# EFFECTS OF SAMPLING INTERVAL ON SPATIAL ANALYSIS OF WATERSHED NITROGEN LOADING[1]

***Shuo-sheng Wu***, E. Lynn Usery, and Michael P. Finn (U.S. Geological Survey, Rolla, MO, USA)

**ABSTRACT:** The purpose of this study is to investigate how spatial patterns and statistics of watershed nitrogen concentration change with sampling interval. The Little River watershed, Georgia, USA, is used as a case area. Point samples at 30-m to 600-m intervals at 30-m lag are taken, respectively, from a 30-m-resolution nitrogen surface that was generated previously with the Agricultural Non-Point Source (AGNPS) pollution model. Basic statistics, variograms, interpolation accuracies, clustering levels, and hot spots are calculated from the samples, and their changes with sampling interval are examined. The results show that the mean, standard deviation, and variogram sills do not have increasing or decreasing trends with increasing sampling intervals while the variogram ranges remain constant. In contrast, the interpolation accuracy, clustering level, and hot spot areas display decreasing trends that approximately follow a power law curve and correspond to the variogram of 30-m-resolution nitrogen. The sampling interval of 360 m appears to be a critical point at which the trends start to level. The trends with sampling interval help researchers and practioners who study watershed pollution to determine appropriate sampling intervals, and the clustering and hot spot analyses are useful for watershed pollution management and resources planning.

## INTRODUCTION

Researchers often need regular samples of environmental variables to model natural processes (Morris, 1999; Tian et al., 2002; Litaor et al., 2002) or as a preliminary effort for optimizing a subsequent sampling scheme (van Groenigen et al., 1997; Simbahan and Dobermann, 2006). Determining appropriate sampling intervals is important because analyses based on different sampling intervals may produce different results. Past studies have explored the effect of sampling interval on the accuracy of digital elevation models (e.g., Li, 1992; Gong et al., 2000; Aguilar et al., 2005; Chaplot et al., 2006), on interpolation accuracy of various soil variables (e.g., Oliver and Frogbrook, 1998; Bourennane and King, 2000; Frogbrook and Oliver, 2000), on spatial patterns and statistics of soil properties (e.g., Western and Blöschl, 1999; Oline and Grant, 2002; Sobieraj et al., 2004; Iqbal et al., 2005), on the estimation of home-range size (e.g., Swihart and Slade, 1985), and on the result of trail impact assessment (e.g., Leung and Marion, 1999). This study investigates how sampling intervals affect the spatial patterns and statistics of watershed nitrogen loading, with the goal to provide guidance for future sampling efforts. Our approach is to systematically take samples at incremental intervals from an AGNPS-generated watershed nitrogen surface to conduct analyses. Using model-simulated data allows us to efficiently test different sampling intervals and to examine the effects. Furthermore, this study intends to investigate the effect of sampling interval on clustering and hotspot analyses of watershed nitrogen loading. Clustering analysis helps determine whether the overall level of nitrogen concentration in a watershed is statistically significant. Hotspot analysis allows one to locate statistically significant high-value clusters (hot spots). These analyses are of importance to watershed pollution management and resources planning. Sampling interval effects on the analysis results is therefore of great concern to researchers and practioners of watershed management.

The next section of the paper (MATERIALS AND METHODS) presents the study area, datasets, and analysis methods in detail. Then the RESULTS section describes results of the analyses of basic statistics, variograms, interpolation accuracies, clustering levels, and hot spots. The DISCUSSION

1 Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

ISBN 978-0976885382
© 2007 American Science Press

section discusses important findings and critical issues of the analyses. Lastly, the CONCLUSIONS section concludes with a summary of the study with emphasis on the major findings.

## MATERIALS AND METHODS

Our study area is the Little River watershed, Georgia, USA, approximately 12 km in width and 35 km in length (FIGURE 1). In a previous study of resolution effects on watershed modeling, a soluble nitrogen concentration surface of 30-m resolution was simulated using the Agricultural Non-Point Source (AGNPS) pollution model (Young et al., 1994, 1995), based on a hypothetical precipitation event and 22 input parameters derived from soil, land cover, and elevation data (Usery et al., 2004). In this study, to investigate the effect of sampling interval, we first selected point samples at 60-m, 90-m, …, to 600-m intervals, respectively, from the simulated nitrogen surface. We then calculated the mean, standard deviation, variogram, interpolation accuracy, clustering level, and hot spots based on the samples. How theses statistics change with sampling interval were examined.

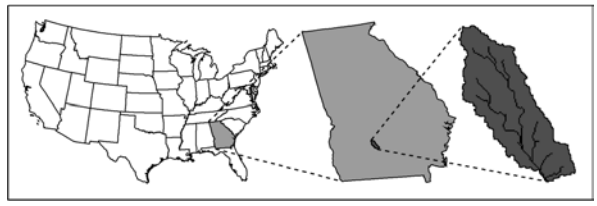

**FIGURE 1. Little River watershed, Georgia, USA**

We used the natural neighbor algorithm for interpolation because it is a simple and robust algorithm that is sufficient to meet our need considering the relatively dense sampling scheme (Boissonnat and Cazals, 2001). We interpolated nitrogen surfaces of 30-m resolution from the point samples and then compared the interpolated surfaces with the original 30-m nitrogen surface. The root mean square error (RMSE) and the total absolute error (TAE) were calculated.

We calculated the Z score of the global Moran's index to represent the overall clustering level, and the Z score of the local Getis-Ord statistic (by cells) to derive hot spots of nitrogen concentrations. Hot spots of Z scores of 1.96 and above, which indicate a 95% confidence interval, are chosen to compare between different sampling intervals. We calculated the clustering and hot spot statistics based on a specified neighborhood distance of 900 m, because this distance is approximately the variogram range (see the RESULTS section) and indicates the limit of spatial interaction.

## RESULTS

The results show that the mean and the standard deviation of nitrogen concentration do not have a consistent trend with increasing sampling intervals (FIGURE 2 and FIGURE 3). The variogram sill does not consistently increase or decrease while the variogram range remains unchanged (approximately 900 m) with increasing sampling intervals (FIGURE 4 and FIGURE 5). The interpolation errors display an increasing trend with increasing sampling interval (FIGURE 6 to FIGURE 8). The overall clustering level and the total hot spot area have a decreasing trend with increasing sampling intervals (FIGURE 9 to FIGURE 11).

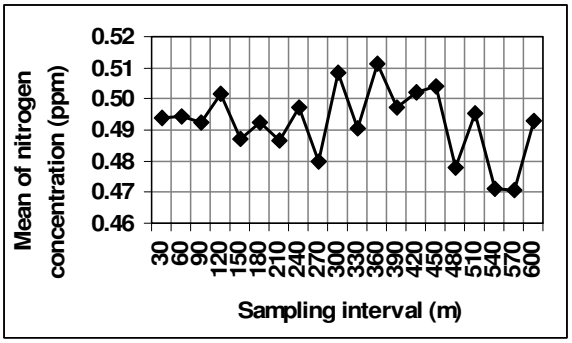


**FIGURE 2. The trend of the nitrogen mean with sampling interval**

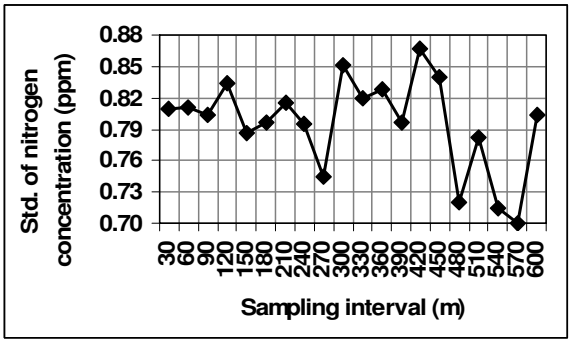


**FIGURE 3. The trend of the nitrogen standard deviation with sampling interval**

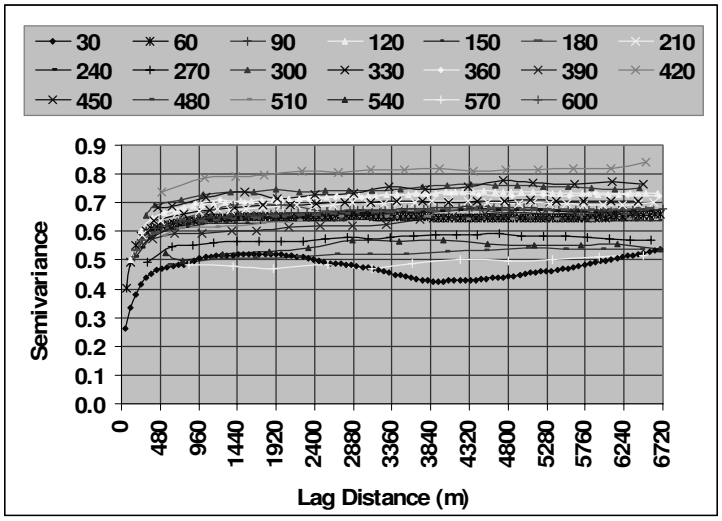


**FIGURE 4. The change of variograms with sampling interval**

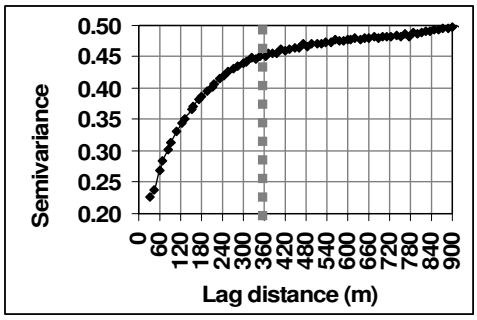


**FIGURE 5. The variogram of the 30-m nitrogen concentration**

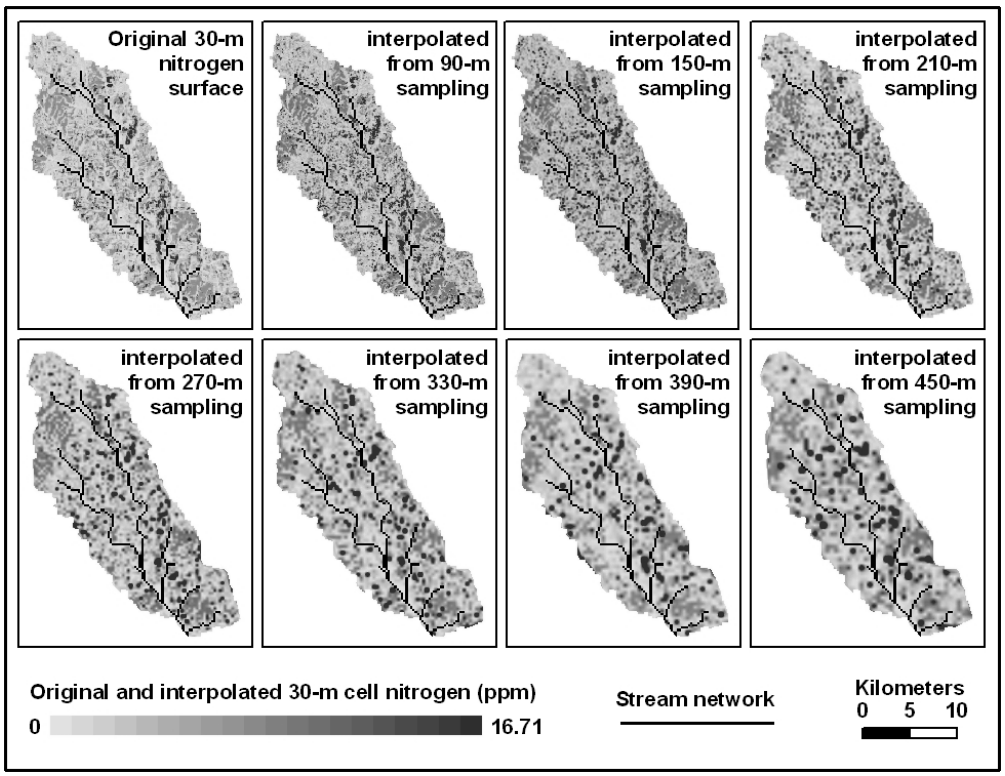


**FIGURE 6. Original and interpolated 30-m nitrogen surfaces based on different sampling intervals**

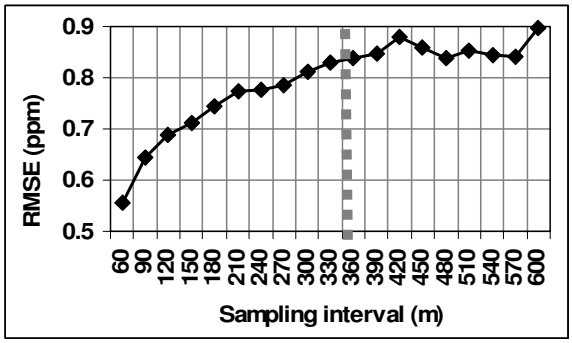


**FIGURE 7. The trend of the root mean square error (RMSE) with sampling interval**

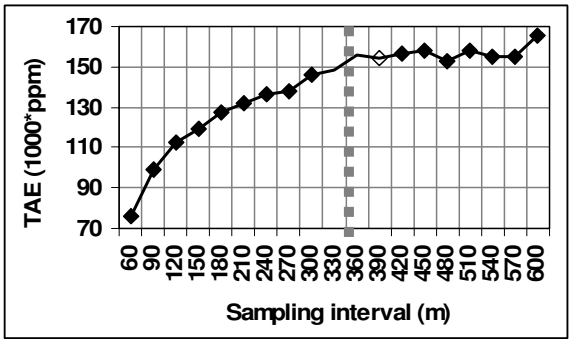


**FIGURE 8. The trend of the total absolute error (TAE) with sampling interval**

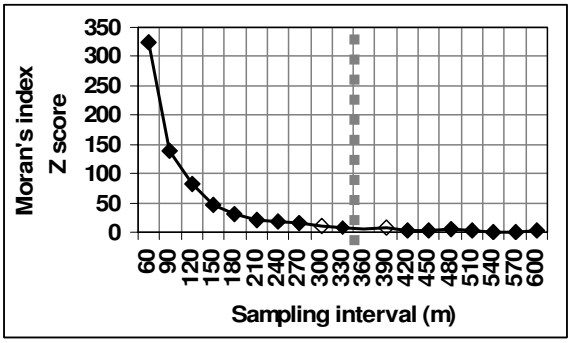


**FIGURE 9. The trend of clustering levels with sampling interval**

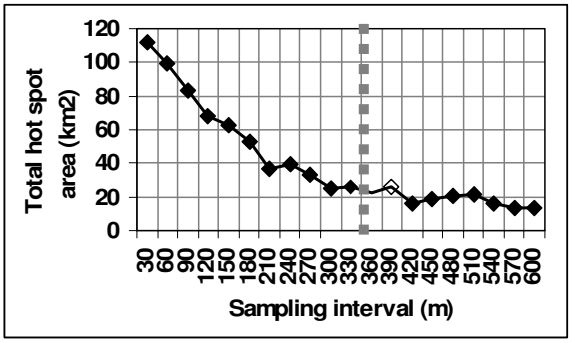


**FIGURE 10. The trend of total hot spot areas with sampling interval**

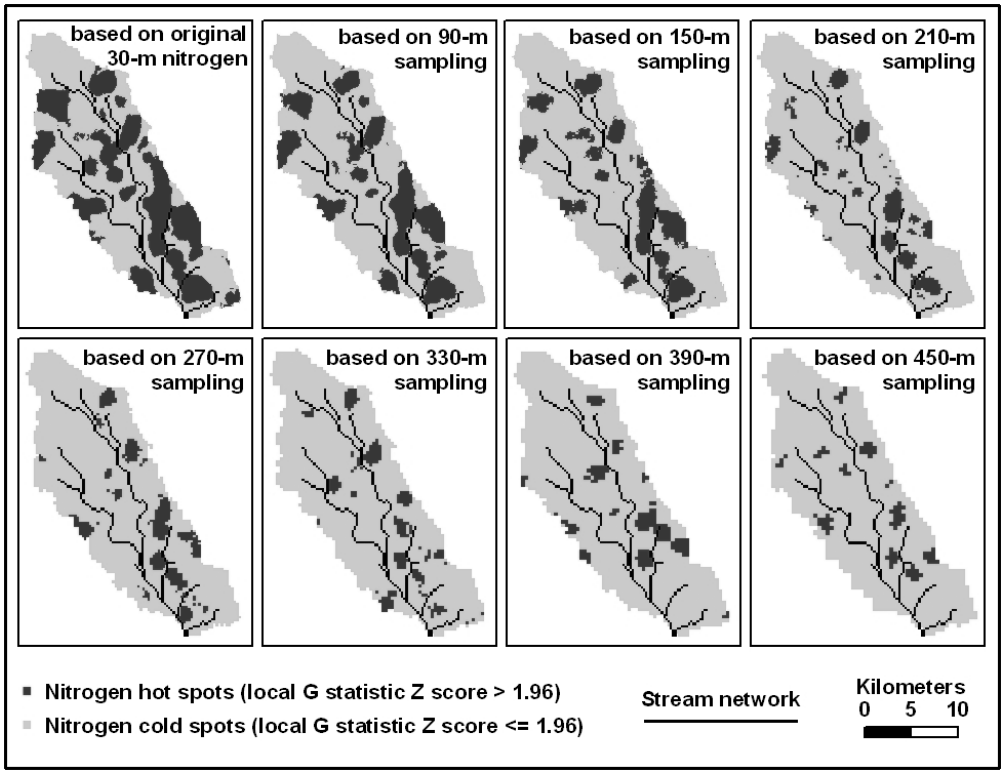


**FIGURE 10. Hot spot patterns of 30-m nitrogen and based on different sampling intervals**

## DISCUSSION

The trends of sample mean and sample standard deviation indicate that estimates from higher sampling densities may not necessarily be more accurate (FIGURE 2 and FIGURE 3). Nevertheless, the trends appear to fluctuate more at larger sampling intervals and, therefore, estimates from smaller sampling intervals may be more robust and reliable.

The comparison of variograms between different sampling intervals demonstrates that sampling intervals do not affect the observed variogram range (FIGURE 4). Nonetheless, the smallest lag available for computing the semivariance in a variogram is the respective sampling interval and, therefore, large sampling intervals may not allow one to construct a representative variogram when the variogram range is relatively small. Furthermore, smaller sampling intervals can produce larger numbers of samples, which generally yield more reliable estimates of the semivariance.

The interpolation error, clustering level, and hot spot areas appear to have similar trends of a power law curve with increasing sampling intervals. The trends also correspond to the variogram of the 30-m nitrogen at equivalent lag distances (FIGURE 5). A sampling interval of 360 m appears to be a critical point at which the trends start to level. The implication is that analyses based on a sampling interval of 360 m will be similar to those based on large sampling intervals (e.g., 600 m) and, therefore, one may prefer using large sampling intervals for cost effectiveness.

## CONCLUSIONS

This study investigates how spatial patterns and statistics of the nitrogen loading in the Little River watershed change with sampling interval, by taking incremental samples at 60-m to 600-m intervals with a 30-m lag from a AGNPS-simulated, 30-m-resolution nitrogen surface. The results show that the mean and standard deviation do not change consistently with sampling interval while the estimates are more robust and reliable at small sampling intervals. Variogram sills do not have consistent trends with changing sampling interval while the variogram ranges remain approximately the same. Small sampling intervals are necessary for constructing representative and reliable variograms at small lags. The interpolation accuracy, clustering level, and hot spot areas have corresponding trends with the 30-m-resolution variogram. A sampling interval of 360 m is a critical point at which the trends start to level and the analysis results become similar. The trends of analysis results with different sampling intervals help researchers and practioners who need to take field samples for studying watershed pollution to choose appropriate sampling intervals for optimizing cost benefits. Moreover, the clustering and hot spots analyses presented in this study are of great value for watershed pollution management and resources planning.

## REFERENCE

Aguilar, F. J., F. Aguera, M. A. Aguilar, and F. Carvajal. 2005. “Effects of Terrain Morphology, Sampling Density, and Interpolation Methods on Grid DEM Accuracy.” *Photogrammetric Engineering and Remote Sensing*. 71(7): 805-816.

Boissonnat, J. D, and F. Cazals. 2001. “Natural Neighbor Coordinates of Points on a Surface.” *Computational Geometry-Theory and Applications*. 19(2-3): 155-173.

Bourennane, H., and D. King. 2000. “Impact of the Sampling Density of the Primary Variable and Precision of the Auxiliary Variable on the Estimation of Soil Properties Using Kriging with an External Drift.” In G. B. M. Heuvelink and M. J. P. M. Lemmems (Eds.), *Proceedings of International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2000)*, pp. 67-74. Delft University press, Amsterdam.

Chaplot, V., F. Darboux, H. Bourennane, S. Leguedois, N. Silvera, and K. Phachomphon. 2006. “Accuracy of Interpolation Techniques for the Derivation of Digital Elevation Models in Relation to Landform Types and Data Density.” *Geomorphology*. 77(1-2): 126-141.

Frogbrook, Z. L., and M. A. Oliver. 2000. "The Effects of Sampling on the Accuracy of Predictions of Soil Properties in Precision Agriculture." In G. B. M. Heuvelink and M. J. P. M. Lemmems (Eds.), *Proceedings of International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2000)*, pp. 225-232. Delft University press, Amsterdam.

Gong, J. Y., Z. L. Li, Q. Zhu, H. G. Sui, and Y. Zhou. 2000. "Effects of Various Factors on the Accuracy of DEMs: An Intensive Experimental Investigation." *Photogrammetric Engineering and Remote Sensing*. 66(9): 1113-1117.

Iqbal, J., J. A. Thomasson, J. N. Jenkins, P. R. Owens, and F. D. Whisler. 2005. "Spatial Variability Analysis of Soil Physical Properties of Alluvial Soils." *Soil Science Society of America Journal*. 69(4): 1338-1350.

Leung, Y. F., and J. L. Marion. 1999. "The Influence of Sampling Interval on the Accuracy of Trail Impact Assessment." *Landscape and Urban Planning*. 43(4): 167-179.

Li, Z. L. 1992. "Variation of the Accuracy of Digital Terrain Models with Sampling Interval." *Photogrammetric Record*. 14(79): 113-128.

Litaor, M. I., T. R. Seastedt, D. A. Walker. 2002. "Spatial Analysis of Selected Soil Attributes across an Alpine Topographic/Snow Gradient." *Landscape Ecology*. 17(1): 71– 85.

Morris, S. J. 1999. "Spatial Distribution of Fungal and Bacterial Biomass in Souther Ohio Hardwood Forest Soils: Fine Scale Variability and Microscale Patterns." *Soil Biology and Biochemistry*. 31: 1375–1386.

Oliver, M. A., and Z. L. Frogbrook. 1998. *Sampling to Estimate Soil Nutrients for Precision Agriculture*. Preceedings No 417, The International Fertiliser Society, Cambridge, UK.

Oline, D. K, and M. C. Grant. 2002. "Scaling Patterns of Biomass and Soil Properties: An Empirical Analysis." *Landscape Ecology*. 17(1): 13-26.

Simbahan, G. C, and A. Dobermann. 2006. "Sampling Optimization based on Secondary Information and its Utilization in Soil Carbon Mapping." *Geoderma* 133(3-4): 345-362.

Sobieraj, J. A., H. Elsenbeer, and G. Cameron. 2004. "Scale Dependency in Spatial Patterns of Saturated Hydraulic Conductivity." *Catena*. 55(1): 49-77.

Swihart, R. K., and N. A. Slade. 1985. "Influence of Sampling Interval on Estimates of Home-Range Size." *Journal of Wildlife Management*. 49(4): 1019-1025.

Tian, Y. H, C. E. Woodcock, Y. J. Wang, J. L. Privette, N. V. Shabanov, L. M. Zhou, Y. Zhang, W. Buermann, J. R. Dong, B. Veikkanen, T. Hame, K. Andersson, M. Ozdogan, Y. Knyazikhin, and R. B. Myneni. 2002. "Multiscale Analysis and Validation of the MODIS LAI Product - I. Uncertainty Assessment." *Remote Sensing of Environment*. 83(3): 414-430.

Usery, E. L, M. P. Finn, D. Scheidt, S. Ruhl, T. Beard, M. Bearden. 2004. "Geospatial Data Resampling and Resolution Effects on Watershed Modeling: A Case Study Using the Agricultural Non-Point Source Pollution Model." *Journal of Geographical Systems*. 6: 289–306.

Van Groenigen, J. W., A. Stein, and R. Zurrbier. 1997. "Optimization of Environmental Sampling Using Interactive GIS." *Soil Technology*. 10(2): 83-97

Western, A. W., and G. Blöschl. 1999. "On the Spatial Scaling of Soil Moisture." *Journal of Hydrology*. 217(3/4): 203–224.

Xiangyun, X., G. Gertner, G. Wang, and A. B. Anderson. 2005. "Optimal Sampling Scheme for Estimation Landscape Mapping of Vegetation Cover." *Landscape Ecology*. 20(4): 375-387.

Young, R. A., C. A. Onstad, D. D. Bosch, and W. P. Anderson. 1994. *Agricultural non-point source pollution model, version 4.03, AGNPS user's guide*. North Central Soil Conservation Research Laboratory, Morris, MN.

Young, R. A., C. A. Onstad, D. D. Bosch, and W. P. Anderson. 1995. "AGNPS: an agricultural nonpoint source model." In V. P. Singh (Ed.), *Computer models of watershed hydrology*, pp.1011–1020. Water Resources Publications, Highlands Ranch, CO.