TRANSIMS Training Course at TRACC

Transportation Research and Analysis Computing Center

Part 3

Population Synthesis based on CENSUS Resources

Dr.-Ing. Hubert Ley
Transportation Research and Analysis Computing Center

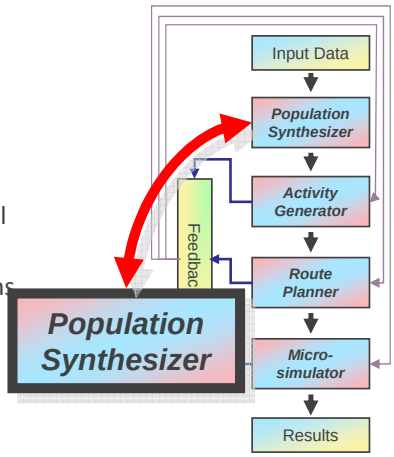
Last Updated: December 9, 2009

Unit **3**

Contents

- Introduction
- CENSUS Bureau Data
- Data Sources
- Data Preparation Tools
- Census Bureau PUMS Data
- Census Bureau SF3 Data
- Placement of Households on the Network
- Algorithms
- IPF Algorithms (Traditional and Two-Step)
- Three-Dimensional IPF Procedure

The TRANSIMS Population Synthesizer



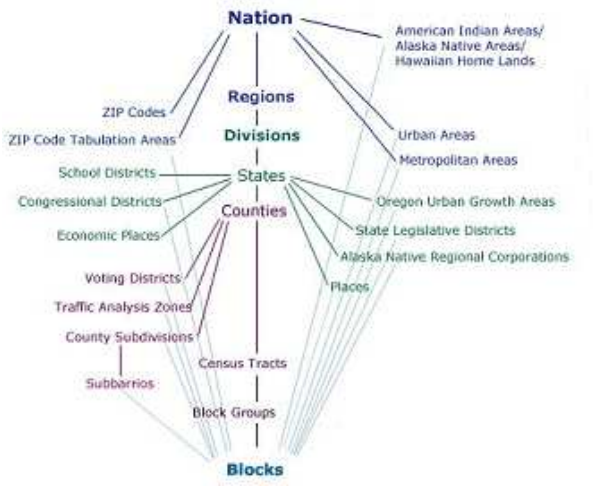
- Mimics regional population ("synthetic population")
  - Demographics closely match real population
  - Households distributed spatially to approximate regional population distribution
  - Household locations determine some of the travel origins and destinations
- Functions of the Population Synthesizer
  - Generation of synthetic households from census data at the block group level
  - Development of each household demographic characteristics (income, members, etc)
  - Placement of each synthetic household on a link in transportation network (activity locations)
  - Assignment of vehicles to each household (sharing vehicles and rides within a household)

Introduction

- The population generator creates synthetic households from census data to model a real population for the entire coverage area
  - Census data is provided for the entire United States
  - The data is made anonymous by the Census Bureau by
    - Providing summarized data on various levels (STF)
    - Providing small but representative subsets of actual data records taken from a larger area (PUMA)
  - The smallest unit of public Census data with sufficient detail is at the
    - block group level (summary data) for ~ 600 to 3,000 people
    - public use microdata area level (sample records) for ~100,000 people
- Develops associated demographic characteristics for each household
- Places each synthetic household on a link in the transportation network
- Assigns vehicles to each household

Population Synthesizer

- The Census data is broken down by summarizing it on a number of levels:
  - Entire United States (10)
  - By State (40)
  - By County (50)
  - By Census Tract (80)
  - By Census Block Group (90)
  - By Census Block (100)
- Independently, samples are provided for PUMAs (public use microdata areas)
  - 1% and 5% sample records
  - both are made anonymous
- The challenge is to reconstruct a representative synthetic population

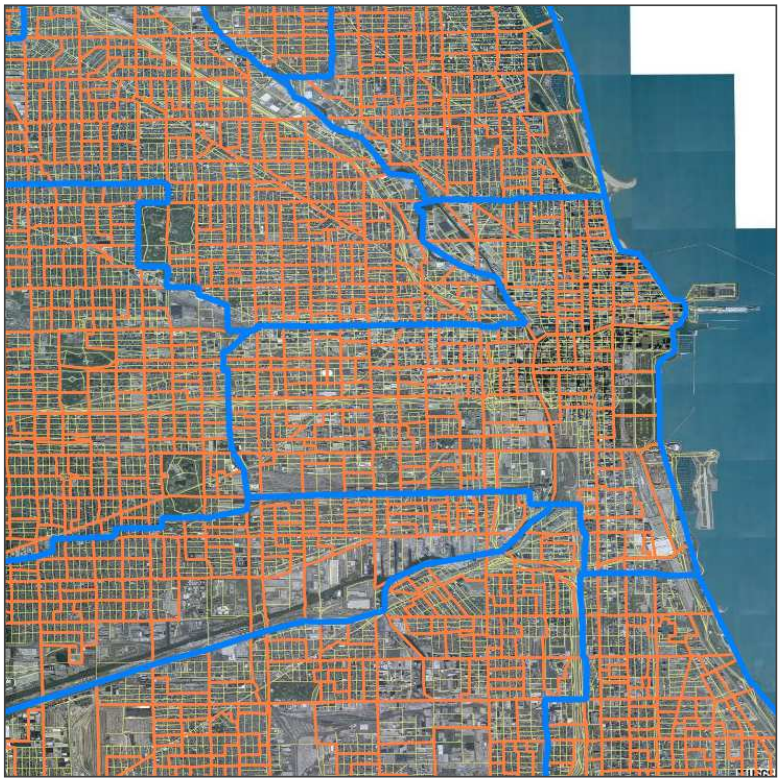


CENSUS Bureau Data

- The following Census Bureau data is of interest for modeling
  - Summary File 3 tables
    - Contain demographic summary tables from Census data for small geographic areas. These one-dimensional summary tables contain information on 100% household demographic variables at the Census Block Group level.
  - PUMS tables
    - Public Use Microdata Sample files consist of a 5% representative sample of almost complete census records from those households contained in a collection of census tracts or other small geographic census areas, which collectively is called a Public Use Micro Area (PUMA).
- A PUMA is constructed so that it contains approximately 100,000 individuals. These files are edited to protect the confidentiality of all individuals, but they have the information necessary to conduct effective research and analysis.

CENSUS Data

- **BLUE**
  PUMS 5% Sample Data
- **ORANGE**
  Aggregate STF3 Data
- **YELLOW**
  Street Network



Data of Interest for TRANSIMS Households

- Characteristic aggregate data per household:
  - HHOLD Household ID
  - STATE State Abbreviation or FIPS code
  - PUMA PUMA number
  - WEIGHT Survey expansion factor
  - PERSONS The number of persons in the household
  - WORKERS The number of workers in the household
  - VEHICLES The number of vehicles available to the household
  - INCOME The household income
  - TYPE The household type
  - HHAGE The age of the head of the household (householder)
  - NUM_LT5 The number of children less than 5 years old
  - NUM_5TO15 The number of children between 5 and 15 years old

Data of Interest for TRANSIMS Households

- Characteristic aggregate data per person:
  - HHOLD — Household ID
  - PERSON — Person number within the household
  - RELATE — The relationship or role of the person in the household
  - GENDER — The gender or sex of the person
  - AGE — The age of the person
  - WORKER — The person's employment status
- Both sets of data (household data and person data) are to be extracted from CENSUS data tables

PopSyn Input Data

- PopSyn is the Version 4 Population Synthesizer
- PopSyn requires the following input data
  - A Zone Data File (Created by ZoneData)
  - A PUMS Household File (Created by PUMSPrep)
  - A PUMS Population File (Created by PUMSPrep)
  - An Activity Location File ("Enriched" using LocationData)
  - A Process Link File (Created by TransimsNet)
  - A Vehicle Type Distribution File
- There are two documents that will help understanding the process
  - The 2000 US Census Data Preparation How-To
  - The Population Synthesizer How-To
  - Both are available from the TRANSIMS web site

TRANSIMS Tools

- PopSyn is the population synthesizer itself
- The input for PopSyn is created using the following tools
  - PUMSPrep
  - SF3Prep
  - ZoneData
  - LocationData
- All data is taken from standard CENSUS files
  - PUMS: http://www2.census.gov/census_2000/datasets/PUMS/FivePercent
  - SF3: http://www2.census.gov/census_2000/datasets/Summary_File_3
- All data is based on the Year 2000 CENSUS

Data Preparation: PUMS Data and PUMSPrep

- H000216551723001011701099979997999799977070   4910786365   4764538933   17748095205
  17383738799830018020000002002090605030101010101020100 00480006100017000001 0   0 030
  0 0   0 0100 0035000000000090 0
  010200020000044100168008000000000211140900024300000243000
- P00021650100001801000010760010110000001014701010010000900220500220000
  00170100000001000000000000000000000000000000000002020202020202200000200000000010020146000
  00000000000000000000000000000000000000000000002222320000000000000000000000-
  0000002000000000000000000000000670001270000000000000000000000000000000000194000000000000242
- P00021650200002202000420780010110000001014701010010000900220320220000
  00170100000001000000000000000000000000000000000002020202020202200000400000000000000006000
  00000000000000000000000000000000000000000000002222320000000000000000000000-
  0000002000000000000000000000000000049000000000000000000000000000000000000049000000000000242
- H000584751723001011701099979997999799977070   4910786365   4764538933   17748095205
  17383738799020006030000010020505060301010101030102000400000500008000000010   0
  01000810030   020380200380100000140 0
  01000103010102200150002300000000021141010008000000080000
- ...

- PUMS data files are compressed files for the entire state
  - Compressed 37MB for Illinois
  - Uncompressed 260MB for Illinois
  - 880,000 records for Illinois
  - Text file with no separators between fields and very long lines

Data Preparation: PUMS Data and PUMSPrep

- Typical information that is being extracted and used from the PUMS data sets

Data	Record Size	Beginning At Field	Description of Data	Allowed Values	Description of Values
RECTYPE	1	1	Record Type	H	Housing Record
PUMA	5	13	Public use microdata area (state dependent)	00100… 99999	PUMA code.
RHHINC	7	141	Household income	0000000	N/A* (GQ**/vacant/no income)
				-999999… 9999999	Total household income in dollars
RWRKR89	1	148	Workers in family in 1989	0	N/A*
				1	No Workers
				2	1 Workers
				3	2 Workers
				4	3 Workers
R18UNDR	1	162	Presence of person under 18 years in household	0	N/A* (No person under 18 in household/GQ**/ vacant)
				1	1 or more person under 18 in household

Data	Record Size	Beginning At Field	Description of Data	Allowed Values	Description of Values
RECTYPE	1	1	Record Type	P	Person Record
RELAT1	2	9	Relationship	00	Householder
				01	Husband/Wife
				02	Son/Daughter
				03	Stepson/Stepdaughter
				04	Brother/Sister
				05	Father/Mother
				06	Grandchild
				07	Other relative
				Non Related	
				08	Roomer/boarder/foster child
				09	Housemate/roommate
				10	Unmarried partner
				11	Other non-relative
				Group Quarters	
				12	Institutionalized person
				13	Other person in group quarters.
SEX	1	11	Sex	0	Male
				1	Female
AGE	2	15	Age	00	Less than 1 year
				01… 89	Age in years
				90	90 or more years old
WORK89	1	122	Worked last year (1989)	0	N/A (less than 16 years old)
				1	Worked last year
				2	Did not work last year

Data Preparation: PUMS Data and PUMSPrep

- Synthetic households in TRANSIMS are divided into three categories:

1. Family households — two or more related persons
2. Non-family households — persons living alone or unrelated persons living together
3. Group quarters — dwellings such as prisons or college dormitories

Data Preparation: PUMS Data and PUMSPrep

```
TITLE                   PUMS Data Extraction
DEFAULT_FILE_FORMAT     TAB_DELIMITED
#---- Input Files ----
PUMS_DATA_FILE          census/PUMS5_17.TXT
#---- Output Files ----
NEW_PUMS_HOUSEHOLD_FILE     demand/Households
NEW_PUMS_POPULATION_FILE    demand/Persons
#---- Parameters ----
HOUSEHOLD_DATA_FIELD_1      HHOLD, INTEGER, 2, 7
HOUSEHOLD_DATA_FIELD_2      STATE, STRING, 10, 2
HOUSEHOLD_DATA_FIELD_3      PUMA, INTEGER, 14, 5
HOUSEHOLD_DATA_FIELD_4      WEIGHT, INTEGER, 102, 4
HOUSEHOLD_DATA_FIELD_5      PERSONS, INTEGER, 218, 2
HOUSEHOLD_DATA_FIELD_6      WORKERS, INTEGER, 0, 1
HOUSEHOLD_DATA_FIELD_7      VEHICLES, INTEGER, 134, 1
HOUSEHOLD_DATA_FIELD_8      INCOME, INTEGER, 251, 8
HOUSEHOLD_DATA_FIELD_9      TYPE, INTEGER, 213, 1
HOUSEHOLD_DATA_FIELD_10     HHAGE, INTEGER, 0, 2
HOUSEHOLD_DATA_FIELD_11     NUM_LT5, INTEGER, 0, 2
HOUSEHOLD_DATA_FIELD_12     NUM_5TO15, INTEGER, 0, 2
POPULATION_DATA_FIELD_1     HHOLD, INTEGER, 2, 7
POPULATION_DATA_FIELD_2     PERSON, INTEGER, 9, 2
POPULATION_DATA_FIELD_3     RELATE, INTEGER, 17, 2
POPULATION_DATA_FIELD_4     GENDER, INTEGER, 23, 1
POPULATION_DATA_FIELD_5     AGE, INTEGER, 25, 2
POPULATION_DATA_FIELD_6     WORKER, INTEGER, 236, 1
STATE_PUMA_LIST         51100, 51200, 51301, 51302
CONVERSION_SCRIPT       census/PUMS_Script.txt
```

- PUMSPrep Control File
  - Refers to the PUMS data file downloaded from the CENSUS web site
  - Creates both a household and a persons (population) file
  - Assigns data from specific column ranges in the PUMS data file to variables that are mostly user-defined
  - Processes data for a specific list of PUMAs that cover the model area
  - Some fields cannot be populated directly but are calculated in a TRANSIMS script (see next page)

Data Preparation: PUMS Data and PUMSPrep

- TRANSIMS Script for PUMSPrep

```
IF (New_Pop.WORKER == 1) THEN
   New_HH.WORKERS = New_HH.WORKERS + 1
ENDIF
IF (New_Pop.RELATE == 1) THEN
   New_HH.HHAGE = New_Pop.AGE
ENDIF
IF (New_Pop.AGE < 5) THEN
   New_HH.NUM_LT5 = New_HH.NUM_LT5 + 1
ELSE
 IF (New_Pop.AGE <= 15) THEN
    New_HH.NUM_5TO15 = New_HH.NUM_5TO15 + 1
 ENDIF
ENDIF
```

- TRANSIMS Scripts are used for a number of different TRANSIMS tools
- This is a special and somewhat limited programming language
- They are somewhat similar to PL/SQL trigger scripts
- Scripts are written as text files and have to be referenced from the correct control file
- The scripts provide variables on input and output to allow powerful pre- and post-processing of data
- Here a script is used for creating aggregate counts per household based on person records in the PUMA data file

Data Preparation: PUMS Data and PUMSPrep

- Sample household file

HHOLD	STATE	PUMA	WEIGHT	PERSONS	WORKERS	VEHICLES	INCOME	TYPE	HHAGE	NUM_LT5	NUM_5TO15
1347	51	100	17	2	1	1	23000	3	32	0	1
4366	51	100	31	2	2	2	201030	5	44	0	0
5847	51	100	31	2	0	1	4	2	18	0	0
7646	51	100	11	3	3	2	42400	5	28	0	0
20464	51	100	22	2	2	2	13200	3	46	0	0

- Sample person file

HHOLD	PERSON	RELATE	GENDER	AGE	WORKER
1347	1	1	2	32	1
1347	2	3	2	10	0
4366	1	1	1	44	1
4366	2	19	1	50	1
5847	1	1	1	18	2
5847	2	16	1	24	2
7646	1	1	1	28	1
7646	2	18	1	28	1
7646	3	18	2	26	1
20464	1	1	2	46	1

- These are sample household and person files from the Alexandria case study generated by PUMSPrep
- Some of the household data has been aggregated from person records
- These are used as inputs for PopSyn at a later stage of the process

Data Preparation: SF3 Data and SF3Prep

- uSF3,IL,000,10,0000001,423440,206752,13570,14070,14137,8776,6760,3577,3408,10882,22548,20936,18161,16123,14823,12572,9074,3030,3500,2129,2420,2794,1922,953,587,216688,14920,12474,12922,7667,6057,3424,3188,10911,23845,21977,19057,17545,16136,14408,10179,3294,3756,2160,2928,4116,2891,1838,995,3811,1890,139,132,119,97,68,112,48,110,200,139,155,196,83,53,104,18,40,13,5,22,12,22,3,1921,125,162,153,65,128,81,22,78,247,156,121,120,149,132,20,44,41,19,11,9,18,0,20,724021,381817,40784,38737,32662,20786,16320,8827,8842,28143,45018,36094,30473,24826,18704,12362,6869,2191,2556,1502,1744,2127,1381,565,304,342204,38487,37271,31815,18414,14356,7485,7587,22505,35703,29844,26553,22557,15959,11523,7283,2480,2995,1666,2098,2623,1671,752,577,249431,125851,18892,14970,12198,6333,4095,2085,2038,6421,10206,9244,9364,8064,5986,4833,3362,1111,1612,825,1005,1373,1048,474,312,123580,18018,14106,12095,6271,4113,2140,2118,6128,9941,8656,8337,7772,5588,4983,3089,1163,1663,884,1163,2102,1491,1016,743,1529141,806477,89661,83155,68705,41180,33512,18086,17287,55162,90223,76251,64796,52237,38751,26640,16699,5437,6908,4333,4935,6326,3657,1568,968,722664,85223,79309,66722,36015,27665,14516,14254,44041,72789,64163,55851,46356,34180,25270,17222,5826,7318,4171,5557,7494,4603,2300,1819
- uSF3,IL,000,10,0000002,418352,204642,13352,13904,13926,8652,6635,3542,3382,10790,22396,20804,18042,16003,14676,12453,8908,2997,3448,2120,2406,2766,1907,953,580,213710,14712,12327,12670,7560,5951,3346,3145,10793,23607,21762,18795,17339,15839,14177,9966,3247,3705,2143,2869,4062,2881,1827,987,3525,1745,137,117,110,94,65,112,48,94,180,125,137,181,69,48,98,18,40,13,0,22,12,22,3,1780,125,158,123,61,117,81,22,74,223,146,105,117,141,128,18,33,36,19,11,7,18,0,17,716684,377614,40407,38303,32245,20550,16103,8697,8740,27888,44609,35673,30154,24486,18443,12235,6812,2173,2538,1481,1730,2118,1372,558,299,339070,38129,36925,31532,18221,14216,7419,7510,22371,35404,29534,26309,22273,15773,11458,7210,2472,2989,1656,2090,2613,1660,746,560,238298,120402,18142,14298,11484,5986,3920,2022,1951,6316,9960,8979,9024,7736,5645,4573,3176,977,1506,784,928,1256,1021,435,283,117896,17272,13514,11452,5911,3956,2093,2070,5917,9634,8392,7967,7345,5188,4679,2887,1109,1548,852,1087,1970,1388,959,706,1508724,795465,88439,81979,67614,40565,33031,17836,17052,54600,89316,75364,63959,51446,38088,26155,16445,5317,6844,4255,4886,6230,3557,1531,956,713259,84155,78230,65679,35415,27265,14357,14115,43617,72065,63441,55260,45702,33580,24973,16935,5743,7223,4111,5497,7357,4504,2262,1773
- ...

- SF3 data files are compressed files for the entire state or by county
  - Compressed 404MB for Illinois
  - Uncompressed approx. 10GB for Illinois
  - 76 county data files for Illinois
  - Text file with separators between fields and very long lines

Data Preparation: SF3 Data and SF3Prep

```
TITLE                   SF3 Data Extraction
DEFAULT_FILE_FORMAT     TAB_DELIMITED

#---- Input Files ----
SF3_GEOGRAPHY_FILE_1    census/ilgeo.uf3
SF3_SEGMENT_FILE_1_1    census/il00001.uf3
SF3_SEGMENT_FILE_1_7    census/il00007.uf3

#---- Output Files ----
NEW_ZONE_DATA_FILE      ZonesSF3

#---- Parameters ----
ZONE_DATA_FIELD_RANGE_1     STATE
ZONE_DATA_FIELD_RANGE_2     COUNTY
ZONE_DATA_FIELD_RANGE_3     TRACT
ZONE_DATA_FIELD_RANGE_4     BLKGRP
ZONE_DATA_FIELD_RANGE_5     P013003..10, HH_AGE1..8
ZONE_DATA_FIELD_RANGE_6     P013012..19, NF_AGE1..8
ZONE_DATA_FIELD_RANGE_7     P014003..8, HH_SIZE2..7
ZONE_DATA_FIELD_RANGE_8     P014010..16, NF_SIZE1..7
ZONE_DATA_FIELD_RANGE_9     P076002..17, HH_INC1..16
ZONE_DATA_FIELD_RANGE_10    P079002..17, NF_INC1..16

SF3_SUMMARY_LEVEL       90
STATE_COUNTY_LIST       51013, 51059, 51510

INPUT_COORDINATE_SYSTEM     LATLONG, MILLION_DEGREES
OUTPUT_COORDINATE_SYSTEM    UTM, 16N, METERS
```

- SF3Prep Control File
  - Refers to the SF3 data files downloaded from the CENSUS web site
  - Creates an intermediate zone data file (SF3 zone data file)
  - Assigns data from specific summary tables in the SF3 data files to variables that are mostly user-defined
  - Processes data for a specific list of counties that cover the model area
  - Reprojection of the geocoding for the centroids for each block group is necessary (TRANSIMS uses UTM)

Association of CENSUS Data with the TRANSIMS Model

- At this point, the CENSUS data is in a format that's suitable for the population synthesizer
- The PUMAs and the SF3 block group data are at this time not associated with each other
- The TRANSIMS network, and specifically the activity locations, are at this point unaware of any of the preparatory work done in the previous steps
- Activity locations need to associated with specific block groups and PUMAs for further processing
- The data files are now being processed to include appropriate geographic coding by placing relational indices into the files
- This is based on GIS shape files available from the CENSUS web site
  - These shape files contain polygons for block groups and PUMAs
- More details can be found in the current PopSyn How-To document

Data Preparation: The ZoneData Tool

```
TITLE                   BlockGroup Data Manipulation
DEFAULT_FILE_FORMAT     TAB_DELIMITED

#---- Input Files ----
NET_ZONE_TABLE          ZonesSF3
BOUNDARY_POLYGON_1      census/arcview/p517_d00.shp
BOUNDARY_POLYGON_2      census/arcview/bg17_d00.shp

CONVERSION_SCRIPT       census/BlockGroup_Script.txt

#---- Output Files ----
NEW_ZONE_TABLE          ZonesData

#---- Parameters ----
COPY_EXISTING_FIELDS    NO

NEW_ZONE_FIELD_1        STATE, INTEGER, 3
NEW_ZONE_FIELD_2        COUNTY, INTEGER, 5
NEW_ZONE_FIELD_3        TRACT, INTEGER, 7
NEW_ZONE_FIELD_4        BLKGRP, INTEGER, 2
NEW_ZONE_FIELD_5        PUMA, INTEGER, 5

NEW_ZONE_FIELD_6        AGE1, INTEGER, 5
NEW_ZONE_FIELD_7        AGE2, INTEGER, 5
NEW_ZONE_FIELD_8        AGE3, INTEGER, 5
NEW_ZONE_FIELD_9        AGE4, INTEGER, 5
NEW_ZONE_FIELD_10       AGE5, INTEGER, 5
NEW_ZONE_FIELD_11       AGE6, INTEGER, 5
NEW_ZONE_FIELD_12       AGE7, INTEGER, 5
NEW_ZONE_FIELD_13       AGE8, INTEGER, 5
NEW_ZONE_FIELD_14       SIZE1, INTEGER, 5
NEW_ZONE_FIELD_15       SIZE2, INTEGER, 5
NEW_ZONE_FIELD_16       SIZE3, INTEGER, 5
NEW_ZONE_FIELD_17       SIZE4, INTEGER, 5
NEW_ZONE_FIELD_18       SIZE5, INTEGER, 5
NEW_ZONE_FIELD_19       SIZE6, INTEGER, 5
NEW_ZONE_FIELD_20       SIZE7, INTEGER, 5

NEW_ZONE_FIELD_21       INCOME1, INTEGER, 5
NEW_ZONE_FIELD_22       INCOME2, INTEGER, 5
NEW_ZONE_FIELD_23       INCOME3, INTEGER, 5
NEW_ZONE_FIELD_24       INCOME4, INTEGER, 5
NEW_ZONE_FIELD_25       INCOME5, INTEGER, 5
NEW_ZONE_FIELD_26       INCOME6, INTEGER, 5
NEW_ZONE_FIELD_27       INCOME7, INTEGER, 5
NEW_ZONE_FIELD_28       INCOME8, INTEGER, 5
NEW_ZONE_FIELD_29       INCOME9, INTEGER, 5
NEW_ZONE_FIELD_30       INCOME10, INTEGER, 5
NEW_ZONE_FIELD_31       INCOME11, INTEGER, 5
NEW_ZONE_FIELD_32       INCOME12, INTEGER, 5
NEW_ZONE_FIELD_33       INCOME13, INTEGER, 5
NEW_ZONE_FIELD_34       INCOME14, INTEGER, 5
NEW_ZONE_FIELD_35       INCOME15, INTEGER, 5
NEW_ZONE_FIELD_36       INCOME16, INTEGER, 5

NEW_ZONE_FIELD_37       HOUSEHOLDS, INTEGER, 6

INPUT_COORDINATE_SYSTEM     LATLONG, DEGREES
OUTPUT_COORDINATE_SYSTEM    UTM, 18N, METERS
```

Data Preparation: The ZoneData Tool

```
Out.ZONE = atoi (Polygon2.STATE) * 100000 + Polygon2.INDEX
Out.PUMA = atoi (Polygon1.PUMA5)
IF (Out.PUMA == 0 OR Out.ZONE == 0) THEN
   RETURN (0)
ENDIF
Out.AGE1 = In.HH_AGE1 + In.NF_AGE1
Out.AGE2 = In.HH_AGE2 + In.NF_AGE2
Out.AGE3 = In.HH_AGE3 + In.NF_AGE3
…
Out.AGE8 = In.HH_AGE8 + In.NF_AGE8
Out.SIZE1 = In.NF_SIZE1
Out.SIZE2 = In.HH_SIZE2 + In.NF_SIZE2
Out.SIZE3 = In.HH_SIZE3 + In.NF_SIZE3
…
Out.SIZE7 = In.HH_SIZE7 + In.NF_SIZE7
Out.INCOME1 = In.HH_INC1 + In.NF_INC1
Out.INCOME2 = In.HH_INC2 + In.NF_INC2
…
Out.INCOME8 = In.HH_INC8 + In.NF_INC8
Out.INCOME9 = In.HH_INC9 + In.NF_INC9
Out.INCOME10 = In.HH_INC10 + In.NF_INC10
…
Out.INCOME16 = In.HH_INC16 + In.NF_INC16
Out.HOUSEHOLDS = Out.SIZE1 + Out.SIZE2 + Out.SIZE3 +
    Out.SIZE4 + Out.SIZE5 + Out.SIZE6 + Out.SIZE7
RETURN (1)
```

- This is an example ZoneData TRANSIMS script
- Creates a unique zone ID from state IDs and other GIS shape file attributes
- Filters illegal data entries out by returning a value of 0 to ZoneData
- Combines household aggregate data with non-household aggregate data
- Calculates household totals
- More details can be found in the PopSyn How-To document

Associating Activity Locations with CENSUS Data

```
TITLE                    Add Census IDs to Activity Locations
DEFAULT_FILE_FORMAT      TAB_DELIMITED
PROJECT_DIRECTORY        ../

#---- Input Files ----
NET_DIRECTORY            ../network
NET_ACTIVITY_LOCATION_TABLE   Activity_Location

BOUNDARY_POLYGON_1       census/arcview/p517_d00.shp
BOUNDARY_POLYGON_2       census/arcview/bg17_d00.shp

CONVERSION_SCRIPT        census/LocationData_Census_Script.txt

#---- Output Files ----

NEW_DIRECTORY            ../network
NEW_ACTIVITY_LOCATION_TABLE   Activity_Location_Final

#---- Parameters ----

COPY_EXISTING_FIELDS     YES
CREATE_NOTES_AND_NAME_FIELDS  YES

NEW_LOCATION_FIELD_1     STATE, STRING, 2
NEW_LOCATION_FIELD_2     PUMA, INTEGER, 10
NEW_LOCATION_FIELD_3     BG_ID, INTEGER, 10

INPUT_COORDINATE_SYSTEM      LATLONG, DEGREES
OUTPUT_COORDINATE_SYSTEM     UTM, 18N, METERS
```

- LocationData Control File
  - Reads an existing activity location data and copies all fields to a new activity location table
  - Creates three new fields in the output file to contain geographic attributes (e.g. membership in PUMAs, block groups, states …
  - The three fields are populated using a TRANSIMS script (see next page)

Associating Activity Locations with CENSUS Data

- TRANSIMS Script for LocationData

```
IF (In.ZONE > 0 AND In.ZONE < 1945) THEN
  Out.BG_ID = atoi (Polygon2.STATE) * 100000 + Polygon2.INDEX
  Out.PUMA = atoi (Polygon1.PUMA5)
  Out.STATE = Polygon2.STATE
ELSE
  Out.BG_ID = 0
  Out.PUMA = 0
  Out.STATE = ""
ENDIF

RETURN (1)
```

- Activity-based modeling makes only sense for “internal” traffic analysis zones
- In Chicago, we have 1945 internal zones
- The script can be modified and provides quite a level of flexibility

- Once the activity locations are aware of their geocoding, the population synthesizer can be run
- This step finalizes the data preparation procedure

The Population Synthesizer Control File

```
TITLE                    Population Synthesis
DEFAULT_FILE_FORMAT      TAB_DELIMITED
PROJECT_DIRECTORY        ../

#---- Input Files ----

NET_DIRECTORY            ../network

NET_ACTIVITY_LOCATION_TABLE  Activity_Location_Final
NET_PROCESS_LINK_TABLE       Process_Link

PUMS_HOUSEHOLD_FILE      Households
PUMS_POPULATION_FILE     Persons

ZONE_DATA_FILE           ZonesData

VEHICLE_TYPE_DISTRIBUTION    inputs/Vehicle_Distribution.txt

#---- Output Files ----

NEW_HOUSEHOLD_FILE       household/Households
NEW_POPULATION_FILE      household/Persons
NEW_VEHICLE_FILE         household/Vehicles
NEW_PROBLEM_FILE         Problems

POPSYN_REPORT_1          PUMS_HOUSEHOLD_SUMMARY
POPSYN_REPORT_2          PUMS_POPULATION_SUMMARY
POPSYN_REPORT_3          SYNTHETIC_HOUSEHOLD_SUMMARY
POPSYN_REPORT_4          SYNTHETIC_POPULATION_SUMMARY

#---- Parameters ----
STATE_PUMA_LIST          51100, 51200, 51301, 51302
PUMS_WEIGHT_FIELD        WEIGHT
##PUMS_VEHICLE_FIELD     VEHICLES
PUMS_AGE_FIELD           AGE
ZONE_DATA_ID_FIELD       ZONE
LOCATION_ZONE_FIELD      BG_ID

ZONE_TOTAL_FIELD_1       HOUSEHOLDS

PUMS_ATTRIBUTE_FIELD_1_1     INCOME
PUMS_ATTRIBUTE_BREAKS_1_1    10000, 15000, 20000, 25000,
                             30000, 35000, 40000, 50000, 60000, 75000, 100000,
                             125000, 150000, 200000
ZONE_FIELD_GROUP_1_1         INCOME

PUMS_ATTRIBUTE_FIELD_1_2     HHAGE
PUMS_ATTRIBUTE_BREAKS_1_2    25, 35, 45, 55, 65, 75, 85
ZONE_FIELD_GROUP_1_2         AGE

PUMS_ATTRIBUTE_FIELD_1_3     PERSONS
PUMS_ATTRIBUTE_BREAKS_1_3    2, 3, 4, 5, 6, 7
ZONE_FIELD_GROUP_1_3         SIZE

STARTING_HOUSEHOLD_ID    1
STARTING_VEHICLE_ID      1

RANDOM_NUMBER_SEED       12332
MAXIMUM_IPF_ITERATIONS   10000
MAXIMUM_IPF_DIFFERENCE   0.0000001

OUTPUT_HOUSEHOLD_FIELDS  HHOLD, STATE, PUMA,
                         PERSONS, VEHICLES, WORKERS, INCOME, HHAGE,
                         NUM_LT5, NUM_5TO15
OUTPUT_POPULATION_FIELDS HHOLD, PERSON, GENDER,
                         AGE, WORKER
```

Population Synthesizer Reports

- PUMS Household Summary

Attribute Name	Total	Average	Minimum	Maximum
PERSONS	32818.00	2.29	1.00	16.00
VEHICLES	21685.00	1.52	0.00	6.00
WORKERS	20836.00	1.46	0.00	9.00
INCOME	1178458950.00	82346.37	-10000.00	912000.00
HHAGE	640829.00	44.78	0.00	93.00
NUM_LT5	2091.00	0.15	0.00	6.00
NUM_5TO15	4011.00	0.28	0.00	6.00

- PUMS Population Summary

Attribute Name	Total	Average	Minimum	Maximum
AGE	1191231.00	36.30	0.00	93.00
GENDER	49545.00	1.51	1.00	2.00
WORKER	32596.00	0.99	0.00	2.00

Population Synthesizer Reports (cont)

- Synthetic Household Summary

Attribute Name	Total	Average	Minimum	Maximum
HHOLD	3098143686.00	39358.50	1.00	78716.00
LOCATION	325396060.00	4133.80	21.00	7734.00
PERSONS	229370.00	2.91	1.00	16.00
PUMA	15730832.00	199.84	100.00	302.00
VEHICLES	174389.00	2.22	0.00	11.00
WORKERS	134074.00	1.70	0.00	9.00
INCOME	6657460978.00	84575.70	0.00	912000.00
HHAGE	3550791.00	45.11	0.00	93.00
NUM_LT5	20727.00	0.26	0.00	6.00
NUM_5TO15	34254.00	0.44	0.00	6.00

- Synthetic Population Summary

Attribute Name	Total	Average	Minimum	Maximum
HHOLD	9195758639.00	40091.37	1.00	78716.00
PERSON	556402.00	2.43	1.00	16.00
AGE	7456021.00	32.51	0.00	93.00
GENDER	346911.00	1.51	1.00	2.00
WORKER	214704.00	0.94	0.00	2.00

Algorithms

IPF Algorithms (Traditional and Two-Step)

- The algorithm adopted by the TRANSIMS Population Synthesizer is based on two types of IPF algorithms:
  - The traditional IPF procedure proposed by Deming and Stephan (1940).
  - The two-step IPF procedure developed by Beckman (1996), known as the modified IPF procedure .
- Traditional Procedure fits only one block group at a time.
- Two-step Procedure can simultaneously consider all block groups that make up a PUMA.
- The two-step procedure makes use of the traditional IPF procedure in its analysis.

Three-Dimensional IPF Procedure

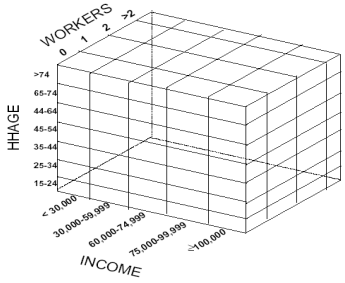
- The same algorithm can be extended for additional dimensions
- The following is a typical example

- Householde Income
  - INC<30,000
  - 30,000 < INC<60,000
  - 60,000 < INC<75,000
  - 75,000 < INC<100,000
  - INC>100,000

- Age of Householder
  - 15 >HHAGE< 24
  - 25 <HHAGE< 34
  - 35 <HHAGE< 44
  - 45 <HHAGE< 54
  - 55 <HHAGE< 64
  - 65 <HHAGE< 74
  - HHAGE>74

- Number of Workers
  - WORKERS = 0
  - WORKERS = 1
  - WORKERS = 2
  - WORKERS > 2

Three-Dimensional IPF Procedure

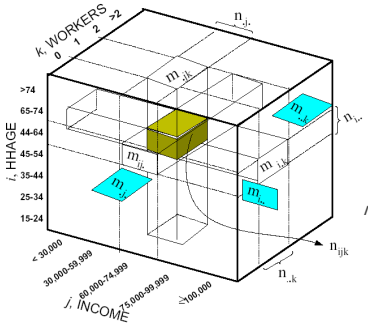


The generalized mathematical formulation developed by Deming *et al.* [1940] for the 3-dimensional matrix of 7×5×4 can be stated as follows:

$$m_{ijk}' = n_{ijk}\frac{m_{i..}}{n_{i..}} \quad \text{(I)}$$

$$m_{ijk}'' = m_{ijk}'\frac{m_{.j.}}{m_{.j.}'} \quad \text{(II)}$$

$$m_{ijk}''' = m_{ijk}''\frac{m_{..k}}{m_{..k}''}. \quad \text{(III)}$$



n = In general, it refers to a cell or a marginal value in the PUMS data,
m = In general, it refers to the marginal value of the STF-3A matrix,
n_{ijk} = Sample frequency from PUMS data falling in the cell n_{ijk} (please see the figure below for illustration),
$m_{.j.}'$ = Marginal data of the updated matrix (from PUMS) for the second (j) dimension,
$m_{..k}$ = Marginal data of the STF-3A file for the third (k) dimension, for example here the third (k) dimension is the WORKERS variable, and
$m_{..k}'$ = Marginal data of the updated matrix for the third (k) dimension.

$m..k$ = Marginal data of the STF-3A file for the third (k) dimension, WORKER
$m.j.'$ = Marginal data of the updated matrix (from PUMS) for the second (j) dimension,
$m..k''$ = Marginal data of the updated matrix for the third (k) dimension.

Two-Step (Modified) IPF

- TRANSIMS uses a two-step IPF procedure instead
- Traditional procedure fits only one block group at a time
- Beckman showed that fitting only one block group at a time may not be entirely correct
  - The sum of the block's STF-3A should also have the same correlation structure as the PUMS data, which equally represents all the blocks in a PUMA
- The two-step IPF procedure can simultaneously consider all block groups that make up the PUMA
- Details on the Two-Step IPF can be found in the TRANSIMS documentation, in particular the
  - The 2000 US CENSUS Data Preparation How-To
  - The Population Synthesizer How-To

PopSyn in a Nutshell

- Extract the household attribute classifications for each PUMA;
- Construct the PUMA cross-classification table for the selected household attributes;
- Retrieve the zone marginal totals for each identified attribute;
- Aggregate the marginal totals for all zones within a PUMA;
- Apply an IPF process to adjust the PUMA cross-classification table based on the aggregate marginal totals;
- Apply a second IPF process to estimate the cross-classification table for each zone using zone marginal totals and the adjusted PUMA cross-classification table as an additional marginal constraint;
- Randomly select a corresponding PUMS household for each household in the zone cross-classification table;
- Randomly assign the household to an activity location within the zone using the specified location weights; and
- Randomly select a vehicle type for each household vehicle and assign the vehicle to the parking lot attached to the household's activity location.

Credits and Acknowledgements

- GIS visualization materials were mostly developed at Argonne based on the TRANSIMS tools developed by AECOM for USDOT
- Chicago road and transit network data used in some of the examples was provided by the Chicago Metropolitan Agency for Planning
- USDOT provided the funding for the development of these training materials
- USDOT provided the funding for the TRACC computing center and the resources necessary to perform these training session
- Some materials have been developed for USDOT by Prof. Antoine Hobeika, Virginia Polytechnic Institute, Civil and Environmental Engineering
- The presentation is loosely based on materials provided by USDOT at a training course in November 2006, and has been updated using the information available in the How-To documentation from the TRANSIMS web site