NSF INTERN Graduate Research Opportunity

Point of Contact Name	Brian Gray
Point of Contact email	brgray@usgs.gov
Title of Opportunity	Models of high-dimensional environmental or ecological data
USGS Center or Program	Upper Midwest Environmental Sciences Center
Internship Location (city, state)	La Crosse, WI (conceivably virtual)
Comments	As this represents a modeling proposal, the intern would need to check in with the supervisor on a regular basis (e.g., weekly) but could otherwise remain at his/her home university. This approach reflects that virtual working arrangements have worked well within the USGS for the previous eight months and that interns working at their home university would better be able to interact with faculty and other students on the topic of the intern's study; such interactions would likely offer the potential for better insights by the student for the USGS. If the student remains offsite, one or more of the USGS responsibilities (e.g., equipment (e.g., laptop)) may not be required.
Creative Summary	Generalized linear mixed models have become popular among scientists and applied statisticians for fitting models of environmental and ecological processes. However, these models are limited in their abilities to address i) unexplained variation at cluster scales (e.g., organism or chemistry variables might be sampled at multiple locations in each of multiple lakes) and ii) multiple outcomes (e.g., multiple species, multiple dissolved nutrients or multiple chemical contaminants). Models that pertain to distributions of multiple species are often termed joint species distribution models. An intern with a statistical or mathematical interest could further develop existing multivariate models or, if a scientific interest, use already-developed multivariate models to tackle natural resource questions that are multivariate in nature. The intern could also employ machine learning models.
Objectives	The objective of this project may be theoretical or applied. A theoretical objective would be statistical and focus on elaborating current methods for making inferences or predictions from multivariate and moderately high-dimensional data, often consisting of regular and irregular time series. Such an approach would entail evaluations of a proposed method using simulated data and an example environmental or ecological dataset. An applied or science-driven approach would focus on using recently-developed computational methods that have seen little application with natural resource questions (e.g., joint species distribution models). A project could also have both theoretical and applied components. An example would be the development and/or use of models with counts from multiple fish species to estimate fish community associations with environmental predictors—or

	similar but with detections of multiple plant species. Or elaborate current methods to address left censoring of multiple chemicals (eg nutrients in freshwater or PCBs in fish tissue).
Intern Duties and Responsibilities - the body of the project description	The intern will develop models appropriate for use with multivariate and clustered environmental or ecological data or use already-developed models to fit such models in a new ecological or environmental setting. Models will be evaluated using statistical or mathematical software and both simulated and measured data. The intern will have the opportunity to work in a natural resource setting, to interact with natural resource scientists, and to share findings with partner natural resource agencies and the scientific community, the latter via one or more peer-reviewed publications.
Key skills required	Familiarity with statistical methods (including probability distributions, and generalized linear models) and/or machine learning, and analytical software (e.g., R or SAS); if a statistical approach will be taken, then the intern will need to have completed graduate-level mathematical statistics courses. Modest familiarity with hydrology, ecology or inorganic chemistry is not required but would be helpful.
Applicable NSF Directorate (check all that apply)	Biological Sciences;Computer and Information Sciences;Geological Sciences;Mathematics, Physics, and Astronomical Sciences;
Keywords	Biology/microbiology/biochemistry;Chemistry/geochemistry;Computer/Data science;Ecology/Ecosystems;Environmental Health;Hydrology;Modeling;Population Dynamics;Statistics;