0,01481365	0,05604141	0,62043968	0,00109646
1	3	1103	1	1074926,53	4257095,99	1600	221300	65197	102993000	1948,10029	0,00622728	0,04133384	0,44460635	0,00129236
1	3	1103	2	1074926,53	4257095,99	1600	221300	65197	102993000	1948,10029	0,00622728	0,04133384	0,44460635	0,00129236
1	3	1103	3	1074926,53	4257095,99	1600	221300	65197	102993000	1948,10029	0,00622728	0,04133384	0,44460635	0,00129236
1	4	1104	1	1195906,9	3281489,82	1675	196100	24210	64686800	1994,25419	0,0149938	0,05072889	0,5803387	0,00137175
1	4	1104	2	1195906,9	3281489,82	1675	196100	24210	64686800	1994,25419	0,0149938	0,05072889	0,5803387	0,00137175
1	4	1104	3	1195906,9	3281489,82	1675	196100	24210	64686800	1994,25419	0,0149938	0,05072889	0,5803387	0,00137175
1	5	1105	1	5707659,05	5487881,2	1575	124400	61327	91815500	2737,05823	0,0102728	0,05977075	0,5485186	0,00204984

Zonal attributes

4. Data limitations (2): Data privacy and information loss

Problem and approach

Age of householder	Persons per household	Income		
		less than 900€	901 - 1,500€	…
< 30	1 person	529	348	…
	2 persons	11	138	…
	3 persons	3	24	…
	4 persons	0	6	…
	5+ persons	0	1	…
30 - 49	1 person	500	525	…
	2 persons	17	216	…
	3 persons	2	85	…
	4 persons	0	23	…
	5+ persons	2	5	…
…	…	…	…	…

4. Data limitations (2): Data privacy and information loss

Aggregation of household types for BLUME

Age of householder	Persons per household	Household income in €					
		less than 900	901 – 1,500	1,501 – 2,000	2,001 – 2,600	2,601 – 3,600	3,601 and more
< 30	1 person	529	348	90	41	13	6
	2 persons	11	138	59	48	41	15
	3 persons	3	24	32	19	12	3
	4 persons	0	6	15	8	5	3
	5+ persons	0	1	5	6	3	1
30 - 49	1 person	500	525	331	200	109	43
	2 persons	17	216	149	153	178	153
	3 persons	2	85	109	106	164	114
	4 persons	0	23	76	86	116	142
	5+ persons	2	5	24	49	41	46
50 - 69	1 person	498	587	252	137	63	50
	2 persons	12	208	226	288	293	288
	3 persons	0	21	29	40	52	91
	4 persons	0	8	10	13	23	56
	5+ persons	0	0	1	6	10	12
70+	1 person	181	633	230	83	31	18
	2 persons	4	73	194	268	154	73
	3 persons	0	5	4	8	7	5
	4 persons	0	0	0	0	0	4
	5+ persons	0	0	0	0	0	1

DLR

4. Data limitations (2): Data privacy and information loss

Aggregation of household types for BLUME

Age of householder	Persons per household	Household income in €					
		less than 900	901 – 1,500	1,501 – 2,000	2,001 – 2,600	2,601 – 3,600	3,601 and more
< 30	1 person	529	348	90	41	13	6
	2 persons	11	138	59	48	41	15
	3 persons	3	24	32	19	12	3
	4 persons	0	6	15	8	5	3
	5+ persons	0	1	5	6	3	1
30 - 49	1 person	500	525	331	200	109	43
	2 persons	17	216	149	153	178	153
	3 persons	2	85	109	106	164	114
	4 persons	0	23	76	86	116	142
	5+ persons	2	5	24	41	41	46
50 - 69	1 person	498	527	262	132	63	50
	2 persons	12	208	226	288	293	288
	3 persons	0	21	29	40	52	91
	4 persons	0	8	10	13	23	56
	5+ persons	0	0	1	6	10	12
70+	1 person	181	633	230	83	31	18
	2 persons	4	73	194	268	154	73
	3 persons	0	5	4	8	7	5
	4 persons	0	0	0	0	0	4
	5+ persons	0	0	0	0	0	1

120 household types

4. Data limitations (2): Data privacy and information loss

Aggregation of household types for BLUME

Age of householder	Persons per household	Household income in €					
		less than 900	901 – 1,500	1,501 – 2,000	2,001 – 2,600	2,601 – 3,600	3,601 and more
< 30	1 person	529	348	90	41	13	6
	2 persons	11	138	59	48	41	15
	3 persons	3	24	32	19	12	3
	4 persons	0	6	15	8	5	3
	5+ persons	0	1	5	6	3	1
30 - 49	1 person	500	525	331	200	109	43
	2 persons	17	216	149	153	178	153
	3 persons	2	85	109	106	164	114
	4 persons	0	23	76	86	116	142
	5+ persons	2	5	24	49	41	46
50 - 69	1 person	498	587	252	137	63	50
	2 persons	12	208	226	288	293	288
	3 persons	0	21	29	40	52	91
	4 persons	0	8	10	13	23	56
	5+ persons	0	0	1	6	10	12
70+	1 person	181	633	230	83	31	18
	2 persons	4	73	194	268	154	73
	3 persons	0	5	4	8	7	5
	4 persons	0	0	0	0	0	4
	5+ persons	0	0	0	0	0	1

4. Data limitations (2): Data privacy and information loss

Result: Aggregated household types

Age of householder	Persons per household	Household income in €					
		less than 900	901 - 1,500	1,501 - 2,000	2,001 - 2,600	2,601 - 3,600	3,601 and more
< 30	1 person	529	348	150			
	2+ persons		183	111	164		
30 - 49	1 person	500	525	331	200	109	
	2 persons		233	149	153	178	196
	3+ persons		117	209	241	321	302
50 - 69	1 person	510	587	252	250		
	2 persons		208	226	288	293	288
	3+ persons					213	159
70+	1 person	185	633	230	83		
	2+ persons			276	276	293	

4. Data limitations (2): Data privacy and information loss

Result: Aggregated household types

Age of householder	Persons per household	Household income in €					
		less than 900	901 - 1,500	1,501 - 2,000	2,001 - 2,600	2,601 - 3,600	3,601 and more
< 30	1 person	529	348	150			
	2+ persons		183	111	164		
30 - 49	1 person	500	525	331	200	109	
	2 persons		233	149	153	178	196
	3+ persons		117	209	211	331	302
50 - 69	1 person	510	287	252	210		
	2 persons		208	226	288	293	288
	3+ persons					213	159
70+	1 person	185	633	230	83		
	2+ persons			276	276	293	

39 household types

6. Conclusions

- Applying model characteristics as data evaluation criteria helps to identify suitable data sources and better assess model quality depending on the chosen data
- Existing data sources (travel surveys, censuses) do not feature sufficient data
- Limitations arise due to survey design or data privacy
- Dwelling characteristics should be considered in travel surveys and mobility attributes in censuses
- Spatial join is one method to complement incomplete data
- Using elaborated methods for aggregating data helps to reduce information loss due to data privacy

Thank you!

Benjamin Heldt
DLR – German Aerospace Center
Institute of Transport Research

Benjamin.Heldt@DLR.de
www.DLR.de/vf