GTAModel V4.0 Population Synthesis

Overview

In order to run GTAModelV4, a ~~forecast year synthetic population is must be provided~~synthetic population for the forecast year must be provided. The algorithm (implemented within XTMF) ~~that is~~ currently used to create the synthetic ~~populations~~population is an unsophisticated reallocation model based ~~upon~~on households found in the TTS 2011 survey. ~~Simply~~Put simply, the households in each planning district are ~~are~~sampled without replacement ~~from~~to fill each zone's forecasted population total. Important demographic information is scraped from the synthetic population and ~~provided in addition to the~~is provided in addition to generated input files Households.csv and Persons.csv files.

This approach comes with ~~multiple~~several disadvantages. For one~~, each planning district's population~~, the population of each planning district ~~are~~is treated as a single homogenous source. This can lead to several issues – one example being that dramatically increasing a zone's population density will not affect which households are selected to populate it. ~~to issues such as in the case of dramatically increasing the population density of a single zone will not change the household structures that would be selected for it.~~ This could cause an over prediction in the number of vehicles and children for the zone as the dwelling types that would be available are not taken into consideration. A second major issue is that the algorithm requires access to the raw TTS population records. This restricts the algorithm's ~~its~~execution to machines that are on the DMG network, behind the firewall.

It is TMG's hope to ~~eventually~~replace this procedure with a superior tool to reduce~~in the future to reduce~~ the dependency ~~on TMG~~on TMG for producing all future year populations. However, it is not recommended to use~~It is not recommended however to take the current~~ GTAModelV4.0.1 with another synthesis procedure ~~and use another procedure~~until the results have undergone extensive validation with the rest of the model system.

Input Files

~~Besides~~Apart from the input files already stored on the DMG machine that runs the algorithm, two files of importance are required. The first file is a ~~is a~~TASHA zonal configuration ~~file~~'Zones.csv', which is required to describe the relationship between zones, planning districts, and loading ~~in~~of the TTS population. The second ~~file~~is a CSV file containing a mapping of the zone (as described in 'Zones.csv') ~~and~~with the future-year population total. The file must contain a header with two columns: ~~the columns being~~"Zone,Population". This file should not ~~contain any commas besides for delaminating the columns~~contain any commas, apart from the delimiter. If you are using Microsoft Excel, please use the MS-DOS CSV file format.

Algorithm

The source code for the algorithm for XTMF 1.3 is publicly available at
https://github.com/TravelModellingGroup/XTMF/blob/XTMF-1.3/Tasha/PopulationSynthesis/TTSPopulationReplicationForcast.cs.

At the start, ~~To start with~~ each household is loaded from the TTS and grouped by the planning district of residency. Each household then stores its own household ~~household~~ expansion factor ~~from~~ taken from TTS. Each person's expansion factor is then normalized within the household so that the sum of each factor ~~of them~~ is equal to the ~~all add up to the~~ population ~~within the~~of the household. Once the population is ready, a parallel process chooses households to assign to each planning district and zone. ~~each planning district is processed in parallel selecting which household structures to assign to each zone. Selection of each household is then processed~~The household selection is performed household-by-household. Each zone within the planning district is iterated over – randomly assigning a household taken from the planning district's pool.

~~one at a time within the planning district by iteration through each zone that requires additional population and randomly assigning a household from the planning district's pool.~~ The random selection is weighted by the expansion factor of the households. If the zone has enough space for the household it is selected and its expansion factor is reduced by one. If the planning district's household pool is depleted~~has no more households available~~ (~~,~~ as would be expected when synthesizing a future year scenario)~~,~~, the expansion factors are restored back to the ~~original household expansion factor~~original value.

Once the population has been selected, the worker categories and occupation totals by zone are collected and stored.

Output Files

The results of the algorithm are household and persons files as described below as well as information for each occupation/employment category required for the Place-Of-Residence-Place-Of-Work models. In order to produce a forecast scenario for GTAModel V4.0.1 additional work is required. The 'Zones.csv' must be updated with the population from the supplied forecast population vector, and the total employment by zone must be produced. The particulars for this ~~change~~ are changed in GTAModel V4.0.2.

Households.csv

This file contains the household level information for each synthetic household with the following columns: HouseholdId, Zone, ExpansionFactor, DwellingType, NumberOfPersons, NumberOfVehicles.

Persons.csv

This file contains the information that describes the individuals in the households with the following columns: HouseholdId, PersonNumber, Age, Sex, License, TransitPass, EmploymentStatus, Occupation, FreeParking, StudentStatus, EmploymentZone, SchoolZone, ExpansionFactor. It is important to note that the EmploymentZone and SchoolZones will only contain a zone number if they are external to the zone system otherwise they will just contain a value of 0.

Worker Categories

This directory contains a file for each occupation employment status category containing the zone, Worker Category pair, and the sum of individuals meeting that criteria (Zone, Worker Category, Ratio of persons in the category). A worker category describes the accessibility to vehicles of persons. A value of 0 would represent a person who does not have a driver's license or a household without a car. 1 would represent a person who lives in a household with at least one car, but where there a more people with a driver's license than the number of cars. 2 is the saturated situation, where there is at least one car per person.

Zonal Residence

This directory contains a file for each occupation employment category containing the total number of persons in the category per zone (Zone, Persons).