USGS NSF GRIP Opportunity

Point of Contact Name:	James Cronin & Donald Schoolmaster
Point of Contact E-mail:	jcronin@usgs.gov; schoolmasterd@usgs.gov
USGS Center:	Wetlands and Aquatic Research Center
Project Title:	Identifying statistical procedures that can be used to correctly estimate ecological relationships: Graphical Causal Modeling for ecology
Project Hypothesis or Objectives:	One of the greatest problems facing ecologists is the increased pressure to rapidly provide decision-support tools for managing novel ecosystems. Moving from a basic science approach to designing decision-support tools requires a methodological shift from identifying potential causal effects (e.g. “does x significantly affect y?”) to generating high quality, unbiased estimates of causal effects (“How much does x change y?”). The solutions to this problem require a bridge that transparently and efficiently guides researchers from basic ecological principles (i.e., our knowledge and beliefs about the world) to appropriate parameter estimation tools. Yet, it is increasingly recognized that conventional methods of ecological inquiry lack this link. The potential success of any given parameter estimation tool depends on how well the causal structure implied by the statistical procedure reflects the causal structure of the process that actually generated the relationship. Unfortunately, the correct causal structure is often unknown and, consequently, it is the role of theory to hypothesize a data-generating process. For systems with potentially complex sets of causal relationships, graphical causal modeling (GCM) can be used to graphically represent and analyze the hypothesized data-generating process, often called a causal diagram (i.e., box and arrow graphs). Developed by computer scientists, mathematicians, philosophers, and statisticians, GCM is a formal framework for analyzing causal diagrams using a simple set of rules to determine whether (1) a specific relationship can be estimated without bias given a set of observed variables and, if so, (2) inform what variables must be included in the statistical model to estimate the relationship. In other words, GCM informs the statistical procedures that can be used to obtain an unbiased estimate of the relationship from experimental and observational data prior to data collection and/or

	parameter estimation. Consequently, GCM has great potential to strengthen communication between ecologists and decision makers while increasing the efficiency and validity of ecological research. This project's objective is to develop a tutorial that (1) helps researchers develop the skills necessary to identify statistical procedures that can be used to correctly estimate relationships among variables and (2) apply these skills to their research. We will work together to develop a tutorial that outlines GCM motivation and methodology and uses the intern's research area as an example. To accomplish this, we will provide a guided exploration of GCM. We will then develop a causal model based on the intern's research area. Finally, we will determine which relationships can be estimated without bias and, if so, which variables must be included in statistical statements. The successful intern's research will address theoretical concepts that repeatedly occur in ecology and evolutionary biology and are relevant to both basic and applied science (e.g., allometry, phylogeny, detection of rare species, calculated biometrics such as diversity, fitness, productivity, host competence). Graphical causal modeling does not require a statistical or quantitative background. It does require dedication to developing and analytically assessing the causal structures that are implied by the theoretical language of a research area.
Duration:	12 months
Internship Location:	Lafayette, LA
Field(s) of Study:	Basic or applied research on the Ecology, Evolution, and Behavior of terrestrial or aquatic systems.
Expected Outcome:	The project will produce (1) a tutorial on GCM's motivation and methodology; (2) a causal diagram representing an ecological relationship that is fundamental to both basic and applied ecology; and (3) GCM analysis of the causal diagram that identifies which statistical procedures can be used to correctly estimate the causal diagram relationships. These products will provide the intern analytical skills that can be applied to both their dissertation and future research projects. The project will also provide a GCM tutorial aimed at USGS researchers.
Special skills/training Required:	The intern should have defined their dissertation topic so that they understand both the theoretical language and methods of data collection of their research area. Previous statistical experience (e.g., probability theory, structural equation modeling, quantitative ecology) is helpful but not necessary.
Duties/Responsibilities:	The intern will be responsible for developing the tutorial by (1) reading and discussing GCM literature, (2) developing a causal

	model, and (3) analytically assessing the causal diagram using GCM.