Position Statement

The inclusion of spatial analytical functionality in general and spatial data analysis in particular in a GIS environment has evolved considerably during the past decades. This started with early efforts during the late 1980s, primarily in academia. Those early implementations linked commercial GIS to commercial statistical software, using various methods of loose and close coupling. Most of the methods implemented at the time remained non-spatial. Since then, major progress has been achieved. Today, there is a proliferation of spatial data analytical tools, toolboxes and plug-ins for GIS, both proprietary and open source. However, most of the analytical extensions are still focused on a traditional desktop environment. In this respect, a critical but underestimated factor in the dissemination and adoption of spatial analytical techniques was the explosive growth of the internet as a means of distributing geographic information. These distribution activities have evolved beyond simply providing geographic data or maps on a web site to an encompassing notion of geoportals, spatial data infrastructures, and distributed geographic information services.

Paralleling the emergence of large repositories of geographic information was the development of the semantic and geospatial semantic web and its associated focus on service oriented architectures. Recent research has stressed the development of ontologies, service descriptions and discovery, much of it under the auspices of the OGC. This has led to standards for the delivery of raster and vector maps and data, such as the Web Map Service and Web Feature Service standards. However, much less progress has been made with regard to analytical services, other than the currently still fairly rudimentary Web Processing Services standard. Also, the scope of analytical techniques implemented in a web environment remains fairly limited, consisting primarily of descriptive statistics, such as local spatial autocorrelation analysis, and proof of concept illustrations.

Spatial econometrics in particular is not well represented in this context. Spatial econometrics is a subset of econometric methods that is concerned with spatial aspects present in cross-sectional and space-time observations. Variables related to location, distance and arrangement (topology) are treated explicitly in model specification, estimation, diagnostic checking and prediction. With a few exceptions, spatial econometric methods are still mostly absent in commercial statistical software. There are
three main software development communities, two of which are centered on a particular platform, one commercial, one open source. One consists of libraries developed for the commercial Matlab software, primarily by James LeSage and Kelly Pace and collaborators. The other is the spdep library for the open source R environment, developed by Roger Bivand and many others. A third community consists of the successors to the early SpaceStat package, i.e., GeoDa, OpenGeoDa and GeoDaSpace, maintained and developed at the GeoDa Center of Arizona State University. Arguably, the latter may be the dominant group in terms of numbers, with the count of GeoDa users now approaching 50,000.

The idea of a spatial econometrics workbench is to leverage the ongoing efforts of the different development groups and move the development and dissemination of this highly specialized software beyond the desktop into a collection of software components delivered over the internet. These components should support the application of advanced methods to the specification and estimation of spatial models in a collaborative environment, based on a service oriented architecture. The workbench uses the same principles as adopted by natural science and engineering cyberinfrastructure initiatives, characterized by a comprehensive and flexible collection of methods, software tools, and support materials.

This effort leverages the open source PySAL library for spatial data analysis, under development at the GeoDa Center and written in the Python language. Specifically, functionality is being delivered in the form of specialized web services that adhere to common and open standards. In addition, methods are provided as web applications that access the services through a web browser. The spatial econometrics workbench is envisaged as a gateway or portal that provides access to a range of techniques, data sets and simulation environments to support both the empirical application of existing methods as well as the testing and evaluation of new methods. A current prototype provides functionality to create, transform and manipulate spatial weights, which are an important part in spatial autocorrelation analysis.

The development of this workbench raises a number of important research questions pertaining to service discovery and chaining, scalability and performance. Important work remains to be done to develop “metadata” for spatial econometric models and methods. This is particularly challenging, since there are typically many different ways in which a given problem may be approached, which is not straightforward to express in a “standard.” In our work, we have found that standard ontologies are ill equipped to deal with this issue and alternative approaches are needed. In particular, effective mechanisms are needed to discover the appropriate atomistic methods and to properly chain them in an effective workflow. In addition, to obtain meaningful performance in the execution of complex computational tasks in a web service implementation requires rethinking of existing algorithms. This involves parallelization where possible and taking advantage of environments such as the cloud or a computational grid to scale up the size of applications to realistic dimensions. To date, relatively little progress has been achieved in this regard and it remains an important research challenge.