Sonoma Technology, Inc.

1360 Redwood Way, Suite C
Petaluma, CA 94954-1169
707/665-9900
FAX 707/665-9800
www.sonomatech.com

DEVELOPMENT OF GRIDDED SPATIAL ALLOCATION FACTORS FOR THE STATE OF CALIFORNIA

TECHNICAL MEMORANDUM
STI-900201/999542-2092-TM

By:
Tami H. Funk
Patricia S. Stiefer
Lyle R. Chinkin
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
California Air Resources Board
2020 L Street
Sacramento, CA 95814

July 10, 2001

This page is intentionally blank.

TABLE OF CONTENTS

Section **Page**

This page is intentionally blank.

LIST OF FIGURES

This page is intentionally blank.

LIST OF TABLES

1. INTRODUCTION

This technical memorandum discusses sources of data and describes in detail methodologies employed to develop gridded spatial allocation factors for the California Regional Particulate Air Quality Study (CRPAQS) and the Central California Ozone Study (CCOS) domains. These spatial allocation factors, which will be used to geographically distribute area and non-road mobile source emissions, were developed from *spatial surrogate data*. Spatial surrogates are economic, demographic, and land cover patterns that vary geographically.

Gridded spatial allocation factors for a 2000 base-year and future years (2005, 2010, 2020) were developed for the entire state of California based on the statewide 4-km grid cell domain defined by the California Air Resources Board (ARB). The definition and extent of the ARB-defined 4-km grid were used to create a 2-km nested grid for which spatial allocation factors were developed. The grid extent and spatial coverage of the grid domain are shown in **Figure 1-1**.

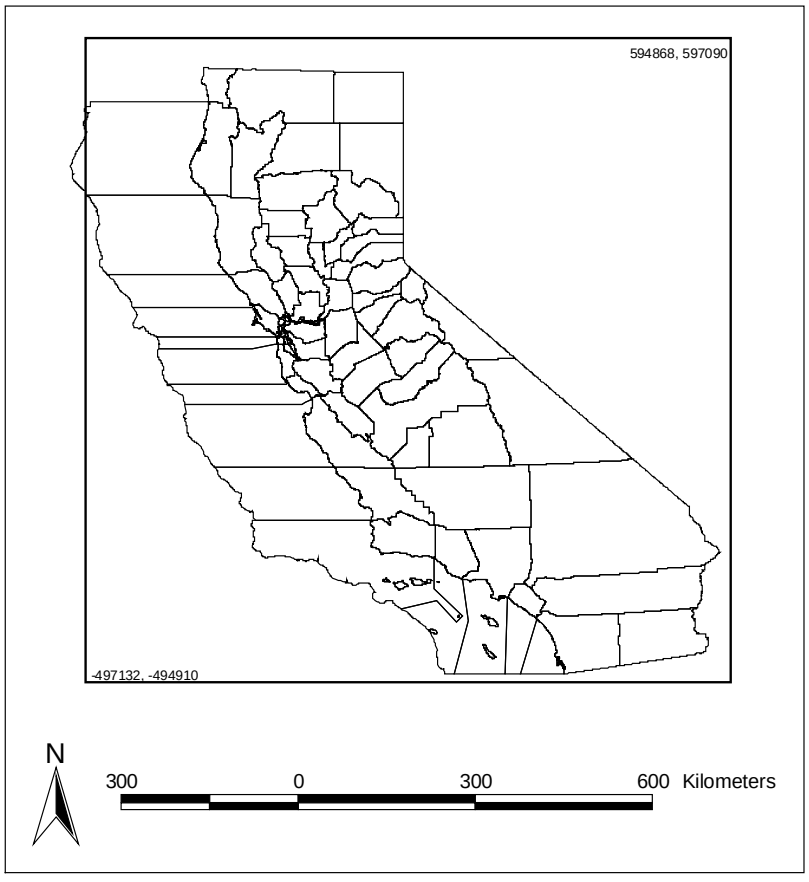


Figure 1-1. Depiction of the CCOS/CRPAQS domain extent (Lambert Conformal Conic projection).

The objectives of the CRPAQS/CCOS gridded surrogate project were twofold: (1) to develop spatial allocation factors of the highest possible quality and spatial resolution, and (2) to develop spatial allocation factors using methodologies and processing schemes that can be used by ARB staff to update surrogate data in the future. To achieve these objectives, a thorough review was performed to identify sources of regional, local, and statewide surrogate data appropriate for use in the development of spatial allocation factors for the CRPAQS/CCOS project; and a Geographic Information System (GIS)-based approach was developed to create a processing scheme that ARB can refer to and build upon in future surrogate work.

The remainder of this technical memorandum is divided into two sections. Section 2 discusses the sources of underlying surrogate data used in this project, and Section 3 provides a detailed discussion of the methodologies and GIS-based processing scheme used to develop the spatial allocation factors. Appendices A through E contain surrogate data information and maps.

2. SOURCES OF SURROGATE DATA

Land use and land cover data are used to develop spatial allocation factors that are associated with specific land uses (e.g., agriculture, feedlots, and recreation areas). For emissions sources associated with residential locations and industrial, commercial, and retail areas, demographic and socioeconomic data are typically used to develop spatial allocation factors. For emissions from sources such as gasoline service stations and drycleaners, facility locations were used to develop spatial allocation factors. A thorough review of available data was carried out to identify existing sources of surrogate data to update and improve upon surrogates used in the past. Consideration was given to the resolution, vintage, and representativeness of surrogate data compared to emissions sources. Several sources of surrogate data were identified including

- Local transportation planning agencies (TPAs)
- Air pollution control agencies
- State and federal agencies
- Commercially available data sources

A technical memorandum, dated November 28, 2000, was delivered to the ARB containing a detailed discussion of the sources of identified surrogate data as well as recommendations for surrogate development.

2.1 LAND USE AND LAND COVER DATA

Several sources of land use and land cover data were identified and evaluated for use in surrogate development. **Table 2-1** contains a summary of the land use and land cover data sets that were used to develop spatial allocation factors.

Table 2-1. Sources of land use and land cover data used to develop spatial allocation factors.

Data Source	Parameter	Resolution	Vintage	Coverage
United States Electronic Yellow Pages (*ProCD Select Phone*)	Autobody shops, dry cleaners, restaurants, gas stations, and wineries	Address locations	1997	Statewide
Environmental Systems Research Institute	Airports, parks, golf courses, hospitals, institutions	Coordinate locations and polygon coverages	1997	Statewide
U.S. Census Bureau (ESRI ADOL version)	Water bodies	Polygon coverages	2000	Statewide
United States Geological Survey	Land use and land cover data for 38 categories	Gridded data	1993	Statewide

Table 2-1. Sources of land use and land cover data used to develop spatial allocation factors.

Data Source	Parameter	Resolution	Vintage	Coverage
ARB CEIDARS Database	Bulk plant locations	Coordinate locations	1999	Statewide
National Atlas	Mine locations	Coordinate locations	1998	Statewide
Bureau of Transportation Statistics	Ports and shipping lanes	Coordinate locations and line coverages	Publication date is 2000; source date varies	Statewide
State Water Resources Control Board	Publicly owned water treatment works locations	Coordinate locations	2001	Statewide
Integrated Waste Management Board	Landfill locations	Coordinate locations	Downloaded from the Internet, no dates	Statewide
StreetWorks	Military bases	Polygon coverages	1995	Statewide
Digital Chart of the World	Elevation data	Polygon coverages	1993	Statewide
California Department of Oil and Gas	Oil and gas well and field locations	Coordinate locations and polygon coverages	1998	Statewide
California Teale Data Center (from ARB)	Urban and rural roads and railroads,	Line and polygon coverages	RR, updated 1991; RDS, updated 1993	Statewide
Department of Water Resources (from ARB)	Agricultural land cover	Polygon coverages	1995	San Joaquin Valley

2.2 DEMOGRAPHIC AND SOCIOECONOMIC DATA

Several sources of statewide and local demographic and socioeconomic data sets were identified and assessed as part of this project. **Table 2-2** provides a summary of the demographic and socioeconomic data sets that were used to develop spatial allocation factors. **Table 2-3** provides a list of the counties covered by each data set. To develop spatial allocation factors of high quality and resolution, local socioeconomic and demographic data were used when available; for rural regions for which local data were not available, the Caltrans STM data were used.

Table 2-2. Sources of statewide and local TPA demographic and socioeconomic surrogate data used to develop spatial allocation factors.

Data Source	Parameter (Years)	Resolution and Coverage
Caltrans Statewide Transportation Model (Caltrans STM)	Population, housing, employment (base and future)	TAZ[a] – data for rural counties only
Association of Bay Area Governments (ABAG) and 1990 U.S. Census	Population, housing, employment (base and future)	Census Tract - San Francisco Bay Area
Sacramento Area Council of Governments (SACOG	Population, housing, employment (base and future)	TAZ[a]- Sacramento Urban Region
Tahoe Regional Planning Agency (TRPA)	Population, housing, employment (base and future)[b]	TAZ[a]- Lake Tahoe Region
Association of Monterey Bay Area Governments (AMBAG)) and 1990 U.S. Census	Population (base and future)	Census Tract - Monterey Bay Area
South Coast Association of Governments (SCAG)	Population, housing, employment (base and future)	TAZ[a]- South Coast Region
Amador County Transportation Commission (ACTC)	Population, housing, employment (base and future)[b]	Growth Allocation Districts (unincorporated areas) and incorporated areas – Amador County
Council of Fresno County Governments (FresnoCOG)	Population, housing, employment (base and future)	TAZ[a] - Fresno County
San Diego Association of Governments (SANDAG)	Population, housing, employment (base and future)	TAZ[a]- San Diego County
San Joaquin Council of Governments (SJCOG)	Population, housing, employment (base and future)[b]	TAZ[a] – San Joaquin County
Tulare County Association of Governments (TCAG)	Population, housing, employment (base and future)	Incorporated and unincorporated areas – Tulare County
Stanislaus Counsel of Governments (StanCOG)	Population, housing, employment (base and future)	Incorporated and unincorporated areas – Stanislaus County
Kern Council of Governments (KernCOG)	Population, housing, employment (base and future)	TAZ[a]- Kern County

[a] TAZ = traffic analysis zone.
[b] Data for future years was interpolated or extrapolated as necessary.

Table 2-3. Counties covered by each of the demographic and socioeconomic data sets listed in Table 2-2.

Data Source	County Coverage
Caltrans STM	Alpine, Butte, Calaveras, Colusa, Del Norte, Glenn, Humboldt, Imperial, Inyo, Kings, Lake, Lassen, Mariposa, Madera, Merced, Mendocino, Modoc, Mono, Nevada, Plumas, east Riverside, east San Bernardino, San Luis Obispo, Santa Barbara, Shasta, Sierra, Siskiyou, Tehama, Trinity, Tuolumne
ABAG	Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, Sonoma
SACOG/TRPA	El Dorado, Placer, Sacramento, Sutter, Yolo, Yuba
AMBAG	Monterey, San Benito, Santa Cruz
SCAG	Los Angeles, Orange, west Riverside, west San Bernardino, Ventura
ACTC	Amador
FresnoCOG	Fresno
SANDAG	San Diego
SJCOG	San Joaquin
TCAG	Tulare
StanCOG	Stanislaus
KernCOG	Kern

2.3 ASSIGNMENT OF SURROGATES TO EMISSIONS SOURCE CATEGORIES

Spatial allocation factors are used to spatially disaggregate countywide emissions into grid cells. Each identified surrogate must be assigned to a source category (or multiple categories) in the emission inventory. This assignment provides a cross reference between the spatial allocation factors and the emission inventory categories. A total of 65 surrogates were developed as part of this project. A listing of all surrogates and spatial allocation factors, and their corresponding spatial surrogate codes (SSC), are contained in Appendix A. Appendix A also includes the surrogate-to-emission inventory cross-reference list. Designating the surrogate-to-emission inventory assignments was an iterative process among Sonoma Technology, Inc. (STI) staff, ARB staff, and local air district staff. Appendix B contains surrogate documentation and/or details about the sources of data.

3. DEVELOPMENT OF GRIDDED SPATIAL ALLOCATION FACTORS

A GIS-based approach was used to develop gridded surrogate data and spatial allocation factors. The GIS-based approach consists of the following steps:

1. Manipulation of all geographic data sets into the required projection and format(s) for the GIS and processing.
2. Manipulation of all tabular data sets into a format required for creating gridded surrogates and spatial allocation factors.
3. Spatial disaggregation of surrogate coverages into grid cells using ArcInfo and customized processing scripts written in Arc Macro Language (AML).
4. Use of Microsoft Access databases to calculate the gridded surrogates and spatial allocation factors for each set of surrogate data.
5. Use of ArcInfo geodatabase capabilities to display gridded surrogate and spatial allocation factor data contained in Access databases.
6. Quality assurance of the gridded surrogate and spatial allocation factor files.
7. Preparation of final databases, map displays, and documentation for ARB.

The remainder of Section 3 provides a detailed discussion of each of the seven processing steps employed to develop the gridded surrogates and spatial allocation factor files.

3.1 OVERVIEW OF THE GIS-BASED DATA PROCESSING SCHEME

A flowchart of the GIS-based data processing scheme is shown in **Figure 3-1**. The processing scheme is comprised of the following five tiers:

Tier 1—Raw geographic and tabular surrogate data: The raw geographic surrogate data files were pre-processed to generate Level 1 ArcInfo layer and map files. The raw tabular surrogate data were pre-processed for importation into the Level_1_Databases.

Tier 2—Level_1_Databases: Level_1_Databases contain all of the raw tabular demographic and socioeconomic data obtained from the individual TPAs. The function of the Level_1_Databases is to place the raw demographic and socioeconomic tabular data into a standardized format for processing.

Tier 3— Level_2_Databases: Level_2_Databases contain all gridded geographic and tabular data necessary for calculating the gridded surrogate and gridded spatial allocation factor files.

Tier 4— Level_3_Databases: The Level_3_Databases are ArcInfo-based geo-databases containing the gridded surrogate data and spatial allocation factor files as well as geographic displays of these data.

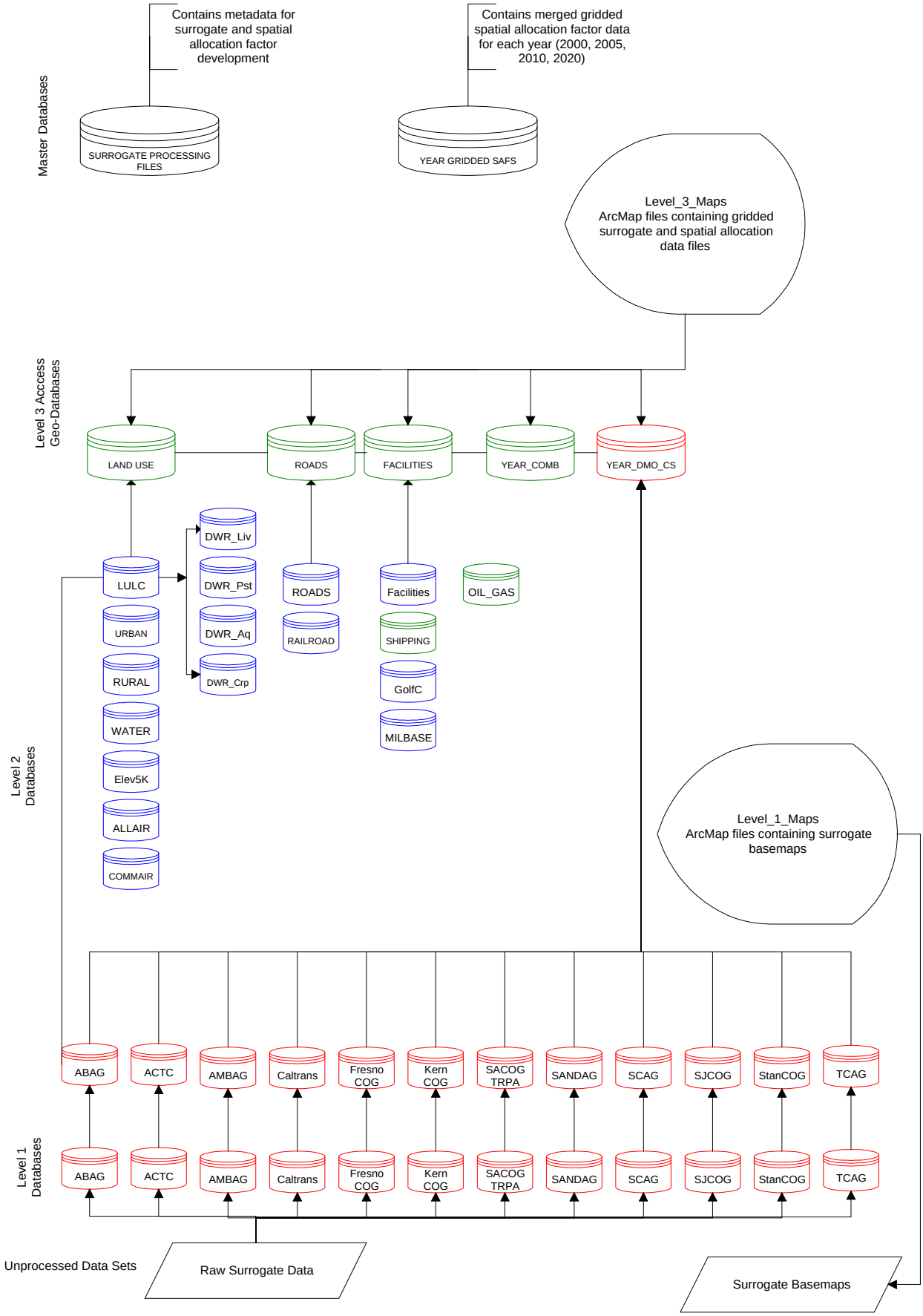


Figure 3-1. Flow diagram of the GIS-based processing scheme used to develop gridded surrogates and spatial allocation factors for the CCOS/CRPAQS domain.

Tier 5—Master Databases: The Master Databases contain all of the tabular spatial allocation factor files and metadata in the format requested by ARB.

The first level of processing involves manipulation and formatting of all geographic surrogate data. The next level of data processing involves preparing the raw tabular surrogate data for importation into the Level_1_Databases. The Level_1_Databases are designed to process the raw data files into a standardized format required by the Level_2_Databases. The Level_1_Databases only apply to the demographic and socioeconomic data sets. The Level_2_Databases perform the gridded surrogate and spatial allocation factor calculations and output files for all surrogate data sets. The Level_3_Databases (as well as some of the Level_2_Databases) are ArcInfo/Access geodatabases which store the gridded surrogate and spatial allocation factor files as feature classes; this feature class storage enables easy display of the data. The final level of processing involves creating the Master Databases which contain all of the final spatial allocation factor files as well as surrogate processing files.

The directories containing all CCOS/CRPAQS gridded surrogate project files are summarized in **Table 3-1**. When working with ArcInfo map files (.mxd), it is important to preserve the directory structure so that the map files can be opened. If files and/or directories are moved from their original locations, it may cause the map files to become unreadable. The directory path that should be established at ARB is E:/Project_Files/Gridded_Surrogates. The directories and sub-directories should correspond to the structure presented in Table 3-1.

Table 3-1. Directories containing all CCOS/CRPAQS gridded surrogate project files.

Directory Name	Contents
Aml_templates	Contains generic AML scripts used for this project.
Basemaps	Contains all surrogate basemaps, gridded basemaps, and AML gridding scripts. The Level 1 ArcInfo layers and mapfiles are also contained in this directory.
Demography	Contains all raw demographic and socioeconomic surrogate data obtained from each local TPA.
Documentation	Contains project documentation files, including this technical memorandum.
Level_1_Databases	Contains all the demographic and socioeconomic surrogate databases.
Level_2_Databases	Contains all the Level_2_Databases used to create gridded surrogate and spatial allocation factors. Also contains ArcInfo feature attribute tables (FAT_Files) and gridded data files used for Level 2 processing.
Level_3_Databases	Contains all the Level 3 geodatabases containing gridded surrogate and spatial allocation factor ArcMap feature classes. Directory also contains the Level 3 ArcMap map files displaying the gridded surrogates and spatial allocation factor data.
Master Databases	Final gridded spatial allocation factor files for base and future years and surrogate information database.

3.2 PREPARATION OF SURROGATE BASEMAPS

GIS data for several demographic, land use, and land cover categories were obtained from various sources in widely varying formats and map projections. The procedures for assembling this data into the surrogate basemaps, and some of the individual data sets processed, are described below. Each basemap was displayed in an ArcInfo mapfile (.mxd); printouts of the basemaps are included in Appendix C.

3.2.1 General Procedures

Following are the general steps for assembling the surrogate basemaps:

- Import the original GIS data into an ArcInfo coverage and re-project it to match the CRPAQS/CCOS grid domain projection.
- Overlay the surrogate coverage with the CRPAQS/CCOS county boundaries to visually verify location.
- Identify and resolve any geometry or placement errors that are beyond acceptable limits.
- Add and populate attribute table fields for the county FIPS codes and standardized cross-reference codes for demographic polygons.

3.2.2 Demographic Basemaps

ARB and STI contacted local TPAs to request demographic forecast data and corresponding GIS boundary files for smaller, sub-county areas such as TAZs, Census Tracts, county planning area boundaries, etc. The Caltrans STM GIS files and data were used for areas where no suitable forecast or GIS data were available.

The GIS data were as varied as the sources and included ArcInfo coverages, ArcView shapefiles, MapInfo native format files, and in a few cases, paper maps. Paper maps were useable where the boundaries depicted corresponded to known boundaries for which GIS data were readily available (i.e., Census Tracts, Caltrans TAZs, etc). Several original data sets or paper maps required additional processing before the final basemap could be completed as listed below:

- Caltrans. A GIS file was available for the 1995 TAZ boundaries; however, the available demographic forecast data were keyed to the 1990 TAZ boundaries, which exist only in paper format. For each county, the paper map was compared to the 1995 GIS file, to which 1990 Census Tract boundaries had been added. Most of the 1990-1995 TAZ boundary differences were resolved by using the underlying Census Tracts as "building blocks" and merging tracts where appropriate to make a new, 1990 TAZ boundary file. In a few locations, 1990 TAZ boundary lines were manually digitized to create the new file.

- Fresno COG. Detailed TAZ boundaries were available in paper format only. These were visually compared to 1990 Census Tract and other boundaries, and TAZ boundary files were constructed by merging the tract and other boundary files and digitizing, as appropriate.
- ACTC. ACTC was able to provide a paper map showing the 5 incorporated areas and 39 growth allocation districts in the unincorporated area of the county for ACTC's 96/97 Regional Transportation Plan. STI scanned the paper map and registered it to a geographic coordinate system. The GIS data files were then digitized.
- ABAG and AMBAG. The 1990 Census Tract boundary files for these two areas were downloaded from the U.S. Census Bureau's Cartographic Boundary Files web site.

Because the demographic and socioeconomic data sets were obtained in many different formats and the geographic boundaries within the files were inconsistent (i.e., TAZ, Census Tracts, community boundaries), a standardized cross-reference field (XREF) was added to each geographic data set. The XREF is a number from 1 to X which makes possible a standard numeric polygon identifier corresponding to each TAZ, Census Tract, or community code.

3.2.3 Land use and land cover basemaps

GIS data for land use and land cover maps were obtained from the various sources listed in Table 2-1. This data varied widely in format and projection. Some original data sets required additional editing or filtering before the final basemap could be completed.

- Recreational use water. Water body polygons from the 2000 census data for 58 counties were merged and manually edited in ArcView before being imported into ArcInfo. The edit process eliminated all dry lakes and retained all lakes and rivers with boat marinas, based on information in *The Thomas Guide® California Road Atlas and Driver's Guide* (1993) and information found on a web site belonging to California Resources Agency, Department of Boating and Waterways.
- Facilities. Locations from the Electronic Yellow Pages (EYP) were geocoded to street addresses or zip code centroids. Sources collocated at a zip code centroid were randomly dispersed throughout the zip code boundary.
- Oil and gas wells. Onshore and offshore oil and gas well locations were available for all districts except District 4, Kern County. For District 4, STI used the field polygons for active fields that fell at least partly in Kern County. For Districts 1, 2, 3, 5, and 7, abandoned wells were filtered out, and all other well sites—active, completed, idle, directional, etc.—were retained.
- Landfills. Only sites flagged as active by the data provider were included. In general, geographic coordinates provided by the data provider were either the actual facility coordinates, or the coordinates of the centroid of the city in the site address. However, the landfill records were not flagged to make this distinction, and about half of the active sites were collocated—two or more landfill records sharing the same coordinates. STI used several geolocation tools to resolve as many collocated addresses as possible. Original coordinates were used for unresolvable locations.

- Publicly owned water treatment works. The original data contained street and city addresses but no geographic coordinates; thus, STI used several geolocation tools to assign coordinates based on street addresses, or zip codes if street addresses could not be located.
- Urban/rural areas. The urban area polygons are based on 1990 U.S. Census Bureau's definitions for Urbanized Areas; the remaining land area outside a 1990 Urbanized Area is rural land. An Urbanized Area is only partly based on population density and consists of the total of the following areas:
  - One or more central places (either a central city within a Census-defined metropolitan area or a city having a population above a certain size)
  - Urban fringe (adjacent, contiguous territory having a density of at least 1,000 persons per square mile)

Surrogate basemaps for all sources of data used to develop gridded surrogates and spatial allocation factors are contained in Appendix C.

3.3 PROCESSING OF TABULAR DATA SETS

The purpose of the Level_1_Databases is to provide a standardized method for defining and computing surrogates from the raw socioeconomic and demographic data. Because data were obtained from various TPAs in many different formats, it was necessary to standardize the data sets and define surrogates prior to gridding the data. Databases were created for each set of socioeconomic and demographic data obtained. There are a total of thirteen databases, all containing the same structure and standardized surrogate information.

Because data were obtained from many different planning agencies, the data sets vary in their content and completeness. In order to standardize the processing of data, demographic and computed surrogate codes were created (DMO1-DMO9, DMO10, DMO12, DMO13 and CS1-CS5). The demographic (DMO) surrogates were developed using the raw data as it exists (or by combining raw data categories) with no mathematical manipulation of the data. The computed surrogates (CS) were developed by implementing scaling and growth factors to account for the spatial patterns of emissions sources that are transient in nature (e.g., construction activity).

The spatial surrogate codes developed were intended to make possible highly resolved socioeconomic and demographic data sets (i.e., figures for population, several categories of housing, and employment data disaggregated by industry sector). However, some data sets only contain numbers for the basic surrogates (population, housing, and employment). In cases where the data sets are comprehensive, there may be 12 unique DMO surrogates. In cases where the data sets contain minimal information, there may only be one unique surrogate (as is the case for AMBAG). In this case, some of the surrogates may be duplicates, in order to create a full set of surrogates for each database to standardize the Level 1 databases and to remain consistent among databases. For example, if only total population data is available, then all DMO surrogates will be duplicates of the total population surrogate, DMO1.

Computed surrogates were developed for construction and architectural coatings activities. Because construction activities are transient (they do not take place in a fixed location), developing spatial allocation factors for these sources is more challenging. Cities tend to expand outward and new building construction usually occurs in the suburban and rural regions surrounding the central business district while maintenance and remodeling occur on existing structures. In order to address the transient nature of construction-related emissions, an approach to spatially allocate construction emission sources using a computed surrogate was employed.

Spatial allocation factors for building (both residential and non-residential) construction activities were computed by taking the difference in each grid cell for retail, non-retail, and housing data between a past year and a future year (e.g., 1995 and 2000, respectively) to obtain the differences between past-year and future-year activities. This difference calculation was used to determine where new construction activity is likely to occur. In addition to determining where new construction may occur, repair and maintenance of existing structures was also considered. Because new construction activity tends to be more equipment-intensive and, consequently, produces more emissions, areas where new construction may occur (based on the difference calculation) were weighted more heavily than areas where little or no new construction may occur.

According to the U.S. Census Bureau's 1997 Economic Census Construction Report (U.S. Census, 1999), approximately 30% of total residential construction expenditures in the state of California are attributed to repair and maintenance of existing residences. The difference in calculation and census figures were combined to arrive at a weighting scheme which weights new construction activity more heavily than repair and maintenance resulting in the greatest amount of emissions being assigned to the grid cells with the highest levels of growth. An example computed-surrogate calculation is shown below.

Computed Surrogate for Residential Construction between years A and B:

$$S_{rc(x)} = (TH_{A(x)} \times 0.30) + (TH_{B(x)} - TH_{A(x)})$$

where:

$S_{rc(x)}$ = residential construction surrogate for grid cell "x"
$TH_{A(x)}$ = total housing value in year A for grid cell "x"
$TH_{B(x)}$ = total housing value in year B for grid cell "x"
0.30 = percentage of total statewide residential construction expenditures spent on repair and maintenance

Spatial Allocation Factor for Residential Construction in 2010:

$$SAF_{rc(x)} = S_{rc(x)} / S_{rc(cnty\ total)}$$

where:

$SAF_{rc(x)}$ = spatial allocation factor for residential construction in grid cell "x"
$S_{rc(x)}$ = computed surrogate value for grid cell "x"
$S_{rc(cnty\ total)}$ = county total surrogate value

Example calculation

Computed surrogate for residential construction between years 2005 and 2010 in grid cell "x":

$$\text{if } TH_{05(x)} = 500;\ TH_{10(x)} = 800;\ \text{and } S_{rc(\text{cnty total})} = 10000$$

$$\text{then, } S_{rc(x)} = (500 \times 0.30) + (800 - 500)$$

$$S_{rc(x)} = 450$$

Spatial allocation factor for residential construction between years 2005 and 2010 in grid cell "x":

$$SAF_{rc(x)} = 450 / 10000$$
$$= 0.045$$

Applying the spatial allocation factor calculated above to countywide emissions estimates for residential construction would result in 4.5% of the emissions being allocated to grid cell "x". When there is no difference, or a negative difference, in housing between years (e.g., $TH_{A(x)} \geq TH_{B(x)}$), the second term in the surrogate equation becomes 0 (or is assumed to be 0 when there is negative growth), and only repair and maintenance on existing structures are considered. This methodology was used to calculate computed spatial allocation factors for residential (CS2), non-residential (CS3) and a combination residential/non-residential activities (CS4). According to the U.S. Census Bureau's 1997 Economic Census Construction Report (U.S. Census Bureau, 2000b), approximately 46% of total non-residential construction expenditures in the state of California are attributed to repair and maintenance. Based on this information, a 46% repair-and-maintenance factor was applied to existing non-residential structures.

A computed surrogate was also created for architectural coating activities (CS1). Figures for residential population and employment were combined with average occupancy data to create a weighting factor for spatially allocating architectural coating emissions. According to the U.S. Census Bureau, the average square space occupancy for residences is 665 ft^2/person and the average space occupancy for employees is 200 ft^2/person (U.S. Census Bureau, 2000a). The equation used to compute the CS1 surrogate is shown below.

Computed surrogate for allocating architectural coatings (CS1):

$$CS1_x = [(\text{Residential population})_x * (665)] + [(\text{Total employment})_x * (200)]$$

where:

$CS1_x$ = computed surrogate 1
$(\text{Residential population})_x$ = residential population for TAZ "x"
665 = residential occupancy (ft^2/person)
$(\text{Total employment})_x$ = total employment for TAZ "x"
200 = business occupancy (ft^2/person)

Level 1 Data Processing

In most cases, the demographic surrogate forecast data were obtained from the same source as the GIS boundary files, except for AMBAG and ABAG. The content of the demographic files, usually MS Excel or text files, varied in the years covered and the number and definition of demographic variables included. Where necessary, STI performed a linear interpolation or extrapolation on all variables to fill in data for the target years of 2000, 2005, 2010, and 2020. All raw demographic variables in the original data set were processed into delimited text files for import into Level_1_Databases. The appropriate county FIPS codes and standardized polygon XREF are assigned to each record in the raw data. The text files were imported into each database, creating raw data tables in the same format in each database. An update query adds a county FIPS code and cross-reference field, if not available in the raw data on import.

Each database contains the raw data tables, a definition table, and a surrogate information table that maps the relevant field from the raw data into the appropriate spatial surrogate names and codes. The surrogate information table also contains a description of the formulas used in the queries to derive or calculate each surrogate from the raw data. Queries in the Level_1_Databases create the output tables, which will be linked into the Level_2_Databases along with the gridded GIS tabular files. The gridded surrogates and spatial allocation factors are calculated in the Level_2_Databases.

The tables contained in each Level_1_Database are listed in **Table 3-2** along with their contents. Each database contains the same standardized queries.

Table 3-2. Tables contained in the Level_1_Databases

Table Name	Description
199X	Standardized surrogate data for a year prior to the 2000 base year.
2000	Standardized surrogate data for 2000.
2005	Standardized surrogate data for 2005.
2010	Standardized surrogate data for 2010.
2020	Standardized surrogate data for 2020.
Definition	File defining the fields contained in the standardized surrogate data files.
SurgInfo	List of each surrogate code and its definition (contingent on what data were available).
XREF	Cross reference code associating a TAZ, Census Tract Code, or City name with a standard numeric reference number XREF.
Miscellaneous Tables	Some of the databases contain other tables containing information that is specific to that data set.

Table 3-2. Tables contained in the Level_1_Databases

Query Name	Description
YEAR_XREF_Update	Updates the standardized data files with the XREF code.
YEAR_DMO	Generates surrogates (DMO1-DMO9) based on definitions in SurgInfo table.
YEAR_CS	Generates computed surrogates (CS1-CS5) based on definitions in SurgInfo table.
YEAR_TOTALS	Calculates the surrogate totals (DMO1-CS5) by county.

3.4 GRIDDING OF SURROGATE BASEMAPS

Development of the gridded surrogate and spatial allocation factor files was carried out using ArcInfo scripts written in AML and the Level_2_Databases. Each surrogate coverage file was spatially disaggregated into grid cells using AML scripts to carry out a sequence of ArcInfo processing steps. Each of the surrogate coverage files was disaggregated to grid cells using the following three types of spatial overlay functions:

- Point-in-polygon – overlaying point location data and the grid domain
- Line-in-polygon – overlaying line coverage and the grid domain
- Polygon-in-polygon – overlaying polygon coverage and the grid domain

Figure 3-2 illustrates of the grid overlays. All AML gridding scripts contain the name of the AML file followed by "_gso". These scripts are located in each of the surrogate basemap directories. The AML scripts are consistent for each surrogate data set. Each AML script prompts ArcInfo to carry out the following processes:

- An overlay is performed which combines the geographic features of the surrogate coverage with the grid coverage and creates a new coverage designated by "_g" in the filename indicating that the surrogate coverage has been gridded.
- The gridded surrogate coverage file (_g) is opened and all records with no spatial overlap between the surrogate and the grid are removed from the file (this feature was added to reduce extraneous data from each file). A new coverage is then generated containing a "_gs" in the filename indicating that the surrogate coverage has been gridded and only the pertinent records have been selected.
- The gridded and selected coverage file (_gs) is opened and a new item (or field) is added to the coverage attribute table. The new field is populated with a unique identification code consisting of the grid cell identifier and the surrogate feature identifier.
- The density of surrogate features (e.g., area, length, or number) contained in each grid cell is calculated and an output file is generated with "_gso" in the filename indicating that the file has been gridded, selected, and output.

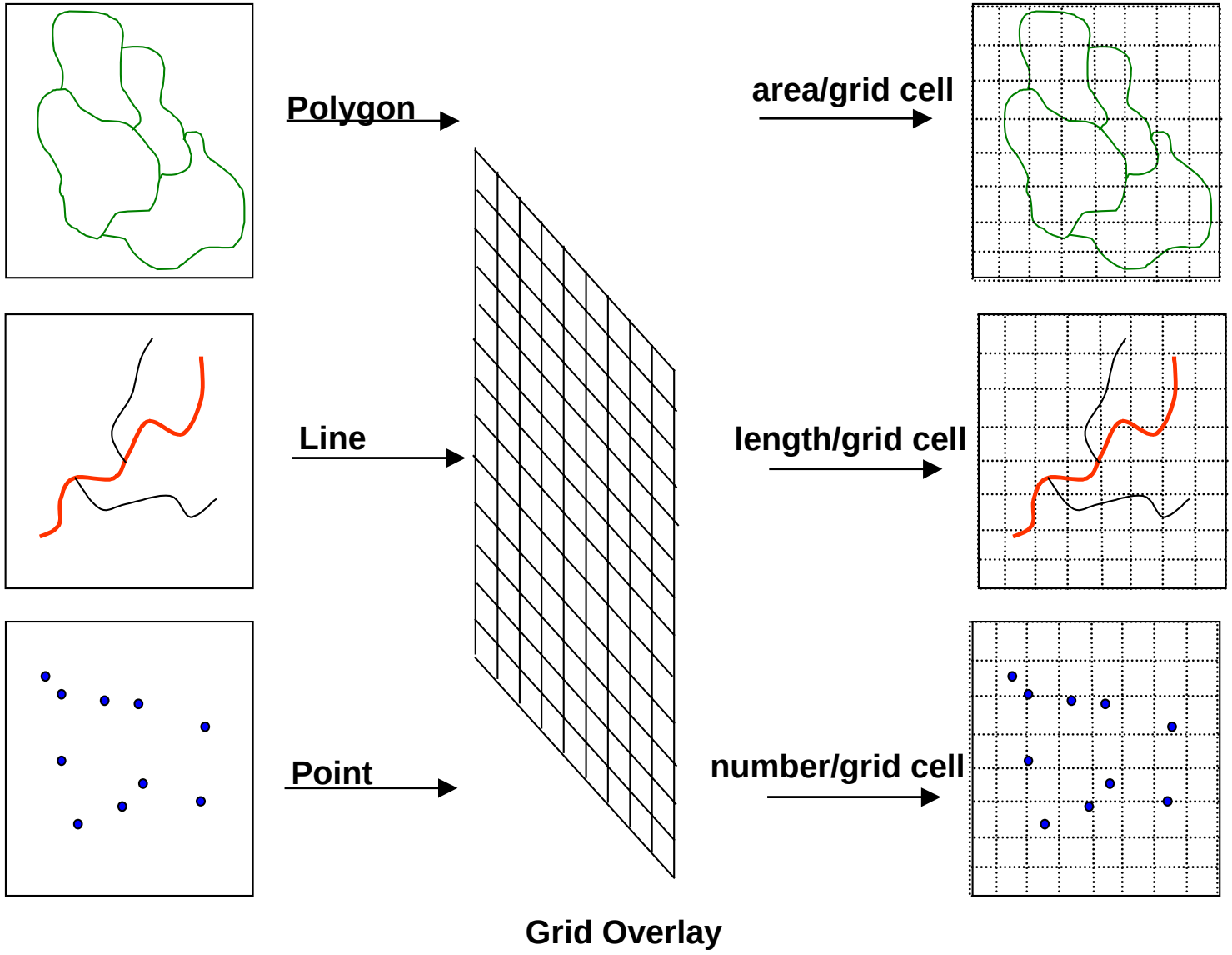


Figure 3-2. Illustration of disaggregating geographic objects into grid cells.

Figure 3-3 includes an example AML gridding script with annotated text next to each command line of the script. The gridded output files were then imported into the Level_2_Databases where they were combined with tabular data (when applicable), and the gridded surrogate and spatial allocation factor files were calculated.

3.5 CALCULATION OF GRIDDED SURROGATES AND SPATIAL ALLOCATION FACTORS

The Level_2_Databases were used to combine the gridded geographic surrogate data with tabular surrogate data (when applicable) and to calculate the gridded surrogate and spatial allocation factor data files. The Level_2_Databases were named based on the type of data they contain. There are four general categories of Level_2_Databases including:

- Databases containing demographic and socioeconomic data
- Databases containing land use and land cover data
- Databases containing transportation routes (i.e., roads and railroads)
- Databases containing area source facility location data

The database category designations are important for Level 3 processing because the Level_3_Databases are linked to the Level_2_Databases according to the type of data contained in each Level_2_Database. The format, filenames, and queries in each of the Level_2_Databases are consistent among the databases and are annotated in all databases.

3.5.1 Level 2 Demographic Databases

The following are Level 2 demographic databases:

- ABAG.mdb (Demographic surrogate data for ABAG)
- ACTC.mdb (Demographic surrogate data for ACTC)
- AMBAG.mdb (Demographic surrogate data for AMBAG)
- Caltrans.mdb (Demographic surrogate data for Caltrans)
- FresnoCOG.mdb (Demographic surrogate data for FresnoCOG)
- KernCOG.mdb (Demographic surrogate data for KernCOG)
- SACOG_TRPA.mdb (Demographic surrogate data for SACOG & TRPA)
- SANDAG.mdb (Demographic surrogate data for SANDAG)
- SCAG.mdb (Demographic surrogate data for SCAG)
- SJCOG.mdb (Demographic surrogate data for SJCOG)
- StanCOG.mdb (Demographic surrogate data for StanCOG)
- TCAG.mdb (Demographic surrogate data for TCAG)

Level 2 demographic databases were constructed with linkages to the tables in the corresponding Level_1_Databases. The demographic and socioeconomic surrogate data (DMO and CS) generated in the Level_1_Databases was used in the Level_2_Databases to create the gridded surrogate and spatial allocation factor files. The Level_2_Databases combine the output of the ArcInfo gridding processes with the tabular surrogate data generated in the Level_1_Databases. **Table 3-3** provides a summary of the tables and queries contained in the Level 2 demographic databases.

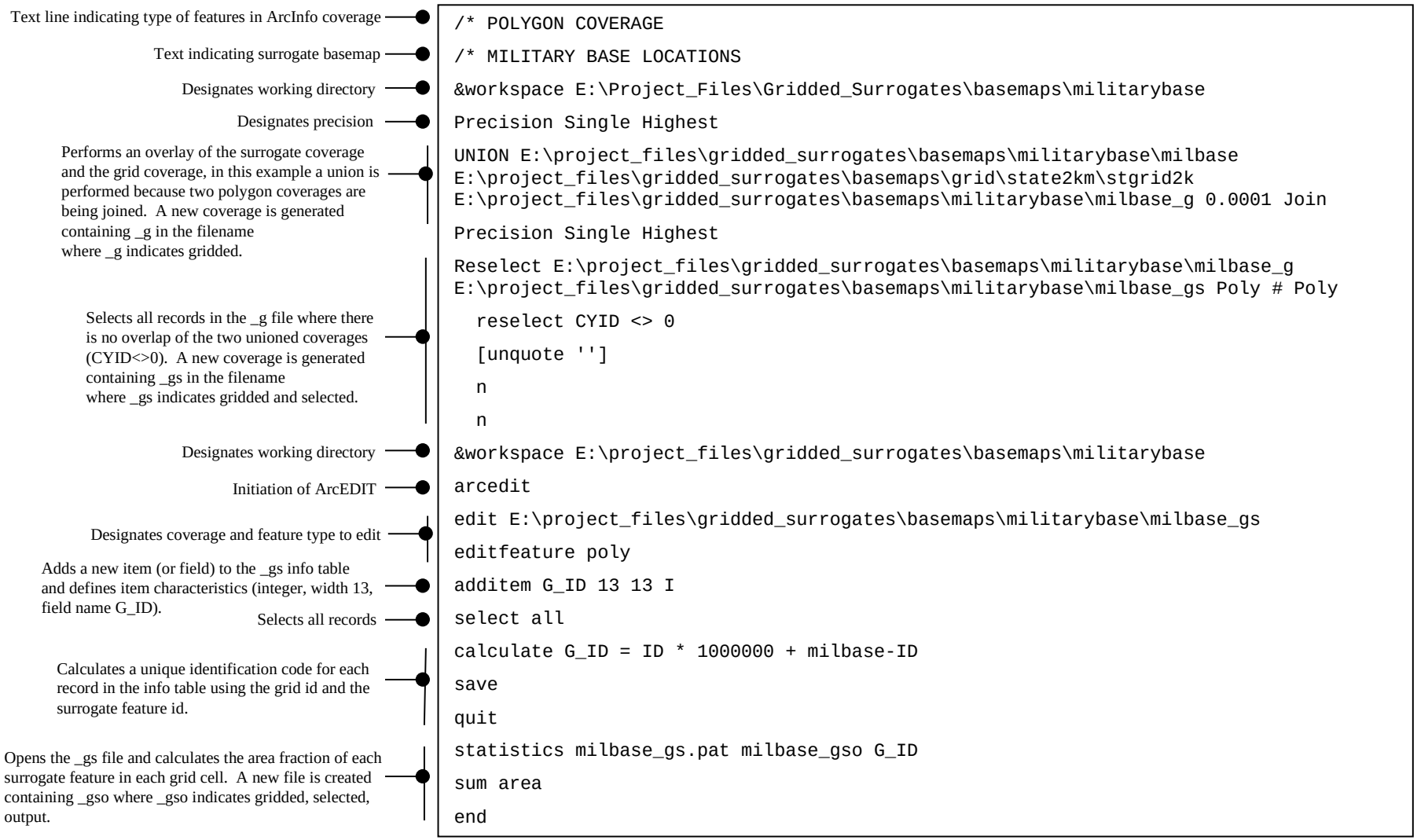


Figure 3-3. Example of an ArcInfo AML script used to carry out gridding processes.

Table 3-3. Summary of tables and queries contained in the Level 2 demographic databases.

Table Name	Description
GRIDDED_INPUT	Gridded surrogate data, output from ArcInfo AML gridding scripts. Contains the area fraction of each TAZ or Census Tract in each grid cell.
YEAR_TOTALS	Contains surrogate totals by county for each base and future year, linked to Level_1_Databases.
YEAR_DMO	Demographic surrogate data (DMO) for each year, linked to Level_1_Databases.
YEAR_CS	Computed surrogate data for each year, linked to Level_1_Databases.
YEAR_GS	Gridded surrogate file.
YEAR_SAF	Gridded spatial allocation factor file.
Query Name	**Description**
Update_Gridded_Input	Updates the gridded input file with GRID_ID and XREF, and calculates area in kilometers.
ZONE_AREAS	Calculates the total area of each TAZ or Census Tract in data set.
YEAR_GS_CALC	A Make Table query that calculates the gridded surrogate file and creates YEAR_GS table.
YEAR_SAF_CALC	A Make Table query that calculates the gridded spatial allocation factor file and creates YEAR_SAF table.

The first step in the Level_2_Database processing was to link the totals, DMO, and CS tables from the Level 1 demographic databases. Once these tables were linked to the Level_2_Database, the gridded output file from the ArcInfo gridding process was imported into the database. The next step in the processing was to update the GRIDDED_INPUT file with three fields: (1) GRID_ID—the grid cell identifier; (2) XREF—the cross reference code representing each TAZ or Census Tract in the surrogate data set; and (3) AREA_KM2—calculates the area (in kilometers) of each TAZ or Census Tract in each grid cell. In order to do this, three new fields (with the appropriate field names) were added to the GRIDDED_INPUT file. Next, the Update_Gridded_Input query was run which updates the new fields with the correct data values. To calculate gridded surrogates and spatial allocation factors, both the fraction of the total TAZ or Census Tract residing in a grid cell and the total area of the TAZ or Census Tract must be known. The ZONE_AREAS query calculates the total area of all TAZ or Census Tract zones.

The gridded surrogate files were calculated using the YEAR_GS_CALC query. The YEAR_GS_CALC query uses the GRIDDED_INPUT, ZONE_AREAS, YEAR_DMO, and YEAR_CS tables to calculate the gridded surrogate values using the following general equation:

$$SSC\_GS = (AF_{(A)(g1)}/AT_{(A)})*SSC\_V_{(A)}$$

where:

SSC_GS = Gridded surrogate value for a given spatial surrogate code (SSC)
$AF_{(A)g(1)}$ = Area fraction of TAZ or Census Tract (A) in grid cell (1)
$AT_{(A)}$ = Total area of TAZ or Census Tract (A)
$SSC\_V_{(A)}$ = Spatial surrogate value for TAZ or Census Tract (A)

The gridded surrogate value is calculated by dividing the area fraction of a TAZ or Census Tract in a grid cell by the total area of the TAZ or Census Tract. The result is the fraction of the total TAZ or Census Tract in the grid cell. The surrogate data value corresponding to that TAZ or Census Tract is then multiplied by the area fraction. This computation results in the surrogate value per grid cell.

Spatial allocation factors are weighted values that indicate what fraction of the county total surrogate value resides in each grid cell. The gridded spatial allocation factors were calculated using the YEAR_SAF_CALC query. The YEAR_SAF_CALC query uses the GRIDDED_INPUT, ZONE_AREAS, YEAR_DMO, YEAR_CS, and YEAR_TOTALS tables to calculate the gridded spatial allocation values using the following general equation:

$$SSC\_GS = [(AF_{(A)(g1)}/AT_{(A)})*SSC\_V_{(A)}]/CT_{SSC}$$

where:

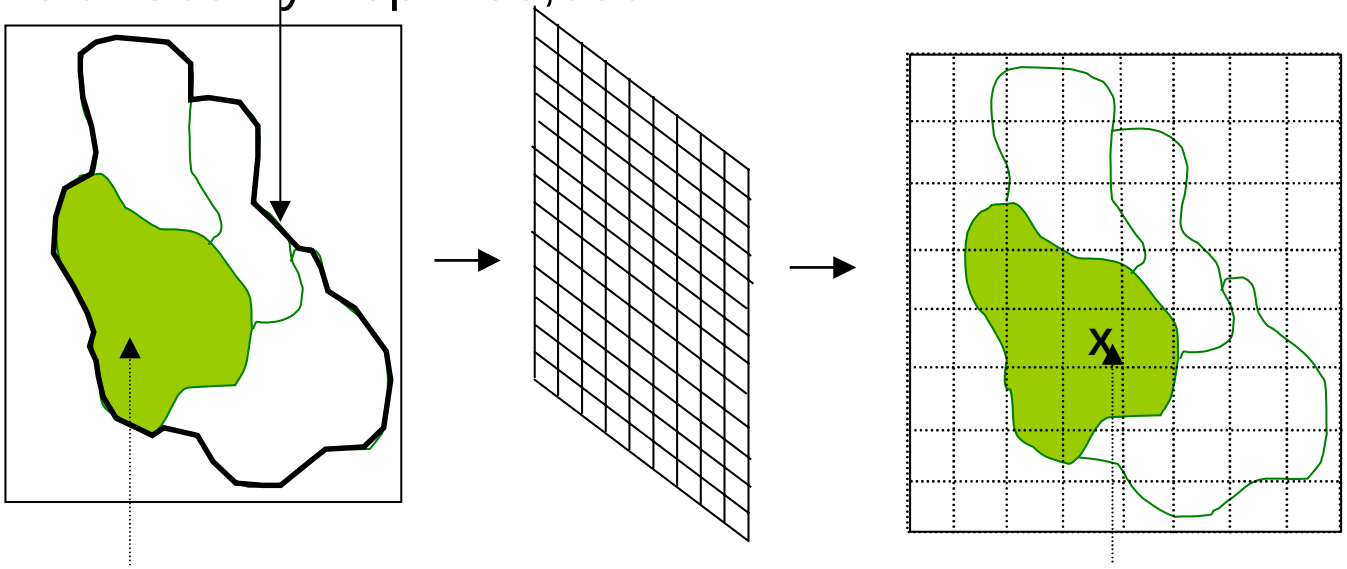
SSC_GS = Gridded surrogate value for a given spatial surrogate code (SSC)
$AF_{(A)g(1)}$ = Area fraction of TAZ or Census Tract (A) in grid cell (1)
$AT_{(A)}$ = Total area of TAZ or Census Tract (A)
$SSC\_V_{(A)}$ = Spatial surrogate value for TAZ or Census Tract (A)
CT_{SSC} = County total spatial surrogate value

The gridded spatial allocation factors are calculated using the same equation for the gridded surrogates and then dividing the gridded surrogate value by the total surrogate value for the county. The result is the fraction of the county total surrogate value in the grid cell. **Figure 3-4** provides an illustrated example of how the spatial allocation factors are calculated.

The output of the gridded surrogate and spatial allocation factor queries are the YEAR_GS and YEAR_SAF tables containing gridded values for each of the demographic databases. These tables are merged and displayed in the Level_3_Databases.

Example Calculation: Population

Total County Pop = 50,000



Pop of TAZ = 5,000

Pop of Grid Cell (x) = 500

SAF = Population of Grid Cell / Population of County

= 500 / 50,000

= 0.01

1% of the county total population resides in grid cell (x)

Figure 3-4. Example of the spatial allocation factor calculation.

3.5.2 Level 2 Land Use and Land Cover Databases

The following are Level 2 land use and land cover databases:

- DWR_Ag.mdb (DWR data for agricultural land)
- DWR_Crp.mdb (DWR data for cropland)
- DWR_Liv.mdb (DWR data for livestock)
- DWR_Pst.mdb (DWR data for pasture land)
- LULC.mdb (Contains both USGS land cover and DWR data)
- RURAL.mdb (Rural land)
- URBAN.mdb (Urban land)
- WATER.mdb (Recreational waters)
- ELEV5K.mdb (Elevation greater than 5000 ft)

The Level 2 land use and land cover databases use the output of the ArcInfo gridding processes to generate the gridded surrogate and spatial allocation factor files. Two land use and land cover data sets were obtained (from the Department of Water Resources [DWR] and the U.S. Geological Survey [USGS]) for agricultural and vegetation land cover. Preference was given to the DWR data set because it contains highly resolved agricultural data for the San Joaquin Valley. The USGS data set was used as a default for regions of the domain that are not covered by the DWR data set. The Level 2 DWR databases are preprocessors to the LULC database. The LULC database includes both the USGS data and the Level 2 pre-processed DWR data. **Table 3-4** provides a summary of the tables and queries contained in the Level 2 land use and land cover databases.

Table 3-4. Summary of tables and queries contained in the Level 2 land use and land cover databases.

Table Name	Description
AG1_GS_DWR	Gridded surrogate data for SSC AG1; table linked to Level 2 DWR_Ag.mdb.
AG2_GS_DWR	Gridded surrogate data for SSC AG2; table linked to Level 2 DWR_Ag.mdb.
AG4_GS_DWR	Gridded surrogate data for SSC AG4; table linked to Level 2 DWR_Ag.mdb.
LU3_GS_DWR	Gridded surrogate data for SSC LU3; table linked to Level 2 DWR_Pst.mdb.
USGS_Surgs_KM2	Gridded surrogate data for the USGS data set; table linked to /basemaps/USGS_LU/USGS_Data_CA.mdb.
Cnty_grid	File containing the area fractions of each county in each grid cell.
Stgrid2k	Polygon attribute table for the Stgrid2k coverage (statewide 2-km grid).

Table 3-4. Summary of tables and queries contained in the Level 2 land use and land cover databases.

Table Name	Description
USGS_DWR_GS	Final gridded surrogate data for the state combining both the USGS and DWR data.
USGS_DWR_SAF	Final gridded spatial allocation factor data for the state combining both the USGS and DWR data.
Query Name	**Description**
0-Build USGS_DWR_GS Table	Make Table query that creates table named USGS_DWR_GS containing all grid cells in the domain.
1-Update_DWR_AG1	Update query that updates all the grid cells with DWR data for SSC AG1 (same for Update queries 2-4)
5-Update_USGS_AG1	Update query that updates all null grid cells with USGS data for SSC AG1 (query selects null grid cells or cells with no DWR data) (same for Update queries 6-12).
USGS_DWR_CNTY	Assigns a county FIPS code to each record in the USGS_DWR_GS table.
CNTY_TOTALS	Computes county totals for each surrogate.
USGS_DWR_SAF_CALC	Make Table query that calculates the spatial allocation factor file.

The first step in the Level 2 land use and land cover database processing was to link the gridded surrogate files from the Level 2 DWR databases and the Level 1 USGS database. The Level 2 DWR databases contain the gridded surrogate data calculated from the gridded output file from ArcInfo. The USGS data were received from USGS in a pre-gridded format. A program was used to re-grid the data to the California statewide 2-km grid domain. The Level 1 USGS database contains the re-gridded data and surrogate file. The Level 2 LULC database contains both the DWR and USGS gridded surrogate data. The queries in the LULC database merge the two data sets giving preference to the DWR data. The first query, 0-Build USGS_DWR_GS Table, creates a blank table with all grid cells in the domain called USGS_DWR_GS.

There are twelve update queries in the database (numbered 1 – 12) that update the USGS_DWR_GS table with surrogate data values for each grid cell. The first four queries update the USGS_DWR_GS table with DWR surrogate data. Queries 5 – 12 update the USGS_DWR_GS table with USGS surrogate data. Queries 5 – 12 are designed to update only grid cells with null values or grid cells with no DWR data. After running the twelve queries, the gridded surrogate file was complete. The spatial allocation factor file, USGS_DWR_SAF, was generated by the USGS_DWR_SAF_CALC query. The spatial allocation factors were

calculated by dividing the gridded surrogate values for the county total surrogate value for each SSC.

The structure and functionality of the URBAN, RURAL, and WATER databases is the same. **Table 3-5** contains the generalized tables and queries contained in each database.

Table 3-5. Summary of tables and queries contained in the Level 2 URBAN, RURAL, and WATER databases.

Table Name	Description
GRIDDED_INPUT	Gridded surrogate data, output from ArcInfo AML gridding scripts. Contains the area fraction of each land use and land cover type (i.e. urban, rural, water) in each grid cell.
Land cover_pat	Polygon attribute table for each land cover type.
SSC_GS	Gridded surrogate data for each SSC.
SSC_SAF	Gridded spatial allocation factor data for each SSC.
Query Name	**Description**
Update_Gridded_Input (1)	Updates GRIDDED_INPUT table with GRID_ID, XREF, and feature area or length.
Update_Gridded_Input (2)	Updates GRIDDED_INPUT table with county FIPS code.
SSC_Totals	Calculates the surrogate totals by county
SSC_GS_CALC	A Make Table query that calculates the gridded surrogate file and creates SSC_GS table.
SSC_SAF_CALC	A Make Table query that calculates the gridded spatial allocation factor file and creates SSC_SAF table.

The first step in the Level 2 URBAN, RURAL, and WATER database processing was to import the gridded data file from the ArcInfo gridding processes. The ArcInfo attribute table for each coverage was also imported into the database. The gridded input table was modified to include the following four fields: GRID_ID, XREF, AREA_KM2, and CNTYFIPS. Update queries 1 and 2 populate these fields with data values. The SSC_Totals query calculates the surrogate totals by county. The SSC_GS_CALC query is a Make Table query that calculates gridded surrogate values by grid cell and creates the SSC_GS table. The SSC_SAF_CALC query is a Make Table query that calculates spatial allocation factors by grid cell and creates the SSC_SAF table. The spatial allocation factor file divides the gridded surrogate values by the county total surrogate value which results in a weighting factor representing the fraction of each surrogate in each grid cell based on the county total. All spatial allocation factors for each county add up to 1 (or 100 if expressed as a percent).

3.5.3 Level 2 Transportation Databases

The Level 2 transportation databases use the output of the ArcInfo gridding processes to generate the gridded surrogate and spatial allocation factor files. The following are Level 2 transportation databases:

- ALLAIR.mdb (All airports)
- COMMAIR.mdb (Commercial airports)
- RAILROAD.mdb (Railroad network)
- ROADS.mdb (Road network)
- SHIPPING.mdb (Ports and shipping lanes)

Table 3-6 provides a summary of the tables and queries contained in the Level 2 transportation databases.

Table 3-6. Summary of tables and queries contained in the Level 2 transportation databases.

Table Name	Description
Surgname_pat	Polygon attribute table for the surrogate coverage contained in each database.
GRIDDED_INPUT	Gridded surrogate data, output from ArcInfo AML gridding scripts. Contains the area fraction or length fraction of each transportation feature (e.g., length of road, area of airport) in each grid cell.
SSC_GS	Gridded surrogate data for specified SSC.
SSC_SAF	Gridded spatial allocation factor data for specified SSC.
Query Name	**Description**
Update_Gridded_Input (1)	Updates GRIDDED_INPUT table with GRID_ID, XREF, and feature area or length.
Update_Gridded_Input (2)	Updates GRIDDED_INPUT table with county FIPS code.
SSC_Totals	Calculates the surrogate totals by county.
SSC_GS_CALC	A Make Table query that calculates the gridded surrogate file and creates SSC_GS table.
SSC_SAF_CALC	A Make Table query that calculates the gridded spatial allocation factor file and creates SSC_SAF table.

The first step in the Level 2 transportation database processing was to add the following four fields to the GRIDDED_INPUT file: GRID_ID, XREF, AREA_KM (or KM2 for polygon features), and CNTYFIPS. These fields were populated using the two Update_Gridded_Input queries. The first update query populates the GRID_ID, XREF, and AREA fields and the second update query uses the feature attribute table (Surgname.pat) to add the CNTYFIPS. The next

step was to calculate the surrogate totals by county using the SSC_Totals query. The gridded surrogate files were calculated using the Make Table query SSC_GS_CALC which generates the table SSC_GS. The spatial allocation factor file, SSC_SAF, was generated by the SSC_SAF_CALC query. The spatial allocation factors were calculated by dividing the gridded surrogate values for the county total surrogate value for each SSC.

3.5.4 Level 2 Facility Location Databases

The Level 2 facility location databases use the output of the ArcInfo gridding processes to generate the gridded surrogate and spatial allocation factor files. The following are Level 2 facility location databases:

- FACILITIES.mdb (Facility locations: bulk plants, POTW, hospitals, institutional buildings, auto body/refinishing shops, drycleaners, gasoline service stations, restaurants/bakeries, wineries, landfills, sand/gravel mines)
- GOLFC.mdb (Golf courses)
- MILBASE.mdb (Military bases)
- OIL_GAS.mdb (Oil and gas wells and fields)

Table 3-7 provides a summary of the tables and queries contained in the Level 2 facility location databases.

Table 3-7. Summary of tables and queries contained in the Level 2 facility location databases.

Table Name	Description
Surgname_pat	Polygon attribute table for the surrogate coverage contained in each database.
GRIDDED_INPUT_SSC	Gridded surrogate data, output from ArcInfo AML gridding scripts. Contains the area fraction or length fraction of each transportation feature (i.e. number of facilities, area of facility) in each grid cell.
SSC_TOTALS	Surrogate totals by county for specified SSC.
SSC_GS	Gridded surrogate data for specified SSC.
SSC_SAF	Gridded spatial allocation factor data for specified SSC.
Query Name	Description
Update_Gridded_Input (1)	Updates GRIDDED_INPUT table with GRID_ID, XREF, and feature area or length.
Update_Gridded_Input (2)	Updates GRIDDED_INPUT table with county FIPS code.

Table 3-7. Summary of tables and queries contained in the Level 2 facility location databases.

Query Name	Description
SSC_Totals	Calculates the surrogate totals by county.
SSC_GS_CALC	A Make Table query that calculates the gridded surrogate file and creates SSC_GS table.
SSC_SAF_CALC	A Make Table query that calculates the gridded spatial allocation factor file and creates SSC_SAF table.

There are three processing methods for the facility location data. The first applies to surrogates that were derived from polygon coverage basemaps (GOLC.mdb and MILBASE.mdb). The second applies to surrogates that were derived from point coverages (FACILITIES.mdb). The third applies to the oil and gas database (OIL_GAS.mdb) which combines point locations and polygon coverage surrogate data.

The processing method involves adding the following four fields to the GRIDDED_INPUT file: GRID_ID, XREF, AREA_KM2, and CNTYFIPS. These fields were populated using the two Update_Gridded_Input queries. The first update query populates the GRID_ID, XREF, and AREA fields, and the second update query uses the feature attribute table (Surgname.pat) to add the CNTYFIPS. The next step was to calculate the surrogate totals by county using the SSC_Totals query. The gridded surrogate files were calculated using the Make Table query SSC_GS_CALC which generates the table SSC_GS. The spatial allocation factor file, SSC_SAF, was generated by the SSC_SAF_CALC query. The spatial allocation factors were calculated by dividing the gridded surrogate values for the county total surrogate value for each SSC.

The second processing method is simple and involves running the SSC_TOTALS_CALC Make Table query that calculates the total surrogate value by county and generates the SSC_TOTALS table. Next, the SSC_GS_CALC Make Table query was run which calculates the gridded surrogate data and generates the SSC_GS table. Lastly, the SSC_SAF_CALC Make Table query was run to calculate the spatial allocation factor data and generate the SSC_SAF table. Spatial allocation factors were calculated by dividing the gridded surrogate value by the county total surrogate value.

The third processing method, which applies to the oil and gas wells database (OIL_GAS.mdb), combines both gridded polygon and point location surrogate data. Locations of oil and gas wells were obtained for all counties in California except Kern. As a default, the locations of oil and gas fields were used in place of well locations for Kern County. The third processing method is simply a combination of the first and second processing methods. The first processing method (discussed above) was used to calculate gridded surrogate data for Kern County. The surrogate tables KG_GS and KO_GS are the Kern county gas and oil fields respectively. The second processing method (discussed above) was used to calculate gridded

surrogate and spatial allocation factor data for oil and gas well point locations for all other counties in California.

The gridded surrogate data for Kern County oil fields was appended to the gridded surrogate data for oil wells in all other counties using the Append_Kern_LOC6_GS query. The LOC6_GS_CALC_2 query was run to group the Kern appended records with all other counties. The surrogate totals by county were calculated using the LOC6_TOTALS query, and the spatial allocation factor file for LOC6 was generated by the Make Table query, LOC6_SAF_CALC. The same procedure was used to process gas wells and fields (LOC11).

3.6 LEVEL 3 DATABASES—DISPLAY OF GRIDDED SURROGATES AND SPATIAL ALLOCATION FACTORS

The Level_3_Databases contain spatial allocation factor data for selected data sets and ArcInfo displays. ArcInfo geodatabases were created by using MS Access to link the gridded surrogate data tables created in the Level_2_Databases to a new Level_3_Database. Queries are contained in Level_3_Databases to merge data sets so that data covering the entire state can be displayed. Each gridded surrogate table was then joined to the 2-km grid coverage in ArcMap to display the values of the individual grid cells across the state. The result of the ArcMap join was saved permanently as a new feature class within the geodatabase. The spatial allocation factors were similarly linked into Level 3 geodatabases and joined in ArcMap. An ArcInfo mapfile (.mxd) was created for each gridded surrogate and for each spatial allocation factor display. A printout of the gridded surrogate displays appears in Appendix D; the spatial allocation factor displays are included in Appendix E. The primary function of the Level_3_Databases is to display and quality assure the gridded surrogate and spatial allocation factor files. The data that is not displayed in the Level_3_Databases is displayed in the Level_2_Databases.

The following are Level 3 ArcInfo geodatabases:

- YEAR_DMO_CS.mdb (DMO and CS surrogate data by year)
- YEAR_DMO_SAF.mdb (DMO and CS spatial allocation factor data by year)
- YEAR_COMB_SAF.mdb (COMB spatial allocation factor data by year)
- FACILITIES.mdb (Facility surrogates and spatial allocation factors)
- LAND_USE.mdb (USGS and DWR land use surrogates and spatial allocation factors)
- ROADS.mdb (Road surrogates and spatial allocation factors)
- SHIPPING.mdb (Shipping/port surrogates and spatial allocation factors)

Combination Spatial Allocation Factors

Several combination (COMB) spatial allocation factors were developed by combining surrogates of different feature types. For example, COMB1 is a combination surrogate for total employment and road density. Total employment is geographically defined by TAZ polygon coverages and road density is a line coverage of the California road network. To combine these two types of land feature data, the two surrogate data files were combined and a new spatial

allocation factor was calculated. The spatial allocation factors for the combined surrogates (COMB1-COMB9) are in the YEAR_COMB_SAF Level 3 database.

The YEAR_COMB_SAF.mdb database was linked to the gridded surrogate files from the Level_2_Databases for each of the components of the combination surrogate. The queries contained in the database perform a series of manipulations and calculations to arrive at the spatial allocation factor files. **Table 3-8** provides a summary of the tables and queries contained in the YEAR_COMB_SAF database.

Table 3-8. Summary of tables and queries contained in the YEAR_COMB_SAF database.

Table Name	Description
Cnty_Grid	File containing the area fractions of each county in each grid cell.
SSC_GS	SSC gridded surrogate file linked from Level_2_Databases.
2000_GS_MERGE	Gridded surrogate file containing DMO and CS data. Linked from Level 3 2000_DMO_CS.mdb.
SSC_TOTALS	Combined surrogate totals by county.
SSC_SAF	SSC spatial allocation factor file.
SSC_SAF_Layer	ArcInfo feature class display (displayed by ArcMap).
Query Name	**Description**
SSC_CALC_1	Make Table query including surrogate data for all components of the combined surrogate. Generates SSC_GS table.
SSC_UPDATE	Update query that updates the surrogate fields of the SSC_GS table with 0 when cells are null.
SSC_TOTALS_CALC	Make Table query that sums the surrogate fields of the SSC_GS table and calculates the combines surrogate totals by county. Generates the SSC_TOTALS table.
SSC_CALC_SAF	Make Table query that calculates spatial allocation factors. Generates SSC_SAF table.
RUN_MERGE	Append query that merges all of the SSC_SAF files together and creates table MERGE_COMB_SAF.

The gridded surrogate files (SSC_GS) for each of the parameters (e.g., population and road density) that make up the combination surrogates were linked from the appropriate Level_2_Databases. The combination surrogates and spatial allocation factors were calculated by first creating a table containing all of the individual surrogate data values that make up the combination surrogate. The SSC_CALC_1 query does this by generating a table (SSC_GS)

containing all grid cells in the domain and the surrogate values associated with each grid cell. The SSC_UPDATE query updates the surrogate fields in the SSC_GS table with 0 when an empty cell appears. To calculate the spatial allocation factors, the county total surrogate values must be calculated. The SSC_TOTALS_CALC query performs this calculation and generates the SSC_TOTALS table. Finally, the spatial allocation factor file (SSC_SAF) was generated by the SSC_CALC_SAF query. The RUN_MERGE query appends each of the individual SSC_SAF tables into one table which was then displayed in ArcMap.

3.7 FUTURE-YEAR SURROGATES AND SPATIAL ALLOCATION FACTORS

Future-year spatial allocation factors were developed for all demographic and computed surrogates (DMO and CS). Appendix A contains a listing of spatial allocation factors created for future years. Future-year surrogates were not created in cases in which the spatial pattern of emissions is likely to remain unchanged or in cases for which there were inadequate data to calculate a spatial allocation factor. In cases for which future-year spatial allocation factors were not developed, the base-year surrogates were assigned to those emissions categories as the default. However, these assignments could be changed by ARB in the future. For example, the existing future-year surrogate assignment for gasoline service stations is the 2000 base-year locations of gas stations. Because it is likely that new gas stations will be built in areas of growing population or development, the future-year surrogate assignment for gas stations could be changed by the ARB to one of the future-year surrogates, such as population, or one of the computed construction surrogates.

Future-year surrogate data were not available for land cover and area source facility surrogates such as agricultural land, gasoline service stations, airports, and auto body shops. Future-year data for these types of surrogates were lacking; therefore, base-year surrogates were used for future years.

3.8 QUALITY ASSURANCE

Quality assurance protocols were carried out during each stage of data processing. Quality assessments of each surrogate basemap were made to confirm that the geographic location and content of each data set was correct within a specified range (as discussed in Section 3.2). The surrogate data tables and the calculation queries in each Level_1_Database were reviewed to ensure that the mapping of raw demographic data to the spatial surrogate scheme was consistent for all 12 demographic data sources. The surrogate and spatial allocation factor data contained in the Level_2_Databases were reviewed, and intermediate queries were created, to verify that the sum of each county's total spatial allocation factor added to 1 (or 100 if expressed as a percent).

A general quality assurance procedure consisted of displaying the basemap, gridded surrogate data, and spatial allocation factor data for each gridded surrogate, or SSC. The basemap display (Appendix C) was overlayed with the gridded display (Appendix D) and the spatial allocation factor map (Appendix E) to verify consistency among each level of data processing. When errors were found, corrections were made; and the quality assurance procedures were repeated until each stage of the data processing had been completed

successfully. Once all spatial allocation factor files had been checked for quality and consistency, the Master Databases were assembled.

3.9 PREPARATION OF FINAL MASTER DATABASES

The following are Master Databases:

- 2000_GRIDDED_SAFS.mdb
- 2005_GRIDDED_SAFS.mdb
- 2010_GRIDDED_SAFS.mdb
- 2020_GRIDDED_SAFS.mdb
- MASTER_SAFS.mdb
- SURROGATE_PROCESSING_FILES.mdb

The YEAR_GRIDDED_SAFS databases contain the gridded spatial allocation factor data for all spatial surrogates and counties by year. The MASTER_SAFS database contains all spatial allocation factor data for all spatial surrogates, counties, and years. The SURROGATE_PROCESSING_FILES database contains information about the spatial surrogates, grid information, and processing version. Tables defining the contents and field definitions for all Master Databases are included in each database.

4. REFERENCES

U.S. Census Bureau (1999) Expenditures for nonresidential improvements and repairs., <http://www.census.gov/const/nrsupp/intro1.html>, October.

U.S. Census Bureau (2000a) 1997 American Housing Survey., <http://www.census.gov/hhes/www/housing/ahs/97dtchrt/tab2-3.html>, September.

U.S. Census Bureau (2000b) Construction and Housing. <http://www.census.gov/prod/www/abs/cons-hou.html>, November.