

# Development of an Automated Pesticide Exposure Analyst for the California's Central Valley

Daniel W. Goldberg<sup>1</sup>, Xinbo Zhang<sup>2</sup>, John P. Wilson<sup>3</sup>, Beate Ritz<sup>4</sup>, Myles G. Cockburn<sup>2</sup>

<sup>1</sup> University of Southern California, Department of Computer Science, GIS Research Lab

<sup>2</sup> University of Southern California, Department of Preventative Medicine

<sup>3</sup> University of Southern California, Department of Geography, GIS Research Lab

<sup>4</sup> University of California, Los Angeles, Department of Epidemiology

## Abstract

Fundamental to many environmental epidemiology studies is the assignment of environmental exposure values to an individual and their use in spatial analysis. This information is typically obtained from spatial intersection between the location of the individual and geographic extent of the environmental factor to be measured. Critical to this type of spatial association (and any results derived there from) is the existence and usability of accurate environmental data describing the environmental factor to be investigated. However, having data in hand that is completely accurate and designed, a priori, to be perfectly useable in one's study is the exception, not the rule.

In this talk, we will detail the automated methodology we have developed for associating historical pesticide exposure values to individuals in the agricultural region of California's Central Valley. In particular we will discuss the technical challenges that we had to overcome due to inaccuracies and inconsistencies in the underlying data sources used for both the estimation of the pesticides applied to an area as well as the reported land use of the area. Using a tiered approach based on classifying types of errors in the underlying data sources, we were able to create a framework that employs a "least possible error" assumption to minimize the error associated with an individual's calculated exposure values. We foresee the open discussion of our approach in this public forum being useful to other environmental modelers, epidemiologists, and spatial data users involved in similar studies that are incorporating any type of spatial data that may not be completely accurate, but still needs to be used with some level of confidence.