# Comparison of interpolation methods for estimating spatially aggregated pollution exposure

Ekaterina I Galkina*

University of Massachusetts, Lowell

Georges G Grinstein†

University of Massachusetts, Lowell

## ABSTRACT

Growing evidence points to serious medical conditions associated with airborne pollutant exposure, including neurologically harmful effects, increases in hospitalizations and mortality due to cardiovascular and respiratory diseases. The US Environmental Protection Agency (EPA) regularly monitors levels of lead, ozone, particulates and other air pollution. However, monitor sites are unequally distributed, necessitating spatial interpolation of pollutant concentrations at poorly sampled locations. Our study focuses on epidemiological data aggregated to only the first 3-digits of the postal ZIP code. Consequently, the varying land area coverage of the information-sensitive residential geographies creates an additional exposure assessment challenge. We compared four common interpolation methods for predicting particulate matter concentrations in the state of California alone, including kernel smoothing, inverse distance weighting (IDW), Voronoi partitioning and kriging. We show that methods that produce a prediction standard error map are the most reliable (i.e. kriging) and that a consensus in the prediction maps can be reached only at the smaller geographical units, i.e. county-level.

**Keywords**: air quality estimation, epidemiology, health, particulate matter, spatial interpolation.

## 1 INTRODUCTION

Air pollution comprises a mixture of particulate matter (PM), organic compounds, metals, and gases. Rises in atmospheric pollution are believed to be responsible for serious health issues and increased hospitalizations due to lung and heart associated diseases, with sometimes fatal consequences. Calderon-Garciduenas et al. showed an association between long-term air pollution exposure and neuroinflammation, altered immune response and central nervous system disease in children and young adults [1]. Through autopsy results, it was discovered that ultrafine PM (<100nm) is able to pass through the blood-brain barrier and translocate into olfactory bulbs and frontal cortex. Motivated by these findings, we aimed to explore the role of air pollutants as a potential accelerator of dementia, a goal which requires an assessment of the study population's level of exposure to PM.

In this case, the study sample included volunteers from the National Alzheimer's Coordinating Center (NACC). In order to protect patient confidentiality, the NACC discloses only the first three digits of the participants' residential ZIP code. To assess the participants' pollution exposure levels, we used existing PM levels, as monitored by the Environmental Protection Agency (EPA), and interpolated levels for areas not directly monitored. In order to determine the best interpolation method, we evaluated several common techniques on a single state. For this purpose, California was chosen due to its large area, environmental diversity (i.e. hills and valleys) high monitor density, and a vast range of pollution levels, as projected by the EPA.

* ekaterina_galkina@student.uml.edu
† grinstein@cs.uml.edu

## 2 APPROACH

### 2.1 Data

The current freeze of the NACC database contains over 23,000 subjects from 34 past and present Alzheimer's Disease Centers (ADCs) nationwide recruited between 2005 and 2014. The participants have a range of diagnoses from cognitively normal to mild cognitive impairment - the prodromal stage of Alzheimer's disease (AD), to possible or probable AD.

Annual 24-hour summary estimates for PM10 (particles less than 10 $\mu$m in diameter) were obtained from the EPA website for 2005 – the year corresponding to the subjects' earliest enrollment date. A point shapefile was created from the EPA monitors' latitude and longitude coordinates using Quantum GIS v1.8.0-Lisboa software (Fig. 1).
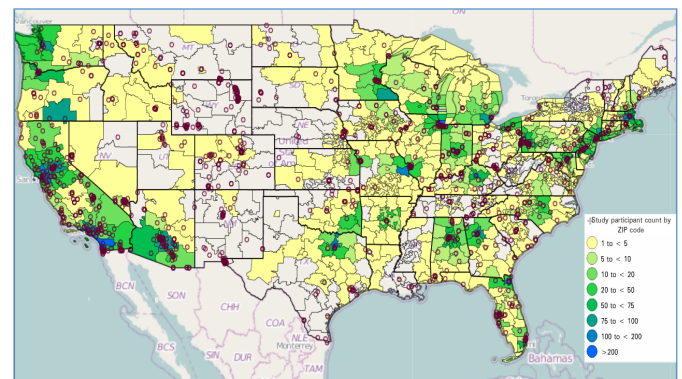


Figure 1: Distribution of study participants by 3-digit residential ZIP code (increasing count from yellow to blue) in relation to PM monitors for 2005 (purple circles). Created using Weave open source web software (iweave.org).

### 2.2 Interpolation Methods

There are two main types of interpolation techniques – deterministic and geostatistical. Deterministic interpolation techniques assign values to locations based on either the extent of similarity or the degree of smoothing. Geostatistical interpolation techniques account for the spatial configuration of the sample points around the prediction location. We compared four different spatial interpolation methods for effectiveness in estimating PM10 concentrations at unsampled locations using ArcGIS v10.2.1 software: kernel smoothing (deterministic), Inverse Distance Weighting (deterministic), Voronoi partitioning (deterministic), and kriging (geostatistical).

## 3 RESULTS

There were a total of 147 PM10 measurements for California EPA monitors for year 2005 after filtering on duplicate locations. The mean PM10 concentration was $25.2 \pm 13.5$. Interpolation results are shown in Figs. 2a-c and 3. Fig. 3 shows how the Voronoi algorithm behaves when a high PM outlier is next to a low PM value; the abrupt change in prediction surface is not seen with other methods.
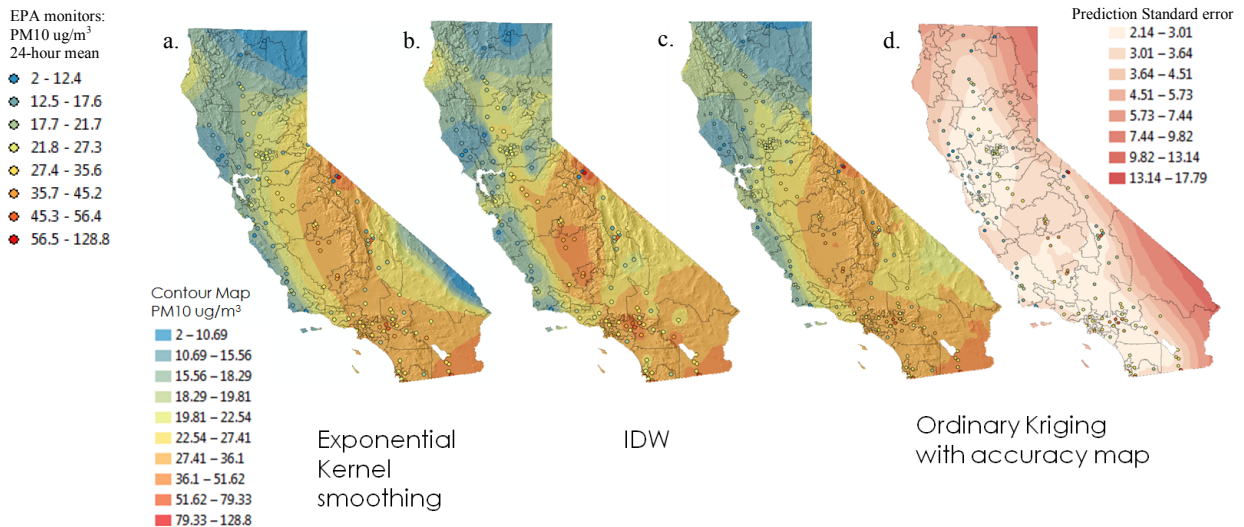
Figure 1: Comparison of ArcGIS interpolation results: a) exponential kernel smoothing, b) Inverse Distance Weighting, c) kriging with d) prediction accuracy map. Borders shown are 3-digit ZIP code boundaries.
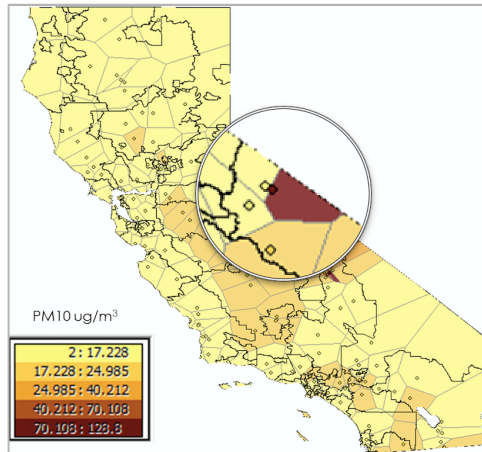


Figure 2: Voronoi partitioning results where an abrupt change in predicted concentration is seen in a region with a high PM outlier (ArcGIS).

## 3.1 Comparison to partial ZIP code boundaries

We examined the Los Angeles county area separately in order to assess the variability of interpolation results and determine if a consensus could be reached (Fig. 4). Indeed, partial ZIP codes encompassing smaller areas showed much more consistent concentrations throughout than larger areas.
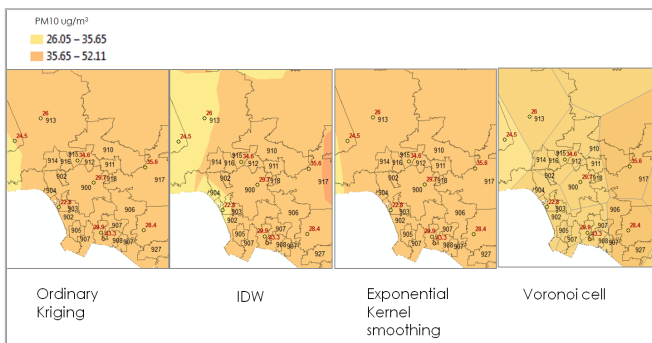


Figure 3: Comparison of the 4 interpolation methods for Los Angeles county to determine whether there is a consensus in the interpolation for smaller geographical regions. 3-digit ZIP codes are indicated in black and PM estimates in red (ArcGIS).

## 4 CONCLUSIONS AND FUTURE WORK

An accurate assessment of a population's pollution exposure is key in correctly determining environmental associations with health outcomes. AD affects 5.5 million people in the US and is the type of disease that takes decades to manifest. Not having found a single genetic cause, researchers turned to studying environmental influences spanning a person's lifetime. The NACC database of AD subjects captures some of these non-genetic risk factors, but lacks air quality estimates. Spatial interpolation, sometimes employed in environmental studies [2], was necessary to define pollutant concentrations at unsampled locations for an unprecedented level of geographical detail – 3-digit ZIP code boundaries.

The aim of the analysis was two-fold: 1) to interpolate air quality estimates of PM10 at poorly sampled locations and 2) to assess how well the estimates overlap with 3-digit ZIP code boundaries. Determination of the best interpolation method was thus based on 1) prediction accuracy, 2) smoothness of the prediction surface, and 3) ability to infer air quality estimates at partial ZIP codes.

Four common interpolation methods were tested on the state of California: exponential kernel smoothing, IDW, kriging, and Voronoi partitioning. IDW showed the smoothest prediction surface transitions. On the contrary, Voronoi partitioning had the most abrupt concentration changes, but was easiest to interpret due to clear PM prediction boundaries created by the cells. Kriging, which incorporates autocorrelation and provides a prediction accuracy map, gives this technique added reliability. Partial residential ZIP codes encompassing smaller regions (i.e. counties) show less variability of predictions warranting the reduction of the study area.

### REFERENCES

[1] Calderon-Garciduenas, et al. (2004). Brain inflammation and Alzheimer's-like pathology in individuals exposed to severe air pollution. Toxicol Pathol, 32(6), 650-658.

[2] Wong, D. W., Yuan, L., & Perlin, S. A. (2004). Comparison of spatial interpolation methods for the estimation of air quality data. J Expo Anal Environ Epidemiol, 14(5), 404-415.