SHINGO IKEDA  
Geospatial Data Processing Engineer  
Gradient  
Email: sikeda@gradientcorp.com  

MATTHEW MAYO  
Geospatial Analyst and Geoscientist  
Gradient  
Email: mmayo@gradientcorp.com  

KIRSTEN ZU  
Epidemiologist  
Gradient  

STEPHANIE KLEINSCHMIDT  
Geospatial Analyst  
Gradient  

Shingo Ikeda  is a geospatial database engineer who has extensive experience in enterprise level GIS database management and software development. He routinely works with large-scale databases and is proficient in all aspects of data management, including data acquisition, conversion, and analysis. He possesses skills and knowledge necessary for raster and vector based GIS / Remote Sensing process design and development for various geospatial projects. Experienced with enterprise level spatial database management using Oracle, PostgreSQL and ArcSDE, he designs and develops web GIS interfaces with technologies such as ArcGIS Server, GeoServer, Python, Flex and JavaScript.

Kirsten Zu  (Ph.D., Sc.D., M.P.H.) is an epidemiologist specializing in the design and statistical analysis of epidemiological studies. She is particularly proficient with survival analysis and handling time varying exposures. Prior to joining Gradient, Zu was a post-doctoral research fellow at the Harvard School of Public Health, where she managed and analyzed substantial datasets from two large, multi-decade prospective cohort studies. Since joining Gradient, Zu has led several original research projects evaluating various health effects of PM2.5 and O3. She has experience in applying for identifiable research data from several government and municipal agencies, including the Massachusetts Department of Public Health, New York City Department of Health and Mental Hygiene, Texas Commission of Environmental Quality, and Texas Department of State Health Services. She also has experience in developing detailed study protocols and obtaining Institutional Review Board (IRB) or independent ethics committee approvals for research involving human subjects. She is also proficient with several statistical software packages, including SAS and R, which are commonly used for data management and statistical analysis.

Matthew Mayo  (M.S, GISP, CPG, PG) is a Geospatial Analyst and Geoscientist with 16 years of experience with spatial database construction and modeling focusing on projects related to hydrogeologic exploration and assessment, and hydrochemical analysis. He has provided technical consulting services to environmental professionals, municipalities, and private entities for wastewater disposal, storm water management, environmental compliance, public water supply, and land redevelopment and asset management projects. He has also provided expert technical support for geologic exploration and GIS services in litigation. Mayo has taught geology and GIS curricula and applications at area Universities and professional programs. He is proficient with numerous geospatial, geologic, and hydrochemical database and modeling software packages and possesses a working knowledge of environmental regulations. Mayo is a certified GIS processional and certified professional geoscientist, received his M.S. from Boston College with distinction in Geology and is
Managing Spatial Big Data and Assessing Positional Accuracy in Analysis of Nationwide Air Quality and Human Health Effects

The spatial distribution of air quality data is a key element in determining causal effects on human health. Developing a database of air quality and atmospheric information on a national scale is critical to establishing a framework on which causal analyses on the relationship between data such as PM2.5 and specific mortality endpoints can be performed. The positional accuracy of the data can have an effect on the final results. The focus of this talk or poster will be to highlight the objectives of the database construction, outline the challenges encountered during the construction process, highlight the spatial component of each of these datasets, provide examples of how the availability and accuracy of the spatial information can be related to the accuracy of the health effects evaluation, and summarize the lessons learned from developing this nationwide spatial database.

To develop this database we downloaded 1999-2013 data on PM2.5 and PM10 from the US EPA’s AQS Technology Transfer Network (total records = 33,589,401). Nationwide county information was retrieved from ESRI (total records = 3143). In addition, we also obtained local hourly meteorological data from the NOAA’s Quality Controlled Local Climatological Data (QCLCD) to generate daily temperature data for each county (total records = 3,200,775). The raw data contains daily PM measurements for most monitor sites and hourly PM measurements for some monitor sites. Some monitor sites have multiple monitor readings under the same monitor site ID.

There were many challenges in developing this database including the acquisition, evaluation, processing, and optimization of numerous “big data” files from different sources. For this effort we needed to find software and hardware solutions including a cluster based database with the ability to load a large volume of data with extremely fast processing and analytical capability.

The endpoint goal of the nationwide spatial database of air quality, weather, health and county data will be to enable users to access and analyze various human health effects due to environmental variables. The use of Big Data architecture in epidemiology will allow algorithms to further perform causal analysis using techniques such as decision tree and deep neural network.