Kirsten Koehler
Assistant Professor
Environmental Health Sciences
Johns Hopkins School of Public Health
615 N. Wolfe Street
Baltimore, MD 21205
kirsten.koehler@jhu.edu
(410) 955-7706

Grant Number: K01OH009886

Project Title: Statistical Analyses for Assessing Space-Time Exposure Data

Project period: 09/01/2011 – 03/31/2015

## Abstract

Technological advancements in exposure assessment, a necessary component of intervention, control, and compliance, have recently increased the accuracy, reliability, and affordability of portable, direct-reading monitors. These monitors can rapidly assess worker exposures to occupational hazards. By coupling the estimated exposure with a known location, an industrial hygienist has the ability to connect exposures to specific sources. Contour plots of the hazard concentration over space, known as concentration maps, have recently been used to assess the spatial variability of hazards. Concentration maps have the potential to be powerful because they are easily comprehensible for managers, exposed employees, and occupational health scientists to locate areas of concern. Reducing or eliminating exposures in these areas will improve worker health. While we believe there is great potential for direct-reading instruments to aid in the identification and mitigation of workplace exposure hazards, it can be dangerous to apply such a methodology without understanding the uncertainties associated with this new form of exposure assessment. To date, no statistical framework has been applied to these maps. The goal of this project was to collect new hazard mapping data with high spatial and temporal resolution and to evaluate several statistical approaches for the analysis of workplace exposure data collected with direct-reading instruments.

There were three specific aims for this project. First, we employed a spatial statistical mapping method to build reliable mapped exposure estimates and compared interpolation methods using a previously collected exposure dataset. Our analysis suggested that Kriging outperformed other spatial interpolation methods and that visual inspection of the experimental variogram and modeled fit were crucial for reliable spatial estimation. Second, we collected comprehensive datasets of hazard concentration as a function of time and space for noise level and, when possible aerosol concentration exposures, at two facilities. The first facility, an engine testing facility, displayed high spatial and temporal variability, while the other, a plastics manufacturing facility, displayed high spatial variability, but low temporal variability. Third, we used the high- resolution datasets to evaluate hazard mapping approaches and data collection methodology and to develop a novel statistical approach to combine data from roving monitors (those that collect short-duration measurements at many locations) and static monitors (those that collect high temporal resolution measurements at few locations).

Analysis of the high spatial- and temporal- resolution dataset provided important insight into the collection of data for hazard mapping. Our analysis suggests that short- duration measurements, often collected for the generation of hazard maps can be misleading, resulting in the display of temporal variability as spatial variability. Because spatial correlation indoors tends to be relatively high, we recommend that in locations were temporal variability is expected to be high, collecting replicate measurements at fewer locations is preferable to collecting a single measurement at more locations when seeking to accurately represent the time-weighted average concentrations in a facility.

**Section 1**

## Key Findings

To inform decision-making by industrial hygienists and management, it is crucial that the maps generated from mapping data are as accurate and representative as possible. Because it has been unclear how many sampling locations are necessary to produce a representative hazard map, researchers typically collect as many points as can be sampled in several hours and interpolation methods are used to produce higher resolution maps. We reanalyzed hazard-mapping datasets from three industrial settings to determine which interpolation methods yielded the most accurate results. The goal was to provide practicing industrial hygienists with practical guidelines to generate accurate hazard maps with "off-the-shelf" mapping software. Kriging was found to be the optimal interpolation method. However, visually verifying the fit of the variogram model was crucial for accurate interpolation by the Kriging method. Exponential and spherical variogram models generally performed better than Gaussian models to fit the experimental variogram. It was also necessary to diverge from some of the default interpolation parameters such as the number of bins used for the experimental variogram and whether or not to allow for a nugget effect to achieve reasonable accuracy of the interpolation for some datasets.

To assess the potential bias that results from the use of short-duration measurements to represent the TWA in a hazard map, noise intensity measurements were collected at high spatial and temporal resolution in two facilities. Static monitors were distributed throughout the facility and used to capture the temporal variability at these locations. Roving monitors (typical of the hazard mapping process) captured spatial variability over multiple traverses through the facility. The differences in hazards maps generated with different sampling techniques were evaluated. Hazard maps produced from sparse, roving monitor data were in good agreement with the TWA hazard maps at the facility with low temporal variability. Estimated values were within 5 dB of the TWA over approximately 90% of the facility. However, at the facility with higher temporal variability, large differences between hazard maps were observed for different traverses through the facility. The temporal variability of noise was found to have a greater influence on map accuracy than the spatial sampling resolution.

We conducted a simulation of occupational hazard concentrations in a hypothetical facility and considered four "sampling" strategies employing higher- and lower- spatial resolutions and single measurements or the average of replicate measurements. Hazard maps were produced from the "sampled" data and compared to the time-weighted average hazard map. These simulations showed that increasing the number of replicates improved the overall comparability of the hazard map produced from the "sampled" data with the time-weighted average hazard map more efficiently than increasing the sampling spatial resolution. However, if accurately capturing peak exposures near sources is of interest, increasing the spatial resolution of the measurements, particularly near sources, was needed.

With Jun Zhu, Career Development Award Mentor, and colleagues from the University of Wisconsin and Colorado State University, we developed new spatiotemporal methodology that combined data from both roving and static monitors to generate hazard maps across space and over time in indoor environments. A novel spatiotemporal statistical model with continuous index in both space and time and model fitting procedure were developed that allowed for fusion of the two types of data in a computationally efficient manner. Our methodology was applied to a subset of the data collected at the engine testing facility and dynamic hazard maps were drawn that interpolated across space and evolved over time. This method proved superior to traditional approaches.

## Translation of Findings

This project has resulted in important contributions to the industrial hygiene (IH) community. Our publications have focused on ways to improve the data collection and spatial interpolation approaches for accurate hazard mapping. Our results suggest that, particularly at facilities with high temporal variability – a feature common at many workplaces, it is important to collect replicate data if hazard maps are intended to be representative of the time-weighted average (TWA) concentration. We recommend that a non-uniform sampling approach be used, whereby spatial resolution is increased near known sources identified by IH personnel to accurately capture the peak concentrations and lower spatial resolution be used further from known sources. By reducing the spatial resolution away from sources, resources should be used to collect multiple measurements at each sampling location, allowing for the analysis of temporal variability and to compute an average concentration for better estimation of the TWA at that location.

We have also provided recommendations for the IH community on the use of off-the-shelf mapping software and developed a new statistical methodology to produce maps of hazard intensity in time and space.

## Outcomes/Impact

This work resulted in recommendations for practicing industrial hygienists on methodology for data collection to produce representative hazard maps and on using off-the-shelf Kriging programs to generate spatial hazard maps. These recommendations will provide industrial hygienists with a statistical method to harness the power of direct-reading exposure monitors – an emerging technology that can address both physical (e.g., noise, radiation, heat) and chemical (e.g., gas, vapor, aerosol) workplace hazards both rapidly and relatively inexpensively considering both spatial and temporal variability.

## Section 2

### Scientific Report

The Specific Aims of this project have not changed from those in the original proposal:

*Specific Aim 1:* Apply and evaluate existing spatial statistical modeling techniques to generate concentration maps based on exposure assessments using portable, direct-reading instruments.

*Specific Aim 2:* Develop comprehensive datasets using direct-reading instruments for two typical exposure scenarios: noise and aerosols.

*Specific Aim 3:* Develop a reference concentration map by applying a spatiotemporal statistical model to the datasets collected in Aim 2 and evaluate the accuracy and precision of the simpler statistical mapping approaches employed in Aim 1.

For Aim 1, a manuscript was published to the peer-reviewed literature with Dr. Thomas Peters at the University of Iowa. In this manuscript we applied existing statistical methods and evaluated the influence of these interpolation methods on hazard maps generated with spatiotemporal exposure datasets. We considered three facilities (data collected by Dr. Peters) and seven interpolation methods and determined that Kriging using an exponential or spherical variogram model outperformed other interpolation methods. We stressed in the manuscript that when using an off-the-shelf Kriging program, it is important to visually verify that the variogram model is appropriately fitting the experimental variogram, particularly for small separation distances. We described the results of several methods to improve the variogram model: fitting several different model types, examining the number of bins used for the experimental variogram, or allowing for a nugget effect when fitting the variogram model. We concluded that the variogram model with the highest $R^2$ or the model that fits the data best for small separation distances should be used for interpolation. In many cases, inverse distance squared weighting performed nearly as well as the more complex interpolation methods. However, we recommend that Kriging be used for hazard mapping because it produces the best linear unbiased prediction.

For Aim 2, Dr. Koehler worked with a M.S. student at Colorado State University (CSU), Kirk Lake, to collect a dense spatiotemporal dataset on four days at a facility in the Front Range of Colorado. Management would not permit us to measure aerosols at this facility, so we focused on noise exposure. This dataset included measurements of noise with static and roving monitors in the production area of the facility. As expected, noise levels were very high (Leq up to 90 dB) at some locations and we recommended hearing protection for all workers in this area of the facility. We found that this facility had large spatial variations, but little temporal variability in noise exposures. A second exposure dataset was collected at an engine testing facility. We collected a spatiotemporal noise and aerosol exposure dataset on two days. This research contributed to the MS thesis to be written by Kirk Lake (Dr. Koehler was the primary thesis advisor) for graduation during Fall 2013. The work also resulted in a publication to The Journal of Occupational and Environmental Hygiene detailing the measurements and assessing the potential bias that results from the use of short-duration measurements to represent the TWA in a hazard map.

For Aim 3, we used two approaches to make recommendations to the community on data collection and data analysis procedures. First, we conducted a simulation study to evaluate the influence of spatial resolution and the use of replicate measures in the generation of hazard maps that are representative of the TWA concentration. In this study, we conducted simulations of hazard intensities with increasingly complex source specifications in a theoretical facility. Theoretical “sampling” campaigns (roving monitoring only) were conducted for each simulation and estimated hazard maps were compared to calculated TWA intensities. Simulated map accuracies were compared to existing hazard mapping datasets (Lake et al. 2015). The goal of this study was to provide recommendations to industrial hygienists on appropriate spatial sampling resolution and needed replicate measures, given preliminary data on the temporal variability within a facility. This work is detailed in a manuscript in preparation for submission to the Annals of Occupational Hygiene by the end of 2015.

Second, we developed a new statistical methodology to use the data collected at the engine testing facility to its fullest capacity. This methodology uses both the roving and static measurements together to produce hazard concentration estimates over time and space, going beyond just the estimation of the TWA. This approach is detailed in a manuscript submitted to the Annals of Applied Statistics.

**Research Development**

Didactic Training during this Career Development Award has included the following courses (in bolded font, below).

| Term | Course Number | Course Title |
|---|---|---|
| Spring 2011 | Grad 544B | Ethical Research |
| Spring 2011 | STAT 523 | Quantitative Spatial Analysis |
| Fall 2011 | ERHS 532 | Epidemiologic Methods |
| Fall 2011 | NR 505 | Concepts in GIS |
| Fall 2011 | STAT 695 | Independent Study: Spatiotemporal Statistics (Dr. Zhu, mentor) |
| Spring 2012 | ERHS 536 | Advanced Occupational Health |
| Spring 2012 | STAT 675K | Topics in Statistical Methods – Bayesian Statistics |
| Fall 2012 | BMS 500 | Mammalian Physiology I |
| Spring 2013 | BMS 501 | Mammalian Physiology II |
| Spring 2013 | STAA 575 | Short Course: Applied Bayesian Statistics |
| Spring 2013 | STAA 544 | Short Course: Mixed Models |

For my Research Training, I have continued working with the statistical programs R, Matlab, WinBugs, and SAS to evaluate existing spatial datasets. During Year 2, I also gave guest lectures in the Industrial Hygiene (ERHS 526) and Aerosols and Occupational Health (ERHS 726) courses at CSU. In August 2013, I began a tenure-track faculty position at the Johns Hopkins Bloomberg School of Public Health in the department of Environmental Health Sciences.

**Publications**

Koehler, Kirsten A., and Thomas M. Peters. "Influence of Analysis Methods on Interpretation of Hazard Maps." *Annals of Occupational Hygiene* (2012).

Lake, K., J. Zhu, H. Wang, J. Volckens, **K.A. Koehler**. Effects of data sparsity and spatiotemporal variability on hazard maps of workplace noise. J Occup Env Hygiene, 12(4): 256-265, 2015.

Ludwig, G., T. Chu, K. Koehler, H. Wang, J. Zhu. Static and roving sensor data fusion for spatio-temporal hazard mapping with application to occupational exposure assessment. Submitted to Annals of Applied Statistics.

Koehler, K., J. Zhu, H. Wang, T. Peters. Sampling strategies for accurate hazard mapping using short-duration measurements. In preparation for submission to Annals of Occupational Hygiene.

This project did not involve any human subjects research.