# *Emissions Modeling Platform Spatial Surrogate Documentation*

**Contract No:** EP-D-12-044
**Work Assignment:** 3-03
**Version No:** 1.0


**Prepared for:** Rich Mason
U.S. EPA OAQPS
C339-02 USEPA Mailroom
Research Triangle Park, NC 27711

**Prepared by:** Zachariah Adelman
Institute for the Environment
The University of North Carolina at Chapel Hill
100 Europa Dr. Suite 490
Chapel Hill, NC 27517-1105


**Date due:** July 1, 2015

# Contents

# Acknowledgements

Ramboll Environ was a collaborator on this Work Assignment and worked extensively on the development and documentation of the oil and gas production surrogates. Ramboll Environ contributed to much of the content in Section 2.6 of this memo. Eastern Research Group, Inc. (ERG) contributed to this work through the development of the Gulf of Mexico platform and support vessel density Shapefile and through the development of oil and gas production Shapefiles and Surrogates. ERG contributed content to Section 2.6 of this memo. The Lake Michigan Air Directors Consortium (LADCO) provided the Midwest shipping lane and Inland Waterways Shapefile. The Georgia Department of Natural Resources provided the rail yard point inventory used for generating the rail yard Shapefile and surrogate. Computer Sciences Corporation (CSC) developed the Great Lakes tugboat zone Shapefile and provided support in the testing and QA/QC of the final surrogate data.


# 1 Introduction

This memo describes the collection, processing, and development of updated geospatial data for calculating spatial surrogates to be used in EPA emissions modeling platforms.

Spatial surrogates are used to allocate county level emission inventories into the rectangular grid cells used by urban and regional scale air quality models. Spatial surrogates are based on data at resolutions different from county-level, such as census tracts or road locations, and can therefore be used to allocate the emissions more specifically than assuming they are uniformly distributed through out the county. For example, motor vehicle emissions from interstate highways can be placed into the grid cells that intersect the highways themselves, and dry cleaning emissions can be allocated to parts of the county that have higher population, or the parts that have more dry cleaners – if data are available on specific locations of dry cleaners. Spatial surrogates consist of values between 0 and 1.0 that specify the fraction of the county emissions that should be allocated to each grid cell that intersects the county. The spatial surrogate fractions for each county sum to 1.0, except the sum may be less than 1 for some counties that intersect the edge and are not completely contained within the modeling domain. The emissions for a specific county within a grid cell are computed by multiplying the surrogate fraction for that county and grid cell by the total emissions for the county (Eyth et al., 2007). A diagram illustrating how spatial surrogates impact the emissions levels within the grid cells of a county is shown in Figure 1.

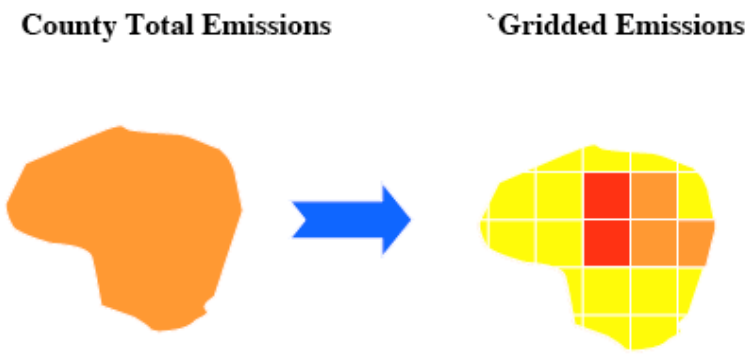


**Figure 1-1. Spatial surrogate schematic**

Spatial surrogates have three components:

1. Weight Shapefile – geospatial dataset used for allocating (or weighting) county-level emission inventory sources to a modeling grid. Examples include roadways, population density, or the point locations of marine ports. This memo describes the weight Shapefile updates completed under this Work Assignment.
2. Data Shapefile – geospatial dataset that corresponds to the administrative unit boundaries of the emission inventory. For this Work Assignment, the data Shapefile defines U.S. county boundaries. EPA specified that the year 2000 county boundaries used for the previous version of the national spatial surrogates be used as the data Shapefile for this Work Assignment.
3. Output grid or polygon – definition of the modeling grid(s) or polygon(s) upon which the spatial surrogates map emission inventories. For this Work Assignment, the three output grids include continental U.S. modeling grids at resolutions of 36-km, 12-km, and 4-km.

For air quality modeling studies prior to 2012, EPA used geospatial data primarily from the year 2000 and earlier for the weight Shapefiles that characterize the spatial distribution of non-point anthropogenic emissions sources. In November 2011 the University of North Carolina Institute for the Environment (UNC-IE) initiated updates to the spatial surrogate and Shapefile database used by the EPA Office of Air Quality Planning and Standards for emissions modeling. By July 2012 many of the original surrogates were updated with weight Shapefiles that reflected North American geospatial data collected since the year 2006. Incremental updates targeted specific inventory sectors, such as offshore commercial shipping, on-road mobile, and on-shore oil and gas production. The data and configurations used to update these surrogates are described here. All of the surrogates described in this memo were generated with the Surrogate Tool[1] of the Spatial Allocator (SA)[2].

The Fall 2011 release of the 2010 U.S. Census population and housing database provided the basis for several updated surrogates[3]. Year 2010 roadway data from TIGER and rail/port data

[1] http://www.cmascenter.org/sa-tools/documentation/4.0/html/srgtool/SurrogateToolUserGuide_4_0.htm

[2] http://www.cmascenter.org/sa-tools/

[3] http://www.census.gov/geo/www/tiger/tgrshp2010/tgrshp2010.html

from the National Transportation Atlas Database (NTAD)[4] also supported several surrogate updates. Several entirely new surrogates were created to support the processing of MOVES on-road mobile sources. The Extended Idle Locations surrogate (205) is designed for MOVES rate per hour (RPH) sources, which estimate the emissions from hotelling for long-haul trucks. The Urban Unrestricted Road (221) and Rural Unrestricted Road (231) surrogates that are designed for MOVES RPD unrestricted road sources are built from a combination of secondary and local roads data. The Off-Network Short-Haul Trucks surrogate (256) is derived from Federal Emergency Management Administration (FEMA) building locations of industrial and commercial buildings. The Off-Network Long-Haul Trucks surrogate (257) uses FEMA building locations for industrial and wholesale trade buildings. The Intercity Bus Terminals (258) and Transit Bus Terminals (259) surrogates use NTAD point locations of bus terminals.

Additional surrogate updates included home heating, building square footage, land-used based surrogates, oil and gas production, commercial businesses, and offshore shipping and ports. The 5-year American Community Survey (ACS) release in 2010 included updated data on home heating distributions by Census block[5]. The FEMA HAZUS-MH v2.0 was released in September 2011 and contains updated building square footage data that are applicable to all of the 500-series EPA surrogates[6]. The 30-m resolution, year 2006 National Landcover Database (NLCD) was used to update the land-use based surrogates, such as total land area, forest land, and crops[7]. Year 2008, 2011, and 2012 oil and gas production data from were aggregated to the continental U.S. (CONUS) modeling grid for different types of oil and gas production. Census block group counts of gas stations and dry cleaners from 2009 were used to create new surrogates for these commercial business categories. Several new datasets of commercial shipping lane activity, tugboat activity, and rail freight density were used to update the surrogates for the aircraft/locomotive/marine inventory sector. Table 1 lists the surrogate classes currently in the EPA emissions modeling platform that have been updated as part of this effort. Table 1 includes the years of the previous and updated geospatial data used to create the different classes of surrogates.

**Table 1. Summary of updates to the EPA surrogate database**

| Shapefile | Surrogates | Current Year | Previous Year |
|---|---|---|---|
| Population and Housing | 100-140 | 2010 | 2000 |
| Home Heating | 150-165 | 2005-2010 | 2000 |
| Road and Rail | 200-280 | 2010 | 2000 |
| Land-use/Land Cover | 300-350 | 2006 | 1992 |
| Building Square Footage | 500-595 | 2000-2006 | 1990 |
| Gas Stations and oil & gas | 600-699 | 2008-2012 | 2000 |
| Shipping | 800-820 | 2010 | 1999-2003 |
| Other industrial and commercial | 850-890 | 2010 | 2000 |

[4] http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/index.html

[5] http://www.census.gov/acs/www

[6] http://www.fema.gov/plan/prevent/hazus

[7] http://www.mrlc.gov/nlcd06_data.php

UNC-IE generated spatial surrogates on the CONUS domain at resolutions of 36-km, 12-km, and 4-km. All of the surrogates were generated on a Lambert Conformal Conic projection (lat_0 = 40° north, lon_0 = 97° west, true_lat1 = 33° north, true_lat2 = 45° north) assuming a spherical earth with a radius of 6,370,000 m. The grid definitions of the three surrogate domains are provided in Table 2.

**Table 2. Surrogate domain definitions**

| Parameter | 36km | 12km | 4km |
|---|---|---|---|
| X-orig | -2,736,000 m | -2,736,000 m | -2,736,000 m |
| Y-orig | -2,088,000 m | -2,088,000 m | -2,088,000 m |
| columns | 148 | 444 | 1332 |
| rows | 112 | 336 | 1008 |

UNC-IE used the gapfilling, normalization, and QA tools of the SA to ensure completeness and quality of the surrogates. Normalizing the surrogates eliminates surrogate values that do not equal 1.0, which can result from rounding errors in the surrogate calculations. The QA steps check for surrogates that do not equal 1.0, missing surrogates, and surrogates that are gapfilled. In addition to these tabulated QA summaries, UNC-IE produced spatial plots of the surrogates on each of the three CONUS domains. The QA tables, spatial plots, and a README file describing the surrogates are packaged in a "QA" directory in the surrogate Linux tar archives. The surrogates and accompanying QA data are available from the U.S. EPA CHIEF[8] website as part of the OAQPS 2008 and 2011 emissions modeling platforms and from the CMAS Center Data Clearinghouse[9].

The list of GIS Shapefiles used for the updated spatial surrogates and details of how these data were preprocessed for generating surrogates are provided in Section 2. An Excel Spreadsheet (US_SpatialSurrogate_Workbook_v093013.xlsx) that accompanies this document includes the detailed specifications of the Shapefiles and surrogates that were collected and updated under this Work Assignment.

# 2 Shapefile Database

## *2.1 U.S. Census-based Surrogates*

### 2.1.1 Population and Housing

The 2010 TIGER/Line database contains U.S. Census population and housing unit counts at the Census block level for each state. UNC-IE downloaded the entire database, merged the state-level data into a national Shapefile, and projected the data to a U.S. national Lambert Conformal projection on a normal sphere. UNC-IE calculated urban and rural areas using Census block groups. Urban was defined as Census block groups that have a population density of at least 1,000 people per square mile and everything else was defined as rural. In addition, Census block groups are classified based on the number of housing units per square kilometers (housing unit

[8] http://www.epa.gov/ttnchie1/

[9] http://www.cmascenter.org/download/data.cfm

density [HUD]) and the attribute AREA_CODE is assigned to census block groups with the following conditions:

1. Urban: HUD >= 1000
2. Suburban: HUD >= 125 and HUD < 1000
3. Exurban: HUD >= 6 and HUD < 125
4. Rural: HUD < 6

The attributes in the file used to compute surrogates include:

- POP2010 – 2010 population
- HU2010 – 2010 housing units
- URBAN – urban population in 2010
- RURAL – rural population in 2010
- HUCH1000 – housing unit changes from 2000 to 2010
- AREA_CODE – census block group codes; 1=urban, 2=suburban, 3=exurban, 4=rural

*Shapefile Name:* pophu_bg2010

*Shapefile Type:* Polygon

*Year*: 2010

*Attributes:* STFID, FIRST_STAT, FIRST_COUN, FIRST_TRAC, FIRST_BLOC, POP2010, HU2010, POP2000, HU2000, HUCH1000, POPCH1000, BG_AREA, URBAN10, POPD10, URBAN, RURAL, AREA_CODE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 100, 110, 120, 130, 131, 132, 133, 134, 137, 140, 165, 400

### 2.1.2 Home Heating Fuels

The American Community Survey (ACS) is a U.S. Census project that collects yearly demographic and housing information from randomly selected households throughout the U.S. The data are aggregated in 5 and 10-year increments to provide Census tract estimates for the statistics collected during the survey. Home heating fuel type from the ACS was used to develop spatial surrogates for home heating sources (i.e. residential wood combustion) in the nonpoint inventory. The ACS 5-year 2010 survey results that were released in 2011 represent data collected from 2006-2010. These data were used to create home heating surrogates.

As a review of the 5-year ACS data for 2010 showed that the number of housing units per Census tract in the ACS are always lower than the number of housing units reported in the 2010 Census. This trend indicates that the ACS only estimates demographic and housing statistics for a random sampling of households in each Census tract and does not represent the entire distribution of households. In order to represent the spatial distribution of home heating sources accurately, the ACS was used to calculate a distribution of heating sources for each Census tract and then these distributions where applied to the 2010 Census Housing Unit Shapefile to construct a home heating Shapefile that is consistent with the number of census tract housing units in the 2010 Census database. For example, if the ACS reports that in Census tract X 75%

of the households use coal for heating and 25% of the households use heating fuel, the 2010 U.S. Census estimate of the total number of households in tract X were multiplied by 0.75 and 0.25 to estimate the distribution of coal and heating fuel use, respectively, for tract X.

The attributes in the file used to compute surrogates include:

- UTIL_GAS - number of housing units using Utility Gas for primary heating
- WOOD - number of housing units using Wood for primary heating
- FUEL_OIL - number of housing units using Fuel Oil for primary heating
- COAL - number of housing units using Coal for primary heating
- LP_GAS - number of housing units using Bottled Gas for primary heating

*Shapefile Name*: heating_fuels_acs0510_c2010

*Shapefil Type:* Polygon

*Year*: 2011

*Attributes*: UTIL_GAS, WOOD, FUEL_OIL, LP_GAS

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates*: 150, 160, 165, 170, 180, 190

## *2.2 Transportation Surrogates*

### 2.2.1 Roadways

UNC-IE merged state-level 2010 TIGER/Line Shapefile data with population density data to create a national file with urban and rural roadways. UNC-IE projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere. The TIGER/Line MTFCC codes S1100 and S1200 were used to define primary and secondary roads, respectively. Urban and rural roadway classifications were calculated by UNC-IE using Census block group population densities (see Section 2.1.1) overlaid onto TIGER roads.

The attributes in the file used to compute surrogates include:

- RDTYPE = road types; 1=urban primary, 2=rural primary, 3=urban secondary, 4=rural secondary

*Shapefile Name*: rd_ps_tiger2010

*Shapefile Type:* Line

*Year*: 2010

*Attributes:* STATEFP, LINEARID, FULLNAME, RTTYP, MTFCC, STFID, URBAN10, RDTYPE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)
*Datum:* NAD83
*Spheroid:* Normal Sphere with radius 6,370,000 meters
*Derived Surrogates:* 200, 210, 220, 221, 230, 231, 240, 250

### 2.2.2 Rail Length

UNC-IE merged state-level 2010 TIGER/Line Shapefile data to create a national railway network file. The TIGER/Line MTFCC codes R1052, R1051, and R1011 were used to define the different classes of rail lines. UNC projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere.

The attributes in the file used to compute surrogates include:

- RRTYPE – rail classes; 1=class 1; 2=class 2; 3=class 3

*Shapefile Name*: rail_tiger2010

*Shapefile Type:* Line

*Year*: 2010

*Attributes:* LINEARID, FULLNAME, MTFCC, RRTYPE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)
*Datum:* NAD83
*Spheroid:* Normal Sphere with radius 6,370,000 meters
*Derived Surrogates:* 260, 270, 280

### 2.2.3 Rail Freight Density

UNC-IE used the 2012 National Transportation Atlas Database (NTAD) freight densities to weight the line Shapefile of rail activity locations. The NTAD RAILTYPE codes were used to define the different classes of rail lines. UNC-IE projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere.

The attributes in the file used to compute surrogates include:

- RAILTYPE – rail classes; 1=class 1,2,3 rail; 2=Amtrak; 3=Commuter rail

*Shapefile Name*: rail_lines_ntad2012

*Shapefile Type:* Line

*Year*: 2012

*Attributes* FRAARCID, MILES, STATEAB, STATEFIPS, CNTYFIPS, STCNTYFIPS, FRAREGIONS, RROWNER1, RROWNER2, RROWNER3, TRKRGHTS1, TRKRGHTS2, TRKRGHTS3, TRKRGHTS4, TRKRGHTS5, TRKRGHTS6, TRKRGHTS7, TRKRGHTS8, TRKRGHTS9, STRACNET, SIGSYS, TRACKS, SUBDIV, FRFRANODE, TOFRANODE, NET, PASSNGR, YARDS, INT_TYPE, DEN09CODE, VERSION, MEDDENS, RAILTYPE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83
*Spheroid:* Normal Sphere with radius 6,370,000 meters
*Derived Surrogates:* 261, 271, 272, 273

### 2.2.4 Rail Yards

Michelle Bergin and Byeong Kim of GA DNR provided a spreadsheet of rail yard locations and emissions collected as part of the Eastern Regional Technical Advisory Committee (ERTAC) railway inventory development and improvement effort (Bergin et al., 2009). UNC-IE converted this spreadsheet to a point Shapefile for generating a rail yard surrogate for the U.S. The surrogate is based on the sum of the NOx and PM2.5 emissions at the rail yard locations in the ERTAC rail yard inventory.

The attributes in the file used to compute surrogates include:

- NOXPMTOT – total of NOx and PM2.5 emissions at each rail yard

*Shapefile Name*: ERTAC_railyard_WRF

*Shapefile Type:* Point

*Year*: 2012

*Attributes* OBJECTID, OWNER1, OWNER2, OWNER3, STATE, FIPS, YARDNAME, LON, LAT, NOX, PM25, NOXPMTOT,

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)
*Datum:* NAD83
*Spheroid:* Normal Sphere with radius 6,370,000 meters
*Derived Surrogates:* 275

### 2.2.5 Bus Terminals

2014 U.S. locations for transportation terminals from the National Transportation Atlas Intermodal Passenger Connectivity Database. This Shapefile was used to define the locations of intercity and transit bus stations for allocating off-network MOVES emissions to modeling grids. This point Shapefile of transportation terminals was obtained from the online National Transportation Atlas Database.

The attributes in the file used to compute surrogates include:

- BUS_INTERC – features for intercity bus terminals
- BUS_TRANSI – features for transit bus terminals

*Shapefile Name*: ipcd

*Shapfile Type:* Point

*Year*: 2014

*Attributes:* FID, Point_ID, CITY, STATE, METRO_AREA, Lon, Lat, FERRY_TRAN, FERRY_INTE, BUS_TRANSI, BUS_INTERC, BUS_CODE, BUS_SUPPLE, RAIL_INTER, RAIL_COMMU, RAIL_HEAVY, RAIL_LIGHT, AIR_SERVICE, INTERCITY_, TRASIT_SE, CBSA_CODE, CBSA_TYPE, MODES_SRV, MODE_BUS, MODE_AIR, MODE_RAIL, MODE_FERRY

*Projection:* Geographic

*Datum:*

*Spheroid:* WGS84

*Derived Surrogates:* 258, 259

### 2.2.6 Extended Idle Locations

Locations of extended idle emissions for tractor-trailers. The base parking area shapefile was provided by EPA and it included 7 different classes of parking locations:

1. State Department of Transportation (DOT) visitor centers
2. DOT welcome centers
3. DOT rest areas
4. DOT weigh stations
5. DOT parking areas
6. Private truck stops
7. Private retail locations, including Walmart and McDonalds

Along with the latitude-longitude coordinates of the parking locations, the database includes attributes indicating the number of parking spots at each location to use for weighting a spatial surrogate. The attributes for the number of parking spots for the different types of locations were pulled from multiple data sources. The EPA database included the number of parking spots for many of the private truck stops but did not include the number of spots for most of the DOT or private retail locations. The following steps were followed to mine parking spot data to develop a complete database of extended idle locations.

State DOT Locations

Nationwide request from all state DOTs for the number of parking spots at state-run locations returned data from the WV, IN, MO, WI, NC, CT, ME, UT, VA, GA, and FL. Many of these states reviewed the default data pulled from the web and updated these data with new information. We reconciled the data provided by the states with the national data provided by EPA. Manual matching was done using Bing satellite imagery for locations in which the parking area exit ramp coordinates were provided rather than the actual parking lot location. The matching yielded a fairly complete database of coordinates and the number (or range) of truck parking spots at DOT locations.

Private Truck Stops

Data downloaded from Pilot, Petro, and TA truck stop company websites include the coordinates and number of truck parking spots at all locations nationally. Merged these data with the EPA database using a 300 meter buffer to exclude duplicates. These data appeared more reliable than the number of parking spot attributes in the EPA database based on manual comparisons with Bing satellite imagery.

Private Retail Locations

The EPA database included coordinates and number of parking spots for some retail and food service areas. Walmarts composed the largest number of private retail locations in the database and did not include any information on the number of truck parking spots at any of the locations.

Gapfilling the Number of Parking Spots

After creating an inventory of the truck parking spot locations from the collected data and merging these data into the EPA database, we used the following rules to fill in the number of truck parking spots:

- If number of spots listed as "<20", set to 20
- If number of spots listed as "20-69", set to 45
- If number of spots listed as ">70", set to 70
- If retail locations listed the number of spots as "unknown", set to 2
- Weigh stations and parking areas all listed the number of spots as "unknown", set to 2
- For rest areas with missing truck parking spots, calculated 1st quartile, median, and 3rd quartile from all known DOT rest area data and used these as low, medium, and high attributes for the number of spots; results: low = 14, median = 18, high = 31
- For truck stops with missing truck parking spots, calculated 1st quartile, median, and 3rd quartile from all known truck stop data and used these as low, medium, and high attributes for the number of spots; results: low = 80, median = 123, high = 188

Develop Shapefile

Created a shapefile from the modified/gapfilled EPA database by converting the point locations to an NAD83 datum and Lamber Conformal Conic projection. Included the attributes truckpark_low, truckpark_med, and truckpark_high to use for weighting spatial surrogates. If the number of truck parking spots at a location was known in the database and there was no need to gapfill, these low/med/high attributes were set to all equal the number of spots at the location.

The attributes in the file used to compute surrogates include:

- trucks_med – median number of truck parking spaces at potential idling locations

*Shapefile Name*: potential_idling_locations_v2_final

*Shapfile Type:* Point

*Year*: 2014

*Attributes:* truck_park, name, address, city, state, zip, latitude, longitude, site_cat, source, directions, alt_source, trucks_low, trucks_med, trucks_high

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 205

## *2.3 Shipping and Port Surrogates*

### 2.3.1 Waterways

2011 National Transportation Atlas Database (NTAD) data for waterway lengths. UNC projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere.

The attributes in the file used to compute surrogates include:

- LENGTH – navigable waterway lengths

*Shapefile Name*: waterway_ntad2011

*Shapefile Type:* Line

*Year*: 2011

*Attributes:* LENGTH, FEATUREID, ANODE, BNODE, LINKNAME, RIVERNAME, AMILE, BMILE, LENGTH1, LENTH_SRC, LINKTYPE, CTRL_DEPTH, WATERWAY, GEO_CLASS, FUNC_CLASS, WTWY_TYPE, CHART_ID, NUM_PAIRS, WHO_MOD, DATE_MOD, HEADING, STATE, FIPS, FIPS2, NONUS, VERSION

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 807

### 2.3.2 Marine Ports Point Locations

2010 NTAD data for marine ports. UNC projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere.

The attributes in the file used to compute surrogates include:

- NONE – default to point locations

*Shapefile Name*: ports_ntad2010

*Shapefile Type:* point

*Year*: 2010

*Attributes:* NAME, LOCATION, ADDRESS, COUNTY, COUNTYFIPS, TOWN, STATE, WATERWAY, PORT_NAME, MILE, BANK, LATDEC, LONG_DEC, OPERATOR1, OWNER, PURPOSE, RWY_CONN, PORTSERIES, SEQNO, LOC_CD, PWDNO, OLDPWD, DOCKCD, NDCCODE, CNGDST, COMMODITY1, COMMODITY2, COMMODITY3, COMMODITY4, REMARKS, DATUM, DEPTH1, DEPTH1A, DEPTH2, DEPTH2A, DEPTH3, DEPTH3A, TOTBERTH1, TOTBERTH2, TOTBERTH3, YEAR, MAPNO, FIRSTNAME, LASTNAME, PHONE, FAX, STFIPS, VERSION

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 800

### 2.3.3 Marine Port Area Emissions

2012 marine port areas weighted by NEI2011 NOx emissions. UNC-IE joined a Shapefile of marine port areas with NEI2011 NOx emissions for SCC 2280002100. UNC projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere.

The attributes in the file used to compute surrogates include:

- ann_value – annual NOx emissions for port vessels from the NEI2011

*Shapefile Name*: Ports_081412_NEI2011_NOx_WRF

*Shapefile Type:* polygon

*Year*: 2012

*Attributes:* NAME,FIPS_1,Area_sqmi,ShapeID,PortID,OID_,country_cd,region_cd,tribal_cod, census_tra,shape_id,scc,emis_type,poll,ann_value,ann_pct_re,control_id,control_me,current_co, cumulative,projection,reg_codes,calc_metho,calc_year,date_updat,data_set_i,jan_value,feb_valu e,mar_value,apr_value,may_value,jun_value,jul_value,aug_value,sep_value,oct_value,nov_valu e,dec_value,jan_pctred,feb_pctred,mar_pctred,apr_pctred,may_pctred,jun_pctred,jul_pctred,aug _pctred,sep_pctred,oct_pctred,nov_pctred,dec_pctred,comment,SHPID

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 820

### 2.3.4 Great Lakes Tugboat Zones


**Figure 2-1. 2011 NEI Great Lakes tugboat zones**

U.S. EPA defined a tugboat operating zone in the Great Lakes as 0.5-2.0 miles offshore (Figure 2). A Shapefile that defines the tugboat zone by county was developed for use in creating a spatial surrogate for emissions in the Great Lakes.

The attributes in the file used to compute surrogates include:

- AREA – area of tugboat zone by county

*Shapefile Name*: TUGvessel_SURROGATE_GREATLAKES

*Shapefile Type:* Polygon

*Year*: 2012

*Attributes*: FIPS, STATE, COUNTY, NAME, ST, STATECTY, AREA,

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 808

### 2.3.5 Gulf of Mexico Platforms

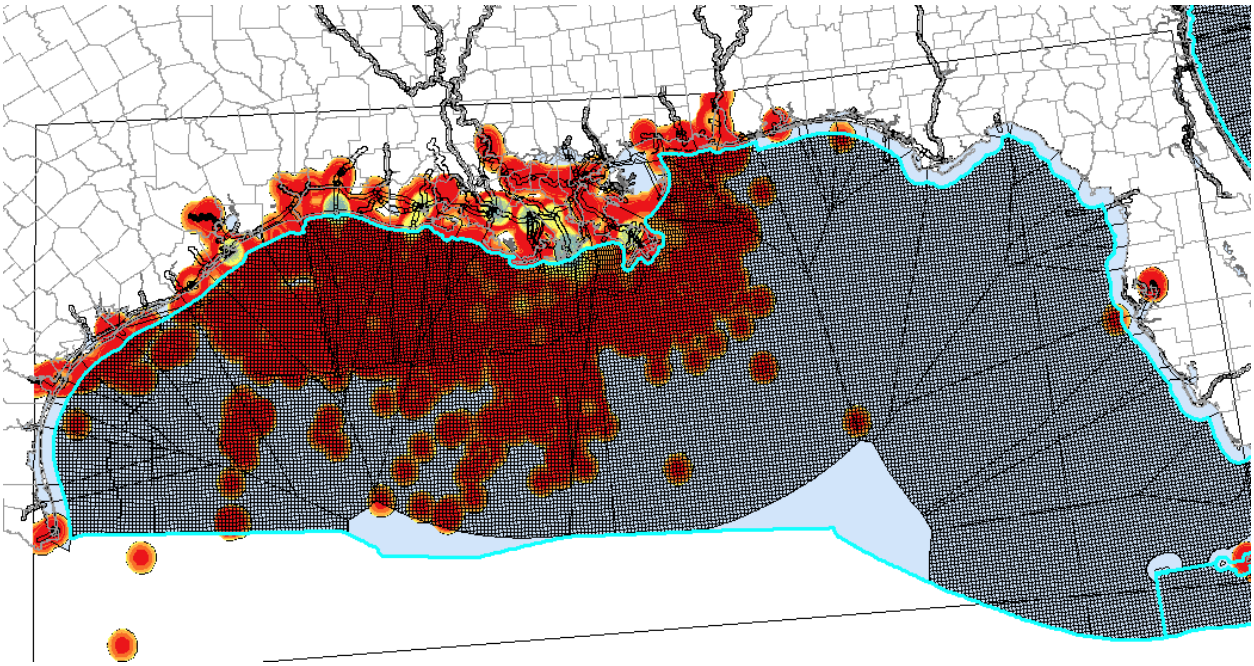

**Figure 2-2. Gulf of Mexico oil and gas platform support vessel densities**

The locations of class 1 and 2 commercial shipping vessels in the Gulf of Mexico were tied to the density of shipping activity surrounding oil and gas platforms (Figure 3). A Shapefile that defines the fraction of total Gulf activity density around the energy development platforms was used to create a Gulf Shipping Lanes surrogate.

The attributes in the file used to compute surrogates include:

- Fraction – fraction of Gulf-wide activity associated with individual oil and gas platforms

*Shapefile Name*: GulfofMexico_SupportVessels_Density_WRF

*Shapefile Type:* Polygon

*Year*: 2010

*Attributes*: GRIDCODE, Area, Fraction

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 803, 802

### 2.3.6 Offshore Commercial Shipping

Offshore shipping lane areas for the Pacific, Atlantic, Gulf of Mexico, and Great Lakes weighted by NEI2011 NOx emissions. UNC-IE joined a Shapefile of offshore commercial shipping lane

areas with NEI2011 NOx emissions for SCC 2280002200. UNC projected the data to a U.S. national Lambert Conformal projection on a normal sphere

The attributes in the file used to compute surrogates include:

- ann_value– annual NOx emissions for C1/C2 vessels from the NEI2011

*Shapefile Name*: ShippingLanes_112812_NEI2011_NOx_WRF

*Shapefile Type:* Polygon

*Year*: 2012

*Attributes*: FIPS,Area_sqmi,ShapeID,OID_,country_cd,region_cd,tribal_cod,census_tra, shape_id, scc,emis_type,poll,ann_value,ann_pct_re,control_id,control_me,current_co, cumulative, projection,reg_codes,calc_metho,calc_year,date_updat,data_set_i,jan_value, feb_value,mar_value,apr_value,may_value,jun_value,jul_value,aug_value,sep_value, oct_value,nov_value,dec_value,jan_pctred,feb_pctred,mar_pctred,apr_pctred,may_pctred, jun_pctred,jul_pctred,aug_pctred,sep_pctred,oct_pctred,nov_pctred,dec_pctred,comment,SHPID

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 805, 806, 802

### 2.3.7 Midwest Shipping and Inland Waterways

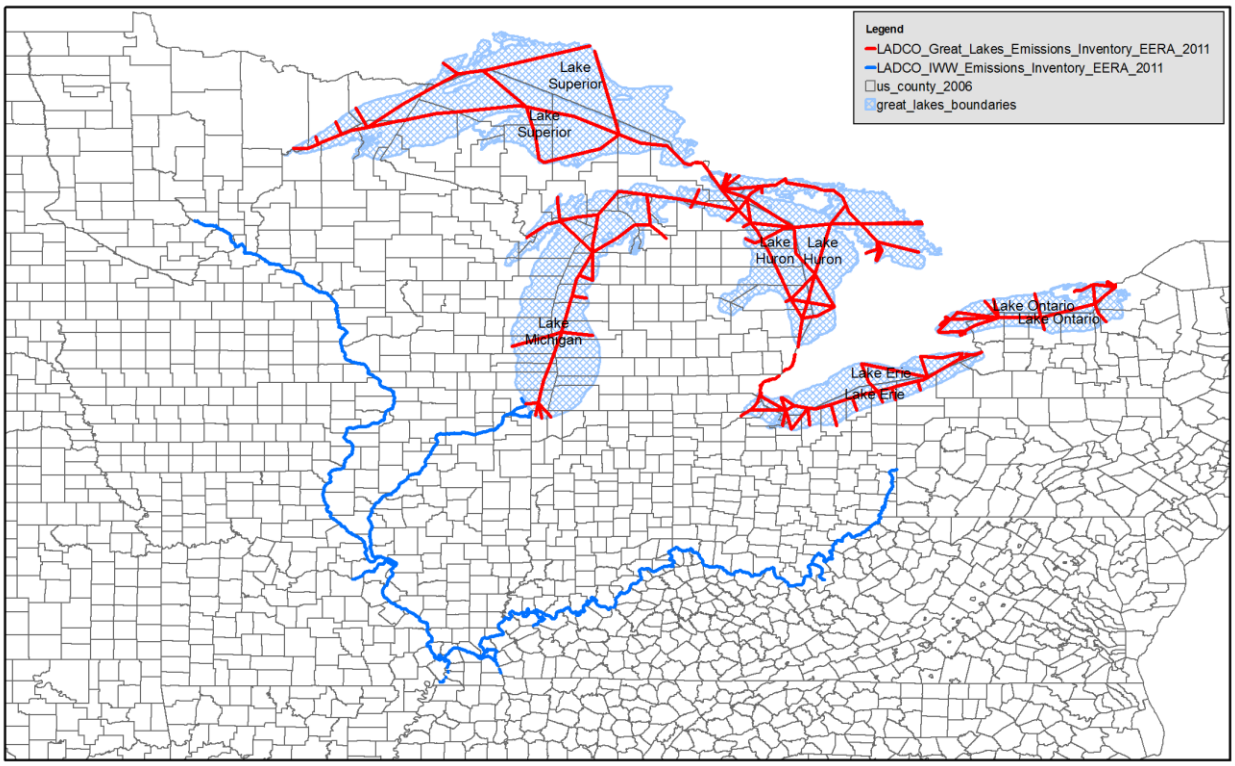


**Figure 2-3. Great Lakes shipping lane and inland waterway inventory**

LADCO provided a 2011 shipping lane and inland waterway emission inventory for the Great Lakes region (Figure 4). UNC projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere and used these data to generate a surrogate for the regional shipping lane emissions.

The attributes in the file used to compute surrogates include:

- G_nonCO2– annual non-CO2 emissions for C1/C2 vessels from the LADCO 2011 inventory

*Shapefile Name*: GreatLakes_IWW_EERA_2011_WRF

*Shapefile Type:* Line

*Year*: 2011

*Attributes*: NWN_ID,G_CO2GAL,G_NOXGAL,G_SOXGAL,G_HCGAL,G_PMGAL, G_BCGAL, G_POMGAL,G_COGAL,seg_km,G_nonCO2,GROUP,Seg_Miles,G_CO2, G_NOX,G_SOX,G_HC,G_PM,G_BC,G_POM,G_CO,

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates*: 812, 802

## *2.4 FEMA Building Footprints*

HAZUS is a nationally standardized methodology that contains models for estimating potential losses from natural disasters. A key component of the HAZUS database is information on building locations, sizes, and classifications. These geospatial data on buildings are useful for allocating inventory data to modeling grids because they contain details on the locations and sizes of different building types. The Federal Emergency Management Agency HAZUS-MH version 2 was released in September 2011 and contains square footage data for different types of buildings throughout the U.S. The building square footage data are used to identify building classifications (i.e. commercial, residential, industrial, institutional) for allocating non-point inventory sources to modeling grids. UNC merged Census-block level data in state Shapefiles into a national Shapefile for select database attributes. UNC used 2002 Census block boundaries when merging these data to maintain consistency with the previous spatial surrogates used by EPA.

The attributes in the file used to compute surrogates include:

- COM# - commercial building square footage
- IND# - industrial building square footage
- RES# - residential building square footage
- EDU# - educational building square footage

- REL# - religious building square footage
- GOV# - government building square footage

*Shapefile Name*: fema_bsf_2002bnd

*Shapefile Type:* Polygon

*Year*: 2011

*Attributes*: COM#, IND#, RES#, EDU#, REL#, GOV#

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 256, 257, 500, 505, 506, 507, 510, 512, 515, 520, 525, 526, 527, 530, 535, 540, 545, 550, 555, 560, 565, 570, 575, 580, 585, 590, 595, 596

## *2.5 Landcover/Landuse Surrogates*

### 2.5.1 Agriculture

2006 National Land Cover Database (NLCD) land cover data was used to create this Shapefile. UNC-IE re-gridded the original 30-meter resolution data to 250-meter resolution using the dominant land use class. UNC-IE extracted agricultural classes 81 and 82 into a 250-meter raster file and then converted the raster file to a polygon Shapefile.

The attributes in the file used to compute surrogates include:

- GRID_CODE – 81 for Pasture or Hay class and 82 for Crop class

*Shapefile Name*: nlcd2006_80s_agri

*Shapefile Type:* Polygon

*Year*: 2006

*Attributes*: GRID_CODE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 310, 318, 319

### 2.5.2 Water and Land

2006 NLCD land cover data was used to create this Shapefile. UNC-IE re-gridded the original 30-meter resolution data to 250-meter resolution using the dominant land use class. UNC-IE extracted water (class 11) and land (class 0) into a 250-meter raster file and then converted the raster file to a polygon Shapefile.

The attributes in the file used to compute surrogates include:

- GRID_CODE – 11=water; 0=land

*Shapefile Name*: nlcd2006_10s_waterland

*Shapefile Type:* Polygon

*Year*: 2006

*Attributes*: GRID_CODE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 340, 346, 350

### 2.5.3 Developed Areas

2006 NLCD land cover data was used to create this Shapefile. UNC-IE re-gridded the original 30-meter resolution data to 250-meter resolution using the dominant land use class. UNC-IE extracted developed land use classes (21 to 24) into a 250-meter raster file and then converted the raster file to a polygon Shapefile.

The attributes in the file used to compute surrogates include:

- GRID_CODE – 21=Open Space; 22=Developed, Low Intensity; 23=Developed, Medium Intensity; 24=Developed, High Intensity

*Shapefile Name*: nlcd2006_20s_developed

*Type:* Polygon

*Year*: 2006

*Attributes*: GRID_CODE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 300, 301, 302, 303

### 2.5.4 Forest Land

2006 NLCD land cover data was used to create this Shapefile. UNC-IE re-gridded the original 30-meter resolution data to 250-meter resolution using the dominant land use class. UNC-IE extracted the forest land use classes (41,42,43, and 90) into a 250-meter raster file and then converted the raster file to a polygon Shapefile.

The attributes in the file used to compute surrogates include:

- GRID_CODE – 40=Forest land

*Shapefile Name*: nlcd2006_40_forest

*Type:* Polygon

*Year*: 2006

*Attributes*: GRID_CODE

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 320

## 2.6 Oil and Gas Production

EPA collected data from multiple data sources to develop spatial surrogates for oil and gas production locations throughout the Continental U.S. Ramboll Environ and ERG worked separately on collecting and compiling active well locations for use in developing spatial surrogates. As some of the well data were harvested from proprietary databases, Ramboll Environ and ERG employed de-identification techniques to mask the information for individual well locations. These techniques primarily involved converting the point well data into polygons with multiple wells in each polygon. A description of the oil and gas production location dataset produced by ERG, which is used in the EPA 2011v2 modeling platform, is provided here. Refer to the previous version of this document for a description of the data collected by Ramboll Environ (Adelman, 2014).

### 2.6.1 Eastern Research Group, Inc. Oil and Gas Production Data (HPDI Database)

The primary activity data source that ERG used for developing oil and gas spatial surrogates were data from Drilling Info (DI) Desktop's HPDI database (ERG, 2014). This database contains well-level location, production, and exploration statistics at the monthly level. Due to a proprietary agreement with DI Desktop, individual well locations and ancillary production data cannot be made publicly available, but aggregated statistics are allowed. The HPDI data represented approximately 80% of the activity data used for these surrogates.

For the remaining 20% of the activity, ERG supplemented the HPDI with additional data from Oil and Gas Commission (OGC) websites. In many cases, the correct surrogate parameter was not available (e.g., feet drilled), but an alternative surrogate parameter was available (e.g., number of spudded wells) and used for generating the spatial surrogates.

In total, ERG compiled over 1.08 million unique well locations from the above data sources. The well locations cover 33 states and 1,193 counties. Each well was uploaded into ArcGIS, and assigned to the associated 2010 census tract identifier and 4-km Continental U.S. grid identifier. Well locations are presented in Figure 2-7.

The oil and gas spatial surrogates distributed with the 2011v2 modeling platform were generated from the ERG database of actual well and production locations. Since these data are proprietary and cannot be redistributed, a set of de-identified Shapefiles that include the number of wells or production units in 4-km grid cells (polygons) will be made available to the public. The surrogates developed from these public Shapefiles will use the number of production units in

each 4-km grid cell as the weighting attribute. Although it will not be possible to exactly reproduce the oil and gas production surrogates in the EPA modeling platform without the proprietary Shapefile data, the de-identified production Shapefiles will provide a close approximation of these data to use for the development of surrogates on customized modeling domains.

For additional details on these Shapefiles and spatial surrogates see ERG (2014).

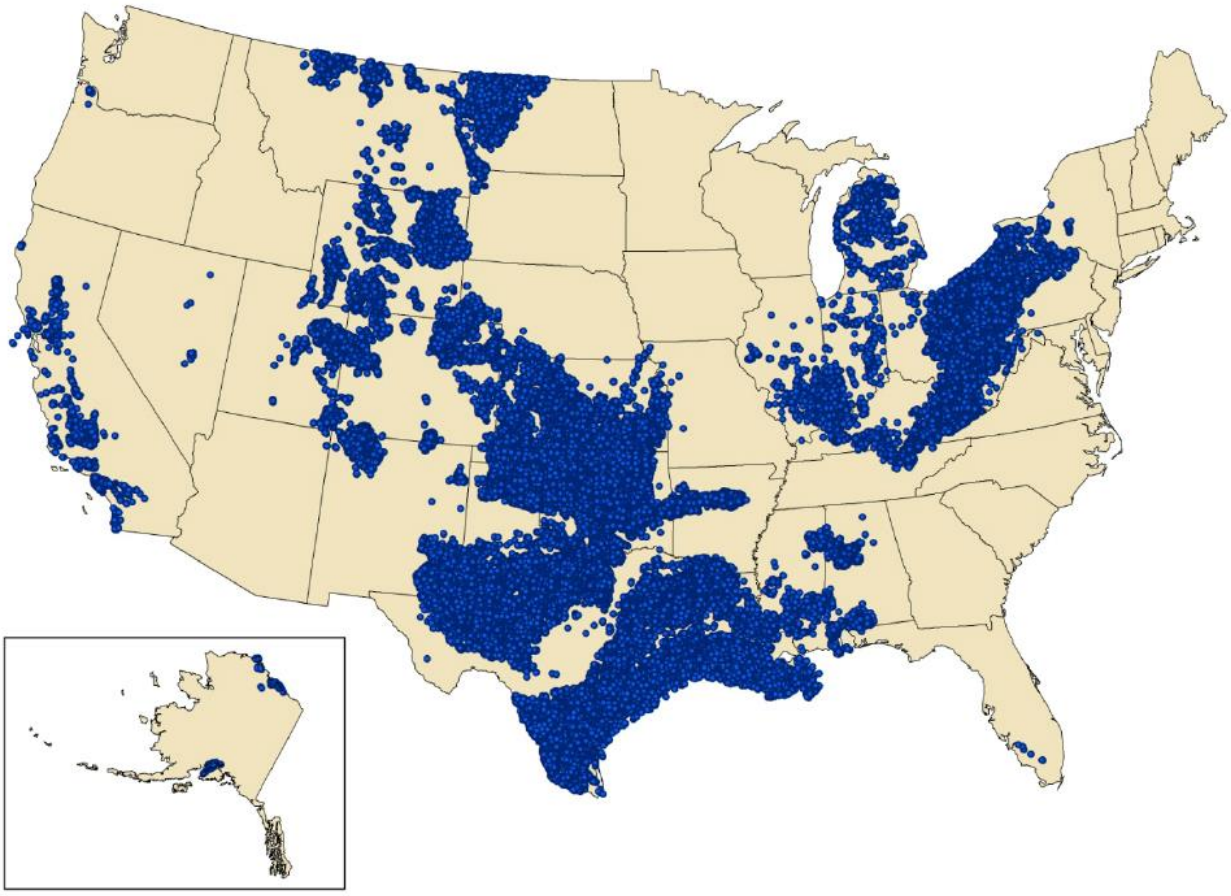

**Figure 2-4. 2012 active oil and gas well locations developed by ERG, Inc.**

The attributes in the public oil and gas Shapefile:

- TBD

*Shapefile Name*: TBD

*Type:* Polygon

*Year*: 2012

*Attributes*: TBD

*Projection:* TBD

*Datum:* TBD

*Spheroid:* TBD

*Derived Surrogates:* 681, 682, 683, 684, 685, 686, 687, 688, 689, 692, 693, 694, 695, 697, 698

## *2.7 Other Industrial and Commercial Activities*

### 2.7.1 Mines

2003 USGS point locations of mineral operations and mines for the following: agricultural uses, construction uses, ferrous metal mines, ferrous metals processing plants, miscellaneous industrial uses, nonferrous metal mines, nonferrous metal processing plants, refractory, abrasive, and other industrial uses, sand and gravel, and crushed stone operations; NOTE: this Shapefile does not include energy minerals (i.e., coal is not included here).

The attributes in the file used to compute surrogates include:

- NONE – point locations of the mining operations are used to derive the surrogate

*Shapefile Name*: mines_usgs2003_wrf

*Shapefile Type:* Point

*Year*: 2003

*Attributes*: SNDGRVX020,COMP_NAME,OPER_NAME,COUNTY,STATE,FIPS, STATE_FIPS,LATITUDE,LONGITUDE,OPER_TYPE,COMMODITY

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 860, 861

### 2.7.2 Quarries

2003 USGS point locations of mineral operations and mines for the following: sand and gravel, and crushed stone operations

The attributes in the file used to compute surrogates include:

NONE – point locations of the mining operations are used to derive the surrogate

*Shapefile Name*: quarries_usgs2003_wrf

*Shapefile Type:* Point

*Year*: 2003

*Attributes*: SNDGRVX020,COMP_NAME,OPER_NAME,COUNTY,STATE,FIPS, STATE_FIPS,LATITUDE,LONGITUDE,OPER_TYPE,COMMODITY

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 862

### 2.7.3 Dry Cleaners

Counts of dry cleaners by Census Block Group in operation during calendar year 2009 were used to update the dry cleaner location surrogate. These data were purchased from the Spatial Insights, Inc. Business Counts Database[10]. As these data are proprietary, this Shapefile cannot be redistributed.

The attributes in the file used to compute surrogates include:

- DRY – number of dry cleaner operations in each Census Block Group

*Shapefile Name*: drycleaner_surrogate

*Shapefile Type:* Polygon

*Year*: 2009

*Attributes*: STFID, FIRST_STAT, FIRST_COUN, FIRST_TRAC, FIRST_BLOC, GEOID, DRY, FIPSSTCO

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 880

### 2.7.4 Gas Stations

Counts of gas stations (with and without convenience stores) by Census Block Group in operation during calendar year 2009 were used to update the gas station location surrogate. These data were purchased from the Spatial Insights, Inc. Business Counts Database[9]. As these data are proprietary, this Shapefile cannot be redistributed.

The attributes in the file used to compute surrogates include:

- GAS – number of gas station operations in each Census Block Group

*Shapefile Name*: gas_station_surrogate

*Shapefile Type:* Polygon

*Year*: 2009

*Attributes*: STFID, FIRST_STAT, FIRST_COUN, FIRST_TRAC, FIRST_BLOC, GEOID, GAS, FIPSSTCO

*Projection:* Lambert Conformal Conic ($X_0$=-97.0, $Y_1$=33.0, $Y_2$=45.0, $Y_0$=40.0)

*Datum:* NAD83

[10] http://www.spatialinsights.com/catalog/product.aspx?product=126&content=1606

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 600

# 3 Known Issues and Future Work

While the Shapefiles and surrogates documented here and in the accompanying spreadsheet represent the best available information for allocating non-point emissions to modeling grids, there are some caveats to these data and issues that should be addressed in the future.

- Some of the surrogates are derived from proprietary Shapefiles that cannot be distributed. These data include:
    - The dry cleaners and gas stations surrogates are generated from data purchased from Spatial Insights, Inc.
    - The oil and gas surrogates (681-698) are generated from data purchased largely from HPDI, Inc. These data were aggregated from points to polygons (4-km grid cells) to obscure the proprietary nature of the data. As described in Section 2.6, this aggregation process results in the loss of some of the original data.
- The total agriculture surrogate (310) is not specific enough for livestock sources and could be improved using a point Shapefile of confined animal feeding operations (CAFOs). Although we were not able to find a national Shapefile of CAFO locations, these data are available from some states. State departments of water quality track CAFO locations for monitoring runoff and water permitting purposes. It will be worthwhile to collect data from as many states as possible for developing a national CAFO surrogate to use for allocating non-point livestock inventories to model grids.
- The high resolution, national NCLD Shapefile breaks the Surrogate Tool by causing memory allocation failures while processing the data (SRGCREATE_ERROR>gpc malloc failure). To avoid this error it is necessary to split the NLCD Shapefiles into subregions, generate surrogates for each subregion, and then paste the data back together to create a national surrogate. We needed to split the NLCD Shapefiles into eight subregions to avoid the memory violations in the Surrogate Tool. The Surrogate Tool needs to be updated to take advantage of the larger memory and multi-processing capabilities of modern computing hardware.

# 4 Surrogate Revision History

- Version 5 (09/2014) Updates
    - The updates in this revision created surrogates for the off-network MOVES processes.
    - Extended Idle Locations (205): Locations of off-network extended idling for the MOVES2014 hoteling emissions mode; built from a combination of private truck stops, retail locations, and DOT rest areas/rest stops/welcome centers. The number of truck parking spots at each location is used as the weight factor for this surrogate

  - Urban Unrestricted (221) and Rural Unrestricted (231) Roads: Locations of on-network local and secondary roads for the MOVES2014 unrestricted RPD emissions mode; merged surrogate that combines urban/rural population as a proxy for local roads with secondary road miles. The merge coefficients for the split between local and secondary roads were derived from national total VMT splits from the NEI2011v1. The secondary/local split for urban unrestricted roads is 0.77/0.23 and for rural unrestricted roads 0.81/0.19.
  - Off-Network Short-Haul (256) Trucks: Start and stop locations for short-haul trucks for the MOVES2014 RPV emissions mode. Surrogate derived from the FEMA building square footage for industrial and commercial buildings, with the exception of theaters and parking garages
  - Off-Network Long-Haul (257) Trucks: Start and stop locations for long-haul trucks for the MOVES2014 RPV emissions mode. Surrogate derived from the FEMA building square footage for industrial and wholesale trade locations
  - Intercity Bus Terminals (258): Start and stop locations for intercity buses for the MOVES2014 RPV emissions mode. Surrogate weights are from the National Transportation Atlas 2014 intercity bus terminal point locations from the Intermodal Passenger Connectivity Database.
  - Transit Bus Terminals (259): Start and stop locations for transit buses for the MOVES2014 RPV emissions mode. Surrogate weights are from the National Transportation Atlas 2014 intercity bus terminal point locations from the Intermodal Passenger Connectivity Database.
  - Education Buildings (506): FEMA building square footage for grade schools and universities. Surrogate designed to represent school bus start and stop locations for the MOVES204 RPV emissions mode.
  - Heavy-Light-Construction Industrial Buildings (507): FEMA building square footage for non-technology industrial areas. Surrogate designed to represent refuse truck start and stop locations for the MOVES 2014 RPV emission mode.
  - Residential – Non-Institutional (526): FEMA building square footage for non-institutional housing (single family homes, mobile homes, and apartment buildings). Surrogate designed to represent motor home start and stop locations for the MOVES 2014 RPV emissions mode.
  - Oil and Gas Production (681-698): Locations of different types of oil and gas production activities.

- Version 4 (09/2013) Updates
  - Shipping lanes surrogate (802): Built as a combination of Great Lakes shipping lanes/Inland waterways, Gulf of Mexico platform densities (803), and offshore shipping locations weighted by NEI2011 NOX emissions from SCC 2280002200 (806), and water (350).
  - Ports weighted by NEI2011 NOx emissions (820): Port areas joined with NEI2011 NOX emissions from SCC 2280002100

  - Oil and gas development surrogates for the following basins: Greater Green River (WY), Uintah (UT), Piceance (CO), Denver-Julesberg (CO), San Juan (NM), Raton (NM), Paradox (CO/NM), Permian (TX/NM), Marcellus (PA/NY/OH/WV/KY), Illinois (IL/IN); surrogates include Gas production at all wells (689), Oil production at all wells (690), Well count of CBM wells (691), Spud count (692), Well count of all wells (693), Oil production at oil wells (694), Well count of all oil wells (694), Gas production at gas wells (696), Oil production at gas wells (697), Well count of gas wells (698), Gas production at CBM wells (699).
  - Gas stations (600) - 2012 gas station point locations weighted by the number of locations in each Census Block Group
  - Dry cleaners (880) - 2012 dry cleaner point locations weighted by the number of locations in each Census Block Group
  - Great Lakes Tug Zone Area (808) - Tug zone from 0.5 - 2.0 miles offshore in the Great Lakes
- Version 3 (10/2012) Updates
  - All NLCD-based (3**) surrogates updated with NLCD 2006 30-m raster files aggregated to 250-m Shapefiles.
    - Includes new med. (301) and high intensity residential (302) surrogates
    - Includes new open space surrogate (303)
    - Total agriculture with orchards/vineyards and orchards/vineyards no longer available
    - Includes new pasture land (318) and crop land (319) surrogates
  - Four new housing surrogates (131-134) from the 2010 US Census data.
  - Three new rail density surrogates (271-273) from 2012 NTAD freight density line Shapefiles
  - New rail yard surrogate (275) from ERTAC rail yard point Shapefile (provided by Georgia DNR)
  - Construction and mining surrogate (861) added as a combination of 0.5 housing change/population + 0.5 mines
  - Quarries surrogate (862) from point locations of mineral operations and mines for sand, gravel, and crushed stone
- Version 2 (06/2012) Updates
  - All FEMA-based (5**) surrogates re-generated with the correct data Shapefile (cty_pophu2k_revised). One artifact of using the incorrect data Shapefile was the inclusion of defunct counties/FIPs or the exclusion of new counties/FIPs in the new surrogates. Dade County, FL and Yellowstone National Park, MT were removed in the new data. Broomfield County, CO, Miami-Dade County, FL, and Boyd County, NE now exist.

- All surrogates use new gapfilling definitions, including being gapfilled with surrogates that used the correct data Shapefile; some of the previous surrogates were gapfilled using FEMA data that used an incorrect data Shapefile
- Two WY counties had surrogates that didn't normalize to 1.0. The Normalization Tool was used to force these two counties to have surrogates that add to 1.0.

# 5 REFERENCES

Adelman, Z. (2014) Emissions Modeling Platform Spatial Surrogate Documentation, Prepared for Rich Mason, US EPA OAQPS, UNC-IE, September 30, 2014.

Bergin, M. et al (2009) ERTAC Rail: A collaborative effort in building a railroad-related emission inventory between Eastern states' air protection agencies and participation with the railroad industry, presented at the 18th Annual Emission Inventory Conference, U.S. EPA, Baltimore, MD, http://www.epa.gov/ttnchie1/conference/ei18/session6/bergin.pdf.

ERG (2014) Technical Memorandum: Modeling Allocation Factors for 2011 Nonpoint NEI, To: Alison Eyth and Dennis Doll, US EPA OAQPS, From: Regi Oomen et al., ERG, February 18, 2014.

Eyth, A. et al. (2007) Accounting for Land Use Changes in Projecting Future-Year Emissions Scenarios, presented at the 16th Annual Emission Inventory Conference, U.S. EPA, Raleigh, NC, http://www.epa.gov/ttnchie1/conference/ei16/session7/eyth.pdf.