Population Synthesizer How-To

This document provides basic information on preparing input data and synthesizing household populations used for TRANSIMS activity-based models. The input data files for Alexandria, Virginia are a subset of the data files provided on the 2000 U.S. Census Website.

Revision History

9/7/2007	Created by AECOM Consult, Inc.
8/9/2008	Updated by AECOM Consult, Inc.

Table of Contents

1.0 Assumptions and Prerequisites
- 1.1 Download Sample Datasets
- 1.2 Initialize the Alexandria Case Study

2.0 Introduction to the TRANSIMS Population Synthesizer

3.0 How to Prepare Household Survey Data
- 3.1 Download Census Data
- 3.2 Select Household and Population Attributes
- 3.3 Run PUMSPrep to Extract the Data

4.0 How to Prepare Zone Summary Data
- 4.1 Select Household Summary Tables
- 4.2 Download Census Data
- 4.3 Run SF3Prep to Extract the Data

5.0 How to Link Census Data to the TRANSIMS Network
- 5.1 Link and Format the Zone Summary Data
- 5.2 Add Census IDs to the Activity Location File

6.0 How to Run the TRANSIMS Population Synthesizer
- 6.1 The PopSyn Control File
- 6.2 Reviewing the Results

7.0 Troubleshooting

8.0 Frequently Asked Questions

1.0 Assumptions and Prerequisites

This document assumes you has installed TRANSIMS Version 4.03 on a Windows or Linux computer system and understand the basic procedures and terminology for executing TRANSIMS programs. Download the TRANSIMS software and documentation from www.transims-opensource.org or http://sourceforge.net/projects/transims. The Installation and Testing How-To addresses the basic TRANSIMS procedures and terminology.

This document also utilizes data distributed by the U.S. Census Bureau for the Year 2000 Census. The Public Use Microdata Sample (PUMS) provides complete household records for a five percent sample of households within Public Use Microdata Areas (PUMAs) that contain approximately 100,000 people. The Summary File 3 dataset provides tables of various household attributes summarized at several levels of geography. Block Groups are geographic subdivisions that contain approximately 1,000 persons.

Access to ArcGIS Version 9.2 and a standard text-editing tool (e.g., vi, Pico, WordPad) or other software that can manipulate text files (e.g., Excel) and ArcView shapefiles is helpful but not necessary.

1.1 Download Sample Datasets

This How-To document uses network and Census information for Alexandria, Virginia. You can download the Alexandria data from www.transims-opensource.org or http://sourceforge.net/projects/transims.

You should create a directory with a name such as

c:\TRANSIMS\Alexandria	(Windows)
/home/TRANSIMS/Alexandria	(Linux)

and unzip the downloaded file to this directory. This process generates the following subdirectories:

batch	
census	
control	
demand	
inputs	/arcview
network	/arcview
results	/arcview
setup	/arcview and /master
survey	

Most of these directories will initially be empty. Batch files found in the setup directory construct the necessary control files and datasets referenced by this case study. They assume the TRANSIMS Version 4.03 executable programs are located in a bin directory at the same level as the Alexandria directory. In this case:

c:\TRANSIMS\bin	(Windows)
/home/TRANSIMS/bin	(Linux)

If this is not the case, the path to the TRANSIMS programs should be added to the operating system's path environment or manually changed in the batch files.

1.2 Initialize the Alexandria Case Study

The setup directory includes a batch file called:

PopSyn.bat	(Windows)
./PopSyn.sh	(Linux)

to prepare the batch procedures and control files needed to execute the sequence of data processing steps described in this document. These procedures assume you have executed the network setup step prior to running the population synthesizer. The setup directory includes a batch file called:

ConvertNet.bat	(Windows)
./ConvertNet.sh	(Linux)

to prepare the batch procedures and control files for constructing the network. Running the batch file added to the batch directory executes the network preparation steps.

Alex.2005.Net.ConvertNet.bat	(Windows)
./Alex.2005.Net.ConvertNet.sh	(Linux)

You may also choose to run the entire population synthesizer process at this point using the batch file added to the batch directory

Alex.2000.Act.PopSyn.bat	(Windows)
./Alex.2000.Act.PopSyn.sh	(Linux)

or wait to execute each program as directed by the following discussion.

2.0 Introduction to the TRANSIMS Population Synthesizer

The PopSyn program uses Census data (or an equivalent small area population forecast) to synthesize households with a set of household characteristics that closely resemble the distribution of households found in a household survey (e.g., the Census PUMS dataset). The program uses a two-step Iterative Proportional Fitting (IPF) procedure to create a cross-classification of household characteristics that match the distribution of key household attributes specified for a specific location (e.g., zone or Block Group). Synthetic households are then drawn at random from the set of representative households in the same cross-classification. The synthetic household inherits all of the household and person characteristics of the selected survey household.

PopSyn allocates the synthetic households to a specific activity location and places the number of vehicles assigned to the household in the parking lot attached to the activity location. This allocation requires that each activity location include a field that associates the location with a geographic area (i.e., zone or Block Group). Each geographic area also needs a field that associates the area with the appropriate set of household survey records, in this case a specific PUMA. Most of this document describes the data processing procedures required to select the desired household attributes from the 2000 Census data and create the reference fields needed to link the Census data to the TRANSIMS network.

Once the input datasets are prepared, the PopSyn program uses the following steps to generate a synthetic population:

- 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.

Though not specifically addressed in this document, it is important to consider how a synthetic population will be generated for a future year when selecting the data items to include as control variables. Future forecasts require updated household type totals (marginals) for each geographic area (i.e., zone or Block Group). If the future forecasts of demographic attributes are limited to only a few household characteristics, it may be desirable to design the population synthesizer to focus only on these characteristics. Alternatively, the PopSyn program can use an existing distribution of household types and a simple forecast of total households by zone to update the marginals for each household attribute used to synthesize the future population.

3.0 How to Prepare Household Survey Data

The U.S. Census Bureau publishes a Public Use Microdata Sample (PUMS) containing complete household records from the 2000 Census for five percent of households within Public Use Microdata Areas (PUMAs). Since PUMAs contain approximately 100,000 people, the Census Bureau can report household specific data without revealing information about a specific household. No other Census dataset provides the disaggregate characteristics of individual households. This makes the PUMS dataset particularly useful for generating synthetic populations.

3.1 Download Census Data

PUMS data can be downloaded free of charge from the U.S. Census Bureau Website at http://www2.census.gov/census_2000/datasets/PUMS/FivePercent. This location provides technical documentation about the PUMS dataset and a subdirectory containing the data for each state. The Virginia subdirectory includes a file called PUMS5_51.TXT that contains the data for Alexandria and all other areas of Virginia. You may choose to download this file to the census subdirectory or use the file with the same name that is included in the Alexandria case study dataset. The case study version is a subset of the Census data file focused on Alexandria.

You may also wish to download the technical documentation from this site or consult the PUMS user manual at http://www.census.gov/prod/cen2000/doc/pums.pdf to determine the appropriate data items to extract from the dataset.

The most useful way of relating the PUMS data to other geography (e.g., Block Group or activity location) is to download the PUMS boundary file in ArcView shapefile format. The U.S. Census Bureau distributes PUMS shapefiles at http://www.census.gov/geo/www/cob/pu5_2000.html. The compressed file called p551_d00_shp.zip contains the Virginia shapefiles. This file can be downloaded and unzipped into the census/arcview directory. This will replace the p551_d00.shp files provided in the Alexandria case study dataset. The case study version is a subset of the Census data file focused on Alexandria. The coordinate system used in the Census files is GCS_North_American_1983.

3.2 Select Household and Population Attributes

The PUMS file contains two types of records: household records and person records. The household records start with an "H" identifier whereas the person records start with a "P" identifier. The SERIALNO field links the household and person records from a given household. Each record includes many fields of limited utility for a TRANSIMS application.

In order to use these data in the TRANSIMS population synthesizer, the household and person records need to be separated into a household file and a population file. The program that performs this task, PUMSPrep, also facilitates the selection and naming of data fields to transfer from the PUMS dataset to the TRANSIMS data files. The decision about which fields to keep in the TRANSIMS file relates to the data needed by the cross-classification table within the population synthesizer and the data needed by the activity generator to match travel survey households to synthetic households.

For this study, we need the following household variables for the population synthesizer and the activity generator:

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

The population variables include:

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

Most of these fields directly map to a data field in the PUMS dataset. Others, such as the age of the householder and the number of children of various ages, can be calculated based on data found in the person records. By studying the PUMS documentation, the following mapping between PUMS data fields and TRANSIMS data fields was established:

TRANSIMS	**PUMS Name**	**PUMS Columns**
[Household Fields]		
HHOLD	SERIALNO	2-8
STATE	STATE	10-11
PUMA	PUMA5	14-18
WEIGHT	HWEIGHT	102-105
PERSONS	PERSONS	106-107
VEHICLES	VEHICL	134-134
TYPE	HHT	213-213
INCOME	HINC	251-258
[Person Fields]		
HHOLD	SERIALNO	2-8
PERSON	PNUM	9-10
RELATE	RELATE	17-18
GENDER	SEX	23-23
AGE	AGE	25-26
WORKER	WRKLYR	236-236

Since the PUMS dataset includes the whole State of Virginia, it is also desirable to identify the list of PUMAs that cover the modeling area. Overlaying the PUMS boundary shapefile and the TRANSIMS network in ArcMap (or any other GIS software) is the simplest way of identifying the PUMAs of interest. (An ArcMap 9.2 project file called Census_Boundaries.mxd is included in the batch directory for your convenience). In this case, four PUMAs (100, 200, 301, and 302) represent the Alexandria case study area.

3.3 Run PUMSPrep to Extract the Data

The PUMSPrep program reads and manipulates one or more PUMS datasets (e.g., files for different states) into TRANSIMS household and population files. The PUMSPrep control file created for the Alexandria case study is located in the control directory with the filename Alex.2000.Act.PUMSPrep.ctl. This file includes the following records:

```
TITLE                       PUMS Data Extraction
DEFAULT_FILE_FORMAT         TAB_DELIMITED
PROJECT_DIRECTORY           ../

#---- Input Files ----
```

```
PUMS_DATA_FILE                census/PUMS5_51.TXT

PUMS_HOUSEHOLD_FILE           NULL
PUMS_POPULATION_FILE          NULL

#---- Output Files ----

NEW_PUMS_HOUSEHOLD_FILE       demand/Alex.2000.Act.PUMS_Households
NEW_PUMS_POPULATION_FILE      demand/Alex.2000.Act.PUMS_Persons

PUMSPREP_REPORT_1             CONVERSION_SCRIPT

#---- 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             inputs/PUMS_Script.txt
```

The STATE_PUMA_LIST defines the four PUMAs included in the output file. The key value is a comma-delimited list of STATE-PUMA codes. In this case, "51" is the FIPS code for Virginia followed by one of the selected Virginia PUMAs (100, 200, 301, and 302).

The household and population field keys identify the field names and data types for the fields added to the output household and population files. The two numbers at the end of each parameter define the starting column number and field length for data copied from the PUMS dataset to the corresponding data field. For example, the SERIALNO field in the PUMS dataset starts in column 2 and ends in column 8. This makes the field 7 characters long, which is why the HHOLD field starts in column 2 with a field length of 7.

Notice that the starting column for the HHAGE field is set to zero. This is a special case where a data field is added to the output file, but there is no field in the PUMS file that can be directly copied to the field. If this field is not specifically set by the conversion script, its output value will be zero.

The conversion script is a user program that processes the sequence of household and person records in the PUMS dataset one record at a time. All of the household and population fields

copied from the PUMS dataset are populated prior to calling the conversion script. The script has access to both the household record and the current person record each time it is called. This enables the script to calculate household attributes based on the characteristics of the people in the household. As the following script shows, this fact is used to population the number of workers, age of householder, and number of children less than 5 years old and between 5 and 15 years old fields.

```
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
```

The PUMSPrep program executes automatically as part of the Alex.2000.Act.PopSyn.bat(.sh) file installed in the batch directory. If you wish to execute the program manually, change to the control directory and type one of the following commands:

```
..\..\bin\PUMSPrep Alex.2000.Act.PUMSPrep.ctl          (Windows)
../../bin/PUMSPrep Alex.2000.Act.PUMSPrep.ctl          (Linux)
```

The output files are written to the demand directory. The first few records of the Alex.2000.Act.PUMS_Households file are shown below:

```
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
```

The first few records of the Alex.2000.Act.PUMS_Persons file are also shown. Note that the calculated fields in the household records accurately summarize the person information shown in the population 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
```

4.0 How to Prepare Zone Summary Data

The PUMS dataset contains detailed household information, but these data do not have geographic specificity. A small area zone geography file is required to distribute the detailed household information to specific activity locations within the TRANSIMS network. This file could be a standard regional zone file provided the zone demographic data include one or more distributions of key household attributes such as household size or income. Since this is often not the case, Census distributions of household attributes by Census Tract or Block Group are used. This section demonstrates how to convert Block Group data from the U.S. Census Summary File 3 to a TRANSIMS zone file for input to the population synthesizer.

4.1 Select Household Summary Tables

Summary File 3 (SF3) provides over 800 summary tables aggregated to 10 levels of geography and stored in 76 data files. Each summary table is identified with a code number and each value in the table is assigned an element number. The table index also has a segment number to identify in which of the 76 data files the table is located.

The first step in extracting data from the SF3 file is to identify the code and segment numbers for the required summary tables. The document found at

http://www2.census.gov/census_2000/datasets/Summary_File_3/0SF3_table_matrix.doc

provides a list of the tables, the variables included, and the table code and segment numbers. For the purposes of this demonstration, four summary tables were selected:

Household Type by Age of Householder
Household Type by Household Size
Family Income in 1999
Non-Family Income in 1999

The following lists the data values included in each table. The values to the right are the table element codes (e.g., P013001) and the segment number (e.g., 01). These are the code numbers used in the TRANSIMS conversion.

```
P13.      HOUSEHOLD TYPE BY AGE OF HOUSEHOLDER [19]
          Universe:  Households
          Total:                                                P013001   01 9
                Family households:                              P013002   01 9
                  Householder 15 to 24 years                    P013003   01 9
                  Householder 25 to 34 years                    P013004   01 9
                  Householder 35 to 44 years                    P013005   01 9
                  Householder 45 to 54 years                    P013006   01 9
                  Householder 55 to 64 years                    P013007   01 9
                  Householder 65 to 74 years                    P013008   01 9
                  Householder 75 to 84 years                    P013009   01 9
                  Householder 85 years and over                 P013010   01 9
                Nonfamily households:                           P013011   01 9
                  Householder 15 to 24 years                    P013012   01 9
                  Householder 25 to 34 years                    P013013   01 9
```

	Householder 35 to 44 years	P013014	01 9
	Householder 45 to 54 years	P013015	01 9
	Householder 55 to 64 years	P013016	01 9
	Householder 65 to 74 years	P013017	01 9
	Householder 75 to 84 years	P013018	01 9
	Householder 85 years and over	P013019	01 9

P14. HOUSEHOLD TYPE BY HOUSEHOLD SIZE [16]
Universe: Households

Total:	P014001	01 9
Family households:	P014002	01 9
2-person household	P014003	01 9
3-person household	P014004	01 9
4-person household	P014005	01 9
5-person household	P014006	01 9
6-person household	P014007	01 9
7-or-more-person household	P014008	01 9
Nonfamily households:	P014009	01 9
1-person household	P014010	01 9
2-person household	P014011	01 9
3-person household	P014012	01 9
4-person household	P014013	01 9
5-person household	P014014	01 9
6-person household	P014015	01 9
7-or-more-person household	P014016	01 9

P76. FAMILY INCOME IN 1999 [17]
Universe: Families

Total:	P076001	07 9
Less than $10,000	P076002	07 9
$10,000 to $14,999	P076003	07 9
$15,000 to $19,999	P076004	07 9
$20,000 to $24,999	P076005	07 9
$25,000 to $29,999	P076006	07 9
$30,000 to $34,999	P076007	07 9
$35,000 to $39,999	P076008	07 9
$40,000 to $44,999	P076009	07 9
$45,000 to $49,999	P076010	07 9
$50,000 to $59,999	P076011	07 9
$60,000 to $74,999	P076012	07 9
$75,000 to $99,999	P076013	07 9
$100,000 to $124,999	P076014	07 9
$125,000 to $149,999	P076015	07 9
$150,000 to $199,999	P076016	07 9
$200,000 or more	P076017	07 9

P79. NONFAMILY HOUSEHOLD INCOME IN 1999 [17]
Universe: Nonfamily households

Total:	P079001	07 9
Less than $10,000	P079002	07 9
$10,000 to $14,999	P079003	07 9
$15,000 to $19,999	P079004	07 9
$20,000 to $24,999	P079005	07 9
$25,000 to $29,999	P079006	07 9
$30,000 to $34,999	P079007	07 9
$35,000 to $39,999	P079008	07 9
$40,000 to $44,999	P079009	07 9
$45,000 to $49,999	P079010	07 9
$50,000 to $59,999	P079011	07 9
$60,000 to $74,999	P079012	07 9
$75,000 to $99,999	P079013	07 9
$100,000 to $124,999	P079014	07 9

```
$125,000 to $149,999                                    P079015   07 9
$150,000 to $199,999                                    P079016   07 9
$200,000 or more                                        P079017   07 9
```

Note that most SF3 tables summarize data for three household types: family households, non-family households, and group quarters. Family households are households with two or more related members; non-family households are households with unrelated persons (or living alone); and group quarters are dwellings such as college dormitories and prisons.

As mentioned earlier, each table in the SF3 file is aggregated to ten different summary levels or geographies. This application extracts data for summary level 90 – Block Groups.

Summary Level	Description
10	United States
20	Regions
30	Divisions
40	State
50	County
60	County Subdivision
70	Place
80	Census Tract
90	Block Group
100	Blocks

4.2 Download Census Data

Block Group summaries can be downloaded free of charge from the U.S. Census Bureau Website at http://www2.census.gov/census_2000/datasets/Summary_File_3. This site includes documentation about the SF3 file structure and subdirectories for each state. Within each state are the 76 data files corresponding to the 76 segment numbers. Of the four summary tables to be extracted, two tables (P013 and P014) are located in segment 1 and two tables (P076 and P079) are located in segment 7. This means that only files va00001_uf3.zip and va00007_uf3.zip need to be downloaded.

In addition to data tables, each state directory includes a geography file that defines relationships between record code numbers and the equivalent geography codes. This file, vageo_uf3.zip, is needed to map the SF3 records to a specific state and county code number in the record selection process. The geography file also contains the coordinates, in latitude-longitude million degrees, of an internal point within the geographic area. This coordinate is used to map the Block Group to the appropriate PUMA.

You may choose to download and unzip the three Virginia files (i.e., va00001_uf3.zip, va00007_uf3.zip, and vageo_uf3.zip) to the census subdirectory or use the files with the same name that are included in the Alexandria case study dataset. The case study version is a subset of the Census data files focused on Alexandria.

A shapefile containing the Block Group boundaries in Alexandria is provided in the census/arcview directory. This file is a subset of the Virginia Block Groups distributed by the U.S. Census Bureau at http://www.census.gov/geo/www/cob/bdy_files.html. In this Website, select the year 2000 option for the Census Block Groups listing and then select the ArcView shapefile data for Virginia. The compressed file called bg51_d00_shp.zip can be downloaded and unzipped into the census/arcview directory. This will replace the bg51_d00.shp files provided in the Alexandria case study dataset. The coordinate system used in the Census files is GCS_North_American_1983.

4.3 Run SF3Prep to Extract the Data

The SF3Prep program reads and manipulates one or more sets of SF3 datasets (e.g., files for different states and/or multiple table files) into a TRANSIMS zone data file. The SF3Prep control file created for the Alexandria case study is located in the control directory with the filename Alex.2000.Act.SF3Prep.ctl. This file includes the following records:

```
TITLE                          SF3 Data Extraction
DEFAULT_FILE_FORMAT            TAB_DELIMITED
PROJECT_DIRECTORY              ../

#---- Input Files ----

SF3_GEOGRAPHY_FILE_1           census/vageo.uf3
SF3_SEGMENT_FILE_1_1           census/va00001.uf3
SF3_SEGMENT_FILE_1_7           census/va00007.uf3

ZONE_DATA_FILE                 NULL

#---- Output Files ----

NEW_ZONE_DATA_FILE             demand/Alex.2000.Act.SF3_Zone_Data

#---- 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, 18N, METERS
```

The STATE_COUNTY_LIST defines the three counties included in the output file. The key value is a comma-delimited list of STATE-COUNTY codes. In this case, “51” is the FIPS code for Virginia followed by the FIPS code for a Virginia county (013 = Arlington, 059 = Fairfax, and 510 = Alexandria). These codes can be identified using the ArcMap 9.2 project file called

Census_Boundaries.mxd included in the batch directory or by searching for county names in the Census documentation.

SF3_SUMMARY_LEVEL = 90 selects summary tables by Block Group. The input and output coordinate system keys instructs the program to convert the centroid of each Block Group from Latitude-Longitude in millions of degrees to the UTM coordinate system used by the TRANSIMS network.

The zone data field range keys map data fields in the SF3 file to data fields in the output TRANSIMS zone file. The first four fields simply copy the state, county, Census Tract, and Block Group codes from the SF3 geography file to the same name in the zone data file. The next six keys copy a range of data fields from specified summary tables to a new set of field names and range values. For example, P013003..10 refers to the values 3 through 10 in table P013. These are the eight household age range values summarized for family household types. These values are output with field names HH_AGE1..HH_AGE8.

The SF3Prep program executes automatically as part of the Alex.2000.Act.PopSyn.bat(.sh) file installed in the batch directory. If you wish to execute the program manually, change to the control directory and type one of the following commands:

```
..\..\bin\SF3Prep Alex.2000.Act.SF3Prep.ctl      (Windows)
../../bin/SF3Prep Alex.2000.Act.SF3Prep.ctl      (Linux)
```

The output file, Alex.2000.Act.SF3_Zone_Data, is written to the demand directory.

5.0 How to Link Census Data to the TRANSIMS Network

At this point in the process, the PUMS household survey and Block Group summary tables are converted to TRANSIMS format, but the Block Group data is not associated with a PUMA and the TRANSIMS activity locations are not associated with a Block Group. This section demonstrates how the ArcView shapefiles are used to post relational indices in each file.

5.1 Link and Format the Zone Summary Data

The ZoneData program adds and manipulates data fields and implements spatial joins within a TRANSIMS zone file. The ZoneData control file Alex.2000.Act.ZoneData_Census.ctl is included in the control directory for adding geographic references and manipulating data field in the Block Group zone file created by SF3Prep. This file contains the following records:

```
TITLE                       BlockGroup Data Manipulation
DEFAULT_FILE_FORMAT         TAB_DELIMITED
PROJECT_DIRECTORY           ../

#---- Input Files ----

NET_DIRECTORY               ../demand
NET_ZONE_TABLE              Alex.2000.Act.SF3_Zone_Data

BOUNDARY_POLYGON_1          census/arcview/p551_d00.shp
```

```
BOUNDARY_POLYGON_2            census/arcview/bg51_d00.shp

CONVERSION_SCRIPT             inputs/BlockGroup_Script.txt

#---- Output Files ----

NEW_DIRECTORY                 ../demand
NEW_ZONE_TABLE                Alex.2000.Act.Zone_Data

ZONEDATA_REPORT_1             CONVERSION_SCRIPT

#---- 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
```

This application reads the two Census boundary files included in the census/arcview directory and redefines all of the data fields generated by the SF3Prep program. The coordinate system

keys indicate that the input shapefiles use latitude-longitude coordinates in millions of degrees. The program converts these coordinates to UTM coordinates to implement the point-in-polygon procedure with the X-Y coordinates in the input zone file. The conversion script then does all of the data manipulation work. As listed below, this script creates a unique zone ID by combining the state code with the record index from the Block Group shapefile and adds the PUMA ID from the PUMS shapefile. If these indices are non-zero, it then creates age, household size, and income distributions by combining the SF3 data for family and non-family household types.

```
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.AGE4 = In.HH_AGE4 + In.NF_AGE4
Out.AGE5 = In.HH_AGE5 + In.NF_AGE5
Out.AGE6 = In.HH_AGE6 + In.NF_AGE6
Out.AGE7 = In.HH_AGE7 + In.NF_AGE7
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.SIZE4 = In.HH_SIZE4 + In.NF_SIZE4
Out.SIZE5 = In.HH_SIZE5 + In.NF_SIZE5
Out.SIZE6 = In.HH_SIZE6 + In.NF_SIZE6
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.INCOME3 = In.HH_INC3 + In.NF_INC3
Out.INCOME4 = In.HH_INC4 + In.NF_INC4
Out.INCOME5 = In.HH_INC5 + In.NF_INC5
Out.INCOME6 = In.HH_INC6 + In.NF_INC6
Out.INCOME7 = In.HH_INC7 + In.NF_INC7
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.INCOME11 = In.HH_INC11 + In.NF_INC11
Out.INCOME12 = In.HH_INC12 + In.NF_INC12
Out.INCOME13 = In.HH_INC13 + In.NF_INC13
Out.INCOME14 = In.HH_INC14 + In.NF_INC14
Out.INCOME15 = In.HH_INC15 + In.NF_INC15
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)
```

The ZoneData program executes automatically as part of the Alex.2000.Act.PopSyn.bat(.sh) file installed in the batch directory. If you wish to execute the program manually, change to the control directory and type one of the following commands:

..\..\bin\ZoneData .Alex.2000.Act.ZoneData_Census.ctl (Windows)
../../bin/ZoneData Alex.2000.Act.ZoneData_Census.ctl (Linux)

The output file, Alex.2000.Act.Zone_Data, is written to the demand directory.

5.2 Add Census IDs to the Activity Location File

The LocationData program adds and manipulates data fields and implements spatial joins within a TRANSIMS activity location file. The LocationData control file Alex.2000.Act.LocationData.ctl is included in the control directory for adding geographic references to the Alexandria activity location file. This file contains the following records:

```
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_2

BOUNDARY_POLYGON_1             census/arcview/p551_d00.shp
BOUNDARY_POLYGON_2             census/arcview/bg51_d00.shp

CONVERSION_SCRIPT              inputs/LocationData_Census_Script.txt

#---- Output Files ----

NEW_DIRECTORY                  ../network
NEW_ACTIVITY_LOCATION_TABLE    Activity_Location_4

LOCATIONDATA_REPORT_1          CONVERSION_SCRIPT

#---- 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
```

This application reads the two Census boundary files included in the census/arcview directory and adds three new data fields to the activity location file. The coordinate system keys indicate that the input shapefiles use latitude-longitude coordinates in millions of degrees. The program converts these coordinates to UTM coordinates to implement the point-in-polygon procedure with the X-Y coordinates in the activity location file. The conversion script uses the matched records to set the new data fields. As listed below, this script creates a unique Block Group ID by combining the state code with the record index from the Block Group shapefile and adds the PUMA ID from the PUMS shapefile. These fields are set only if the zone number assigned to the activity location is one of the internal zone numbers.

```
IF (In.ZONE > 0 AND In.ZONE < 70) 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)
```

The LocationData program executes automatically as part of the Alex.2000.Act.PopSyn.bat(.sh) file installed in the batch directory. If you wish to execute the program manually, change to the control directory and type one of the following commands:

..\..\bin\LocationData Alex.2000.Act.LocationData.ctl	(Windows)
../../bin/LocationData Alex.2000.Act.LocationData.ctl	(Linux)

The output file, Activity_Location_4, is written to the network directory.

6.0 How to Run the TRANSIMS Population Synthesizer

The population synthesizer takes marginal distributions of key demographic attributes by zone (e.g., Block Group) and detailed household information from the PUMS household survey to synthesize households, population, and vehicles at TRANSIMS activity locations. This section discusses the inputs and outputs of the PopSyn program.

6.1 The PopSyn Control File

The PopSyn control file Alex.2000.Act.PopSyn.ctl is included in the control directory. This file contains the following records:

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

#---- Input Files ----

NET_DIRECTORY                  ../network

NET_ACTIVITY_LOCATION_TABLE    Activity_Location_4
NET_PROCESS_LINK_TABLE         Process_Link_2

PUMS_HOUSEHOLD_FILE            demand/Alex.2000.Act.PUMS_Households
PUMS_POPULATION_FILE           demand/Alex.2000.Act.PUMS_Persons

ZONE_DATA_FILE                 demand/Alex.2000.Act.Zone_Data

VEHICLE_TYPE_DISTRIBUTION      inputs/Vehicle_Distribution.txt

#---- Output Files ----

NEW_HOUSEHOLD_FILE             demand/Alex.2000.Act.Households
NEW_POPULATION_FILE            demand/Alex.2000.Act.Persons
```

```
NEW_VEHICLE_FILE              demand/Alex.2000.Act.Vehicles
NEW_PROBLEM_FILE              results/Alex.2000.Act.PopSyn_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

LOCATION_WARNING_FLAG         FALSE
```

The control parameters identify the list of STATE-PUMA codes to process, the weight field that expands the survey household to the PUMA universe, and the field names that link the PUMS, zone, and activity location files. Note also that two options are available for generating vehicles. The PUMS_VEHICLE_FIELD specifies the number of vehicles generated for each household. Alternatively, the PUMS_AGE_FIELD generates a vehicle for each household member greater than or equal to 16 years of age. This option can help to avoid vehicle availability problems in the activity generator by providing each potential driver with a vehicle.

The PopSyn program can process several types of households simultaneously (e.g., family households, non-family households, and group quarters). This application, however, processes all of the households as a single type. In other words, all of the ZONE_TOTAL_FIELD_*, PUMS_ATTRIBUTE_FIELD_*, and ZONE_FIELD_GROUP_* keys has "_1" as the first

value. The ZONE_TOTAL_FIELD_1 specifies that the total number of households generated is defined by the HOUSHOLDS field in the zone data file. If this were a future year model application, a different field (e.g., HH2020) might be specified. All of the distribution values will be factored to the new household total.

The PUMS_ATTRIBUTE_FIELD_*, PUMS_ATTRIBUTE_BREAKS_*, and ZONE_FIELD_GROUP_* fields define the cross-classification table for the household match. In this case, three control variables are specified: household income, the age of the householder, and household size. The field keys identify the field names of these variables in the PUMS and zone files. The breaks keys define the break points between attribute values in the PUMS file that corresponds to a summary category in the zone file. For example, age breaks are defined at 25, 35, 45, 55, 65, 75, and 85 years of age. This corresponds to the summary categories provided in the SF3 table P013. The program will expect to find eight fields in the zone data file corresponding to these age breaks with field names AGE1, AGE2, … AGE8.

The output household and population fields are a list of fields from the input PUMS household and population files that are copied to the output household and population files. Note that variables can be used in the population synthesis process, but not saved to the output files.

6.2 Reviewing the Results

The PopSyn program executes automatically as part of the Alex.2000.Act.PopSyn.bat(.sh) file installed in the batch directory. If you wish to execute the program manually, change to the control directory and type one of the following commands:

```
..\..\bin\PopSyn Alex.2000.Act.PopSyn.ctl        (Windows)
../../bin/PopSyn Alex.2000.Act.PopSyn.ctl        (Linux)
```

In either case, a printout file called Alex.2000.Act.PopSyn.prn is created in the control directory. First check the household, population, and vehicles generated by each household model. This report lists the households generated within each PUMA. Since the vast majority of the Alexandria activity locations are within PUMA 200, these results are logical.

```
Household Model #1
       PUMA 51100 Households = 8967, Population = 26189, Vehicles = 20415
       PUMA 51200 Households = 61024, Population = 173255, Vehicles = 131874
       PUMA 51301 Households = 5618, Population = 19674, Vehicles = 14439
       PUMA 51302 Households = 3107, Population = 10252, Vehicles = 7661
Model #1 Total Households = 78716, Population = 229370, Vehicles = 174389
```

Now check the average attribute values found in the PUMS household and population files with the average attribute values found in the synthetic population. You would expect the number of vehicles per household to be higher because this application generated a vehicle for each driver, but the other values are not as similar as might normally be expected. This is mostly due to the limits of the Alexandria study area. A regional application is much less likely to have major portions of several PUMAs outside of the geographic area of the model. What these values show is how Alexandria is different from other areas in Northern Virginia.

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

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

7.0 Troubleshooting

The PopSyn program typically does not generate a significant number of problems. Activity location problems may occur if the locations are outside of the area covered by the Block Group or PUMS data.

A more likely problem is a mismatch between the number of fields implied by the PUMS attribute breaks and the number of fields in the zone data file. There should always be one less value in the breaks key then in the zone data file. The program assumes the first zone data field corresponds to the data less than the first break value. It also assumes that the last field in the zone data file corresponds to all data greater than or equal to the last break value. For example, income breaks of 20,000, 40,000, and 80,000 require four income fields in the zone data file.

INCOME1 = 0 to 20,000; INCOME2 = 20,000 to 40,000; INCOME3 = 40,000 to 80,000; and INCOME4 = 80,000+.

8.0 Frequently Asked Questions

How do I create a synthetic population for a different analysis year?

> First, a socio-economic forecasting process estimates the total number of households or the number of households by one or more demographic attributes (e.g., households by income quartile). These forecasts are distributed to traffic analysis zones or to Block Groups and posted on the zone data file input to the population synthesizer. If only the total number of households is changed, only the ZONE_TOTAL_FIELD_* in the PopSyn control file need be updated to point to the future household total. If the distribution of households by a demographic attribute changes, the distribution fields for that attribute in the zone data file need to be updated.