Environmental and Residential Population Analysis Multisite Tool User Guide

2-14-2025

Table of Contents

What is the Environmental and Residential Population Analysis Multisite Tool?

The Environmental and Residential Population Analysis Multisite (ERPAM) tool lets you easily and quickly compile residential and environmental information summarized across hundreds or thousands of places, all at the same time.

The ERPAM tool provides users with a multisite report, which shows an overall summary of environmental and residential conditions for the average resident aggregated across all analyzed locations.

This multisite information allows environmental and residential analyses to move beyond looking at a small number of indicators or a few residential populations to a more comprehensive picture of environmental conditions and populations near a whole set of facilities or in other types of places. Consequently, one might use this information to better characterize the areas relevant to a proposed action or initiative.

Using the Web Application

The home page shows ways to specify locations for analysis (on the left), weblinks at the upper right, a map in the center, and buttons below that for setting the distance, naming the analysis, and starting the analysis.

Site Selection - How do you Specify the Places you Want to Analyze?

Locations can be defined in a few ways when using the ERPAM Tool.

The tool can summarize the following:

- **Indicators near any set of points**
  You can see information for residents near all the points you provide locations for. This can show you information about residential populations who live in any of the communities close to certain types of regulated facilities, for example.
- **Indicators within any areas you have defined on a map**
  You can get a report on residents in certain areas if you have a list of polygons that represent areas of interest.

ERPAM lets you specify locations in any of these ways:

- **Points (latitude/longitude):** you can upload a table of points identified by their latitude and longitude geographic coordinates.
- **Shapes (a shapefile):** You can provide a shapefile with polygons that define specific boundaries like neighborhoods or other types of areas.

- **Industry (industry name or NAICS code)**: You can specify regulated facilities within industry categories of various types (using NAICS industrial codes), by picking an industry type from a list in the on-screen menu.

How would you like to identify locations?

To start picking locations, make sure ERPAM knows what kinds of places you are trying to analyze and how you will specify them. By default, when you open the app, it assumes you will be uploading a table of points identified by their latitudes and longitudes. If you want a different approach, you must specify that using the pulldown menus.

There are a few ways you can specify the locations you want to analyze. The two overall approaches are to upload a table of specific places, or to pick a whole category of places from a pulldown menu of categories like industry types.

After you have specified the type of location information you are using, then you can click Browse to upload the locational data or click in the NAICS window to scroll or type names of industries or code numbers.

Upload specific locations – What type of data are you uploading?

If you choose to upload a spreadsheet of locations, the data must have specific formatting and information as described below.

- The list of sites must be in the first tab (worksheet).
- The first row must be the column names, starting with column A.
- All other rows must be one site per row.
- The first row must identify them as point coordinates ("lat" and "lon" header names to denote latitude and longitude) or as "REGISTRY_ID" for regulated facility ID numbers (which the tool converts to points because it knows where each facility is).

You can also upload a shapefile of locations as explained below.

Latitude/Longitude file upload

Choose this option to upload a file with lat-long coordinates.

- Required filetype: .csv, .xls, or .xlsx
- Required Columns: lat, lon (but some aliases will work too, such as "latitude" and "longitude")

Note: An example of a lat/lon file with correct fields and formatting is provided via a link in the tool. Click to download, save, and then click Browse to try uploading that saved file as an example for trying out the tool.

Facility IDs (FRS identifiers)

The regulated facilities are stored in a database/tool called the Facility Registry Services (FRS). ERPAM uses a regularly updated snapshot of that database to store the latitude and longitude and

name, etc. of each active regulated site, based on a unique ID called the FRS ID or Registry ID. For example, the ECHO tool shows FRS IDs and lets you search by ID.

- Required filetype: .csv, .xls, or .xlsx
- Required Columns: REGISTRY_ID (but some aliases are recognized such as "regid" and as a last resort "siteid")

Note: An example of the Registry IDs file is provided via a link in the tool. Click to download, save, and then click Browse to try uploading that saved file as an example for trying out the tool.

Shapefile of polygons

Another option is to upload a shapefile that contains polygons. These polygons might represent neighborhood boundaries, areas defined by air quality modeling results, or places within some buffer distance of a rail line or pipeline, etc. ERPAM does not provide the shapefiles – you must create or obtain shapefiles elsewhere and then upload them to the tool for analysis of residents inside each polygon.

- Required files: .zip OR .shp, .shx, .dbf, .prj
- Required fields: geometry

Note: An example of a shapefile and the necessary components is provided via a link in the tool. Click to download, save, and then click Browse to try uploading that saved file as an example for trying out the tool.

Use the "Browse" button to upload a file

After you have specified the type of location information you are using, then you can click Browse to upload a spreadsheet or shapefile.

Note that trying to click "Browse" to upload a file will not work if you are trying to upload a type different than what is specified by the pulldown menus. For example, if you want to upload a shapefile, you must click in the pulldown menus to indicate that is the type you want.

Select a category of locations – How would you like to select categories?

Instead of providing a list of specific places, you can pick a category like all the facilities that are classified as one type of regulated site.

By Industry (NAICS) Code

Industries are classified into types and subtypes using NAICS codes that correspond to industry names. If you want to analyze all the regulated active facilities that are tagged as being in a particular industry, you can do so by NAICS.

In the box labelled "SELECT INDUSTRY OF INTEREST" you can click and then either browse or start typing a name or number. Industry names and codes will appear, along with counts of how many regulated facilities have been identified as under that NAICS. It is important to note

that a very large percentage of regulated sites in the Facility Registry Services (FRS) database do not have any NAICS information, so a set of facilities found using this approach cannot be considered a comprehensive list of all the regulated sites in a given NAICS.

The NAICS and FRS data used in ERPAM are snapshots of the information provided on active facilities in the FRS database.

Note: The FRS data used here were updated in August 2024.

See all subcategories of NAICS? (Basic or Detailed List)

This button helps you efficiently browse or search through the industry categories.

The basic/detailed radio button helps you quickly see which industry groups you might be looking for before you switch to a detailed view in case you need to pick a narrow/specific industry. It is recommended that you use the basic list to start exploring industry categories, given the sheer number of NAICS codes and subcategories.

However, if you know the specific NAICS code you want, you may want to start off by viewing the detailed list and just start typing a few letters to search for and quickly jump to the exact industry you are looking for.

A given regulated facility may be identified as being in only one NAICS code, but that code might be a broad category (e.g., a 3 or 4-digit code) or narrow category (e.g., a 6-digit code). The NAICS codes are organized like a tree-like hierarchy with subgroups and sub-subgroups. The short codes are 2 digits and represent extremely broad categories such as “Manufacturing.” The longer codes are very specific industries, such as “Walnut Farming.”

Add all subcategories of NAICS? (Yes or No)

This button helps ensure inclusion of all facilities you intend to include in the analysis.

Typically, you want to choose “Yes” to include subcategories. This means if you choose code 123, the tool will also pull in for analysis all the facilities under the umbrella of that broader category (e.g., codes 1234, 123451, 123452, etc.).

If you choose “No,” you run the risk of excluding sites you intended to include (e.g., a facility tagged as being in industry 123451 often is not explicitly tagged as also being under the broader 12345, 1234, 123, 12 codes).

Name of Your Analysis

You can type a name for your analysis, and that will appear in the results report and downloaded spreadsheet of results.

Distance from Site

Using the slider control, you can specify the distance from each point (the radius of the circle around each point) or the buffer (extra distance from the edges) added to polygons. For example, if you set it to 3.1 miles as the radius and use points defined by latitude/longitude, the analysis will include all residents within 3.1 miles (about 5 kilometers) of any one or more of the specified points. If you use a shapefile of polygons and add a buffer of 1 mile, the analysis will

account for residents inside each polygon plus any residents who are within 1 mile of any edge of a polygon.

Verifying the Uploaded Sites

Total location(s) uploaded information box

After you upload (or otherwise specify) locations, an information box shows how many sites were uploaded or selected and if any have invalid information such as bad or missing lat/lon data. The results table of all sites should reflect all the uploaded locations but will only provide results for the ones with valid location information.

Note: The ERPAM Report will not provide indicator results if a site is too small relative to the local population density, because it will not be able to adequately estimate how many residents are at that location.

Interactive Map

The map on the site selection page and the map in the ERPAM Report are "interactive" when viewed on the tool website and when used as a saved .html file when you download the report. This means you can interact with the map on the webpage or in the report, by clicking on it.

- Click + or – to zoom in/out the map, or use the scroll wheel on a mouse
- Click and drag on the map to move right/left/up/down
- Click on a point or polygon to see a popup window with information on the site.
- Click the small x at the upper right of a popup window to close that window.

Review selected sites button

- Click to see a table of the uploaded information.
- Click CSV or Excel to download that information in case you want to save it along with the assigned id column that is automatically created to keep track of the sites to be analyzed.
- Click Dismiss (bottom of the window) to close the view of the table.

Start Analysis button

Once you have selected sites to analyze and chosen a distance (optionally naming the analysis being performed), click the "Start Analysis" button. A progress bar appears at the lower right to show you it is still working on the results. If you are analyzing hundreds or thousands of sites, you may have to wait longer to get results. For example, analyzing 1,000 sites might take up to 30 seconds (less time for shorter distances). For a radius of about 3 miles, the tool may require roughly about one minute for 2,000 sites.

See Results tab

As soon as the analysis is finished, the "See Results" tab opens to display the summary as a Multisite Report.

Community Report tab / Multisite Report

This page features the overall summary results aggregated across all the analyzed sites. It counts each resident only once even if they are near two or more of the analyzed sites. This lets the tool correctly report values for the average resident who is at any one or more of the sites.

The report generated contains the following:

- **Barplot:** A barplot is provided in the Multisite Report, and this may be the first place you want to look because it summarizes overall results.
- **Tables:** Tables of data summarizing a "Value" for each indicator that is the indicator value for the average resident, across all residents at the locations analyzed.
- **Map:** The report also provides a map that is interactive (you can click in it to move around and get info on a single site, etc.).
- **Download Button:** You can download the report to save or share it, using the download button. A snapshot copy of it is also found in the downloaded detailed spreadsheet in the detailed results tab.

Barplot

The barplot is a very useful way to immediately see if any of the analyzed residential groups present in the analyzed locations are found at levels above overall national levels. It uses **RATIOS** to make it easy to see which indicators are relatively high and how high they are relative to reference areas (US or State). A heatmap approach was used, showing the bars in different colors to help quickly interpret notable results, as defined below.

- Gray bars show data where the ratio is no more than 1. This means the indicator value in analyzed locations is the same as or lower than the overall average value nationwide (or statewide).
- Yellow bars indicate a ratio over 1 but no more than 2, meaning the indicator value in analyzed locations is higher than the value nationwide or statewide, so the group is "overrepresented" in these locations or the indicator is "above average," but not by a factor of 2 (i.e., it is less than twice the overall average).
- Orange means a ratio over 2 (more than twice the overall average) but no more than 3.
- Red means a ratio over 3x, or at least triple the overall value.

Tables of Results

The ERPAM report has a table of residential population indicators, a table of environmental indicators and a few additional statistics such as information about languages spoken, etc.

One column summarizes the "Value" for each indicator, meaning the indicator value for the average resident, across all residents at the locations analyzed.

What the "State" Average or Percentile Means

Because the ERPAM report summarizes information aggregated across sites that may be in many States, the percentiles and averages are not for just one State. They reflect the mix of States for the population analyzed. The average shown is the average of the State averages where each State is weighted by the share of all analyzed residents who are in that State. The percentile shown is a weighted average of State percentiles across the analyzed residents.

Interactive Map

The map on the site selection page and the map in the ERPAM Report are "interactive" when viewed on the tool website and when used as a saved .html file when you download the report. This means you can interact with the map on the webpage or in the report, by clicking on it.

- Click + or – to zoom in/out the map, or use the scroll wheel on a mouse
- Click and drag on the map to move right/left/up/down
- Click on a point or polygon to see a popup window with information on the site.
- Click the small x at the upper right of a popup window to close that window.

Ratios to Averages

The ERPAM Tool uses ratios as another way to quickly communicate how high each indicator is. The ratios provide a useful perspective on the results, similar to how percentiles also give a sense of how high each indicator is. Just like percentiles, the ratios compare each indicator to the whole State or the whole US. By putting all indicators into one common scale, percentiles or ratios make it easy to compare across several indicators and to see which ones are relatively elevated or notable. Barplots in the tool show ratios for a few residential or environmental indicators side by side, where the highest bars have the highest ratios and are the most elevated relative to the US or State overall. The detailed tables provide all the ratios as well as the averages.

Each ratio is calculated by the tool as the value of the indicator for the average resident in areas analyzed divided by the overall value nationwide or statewide. For example, if the "Ratio to US avg %Hispanic" is 1.5, it means the %Hispanic among residents in the analyzed locations is 1.5 times as high as %Hispanic for the whole US. If the US overall is 18% Hispanic, a ratio of 1.5 means that almost 27% of residents in the analyzed locations were Hispanic (because 18% x 1.5 = 27%).

Download Report button – an interactive map and report in one file

The download button is located at the bottom of the report screen – You must scroll down to find it. Click the button to download and save a file that has an interactive version of the ERPAM Report. You can reuse that saved file later to look back at results overall or for each site, but you also can then email that .html file to anyone else and they will be able to use the interactive map that is stored inside that file, including seeing the popup windows.

Details tab

The details tab shows results that are more detailed than the summary report provides. It contains tabs named Site-by-site Table, and Plot Average Scores. Each is explained below.

Site-by-Site Table

This tab shows an interactive "preview" table that you can use, but you can also just skip to clicking the button that lets you save the results using the **DOWNLOAD RESULTS TABLE** button.

Using the Interactive preview table

The interactive table of site-by-site results has one row per site analyzed, and one column for each of the key indicators. It does not show all the indicators you can see in the downloadable results table – only some are shown in the web table view. You can scroll down and to the right to see more of the table.

Click on a column header to sort on that column and click again to change the order of sorting. For example, you can scroll to the right to see the column for % low income and then click that column header (the name of the column) to sort and see which sites have the highest scores for % low income.

Type in a box at the top of a column (under the header) to filter on that column. For example, you can view only results in one State or one EPA Region by clicking in that box.

The Barplot Report column is the only column not found in the spreadsheet download. If you click the button, it lets you download and then view a report that has two barplots with **RATIOS** for that one location: ratios to the US average and ratios to the State average, for each of several residential indicators. This gives a quick view of which, if any, residential indicators is notable at that one location.

The ECHO report column has links only when you have provided facility IDs in the process of specifying sites to analyze.

Download Results Table (as Spreadsheet) button

Downloading the results table is highly recommended – it is the only way to save your detailed results and contains a lot of useful information in several tabs.

Click the button to download a spreadsheet of results. Please wait for the download to complete as it takes several seconds to create the tables and save the file to your computer.

Understanding the Excel Spreadsheet (Downloaded File)

The downloaded spreadsheet has several tabs of information. The *Overall* tab provides the same summary aggregated numbers as the ERPAM Report described above. The *Each Site* tab has one row per site and one column per indicator. The two tabs have the same columns. Other tabs provide plots, a snapshot of the summary report, and notes about the analysis (how many sites, etc.). The tabs are explained below.

Excel or another similar tool will allow you to sort, copy/paste, etc. or use pivot tables, filters, or formulas to create custom summary metrics or plots.

In the Each Site or Overall tab, the color-coded cells form a "heatmap" that lets you quick spot higher scores. For the percentiles, the color coding is yellow/orange/red indicate 80th/90th/95th percentiles, respectively.

The heatmap for RATIOS columns uses yellow/orange/red the same way the ratio barplots use them: yellow/orange/red indicate >1x up to 2x/ 2x-3x / >3x the average, respectively.

There are many columns in the "*Each Site*" or "*Overall*" tab, each clearly labelled with a long name that is a full description or definition of the indicator. The columns include raw indicator scores, percentiles, ratios, averages, counts, and other miscellaneous information such as the State name, number of sites or blocks or blockgroups nearby, etc. See Appendix for some additional details.

- The id column is created by the tool to assign a unique identification number to each analyzed site. They appear in the same order as was provided in any uploaded table of sites.
- The *Community Report* tab has a snapshot of the ERPAM Report (not interactive), for convenience.
- The *Overall* tab shows the overall results (from the ERPAM Report), and has the same columns as the "Each Site" tab, but has only 1 row because it represents the average residents across all sites.
- The *Overall2* tab shows the overall results (from the ERPAM Report) but in a tall format.
- The *thresholds* tab shows how many of the indices exceed the 80th percentile, at each site.
- The *plot_distances* tab in the spreadsheet shows the average distance from the analyzed sites, for each residential group, with the areas analyzed.
- There is also a *Notes* tab with information about the analysis such as the number of sites and how much the total population is concentrated at a few key sites.

Plot Average Scores (Barplots)

There is a page with plots (charts) that is shown if you select See Results > Details > Plot Average Scores. The charts show how each residential group or environmental stressor, in the analyzed locations, compares to its US average. These are RATIOS.

These barplots let you quickly check which indicators of certain types are relatively elevatedand you can select indicator types and data types.

Select the Data Type by clicking either “Ratios to US” to see **RATIOS** or choose “Raw Data” to see the actual scores such as % low income or the PM2.5 concentration in air. The Raw Data option shows you the indicator values for the average resident and the average site, as well as for the US overall.

Average Person versus Average Site

Average site: The average site's average resident (the average resident's score is calculated at each site as the site-specific population-weighted mean, and then the arithmetic mean of those site-specific scores is calculated)

Average person at these sites: The average person among all the residents who are at any one or more of the sites, counting each person only once even if they live near more than one site.

Note: The average site may have a higher indicator value than the average person (resident) at those sites. This is common because some sites are in high population density locations, so they account for most of the analyzed population, and they drive the results overall when looking at the average resident. The average of sites will equally weight each site, including those with very few residents. The average person will equally weight each resident, even if most of them live at only a few sites in more urbanized areas.

Appendix

Additional Details on Calculations

How do the Calculations Work?

The ERPAM tool runs either a polygon-based or proximity-based analysis at each location, using an efficient method of spatial indexing to quickly identify nearby residents, for example.

The analysis features in are powered by a software toolkit developed by the United States Environmental Protection Agency (US EPA). That toolkit, called ERPAM, is also available as an open-source software package for researchers, developers, and data analysts.

Note: Some numbers as shown on the ERPAM report for a single location will in some cases appear very slightly different than in multi-site reports. All numbers shown in both types of reports are estimates, and any differences are well within the range of uncertainty inherent in the American Community Survey data. Slight differences are inherent in very quickly calculating results for multiple locations.

-