Nathan Piekielek
Geospatial Services Librarian
Head, Donald W. Hamer Center for Maps & Geospatial Information
The Pennsylvania State University Libraries
Email: nbp104@psu.edu

Position Paper on Replicability and Reproducibility

1) What forms of failure to replicate exist in the geospatial sciences? Can a formal framework be devised?
2) In what areas of geospatial research is the danger of non-replicability most severe?
3) What mechanisms can be used to avoid or minimize the danger of such failures?
4) Do we expect the results of model calibrations to be constant over space and if not, what are the implications for spatial analysis?
5) How should R&R be incorporated into the design and implementation of future spatial software?
6) How should students be made aware of these issues?
7) What follow-on activities might draw greater attention to these issues?

I understand replicability and reproducibility (R&R) to be two distinct scientific concepts that were largely developed by researchers who were trying to make sense of the results of field and laboratory experiments often through statistical analyses. This understanding comes from my background and training in the statistical/experimental fields of ecology and biology. In my view, replicability refers to the ability to achieve similar research results under similar conditions as one might attempt to do in a randomized block study design in a field or laboratory trial where independent variables are manipulated as treatments and the effect on the dependent variable is observed while attempting to control for all other confounding factors. Replication as a concept that evolved as we learned that even in the laboratory we cannot control everything and therefore the expectation is that we attain more or less the same results from randomized block replicates. Through the use of population-level statistical models to quantify the treatment effect (controlling statistically for random and block effects), it is thought that having more random replicates will produce a model with greater generalizability due to the variety of random contextual conditions under which the experiment was conducted. Reproducibility on the other hand, is a much higher standard that refers to the ability to achieve the exact same experimental result under the same conditions. Until recently, reproducibility was largely an aspirational concept in experimental fields rather than something that was seen as achievable. Computational experimentation changed our view of reproducibility so that now it is seen as achievable, although it remains a real challenge for a variety of reasons including poor documentation and a general lack of software preservation. The geospatial sciences have been slow to adopt R&R concepts because the geography discipline (that is the academic home of GIScience), has little history of experimentation, and until recently the tools of GIScience have been largely singular (i.e. ESRI products), with little influence of the computational sciences and/or other fields that were incorporating simulation experimentation into their repertoire of well-accepted

methods. As the existence of this workshop suggests, we are presently at a crossroad where there is ample opportunity to make progress on R&R in geospatial research and tools.

From Tobler's first law, geography has long recognized geographic position and context as important factors in understanding most phenomenon. It follows that replication (in the sense described above), in the geospatial sciences has been rare while in other fields (e.g. ecology), location has played a prominent role causing confusion in a variety of long-standing and high-profile arguments about foundational theory. In fact, if the ability to replicate across space were greater, (i.e. geographic position and context were less important), the discipline of geography would either look very different than it does, or perhaps not exist at all – why would we need geography? Instead, we observe autocorrelation of varying strengths that can either be scrubbed from statistical models that aim to elucidate cause and effect, or the autocorrelation itself modeled as important as is the case in many geoscience applications. In either case, the real danger to scientific inquiry and knowledge creation is in obscuring how autocorrelation was handled so that on the one hand geographers fail to consider central tenants of experimentation, replication, and generalizability in their investigations; and on the other hand, scientists from other disciplines fail to consider the effect of geographic context in their experimentation. Software and computational tools play an important role in how R&R is handled in the geospatial sciences.

In the spirit of reproducibility, transparency in the research process, and maintaining public trust it seems intuitive that one should be able to produce the same computational results given the same data inputs and analysis methods. In practice, this is often a challenge (I have been unable to reproduce the research results of others in three out of three tries), and is often a failure of software tools and sometimes of researchers themselves (i.e. human error). In proprietary geospatial software tools (i.e. ESRI products), there is a tension between ease of use and reproducibility and self-documentation. Even if geospatial software has the ability to self-document, the number of short-cuts and on-the-fly adjustments that are baked into the tool are usually so great that they seriously compromise reproducibility. The only successful cases of reproducibility in geospatial research with which I am familiar all used self-documenting free and open-source geospatial tools (i.e. spatial libraries of R statistical). Scientific workflow softwares provide substantial gains in ease of use and robust documentation, although they place a substantial burden on software preservation to ensure reproducibility. There is no silver-bullet to ensure reproducibility, only ways of shifting the burden back and forth between softwares, users, and those involved in preservation.

The solution(s) requires us to take a long-view and train geography students (i.e. the next generation of spatial experts), on the concepts of R&R and computational experimentation as well as in foundational skills like coding and documentation. They in turn will likely pressure geospatial software developers to produce the tools necessary to make advances in geo R&R.

*A small partial grant (~$500) towards travel and lodging would help me to attend this workshop, although may not be required depending on my travel funding next calendar year?