**DS421 RET DATA AVAILABILITY REPORT  
August 5th, 2019  
Katherine Wolf**

**QUESTIONS**

* **Years?** How to deal with time? I’ve included available years with the variables.
* **Different thresholds for mandatory criteria pollutant reporting by air district in CA** (check for facilities reporting less than 10 tons of a criteria pollutant in a year and exclude those from the dataset?)
* **Mobile sources? Wind/topography/temperature/land-use factors?**
  + Increase complexity exponentially
  + Without them, incomplete picture of air pollution exposure
  + Land-use-regression models would be the extreme version
* **Private schools?**
  + We could have enrollment numbers and race poverty breakdowns for private schools as well, albeit from a different source
  + Private schools could include students whose families live out of state (if we’re using schools as a proxy for population characteristics)
  + Public schools overrepresent some groups without them (low-income and non-white students)?
  + No poverty metric data available for private schools, to my knowledge
* **Model-building help**
* **Can also bump questions to Rachel’s lab on Thursday (1:30 to 3 p.m.)**

**MODEL BEGINNINGS**

**VARIABLES WE CAN HAVE**

* Total emissions of pollutant at point source
* Distance between student at school and point source
* Distance between student at school and operating monitor
* Free (or free and reduced-price? or run both ways?) lunch status counts of students by school (proxy for poverty) (public schools only)
* Race/ethnicity counts of students by grade (public only—private has counts by school)
* Measured ambient pollutant concentrations at the operating monitors. Not sure how to incorporate these, but we do have them.

**HYPOTHESES-ISH**

* **Null-ish hypothesis:** In an ideal, just world, proximity to an emissions source should correlate with proximity to an emissions monitor, i.e., a student with higher exposure risk for a pollutant (here, roughly, some combined weight of proximity to potential emissions sources and total emissions from those sources) should be closer to a monitor for that pollutant than a student with less risk, regardless of race or poverty status.
* **Alternative hypothesis:** Students of non-white race/ethnicity and/or in poverty are closer to emissions sources and farther from monitors than white and/or not-in-poverty students.

**NOTES**

* PROBLEM: Some factors definitely considered in monitor siting and operation decisions (history of being in compliance with air pollution regulations, emissions from mobile sources, etc.) might not be in the model if we go the simple route. Do these matter? Or do we just care about where students ended up in relation to monitors and pollution sources?
* THOUGHT: Private versus public school-dom might also be interesting as its own distinction.

**DATA SOURCES**

**SCHOOLS**

* Enrollment
  + Public: California Department of Education
    - Variables:
      * Race/ethnicity student counts by gender by grade  
        <https://www.cde.ca.gov/ds/sd/sd/filesenr.asp>
        + Races:

Code 0 = Not reported

Code 1 = American Indian/Alaska Native, not Hispanic

Code 2 = Asian, not Hispanic

Code 3 = Pacific Islander, Not Hispanic

Code 4 = Filipino, Not Hispanic

Code 5 = Hispanic or Latino

Code 6 = African American, not Hispanic

Code 7 = White, not Hispanic

Code 9 = Two or More Races, Not Hispanic

* + - * Free and free- or reduced-price lunch-eligible student counts and percentages by school (not by race, gender, or grade)  
        <https://www.cde.ca.gov/ds/sd/sd/fsspfrpm.asp>
    - Frequency: 1984–85 (enrollment), 1988 (lunch data) to 2018–19
  + Private: National Center for Education Statistics  
    <https://nces.ed.gov/ccd/elsi/>
    - Variables:
      * Enrollment counts by grade OR race (not joint)
        + American Indian/Alaska Native
        + Asian/Pacific Islander [sic]
        + Hispanic
        + Black
        + White
        + Hawaiian Native/Pacific Islander [sic]
        + Two or More Races
      * No school lunch data
    - Odd-numbered years only (1997–98 through 2015–16)
* Locations
  + School campus GIS database from <http://www.californiaschoolcampusdatabase.org/> (might have better location estimates than the state-provided longitudes and latitudes, but only includes public schools)

**AIR POLLUTION EMISSIONS**

* California Air Resources Board
* Stationary point sources

<https://ww3.arb.ca.gov/ei/ei.htm> (general emissions inventory page)

<https://www.arb.ca.gov/app/emsinv/facinfo/facinfo.php> (stationary facility data page)

<https://ww3.arb.ca.gov/ei/tools/pollution_map/#dataTab> (mapping tool for 2008 onward)

* + Variables and years
    - Criteria pollutants (PM2.5, PM10, lead, ozone, CO, NOx, SOx): 1987, 1990, 1993, 1995–2017 (2008–2017 using mapping tool)
    - Toxics/hazardous air pollutants (kinda everything else): 1996–2017 (2008 onward using mapping tool for select pollutants)
  + Reporting inconsistencies
    - Regulatory changes over time
      * 1990 (federal Clean Air Act amendments)
      * 2011 federal reporting requirements change for PM requiring condensable particulate matter reporting (resulting in higher PM levels reported under same actual emissions levels)
      * 2015 Federal Air Emission Reporting Requirements changed, effective 2016 (?)
    - Pollutant-specific issues
      * Facilities required to report criteria pollutant emissions if they emit 10 tons or more per year of a single criteria pollutant in California (reporting changed for PM2.5/PM10 in 2011) as of July 1989, BUT
      * Individual CA air districts can decide to use even lower reporting thresholds, so those districts will show more facilities emitting/higher emissions than districts with higher thresholds (@#$%!). ARB 617 law implementation should fix this discrepancy in the next year or two, but not in time for this study.
      * Federal toxics reporting requirements can change yearly.
* Mobile sources?
  + <https://ww3.arb.ca.gov/msei/categories.htm>
  + Example: EMFAC models for on-road sources
    - 2000 onward
    - Estimates motor vehicle pollutant emissions at city-ish level as function of vehicle emissions rates and estimated vehicle activity.

**AIR POLLUTION MONITORS**

* Criteria pollutant monitor sites, concentrations, and operating dates:

<https://www.arb.ca.gov/aqmis2/aqdselect.php?tab=specialrpt>

* Toxics pollutant monitor sites, concentrations, and operating dates:  
  <https://www.arb.ca.gov/adam/toxics/toxics.html>
* Daily averages from 1987 at least onward