Gone with the Wind:
Federalism and the Strategic Location of Air Polluters

James E. Monogan III, David M. Konisky, & Neal D. Woods

Codebook

Here we present a description of the variables from the two principal data files we analyze. The first file (`majorAirAllDist.csv`) consists of 36,972 pooled sites of air and hazardous waste polluters; this file is called by `lower48control.R` to reproduce our primary sets of results in the print article. The second file (`anyPowerPlantDist.csv`) consists of 26,034 pooled sites of power plants (all air polluters) and hazardous waste polluters; this file is called by `powerPlantsModel.R` to reproduce additional appendix results that subset power plants based on when they were sited. Each bullet point represents one of the two data sets, and variables are listed below.

- **File: majorAirAllDist.csv.** Locations of major air polluters and large quantity generators of hazardous solid waste.

  fips: FIPS code for the county a site is in.

  loc_state: Two-letter postal abbreviation for the state the facility is located in.

  AFSalt: ID variable for each facility. Based on the Air Facility System database ID.

  reg_id: EPA regulated site ID number for the facility.

  fac_name: Name of the facility.

  airtotal: Total pounds of toxic air emissions in 2010.

  air: Indicator variable coded 1 if the site is a major air polluter, and 0 if the state is a non-air polluting large quantity generator of hazardous waste.

  tsdf: Indicator variable coded 1 if the site is a hazardous waste treatment, storage, and disposal facility, and 0 otherwise.

  lqg: Indicator variable coded 1 if the site is a large quantity generator of hazardous waste, and 0 otherwise.

  duplicate: Residual variable from the data management program. (All values are coded 0 or missing.)

  longitude: Longitude coordinate for the facility's location.

latitude: Latitude coordinate for the facility's location.

eastings: Coordinate for the facility's location in eastings, in meters, based on ESRI projection 102003 (USA contiguous Albers equal area conic) using the North American Datum of 1983 (NAD83).

northings: Coordinate for the facility's location in northings, in meters, based on ESRI projection 102003 (USA contiguous Albers equal area conic) using the North American Datum of 1983 (NAD83).

angle.full: Forecast of the prevailing wind direction at the site, in radians, based on the kriging model fit over the 299 weather stations.

angle.mod: First component of the wind direction forecast, in radians, based on a quadratic latitude-and-longitude referenced trend term.

angle.error: Second component of the wind direction forecast, in radians, based on weighted residuals from the trend model at the 299 weather stations, with closer weather stations receiving more weight.

down: Distance from the site to the downwind border in kilometers.

sDown: Scaled distance to the downwind border, with values ranging from 0 at the border to 1 at the site farthest from the border.

up: Distance from the site to the upwind border in kilometers.

sUp: Scaled distance to the upwind border, with values ranging from 0 at the border to 1 at the site farthest from the border.

east: Distance from the site to the due east border in kilometers.

sEast: Scaled distance to the due east border, with values ranging from 0 at the border to 1 at the site farthest from the border.

west: Distance from the site to the due west border in kilometers.

sWest: Scaled distance to the due west border, with values ranging from 0 at the border to 1 at the site farthest from the border.

SIC_CODES: Standard industrial classification (SIC) codes for the facility. Variable is in string format, and one facility may have multiple codes.

sic1: The first-listed SIC code for the site, in integer format.

power: Indicator coded 1 if the facility is a power plant, 0 otherwise.

state: Full name of the state in which a site is located.

statenum: Index of the state in alphabetical order. (I.e., 1=Alabama, 50=Wyoming.)

gi_tot: Green index based on 67 separate indicators of state environmental policy effort (Hall & Kerr 1991).

gi_net: Hall & Kerr's green index, removing measures of environmental primary assumption.

ed_locational: Saiz's measure of locational economic development based on a content analysis of 2756 state-level programs (Saiz & Clarke 2013).

env_ig: Gray and Lowery's (2001) data on the number of registered environmental lobbyists. This measure averages the 1990 and 1997 data.

scaledIG: Gray and Lowery's environmental interest group data, scaled by the state's Gross State Product.

new: Indicator variable coded 1 if the facility was constructed after 2005, 0 otherwise.

public: Indicator variable coded 1if the facility is operated by any level of government, 0 otherwise.

windSpeed: Average wind speed in the state a site is located in.

county: Name of the county a site is located in.

popDens: Population density of the county a site is located in.

- **File: anyPowerPlantDist.csv.** Locations of power plants and large quantity generators of hazardous solid waste.

fac_name: Name of the facility.

loc_state: Two-letter postal abbreviation for the state the facility is located in.

reg_id: EPA regulated site ID number for the facility.

online_year: Year that a power plant went online.

air: Indicator variable coded 1 if the site is a power plant (and hence a major air polluter), and 0 if the state is a non-air polluting large quantity generator of hazardous waste.

tsdf: Indicator variable coded 1 if the site is a hazardous waste treatment, storage, and disposal facility, and 0 otherwise.

lqg: Indicator variable coded 1 if the site is a large quantity generator of hazardous waste, and 0 otherwise.

duplicate: Residual variable from the data management program. (All values are coded 0 or missing.)

longitude: Longitude coordinate for the facility's location.

latitude: Latitude coordinate for the facility's location.

eastings: Coordinate for the facility's location in eastings, in meters, based on ESRI projection 102003 (USA contiguous Albers equal area conic) using the North American Datum of 1983 (NAD83).

northings: Coordinate for the facility's location in northings, in meters, based on ESRI projection 102003 (USA contiguous Albers equal area conic) using the North American Datum of 1983 (NAD83).

angle.full: Forecast of the prevailing wind direction at the site, in radians, based on the kriging model fit over the 299 weather stations.

angle.mod: First component of the wind direction forecast, in radians, based on a quadratic latitude-and-longitude referenced trend term.

angle.error: Second component of the wind direction forecast, in radians, based on weighted residuals from the trend model at the 299 weather stations, with closer weather stations receiving more weight.

down: Distance from the site to the downwind border in kilometers.

sDown: Scaled distance to the downwind border, with values ranging from 0 at the border to 1 at the site farthest from the border.