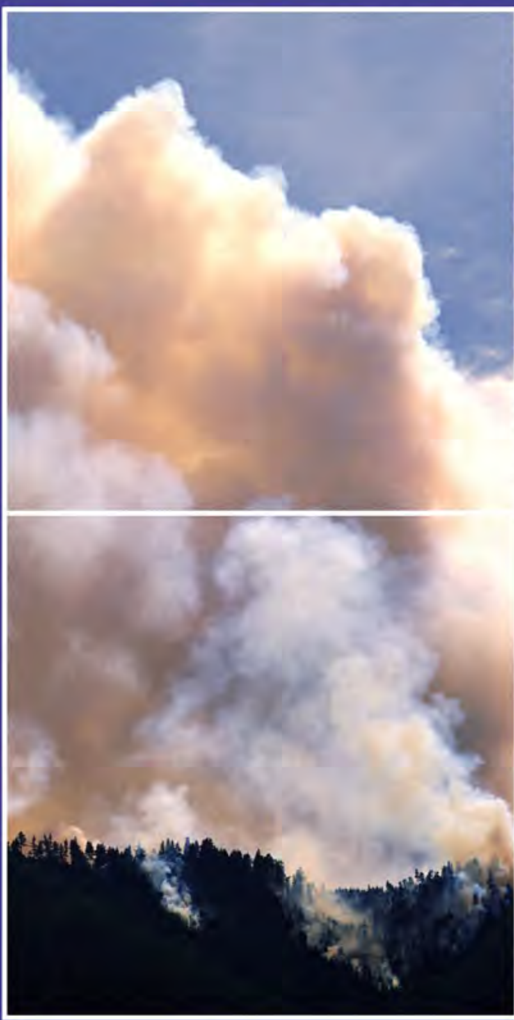
United States
Environmental Protection
Agency

EPA/600/R-14/305 September 2014 www.epa.gov/ord



Environmental Quality Index

Overview Report

Office of
Research and Development
National Health and
Environmental Effects
Research Laboratory
Environmental Public Health Division

United States
Environmental Protection
Agency

EPA/600/R-14/305 | September 2014 | www.epa.gov/ord

ENVIRONMENTAL QUALITY INDEX
Overview Report

Office of Research and Development
National Exposure Research Laboratory

Project Personnel

Danelle T. Lobdell, U.S. Environmental Protection Agency (EPA), Office of Research and Development (ORD), National Health and Environmental Effects Research Laboratory (NHEERL)

Jyotsna Jagai, University of Illinois at Chicago, Oak Ridge Institute for Science and Education (ORISE) Faculty Grantee

Lynne C. Messer, Portland State University, Support Contractor

Kristen Rappazzo, University of North Carolina (UNC), Department of Epidemiology, ORISE Grantee

Shannon Grabich, UNC, Department of Epidemiology, ORISE Grantee

Christine L. Gray, UNC, Department of Epidemiology, ORISE Grantee

Kyle Messier, Student Services Contractor

Genee Smith, Student Services Contractor

Suzanne Pierson, Innovate!, Inc., Geographic Information Systems (GIS) Contractor Support

Barbara Rosenbaum, Innovate!, Inc., GIS Contractor Support

Mark Murphy, Innovate!, Inc., GIS Contractor Support

Acknowledgments

External Peer Reviewers

Angel Hsu, Yale University, School of Forestry and Environmental Studies

Paul D. Juarez, University of Tennessee Health Science Center, Department of Preventive Medicine

Peter H. Langlois, Texas Department of State Health Services, Birth Defects Epidemiology and Surveillance Branch

Internal Peer Reviewers

Jane Gallagher, U.S. EPA, ORD, NHEERL

Thomas Brody, U.S. EPA, Region 5

Lisa Smith, U.S. EPA, ORD, NHEERL

Table of Contents

List of Figures

List of Maps

List of Tables

1.0 Introduction

A better way to calculate overall environmental quality is needed for researchers who study the environment and its effects on human health. This report is an overview of how the environmental quality index (EQI) was developed for all counties in the United States for the period 2000-2005. The EQI represents five areas (called "domains") of the environment ([1] air, [2] water, [3] land, [4] built, and [5] sociodemographic). In addition to the EQI, there is an index for each of the five domains. The EQI accounts for environmental differences between urban and rural areas by grouping counties into one of four rural-urban continuum codes (RUCCs), ranging from highly urban to rural-isolated areas.

The EQI was developed in four steps: (1) The five domains were identified, (2) data for each of the five domains were located and reviewed, (3) environmental variables were developed from the data sources, and (4) data were combined in each of the environmental domains; then these domain indices were used to create the overall EQI. The EQI relied on data sources that are mostly available to the public. The approach to creating the EQI is outlined, so others can repeat the steps for their own unique areas of interest.

This report gives an overview of the EQI. A companion report, Creating an Overall Environmental Quality Index, Technical Report, provides the detailed methodology and results. The variables, EQI, domain-specific indices, and EQI stratified by rural-urban data are available publically at the U.S. Environmental Protection Agency's (EPA's) Environmental Dataset Gateway. Also, an interactive map of the EQI is available at EPA's GeoPlatform.

Background

The assessment of environmental exposures for human health is changing, and new methods constantly are being developed. Exposures (both good and bad) that affect human health happen at the same time, but understanding their combined impact is difficult. For example, negative environmental features, such as landfills and industrial plants, often are located in neighborhoods with a high percentage of minority and poor residents.[1-7] On the other hand, high-income neighborhoods often have features that promote health, such as parks, health clubs, and well-stocked grocery stores.[8,9] Yet, no single exposure can be held responsible for good or poor health. It is not just good quality air or high income that produces health because many other exposures promote good health as well.

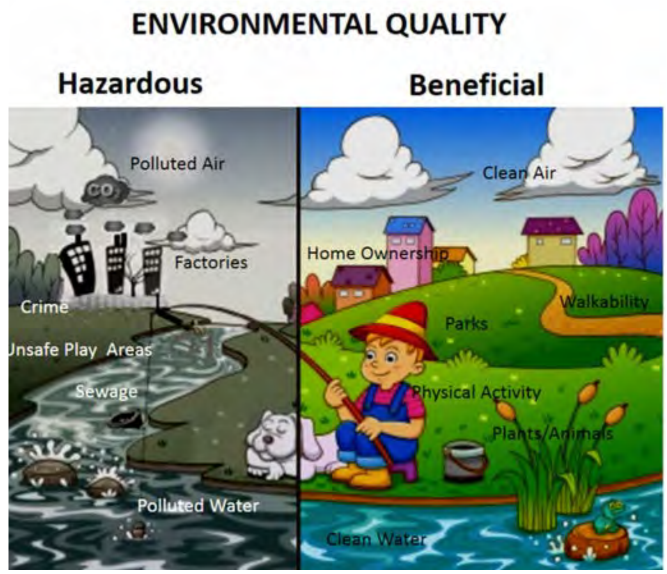


Figure 1. Conceptual environmental quality—hazardous and beneficial aspects.

One limitation to current methods in environmental health research is the focus on single-exposure types. Well-designed environmental health studies face a trade-off: Either researchers can collect a lot of high-quality data on only a few participants because collecting detailed exposure data is expensive and time-consuming, or researchers can collect less-detailed exposure data on a larger number of study participants because, the more participants in a study, the more expensive it is to conduct. This trade-off makes it impossible to account for many exposures that study participants might experience in addition to the main exposures of interest.

An index that summarizes many variables into a single variable is one approach that could improve statistical efficiency and still account for many environmental exposures at once. The index then could be used to identify areas with different levels of environmental quality. Clusters of negative environmental exposures could be identified and linked to health outcomes.

Conceptually, an EQI accounts for the multiple domains of the environment that encompass an area where humans interact (see Figure 1). These domains include chemical, natural, built, and sociodemographic environments that have both positive and negative influences on health. People move in and out of these positive and negative influences. Also, the positive and negative influences may even be co-located. As a result, the EQI examines both adverse health outcomes and protective health events.

Purpose

A better estimate of overall environmental quality is needed. It will improve the understanding of the relationship between environmental conditions and human health. Thus, an EQI was developed for all counties in the United States. The EQI uses indicators from the chemical, natural, built, and social environment. The EQI is composed of five environmental domains: (1) air, (2) water, (3) land, (4) built, and (5) sociodemographic.

Uses of EQI

The EQI was designed to be used in two main ways: (1) to represent "environmental quality" in research designed to assess the relationship between environmental quality and human health outcomes and (2) as a variable to account for surrounding conditions for researchers interested in a specific environmental exposure (e.g., exposure to pesticides) and human health outcomes (e.g., cancer). However, other uses of the data are expected by different end users, such as local, county, State and Federal governments, nongovernmental organizations, and academic institutions.

The EQI holds promise for improving environmental estimation in public health because it describes the surrounding county-level conditions to which residents are exposed. Use of the EQI will help public health researchers investigate the cumulative impact of many diverse environmental domains. The EQI was developed to help understand which domains (air, water, etc.) contribute the most to the overall environment. It also may be important for policymakers and environmental health workers to have information specific to the domains. Thus, domain-specific indices also were created. Each domain-specific index can be helpful to understand which domain is making the biggest contribution to the total environment in that particular county. This also can be expanded to understanding environmental differences by urban or rural status. In addition, researchers can use the EQI to control for environmental quality in their studies of specific exposures on health outcomes, adding environmental context to isolated exposures.

Another potential use of the EQI is for the comparison of county environmental quality across the United States. The EQI can be used to identify counties having a greater burden of poor health because of poor environmental quality and to see the important environmental domains contributing to an individual county's environmental quality. With the EQI currently at county level, environmental injustice may be difficult to tease out; however, the methods applied may be used to make local EQIs for smaller geographical areas.

2.0
Construction of the EQI

Domain Identification

Approach

Three sources were used to identify EQI domains:

1. EPA's Report on the Environment (ROE),[10]
2. an environmental health literature review (searches for published papers reporting on "environment" and "infant mortality"), and
3. expert consultation.

The ROE served as the starting point for the EQI. The media chapters from the ROE were used to identify environmental domains, data sources, and variables. Three domains were identified: (1) air, (2) water, and (3) land.

After reviewing the ROE, studies of environmental effects on infant mortality were reviewed. This enabled exploration of environmental domains using an indicator of national health and well-being. To be thorough, publications that came up in many searches were used to find more references. A broader definition of "environment" emerged.

Based on the literature search, the built and sociodemographic environments were explored. Negative environmental exposures have been associated with social exposures. A social epidemiologist and other experts were consulted to help create a broader definition of "environment" for the EQI.

Summary of Activities

Based on the three sources, (1) the ROE, (2) literature review, and (3) experts, five environmental domains were identified and developed for the EQI: (1) air, (2) water, (3) land, (4) built, and (5) sociodemographic.

Data Source Identification and Review

Approach

Predetermined categories were identified to represent each domain. Based on these categories, data were gathered for each domain (air, water, land, built, and sociodemographic) for all 3141 counties in the United States. The process included the following steps:

- find EPA and non-EPA environmental data sources;
- summarize the data sources in terms of availability, data quality, spatial and temporal coverage, storage requirements, and how to access the data;
- decide the most appropriate data sources for each domain; and
- obtain the identified datasets.

Possible data sources for each of the five domains were found using Web-based search engines (e.g., Google), site-specific search engines (e.g., Federal and State data sites), scientific data sources (e.g., PubMed, ScienceDirect, TOXNET), and personal communication from data owners. Data available for all U.S. counties for the years 2000-2005 was wanted. An inventory of all the found data sources was created.

Several criteria were used to assess data sources. Three key criteria included (1) data representing the predetermined category, (2) data quality, and (3) data coverage (available across the United States, including Hawaii and Alaska). Other factors were the ability to aggregate data at the county level and having data within the 2000-2005 time period. Ideally, data would be available every year from 2000 to 2005.

Summary of Activities

The overall data inventory is available at EPA's Environmental Dataset Gateway. Table 1 lists and describes the data sources that were used to make the EQI. An overview of the number of data sources kept for each domain is presented below.

Air Domain

Three data categories were considered: (1) monitoring data, (2) emissions data, and (3) modeled estimates representing concentrations of either criteria air pollutants or hazardous air pollutants (toxics). Twelve data sources were identified, and seven were considered for the EQI. Two were used for the air domain of the EQI because they were the most complete.

Water Domain

Five broad data categories within the water domain were identified: (1) modeled, (2) monitoring, (3) reported, (4) surveyed/studied and (5) miscellaneous data. Eighty data sources were identified. Five were used for the water domain of the EQI.

Land Domain

Land domain data sources were grouped into four categories: (1) agriculture, (2) industrial facilities, (3) geology/mining, and (4) land cover. Eighty sources were identified. Eleven were kept and used in the land domain of the EQI: two from agriculture, seven from facilities, and two from geology/mining.

Sociodemographic Domain

The sociodemographic domain is represented by crime and socioeconomic data. Only two data sources were kept for the sociodemographic domain of the EQI.

Table 1. Sources of Data for Air, Water, Land, Built-Environment, and Sociodemographic Domains for Use in the Environmental Quality Index

Air Domain			
Source of Data	**Description**	**Strengths**	**Limitations**
Air Quality System[11]	Repository of ambient air quality data, including both criteria and hazardous air pollutants (HAPs)	Measured values; network of criteria air pollutant monitors is substantial; measurement occurs regularly and is synchronized; data are audited for accuracy and precision.	The HAP network is sparse; some counties have no monitors, necessitating interpolation of concentrations for unmonitored locations.
National-Scale Air Toxics Assessment[12]	Estimates of hazardous air pollutant concentrations using emissions information from the National Emissions Inventory and meteorological data input into the Assessment System for Population Exposure Nationwide model	Validated models; coverage for all U.S. counties; majority of HAPs included.	Data are available at 3-year intervals; may underestimate concentrations; uses simplifying assumptions when information is missing or of poor quality; changes in methodology may result in different estimates between years.
Water Domain			
Source of Data	**Description**	**Strengths**	**Limitations**
Watershed Assessment, Tracking and Environmental Results Program Database/Reach Address Database[13]	Collection of EPA water assessments programs, including impairment, water quality standards, pollutant discharge permits and beach violations	Only database maintaining information on EPA Clean Water Act regulations	Data maintained and provided by States and, therefore, difficult to compare across States and not consistently reported with respect to temporal reporting and type of data reported across States.
National Contaminant Occurrence Database[14]	Samples both regulated and unregulated contaminants in public water supplies; maintained by EPA to satisfy statutory requirements for Safe Drinking Water Act	Provides measures for several chemicals and pathogens that are not measured elsewhere	Data provided by public water supplies; therefore, need to use spatial aggregation to get county-level estimates
Estimates of Water Use in the United States[15]	County-level estimates of water withdrawals for domestic, agricultural, and industrial use calculated by the U.S. Geological Survey	County-level estimates	Estimated based on various data sources
Drought Monitor Data[16]	Geographic information systems raster files reporting weekly modeled drought conditions. A collaboration that includes the National Atmospheric and Oceanic Administration, the U.S. Department of Agriculture, and academic partners.	Weekly coverage for the entire country	Modeled data; raster data, therefore, required spatial aggregation.
National Atmospheric Deposition Program[17]	Measures deposition of various pollutants, such as calcium, sodium, potassium, and sulfate, from rainfall	Weekly coverage for the entire country	Data not at the county level and required spatial interpolation.

Table 1. (continued) Sources of Data for Air, Water, Land, Built-Environment, and Sociodemographic Domains for Use in the Environmental Quality Index

Land Domain			
Source of Data	**Description**	**Strengths**	**Limitations**
National Pesticide Use Database: 2002[18]	Delineates State-level pesticide usage rates for cropland applications; contains estimates for active ingredients, of which 68 are insecticides, and 22 are other pesticides.	Provides a measure of pesticide usage	Pesticide rates only available at the State level for contiguous states; noncropland uses are not included.
2002 Census of Agriculture Full Report[19]	Summary of agricultural activity, including number of farms by size and type, inventory and values for crops and livestock, and operator characteristics	Can be used to approximate land- and water-related agricultural outputs (e.g., potential pesticide burden per acre, potential exposure to cattle, dust, etc.)	Not direct measures of pesticides or probable exposures
EPA Geospatial Data Download Service[20]	Maintained by EPA and provides locations of and information on facilities throughout the United States; different datasets within this database are updated at different intervals, but most are updated monthly; no set spatial scale across datasets. Some provide addresses, some geocoded addresses, etc.	Indicators for major facilities (e.g., Superfund sites;[21] Large Quantity Generators;[22] Toxics Release Inventory;[23] Resources Conservation and Recovery Act Treatment, Storage, and Disposal Facilities and Corrective Action Facilities;[24] Assessment, Cleanup, and Redevelopment Exchange Brownfield sites;[25] and Section Seven Tracking System pesticide producing site locations[26]) are available.	Contains much more information than just the facilities, type, and location; for example, Standard Industrial Classification System and North American Industry Classification System codes, Native American jurisdictions, interest type, etc.
National Geochemical Survey[27]	Geochemical data (arsenic, selenium, mercury, lead, zinc, magnesium, manganese, iron, etc.) for the United States based on stream sediment samples	Provides county-level means and standard deviations for each element; sampled data interpolated over nonsampled space results in variance estimates.	Includes data from several surveys; therefore, sampling locations and number of samples available vary by location.
Map of Radon Zones[28]	Identifies areas of the United States with the potential for elevated indoor radon levels; maintained by EPA	Each U.S. county is assigned to one of three radon zones based on radon potential.	Data are not actual measurements of radon, and only three levels of radon potential reduce possible county-level variability.
Sociodemographic Domain			
Source of Data	**Description**	**Strengths**	**Limitations**
U.S. Census[29]	County-level population and housing characteristics, including density, race, spatial distribution, education, socioeconomics, home and neighborhood features, and land use	Uniformly collected and constructed across the United States and can be used for construction of a variety of different variables	Decennial census available every 10 years; sample data are available at more frequent (e.g., 1-, 3-, and 5-year) intervals; may underestimate concentrations; uses simplifying assumptions when information is missing or of poor quality
Uniform Crime Reports[30]	County-level reports of violent crime	General estimate of public safety exposure	Reporting may differ across geography

Table 1. (continued) Sources of Data for Air, Water, Land, Built-Environment, and Sociodemographic Domains for Use in the Environmental Quality Index

Built-Environment Domain			
Source of Data	**Description**	**Strengths**	**Limitations**
Dun and Bradstreet North American Industry Classification System codes[31]	Description of physical activity environment (recreation facilities, parks, physical-fitness-related businesses) food environment (fast-food restaurants, groceries, convenience stores) education environment (schools, daycares, universities) per county	Detailed, thorough data; geocoding to county level is likely accurate; ongoing updates.	Proprietary data; not publicly available
Topologically Integrated Geographic Encoding and Referencing[32]	Road type and length per county	National coverage	Different road types may not be equivalent across U.S. geography; confer different exposure risks.
Fatality Annual Reporting System[33]	Annual pedestrian-related fatality per 100,000 population; maintained by National Highway Safety Commission	County-level reports and annual updates	Pedestrian fatalities result from diverse types of events and are not well captured in the database.
Housing and Urban Development Data[34]	Housing authority profiles provide general housing details (low-rent and subsidized/section 8 housing); information updated by individual public housing agencies.	Complete data source for unique element of the urban built environment	Not all counties contain housing authority properties; when the value for housing authority = 0, no housing authority property is present.

Built-Environment Domain

Built-environment data sources were grouped by categories: traffic-related, transit access, pedestrian safety, access to various business environments (such as the food, recreation, health care, and educational environments), and the presence of subsidized housing. Twelve data sources were identified, and four were kept for the built-environment domain of the EQI: (1) one traffic-related, (2) one for pedestrian-safety, (3) one for use in the various business environments (physical activity, food, health care, and educational), and (4) one for subsidized housing.

Variable Construction

Approach

After researching and choosing data sources, variables were created to represent each of the five domains ([1] air, [2] water, [3] land, [4] sociodemographic, and [5] built environment]. New variables were created because raw data sources were not always appropriate for statistical analysis. For example, a data source might provide the count of Superfund sites in a county, but that raw count is not terribly informative for environmental health research because counts likely vary by the number of people who live in a county. Therefore, a population-adjusted count or rate variable is created, where the count of Superfund sites in a county is adjusted for the number of people who live in that county.

The process for creating variables was to

- make variables for each domain for each available year of data (2000-2005),
- look for pairs or groups of variables that are giving the same information statistically and decide which of the variables best represents the environmental domain (and remove the extra variables),
- look for missing data,
- look at the distribution and statistical properties of each variable and decide how it should be scaled for analysis, and
- average variables from 2000-2005 for each county.

Table 2 provides a listing of variables for each domain. Appendix II in Creating an Overall Environmental Quality Index, Technical Report lists all the variables considered for the EQI. It also lists which variables were kept and why others were not kept.

Table 2. List of Variables by Domain Included in the Environmental Quality Index

Domain	Variable Definition
Air	
	Particulate matter under 10 µm in aerodynamic diameter
	Particulate matter under 2.5 µm in aerodynamic diameter
	Nitrogen dioxide
	Sulfur dioxide
	Ozone
	Carbon monoxide
	1,1,2,2-tetrachloroethane
	1,1,2-trichloroethane
	1,2-dibromo-3-chloropropane
	2,4-toluene diisocyanate
	2-chloroacetophenone
	2-nitropropane
	4-nitrophenol
	Acetonitrile
	Acetophenone
	Acrolein
	Acrylic acid
	Acrylonitrile
	Antimony compounds
	Benzidine
	Benzyl chloride
	Beryllium compounds
	Biphenyl
	bis-2-ethylhexyl phthalate
	Bromoform
	Cadmium compounds
	Carbon disulfide
	Carbon tetrachloride
	Carbon sulfide
	Chlorine
	Chlorobenzene
	Chloroform
	Chloroprene
	Chromium compounds
	Cresol/cresylic acid
	Cumene
	Cyanide compounds

Domain	Variable Definition
Air	
	Dibutylphthalate
	Diesel engine emissions
	Dimethyl formamide
	Dimethyl phthalates
	Dimethyl sulfate
	Epichlorohydrin
	Ethyl acrylate
	Ethyl chloride
	Ethylene dibromide
	Ethylene dichloride
	Ethylene glycol
	Ethylene oxide
	Ethylidene dichloride
	Glycol ethers
	Hexachlorobenzene
	Hexachlorobutadiene
	Hexachlorocyclopentadiene
	Hexane
	Hydrazine
	Hydrochloric acid
	Isophorone
	Lead compounds
	Manganese compounds
	Mercury compounds
	Methanol
	Methyl isobutyl ketone
	Methyl methacrylate
	Methyl chloride
	Methylhydrazine
	Methyl tert-butyl ether
	Nitrobenzene
	N,N-dimethylaniline
	o-toluidine
	Polycyclic organic matter/polycyclic aromatic hydrocarbons
	Pentachlorophenol
	Phosphine
	Phosphorus
	Polychlorinated biphenyls

Table 2. (continued) List of Variables by Domain Included in the Environmental Quality Index

Domain	Variable Definition
Air	
	Propylene dichloride
	Propylene oxide
	Quinoline
	Selenium compounds
	Styrene
	Tetrachloroethylene
	Toluene
	Trichloroethylene
	Triethylamine
	Vinyl acetate
	Vinyl chloride
	Vinylidene chloride
Domain	**Variable Definition**
Water	
	Percent of stream length impaired in county
	Sewage permits per 1000 km of stream in county
	Industrial permits per 1000 km of stream in county
	Stormwater permits per 1000 km of stream in county
	Number of days closed per event in county, 2000-2005
	Number of days per contamination advisory event in county, 2000-2005
	Number of days per rain advisory event in county, 2000-2005
	Percent of population on self supply, average 2000 and 2005
	Percent of public supply population that is on surface water, average 2000 and 2005
	Calcium precipitation weighted mean
	Magnesium precipitation weighted mean
	Potassium precipitation weighted mean
	Sodium precipitation weighted mean
	Ammonium precipitation weighted mean
	Nitrate precipitation weighted mean
	Chloride precipitation weighted mean
	Sulfate precipitation weighted mean
	Total mercury deposition
	Percent of county in extreme or exceptional drought (intensity levels D3 and D4, respectively)
	Arsenic
	Barium
	Cadmium
	Chromium
	Cyanide
	Fluoride
	Mercury (inorganic)
	Nitrate
	Nitrite
	Selenium
	Antimony
	Beryllium
	Thallium
	Endrin
	Lindane
	Methoxychlor
	Toxaphene
	Dalapon
	di(2-ethylhexyl) adipate
	Oxamyl (Vydate)
	Simazine
	di(2-ethylhexyl) phthalate
	Picloram
	Dinoseb
	Hexachlorocyclopentadiene
	Carbofuran
	Atrazine
	Alachlor
	Heptachlor
	Heptachlor epoxide
	2,4-Dichlorophenoxyacetic acid
	Hexachlorobenzene
	Benzo[a]pyrene
	Pentachlorophenol
	1,2,4-Trichlorobenzene
	Polychlorinated biphenyls
	1,2-Dibromo-3-chloropropane
	Ethylene dibromide
	Xylenes
	Chlordane
	Dichloromethane (Methylene chloride)
	1,2-Dichlorobenzene (o-Dichlorobenzene)
	1,4-Dichlorobenzene (p-Dichlorobenzene)

Table 2. (continued) List of Variables by Domain Included in the Environmental Quality Index

	Vinyl chloride
	1,1-Dichloroethylene
	trans-1,2-Dichloroethylene
	1,2-Dichloroethane (Ethylene dichloride)
	1,1,1-Trichloroethane
	Carbon tetrachloride
	1,2-Dichloropropane
	Trichloroethylene
	1,1,2-Trichloroethane
	Tetrachloroethylene
	Benzene
	Monochlorobenzene (Chlorobenzene)
	Toluene
	Ethylbenzene
	Styrene
	Alpha particles
	cis-1,2-Dichloroethylene
	Silvex
Domain	**Variable Definition**
Land	
	Harvested acreage
	Irrigated acreage
	Farms per acre
	Manure applied
	Chemicals used to control nematodes
	Chemicals used to control disease
	Chemicals used to defoliate/control growth/thin fruit
	Animal units
	Herbicides
	Fungicides
	Insecticides
	Arsenic
	Selenium
	Mercury
	Lead
	Zinc
	Copper
	Sodium
	Magnesium
	Titanium
	Calcium
	Iron
	Aluminum
	Phosphorus
	Facilities per county population
	Radon zone
Domain	**Variable Definition**
Sociodemographic	
	Percent renter occupied
	Percent vacant units
	Median household value
	Median household income
	Percent persons with income below the poverty level
	Percent who do not report speaking English
	Percent earning greater than high school education
	Percent unemployed
	Percent work outside county
	Median number rooms per house
	Percent of housing with more than 10 units
	Mean number of violent crimes per capita
Domain	**Variable Definition**
Built Environment	
	Proportion of roads that are highways
	Proportion of roads that are primary streets
	Traffic fatality rate
	Percent of population using public transport
	Vice-related businesses
	Entertainment-related businesses
	Education-related businesses
	Negative-food-related businesses
	Positive-food-related businesses
	Health-care-related businesses
	Recreation-related businesses
	Transportation-related businesses
	Civic-related businesses
	Total subsidized housing units

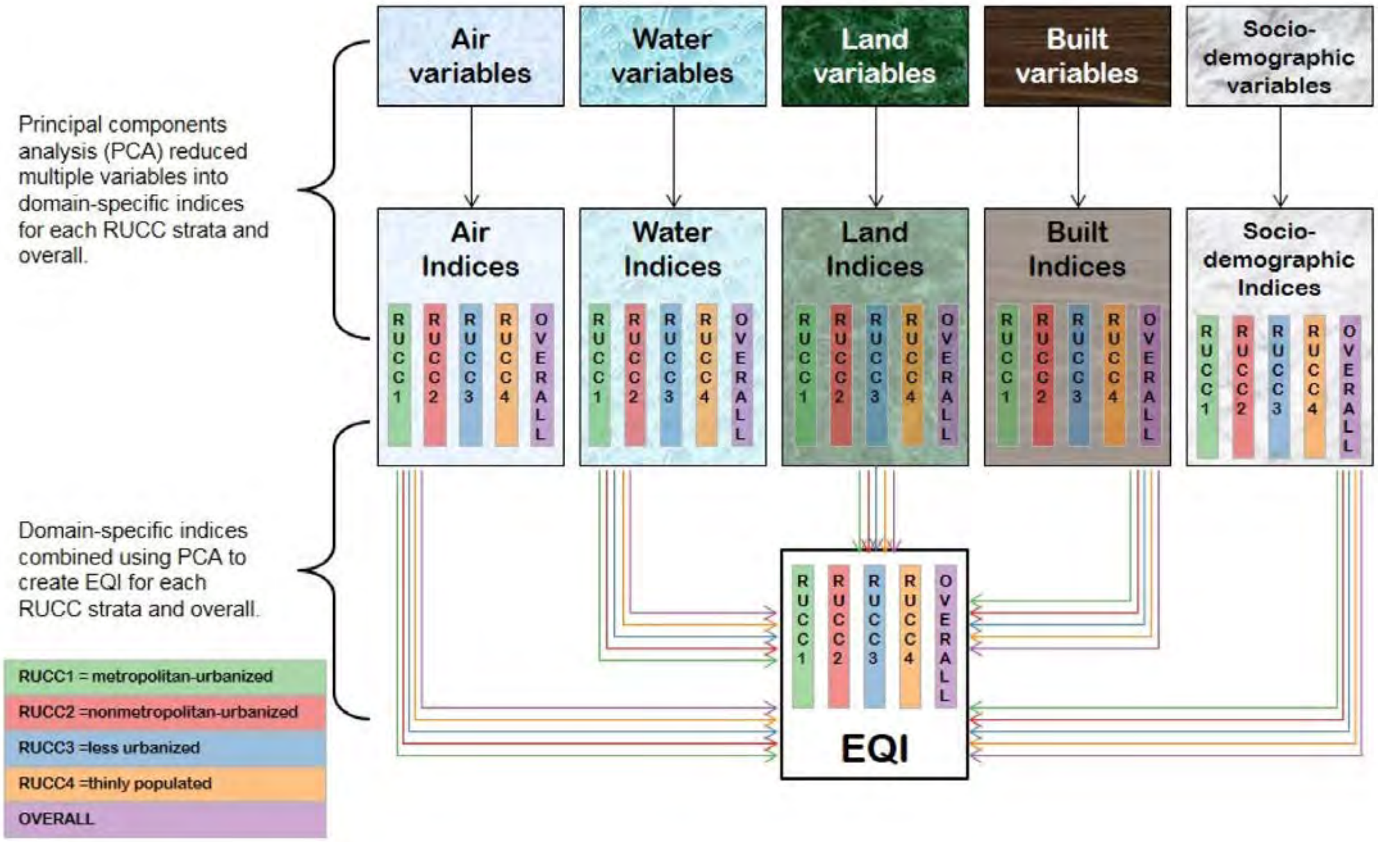


Figure 2. Principal component analysis for the Environmental Quality Index (EQI). All counties included with four rural-urban continuum codes (RUCCs).

Summary of Activities

New variables were created for each domain. These variables were created using data relevant to that domain. The variable characteristics were checked to make sure they were created in a way that would make sense statistically and would work with the chosen variable reduction method.

Data Reduction and Index Construction

Approach

After variables were created, they were combined into a single index (the EQI) using statistical methods. Each domain has its own index (air domain index, water domain index, etc.). Next, each of the domain-specific indices was used to create the overall EQI. The statistical process used to add these variables together is called principal component analysis (PCA). Figure 2 shows the steps that include

- use PCA on the variables in each domain to keep the most important piece of information for each domain index,
- use PCA on the domain indices to keep the most important information for the overall EQI, and
- group counties by their RUCC and repeat the two steps above for each RUCC group.

PCA

PCA is a statistical method that combines information from many variables into one summary variable, called an index. This "reduction" of many variables into one is useful because the one variable can be used in a statistical analysis of health outcomes, instead of trying to include hundreds of separate variables at the same time.

PCA was chosen to turn many variables into one index for a few reasons. It puts different variables into the same format (it "standardizes" them), so they can be added together. It provides each variable a measure of relative importance, or "weight", in its relationship to all the other variables included in the PCA. This weight is important for understanding which variables seem the most important for explaining the index. It takes into account how much of a variable is present, or its prevalence, in the overall environment. PCA then creates a single variable that can be used in other models. Researchers also can use the PCA values for each variable to understand differences in variables.

The domain-specific indices and the EQI were created for each county in the United States. The four RUCC groups were used to account for differences in rural versus urban areas. There were originally nine RUCC codes. Those nine were combined to make four RUCCs for the EQI: (1) RUCC1 represents metropolitan-urbanized = codes 1+2+3; (2) RUCC2 nonmetropolitan-urbanized = 4+5; (3) RUCC3 less urbanized = 6+7; and (4) RUCC4 thinly populated

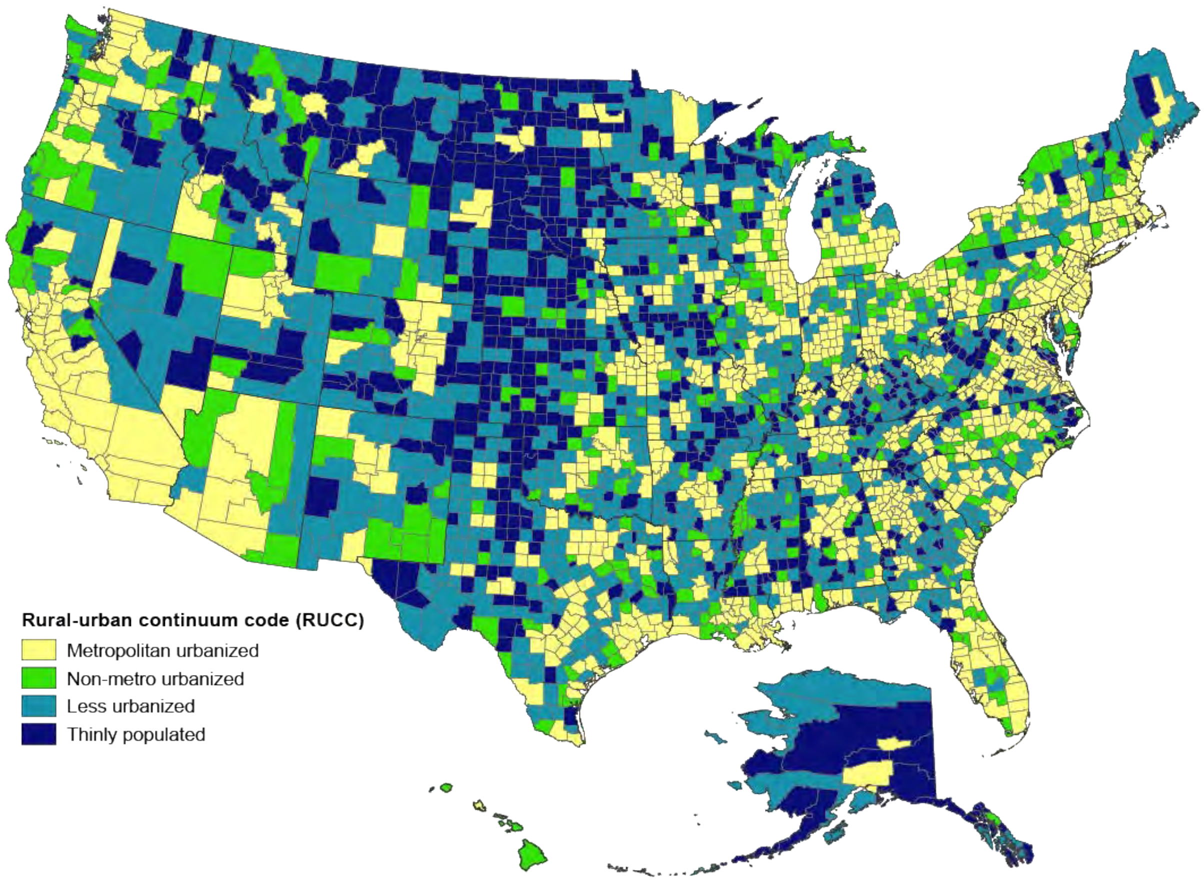


Figure 3. Rural-urban continuum codes (RUCCs) for all counties in the United States.

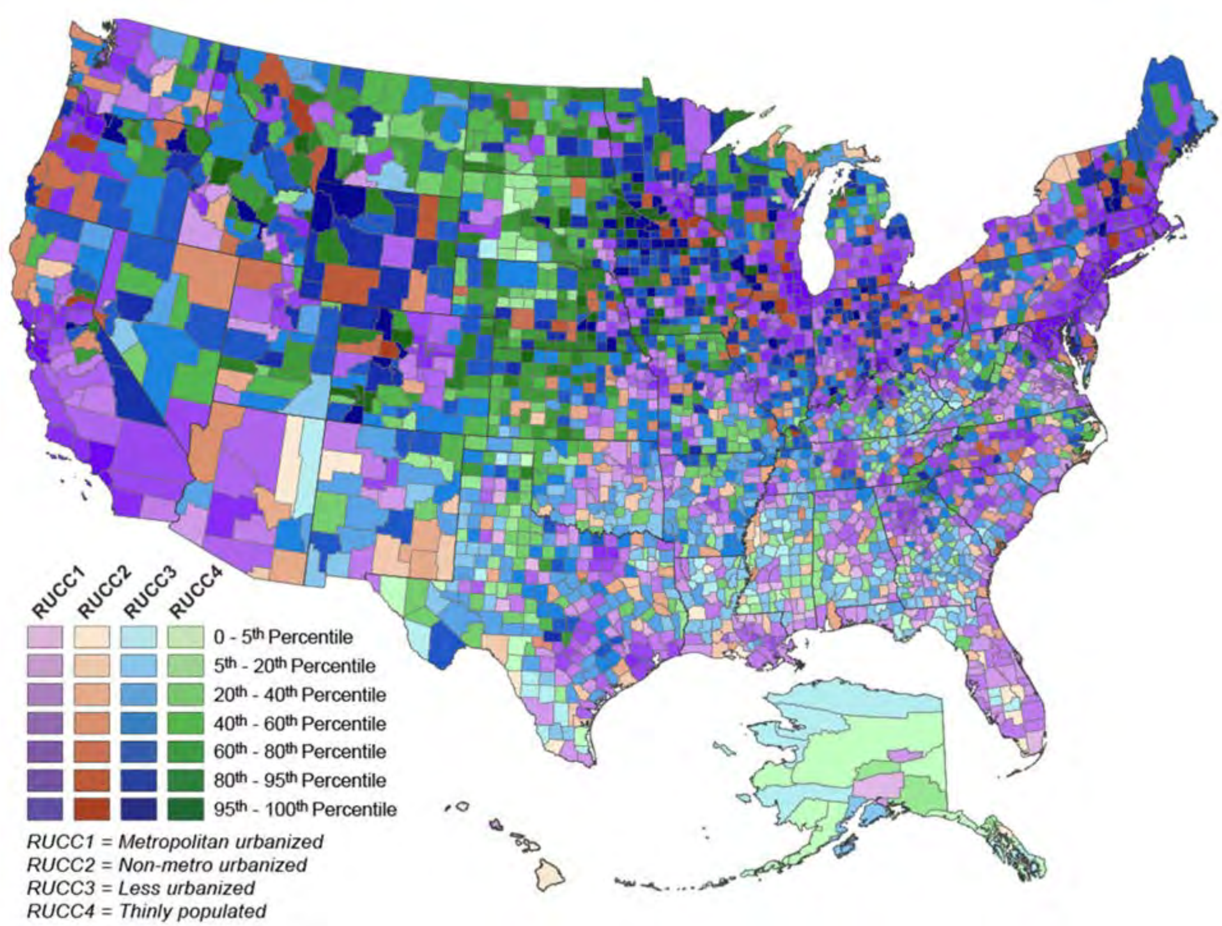


Figure 4. Map of the Environmental Quality Index by rural-urban continuum codes (RUCCs).

Table 3. Weights for Each Domain's Contribution to the Environmental Quality Index for 3141 U.S. Counties (2000-2005) and for the Counties Stratified by Their Rural-Urban Status (RUCC code)

	Metropolitan-Urbanized (RUCC1)	Nonmetropolitan-Urbanized (RUCC2)	Less Urbanized (RUCC3)	Thinly Populated (RUCC4)	OVERALL
Number of Counties	1089	323	1059	670	3141
Air Domain Index	0.5063	0.3343	0.1609	0.0285	0.4867
Water Domain Index	0.2757	0.2958	0.2981	0.1347	0.2618
Land Domain Index	0.4379	0.5506	0.5503	0.5785	0.3887
Sociodemographic Domain Index	0.4538	0.5963	0.5675	0.6263	0.5077
Built-Environment Domain Index	0.5196	0.3769	0.5102	0.5041	0.5345

Because PCA analyzes total, not shared, variance, the weights need not total 1.0.

(rural) =8+9 (see Figure 3).[35-38] The index-creation process was repeated for those four RUCC groups, leading to an overall EQI and five domain-specific indices for each RUCC group.

Results

For detailed results, consult Creating an Overall Environmental Quality Index, Technical Report.

Description of EQI

For EQI scores in RUCC groups, higher values suggest worse environmental quality, and lower values suggest better environmental quality. Figure 4 provides a map of the EQI by RUCC divided into percentiles, where the lower percentiles represent better environmental quality, and the higher percentiles represent worse environmental quality. The bulk of counties had EQI scores in the better range.

Additionally, Appendix I contains county maps for the nonstratified EQI and domain-specific indices, RUCC-stratified EQI, and RUCC-stratified domain-specific-indices. All indices were grouped into percentiles.

Domain-Specific Index Description

The way in which the domain-specific indices contributed to the EQI differed depending on how rural or urban the county was (Table 2). In the most urban areas (RUCC1), the built-environment domain had the most influence (0.5196, the weight associated with the built environment, is the largest number for the RUCC1 column from Table 2.). For the nonmetropolitan-urbanized areas (RUCC2), the sociodemographic and land domains had the most influence, and the water domain had the least influence. The air domain was the least influential for the less urbanized counties (RUCC3). In the most thinly populated counties (RUCC4), the sociodemographic and land domains were the most influential.

For the nonstratified EQI, the built and the sociodemographic domains had the most influence (0.5345 and 0.5077, respectively). The air domain also had a fair amount of influence, and the water domain had the least.

3.0 Discussion

An EQI was developed for all counties (N=3141) in the United States. This EQI includes five environmental domains: (1) air, (2) water, (3) land, (4) built, and (5) sociodemographic. For each domain, variables were created from many data sources. Then, domain-specific indices and an EQI were created using PCA. The EQI also is divided into four RUCC groups to account for rural-urban differences. The PCA shows that environmental quality is driven by different domains in rural versus urban areas.

Strengths and Limitations

Data

Data sources represented each of the five environmental domains. Documentation for each data source was good. Even though many data sources were found, gaps in the data remain.

The EQI is useful for representing the overall surrounding environment. It is not as useful for describing specific environments. If there were no data available for an important part of the environment, then the EQI was unable to capture that part. Areas, either counties or domains, with little data were not represented as well as areas with a lot of data.

It is difficult to find environmental data sources that fully cover all areas at all time intervals. Most data were not collected often enough. This is why an EQI covering 6 years was developed. If more data were collected more often, there would be an EQI for each year.

When counties had data values that were missing, information on those variables had to be estimated. This makes it harder to understand how pollutants affect urban and rural areas differently. Although many of the environmental data points were collected in smaller areas than counties (e.g., for a municipality or city), most are not maintained in a single source, such as a State or county data repository. National repositories for some domains exist (e.g., water, air), but no built-environment repository (for transit, walkability/physical activity, presence of sidewalks, or pedestrian lighting) is available. Cities or towns with less money may not be able to collect these data. Thus, data were available at different levels across the United States.

PCA Methodology

Using PCA had limitations. Normality is an important statistical assumption for PCA. Some data had to be scaled to be made normal. Scores from a PCA also can be hard to interpret. Outliers in the data also can be a limitation. However, with 3141 counties and proper statistical checks, this is not a big problem for the EQI.

Using PCA was also a strength of this project. PCA enabled a lot of variables to be combined into a single index. The EQI is standardized. This means it can be compared to other EQIs created in other countries or at different levels (e.g., city instead of county). Another strength is that PCA has been used to make other indices.[39, 40]

Application

The EQI was focused solely on the outdoor environment. This may not be the most relevant exposure in relation to human health and disease. The EQI is at the county level, not the individual level. This means it can be used to see which counties are less healthy environments. It will not be good at predicting which people are likely to have certain diseases.

Other Environmental Indices

The EQI is unique. Most other EQIs focus on one environmental domain (e.g., Air Quality Index[41]) or a specific type of activity (e.g., Pedestrian Environmental Quality Index[42]) or vulnerability (e.g., Cumulative Environmental Vulnerability Assessment,[43] heat vulnerability index[44]). State-specific indices also exist, (e.g., CalEnviroScreen 1.0,[45] Virginia Environmental Quality Index[46]), but they often cannot be compared to other States because the data are different.

Other indices are at a larger spatial resolution, usually at the country level. Country-level indices include the Environmental Sustainability Index[39] and the Environmental Vulnerability Index.[47]

Conclusions

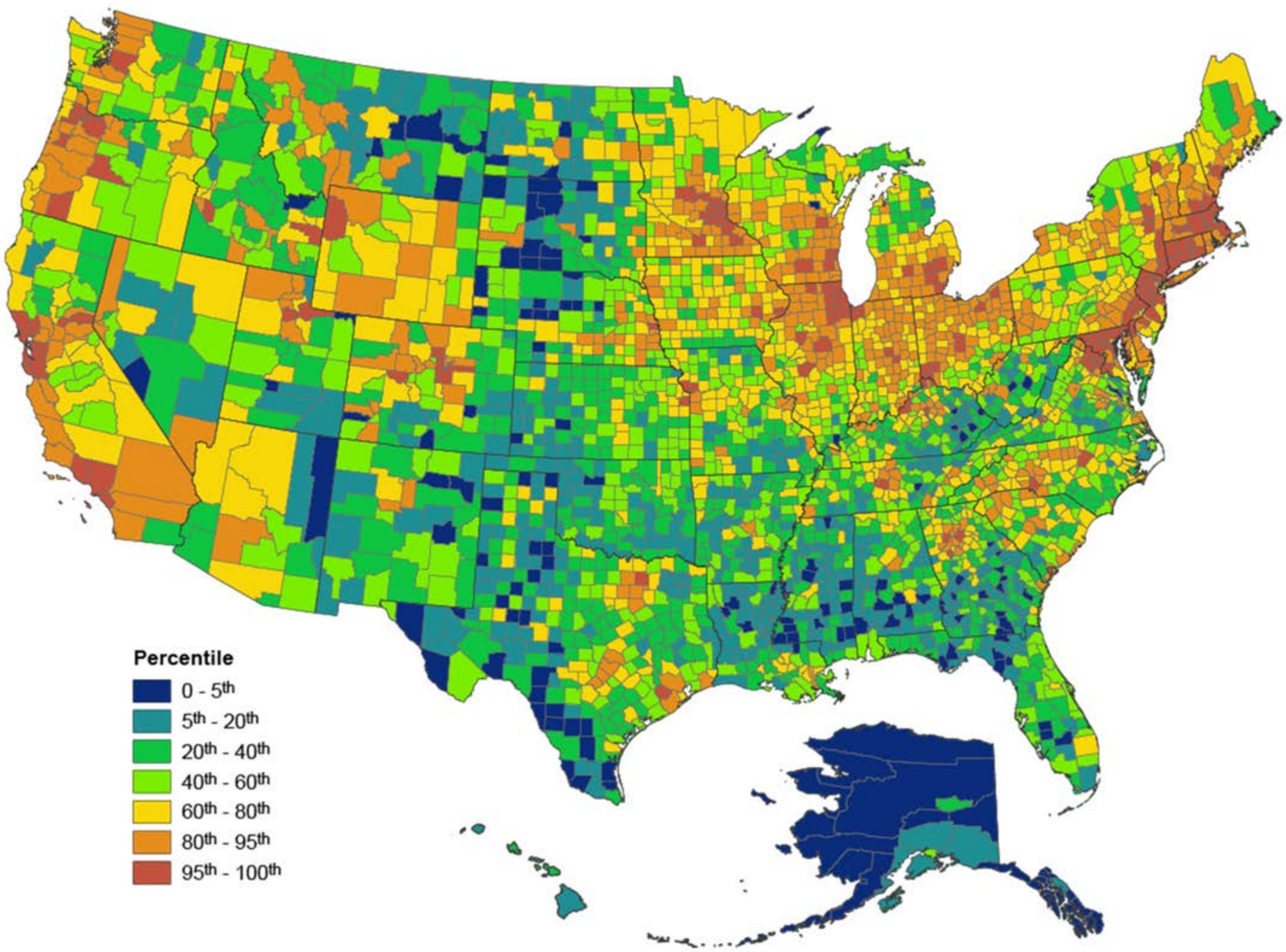
The EQI was constructed for all 3141 counties in the United States. The EQI has five environmental domains: (1) air, (2) water, (3) land, (4) built, and (5) sociodemographic. It is divided into four rural-urban groups. The methods can be repeated by others, and the data are available to the public. The EQI is a first step for looking at many environmental exposures at once. The EQI can be used as a measure in environmental health research. This broad effort uses many factors that work together to impact environmental quality and public health. Updates to the EQI for 2006-2010 are planned. Looking at smaller geographic areas also is planned.

4.0 References

1. Payne-Sturges, D., et al., Workshop summary: connecting social and environmental factors to measure and track environmental health disparities. Environ Res, 2006. 102(2): p. 146-53.
2. Mohai, P., et al., Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: evidence from the Americans' Changing Lives Study. Am J Public Health, 2009. 99 Suppl 3: p. S649-56.
3. Payne-Sturges, D. and G.C. Gee, National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. Environ Res, 2006. 102(2): p. 154-71.
4. Fan, A.M., G. Alexeeff, and S.B. Harris, Cumulative risks and cumulative impacts of environmental chemical exposures. Int J Toxicol, 2010. 29(1): p. 57.
5. Martuzzi, M., F. Mitis, and F. Forastiere, Inequalities, inequities, environmental justice in waste management and health. Eur J Public Health, 2010. 20(1): p. 21-6.
6. Johnson, B.L. and S.L. Coulberson, Environmental epidemiologic issues and minority health. Ann Epidemiol, 1993. 3(2): p. 175-80.
7. Norton, J.M., et al., Race, wealth, and solid waste facilities in North Carolina. Environ Health Perspect, 2007. 115(9): p. 1344-50.
8. Larson, N.I., M.T. Story, and M.C. Nelson, Neighborhood environments: disparities in access to healthy foods in the U.S. Am J Prev Med, 2009. 36(1): p. 74-81.
9. Lovasi, G.S., et al., Built environments and obesity in disadvantaged populations. Epidemiol Rev, 2009. 31: p. 7-20.
10. United States Environmental Protection Agency (EPA), EPA's 2008 Report on the Environment, 2008: Washington, DC.
11. United States Environmental Protection Agency (EPA). The Ambient Air Monitoring Program. 2010 [cited 2010 July 15]; Available from http://www.epa.gov/air/oaqps/qa/monprog.html.
12. United States Environmental Protection Agency (EPA). National Air Toxics Assessments. 9/10/2010]; Available from http://www.epa.gov/ttn/atw/natamain/.
13. United States Environmental Protection Agency (EPA). Watershed Assessment, Tracking and Environmental Results (WATERS). [cited 2010 August 26]; Available from http://www.epa.gov/waters/.
14. United States Environmental Protection Agency (EPA). National Contaminant Occurrence Database (NCOD). 2010 [cited 2010 August 26]; Available from http://water.epa.gov/scitech/datait/databases/drink/ncod/databases-index.cfm.
15. United States Geological Survey (USGS). Estimated Use of Water in the United States. [cited 2010 August 26]; Available from http://water.usgs.gov/watuse/.
16. National Drought Mitigation Center (NDMC). Drought Monitor Data Downloads. August 26, 2010 [cited 2010 August 26]; Available from http://www.drought.unl.edu/dm/dmshps_archive.htm.
17. National Atmospheric Deposition Program. National Atmospheric Deposition Program. [cited 2010 August 26]; Available from http://nadp.sws.uiuc.edu/.
18. Gianessi, L. and N. Reigner, Pesticide Use in U.S. Crop Production: 2002. Insecticides & Other Pesticides, 2006, CropLife Foundation: Washington, DC.
19. United States Department of Agriculture (USDA). 2002 Census of Agriculture full report. 2002 [cited 2010 August 26]; Available from http://www.agcensus.usda.gov/Publications/2002/index.asp.
20. United States Environmental Protection Agency (EPA). EPA Geospatial Data Download Service. 2013 [cited 2013 September 10]; Available from http://www.epa.gov/envirofw/geo_data.html.
21. United States Environmental Protection Agency (EPA). Superfund National Priorities List (NPL) Sites. 2010; Available from http://www.epa.gov/superfund/sites/npl/index.htm.
22. United States Environmental Protection Agency (EPA). Resource Conservation and Recovery Act (RCRA) Large Quantity Generators (LQG). 2010 [cited 2010 August 26]; Available from http://www.epa.gov/osw/hazard/generation/lqg.htm.
23. United States Environmental Protection Agency (EPA). Toxics Release Inventory (TRI) Sites. 2010 [cited 2010 August 26]; Available from http://www.epa.gov/tri/.
24. United States Environmental Protection Agency (EPA). Resource Conservation and Recovery Act (RCRA) Treatment, Storage, and Disposal Facilities (TSD) and (RCRA) Corrective Action Facilities. 2010 [cited 2010 August 26]; Available from http://www.epa.gov/osw/hazard/tsd/index.htm.

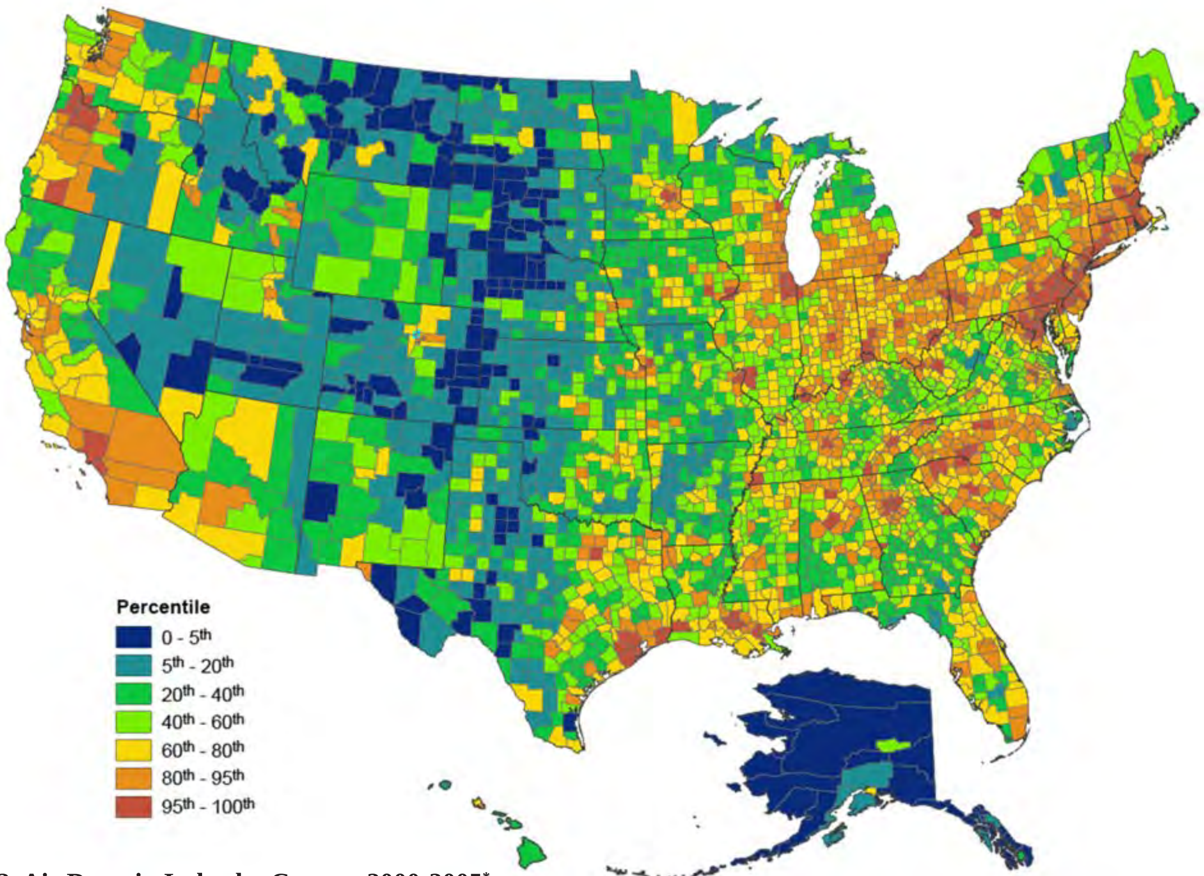
25. United States Environmental Protection Agency (EPA). Assessment, Cleanup, and Redevelopment Exchange (ACRES) Brownfield Sites. 2010 [cited 2010 August 26]; Available from http://www.epa.gov/brownfields/.
26. United States Environmental Protection Agency (EPA). Section Seven Tracking System (SSTS) Pesticide Producing Site Locations. 2010 [cited 2010 August 26]; Available from http://www.epa.gov/compliance/data/systems/toxics/sstsys.html.
27. United States Geologic Services (USGS). National geochemical survey. 2010 [cited 2010 August 26]; Available from http://tin.er.usgs.gov/geochem/doc/averages/countydata.htm.
28. United States Environmental Protection Agency (EPA). Map of radon zones. 2010 [cited 2010 August 26]; Available from http://www.epa.gov/radon/zonemap.html.
29. United States Census Bureau. 2000; Available from (http://factfinder.census.gov).
30. Federal Bureau of Investigation (FBI). Uniform Crime Reports. 2010 [cited 2010 August 26]; Available from http://www.fbi.gov/ucr/ucr.htm.
31. Dun and Bradstreet. Dun and Bradstreet Products. 2010 [cited 2010 August 26]; Available from http://www.dnb.com/us/dbproducts/product_overview/index.html.
32. United States Census Bureau. Topologically Integrated Geographic Encoding and Referencing. 2010 [cited 2010 August 26]; Available from http://www.census.gov/geo/www/tiger/.
33. National Highway Traffic Safety Administration (NHTSA), N.C.f.S.a.A.N. Fatality Analysis Reporting System (FARS). 2010 [cited 2010 August 26]; Available from http://www.nhtsa.gov/people/ncsa/fars.html.
34. United States Department of Housing and Urban Development. Multifamily Assistance and Section 8 Contracts Database. [cited 2012 November 28]; Available from http://portal.hud.gov/hudportal/HUD?src=/program_offices/housing/mfh/exp/mfhdiscl.
35. Langlois, P.H., et al., Occurrence of conotruncal heart birth defects in Texas: a comparison of urban/rural classifications. J Rural Health, 2010. 26(2): p. 164-74.
36. Messer, L.C., et al., Urban-rural residence and the occurrence of cleft lip and cleft palate in Texas, 1999-2003. Ann Epidemiol, 2010. 20(1): p. 32-9.
37. Langlois, P.H., et al., Urban versus rural residence and occurrence of septal heart defects in Texas. Birth Defects Res A Clin Mol Teratol, 2009. 85(9): p. 764-72.
38. Luben, T.J., et al., Urban-rural residence and the occurrence of neural tube defects in Texas, 1999-2003. Health Place, 2009. 15(3): p. 848-54.
39. Emerson, J., et al., 2012 Environmental Performance Index and Pilot Trend Environmental Performance Index - Full Report, Yale Center for Environmental Law and Policy, Editor 2012, Yale University, Columbia University.
40. Messer, L.C., et al., The development of a standardized neighborhood deprivation index. J Urban Health, 2006. 83(6): p. 1041-62.
41. AirNow. Air Quality Index. [cited 2013 August1]; Available from http://www.airnow.gov/?action=aqibasics.aqi.
42. San Francisco Department of Public Health. Pedestrian Environmental Quality Index. [cited 2013 August 1]; Available from http://www.sfphes.org/elements/24-elements/tools/106-pedestrian-environmental-quality-index.
43. Huang, G. and J. London, Cumulative environmental vulnerability and environmental justice in California's San Joaquin Valley. Int J Environ Res Public Health, 2012.
44. Reid, C., et al., Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study. Environ Health Perspect, 2012. 120: p. 715-720.
45. California Environmental Protection Agency, California Communities Environmental Health Screening Tool, Version 1 (CalEnviroScreen 1.0), 2013.
46. Studies, V.C.f.E. Virginia Environmental Quality Index. [cited 2013 August 9]; Available from http://www.veqi.vcu.edu/.
47. EVI Official Global Website. The Environmental Vulnerability Index. 2013 [cited 2013 August 1]; Available from http://www.vulnerabilityindex.net/.

Appendix I:
County Maps of Environmental Quality Index

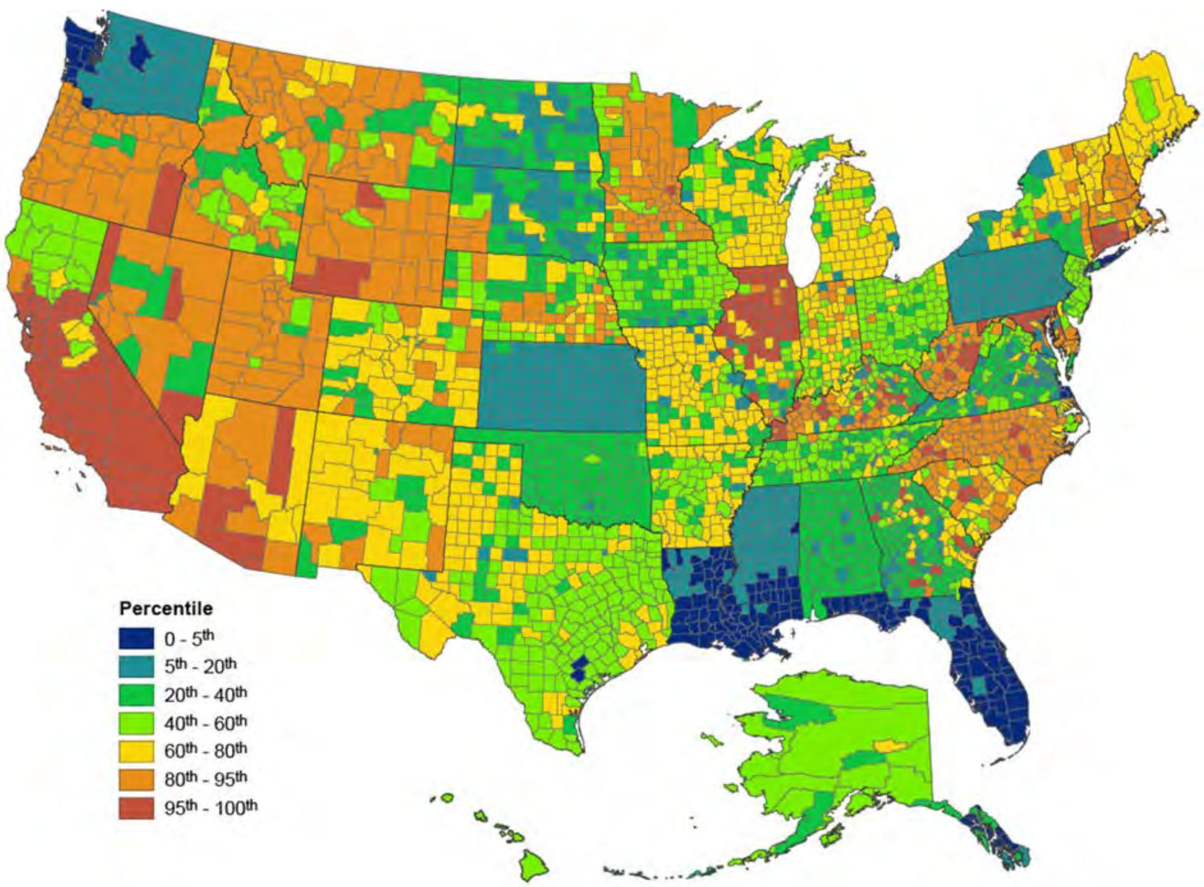


Map 1. Environmental Quality Index by County, 2000-2005.*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality

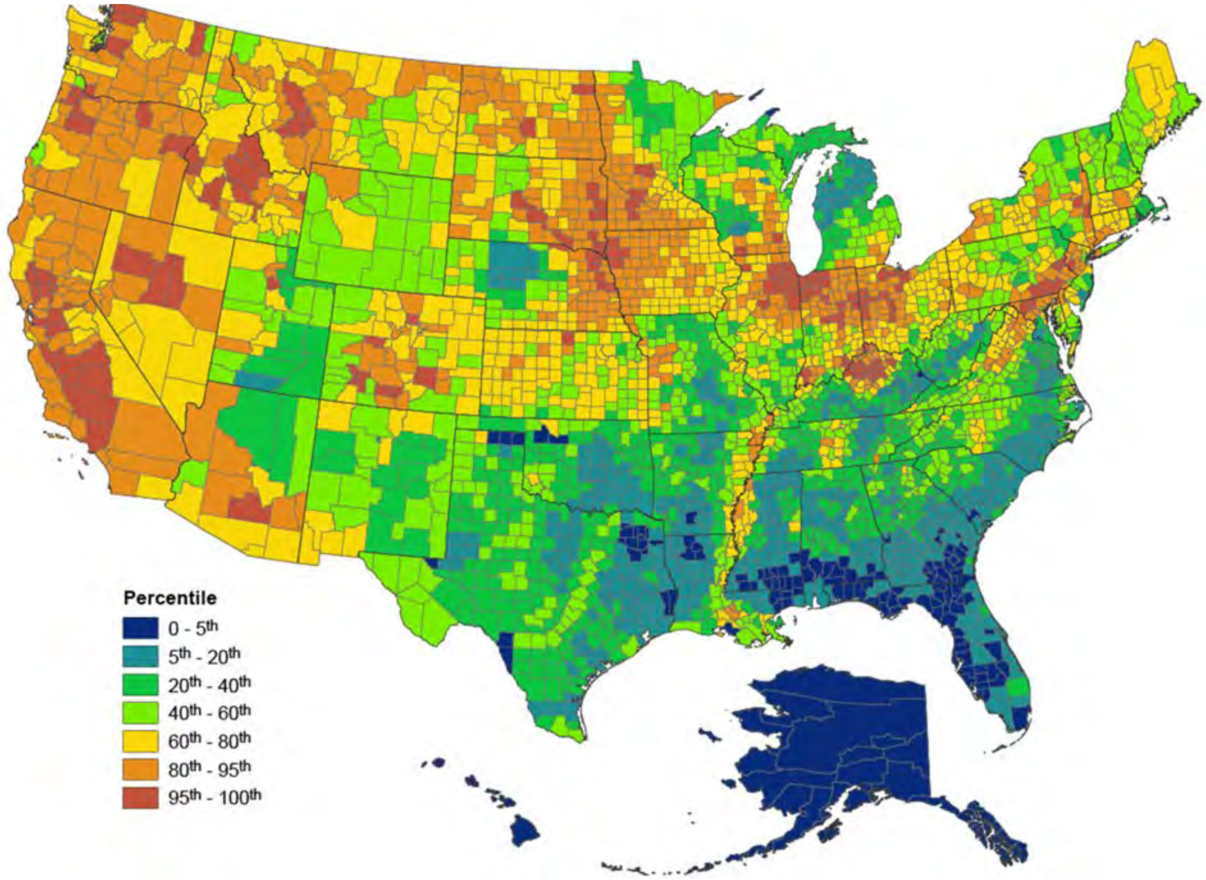


Map 2. Air Domain Index by County, 2000-2005*

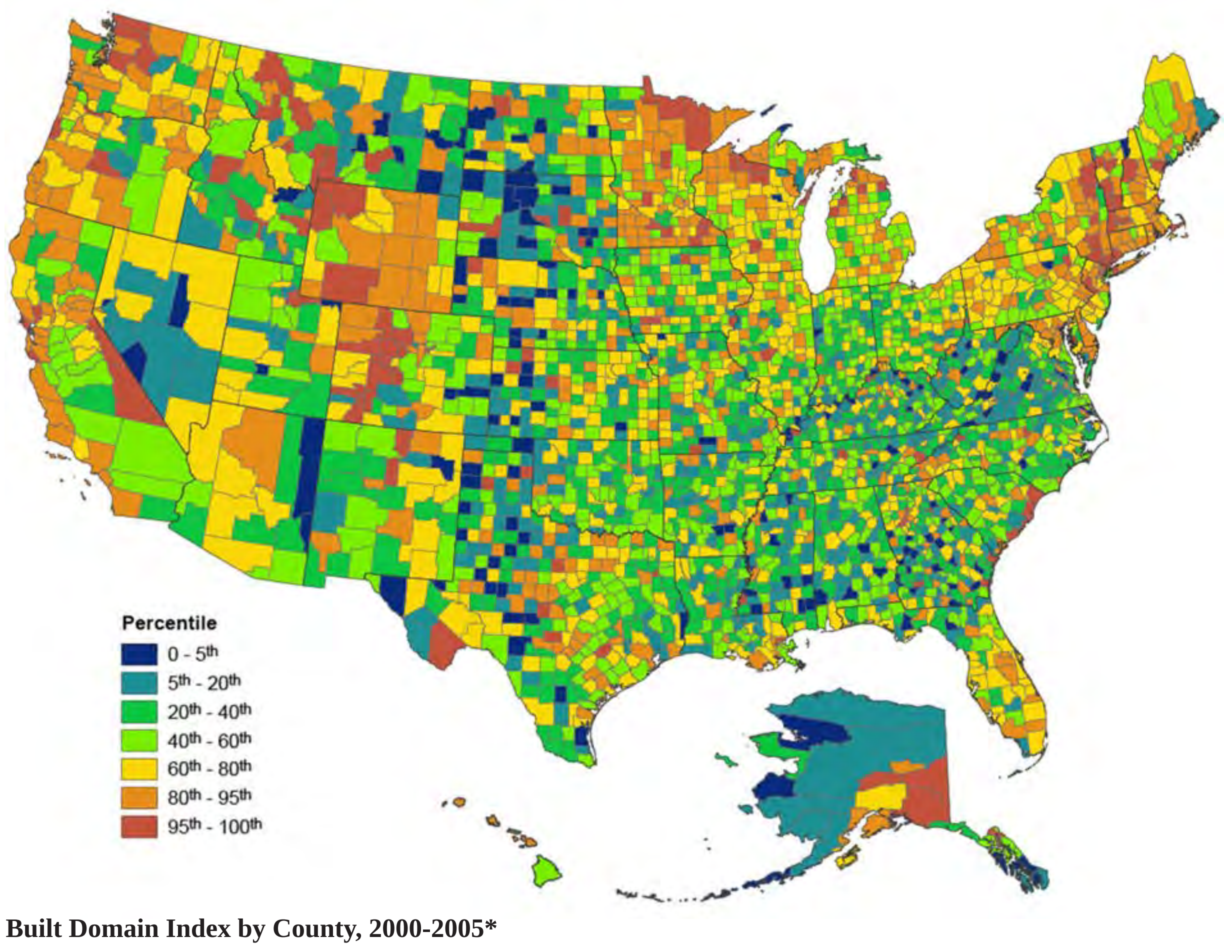


Map 3. Water Domain Index by County, 2000-2005*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality

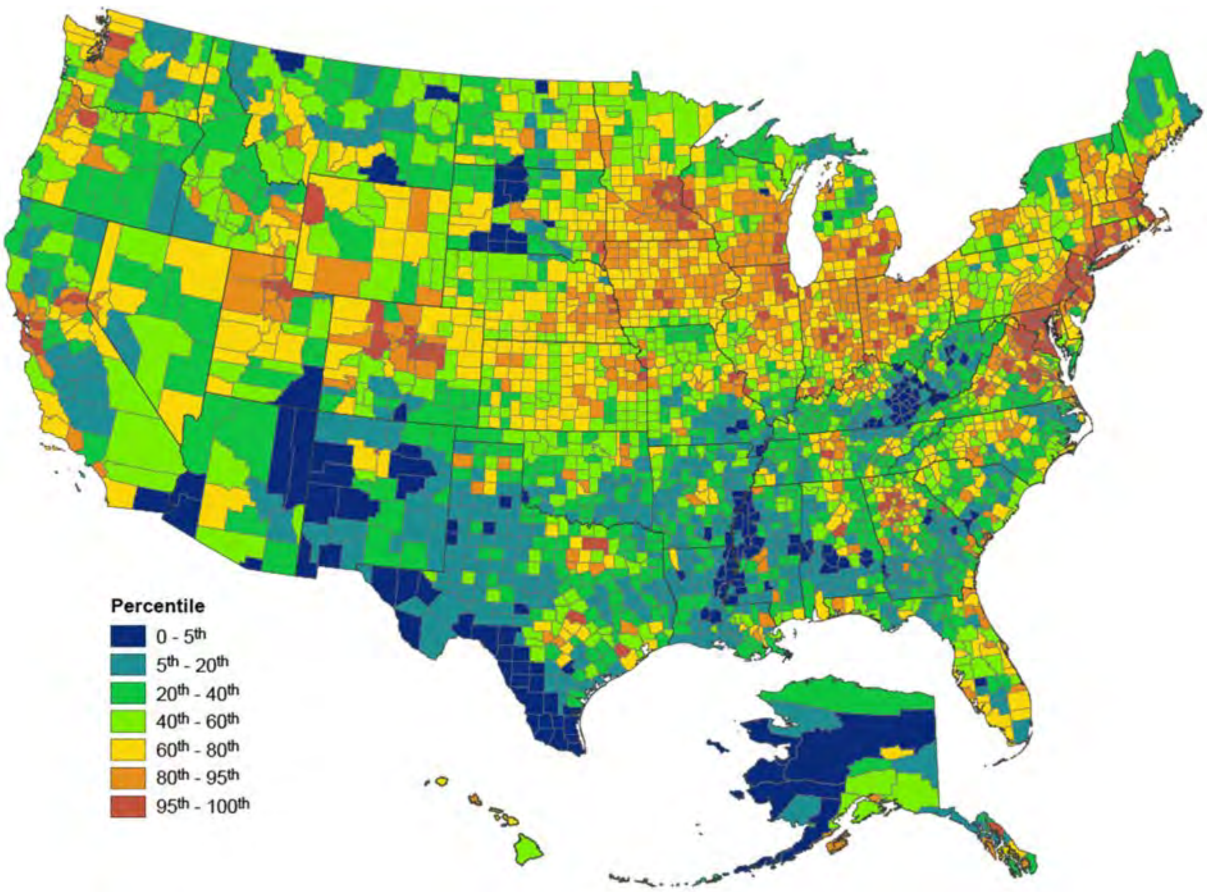


Map 4. Land Domain Index by County, 2000-2005*



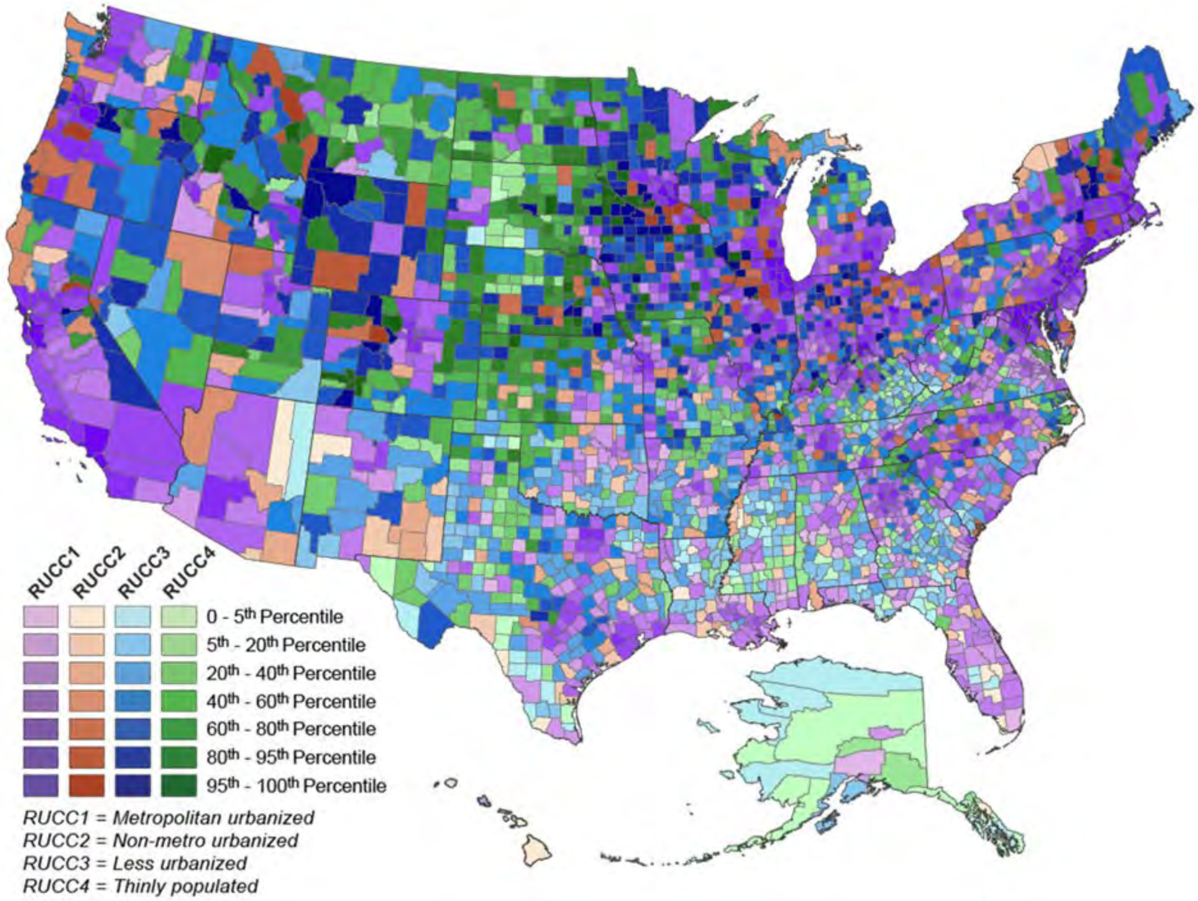
Map 5. Built Domain Index by County, 2000-2005*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality

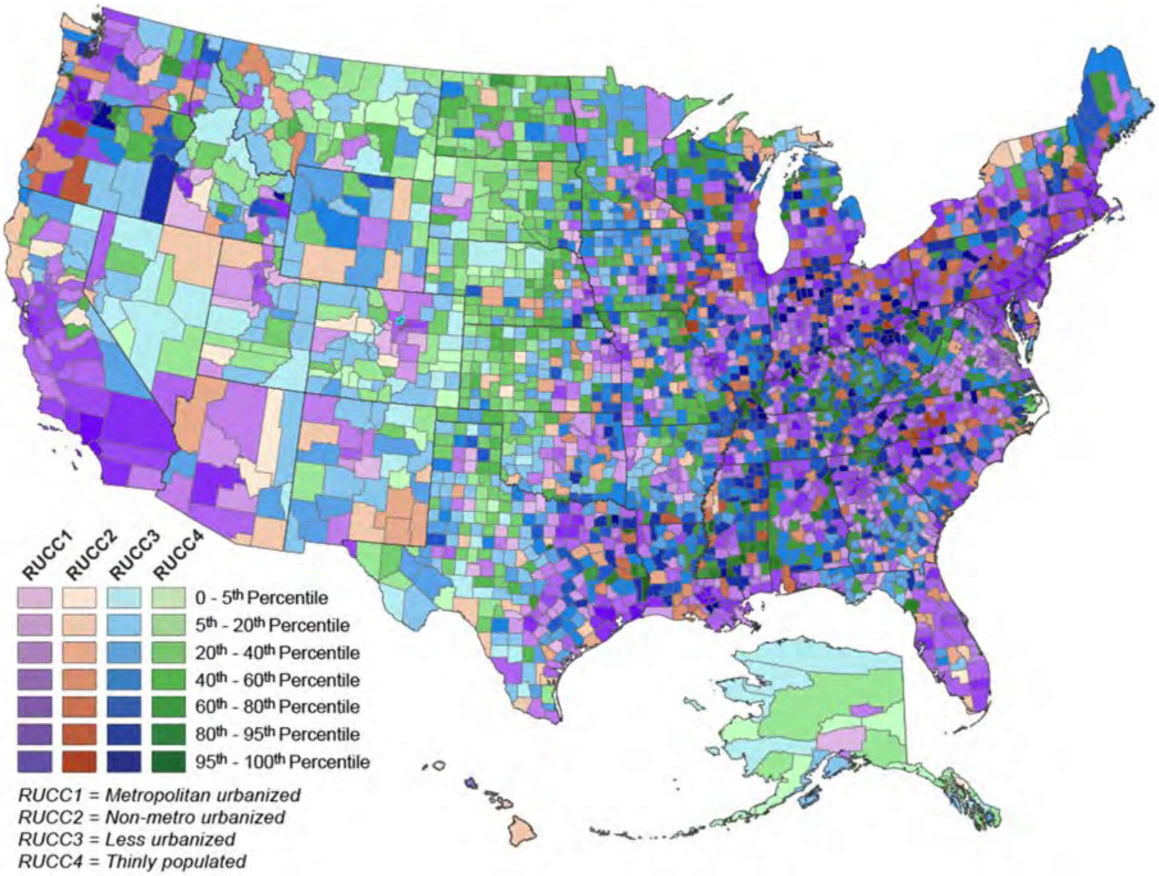


Map 6. Sociodemographic Domain Index by County, 2000-2005*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality

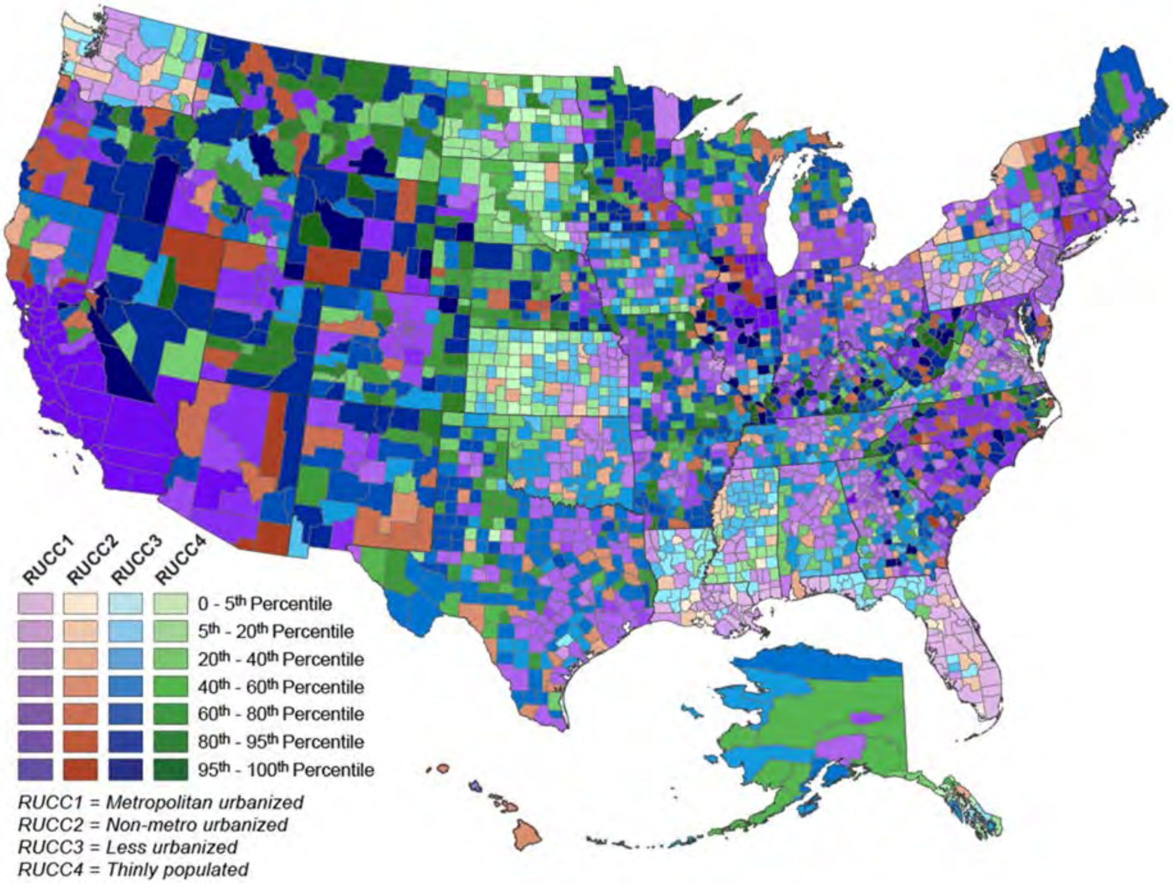


Map 7. Environmental Quality Index Stratified by Rural-Urban Continuum Codes by County, 2000-2005*

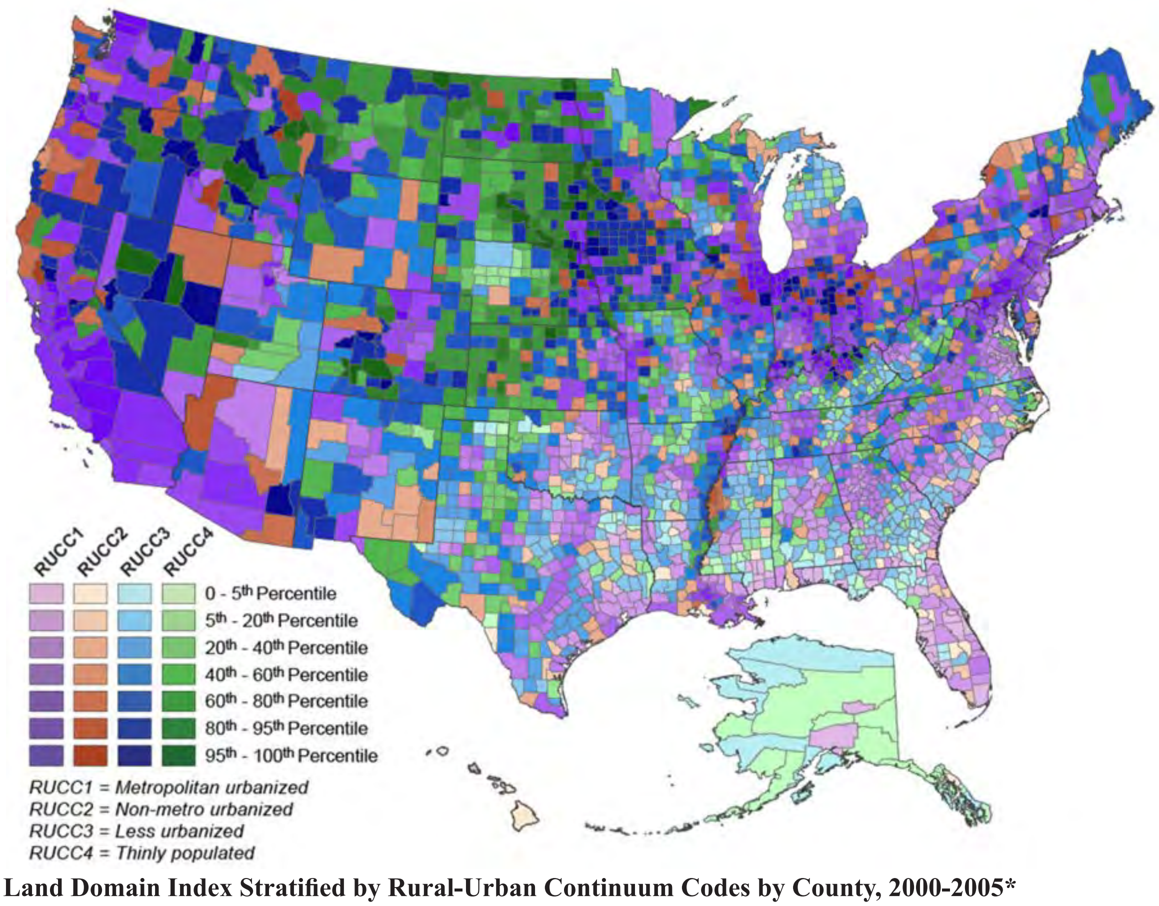


Map 8. Air Domain Index Stratified by Rural Urban Continuum Codes by County, 2000-2005*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality

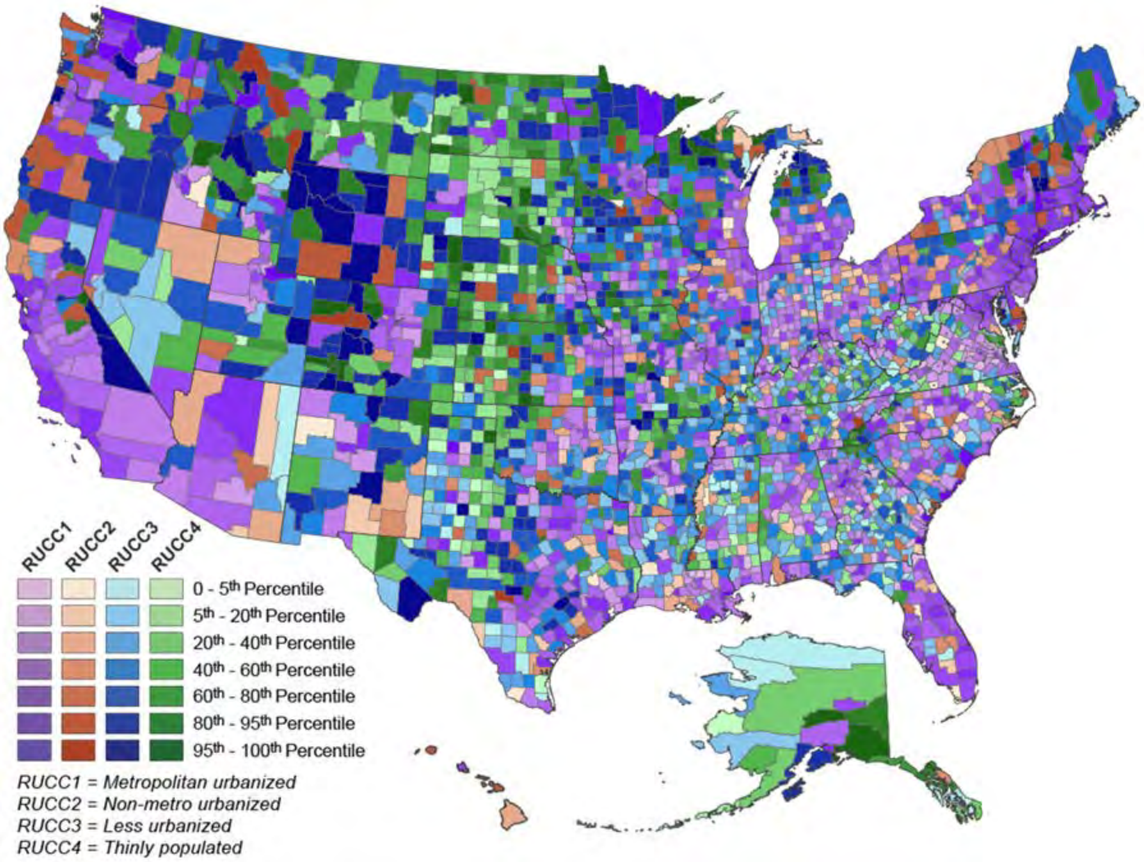


Map 9. Water Domain Index Stratified by Rural-Urban Continuum Codes by County, 2000-2005*

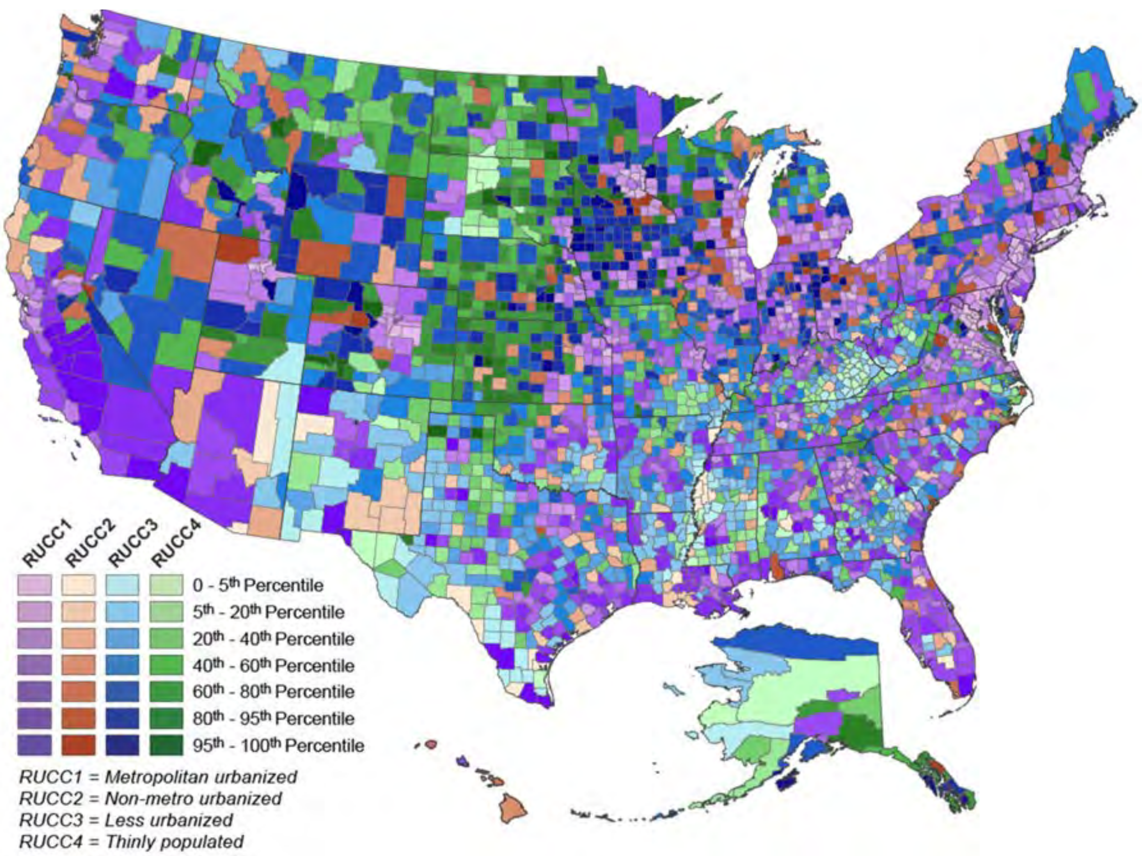


Map 10. Land Domain Index Stratified by Rural-Urban Continuum Codes by County, 2000-2005*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality



Map 11. Built Domain Index Stratified by Rural-Urban Continuum Codes by County, 2000-2005*



Map 12. Sociodemographic Domain Index Stratified by Rural-Urban Continuum Codes by County, 2000-2005*

* Higher EQI values suggest worse environmental quality, and lower EQI values suggest better environmental quality

Appendix II: Quality Assurance

The approved National Health and Environmental Effects Research Laboratory (NHEERL) Environmental Public Health Division (EPHD) Intramural Research Protocol for this project is “Creating an Overall Environmental Quality Index,” with Document Control Number IRP-NHEERL/HSD/EBB/DL/2008-01r1. An internal EPA review of this report was conducted in August 2003 by Lisa Smith, NHEERL Gulf Ecology Division; Jane Gallagher, NHEERL EPHD), and Tom Brody (Region 5). An external peer review was conducted in July 2014 by Angel Hsu, Yale University, School of Forestry and Environmental Studies; Paul D. Juarez, University of Tennessee Health Science Center, Department of Preventive Medicine; and Peter H. Langlois, Texas Department of State Health Services, Birth Defects Epidemiology and Surveillance Branch.

The data sources used to create the EQI and the criteria used to select the data sources are mentioned in Creating an Overall Environmental Quality Index, Technical Report (Technical Document), in Part II: Data Source Identification and Review. Additional information about the sources can be found in the Technical Document in Appendix I and Appendix II. Table 1 in this report provides the strengths and limitations of the sources used in the EQI.

Information about uses of the EQI, as well as strengths and limitations of the EQI, is located in the Discussion of this report.



SCIENCE

United States
Environmental Protection
Agency

Office of Research and Development (8101R)
Washington, DC 20460

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT NO. G-35

Official Business
Penalty for Private Use
$300

Recycled/Recyclable Printed on paper that contains a minimum of 50% postconsumer fiber content processed chlorine free