Spatial Crop Yield and Soil Water Content: Measurement, Scaling and Topographic Analyses

T. R. Green[a], R. H. Erskine[a,b], A. Martinez[b], M. R. Murphy[a],
L. R. Ahuja[a], J. D. Salas[b] and J. A. Ramirez[b]

[a] *USDA, Agricultural Research Service (ARS), Fort Collins, Colorado, USA* (green@gpsr.colostate.edu)
[b] *Colorado State University, Department of Civil Engineering, Fort Collins, CO, USA*

Abstract: The need to transfer information across a range of space-time scales (i.e., scaling) is coupled with the need to predict variables and processes of interest across landscapes. Agricultural landscapes offer a unique set of problems and space-time data availability with the onset of satellite-based positioning and crop yield monitoring. The present study addresses quantification of the spatial variability of rainfed crop yield and near-surface soil water over a farm field using three methods: 1) geostatistical and fractal analyses; 2) multiple linear regression (MLR) using topographic attributes for explanatory variables; and 3) nonparametric estimation by Spatial Artificial Neural Networks (SANN). Method 1 is useful for scaling the spatial moments of each variable to determine appropriate scales of measurement and management. Methods 2 and 3 take advantage of empirical and process knowledge of topographic controls on water movement and microenvironments, where topographic attributes estimated from a digital elevation model at some scale (10 m by 10 m here) help explain the observed spatial variability in crop yield. Soil water (top 30 cm) displays more random spatial variability, and its dynamic nature makes it difficult to predict in both space and time. Despite such variability, spatial structure is evident and can be approximated by simple fractals out to lag distances of about 400 m. The SANN technique is more flexible than point-to-point parametric correlations, including the use of spatial activation functions for interpolation within a field. Using topographic attributes as input, SANN provides a minimum prediction error of 0.57 relative root mean squared error (RRMSE) for crop yield in 1997, which explains 68% of the spatial variance compared with 48% for MLR. Within field interpolation (adding the easting and northing as inputs) reduced the RRMSE to 0.47 ($R^2 = 0.78$).

Keywords: Spatial variability; Scaling; Fractals; Regression; Neural networks

1. INTRODUCTION

Variability in crop yield stems from nonlinear spatial interactions between numerous factors, including topographic relief, precipitation, soil hydraulic properties, nutrients and organic matter. Such factors cause variability at nested scales that may give rise to a self-similar pattern of variation described by fractal geometry [Mandelbrot 1977; Burrough 1981; Green et al. 1999]. Wood [1995] and Peters-Liddard et al. [2001] used remotely sensed data and modeled soil moisture to demonstrate (multi)fractal behavior. Fractal analyses characterize the different scales of spatial variability, which may be useful for determining the optimum sizes of management zones. There is further need for prediction of spatial patterns of crop yield based on independent spatial data. Landscape topography reveals spatial patterns similar to crop yield; thus computed topographic attributes may help quantify yield patterns.

The field used in this study is part of the Lindstrom farm located in eastern Colorado, USA (40.37N, 103.13W). We collected spatial yield data for winter wheat in 1997 and Foxtail Millet in 1999. The average pan evaporation is approximately 1600 mm per growing season, while the average annual precipitation (1961-90) is only 440 mm [Peterson et al. 2000]. The terrain in northeastern Colorado is generally undulating with aeolian deposits of silt- and sand-sized material mantling sedimentary rock (primarily sandstone) and fluvial deposits of the South Platte River Basin. The relief is relatively pronounced in the study field, where the elevation ranges from approximately 1365 m to 1386 m, with slopes exceeding 6%. The unconsolidated sediment and soils are relatively thick (at least 3 m) with little or no surface expressions of groundwater or perched water in the root zone. Thin calcareous horizons have been sampled at depths of 20-50 cm, but soil horizons are not very pronounced otherwise.

2. METHODS

2.1 Measurements

Crop yield monitoring on the Lindstrom Farm began in 1997, and more concerted efforts to measure the soils, soil-water content and topography began in 1999. The present discussion focuses on spatial measurement of elevation, near-surface water content, and crop yield.

2.1.1 Elevation

Elevation data were collected using dual-frequency, real-time kinematic GPS mounted to an all-terrain vehicle. Raw elevation data (approximately 5m resolution) were interpolated to a 10m by 10m grid digital elevation model (DEM). The experimental variogram, out to a separation distance of 15 meters, was fit to a Gaussian model. Cross-validation of this interpolation method resulted in an RMSE of 0.036 m.

2.1.2 Crop yield

Winter wheat was harvested in July 1997 using a 9-m wide combine head and measured with a yield monitor (MicroTrak Systems Inc.) linked to GPS with satellite (OmniStar™) differential correction. Values from irregular yield polygons were interpolated using Kriging onto the 10m DEM.

Foxtail Millet (hay) was baled in the field. A sample of the hay bales were measured for volume, water content and weight to determine the dry weight per bale. Bale locations were located with a GPS to determine the harvest area.

2.1.3 Surface soil moisture

Soil water contents in the top 30 cm were measured using time domain reflectometry (TDR) in the spatial pattern shown in Figure 1 for each sampling period (typically two days). The TDR probes were inserted using a hydraulic coring device (Giddings Machine Co.) mounted on a 6-wheel ATV (John Deere Gator™) similar to Western and Grayson [1998]. A satellite differential GPS (XRS™ by Trimble) was used for both navigation to sample points and communication with the TraseBE™ TDR instrument (Soil Moisture, Inc.), making rapid measurements possible.

2.2 Data Analyses

2.2.1 Computing topographic attributes

Topographic attributes can be computed using a DEM and GIS software. Estimates of slope and specific contributing area (SCA), defined as the upslope contributing area per unit contour upslope contributing area per unit contour length, were computed using TARDEM [Tarboton, 2000] implementing the D_{inf} method for flow direction [Tarboton, 1997], with and without sink filling. The SCA values were log-transformed, as per Western *et al.* [1999], to remove the extreme skewness in the frequency distribution. The aspect and plan and profile curvatures were computed using TAPES-G [Gallant and Wilson 1996]. The total curvature was computed using a central finite difference.

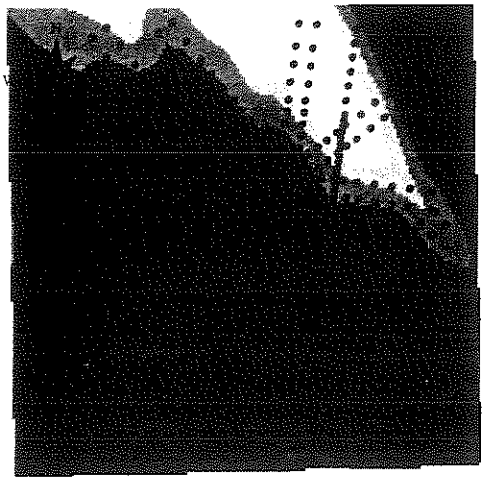

Figure 1. TDR measurement locations (symbols) on a map of elevation (darkest = 1365m, lightest = 1386m).

2.2.2 Experimental variograms

Analyses of experimental variograms and model fits were performed using S+ Spatial Analyst™. Power-law variograms without any nugget were used for all data types (elevation, crop yield, and soil water content) to facilitate consistent fractal analyses below. In the case of elevation, however, Gaussian variograms fit the data best. To test for anisotropy, the procedure was repeated with directional variograms for angles of 10, 55, 100 and 145 degrees from North (see patterns in Figure 1).

2.2.3 Fractal analysis

There are numerous methods to test whether spatial variability follows a fractal behavior [e.g., Barton and Lapointe 1995]. In this study, we used the approach of Burrough [1983]. Briefly, the characteristics of fractional Brownian motion are that sequential increments ($Y(x+h)$-$Y(x)$) must have a Gaussian distribution with zero mean and variance σ^2, and have a variogram that can be adequately described by:

$$2\gamma(h) = E[Y(x+h) - Y(x)]^2 = \sigma^2 h^{2H} \quad (1)$$

where $Y(x+h)$ and $Y(x)$ are values of yield (or other variables) at locations $x+h$ and x, E is the expectation operator, and H is a scaling power parameter known as the Hurst exponent. It follows

from Eq. 1 that a good linear fit on log-log plot of the semivariogram versus lag h indicates fractional Brownian motion. The fractal dimension $D = 2\text{-}H$, where H is the slope of the fitted straight line divided by 2. Smaller values of D indicate that the process is more persistent or smooth.

2.2.4 Multiple linear regression (MLR)

Wheat yield for 1997 was regressed on topographic attributes computed from the DEM. Univariate correlations were also computed for soil moisture for 1999-2000 using standard procedures.

2.2.5 Spatial analysis neural networks (SANN)

The algorithm used herein is a simplified version of the SANN algorithm [Shin and Salas 2000], implemented by Martinez et al. [2001] for spatial interpolation. It consists of four layers, namely: an input layer, Gaussian Kernel Function (GKF) layer, summation layer, and estimation layer, where neurons or nodes between layers are interconnected in the feed-forward direction.

The input layer passes the input coordinate vector to the GKF layer without any weighting. The GKF layer consists of N GKF nodes that represent the receptive field or influence region of each observed vector using a Gaussian transfer or activation function. The response of each GKF node is a function of the Euclidean distance from the center to the input vector, reaching its maximum when the input vector is placed at the GKF node, and decreasing exponentially outward from there. The outputs of the GKF nodes are passed to the summation layer through weighted connections, and outputs from the summation nodes are passed to the estimator layer, which consists of 1 node.

This simplified SANN algorithm has two operational modes: a training mode and an interpolation mode. Training, as defined here, consists of model fitting and validation on separate data sets using multiple realizations of 1000 points sampled from the spatial crop yield of over 6400 points. We used ten different realizations for each set of input (explanatory) variables.

3. RESULTS AND DISCUSSION

3.1 Space-Time Data

The soil water data collected in 1999 is summarized by the histograms in Figure 2. During the early growth stages in June (a), the crop had not yet used much stored water, and the distribution was skewed to the left. By mid-July (b) the distribution was more Gaussian, and late-July rainfall increased soil water in early August (c). Finally, after harvest and little rain, the soils were dry, and the distribution skewed to the right.

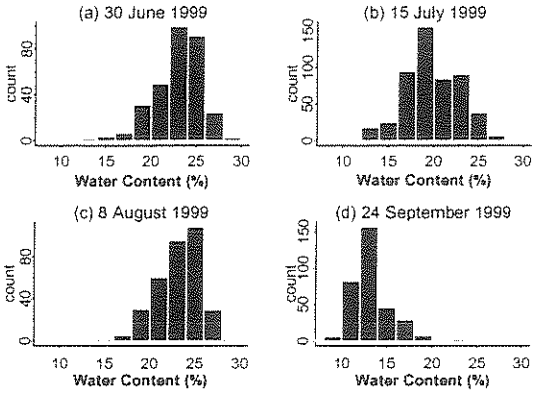

Figure 2. Frequency histograms of measured water content (30 cm TDR) for four sample dates (a-d).

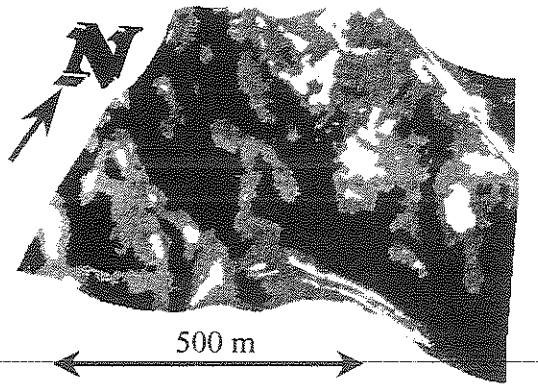

Figure 3. Winter wheat grain yield overlaid on the topography of the quarter section (65 ha). Dark areas are high (up to 7.3 Mg/ha) and light areas are low yielding (< 0.9 Mg/ha).

3.2 Experimental Variogram Analysis

The spatial statistical structures of elevation, crop yield and TDR data are analyzed using experimental variograms and model fits. Experimental variograms include data pairs in a 45 degree view. The Gaussian variogram model fits the experimental variograms of elevation best over the full range due to an apparent sill or periodic behavior. There is significant anisotropy due to the prevailing summit and swale directions toward the SE (Figure 3). Some of this effect could be removed with Universal Kriging, but not using a simple linear trend. Furthermore, non-stationary behavior is expected for this domain size, and required for the power-law variogram model.

Variograms for winter wheat in 1997 are shown in Figure 4. The multiplier a and exponent b of the power law are shown below each variogram. The power-law model fits the omni-directional data (a)

well with its monotonic structure and small nugget. Foxtail millet data for 1999 also displays spatial structure that is fit well by the power-law model (not shown). The omni-directional fit to all data is very good, and gives parameters that are very similar to the 100 degree direction. In fact, the bearing of the baler path was approximately 145°, which may be responsible for much of the observed anisotropy.

TDR variograms also displayed anisotropy, but Figure 5 shows only the omni-directional variograms for each sample date in 1999. The power law fits all but the June 24 sample (a) relatively well, with the most structure and best fits when it was relatively wet (b) and (d).

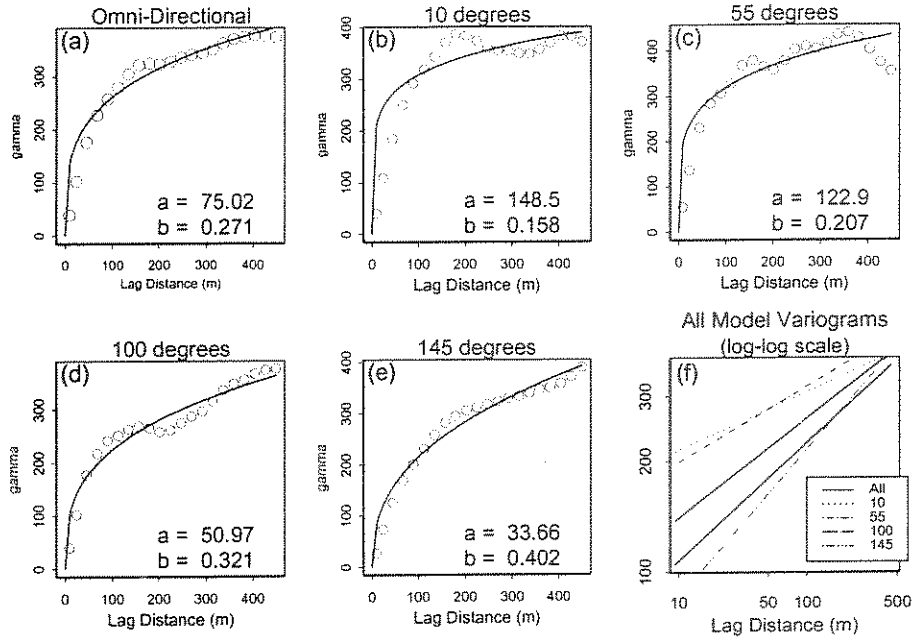

Figure 4. Winter Wheat yield data (1997) showing experimental variogram and model fits using power law (for fractal analysis below) for omni-directional (a) and directional cases (b)-(e). All five model variograms are show together on a log-log plot (f).

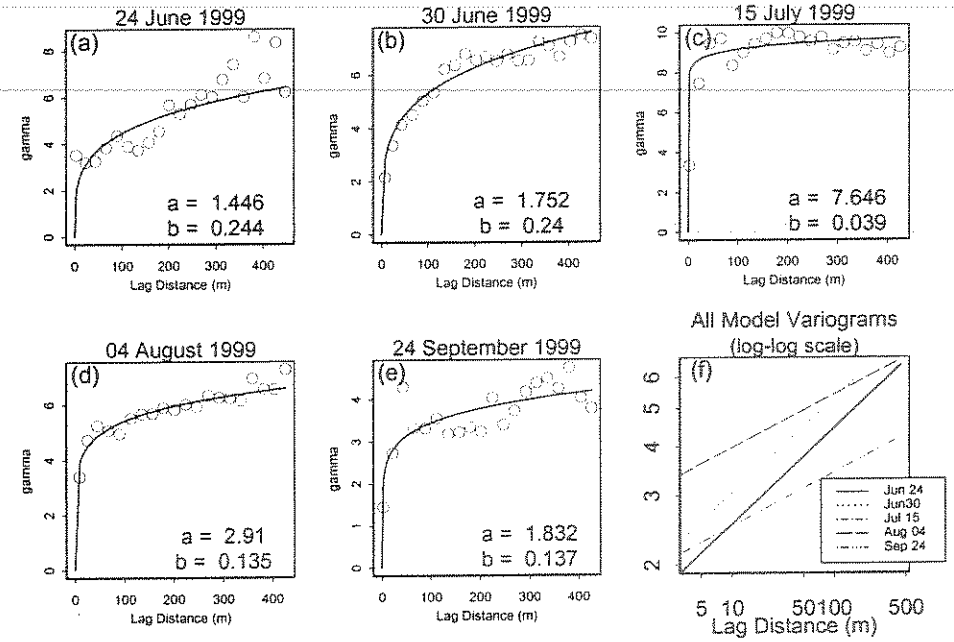

Figure 5. Soil water content for five sample dates (a)-(e), and a composite of all five dates on a log-log plot (f).

3.3 Scaling with Fractals

The omni-directional fractal dimensions range from 1.47 for elevation to 1.98 for soil water in July 1999. For wheat yield in 1997, D=1.86, and for millet in 1999, D=1.83. Both display much less spatial persistence than elevation, but more than water content. Also, Figure 6 shows the variability in D and mean water content with time.

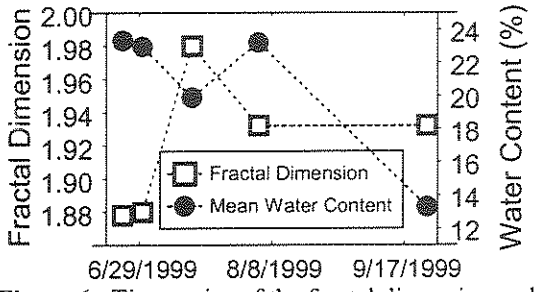

Figure 6. Time series of the fractal dimension and mean values of spatial TDR water content.

3.4 Predicting Crop Yield and Soil Water Content from Topographic Attributes

Table 1 gives detailed results from univariate linear regressions between the TDR data and topographic attributes for each sample date. The correlations are generally low in this semi-arid landscape. The correlation coefficient (r) values for the strongest explanatory variables are plotted versus the mean water content for each date in Figure 7. Spatial water content is most closely related to these topographic attributes during the wettest condition.

MLR and SANN methods were used on the 1997 wheat yield data versus topographic attributes. The prediction errors are given in Table 2 for the best combinations of input (explanatory) variables. These are mean relative RMSE values of prediction from ten realizations of 1000 points for training/correlation and 1000 points for validation. Figure 8 shows that MLR out-performed SANN for one topographic input only, and SANN kept improving with more inputs. The last row in Table 2 also shows the benefit of SANN for interpolation within a field, where the spatial coordinates (x,y) are added as inputs. It is instructive to compute $R^2 = 1 - RRMSE^2$ where RRMSE is the relative root mean squared error. For prediction from topographic attributes alone using five inputs, $R^2 = 0.68$ for SANN and $R^2 = 0.48$ for MLR.

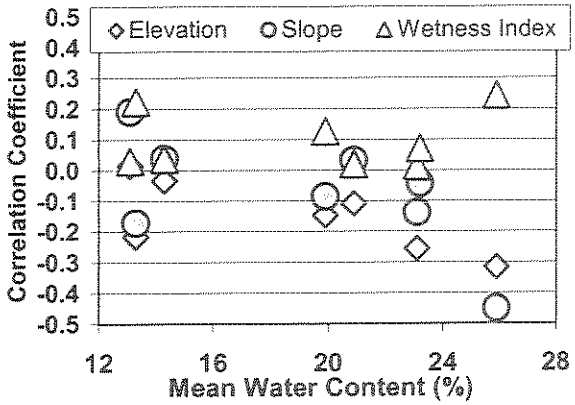

Figure 7. Values of the correlation coefficient (r) between TDR water content and three primary topographic attributes versus the spatial mean of water content for each sampling date (see Table1).

4. CONCLUSIONS

Measurement of spatial crop yield and soil water content described here provides data needed for scaling analyses and testing of spatial regression methods. Although water content data is "noisy" and results in low r values from univariate regressions with topographic attributes, there is spatial structure at the field scale. MLR and SANN methods using topographic attributes proved useful for predicting spatial crop yield, and SANN demonstrated a distinct advantage for multiple input variables.

5. ACKNOWLEDGMENTS

We thank Gilbert Lindstrom for providing access to his farm for data collection. Gale Dunn, Marvin Shaffer, and Jan Cipra helped with yield mapping.

Table 1. Results of univariate regressions of water content on topographic attributes, where **Z** is elevation, **S** is slope, **Pro C** and **Plan C** are profile and plan curvatures, **ln(a)** is specific contributing area, and **W.I.** is the wetness index defined as **ln(a/S)**.

Sample		WaterContent (%)		Correlation Coefficient (r)					
Date	#	mean	St.Dev.	Z	S	ProC	Plan C	ln (a)	W.I.
30-Jun-1999	504	23.1	2.6	-0.26	-0.14	0.19	0.08	0.13	0.01
15-Jul-1999	526	19.9	3.1	-0.15	-0.09	0.00	0.04	0.00	0.13
4-Aug-1999	338	23.2	2.5	-0.06	-0.04	0.08	0.00	0.02	0.07
24-Sep-1999	589	13.3	2.1	-0.22	-0.17	0.11	0.06	0.05	0.22
27-Apr-2000	339	25.9	1.8	-0.32	-0.45	0.29	0.17	0.35	0.24
23-May-2000	578	20.9	2.9	-0.11	0.03	0.03	-0.03	0.00	0.02
20-Jun-2000	591	13.1	2.3	0.01	0.19	0.13	0.05	0.07	0.03
25-Jul-2000	598	14.3	2.3	-0.03	0.04	0.05	0.05	-0.06	0.04

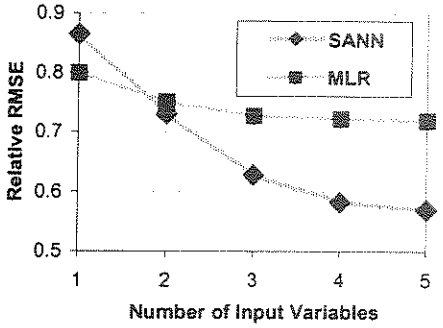

Figure 8. Comparison of the prediction errors of neural networks (SANN) and regression (MLR) versus the number of input (explanatory) variables.

Table 2. Relative RMSE (RRMSE) of prediction values using Multiple Linear Regression (MLR) and Spatial Analysis Neural Networks (SANN). **Z** is elevation (relief), **S** is Slope, **C** is total Curvature, **A** is cosine(Aspect), and **SCA** is ln(Specific Contributing Area), where the subscript denotes local Sinks (**S**) or Filled (**F**).

Input Variables		RRMSE	
Number	Names	SANN	MLR
1	Z	0.931	0.871
	S	0.865	0.800
	C	0.973	0.970
	A	0.997	0.973
	SCA_S	1.013	0.937
	SCA_F	1.003	0.929
2	Z, S	0.730	0.750
	Z, A	0.858	0.830
	S, C	0.808	0.768
	S, A	0.848	0.798
	S, SCA_F	0.770	0.762
3	Z, S, C	0.675	0.736
	Z, S, A	0.655	0.743
	Z, S, SCA_F	0.628	0.728
4	Z, S, C, SCA_F	0.616	0.724
	Z, S, A, SCA_F	0.582	0.723
	Z, S, C, A	0.604	0.730
5	Z, S, C, A, SCA_F	0.570	0.720
5 + X, Y	Z, S, C, A, SCA_F	0.472	0.719

6. REFERENCES

Barton, C.C. and R.R. LaPointe, *Fractals in Petroleum Geology and Earth Processes*, Plenum, New York, 1995.

Burrough, P.A., Fractal dimensions of landscape and other environmental data. *Nature* 294: 240-242, 1981.

Burrough, P.A., Multiscale sources of spatial variation in soil, I. The application of fractal concepts to nested levels of soil variation. *J. Soil Sci.* 34, 599-620. 1983.

Gallant, J.C. and J.P. Wilson, TAPES-G: A grid-based terrain analysis program for the environmental sciences, *Computers & Geosci.*, 22(7):713-722, 1996.

Green, T.R., M.H. Nachabe, L.R. Ahuja, M.R. Murphy, J.C. Ascough II and M.J. Shaffer, Preliminary fractal analysis of crop yield and experimental design for modeling space-time variability under dryland agriculture, *Proc. 19th AGU Hydrology Days*, 16-20 August 1999, Fort Collins, CO, H.J. Morel-Seytoux, ed., pp. 187-198, 1999.

Mandelbrot, B., *Fractals: Form, Chance, and Dimension*. San Francisco: W. H. Freeman & Co., 1977.

Martinez, A., J.D. Salas and T.R. Green, Testing and Application of Spatial Analysis Neural Networks: Sensitivity to Structural Parameters, *Proc. 21st AGU Hydrology Days*, 2-5 April 2001, Fort Collins, CO, J.A. Ramirez, ed., pp. 176-186, 2001.

Peters-Lidard, C.D., F. Pan and E.F. Wood: A re-examination of modeled and measured soil moisture spatial variability and its implications for land surface modeling. In: Nonlinear Propagation of Multi-Scale Dynamics through Hydrologic Subsystems, M. Sivapalan, et al. (eds.): *Advances in Water Resources* (Special Issue), in press, 2001.

Peterson, G.A., et al., Sustainable dryland agroecosystem management, Tech. Bull. TB00-3, Agric. Exp. Stn., Colorado State Univ., Fort Collins, CO, 83 pp.

Shin, H.S and J.D. Salas, Spatial analysis of hydrologic and environmental data based on artificial neural networks, In: Artificial Neural Networks in Hydrology, *Water Science and Technology* Vol. 36, R.S. Govindaraju and A.R. Rao (eds.), Kluwer Academic Publishers, Amsterdam, 2000.

Tarboton, D.G., A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resources Research*, 33(2): 309-319, 1997.

Tarboton, D.G., TARDEM: a suite of programs for the analysis of digital elevation data, http://www.engineering.usu.edu/dtarb/tardem.html, May5, 2000.

Western, A.W., and Grayson, R.B., The Tarrawarra data set: Soil moisture patterns, soil characteristics and hydrological flux measurements. *Water Resources Research*, 34(10): 2765-2768, 1998.

Western, A.W., R.B. Grayson, G. Blöschl, G.R. Willgoose and T.A. McMahon, Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resources Research*, 35: 797-810, 1999.

Wood, E.F., Scaling behaviour of hydrological fluxes and variables: empirical studies using a hydrological model and remote sensing data, in Scale Issues in Hydrological Modelling, J. Kalma and M. Sivapalan (eds.), Wiley & Sons, Chichester, pp. 89-104, 1995.