APPENDIX H

POPULATION ANALYSIS METHODOLOGY

POPULATION ANALYSIS METHODOLOGY

New York/New Jersey/Philadelphia Metropolitan Airspace Redesign Project EIS

11/28/2005

Table of Contents

Introduction

The methodology, analysis and forecast population for the New York/New Jersey/Philadelphia Airspace Enhancement EIS was developed from January 2005 to September 2005. To support noise impact analysis in the EIS, future year population projections to the Census Block (CB) level of detail were required. The forecast effort was required since the 2000 U.S. Census information is not considered to be current enough to support a reasonable analysis of the potential noise impacts for the EIS alternatives on the study area population. This analysis provides CB forecasted population information for all CBs in the study area for years 2006 and 2011.

Overall, the study area consisted of more than 400,000 Census Blocks. Once blocks with no population were taken out, the actual number of blocks analyzed was 323,708.

The initial forecasting effort focused on only forecasting Total Population by CB. Due to the limited demographic information available at the CB level of geography, study area projections were initially collected only at the census block group level (BG). Subsequently, demographic information collected at the Census Block Group (BG) level was dispersed down to the Census Blocks (CBs) located within each Block Group.

The methodology utilized for this study is an enhancement of the methodologies used for similar previous studies carried out for the Potomac Consolidated TRACON Airspace Redesign EIS and the Midwest Airspace Enhancement EA.

Data

The major items of data utilized in this report consist of:

- Census Geography
- Census Data
- Population Projection Data
- Land Use Geography

Census Geography

Census Geography was obtained from Environmental Systems Research Institute (ESRI). This consisted of 2000 Census Block and Block Group boundary shapefiles. Since the population projection data assumed no change in Census Geography from the year 2000 geography – these shapefiles were consistent with all the projection data.

A key element of these shapefiles is that they contained the Census Block Number As an identifier for each Census Block shape. In addition it is possible to derive the Census Block Group from the Census Block Identifier. Thus, the population projection data can be linked to the Census Geography through these identifiers. In consequence, the population projection data can then be linked to the original Census Data.

A second Census Geography dataset was also collected. This consisted of 1990 Census Block and Block Group boundary shapefiles from ESRI. These files were utilized for fine-tuning the population distribution algorithms described in the Analysis section of this paper.

Census Data

Census Data consisting of Population by Census Block for year 2000 was also obtained from ESRI. This data included Total Population for each individual Census Block, which was a key ingredient in the analysis.

In addition Census data for year 1990 was also collected to augment the analysis. This data was collected from the Social Science & Government Data Library of the University of California, Berkeley.

Population Projection Data

The population analysis for the EA required population projection data for the years - 2006 and 2011.

Following up on the previous work done, a quick review indicated that the situation described in those reports (about the available data sources) for this kind of data was still valid. Thus, the same vendor – Applied Geographic Solutions (AGS), was selected.

AGS was able to provide Total Population data at the Census Block Group Level for the years 2004/2009/2014. Thus, the data would need to be dispersed down to the Census Block level.

Land Use

Land use is a critical element for successful dispersion of the Census Block Group data down to the Census Block data. The key issue is distinguishing residential from non-residential land use so that correct population predictions can be made as to where the population is likely to be located in the future. Land use data was obtained from the US Geological Survey's (USGS) National Land Cover Dataset (NLCD). The NLCD was released in 2000 and used satellite imagery collected in the mid-1990s. The imagery was merged with local data to come up with 22 land use classifications. Although, the NLCD provides a good base for data analysis it was felt for this study that better local land use data would improve the quality of analysis. Thus, the study area was divided into multiple sub-areas for the purposes of land use. This is shown below:

Land Use Data Source by Study Sub-Areas

Study Sub-Area	Land Use Data Source
New York City	NYC Dept. of City Planning Land Use Data – 8/2004
Westchester County	Westchester County Generalized Land Use – 1996
City of Philadelphia	Delaware Valley Regional Planning Commission GIS Land Use – 2000
New Jersey	New Jersey Dept. of Environmental Protection Land Use – 1995/97
Connecticut	USGS NLCD
Delaware	USGS NLCD
New York	USGS NLCD
Pennsylvania	USGS NLCD

This analysis included an effort to identify all land use areas that could be expected to stay non-residential for the duration of the study period (to 2011). These include Parks, Wetlands, Federal Lands, Historical landmarks, etc. These were separately obtained and processed and used to set an upper limit on how much land use change could potentially occur in individual blocks. The following table shows the types of data and their sources for this kind of land use:

Land Use Data Source for Permanent Non-Residential Land Use

Land Use Type	Land Use Data Source
Wetlands	NJ Dept of Environmental Protection & US FWS National Wetlands Inventory
Parks	ESRI/GDT
Federal Lands	ESRI/USGS National Atlas Federal & Indian Land Areas
Landmarks	ESRI/GDT
Water	ESRI

Analysis

This section consists of a description of analysis steps, key algorithms and data transformations utilized in the analysis.

Steps

1 **Source Data Metadata Completion:** Obtain all source data from the variety of sources identified in the data section. Complete metadata on all source data (where available). Conform to FGDC format if it is geographic data. This ensures that we can clearly document our data sources.

2 **Source Data Re-projection:** Re-project all source data if necessary so that we are dealing with one spatial projection for all subsequent analysis.

3 **Source Data Editing:** Perform limited editing on source data. Because of the size of the study area and the nature of the data, it is not possible to do large scale data editing and quality checking. However, certain obvious flaws can be removed.

4 **Geographical Dataset Preparation:** Multiple datasets were prepared. These consisted of merging the local area datasets such as data from New York City, Westchester County, etc. with the larger study area based on NLCD data. The preparation step consisted of clipping and merging the geographical areas to get just the needed datasets of census and land use geography.

5 **Land Use Data Transformation:** Land use classifications came in different varieties based on data source. These were all conformed to a three class scheme – residential single-family, residential multi-family and non-residential based on the descriptions of each original land use class. For instance, if the original classes of land use were agricultural or industrial, they were all re-

classed as non-residential. Finally, non-residential land use such as parks & wetlands were separately merged into this data creating a fourth category – 'permanent non-residential land-use'. These areas were forecast to not change to residential land-use during the study period irrespective of how much growth took place in their vicinity.

6 **Census Block Residential/Non-Residential Area Calculation:** The land use data is spatially merged into the census block geographies to get residential/non-residential land use area per census block.

7 **Census Block Group Geography and Population Forecast Data Processing:** The Census Block Group is extracted from the census block identifier and all census blocks belonging to a census block group are merged spatially to form Census Block Group geographies. In addition, relevant data such as residential land area is aggregated from the individual census blocks to get data for the block groups. The Census Block Group geography is then attached to the initial population forecast data from AGS using the Block Group identifier as the relational key.

8 **Forecast Population Dispersal Block Group to Block:** The population block group forecast data is dispersed down to the census block based on an enhancement of the original 'Population Change Dispersal' algorithm used for the Midwest EA study. The modified algorithm is described later on in this section.

9 **Interpolation Forecast years to Study years:** Since the available forecast data (2004/2009/2014) did not match the study year needs (2006/2011) it was interpolated using an interpolation algorithm described elsewhere in the document.

10 **Final Data File Preparation:** In this step the individual census block centroids (a geographical point in the middle of the census block) were calculated. Finally, the data was exported to a .dbf file format, file descriptions were attached and data and descriptions were transmitted.

Algorithms

Population Dispersal Block Group to Block

The problem here is to disperse population data available at a larger geographical area to a number of smaller geographical units within it. This is necessary since the population forecast data is not available for the smaller unit (census block) but is only available for the larger unit (census block group).

The Equal Area & the Equal Area with Control Zones Methods: The Equal Area method simply takes a population at the block group level and prorates it to the individual blocks based on their individual areas. The problem with this approach is that it does not take into account existing land use at all. Thus, if the largest area block within a block group happens to be a park – the Equal Area method would still assign the largest percentage of population to it. Instead, the Equal Areas with Control Zones method could be used. This method takes into account current land use and distributes population based on residential land use area, not the total area of each block. Thus, in the above example although the block has a large total area, it would have a small or none residential land use area. Therefore, it would get very little or no population distributed to it.

Population Change Dispersal using Equal Areas with Control Zones Method: The Equal Areas with Control Zones method needs to take into account current population densities in addition to current land use - especially for low population growth areas such as inner cities. A conservative option is to only distribute the change in population based on residential land use. This method is called Population Change Dispersal using Equal Areas with Control Zones. The Population Change Dispersal method distributes just the *change* (not the total) in the Block Group's population from the previous period according to existing land use. Thus, if a block group is forecasted to undergo no population change, the individual blocks within the block group will keep their current population. This technique results in smooth transitions in population within block groups. It may result in an over-estimation of population for very high growth areas. In any case, very high growth areas would tend to see a change in current land use as well and would require a technique to project land use change within individual block groups. This analytical modification was developed for the current New York/New Jersey/Philadelphia Airspace study.

Population Change Dispersal using Equal Areas with Control Zones and Population Growth Categories Method In order to predict at what threshold of population increase land-use change within a block group would be likely – the 1990 census geographies were compared to the 2000 census geographies for the study area at the block group level. The analysis showed that overwhelmingly – the block group geographies changed due to population increase. Typically, a large block group in 1990 would be split up into multiple smaller block groups as the population in the block group changed significantly and new roads, etc were added. This is in keeping with census practices. Thus, change in census geography could be used as a surrogate for change in land use. A statistical spatial analysis of population vs. census geography change between 1990 and 2000 then yielded 20% change in population as the critical threshold in predicting land use change. This threshold was utilized to categorize the block groups as low population change and high population change block groups. The population dispersal techniques used for low change block groups were different from the ones used for high change block groups. These are described as follows:

- **Low predicted population change algorithm:** For block groups categorized as low forecast population change (<20% over 10 years), the methodology of 'Population Change Dispersal using Equal Areas with Control Zones' described above was utilized. The rationale for this being that no significant land-use change during the study period was being forecast for these block groups and so the use of existing residential land use as the determining factor in how population would change was appropriate.

- **High predicted population change algorithm:** For block groups categorized as high forecast population change (>20% over 10 years), the methodology of 'Population Change Dispersal using Equal Areas with Control Zones' described above was not utilized. This strategy was modified to use not the current residential area of the block. Instead we utilized the potential maximum area within the block that could change to residential in the future. The rationale for this being that significant land-use change during the study period was being forecast for these block groups and so the use of existing residential land use as the determining factor in how population would change was not an appropriate gauge. Instead how much land was potentially available for residential development was a better indicator.

Forecast Year Interpolation

Forecast year interpolation was required since the data obtained from the vendors did not conform to the study year needs. The technique used for interpolation is described below:

Since 4 data points were available in all cases (for each individual census block), a cubic polynomial fit of the form:

$$ax^3 + bx^2 + cx + d$$

would pass through all the known points exactly. A Least-Square fit using this form of the equation was applied to each census block's data, resulting in a unique prediction equation (with known a, b, c and d values) for each census block. The unknown years were then calculated using this equation.

This form of interpolation works quite well when the data requiring interpolation is close to a known point. In our case all unknown years were within two years of a known year. In addition, the rate of change in population for most census blocks is not very large. Thus, this form of interpolation is a good approximation.

GIS Data Processing for Environmental Justice Analysis

Two types of data were processed for this application:

- Total Non-Hispanic White Population
- Median Income

Their processing is described below.

Total Non-Hispanic White Population:

This was processed in a manner similar to the processing for the total population. Individual ‘Block level’ values for ‘Total Non-Hispanic White Population’ were available for Census 1990 & 2000 as well as ‘Block Group level’ projection years of 2004/2009/2014. The projection years were dispersed down to the ‘Block level’ using the ‘Population Change Dispersal using Equal Areas with Control Zones and Population Growth Categories’ method described in the main report. These years were then interpolated using the interpolation method described in the main report to get 2006/2011 analysis year data. Finally, this ‘Total Non-Hispanic White Population’ data was subtracted from the previously obtained ‘Total Population’ data to obtain the projected ‘Block level’ minority data.

Median Income:

Because of privacy concerns, the Census does not publish ‘Block level’ median income data. Thus, the process described above or any other process that utilizes historical data to make future projections at the block level cannot be utilized. The projection data was for the years 2004/2009/2014 at the ‘Block Group level’. This data was then assumed to apply to each individual block within the particular block group and interpolated to get the 2006/2011 analysis year median income data.