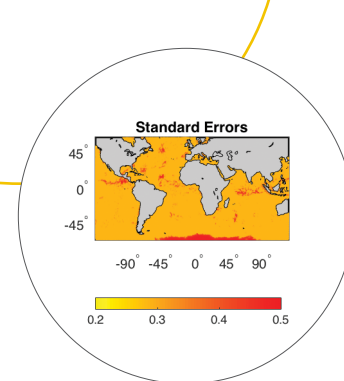


# Data Analysis

## Kang team creating novel modeling methods



Whether it's the ability to predict and respond to natural disasters, analyze brain imaging data, or understanding social network information, scientific researchers and engineers are increasingly turning to high performance computing (HPC) to tackle design obstacles or study real-world phenomena.

The University of Cincinnati's Emily Kang, Ph.D., is carrying out extensive simulation studies through the Ohio Supercomputer Center's resources to create methods and algorithms that will reduce uncertainties in computer modeling and make the process easier and more efficient for researchers.

"The methods and algorithms we have developed can be used to analyze data in many fields, including but not limited to geography, climate science, agriculture, biomedical sciences, and marketing," said Kang, an associate professor in the Department of Mathematical Sciences in the McMicken College of Arts and Sciences. "We are interested in studying how variables are associated with each other or using spatial/spatio-temporal dependence to improve prediction."

Kang has collaborated with researchers in many different research fields to develop robust and efficient statistical models that allow scalable memory and computational complexity, including efficient algorithms to fit high-dimensional social network data and flexible models to analyze massive global remote sensing data, which can be widely used in marketing and environmental engineering.

"Creating a novel method requires both theoretical justification and empirical demonstration," Kang said. "Although our project is viewed as more like fundamental research, we need to carry out extensive simulation studies to validate our new methods and compare their performance with existing ones."

Using software such as Matlab, Julia and R on OSC's Owens Cluster, Kang's analyses requires processing a huge amount of data.

"Using the supercomputer helps my group implement these numerical studies effectively and efficiently," she said. "OSC has enabled us to run simulation replicates in parallel and to easily implement scalable algorithms, which have been essential for the success of our project."

To complete all the studies in many different areas of research, Kang uses multiple students focusing on different topics in the project.

"OSC has enabled them all to investigate different problems and implement their own numerical studies simultaneously," Kang said. "It is essential for my students, the next generation of data scientists, to have the training and experience with high performance computing resources." •

**PROJECT LEAD** // Emily Kang, Ph.D., University of Cincinnati  
**RESEARCH TITLE** // Statistical models for massive data with complex dependence structure  
**FUNDING SOURCE** // University of Cincinnati  
**WEBSITE** // [emilystat.wixsite.com/gdads](http://emilystat.wixsite.com/gdads)

