

# **CRC Report No. A-119**

## **HIGH-RESOLUTION INVENTORY DATA EXTRACTOR AND SOURCE APPORTIONMENTS REGROUPING TOOL DEVELOPMENTS: CRC PROJECT A-119**

**Final Report**

**March 2020**



**COORDINATING RESEARCH COUNCIL, INC.  
5755 NORTH POINT PARKWAY • SUITE 265 • ALPHARETTA, GA 30022**

**The Coordinating Research Council, Inc. (CRC) is a non-profit corporation supported by the petroleum and automotive equipment industries. CRC operates through the committees made up of technical experts from industry and government who voluntarily participate. The four main areas of research within CRC are: air pollution (atmospheric and engineering studies); aviation fuels, lubricants, and equipment performance; heavy-duty vehicle fuels, lubricants, and equipment performance (e.g., diesel trucks); and light-duty vehicle fuels, lubricants, and equipment performance (e.g., passenger cars). CRC's function is to provide the mechanism for joint research conducted by the two industries that will help in determining the optimum combination of petroleum products and automotive equipment. CRC's work is limited to research that is mutually beneficial to the two industries involved. The final results of the research conducted by, or under the auspices of, CRC are available to the public.**

**CRC makes no warranty expressed or implied on the application of information contained in this report. In formulating and approving reports, the appropriate committee of the Coordinating Research Council, Inc. has not investigated or considered patents which may apply to the subject matter. Prospective users of the report are responsible for protecting themselves against liability for infringement of patents.**



**INSTITUTE FOR  
THE ENVIRONMENT**

# **High-resolution inventory data extractor and source apportionments regrouping tool developments: CRC Project A-119**

## **Final Report**

Prepared for:

**The Coordinating Research Council, Inc.**

5755 N. Point Pkwy Suite 265

Alpharetta, GA 30022

Prepared by:

**Bok H. Baek, Ph.D.**

Institute for the Environment

The University of North Carolina at Chapel Hill



Updated on:

**March 12, 2020**

# Contents

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| 1. Executive Summary.....                                                  | 3  |
| 2. Task 1: Source Apportionment Regrouping Tool: “SA_Regroup”.....         | 4  |
| 2.1 SA_Regroup Tool Development.....                                       | 4  |
| 2.2 What is in the package? .....                                          | 6  |
| SA_Regroup Run Script: “sa_regroup.csh” .....                              | 6  |
| 2.3 Run Scripts.....                                                       | 7  |
| Use Case Location .....                                                    | 7  |
| 2.4 SA_Regroup Source Code and Makefile.....                               | 8  |
| 3. Task 2: High-resolution Windowed Emissions Extractor: “SubDomain” ..... | 9  |
| 3.1 Sub-Domain Inventory Extractor Development .....                       | 9  |
| 3.2 What is in the package? .....                                          | 10 |
| “Subdomain_extraction.py”: Domain GIS Shapefile Generator .....            | 11 |
| SMOKE Inventory Extract Run Scripts.....                                   | 12 |
| 3.3 Python Wrapper Run Script: “run_subdomain_inv_extractor.py” .....      | 14 |
| Use Case Location .....                                                    | 15 |
| 4. How to Integrate into an existing NEI Emissions Modeling Platform.....  | 16 |
| 5. Run Date Setting: run_settings.txt.....                                 | 16 |
| 6. SMOKE Report Configuration Template.....                                | 16 |

## 1. Executive Summary

Under the Coordinating Research Council (CRC) project no. A-119, UNC has prepared the modeling tools that can improve modeling techniques and data availability for source characterization for states, Multi-Jurisdictional Organizations (MJO), and other air quality/emissions modeling research groups.

Chemical Transport Models (CTM) and related ambient measurements are essential to understand the relationships of different emission sources for the public health impact studies. Because of rapid changes in emissions between the years, it is crucial to perform the detail assessment studies to incorporate these spatio-temporal sensitive changes into the CTM modeling applications.

In this project, we developed two modeling supporting tools that can assist modelers and non-modelers to easily generate the source-sector-specific detail emission inventory reports from the targeted modeling region, optionally defined by the users. We first developed the tool to help users to understand the shift in source contributions to local air quality issues. It will allow users to quickly regroup existing source apportionment ready emission inputs across all sectors for the newest CTM source apportionment modeling. The following tool will allow both experienced and non-modelers to easily extract critical high-resolution inventory data from the official National Emissions Inventory (NEI) from U.S. EPA. It can be compared to measurement data, including both terrestrial and remote sensing technologies.

These two modeling tools listed below are packaged with the complete stand-alone use-cases for users to successfully install the tools on their Emissions Modeling Platforms (EMP). It comes with the complete inputs, run script and tools:

1. Source Apportionment Regrouping Tool: “SA\_Regroup”
2. Sub-domain Emission Inventory Extractor: “SubDomain”

This technical final report can be served as User’s Guide Documents for these two tools listed above. It describes the details of all input and output files along with various run scripts and tool source codes for users. Chapter 2 describes how to use the “SA\_Regroup” tool to develop the regrouped Source Apportionment (SA) emissions inputs for CTM. Chapter 3 is for the “SubDomain” Inventory extractor tool that allows users to easily generate the detail hourly gridded emissions from the user-defined sub-modeling domain area. The final chapter 4 will show you how to integrate these tools package into an existing NEI EMP distributed by OAQPS U.S. EPA.

## 2. Task 1: Source Apportionment Regrouping Tool: “SA\_Regroup”

The current EPA provides states and Multi-Jurisdictional Organizations (MJO) predefined spatial areas (states/counties) as SA groups. These groups allow policymakers to show the various impacts of different jurisdictions on a single monitoring receptor location. Alternatively, users can define source apportionment groups based on inventory categories so the modeler can demonstrate the impact of certain sectors on a monitoring location. However, a problem with the current method of using SMOKE modeling system is that it requires the user to choose between the two approaches without mixing the schemes.

The UNC Institute for the Environment (UNC-IE) developed a tool, SA\_Regroup that enables apportionment by both spatial areas and inventory categories. The tool allows users to regroup the Source Apportionment (SA) ready emissions input data files for air quality models like the Community Multiscale Air Quality (CMAQ) and Comprehensive Air Quality Model with Extension (CAMx).

### 2.1 SA\_Regroup Tool Development

In 2014, the UNC-IE and EPA implemented a new feature in the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system to generate SA-ready sector-specific emissions output files for CMAQ and CAMx to understand the relationship between the emission sources and the receptors. The SA approach will allow users to identify and quantify the sources of ambient air pollutants and their effects at a site (receptor), primarily based on species concentration measurements and modeling. The details of SA output files are available from the link below:

<https://www.emascenter.org/smoke/documentation/4.7/html/ch06s14s03.html>

The “SA\_Regroup” tool will read existing sectors/group\_numbers SA grouped output files from SMOKE (Table 1) along with the new SA regrouping mapping input file called “REGROUP\_GROUPS” created by LADCO (Table 2) and then alter the SA group\_numbers based on a regroup mapping table that holds the list of desired input source sectors and desired regroup numbers (Figure 1).

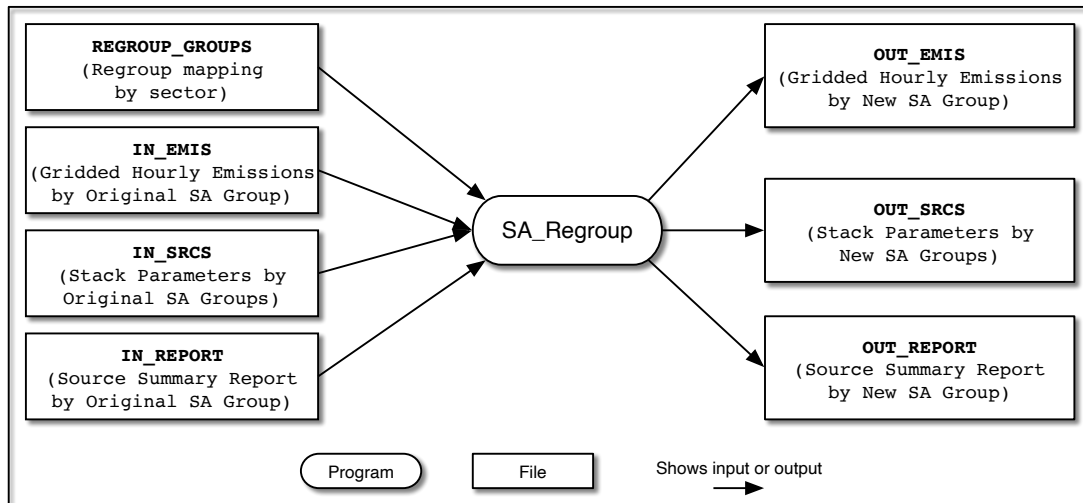


Figure 1. SA\_regroup Input and Output Files

Table 1. The list of SA-related output files from SMOKE modeling system.

|                 |                   |                                                                                                                     |
|-----------------|-------------------|---------------------------------------------------------------------------------------------------------------------|
| SGINLNTS[_S _L] | I/O API<br>NetCDF | Emissions by source group, formatted as CMAQ in-line emissions data, produced if SMK_SRCGROUP_OUTPUT_YN is set to Y |
| SRCGROUPS_OUT   | I/O API<br>NetCDF | Source group information, produced if SMK_SRCGROUP_OUTPUT_YN is set to Y                                            |
| SRCGRP_REPORT   | ASCII             | Source group report file, produced if SMK_SRCGROUP_OUTPUT_YN is set to Y                                            |

Table 2 is the format of REGROUP\_GROUPS input file.

| Lines | Position | Type    | Description                                                |
|-------|----------|---------|------------------------------------------------------------|
| 1+    | 1        | Chars   | Group ID Description                                       |
|       | 2        | Chars   | Sector name 1 [ex: nonpt]                                  |
|       | 3        | Chars   | Sector name 2 [ex: ptnonpipm]                              |
|       | ...      | Chars   | Sector name #                                              |
| 2+    | 1        | Integer | Group ID from the original SMOKE source group output files |
|       | 2        | Integer | New group ID of Sector 2 [ex: nonpt]                       |
|       | 3        | Integer | New group ID of Sector 2 [ex: ptnonpipm]                   |

| Lines | Position | Type    | Description              |
|-------|----------|---------|--------------------------|
|       | ...      | Integer | New group ID of Sector # |

While all original SA group IDs from the SMOKE output files will be regrouped to new sector-specific SA group ID based on the regroup mapping (REGROUP\_GROUPS) input file (Table 2).

The SA Regroup tool will remap the sector specific source group ID starting column #2 to the new source group ID from column #1 from REGROUP\_GROUPS input file. Any group IDs listed in column #1 will be remapped and output to the regrouped output files (OUT\_EMIS, OUT\_SRC, and OUT\_REPORT) listed in Figure 1.

## 2.2 What is in the package?

For a better consistency between UNC-IE and LADCO modeling system, UNC-IE developed the SA Regroup tool and its run scripts based on the U.S. EPA's National Emissions Inventory (NEI) Emissions Modeling Platform (EMP) system. **To run the tool successfully, users must have the full set of SA grouped output files from the SMOKE modeling system ready for the SA\_Regroup tool (Table 1).** The list of input and output files is shown in Figure 1.

### Run scripts:

```
$INSTALL_DIR/LADCO_report/scripts/sa_regroup/
```

- README
- sa\_regroup.csh
- run\_sa\_regroup.batch.csh
- SA\_REGROUP\_mapping\_input.csv (REGROUP\_GROUPS)

### Input files

- **REGROUP\_GROUPS:** SA Regroup Mapping Input File [**New**]
- **IN\_EMIS:** Gridded hourly emissions from original SA group output file from SMOKE
- **IN\_SRCS:** Stack parameters from original SA group output file from SMOKE
- **IN\_REPORT:** SA summary report from original SA group output report from SMOKE

### Output files

- **OUT\_EMIS:** Regrouped gridded hourly emissions from SA\_Regroup tool
- **OUT\_SRCS:** Regrouped stack parameters from SA\_Regroup tool
- **OUT\_REPORT:** Regrouped SA summary report from SA\_Regroup tool

### SA Regroup Run Script: "sa\_regroup.csh"

Once all sector-specific, SA-ready SMOKE output files are ready, users need to configure the following flags and environment variables in the run script, which are:

---

```

setenv MRG_SOURCE    $argv[1]      ! "mrg_source - A|B|M|P"
setenv MRG_SECTOR    $argv[2]      ! "sector - ag, nonpt, ptnonipm, ..."
setenv MRG_SDATE     $argv[3]      ! "date - YYYYMMDD"

setenv REGION_ABBREV  ${GRID}      # Define the modeling domain of input file

setenv REGROUP_GROUPS "SA_REGROUP_mapping_input.csv"

setenv IN_REPORT     ${INSTALL_DIR}/...../srcgrp_report_mole....txt
setenv IN_SRCS       ${INSTALL_DIR}/...../stack_groups_out....ncf
setenv IN_EMIS       ${INSTALL_DIR}/...../sginln_mole_....ncf

```

---

## 2.3 Run Scripts

Note that users must set the NEI EMP home directory, `$INSTALL_DIR`, by updating the main directory definition file called “`directory_definitions.csh`” located under:

```
> cd LADCO_reports/scripts/
```

To generate user-defined regrouped SA emissions output files for CAMx, users need to must go to the run script directory located at:

```
> cd $INSTALL_DIR/LADCO_reports/scripts/sa_regroup/
```

And invoke the the “`sa_regroup.csh`” run script as follows:

```
> ./sa_regroup.csh $MRG_SOURCE $MRG_SECTOR $MRG_SDATE
```

### Use Case Location

As a part of deliverables, we have created the use case for two NEI EMP sectors (nonpt and ptnonipm). These two sectors run can be executed interactively:

For nonpt (Area Nonpoint sources sector)

```
> ./sa_regroup.csh A nonpt 20160606
```

For ptnonipm (Point ptnonEGU sources sector)

```
> ./sa_regroup.csh P ptnonipm 20160606
```

Or, users can run the tool using the batch run script that can invoke multiple sectors and multiple consecutive dates:

```
> ./run_sa_regroup.batch.csh
```

Note that users must list any desired sector names in these run scripts knowing which source category (e.g., A, M, P, B) should be assigned to each inventory sector. Also, update the date information (e.g., year, mon, sday, and eday) to run multiple consecutive dates in a single execution. Because some sectors only ran for representative dates, only those dates will the output files be generated.

Once all the use case run scripts are executed, you can find the output files from SA\_Regroup tool at:

```
$INSTALL_DIR/LADCO_reports/smoke_out/LADCO_reports/12US1/cmaq_cb6/${SECTOR  
}/regroup/
```

Also, the output files can be compared against to the ones from the **answer** directory located at:

```
$INSTALL_DIR/LADCO_reports/smoke_out/LADCO_reports/12US1/cmaq_cb6/${SECTOR  
}/regroup/answers
```

## 2.4 SA\_Regroup Source Code and Makefile

As a part of deliverables, we have uploaded the SMOKE SA Regroup source code and Makefile in the package called “SA\_Regroup\_source\_code.tar.gz”. The following files are included in this package:

1. emutil/saregroup.f
2. Makefile

### 3. Task 2: High-resolution Windowed Emissions Extractor: “SubDomain”

There is a substantial body of work in data analytics to compare measurement results against emissions inventories. However, the problem often is that the inventories from large databases are very complex and difficult to compile. The existing model inventory to satellite comparisons simplifies to regional ratios of pollutants. The objective of this task is to develop a tool that allows for more complex comparisons that would include temporal comparisons between the different data sources. UNC-IE developed an inventory data extractor that can provide user-defined windowed high-resolution inventory data that can be compared to measurement data including both terrestrial and remote sensing technologies.

In general, the air quality model-ready, gridded, and chemically speciated hourly inventory emissions for all emission sectors are generated by the SMOKE modeling system. It can generate various types of high-quality inventory emission data for various users. However, it requires a substantial amount of resources and can be time-consuming without the proper support of experienced emissions modelers.

Under this task, UNC-IE developed a utility package called “**Sub-Domain**” that extracts temporally and spatially resolved high-quality inventory data that non-experts can easily use without much of emissions modeling understanding.

#### 3.1 Sub-Domain Inventory Extractor Development

The “**Sub-Domain**” Inventory Extractor is developed based on all the intermediate output files from the SMOKE modeling system. Instead of massive state/county/scc/cell/hourly summary reports for each day, this will minimize the size and number of input files required to generate the high-resolution inventory data required for this project. There is a Quality Assurance program in SMOKE called “**Smkreport**.” It is a powerful tool that can generate various types of QA summary reports based on all the intermediate output files from other SMOKE programs like Smkinven, Grdmat, Temporal, Spcmat, and more. The detail information of Smkreport can be found in [Chapter 7](#) of the SMOKE user’s document

(<https://www.cmascenter.org/smoke/documentation/4.6/html/ch07s03s02.html>).

Within its functionalities, Smkreport can not only generate high-resolution emissions like chemically speciated and gridded hourly emissions that one expects from an air quality modeling system, but also extract sub-domain region of emissions from the original modeling domain by defining the lower-left corner (LLC) and the upper-right corner (URC) of grid cells locations by column and row. In this project, we used this pre-existing functionality to extract the high-resolution gridded hourly emissions from the original modeling domain based on the user-defined sub-domain region. However, the main focus of this sub-domain inventory extractor development is for the community of people who are not familiar with the concept of the emissions modeling system. So, defining the LLC and URC grid cell locations can be very challenging, and in most cases, it is not applicable. Thus, we have developed a python program that can easily understand user-defined LLC and URC latitude and longitude coordinates instead of grid cell row and column locations. Also, this python program can generate the sub-domain modeling domain polygon GIS shapefiles based on user-defined sub-domain regions. GIS community can then overlay this

shapefile on the map with the gridded emissions produced by the Smkreport program in the SMOKE modeling system.

The SMOKE modeling system consists of various core programs that can import, temporally and spatially allocate, and chemically speciate emission inventory data to generate high-quality air quality model-ready, gridded, and speciated hourly emissions. Each SMOKE core program generates a highly condensed matrix of intermediate output files in binary format. Sub-Domain InvExtractor will use these intermediate binary output files from the SMOKE core programs to generate the high-resolution user-defined sub-domain region inventory dataset quickly. Figure 1 shows the input and output files of the Sub-Domain InvExtractor program.

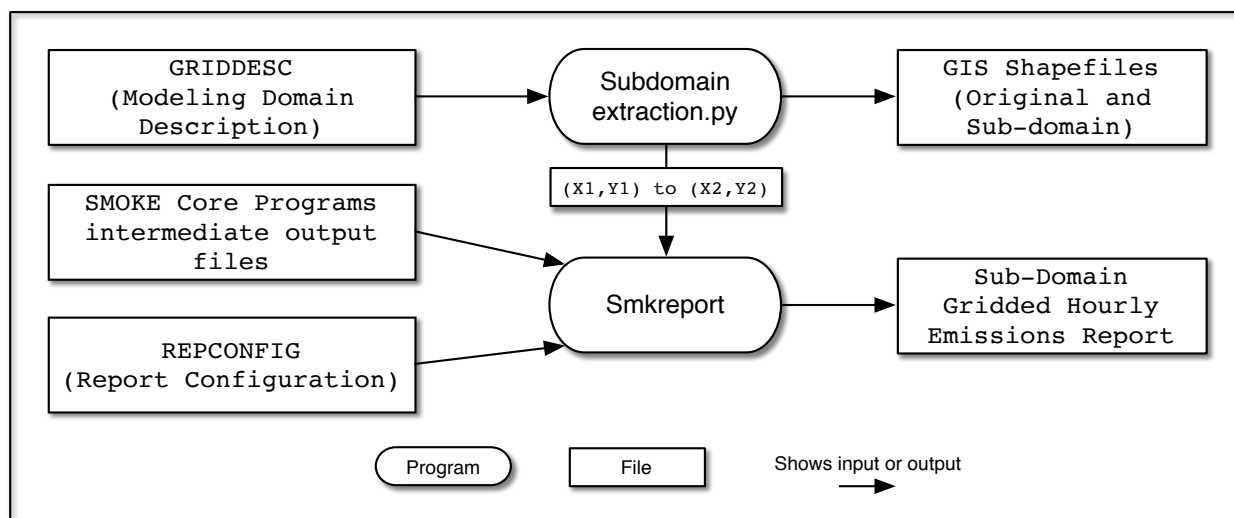


Figure 2. Sub-Domain InvExtractor Input and Output Files.

### 3.2 What is in the package?

Similar to the SA\_Regroup tool package, we developed the sub-domain extraction python tool and run scripts based on the U.S. EPA's NEI EMP system. Sub-Domain InvExtractor run script will use all the intermediate output files located under \$ INSTALL\_DIR/\$CASE/intermed directory.

**NOTE: Users must complete the successful SMOKE NEI EMP runs for all sectors and all 12 months prior to the use of this tool.**

To successfully execute this tool, users must integrate the following listed files into their existing NEI EMP system.

- **New Sample Sub-domain region Smkreport summary report:**  
\$INSTALL\_DIR/LADCO\_report/reports/subdomain/rep\_\${sector}\_\${month}...
- **New Smkreport report configuration template for SUBGRID region:**  
\$INSTALL\_DIR/ge\_dat/smkreprot/repconfig/subdomain
- **New Sub-domain InvExtractor related run scripts**  
\$INSTALL\_DIR/smoke\_v4.5/scripts/emf/smk\_monthly\_emf\_rep\_subdomain.csh  
\$INSTALL\_DIR/smoke\_v4.5/scripts/run/qa\_run\_subdomain\_v1.csh  
\$INSTALL\_DIR/LADCO\_report/scripts/run\_settings.txt  
\$INSTALL\_DIR/LADCO\_report/scripts/run\_settings\_README.txt  
\$INSTALL\_DIR/LADCO\_report/scripts/subdomain/

All sub-domain related inventory extract SMOKE run scripts, python GIS shapefile generator, and wrapper python program are located at:

```
> cd $INSTALL_DIR/LADCO_reports/scripts/subdomain/
```

- **run\_subdomain\_inv\_extractor.py:** Sub-domain wrapper python program
- **subdomain\_extraction.py :** Sub-domain GIS shapefile generator (X1,Y1 to X2,Y2)
- **run\_smkreport\_subdomain\_nonpoint.csh:** Area sector sub-domain SMOKE run scripts
- **run\_smkreport\_subdomain\_point.csh:** Point sector sub-domain SMOKE run scripts
- **run\_smkreport\_subdomain\_onroad.csh:** Onroad sector sub-domain SMOKE run scripts
- **subs :** Substituting strings c-shell script

### **“Subdomain\_extraction.py”: Domain GIS Shapefile Generator**

Once all the packaged files are copied to the right location to LADCO’s EMP system, the user must install the python3 with the following libraries to execute the sub-domain shapefile generator python program successfully. There are two ways to install these python libraries. \

- If you have admin right, then just do the following:
  - a. pip3 install pandas
  - b. pip3 install geopandas
  - c. The above 2 package will install numpy, shapely and pyproj too.
- If you don’t have the admin right on your machine:
  - a. pip3 -m venv env [env is a local directory you want the virtual environment to install]
  - b. source ./ven/bin/activate.csh [This will activate the virtual environment]
  - c. You will see (venv) before your command prompt
  - d. pip3 install pandas
  - e. pip3 install geopandas
  - f. The above 2 package will install numpy, shapely and pyproj too.

A sub-domain GIS shapefile generator will create two polygon shapefiles for two modeling domains (original and sub-domain) as well as the subdomain row and column location output file called “**subdomain\_location.txt**” based on user-defined LLC and URC latitude and longitude coordinates. Once the two corners of latitude/longitude coordinates (LLC and URC) are provided, the **Subdomain\_extraction.py** tool will first identify the grid cells based on their coordinates. It will then draw the nested sub-domain box internally to extract the gridded hourly inventory pollutant emissions in CSV format using the **Smkreport** program.

### **SMOKE Inventory Extract Run Scripts**

Because there are distinct differences in run settings between area (A), point (P) and mobile (M) source categories, we created three source category-specific run scripts to extract the user-defined sub-domain region gridded hourly emissions.

- `run_smkreport_subdomain_nonpoint.csh`: **Area Nonpoint InvExtractor** run script
- `run_smkreport_subdomain_point.csh` : **Point InvExtractor** run script
- `run_smkreport_subdomain_onroad.csh`: **Mobile Onroad InvExtractor** run script

Once all sector-specific SMOKE runs are complete under the U.S. EPA EMP system, users need to configure the following flags and environment variables in the run script, which are:

```
-----
## Define the processing sector name
setenv SECTOR $argv[1]

## Emissions modeling year/month/date
# (i.e. meteorological year, not necessarily the inventory year"
setenv EPI_STDATE_TIME "$argv[2] 00:00:00.0"
setenv EPI_ENDATE_TIME "$argv[2] 23:59:00.0"
setenv BASE_YEAR `echo $EPI_STDATE_TIME | cut -d\ - -f1`
setenv RUN_MONTHS `echo $EPI_STDATE_TIME | cut -d\ - -f2`
setenv SPINUP_DURATION "0"

# Define the location of input files
setenv GRIDDESC "${GE_DAT}/gridding/griddesc_lambertonly_08dec2017_v0.txt"
setenv HOLIDAYS "${GE_DAT}/temporal/holidays_13feb2017_v1.txt"

# Define the original modeling domain name
setenv GRID $REGION_IOAPI_GRIDNAME # original domain name
setenv SUBDOMAIN_NAME "Chicago_region" # user-defined subdomain name

# Define the sub_domain name
setenv SW_LAT $argv[3] # Define the lower-left corner Latitude
setenv SW_LON $argv[4] # Define the lower-left corner Longitude
setenv NE_LAT $argv[5] # Define the upper-right corner Latitude
setenv NE_LON $argv[6] # Define the upper-right corner Longitude
setenv SHAPE $argv[7] # -s: Generate original/subdomain shapefiles
setenv KML $argv[8] # -k: Generate subdomain KML file

# Sector-specific representative dates setup: Varying by sector
setenv L_TYPE "week"
```

```
setenv M_TYPE "week"
setenv RUN_HOLIDAYS "Y"
```

---

Prior to define the modeling date, users must complete the successful SMOKE NEI EMP runs for all sectors for all 12 months prior to the usage of this subdomain extraction tool. U.S. EPA NEI EMP inventory sector has its preset modeling representative days to enhance the computational time and storage space based on their own characteristics. Those sector-specific **representative** modeling dates are called SMOKE merging dates are listed below:

- "aveday" = One representative day per month
- "mwdss" = Four representative days per month: [Mon, Weekday (Tue-Fri), Sat, Sun]
- "week" = Seven representative days per month: one for each day of the week
- "all" = No representative days; each day is processed individually

As a part of package, we generated SMOKE merging dates input files from modeling year 2011 to 2016. If users wish to generate other year-specific SMOKE merging dates input files, please follow the detail instruction described in the READ document

“README\_create\_merge\_dates” to invoke the python run script

“create\_merge\_dates\_ann.py” located at:

```
> $INSTALL_DIR/smoke4.5/scripts/smk_merge_dates/
```

To generate year-specific SMOKE merging dates files, user can invoke the

“create\_merge\_dates\_ann.py” python run script:

```
> create_merge_dates_ann.py YYYY
```

To define the sub-domain region in the original modeling domain, users must define the four variables to draw the region with LLC and URC latitude and longitude coordinates as arguments:

```
setenv SW_LAT    $argv[3]      # Define the lower-left corner Latitude
setenv SW_LON    $argv[4]      # Define the lower-left corner Longitude
setenv NE_LAT    $argv[5]      # Define the upper-right corner Latitude
setenv NE_LON    $argv[6]      # Define the upper-right corner Longitude
```

Also, users can optionally generate subdomain GIS shapefile and/or KML shapefile using the optional commands: “-s” for GIS shapefile and “-k” for KML shapefile generations.

Users must generate sub-domain GIS shapefile using the GIS domain shapefile generator python program called “subdomain\_extraction.py” at least once in the beginning. Once the original and sub-domain shapefiles are generated by the python tool, users can optionally turn off this feature to speed up the high-resolution Smkreport sub-domain reports generation simply by removing “-s” for GIS shapefile and “-k” for KML shapefile.

```
setenv SHAPE     $argv[7]      # -s: Generate original/subdomain shapefile
```

```
setenv KML          $argv[8]          # -k: Generate subdomain KML file
```

Because the setting for each sector is unique, we have developed the sector-specific run scripts based on the consistent setting from the U.S. EPA based on the week and holiday settings (L\_TYPE and RUN\_HOLIDYS).

Unlike other sectors, on-road mobile sectors like RPD, RPV, RPH, and RPP require one additional flag to define whether output the activity (like VMT, VPOP, and HOTELLING) or inventory pollutants. To generate hourly emissions instead of activity, users need to set “MTMP\_OUTPUT\_YN = Y” to hourly emissions generated by **Movesmrg** program based on activity data and MOVES emission factors.

### 3.3 Python Wrapper Run Script: “run\_subdomain\_inv\_extractor.py”

The run\_smkreport\_subdomain.py is the main sub-domain inventory extractor run script. It provides user a way single entry to run sector-specific SMOKE C-shell run scripts.

User can invoke the run\_smkreport\_subdomain.py by just run it without any command line options and the program will show you what command line options are need. There are six entries which are mandatory and two are optional. The program will base on the sector name you entered to call the proper sector-specific SMOKE run script and generate the sub-domain inventory summary report.

As an example, user can invoke the the “run\_subdomain\_inv\_extractor.py” run script as follow:

```
> ./run_subdomain_inv_extractor.py -sector nonpt -date '2014-01-01' -  
subdomain subdomain -lllat 41.3 -lllon -88.3 -urlat 42.0 -urlon -87.0 -s 1  
-k 1
```

Following six arguments are mandatory:

- **-sector:** Sector name from SMOKE run: “all” for all sectors
- **-date:** Modeling date in format of “YYYY-MM-DD”
- **-subdomain:** Subdomain name for extracted domain
- **-lllat:** Lower left corner latitude
- **-lllon:** Lower left corner longitude
- **-urlat:** Upper right corner latitude
- **-urlon:** Upper right corner longitude

Following two arguments are optional:

- **-s:** 1 or 0 (1 – Create shapefile, 0 – don’t create shapefile. Default is 0)
- **-k:** 1 or 0 (1 – Create kml file and this will create a shapefile too, Default is 0)
- **-h:** Commands + option help command

This wrapper python program will gather all the necessary information and feed all the necessary information to the sector-specific SMOKE run scripts and subdomain\_extraction.py to generate a sub-domain report.

If you need the details on how to run “`run_subdomain_inv_extractor.py`” python wrapper run program, use the help command:

```
> ./run_subdomain_inv_extractor.py -h
```

### Use Case Location

As a part of deliverables, we have created the use case for three NEI EMP sectors (nonpt, ptagfire, and RPH [RatePerHour]). These three sectors’ run scripts can be executed interactively:

#### **For nonpt (Area Nonpoint sources sector)**

```
> ./run_subdomain_inv_extractor.py -sector nonpt -date '2014-01-01' -  
subdomain subdomain -lllat 41.3 -lllon -88.3 -urlat 42.0 -urlon -87.0 -s 1  
-k 1
```

#### **For ptagfire (Point agricultural fire sources sector)**

```
> ./run_subdomain_inv_extractor.py -sector ptagfire -date '2014-01-01' -  
subdomain subdomain -lllat 41.3 -lllon -88.3 -urlat 42.0 -urlon -87.0 -s 1  
-k 1
```

#### **For RPH (Mobile Onroad RatePerHour sources sector)**

```
> ./run_subdomain_inv_extractor.py -sector RPH -date '2014-01-01' -  
subdomain subdomain -lllat 41.3 -lllon -88.3 -urlat 42.0 -urlon -87.0 -s 1  
-k 1
```

Note that users must list any desired sector names in this run scripts knowing which source category (e.g., A, M, P, B) should be assigned to each inventory sector. Also, update the date information (e.g., year, mon, sday and eday) to run multiple consecutive dates in a single execution. The tool will generate the requested date-specific sub-domain gridded emissions output filename. However, the actual date in the output file will be based on the representative date since some sectors only ran for representative dates.

Once all the use case run scripts are executed, you can find the output files from Sub-Domain InvExtractor tool at:

```
>cd $INSTALL_DIR/LADCO_reports/reports/subdomain
```

Also, the output files can be compared against to the ones from the answer directory located at:

```
>cd $INSTALL_DIR/LADCO_reports/reports/subdomain/answers
```

## 4. How to Integrate into an existing NEI Emissions Modeling Platform

To integrate these new tools into an existing NEI Emissions Modeling Platform, users need to perform the following few steps:

- Copy this new subdomain-related EMF run script to an existing EMF run script directory:

```
$INSTALL_DIR/smoke4.5/scripts/emf/smk_monthly_emf_rep_subdomain.csh  
$INSTALL_DIR/smoke4.5/scripts/run/qa_run_subdomain_v1.csh
```

- Copy this new Smkreport REPCONFIG input files custom-designed for Subdomain tool:

```
$INSTALL_DIR/ge_dat/smkreport/reconfig/subdomain/  
  a. For Area sources. : reconfig_area_subdomain_v1.txt  
  b. For Point sources : reconfig_point_subdomain_v1.txt  
  c. For Mobile sources: reconfig_onroad_subdomain_v1.txt  
  d. Custom built REPCONFIG: reconfig_subdomain.custom.txt
```

- Copy all new run scripts for Subdomain and SA regroup tools:

```
$INSTALL_DIR/LADCO_reports/scripts/subdomain/*  
$INSTALL_DIR/LADCO_reports/scripts/sa_regroup/*
```

## 5. Run Date Setting: run\_settings.txt

The user can define the modeling dates using the run setting input file located at \$INSTALL\_DIR/LADCO\_report \${CASE}/scripts directory. The detail information on how to setup the run setting input file can be found in “run\_settings\_README.txt”.

For an example, the following setting will invoke onroad\_cb6orig sector only from January 1<sup>st</sup> to 4<sup>th</sup>, 2014. The first line setting will enable running Smkreport run for everyday for modeling domain for 12US2. However, the 2<sup>nd</sup> line setting will limit for onroad\_cb6orig sector to run only from January 1<sup>st</sup> to 4<sup>th</sup>, 2019.

```
=====
Sector, Grid, Name, Part, Start Date, End Date, Value
all, 12US2, smkreport, PART2, 0, 0, Y
onroad_cb6orig, 12US2, smkreport, PART2, 01/05/2014, 01/31/2014, N
=====
```

## 6. SMOKE Report Configuration Template

This section describes the default Smkreport report configuration template file. By default, the current REPCONFIG input file is configured to output the most detailed, high-resolution source-level gridded hourly emissions. We use the /DEFINE SUBGRID/ function built in Smkreport

to define the sub-domain based on the user-defined LLC and URC latitude and longitude coordinates which converted to LLC Column (X1) and Row (Y1) and URC Column (X2) and Row (Y2) by the sub-domain GIS shapefiles generator discussed in next section (Figure 1).

Users also can output different types of reports by sub-domain by updating the REPCONFIG template input file. It can optionally generate chemically speciated model species along with main inventory pollutants (i.e., NO<sub>x</sub>, VOC, PM<sub>2.5</sub>, Benzene, etc.) by using the **"SPECIATION MASS"** and **"BY"** commands in the REPCONFIG input file located at:

```
$INSTALL_DIR/ge_dat/smkreprot/reconfig/subdomain/
```

Please check out the details of how to update the REPCONFIG file from Chapter 7 in SMOKE user's document (<https://www.cmascenter.org/smoke/documentation/4.6/html/ch07s03s02.html>).

```
=====
SMK_SOURCE A
```

```
/DEFINE SUBGRID/ MY_GRID
(X1,Y1) TO (X2,Y2)
/END/
```

```
/DELIMITER/ ,
```

```
/NEWFILE/ REPORT30
/CREATE REPORT/
TITLE Stationary area
SELECT SUBGRID MY_GRID
TEMPORAL
GRIDDING
BY CELL
BY HOUR
BY SOURCE
NUMBER E13.5
UNITS ALL tons/day
/END/
=====
```