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: 097207

Multidimensional Tool for the Visualization of Spatiotemporal Variance in Soil Moisture

Brett M. Nekolny, Mechanical Engineering

Graduate Research Assistant, 1620 Howe Hall, Ames, IA, 50011, bnekolny@iastate.edu

Kristin L. Crawford, Aerospace Engineering

Research Assistant, 1620 Howe Hall, Ames, IA, 50011, kleigh@iastate.edu

Joseph S. Goering, Mechanical Engineering

Research Assistant, 1620 Howe Hall, Ames, IA, 50011, jgoering@iastate.edu

Stephanie S. Kaphingst, Mechanical Engineering

Research Assistant, 1620 Howe Hall, Ames, IA, 50011, stephkap@iastate.edu

Amy L. Kaleita, Agriculture and Biosystems Engineering

Associate Professor, 205 Davidson Hall, Ames, IA, 50011, kaleita@iastate.edu

Eliot H. Winer, Mechanical Engineering

Associate Professor, 1620 Howe Hall, Ames, IA, 50011, ewiner@iastate.edu

Written for presentation at the
2009 ASABE Annual International Meeting
Sponsored by ASABE
Grand Sierra Resort and Casino
Reno, Nevada
June 21 – June 24, 2009

Mention any other presentations of this paper here, or delete this line.

Abstract. *With water a precious resource, it is important to understand factors affecting soil moisture. Current research focuses on understanding this relationship; unfortunately these methods are specialized in their applications or overwhelm the user with information making correlations difficult to comprehend. Often, numerical results provide understanding of prominent correlations but miss subtle relationships, hindering subsequent decisions. This project aims to develop a decision making tool combining numerical analysis with visualization techniques to provide the user with the* *information to analyze soil moisture's spatial and temporal variability. Current work has shown that self-organizing maps are effective for displaying comprehendible relationships to the user.*

The abstract is often the only part of the paper to be read, so include your major findings in a useful and concise manner. Include a problem statement, objectives, brief methods, quantitative results, and the significance of your findings. The abstract should be no more than 250 words long.

Keywords. Self-Organizing Map, Spatial Variability, Evapotranspiration, Clustering.

Introduction

The ability to easily understand the relationship between soil moisture patterns and landscape properties is of great interest to both industry and the scientific community. While previous research has given some ability to understand these soil moisture patterns, the models to do so are complicated; they require extensive input data, and the outputs difficult to comprehend. One opportunity for improvement on this research can be harnessed using advanced multidimensional analytical and visualization techniques as means for enhancing and building the understanding the data. Analytical techniques are instrumental in extracting relationships from within complex data sets, but these relationships can be further uncovered and understood quicker and easier by accompanying such analytical tools with state-of-the-art visualization research.

Keeping in mind the objective of modeling and predicting soil patterns and properties, there have been many researchers who have helped expand this cause. This paper describes the beginnings of a tool that combines some of these methods to develop a decision support device which can guide a user in up-scaling point data or downscaling areal average data in order to develop high-quality maps of soil moisture at an appropriate resolution for the user's application. A tool like this would aide in designing better ground-based observation systems, and will facilitate bridging the gap between limited point-based observations to satellite pixel-scale average areas. Some of the methods which have promise to continue expanding the knowledge of soils are PCA, SOMs, topology preserving neural networks, and remote sensing.

Principal component analysis (PCA) is an analytical tool that has been widely used in distilling large datasets, including environmental systems data. For example, Kooch et al. (2008) used PCA to determine the soil factors most influential in determining overall soil type. The success shown in this study and other PCA research demonstrates its utility as an analytical tool especially in determining the most prevalent factors influencing soil moisture. Unfortunately, in higher dimensional data, some accuracy is lost and relationships can go unseen without the visual aid of other techniques. Therefore, new techniques incorporating advanced visualization tools will add significant value to more traditional analytical techniques like PCA.

One such visualization technique is the self-organizing map (SOM), developed by Kohonen (1989). This technique uses neural networks to construct simple visualizations of high-dimensional data. From the successes of Mele and Crowley (2008) and Kothari and Shafiqul (1999), we have examples of self-organizing maps being applied soil hydrological data. These researchers have shown meaningful correlations within the data but were unable to develop dependable methods for varying inputs. It is possible to expand their research to apply these methods to classifying the soil moisture in a given land area using data gathered from land and ground based observations.

Kiviluoto (1996) and Bação et al. (2004) used SOMs in landscape analysis retaining the topological information of spatial data, rather than assuming that data points were not spatially arranged. This method, GEO-SOM, is a modification of the traditional SOM. It also helps eliminate the confusion surrounding the spatiality of a SOM and the spatiality of the given plot of land being analyzed. This method will be used to further classify the soil moisture clusters and allow for specific ground based validation locations.

The most powerful approach to the problem of understanding the factors affecting soil moisture is likely a visual approach with a data mining method, and future work will show whether simply a visual approach is sufficient or if a combined analytical approach is needed. A specific example of an analytical approach is shown in the research of Annas, Kanai, and Koyama

(2007) have combined previous research areas to design a tool for classifying fire risk in forest regions. Their method uses both PCA and SOMs to build an instrument with the benefits of an analytical technique to draw out correlations in a large and complex data set, and visualization maps to display the strongest of these correlations extracted by the statistical method. While the details and problems of fully combining these tools have yet to be researched, such as the indistinctness of the SOM clusters showing correlations in the data, the concept has proven to be a very flexible method with the ability of correctly displaying the complexity of the data.

Despite the success of previous research, many challenges and opportunities remain. Many of the previous researchers have performed promising research in specialized areas, and some work has even bridged the gaps between these specialized areas, but more investigation needs to be done to determine the most promising combination of techniques. The intent of this research is to introduce self-organizing maps as a soil moisture classifier and identifier, with the end goal providing a decision making tool.

Methods

Distributed near-surface soil moisture data was collected across an agricultural field, Brooks Field Figure 1, just southeast 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 Canisted 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 42 locations, which were distributed on a regular grid with a 50-m interval. Elevation data was obtained by LiDAR survey at a 2-m resolution. Additionally, soil chemistry data (nitrogen content, carbon content, pH levels) at X sites were available from work done in the same field as part of a different project. Soil chemistry data was used in place of soil textural information because although the latter has a direct influence on soil hydrologic behavior, data on the former was already available for our study site, and is indirectly linked to soil texture.

Figure 1: Brooks Field

This research has focused on using principle component analysis and self-organizing maps to generate and understand meaningful trends between multiple variables with a goal of dimensionality reduction. In order to incorporate all the datasets into the training of the map it was necessary to consolidate the data into collocated points, which required interpolation of soil chemistry data to the moisture sampling locations. Interpolation was executed using Kriging (Malvic and Durekovic, 2003). For the elevation data, the nearest elevation data point to the soil moisture sampling locations was used.

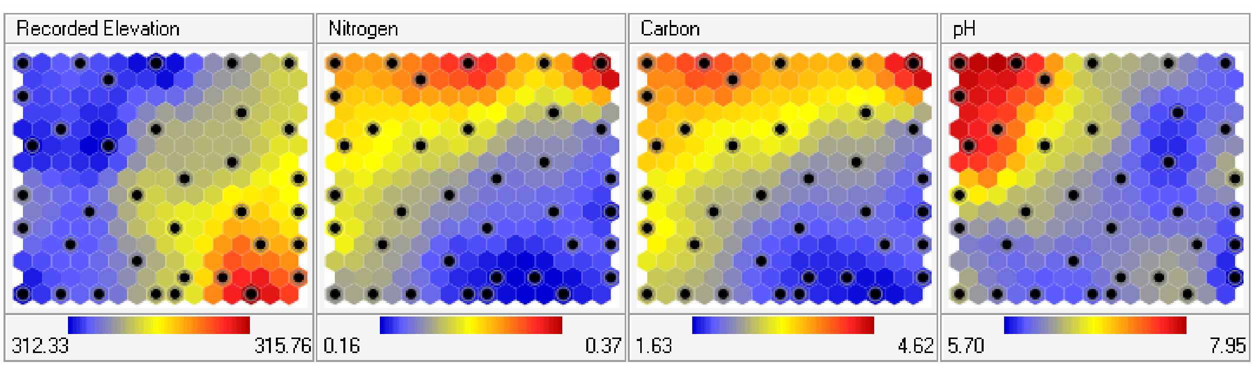
While evaluating the benefits and possibilities of PCA, exploration into the potential of self-organizing maps (SOMs) ensued. We are using self-organizing maps to explore the trends between the variables involved in various days' soil moisture levels in a visual manner. We are also using these maps to cluster the influences affecting soil moisture, such as topographical influences, soil chemistry influences, and will expand these inputs to more detailed topographical details and electrical conductivity data. Lastly, we are using these maps to cluster the soil moisture patterns for various seasons to observe the variability of the soil moisture levels and locations from season to season.

The raw data maps are being used to visualize the variability in 'wet' and 'dry' day's trends in relation to the input variables, and justification for some of these trends will be explained later. The soil moisture observations classified as 'wet' and 'dry' were selected based upon the precipitation levels during and prior to the chosen days. The chosen observations were of two high precipitation days, and two less than average dry condition days. These two variations in moisture levels can provide a basis for comparison of the trend between elevation and moisture in variable conditions

As explained the SOM can be used to cluster the influencing variables of desired qualities through organization and validation. The organization will be shown in the results, and is generated through the input of all the variables available that affect the desired result (soil moisture). The cluster map can then be validated by inputting soil moisture levels into the cluster map and observing the uniformity of soil moisture levels in each cluster. Finally, the SOMs is being used to observe the variability of soil moisture from seasons to season by inputting the soil moisture levels into the SOM, organizing them into cluster.

Results and Discussion

The data available from the Brooks farm allowed for training of raw data maps, and SOM clustering maps to visualize the data available its hidden correlations. In Figure 2 below, these raw data maps present relationships between various factors present in the soil to moisture relationship. These factors are more easily understood because of the visual presentation of the results, allowing the viewer to understand the specific correlations between the soil chemistry factors and elevation values to various moisture conditions. Below, in Figure 2, the SOMs have visually reinforced the inverse relationship of elevation and soil moisture, but this example shows the possibility of expansion from a two-dimensional case to a multidimensional relationship while still comprehending the interactions between variables.



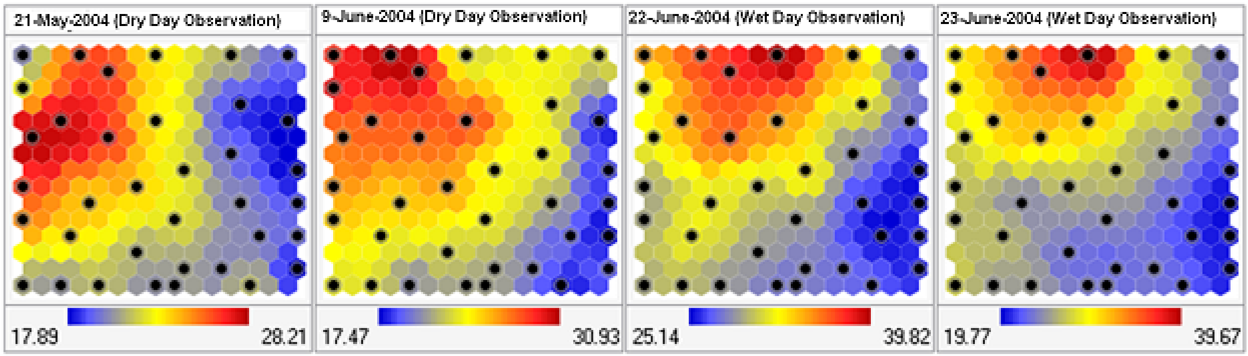


Figure 2: Self-organizing Map of Soil Data (two wet days and two dry days)

This figure shows individual self organizing maps for each input variable into the algorithm. The variability in the trends between the wet and dry days versus the other factors exposes a deeper relationship between the elevation and soil moisture, specifically in varying conditions. The explanation of this variability could surround both soil chemistry data (nitrogen, carbon, pH, and electrical conductivity) and soil landscape data (elevation, easting and northing coordinates, land slope, and land curvature).

Further investigation into Figure 2 can uncover a similarity between the nitrogen and carbon contents of the soil, as they have similar color distributions. These similar color diffusions highlight a connection of the two variables, meaning they have a similar trend throughout the land area.

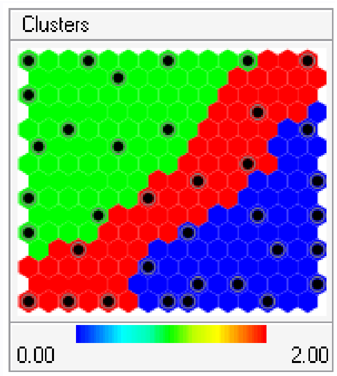


Figure 3: Cluster Map of Soil Data

The figure to the right is another resulting map extracted from the data. Figure 3 is a cluster map formed by training the first four variables from Figure 2 (elevation, nitrogen, carbon, and pH). This cluster map is an example of the dimensionality reduction possibilities that self-organizing maps contain. This cluster map highlights the important observation that there are three 'categories' of area in the given data set, three clusters were chosen for simplicity of the figure. Within each category, the data points will contain very similar values for each position on the map. These specific 'nodes' or hexagons on the cluster map can be correlated to the same spatially positioned 'nodes' on the maps in Figure 1. Generally, the green cluster contains locations of higher soil moisture, higher nitrogen, carbon, and pH values, and lower elevation. Conversely, the blue cluster contains locations with lower soil moisture, lower nitrogen, carbon, and pH values, and higher elevation. Lastly, the red area contains moderate levels of all the variables. As explained earlier, this dimensionality reduction shows large groupings within the data set, and the variables mimicking these groupings are the most influential factors affecting the data as a whole, specifically the elevation and pH are important factors.

Future Work

While the previous results display variability in soil moisture from a 'wet' to a 'dry' observation day, there are still more results that need to be gathered. One of the next steps in this research is to compile a group of cluster maps for each observed season in order to observe the variability from season to season. This will reinforce the factors in addition to elevation that control the soil moisture levels.

The self-organizing map creates complications in the comprehension process because users would like to believe that the rectangular map generated follows the same spatial structure as the physical land map of the input data when in fact they do not relate. One solution to this problem is introducing topology preservation methods into self-organizing maps such as the expansion of the generative topographic map (Bishop et al., 1998), or even more applicable the GEO-self-organizing map (Kiviluoto, et al., 1996; Bação, et al., 2004). This would alleviate this problem of unrelated spatial structure over a physical land map. In order to incorporate this method, the colored clusters from the clustering map will be overlaid onto the topographic map. It will then be possible to display each season and compare the distribution of soil moisture from season to season.

In addition to this visual representation, it is important to supplement with analytical tools which will reinforce the findings observed with the self-organizing maps. These two tools could be combined in an interface to optimally show the trends on a topographic map. Principle component analysis would aid this application, being a numerical method, because it would generate values to inform the program which data points have similar trends without confusing the user with an overwhelming numerical output.

Conclusion

Understanding and predicting soil moisture patterns has a number of applications in hydrology and agriculture levels. This information would allow a better understanding of the land for agriculture, weather, and space exploration. As shown in the results, soil moisture levels are not directly explained by elevation alone, but there are underlying factors that affect soil moisture. Using self-organizing maps, these factors can be extracted and understood using minimal data input.

Acknowledgements

This research was supported by the Iowa Space Grant Consortium and the National Aeronautics and Space Administration (NASA). The authors also expressed sincere thanks to the Spatial Data Analysis Lab at Iowa State University for the collection of ground observation data, and the Virtual Reality Applications Center at Iowa State University for the use of space and resources.

References – *Please do not use hyperlinks*--

Annas, S., Kanai, T., & Koyama S. 2007. Principal Component Analysis of Self-Organizing Map for Visualizing and Classifying Fire Risks in Forest Regions. *Agriculture Information Research,* 16(2): 44-51.

Bação, F., Lobo, V., & Painho, M. 2004. Geo-Self-Organizing Map (Geo-SOM) for Building and Exploring Homogeneous Regions. *Lecture Notes in Computer Science, 3234*: 22-37.

Bishop. C., Svensen, M., & Williams, C. 1998. GTM: The Generative Topographic Mapping. *Neural Computation,* 10:215-234.

Chang, D., & Islam, S. 2000. Estimation of Soil Physical Properties Using Remote Sensing and Artificial Neural Network. *Remote Sensing of Environment.* 74:534-544.

Kiviluoto, K. 1996. Topology preservation in self-organizing maps. *IEEE International Conference on Neural Networks,* 1:294-299.

Kohonen, T. 1989. The self-organizing map. *Neurocomputing*. 21:1-6.

Kooch, Y., Jalilvand, H., Bahmanyar, M., & Pormajidian, M. 2008. The Use of Principal Component Analysis in Studying Physical, Chemical, and Biological Soil Properties in Southern Caspian Forests (North of Iran). *Pakistan Journal of Biological Sciences, 11*(3):366-372.

Kothari, R., & Shafiqul, I., 1999. Spatial Characterization of Remotely Sensed Soil Moisture Data Using Self Organizing Feature Maps. *IEEE Transactions on Geoscience and Remote Sensing,* 37(2):1162-1165.

Levine, E.R., Kimes, D.S., Sigillito, V.G. November 1996. Classifying soil structure using neural networks. *Ecological Modeling.* 92:101-108.

Lourenco, F. 2005. Exploratory Geospatial Data Analysis Using Self-Organizing Maps. MS thesis. Lisbon, Portugal: University Nova of Lisbon, Geographical Information Systems Dept.

Malvic, T., & Durekovic, M., 2003. Application of methods: Inverse distance weighting, ordinary kriging, and collocated cokriging in porosity evaluation, and comparison of results on the Benianci and Stari Gradac fields in Croatia. *NAFTA,* 54(9):331-340.

McCabe, M., & Wood, E. 2006. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. *Remote Sensing of Environment,* 105:271-285.

Mele, P., & Crowley, D. 2008. Application of self-organizing maps for assessing soil biological quality [Electronic version]. *Agriculture, Ecosystems and Environment,* 126:139-152.

Schultz, A., Wieland, R. 1997. The use of neural networks in agroecological modelling. *Computers and Electronics in Agriculture.* 18:73-90.

Zhang, L., Miklas, S., Mustafa, A., Harrington, R. 2009. Application of the self-organizing map as a prediction tool for an integrated constructed wetland agroecosystem treating agricultural runoff. *Bioresource Technology.* 100:559-565.

Zhang, L., Scholz, M., Mustafa, A., Harrington, R. 2004. Application of the self-organizing map as a prediction tool for an integrated constructed wetland agroecosystem treating agricultural runoff. *Hydrological Processes.* 18:3235–3254.