Technical Vignette 1: Why kriging in ArcGIS may be a bad idea: A statistician's perspective

Christopher Paciorek, Department of Biostatistics, Harvard School of Public Health

Version 1.0: January 2008

Below I outline a number of drawbacks in using ArcGIS to do kriging compared to using statistical software packages. In general, for exploratory analysis and straightforward model fitting, I would recommend using mgcv (the gam() function) or SemiPar (the spm() function) in R to fit penalized likelihood smoothing models rather than doing kriging. Both mgcv and SemiPar allow for additive models with the inclusion of additional covariates (as linear or smooth terms) and for count and binary data and both provide estimates of uncertainty. Neither relies heavily on ad hoc fitting methods used in the ArcGIS kriging routines. An alternative that actually implements kriging but using maximum likelihood to fit the covariance parameters is the geoR package in R. Also note that if one wants to use the GIS software to create maps, the results from mgcv and SemiPar can be imported into the GIS software.

Some drawbacks to kriging in ArcGIS:

1. Not likelihood-based: ArcGIS implements kriging via the variogram fitting approach. In the last two decades, fitting kriging models (i.e., models with latent Gaussian process structure) via maximum likelihood and Bayesian methods have become more popular. Many statisticians now favor these latter approaches because they are grounded in standard likelihood theory, the framework for much statistical work.

2. Many parameters and user choices: ArcGIS kriging relies heavily on ad hoc fitting methods that involve a number of parameters and user choices in how to do the fitting. The degree of smoothness in the surface estimated by SemiPar and mgcv is controlled by a single parameter, rather than a number of parameters in the kriging approach. While having multiple parameters and choices might seem to allow for a more flexible fit to the data, with the kind of smoothing done for standard spatial data, there is effectively only one parameter that controls the degree of smoothing, so the kriging approach introduces additional complexity with little change in the ability to fit a surface. The exception to this comment is the ability of the kriging methodology to account for directionality. mgcv and SemiPar do not allow for anisotropy, so strongly directional surfaces may be better fit via kriging.

3. Issues with non-normal data: There are now a variety of likelihood and penalized likelihood approaches to fitting spatial models and spatio-temporal models that allow for count and binary outcomes and for easily including other covariates in the mean structure of the model. In particular, thin plate spline models are implemented in an additive modeling framework

in mgcv and SemiPar. The generalizations of kriging to generalized outcomes such as implemented in ArcGIS are not based on standard generalized linear statistical models, while the mgcv and SemiPar approaches are firmly grounded in the GLM and GLMM statistical modeling framework and use the Poisson and binomial likelihoods.

4. Inability to deal with additional covariates: ArcGIS is not set up to handle additional covariates in the mean structure of the model. In contrast mgcv and SemiPar are explicitly designed to fit models of the form: $y_i \sim \mathcal{N}(\mu + \sum f_k(x_{k,i}) + g(s_i), \sigma^2)$ where the $f_k(\cdot)$ terms are smooth regression functions of additional covariates and $g(\cdot)$ represents the spatial structure in the model. The $f_k(\cdot)$ terms can also be specified to be linear terms of the additional covariates.

5. GUI-based: Kriging in ArcGIS is done through the graphical user interface, while R allows one to fit many models using scripts. This allows automation of analyses. A caveat here is that I believe there is some ability to set up scripts in ArcGIS, but scripting is at the core of what R is.

6. Lack of transparency: The underlying code by which the ArcGIS kriging is done is not accessible (to my knowledge), and it can take some effort to understand exactly what is being done based on the documentation. The online documentation has limited technical detail. More extensive information is available in:

    - Johnston, K. et al. 2001. Using ArcGIS geostatistical analyst. Redlands, CA: Environmental Systems Research Institute.
    - McCoy, J. et al. 2001. Using ArcGIS spatial analyst. Redlands, CA: Environmental Systems Research Institute.
    - The ESRI website has some technical reports, most of which are somewhat tangential to the methods implemented in ArcGIS, but a couple of which provide further details on the fitting methods.

If you do want to do kriging in ArcGIS, despite my concerns above, I have created awebsite (based on ArcGIS 8.3) with some guidance on how to do it. Be forewarned that this information is based on ArcGIS 8.3 and was written in 2004 or so, so it is somewhat outdated.