6-2008

Genetic algorithm for parameter and scale selection to predict soil moisture patterns

Lingyuan Yang
Iowa State University

Amy L. Kaleita
Iowa State University, kaleita@iastate.edu

Genetic algorithm for parameter and scale selection to predict soil moisture patterns

Abstract
Soil moisture is a critical component of hydrological processes, and its spatio-temporal distribution depends on many geographical factors (such as elevation, slope, and aspect, etc.). Each of the factors is likely influential over a different scale and to a different degree. Near-surface soil moisture data were collected across a working 10-ha field southwest of Ames, IA in growing seasons of 2004 to 2007. A genetic algorithm is developed to compare geographical factors to the moisture patterns over a range of scales. The genetic algorithm will develop a model in which each factor is computed over a different scale for use in prediction of reference variable. Optimized scales for each parameter are arrived at through successive generations, including crossover and mutation of this evolutionary algorithm. Using this approach, not only are the primary influential relationships uncovered, but the most appropriate scale for comparison to moisture pattern is identified. The results of this analysis can be used to predict the spatio-temporal patterns of soil moisture across a region a priori.

Keywords
Evolutionary computation, scaling, soil water content

Disciplines
Agriculture | Bioresource and Agricultural Engineering

Comments
This is an ASABE Meeting Presentation, Paper No. 083733.

2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation

Paper Number: 083733

Genetic algorithm for parameter and scale selection to predict soil moisture patterns

Lingyuan Yang, Graduate Research Assistant

Department of Agricultural & Biosystems Engineering, Iowa State University

Amy Kaleita, Assistant Professor

Department of Agricultural & Biosystems Engineering, Iowa State University

Written for presentation at the
2008 ASABE Annual International Meeting
Sponsored by ASABE
Rhode Island Convention Center
Providence, Rhode Island
June 29 – July 2, 2008

Abstract. *Soil moisture is a critical component of hydrological processes, and its spatio-temporal distribution depends on many geographical factors (such as elevation, slope, and aspect, etc.). Each of the factors is likely influential over a different scale and to a different degree. Near-surface soil moisture data were collected across a working 10-ha field southwest of Ames, IA in growing seasons of 2004 to 2007. A genetic algorithm is developed to compare geographical factors to the moisture patterns over a range of scales. The genetic algorithm will develop a model in which each factor is computed over a different scale for use in prediction of reference variable. Optimized scales for each parameter are arrived at through successive generations, including crossover and mutation of this evolutionary algorithm. Using this approach, not only are the primary influential relationships uncovered, but the most appropriate scale for comparison to moisture pattern is identified. The results of this analysis can be used to predict the spatio-temporal patterns of soil moisture across a region a priori.*

Keywords. Evolutionary computation, scaling, soil water content

(The ASABE disclaimer is on a footer on this page, and will show in Print Preview or Page Layout view.)

Introduction

Soil moisture is a critical component of hydrological, biochemical, and geomorphologic processes, as well as crop growth. In recent years, much attention has been focused on the spatial variability of surface soil moisture at various scales (e.g. Starks et al., 2006). There are a variety of applications, including studying the spatial variability of soil moisture in order to better understand the nature of soil moisture within a satellite pixel (footprint) (Famiglietti et al., 1999).

There are two major measurement approaches for capturing the soil moisture variability, ground-based measurement and remote sensing. Studies of soil moisture patterns using ground-based measurements require detailed and accurate information, which can be time- and resource-consuming to generate with *in situ* techniques due to the spatial and temporal variability of soil moisture. Remote sensing, used for measuring soil moisture, gives the spatial average of soil moisture over a large area known as a footprint, in which only predominant soil and vegetation types are usually used for calibration purposes (Mohanty and Skaggs, 2001). It requires some form of groundtruthing using *in situ* techniques to convert the remotely sensed signal into a soil moisture estimate and error accounting; this process is known as validation or calibration. But the interpretation and validation of the remotely sensed signal is hampered by the high spatial variability of soil moisture (Cosh et al., 2004; Western et al., 2004), as well as by the mismatch in scales between satellite footprints and a ground sample (Western and Blöschl, 1999), the latter is usually 8-10 orders of magnitude smaller than the former.

Soil moisture variability has been related to many factors. Mohanty and Skaggs (2001), studying data from the Southern Great Plains Hydrology 1997 (SGP97) field experiment, suggest that further studies are required to understand soil moisture dynamics as related to soil, vegetation, and topographic features. Jacobs, et al. (2004) summarized how soil heterogeneity, land cover, topography affect soil moisture content at different scales and through different mechanisms during the 2002 Soil Moisture Experiment field study (SMEX02). It was stated that soil heterogeneity affects soil moisture content thorough variations in soil texture, soil water holding capacity, soil color, soil water retention, and pixel- and pore-scale hydraulic properties; that land cover influences runoff, interception and evapotranspiration processes, and in turn, the soil moisture dynamics; and that topography plays a dominant role for small catchments and hillslopes (Jacobs, et al., 2004). Hawley et al. (1983) indentified that topography as the most significant feature in soil moisture spatial structure at a field scale.

Most of these studies mainly focused on the qualitative explanation of the relationship between soil moisture variability and the influential factors. However, there are few studies which investigate the soil moisture variability as a function of the variability of those influential factors affecting the soil moisture. Neither are the scales at which the influential factors have impact on soil moisture considered. Since not only the influential factors are considered, but the optimal combination of scales for these factors is included, it is difficult to identify the best factors at the best scales for the functions. Therefore, an effective optimization method is needed.

Genetic Algorithms (GAs) are adaptive heuristic search algorithms premised on the evolutionary ideas of natural selection and genetics. The basic concept of GAs is designed to simulate the Darwinian concept of “survival of the fittest.” GAs are computationally simple yet powerful to provide robust search for difficult combinational search problems in complex spaces, without being stuck in local extremes (Goldberg, 1989a; Tang, et al., 1999; Steward, et al., 2005). Therefore, GAs are powerful alternative tools to traditional optimization methods (Goldberg, 1989a).

In order to examine how the variability of soil moisture patterns can be attributed to easily observed static variables that drive soil moisture evolution, the objective of this study is to: 1) identify and represent the common spatial pattern of near-surface soil moisture at the field scale in different growing seasons; and 2) develop and use a genetic algorithm to investigate the soil moisture variability as a function of the influential topographic factors (slope, aspect, elevation and curvature) and their interactions at various scales. In doing so, we hope to identify not only the most important topographic factors for soil moisture pattern development, but also the operative scales of each of those factors.

Location and Data

Distributed near-surface soil moisture data was collected across an agricultural field, Brooks Field, just southwest of Ames, IA, during the 2004, 2005 and 2007 growing seasons. Brooks is a 10-ha corn-soybean rotation field with moderate topographic variation. According to NRCS web soil survey (http://websoilsurvey.nrcs.usda.gov/app/), Brooks has 5 soil types: Nicollet loam, Harps loam, Webster clay loam, Clarion loam, and Canisteo clay loam. There are a total of 25 days (from mid-May to early July) in 2004, 30 days (from early-May to Mid-August) in 2005, and 29 days (late May to Mid-August) of data in 2007 available for analysis. Three volumetric moisture content readings for 0-6 cm depth were taken using a Theta probe moisture meter at each of 78 locations, which include 42 regular grid nodes with a 50-m interval and two transects with a 5-m interval.

Elevation data was obtained by LiDAR survey at a 2-m resolution at Brooks Field. These data were used to create maps of topographic indices at scales from 4 to 120 meters. 2-m DEM (digital elevation model) was generated in ArcGIS, then aggregated to larger scales (4 m, 8 m …… 120 m). Slope, aspect, curvature at the various scales were calculated from the individual DEMs. Sampling locations for soil moisture are shown in Figure 1, along with elevation data and soil types.

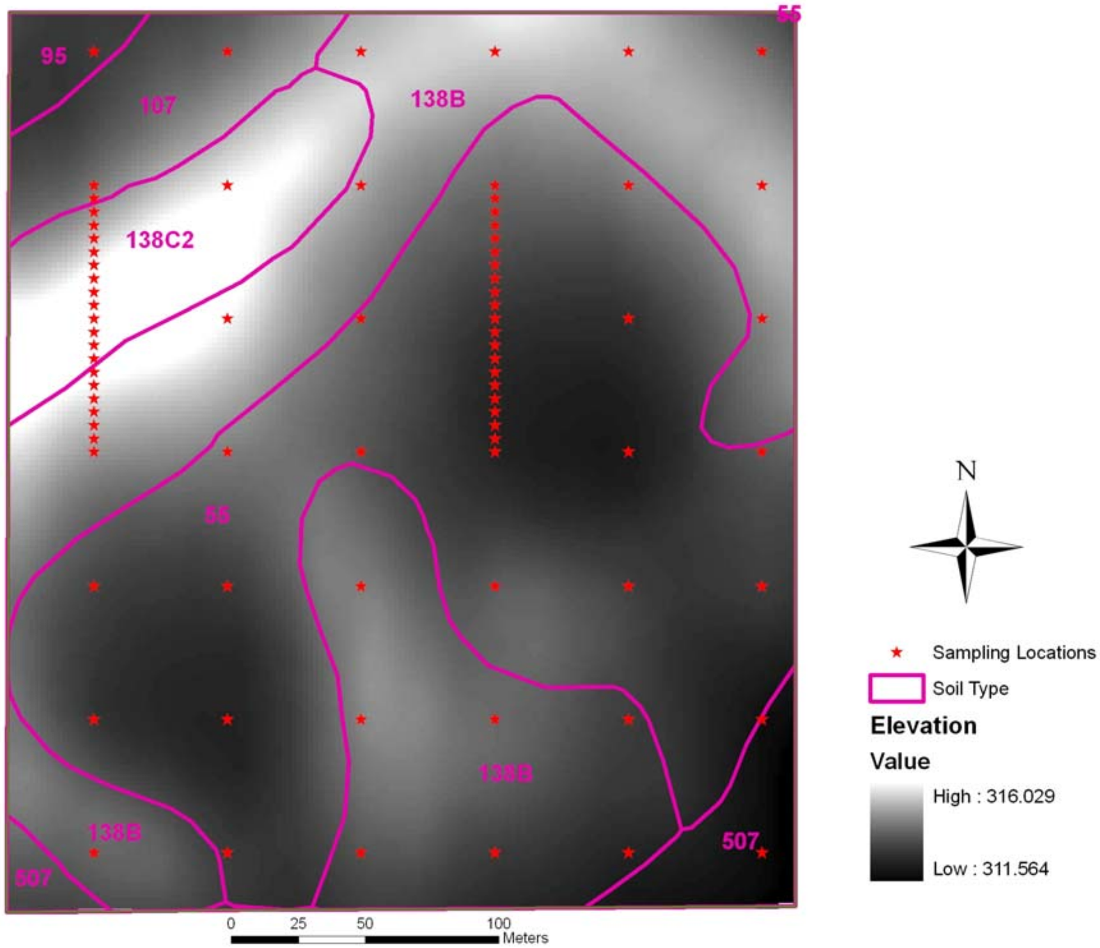

Figure 1. Soil Moisture Sampling Locations at Brooks Field. Elevation data (shading) is given in meters. Soil type indices: 55: Nicollet loam, 1 – 3% slopes; 95: Harps loam, 1 – 3% slopes; 107: Webster clay loam, 0 – 2% slopes; 138B: Clarion loam, 2 – 5% slopes; 138C2: Clarion loam, 5 – 9% slopes, moderately eroded; 507: Canisteo clay loam, 0 – 2% slopes.

Methods and Analysis

Spatial pattern of soil moisture

The surface soil moisture dataset was analyzed using the methods provided by Vachaud et al. (1985) to determine the spatial patterns of soil moisture at the field scale in the monitoring period (Yang, et al., 2007). The mean relative difference $\bar{\delta}_i$ and its standard deviation $\sigma\left(\bar{\delta}_i\right)$ were calculated. These are defined by equations 1 through 3:

$$\delta_{ij} = \frac{\theta_{ij} - \bar{\theta}_j}{\bar{\theta}_j} \qquad (1)$$

$$\bar{\delta}_i = \frac{1}{m}\sum_{j=1}^{m}\delta_{ij} \qquad (2)$$

$$\sigma(\overline{\delta}_i) = \left[\sum_{j=1}^{m} \frac{\delta_{ij} - \overline{\delta}_i}{m-1} \right]^{1/2} \qquad (3)$$

where , θ_{ij} is the moisture at the i th location on the j th sampling occasion, $\overline{\theta}_j$ is the field average of all θ_{ij} on the j th sampling occasion, and m is the total number of sampling occasions. Locations with $\overline{\delta}_i$ close to zero indicate sites having average soil moisture close to field mean, whereas Locations with negative mean relative difference underestimated the field average soil moisture, which are relatively drier, and locations with positive ones overestimated the field average, which are relatively wetter. Locations with small $\sigma(\overline{\delta}_i)$ are considered to be temporally stable, because they have relatively consistent behavior over time with respect to the field average soil moisture content.

It is notable that the ranges of mean relative difference of soil moisture in Brooks Field vary between -14.5% and 24.0%, and one standard deviation ($\sigma(\overline{\delta}_i)$) was fairly uniform changing from 6.6% to 15.6% (shown in Figure 2). In general, the relatively drier locations were more stable than the relatively wetter locations.

The mean relative difference will be used as the intended output of GA described below.

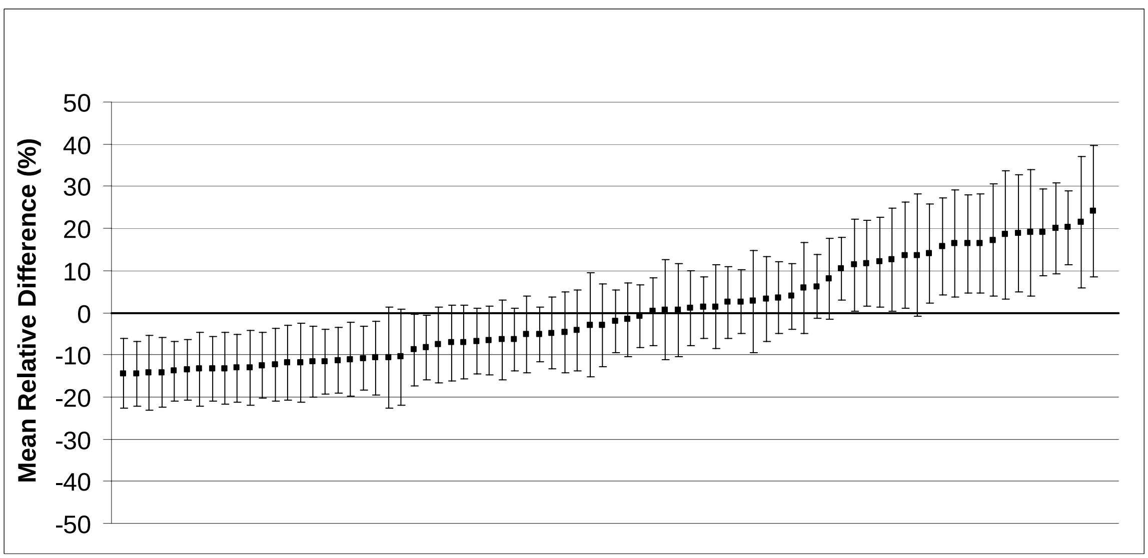

Figure 2. Ranked mean relative difference of soil moisture with one standard deviation error bars for three growing seasons in 2004, 2005, and 2007.

Genetic algorithm structure

In order to determine the role of topography in the development of the common spatial distribution of soil moisture in the Brooks field, a genetic algorithm loosely based on the range operator-enabled GA developed by Steward et al. (2005) and Kaleita et al. (2006), was employed. The ultimate goal of the GA was to develop a model for prediction of mean relative difference based on a combination of important topographic indices computed at the most appropriate scale, respectively.

The technical details of the GA are as follows:

- Chromosome: Chromosomes are the abstract representations of candidate solutions to an optimization problem evolving toward better solutions. In this study, solutions are the predictive models for mean relative difference based on topographic data. The chromosomes were composed of 18 genes that described variables which were selected by MLR (Multi-Linear Regression) models. There are two parts of genes in each chromosome. The first part, including 4 genes, was encoded in integers representing: (1) scale of slope, (2) scale of aspect; (3) scale of elevation; (4) scale of curvature; the second part, including 14 genes, was encoded in binary string representing: (5) – (18) whether the variables (slope, aspect, elevation, and curvature), their squares, and their interactions (or cross products) were selected in the model or not. “1” represented selected, and “0” meant not selected.

- Population size: One of the advantages of genetic algorithms over traditional optimization and search procedures is that GAs search from a population of solutions, not a single solution. The rule of thumb, suggested by Goldberg (1989b), a population size approximately equal to the chromosome length. For our study with a chromosome length of 34 bits, a population size of 36 was used.

- Selection: During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. In this study, the population of 36 chromosomes was ranked by fitness value. Absolute fitness replacement was implemented by sorting the population by fitness and discarding the less fit half of the population. The upper half was then replicated to form a new lower half, processed for the following crossover and mutation.

- Crossover and mutation: The next step is to generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination), and/or mutation. For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a "child" solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each child, and the process continues until a new population of solutions of appropriate size is generated. At each generation, the population of 36 individuals was sorted by fitness, and the less fit half of the population was discarded. The upper half was then replicated to form a new lower half, and children were generated using single-point crossover. Each child was selected for mutation. The mutation rate (the number of genes per individual that can mutate) was variable, starting with a very high mutation rate in order to make the program search in a broader region and then using a low mutation rate near the end. When an individual was mutated, either the first part of the child, or the second part, or both of them was mutated.

- Fitness: A fitness value is a particular type of objective function value that quantifies the optimality of a solution (that is, a chromosome) in a genetic algorithm so that that particular chromosome may be ranked against all the other chromosomes. The fitness for an individual was the RMSECV (Root Mean Square Error of Cross Validation) associated with that individual calculated through a MLR cross-validation procedure. Analysis was replicated seventeen times with randomly generating seventeen different starting populations. Cross validation was accomplished using leave-one-out procedure. Model performance was measured using RMSECV. The model with the minimum RMSECV in the population was the best one.

- Stopping Criteria: The generational process is repeated until a stopping condition has been reached. This genetic algorithm stopped when one of these two conditions was satisfied: first, the program terminated when more than half of the population contained duplicate individuals; second, the process stopped if the generation number was greater than 200.

Results and Discussion

The multi-linear regression was applied between mean relative difference and the topographic factors as well as their interactions. MLR analysis gave models with a good fit to the mean relative difference. For the selected best 20 models from 20 replications, R^2 values were all above 0.85 (shown in Table 1 and Figure 3). The RMSECV of models was ranging from 4.10 to 4.30% (Table 1).

The times of being selected by the models for topographic variables and their interactions were different for each one of them (as Table 2). Therein, the slope, curvature, square of aspect, square of curvature, the cross product of slope and aspect, the cross product of slope and elevation, the cross product of slope and curvature, the cross product of aspect and elevation, and the cross product of elevation and curvature, were mostly selected. And they were selected by 18, 16, 16, 18, 16, 18, 17, 18, and 17 times out of 20 models, respectively.

GA identified that the most selected scales for slope, aspect, elevation, and curvature were 28, 16, 28, and 72 meters, summarized as in table 1 and figure 4. Although the elevation was only selected 5 times out of 20 best models, the interactions between elevation and all three other factors (slope, aspect, and curvature) were most frequently selected by the models. The results suggested that elevation at 28-m scale might have less impact on the soil moisture pattern by itself, but it highly influenced the soil moisture pattern with the combination with other topographic factors. As the results indicated that not only the slope at 28 m was important in predicting the soil moisture pattern, but the interactions of slope at this scale with all other factors were important, too. Curvature and curvature square at 72 m were also indentified as important factors for soil moisture pattern prediction.

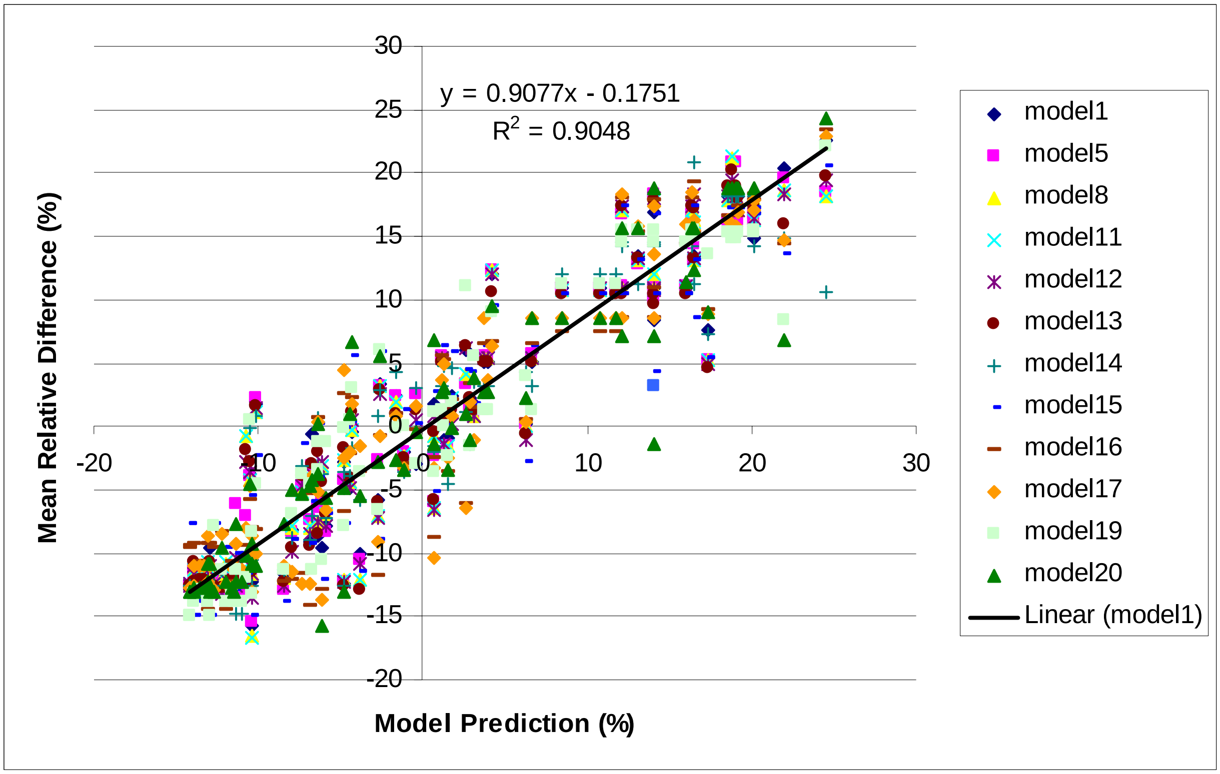

Figure 3. The best model predicted mean relative differences by Multi-Linear Regression (MLR) and the measured mean relative difference in Brooks Field.

Table 1. Fitness values of the model and scales for each topographic factor of the best 20 models identified by GA

Fitness of the Model (%)	R-square (Model Prediction vs. Measured MRD)	Selected scales of slope (m)	Selected scales of aspect (m)	Selected scales of elevation (m)	Selected scales of curvature (m)
4.10	0.90	28	32	28	72
4.10	0.90	28	32	28	72
4.10	0.90	28	32	28	72
4.10	0.90	28	32	28	72
4.12	0.89	28	20	28	72
4.12	0.89	28	20	28	72
4.13	0.89	28	16	28	72
4.14	0.89	28	16	28	72
4.14	0.89	28	16	28	72
4.14	0.89	28	16	28	72
4.14	0.89	28	16	28	72
4.20	0.89	28	8	28	72
4.21	0.89	28	8	28	72
4.27	0.89	28	48	72	84
4.29	0.85	28	104	28	72
4.29	0.88	116	8	16	72
4.29	0.89	116	4	88	72
4.29	0.89	116	4	88	72
4.30	0.88	28	80	60	84
4.30	0.87	116	80	52	56

Table 2. Times of being selected for the topographic factors and their interaction in 20 best models identified by GA

S	A	E	C	S^2	A^2	E ^2	C^2	S*A	S*E	S*C	A*E	A*C	E*C
18	10	5	16	7	16	6	18	16	18	17	18	3	17
Note: S: slope; A: aspect; E: elevation; C: curvature													

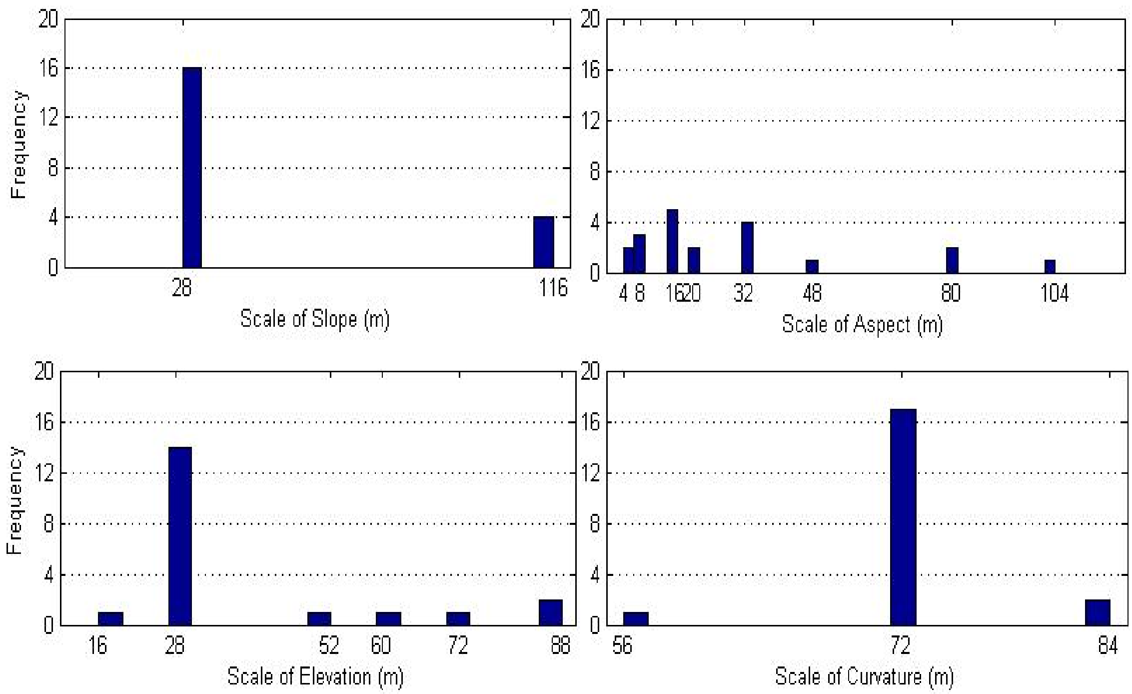

Figure 4. Histograms of selected scales for each topographic factor in 20 best models identified by GA

Conclusions

There is common spatial pattern of soil moisture in the study field. Mean relative difference can represent field scale most common soil moisture pattern, which explained 61% of total variance. The mean relative difference can be used to represent the relative field soil moisture conditions, relatively drier or wetter, or under the average condition. And it summarizes the spatial pattern through the observing seasons, rather than each day.

The genetic algorithm developed in this study not only found the optimal scales for all the influential topographic variables, but identified the most influential variables in the regression models. The most common spatial pattern of soil moisture at Brooks Field was highly influenced slope, curvature, aspect square, curvature square, as well as the interactions of slope and aspect, of slope and elevation, of slope and curvature, of aspect and elevation, of elevation and curvature, and each factor was influential at different scale. The optimal scales found out in this field indicated a coarse resolution elevation data would be sufficient to predict the spatial-temporal pattern of soil moisture at Brooks Field, and the finer resolution Lidar elevation data (for example 2 m we used in this study) is not necessary, which would save a lot of resources.

Acknowledgements

This work was supported by NASA grant NNG06GC63G from the Land Surface Hydrology program, and by the Iowa State University Environmental Science Fellowship and the Iowa Space Grant Consortium Program for Space-borne and Earthbound System Sustainability.

References

Cosh, M.H., Jackson, T.J., Bindlish, R., Prueger, J.H., 2004. Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. Remote Sensing of Environment 92: 427-435.

Famiglietti, J.S., Devereaux, J.A., Laymon, C.A., Tsegaye, T., Houser, P.R., Jackson, T.J., Graham, S.T., Rodell, M., van Oevelen, P.J., 1999. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. Water Resources Research 35 (6): 1839-1851.

Goldberg, D. E. 1989a. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, Mass.: Addison-Wesley Publishing Co., Inc.

Goldberg, D. E., 1989b, Sizing populations for serial and parallel genetic algorithms. In Proc. 3rd Int. Conference on Genetic Algorithms, 70-79, San Mateo, Calif: Morgan Kaufman.

Hawley, M.E., Jackson, T.J. and McCuen, R.H., 1983. Surface soil moisture variation on small agricultural watersheds. Journal of Hydrology 62: 179–200.

Johnson, R. and Wichern D., 2002. Applied Multivariate Statistical Analysis, 5th ed., Prentice-Hall, Upper Saddle River, N. J.

Kaleita, A.L., Steward, B.L., Ewing, R.P., and Ashlock, D.A., Westgate, M.E., Hatfield, J.L., 2005. Novel analysis of hyperspectral reflectance data for detecting onset of pollen shed in maize, Transactions of the ASAE 49 (6): 1947-1954.

Mohanty, B.P., Skaggs, T.H., 2001. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. Advances in Water Resources 24: 1051-1067.

Starks, P.J., Heathman, G.C., Jackson, T.J., Cosh, M.H., 2006. Temporal stability of soil moisture profile. Journal of Hydrology 324: 400-411.

Steward, B.L., Kaleita, A.L., Ewing, R.P., and Ashlock, D.A., 2005. Genetic algorithms for hyperspectral range operator selection, ASABE paper No. 053063. St. Joseph, Mich.: ASABE.

Tang, L., Tian, L.F., and Steward, B.L., 2000, Color image segmentation with genetic algorithm for in-field weed sensing, Transactions of the ASAE 43: 1019-1027.

Western, A.W. and Blöschl, G., 1999. On the spatial scaling of soil moisture. Journal of Hydrology 217: 203-224.

Western, A.W., Zhou S., Grayson, R.B., McMohon, T.A., Blöschl, G., Wilson, D.J., 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. Journal of Hydrology 286: 113-134.

Yang, L. and Kaleita. A.L., 2007. Understanding spatio-temporal patterns of soil moisture at the field scale, ASABE paper No. 072108. St. Joseph, Mich.: ASABE.