

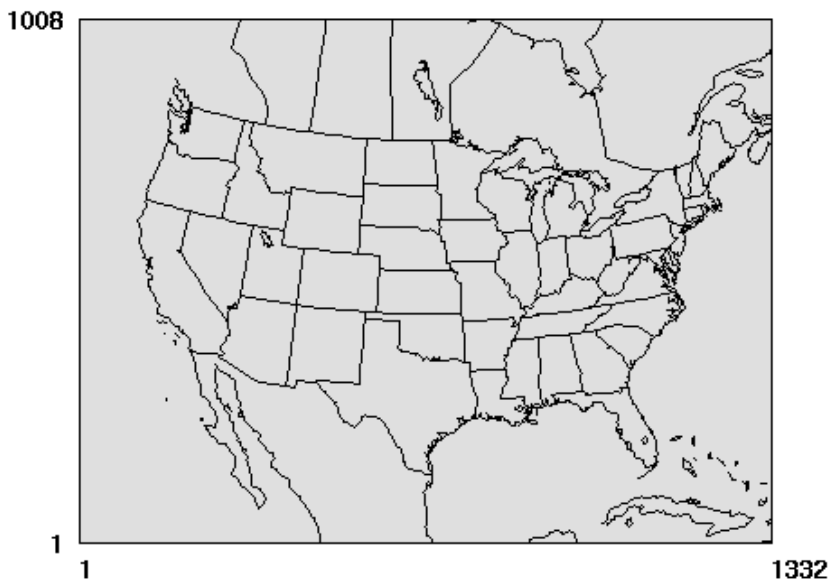
Process for Developing SMOKE-Ready Surrogate Files

Gridding surrogates are used to spatially allocate emission sources from a coarse geographic area to finer grid cells used for modeling. There can be hundreds of unique source categories in an emissions inventory, which is typically developed for counties, states, or other areas. The exact location of most major emission sources is known and their geographic coordinates are usually contained in the inventory. These usually are referred to as major point sources and include electric utilities and major industrial facilities. However, other emission sources are estimated for the entire county or other area as an aggregate since the exact locations of each source are not included in the modeling inventory. Surrogates are human activities or land use information that are used to represent a more precise location of emission source category groups. A gridded surrogate ratio is the ratio of the amount of a surrogate in a modeling grid cell to the total amount of that surrogate in a county. Grid cell emissions are calculated by multiplying the cell's gridded surrogate ratio by the county emissions.

The purpose of this project was to identify data sources that could be used as surrogates, grid the surrogates to a 4 km grid, and to create surrogate cross-references and spatial surrogate ratios that could be used with the SMOKE emissions processing system for a specified 4 km grid, a 12 km grid and a 36 km grid. The projection and 4 km grid specifications are provided below. Although part of Mexico is included in the grid, no surrogates for Mexico were created as part of this project. Only surrogates for the U.S. and Canada were created. This project resulted in the creation of approximately 60 different surrogates for U.S. emissions and approximately 40 for Canadian emissions. Tables 1 and 2 (at the end of this document) summarize the available surrogates for the U.S. and Canada, respectively.

4km surrogate grid

LCP Center -97, 40 (33,45)
SW corner = -2736, -2088 km



Projection information:
Projection: Lambert
Units: Meters

Parameters: Spheroid Sphere (This is internally defined in ArcInfo® as Sphere of Radius 6370997 meters)
1st Standard Parallel: 33 0 0.000
2nd Standard Parallel: 45 0 0.000
Central Meridian: -97 0 0.000
Latitude of Projection's Origin: 40 0 0.000
False easting (meters): 0.00000
False northing (meters): 0.00000

4 km Grid Specifications

Projection	Units	X Origin	Y Origin	X Cell Length	Y Cell Length	Columns	Rows	Alpha	Beta	Gamma	X Center	Y Center
Lambert Conformal	Meters	-2736000	-2088000	4000	4000	1332	1008	33	45	-97	-97	40

The 12 and 36 km grids are in the same projection and cover the same grid domain.

The SMOKE spatial surrogate files for U.S. emission sources were developed in order to provide surrogate data for every county regardless of whether the county has that geographic feature. For example, even though most counties in the U.S. have no orchards/vineyards, the orchards/vineyard surrogate is complete for all U.S. counties. This is because for ALL counties with no orchards/vineyards, we gap-filled using the agriculture land surrogate. We denote agriculture land as a “secondary surrogate” for orchards/vineyards. Users need to be aware that for gap-filled surrogates, the actual surrogate name (e.g., orchards/vineyards) is only applicable to counties that have that geographic feature. For counties without that geographic feature, the surrogate is actually the secondary or tertiary surrogate listed in the surrogate codes spreadsheet. The gap-filling spreadsheet lists, for each surrogate, which counties were gap-filled. Users who would like to change secondary surrogates or remove them can do so by either removing the data based on the gap-filling spreadsheet information or by re-computing the gridded ratios using the geodatabase containing the gridded information (described in step 1 below). In this second method, you will generate gridded ratios for counties with surrogate data only; if you would like to use secondary surrogates for counties without a particular surrogate feature, you will need to fill in this information. We provide the geodatabase through an ftp site you can get to from EPA’s CHIEF website – go to <http://www.epa.gov/ttn/chief/emch/spatial/newsurrogate.html>. Scroll down to “data files” to get to the ftp site.

The SMOKE-ready cross-reference (surrogate assignments to SCC codes) file for the U.S. was originally developed for a specific 1999 National Emissions Inventory (NEI) data set. This NEI data set consists of:

- (1) version 3 final (released Summer/Fall 2003) NEI for hazardous air pollutants (HAPs) and
- (2) version 2 final (released Fall 2002) NEI for criteria pollutants

However, we added, when preparing the May 2004 cross-reference file, additional SCC codes that reflect some (but not necessarily all) categories in the version 3 of the 1999 NEI and the draft 2002 NEI. Users of these latter inventories or other inventories should check to see if all county-level SCCs are covered by the cross-reference file for the U.S. that we supply. Revisions to the cross-reference file may be needed for other inventories if they contain emission sources (SCCs) that do not appear in the data sets described above.

The base data used for developing the SMOKE-ready gridded surrogates are contained in ArcGis® shape files. These files are located on EPA's CHIEF website – go to <http://www.epa.gov/ttn/chief/emch/spatial/newsurrogate.html> and scroll down to “data files” to get to the ftp site. The data sources are described in the first spreadsheet of the Surrogate_Documentation_Workbook.xls file. There is also a data dictionary on the ftp site that provides the projection parameters. The following shape files in this directory are not currently used for U.S. surrogates: us_nat_gas, (natural gas compressor stations and production plants), us_nf (national forests), and ind3b_draft16_no_ak_ok_trust_wgs84_dd (tribal land). In addition, the airport_point shape file is only used to gap fill the airport area surrogate. Furthermore, a number of the data sources listed in the “Sources of 2000 US Surrogates” are not used by themselves to make surrogates (e.g., theaters) however, these data may be used in combination with other data on the shapefile to make a surrogate (e.g., theaters are a component of the commercial land surrogate).

SMOKE-ready gridded surrogate files

The general process for creating the U.S. SMOKE-ready gridded surrogate files from the ArcGis shape files is as follows (see also Appendix A for an example):

1. Overlay the grid on the surrogates. Generate the 4 km grid polygons with the specifications above and spatially overlay (intersect) the grid onto the surrogate area polygons or points using ArcGis/Arcview 8.2. The resulting geodatabase contains, for each surrogate, the county FIPS code, the grid column and row number, and the amount (area, miles or count) of the county's portion of the surrogate in that cell.

Surrogates for a gridded domain with specifications different than those in the table above can be developed using this same process. A new grid would need to be generated in ArcGis using the new domain specifications and then overlaid onto the surrogate base data polygons. The base data shape files would need to be projected to the new specifications prior to the overlay.

2. Use PC SAS to extract and convert each geodatabase table generated by ArcGis to a SAS dataset. This was done manually using the import option from the drop-menu in SAS but a program can be written to perform the conversions in a batch mode. Each table contains the gridded area, miles or count in each county for a specific surrogate. The variables include FIPS code, column number, row number and area, miles or count.
3. Calculate surrogate ratios. Surrogate ratios are calculated for each surrogate using a series of SAS programs (surrogate_ratio_programs.tar) that are run under UNIX using a batch file called all.bat. The programs sum the surrogates for each county and calculate each the ratio by dividing the county cell surrogate value by the total county surrogate value. Combination surrogates where both are of the same type (i.e., Heavy and High Tech Industrial are both area) were summed prior to calculating the ratio. Combination surrogates with unlike data (i.e., 3/4 Roadway Miles plus 1/4 Population are line and area data) were summed after calculating the ratios and then normalized.

The surrogate cross-reference code is also assigned here. The output of this step is one SAS dataset containing the surrogate ratios for each unique surrogate code.

4. Gap-fill surrogates for counties missing data. There will be many instances where inventory emissions exist for a particular county but there is no data, for that county, for the surrogate assigned. For example, a county with class 1 locomotive emissions may not have data for the class 1 railroad surrogate. In this case we have selected to incorporate, within the assigned surrogate, a different source of data (a different surrogate) for that particular county. We incorporate secondary surrogates even if there is no emission source that requires it for that particular county. We denote this process as “gap-filling.” All surrogates resulting from the gap-filling process have ratios for all counties.

For each surrogate, we assign a secondary or tertiary surrogate where needed for gap-filling. For the class 1 railroad surrogate example mentioned above, we chose total railroads as the secondary surrogate. The secondary or tertiary surrogate chosen would be the same across all counties and apply to all SCCs that use the particular primary surrogate. The program `gap_fill.sas` performs the gap-filling. The program will pull in and substitute the secondary surrogate for counties where the primary surrogate is missing. Tertiary surrogates will then be assigned to those counties that are still without surrogates.

The program `gap_filling.sas` uses SAS macros to identify counties having no values for each surrogate. It then assigns the data based on the appropriate secondary or tertiary surrogate to these counties. A list of the counties that require secondary and tertiary surrogates is generated as the program runs. We use this list to provide an excel workbook that lists the specific counties gap-filled for each surrogate. A check to see that surrogate ratios for each county sum to approximately 1.00 is also performed in the program. Ratios will not always sum exactly to 1.00 due to rounding. However, SMOKE will normalize surrogates greater than 1.00. Canadian surrogate ratio totals will sum to less than 1.00 for those provinces not entirely within the grid domain. A SAS dataset for each surrogate is created at the completion of the program. Note that Canadian surrogates are not gap-filled.

5. Create SMOKE-formatted spatial surrogate files. The program `SMOKE_surrogate.sas` combines the surrogate data created in step 4 and creates a SMOKE-formatted surrogate file for the 4 km grid. This program also uses SAS macros to specify the surrogate code and the ratio variable name.

The “surgtool” SMOKE utility program was used to develop the 12 km and 36 km surrogate files from the 4 km file. Environment variables need to be set for `INFILE`, `OUTFILE`, and `GRIDDESC`. The `GRIDDESC` variable refers to the standard SMOKE `G_GRIDPATH` description. (See SMOKE documentation for further details). This program can also be used to create surrogate profile tables with other grid resolutions and domains provided the Lambert Conformal projection specifications are the same as those in the 4 km grid. A SAS program was written to normalize the surrogate ratios output from `surgtool` since rounding produced values that summed to less than 1.

Assignment of surrogates to SCCs for U.S. Emission Sources

SCC codes from the 1999 NEI data set described above were put into a spreadsheet along with the SCC description; additional SCC codes were added in May 2004, and some assignments were changed. We manually examined each SCC description and chose the appropriate surrogate to be used. We examined the surrogates used previously and, in some cases, a visual distribution of the surrogate information. In some cases, we consulted inventory developers. For most of the industrial and commercial-related SCCs, we matched the SIC code in the SCC definition with the appropriate Federal Emergency Management Agency (FEMA) commercial or industrial land use category. For example, 2304050000, 'Industrial Processes; Secondary Metal Production: SIC 33; Nonferrous Foundries (Castings); Total', we assigned the surrogate "Metals and Minerals industrial land use (IND4)" because this surrogate is based total square footage of all buildings classified as Metals/Minerals Processing Factories: SIC Codes: 10, 12, 13, 14 and 33. However, with some commercial-related SCCs, in particular those involved in miscellaneous food and commercial cooking categories, we used total commercial building square footage. For example, for SCC = 2302080000, 'Industrial Processes; Food and Kindred Products: SIC 20; Miscellaneous Food and Kindred Products; Total' and for SCC= 2302002000, 'Industrial Processes; Food and Kindred Products: SIC 20; Commercial Cooking - Charbroiling; Charbroiling Total' we assigned the surrogate "Commercial land", because we feel that these operations would be related to where commercial buildings would be geographically located.

The resulting surrogate assignment spreadsheet called `surrogate_assignments_us` links each SCC from the 1999 NEI data set to a surrogate. The SMOKE-formatted surrogate cross-reference file is based on this spreadsheet. The following paragraphs discuss particular surrogate assignment issues.

One issue we had in the assignments was in deciding the level of detail for which to use the data we collected. For agriculture for example, the National Land Cover Data (NLCD) distinguishes between "row crops and grains," "fallow land," and "pasture". After visually looking at these more detailed land classes and reviewing the accuracy assessment (<http://landcover.usgs.gov/accuracy/>), we decided to aggregate the more detailed agricultural land classes together into an "Agriculture with Orchards/Vineyards" and "Agriculture without Orchards Vineyards". We did not do such an analysis of the FEMA (e.g., IND1, IND2, etc.) surrogates, and instead are using the level of detail available. This contributes to the large number of surrogates we have, which in turn contributes to the large file sizes and the need for gap-filling. A potential next step for this project would be to analyze the differences between using aggregated FEMA data (e.g., total industrial land) versus the current approach of using the more detailed FEMA data. Further, alternatives to gap-filling may be pursued. Finally, priority categories could be examined to further explore any available surrogate information, even if the information covers a small part of the domain.

For construction-related sources, we felt that either a population or housing change would be a possible indicator of construction. We decided on housing change as we felt there would be less "spikes" in housing than in population. We also had readily available 1990 and 2000 housing data (from the census). Since we noticed that many urban areas had no housing change but would still likely have construction we decided to include population in this surrogate, so that the resulting surrogate is a combination of both housing change (between the 1990 and 2000 census) and population.

For airport-related sources, we originally did not develop surrogates or assignments because we didn't think these would be needed. This is because SMOKE 1.5 (and later versions) contains a new "area-to-point" process that allows SMOKE to grid these sources based on a database of airport locations and allocation factors. This airport location and allocation factor database was developed as part of a separate project (EMS-HAP version 3) and will be documented in the User's Guide (Appendix C), which is not yet published at this date. A paper summarizing the development of this approach will be presented at the 2004 Emission Inventory Conference (**International Emission Inventory Conference, "Working for Clean Air in Clearwater,"** Clearwater, Florida, June 7 - 10, 2004.) and will be available on the conference website (<http://www.epa.gov/ttn/chief/conferences.html>).

Two issues contributed to our decision to develop airport surrogates:

- We wanted to provide the flexibility for users to use different emission processing systems which may not have the "area-to-point" capability,
- We wanted to enable users to utilize the new airport boundary data we found which allows emissions associated with a particular airport to be gridded over several different grid cells rather than just the single grid cell containing the airport centroid. This would be applicable in cases where the grid cells are small and the airport is large.

Because we did not feel the airport area data we obtained was complete (it had 2400 airports), we also developed an airport points surrogate (for gap-filling) and a military aircraft surrogate (which provided point locations of military airports).

It has recently (October 2003) come to our attention that the airport location data in the airport area shapefile and the airport point shape file is not consistent for about 35 counties with the airport location data (county allocations) used in developing the aircraft emission estimates for the 1999 NEI versions 2 and 3. For example, the inventory has the Philadelphia International Airport (PHL) aircraft emissions inventoried in Philadelphia County whereas the airport point and shapefiles place this airport in Delaware County. In reality, the airport is in both counties (although most of it is in Delaware County). This inconsistency is a data limitation for the airport area and airport point shape files. This inconsistency was fixed in the ancillary files developed to support the SMOKE area-to-point approach, so if you are using SMOKE, we recommend you use that approach to allocate airport-related emissions.

Changes to surrogate information in May 2004

Up until and including May 2004, full scale replacements of the originally posted (August 2003) "new surrogate data" have been made, and the older information is no longer available. While most of the replacements have been due to correcting errors in the original information, this latest set of files also represents improvements to two particular surrogates: ports and navigable waterways. Efforts made to improve the ports and navigable waterways surrogates resulted from a better understanding of how county-level emissions were developed for the 1999 NEI. These efforts are discussed in the below paragraphs.

Between October 2003 and January 2004, we embarked on a project to improve upon particular surrogates applied to locomotive and marine vessel-related emission categories by using consistent geographic data that were used in developing the 1999 National Emission Inventory (Version 3) for these categories. For commercial marine emissions on navigable

waterways and for locomotives, the NEI allocated emissions to counties based on activity information associated with the waterway and railroad links.

Railroad Surrogate- Not Updated

We determined that there was no significant gain in revising the railroad surrogate because of a number of reasons (we were not able to obtain the appropriate shape files used in the NEI development; the data on rail activity we obtained had a large number of links without activity data and appeared high, etc.) and we chose to stick with the previously developed (October 2003) surrogate which is based on number of railway miles.

Updated Ports Surrogate

For ports, we developed an improved shape file that contained port locations for counties which previously did not have any port information. We determined that counties for which emission estimates were made did not contain point data on ports in the original shape file. The original (pre May 2004) shape file was from the Bureau of Transportation Statistics (BTS) (<http://www.bts.gov/gis/>). We determined the basis for the commercial marine vessel port emissions using documentation (“Documentation for Aircraft, Commercial Marine Vessel, Locomotive, and Other Nonroad Components of the National Emissions Inventory, Volume I: Methodology - October 2003”, available at ftp://ftp.epa.gov/pub/EmisInventory/finalnei99ver3/criteria/documentation/nonroad/99nonroad_voll_oct2003.pdf) and discussing the approach with the inventory developers. As a result, we sought additional port locations in those counties with port emissions that we could add to the original BTS shape file. We identified a contact from the Tennessee Valley Authority, Carolyn Koroa, Geographic Analyst, Navigation & Hydraulic Engineering who emailed us (11/13/2003) a shape file containing additional port data (principalports2001). We added ports only for counties for which the commercial marine vessel port emissions were computed by the EPA methodology described in the above mentioned documentation and for which the original BTS file contained no ports. Note that even after adding ports to the file, there were still counties with 1999 NEI port emissions that did not contain port locations.

Updated Navigable Waterways Surrogate

The original (October 2003) navigable waterways surrogate was based on the miles of waterways obtained from a BTS shape file obtained from http://www.bts.gov/gis/download_sites/ntad02/newusdownloadform.html. We improved this surrogate by using activity data (activity being the number of short tons of cargo carried in a particular waterway) that is associated with the waterway segments. Activity data from the US Army Corps of Engineers (2001 Waterway Network Link Commodity Data) were used. (<http://www.iwr.usace.army.mil/ndc/data/>). The activity information was joined to the BTS spatial coverage based on the water ID which was common to both files. We determined this would be a more accurate way to spatially allocate navigable waterway commercial marine emissions since this was the approach used in allocating national-level emissions to counties. Note however, that because we were unable to obtain the actual shapefiles used in the NEI development, we are consistent in approach only, not in the actual data. We examined the new navigable waterway activity surrogate to make sure it made sense. We looked at 36km and 4km grid cell values and found very large differences with the 4km grid cells. For the 36km grid example, we chose a county with large differences (FIPS=22051, which is Jefferson Parish, Louisiana) and looked at surrogates based on the length of waterway versus the new activity approach.

For the length of the waterway surrogate, most of the grid cells in the county were receiving the same ratio value, whereas the activity weighted approach created higher ratio values in certain grid cells. The grid cells with higher ratios were associated with the Mississippi Rivers and Gulf intracoastal waterways, whereas a grid cell with no activity (zero ratio) was a waterway segment feeding Lake Ponchartrain. Previously, a waterway segment feeding Lake Ponchartrain would have received a similar emission ratio because it had a length comparable to the Mississippi River in this particular county. By using activity to differentiate between these two waterway segments, the new activity-weighted surrogate now concentrates the emissions along the Mississippi river (or in other parts of the country, rivers that have shipping activity.) This appears to be a more valid way to distribute the counties' marine/waterway emissions because one would expect significantly more emissions to be associated with grid cells containing the larger, more traveled, waterways than the smaller ones.

Canadian Surrogates

A different set of surrogates was used for Canada to be compatible with their emissions inventory. The base data for most of the surrogates are proprietary and were not available to the US EPA for gridded surrogate development. Shape files of Canadian 10km gridded surrogates were provided by Environment Canada and then reprojected to Lambert Conformal using the grid specifications above and regridded to 4 km. These were developed for use with province-level emissions. The document "Canada Emissions Distribution Techniques," provided by Environment Canada, describes the techniques used by Environment Canada in the development of their 10 km gridded surrogates. The process for creating the surrogate ratios and profile table is the same as the US procedure above except that gap-filling was not performed for the Canadian surrogates. Because the surrogates provided by Canada were for Province-level emissions, our surrogates are also designed for Province-level emissions.

A table of SIC-based source category -to- surrogate relationships was provided by Environment Canada and was used to develop the surrogate cross-reference file with the exception of onroad mobile emissions. With the exception of the vehicle type surrogates described below, the cross-reference file for Canada was created by merging Canada_SIC_SCC.xls (the SIC to SCC crosswalk) with the file Canadian_surrogates.xls (the SIC to surrogate crosswalk). Both of these crosswalk files were provided by Environment Canada.

Since the onroad emission source categories provided by Canada were only stratified by vehicle type and the surrogate data Canada supplied is based only on road class, it was necessary to develop vehicle type surrogates. These surrogates were developed as weighted roadway miles across all road classes. The weights are different for each vehicle type, and are based on the amount of VMT that vehicle travels on each of the road types. Surrogates for the seven vehicle types (LDGV, LDGT, HDGV, MC, LDDV, LDDT, and HDDV) were developed by applying the US percentages of vehicle type - roadway type ratio to the four Canadian road classes.

Vehicle Type – Roadway Ratios

Vehicle Type	Highway (US. Rd Class: freeways/interstates)	Primary Road (US. Rd Class: Arterials)	Secondary Road (US. Rd Class: collectors)	Residential Street (US. Rd Class: local)
LDGV	.302	.425	.145	.128
LDGT	.301	.425	.147	.128
HdGV	.287	.418	.171	.124
MC	.297	.423	.153	.127
LDDV	.302	.425	.145	.128
LDDT	.301	.425	.147	.128
HDDV	.277	.413	.189	.121

Canadian surrogate files contain the ratio of the grid cell surrogate to the total province surrogate value. Ratios will sum to less than 1 for those provinces not entirely within the unified grid domain. The surrogate profiles identified above can only be used with a province level emission inventory.

Quality Assurance

The following steps were taken during the development of the gridded surrogate files to report and correct potential errors.

- The gap-filling program reports instances where ratios that do not sum to approximately 1 for each county. These reports were reviewed for each surrogate and none were found.
- Canadian surrogate programs also reported instances where surrogate ratio sums did not exceed 1 for each province. There were a few cases which were renormalized.
- The gap-filling program also reports all counties with emissions having a primary, secondary, or tertiary gridded surrogate.
- Maps of the 4km gridded surrogates generated using PAVE software were visually reviewed for reasonableness. (Note this was done on July/August 2003 version of surrogates. It is assumed that the corrections made would still deem the surrogates reasonable)
- SMOKE was run with the 1999 NEI and the new 36 km surrogates to ensure that the gridding module did not generate warnings or errors. PAVE plots of gridded hourly emissions for one day were also reviewed. Note this was done on the July/August 2003 version of surrogates. It is assumed that the corrections made would have the same results.)

Note: SAS is a registered trademark of SAS Institute, Inc.; ArcInfo and ArcGIS are registered trademarks of Environmental Systems Research Institute, Inc.

Table 1. Available U.S. Surrogates For Lambert Conformal Projection Grid
LCP center: -97,40 (33,45), SW corner = -2736, -2088 km
 Abbreviations used in Table are listed below

SURROGATE	CODE	DESCRIPTION	Secondary SURROGATE used for gap-filling	Tertiary SURROGATE used for gap-filling	Quaternary SURROGATE Used for gap-filling	Approx. Number of counties that needed gap-filling (conterminous US)
Population	100	Number of people, U.S. Census, 2000				0
Housing	110	Number of housing units, U.S. Census, 2000				0
Urban Population	120	Amount of population considered "urban", based on U.S. Census, 2000	Population			712
Rural Population	130	Amount of population considered "rural", based on U.S. Census, 2000	Population			33
Housing Change and Population	140	Combination of equal weightings of 1) Positive change in gridded housing from 1990 to 2000 (census), and 2) Population (census 2000)				
Residential Heating - Natural Gas	150	Number of units with home heating as utility gas, from Census 2000	Housing			8
Residential Heating - Wood	160	Number of units with home heating as wood from Census 2000	Housing			34
Residential Heating - Distillate Oil	170	Number of units with home heating as fuel oil from Census 2000	Housing			166
Residential Heating - Coal	180	Number of units with home heating as coal from Census 2000	Housing			1790
Residential Heating - LP Gas	190	Number of units with home heating as bottled gas from Census 2000				
Urban Primary Road Miles	200	Road miles of urban primary roads (TIGER 2000, CFCC classes A11-A18, A21-A28 plus A63, overlaid with US Census 2000 urban areas)	Total Road Miles	Population		896
Rural Primary Road Miles	210	Road miles of rural primary roads (TIGER 2000, CFCC classes A11-A18, A21-A28 plus A63, overlaid with US Census 2000 rural areas)	Total Road Miles			182
Urban Secondary Road Miles	220	Road miles of urban secondary roads (TIGER 2000, CFCC classes A31-A38 plus A63, overlaid with US Census 2000 urban areas)	Total Road Miles			979
Rural Secondary Road Miles	230	Road miles of rural secondary roads (TIGER 2000, CFCC classes A31-A38 plus A63, overlaid with US Census 2000 rural areas)	Total Road Miles	Population		101
Total Road Miles	240	Sum of urban primary, rural primary, urban secondary and rural secondary (doesn't include local roads)	Population			2
Urban Primary plus Rural Primary	250		Total Road Miles	Population		166
3/4 Roadway Miles plus 1/4 population	255	Combination of population surrogate (Census 2000) and total road miles surrogates				

SURROGATE	CODE	DESCRIPTION	Secondary SURROGATE used for gap-filling	Tertiary SURROGATE used for gap-filling	Quaternary SURROGATE Used for gap-filling	Approx. Number of counties that needed gap-filling (conterminous US)
		(Tiger 2000), weighted ¾ for roadway and ¼ for population				
Total Railroad Miles	260	Sum of railroad miles of class 1, class 2&3, and unknown classified railroads from BTS/NTA Data from 2001	Total Road Miles	Population		367
Class 1 Railroad Miles	270	railroad miles of class 1 railroads from BTS/NTA Data from 2001	Total Railroad Miles	Total Road Miles	Population	886
Class 2 and 3 Railroad Miles	280	railroad miles of class 2&3 railroads from BTS/NTA Data from 2001	Total Railroad Miles	Total Road Miles	Population	1586
Low Intensity Residential	300	Land area classified as low intensity residential from NLCD (1992 Landsat imagery)	Single Family Dwellings (RES1)			6
Total Agriculture	310	Sum of land area of the following NLCD (1992 Landsat imagery) classes: pasture/hay, grains, row crops, fallow land and orchards/vineyards	Rural Land Area			15
Total Agriculture without Orchards/Vineyards	311	Sum of land area of the following NLCD (1992 Landsat imagery) classes: pasture/hay, grains, row crops and fallow land	Rural Land Area			16
Orchards/Vineyards	312	Land area classified as orchards/vineyards from NLCD (1992 Landsat imagery)	Total Agriculture	Rural Land Area	Land Area	2901
Forest Land	320	Land area classified as forest from NLCD (1992 Landsat imagery)	Rural Land Area			67
Strip Mines/Quarries	330	Land area classified as strip mines/quarries from NLCD(1992 Landsat imagery)	Mines	Rural Land Area	Land Area	898
Land	340	All area not classified as open water from NLCD (1992 Landsat imagery)				
Water	350	Areas classified as open water from NLCD (1992 Landsat imagery)	Navigable Waterway Miles			12
Rural Land Area	400	Land Areas not designated as an urbanized area or an urban cluster by Census 2000	Land			12
Commercial Land	500	Sum of FEMA building square footage in each 1990 census tract, from the following categories: COM1, COM2, COM3, COM4, COM5, COM6, COM7, COM8 and COM9				
Industrial Land	505	Sum FEMA building square footage in each 1990 census tract, from the following categories: IND1, IND2, IND3, IND4, IND5 and IND6				
Commercial plus Industrial	510	Sum of FEMA building square footage in each 1990 census tract, from all categories that make up commercial and industrial land surrogates				
Commercial plus Institutional Land	515	Sum of FEMA building square footage in each 1990 census tract, from all categories that make up commercial land plus the following FEMA categories: RES5, RES6, EDU1, EDU2, and REL1				
Commercial plus Industrial plus	520	Sum of FEMA building square footage in each 1990 census				

SURROGATE	CODE	DESCRIPTION	Secondary SURROGATE used for gap-filling	Tertiary SURROGATE used for gap-filling	Quaternary SURROGATE Used for gap-filling	Approx. Number of counties that needed gap-filling (conterminous US)
Institutional		tract, from all categories that make up commercial land and industrial land plus the following FEMA categories: RES5, RES6, EDU1, EDU2, and REL1				
Golf Courses plus Institutional plus Industrial plus Commercial	525	Combination surrogate made up of equal weightings of 1) number of golf courses from ESRI golf course data, and 2) Institutional plus industrial plus commercial surrogate	Population			
Residential_High Density (RES3 + RES4 + RES5 + RES6)	530	Sum of Federal Emergency Management Agency (FEMA) building square footage in each 1990 census tract, from the following FEMA residential categories: RES3, RES4, RES5 and RES6.				
Residential (RES1-4) + Commercial + Industrial + Institutional + Government	535	Sum of FEMA building square footage in each 1990 census tract, from all FEMA categories making up the commercial and industrial surrogates plus the following FEMA categories: RES1, RES2, RES3, RES4, EDU1, EDU2, REL1, GOV1, GOV2				
Retail Trade (COM1)	540	FEMA building square footage in each 1990 census tract, from retail trade (COM1)				
Personal Repair (COM3)	545	FEMA building square footage in each 1990 census tract, from personal repair (COM3)	Commercial land			1
Retail Trade (COM1) plus Personal Repair (COM3)	550	Sum of FEMA building square footage in each 1990 census tract, from retail trade (COM1) and personal repair (COM3)				
Professional/Technical (COM4) plus General Government (GOV1)	555	Sum of FEMA building square footage in each 1990 census tract, from Professional/Technical (COM4) and General Government (GOV1)	Commercial land			
Hospital (COM6)	560	FEMA building square footage in each 1990 census tract, from hospitals (COM6)	Medical Office/Clinic (COM7)	Commercial Land		466
Medical Office/Clinic (COM7)	565	FEMA building square footage in each 1990 census tract, from Medical Office/Clinic (COM7)	Hospital (COM6)			11
Heavy and High Tech Industrial (IND1 + IND5)	570	Sum FEMA building square footage in each 1990 census tract, from Heavy and High Tech Industrial (IND1 + IND5)	Total Industrial			15
Light and High Tech Industrial (IND2 + IND5)	575	Sum of FEMA building square footage in each 1990 census tract, from Light and High Tech Industrial (IND2 + IND5)	Total Industrial			5
Food, Drug, Chemical Industrial (IND3)	580	FEMA building square footage in each 1990 census tract, from Food, Drug, Chemical Industrial (IND3)	Industrial Land			115
Metals and Minerals Industrial (IND4)	585	FEMA building square footage in each 1990 census tract, from Metals and Minerals Industrial (IND4)	Industrial Land			207
Heavy Industrial (IND1)	590	FEMA building square footage in each 1990 census tract, from Heavy Industrial (IND1)	Total Industrial			15

SURROGATE	CODE	DESCRIPTION	Secondary SURROGATE used for gap-filling	Tertiary SURROGATE used for gap-filling	Quaternary SURROGATE Used for gap-filling	Approx. Number of counties that needed gap-filling (conterminous US)
Light Industrial (IND2)	595	FEMA building square footage in each 1990 census tract, from Light Industrial (IND2)	Total Industrial			5
Industrial plus Institutional plus Hospitals	596	Sum FEMA building square footage in each 1990 census tract, from all FEMA categories making up the industrial surrogate plus the following categories: RES5, RES6, EDU1, EDU2, and REL1 plus hospitals (COM6)				
Gas Stations	600	Number of gas stations from Spatial Insights – business counts data base based on year 2000 data	Commercial Land			105
Refineries and Tank Farms	650	Number of refineries and tank farms from FEMA Database published 1999	Industrial Land			2604
Refineries and Tank Farms and Gas Stations	675	Sum of gas stations, refineries and tank farms	Commercial plus Industrial			104
Airport Area	700	Area of airports from ESRI CD containing 2400 airports, 1999	Airport Point	Population		1316
Airport Point	710	Number of airports based on a combination of a variety of data sources including the Federal Aviation Administration and BTS, data ranges in vintage from 1996 to 2001	not gap-filled because not used as a surrogate			
Military Airports	720	Number of military airports from BTS data from 2001	Airport Point			2773
Marine Ports	800	Number of marine ports from a shapefile combining data from several sources:	Navigable Waterway Miles	Water	Land Area	2646
Navigable Waterway Activity	810	Miles and activity of waterways - navigable inland and intracoastal waterways.	Navigable Waterway Miles	Ports	Water	2313
Golf Courses	850	number of golf courses from the USGS geographic names information system, 2000.	Housing			2219
Mines	860	Number of mines from the Minerals Information Team of USGS, 1998.	Strip Mines/Quarries	Rural Land Area		998
Wastewater Treatment Facilities	870	Number of Waste Water Treatment plans from a FEMA databased containing over 1200 facilities (published 1999)	Commercial plus Industrial			2572
Drycleaners	880	Number of drycleaners from a zip code business patterns database from the US Census Bureau Economics and Statistics Administration containing 1997 data	Commercial Land			307
Commercial Timber	890	Number of possible timber removal locations from a US Forest Service Forest Inventory Analysis – data from mid to late 90's	Forest Land	Rural Land Area		513

Abbreviations Used in Above Table:

CFCC: census feature class code. Used to identify the most noticeable characteristic of a feature.

CFCC descriptions:

A11 Primary road with limited access or interstate highway, unseparated

A12 Primary road with limited access or interstate highway, unseparated, in tunnel
A13 Primary road with limited access or interstate highway, unseparated, underpassing
A14 Primary road with limited access or interstate highway, unseparated, with rail line in center
A15 Primary road with limited access or interstate highway, separated
A16 Primary road with limited access or interstate highway, separated, in tunnel
A17 Primary road with limited access or interstate highway, separated, underpassing
A18 Primary road with limited access or interstate highway, separated, with rail line in center
A21 Primary road without limited access, US highways, unseparated
A22 Primary road without limited access, US highways, unseparated, in tunnel
A23 Primary road without limited access, US highways, unseparated, underpassing
A24 Primary road without limited access, US highways, unseparated, with rail line in center
A25 Primary road without limited access, US highways, separated
A26 Primary road without limited access, US highways, separated, in tunnel
A27 Primary road without limited access, US highways, separated, underpassing
A28 Primary road without limited access, US highways, separated, with rail line in center
A63 Access ramp, the portion of a road that forms a cloverleaf or limited access interchange
A31 Secondary and connecting road, state highways, unseparated
A32 Secondary and connecting road, state highways, unseparated, in tunnel
A33 Secondary and connecting road, state highways, unseparated, underpassing
A34 Secondary and connecting road, state highways, unseparated, with rail line in center
A35 Secondary and connecting road, state highways, separated
A36 Secondary and connecting road, state highways, separated, in tunnel
A37 Secondary and connecting road, state and county highways, separated, underpassing
A38 Secondary and connecting road, state and county highway, separated, with rail line in center

BTS: Bureau of Transportation Statistics

BTS/NTA: Bureau of Transportation Statistics/National Transportation Atlas

NLCD: National Land Cover Data

FEMA: Federal Emergency Management Agency

USGS: U.S. Geological Survey

FEMA Building Square Footage Categories

RES1 Residential, Single Family Dwelling

RES2 Residential, Mobile Home

RES3 Residential, Mult Family Dwelling

RES4 Residential, Temporary Lodging (SIC code 70)

RES5 Residential, Institutional Dormitory

RES6 Nursing Home (SIC 805, 8052, 8059)

COM1 Commercial, Retail Trade (SIC 52, 53, 54, 55, 56, 57, 59)

COM2 Wholesale Trade (SIC 42, 50, 51)

COM3 Personal/Repair Services (SIC 72, 75, 76, 83, 86)

COM4 Prof./Technical Services (SIC 40, 41, 44, 45, 46, 47, 49, 61, 62, 63, 64, 65, 67, 73, 78 [except 7832], 81, 87, 89)

COM5 Banks (SIC 60)

COM 6 Hospital (SIC 8062, 8063, 8069)

COM7 Medical Office/Clinic (SIC 80 [except 8051, 8052, 8059, 8062, 8063, 8069])

COM8 Entertainment & Rec. (SIC 48, 58, 79 [except 7911], 84)

COM9 Theaters (SIC 7832, 7911)

COM10 Parking

IND1 Heavy (SIC 22, 24, 26, 32, 34, 35 [except 3571, 3572], 38, 39)

IND2 Light (SIC 23, 25, 27, 30, 31, 36 [except 3671, 3672, 3674], 38, 39)

IND3 Food/Drugs/Chemicals (SIC 20, 21, 28, 29)

IND4 Metals/Minerals Processing (SIC 10, 12, 13, 14, 33)

IND5 High Technology (SIC 3571, 3572, 3671, 3672, 3674)

IND6 Construction (SIC 15, 16, 17)

AGR1 Agriculture (SIC 01, 02, 07, 08, 09)

REL1 Church/Non/Profit Offices (SIC 86)

GOV1 General Services (SIC 43, 91, 92 [except 9221, 9224]), 93, 94, 95, 96, 97)

GOV2 Emergency Response (SIC 9221, 9224)

EDU1 Schools (SIC 82 [except 8221, 8222])

EDU 2 Colleges/Universities (SIC 8221, 8222)

Table 2 Available Surrogates for Canadian Province-Level Emissions

Canadian Surrogates - base data for each of these were provided by Canada

Total population	900
Total dwelling	901
Agricultural and related service industries	902
Fishing and trapping industries	903
Lgging and forestry industries	904
Mining (including milling), quarrying and oil well industries	905
Manufacturing industries	906
Construction industries	907
Transportation and storage industries	908
Communication and other utility industries	909
Wholesale trade industries	910
Retail trade industries	911
Finance and insurance industries	912
Real estate operator and insurance agent industries	913
Business service industries	914
Government service industries	915
Educational service industries	916
Health and social service industries	917
Accommodation, food and beverage service industries	918
Other service industries	919
Combination of Wholesale trade and Retail trade industries	920
Combination of Agricultural and Manufacturing industries	921
Combination of Forestry and Manufacturing industries	922
Combination of Mining and Manufacturing industries	923
Combination of Communication industries and Total Dwelling	924
Combination of Construction industries and Total Dwelling	925
Combination of Government industries and Total Dwelling	926
Ferry corridors.	927
Marine corridors.	928
Flight corridors, heavy jets.	929
Flight corridors, ligh & medium jets.	930
Minor airport location 50km buffer.	931
Rail corridors.	932
Forest fires.	933
Paved Roads	941
Unpaved Roads	942

Canadian Surrogates - computed by EPA from base data provided by Canada

Light Duty Gasoline Vehicles (LDGV)	934
Light Duty Gasoline Trucks (LDGT)	935
Heavy Duty Gasoline Vehicles (HDGV)	936
Motorcycles (MC)	937
Light Duty Diesel Vehicles (LDDV)	938
Light Duty Diesel Trucks (LDDT)	939
Heavy Duty Diesel Vehicles (HDDV)	940

APPENDIX A: Example Process for Creating gridded surrogates in GIS

Using population as an Example:

(see below figure as visual aid)

Overlay the population file with the chosen grid (make sure they both have the same projection)

Command Used: IDENTITY (other commands could also be used such as INTERSECT)

Calculate the new area of the polygons.

Commands Used: ADDITEM and CALCULATE

Calculate the ratio of the new area to the original block group area.

(New Polygon Area in hectares / AREA-H) = RATIO

Commands Used: ADDITEM and CALCULATE

Calculate new Population for each polygon

Commands Used: ADDITEM and CALCULATE

New Polygon Population = (POP2000 * RATIO)

Add a field that contains both the fips code and the column and row so that population can be summed by this field (CRFIPS) ie: FIPSSTCO = 48201 col = 125 row = 101 --- CRFIPS = 12510148201

Commands: ADDITEM and CONCATENATE (fips, col, and row are all defined as text)

Run statistics to sum New Polygon Population using CRFIPS as the case item.

Run statistics in original population file (pre-overlay) to sum population by county.

Below steps can be completed in SAS:

Create population surrogate ratios for each CRFIPS by dividing

Population in Grid Cell / County Population.

