Detection of Anomalous Crop Condition and Soil Variability Mapping Using a 26 Year Landsat Record and the Palmer Crop Moisture Index

E.R. Venteris^{}, J.D. Tagestad, J.L. Downs, and C.J. Murray*

Pacific Northwest National Laboratory, P.O Box 999, MSIN K9-33, Richland, WA 99352, United States

*Corresponding Author

Erik R. Venteris, Ph.D.
Soil Analytics Scientist
Emerging Leaders in Science – Big Data
Monsanto Company
Trait Field Solutions

Formerly at
Pacific Northwest National Laboratory
P.O Box 999, MSIN K9-33
Richland, WA 99352
erik.venteris@pnnl.gov

ABSTRACT

Cost-effective and reliable vegetation monitoring methods are needed for applications ranging from traditional agronomic mapping, to verifying the safety of geologic injection activities. A particular challenge is defining baseline crop conditions and subsequent anomalies from long term imagery records (Landsat) in the face of large spatiotemporal variability. We develop a new method for defining baseline crop response (near peak growth) using the normalized difference vegetation index (**NDVI**) from 26 years (1986-2011) of Landsat data for 400 km² surrounding a planned geologic carbon sequestration site near Jacksonville, Illinois. The normal score transform (y_{NDVI}) was applied on a field by field basis to accentuate spatial patterns and level differences due to planting times. We tested crop type and soil moisture (Palmer Crop Moisture Index (CMI)) as predictors of expected crop condition. Spatial patterns in y_{NDVI} were similar between corn and soybeans—the two major crops. Linear regressions between y_{NDVI} and the cumulative CMI (CCMI) exposed complex interactions between crop condition, field location (topography and soils), and annual moisture. Wet toposequence positions (depressions) were negatively correlated to CCMI and dry positions (crests) positively correlated. However, only 21% of the landscape showed a statistically significant ($p < 0.05$) linear relationship. To map anomalous crop conditions, we defined a tolerance interval based on y_{NDVI} statistics. Tested on an independent image (2013), 63 of 1,483 possible fields showed unusual crop condition. While the method is not directly suitable for crop health assessment, the spatial patterns in correlation between y_{NDVI} and CCMI have potential applications for pest damage detection and edaphological soil mapping, especially in the developing world.

Keywords – monitoring, geologic injection, NDVI, corn, soybean, soil mapping

1. Introduction

We developed a new analysis method to quantify the field-scale spatial variability of crop condition and its temporal response to inter-annual soil moisture variations (Jaynes et al., 2003) for monitoring and mapping applications. The initial motivation for this study was to define a baseline crop condition to enable detection of potential damage from leakage associated with geologic injection activities such as secondary oil recovery (“fracking”) and carbon sequestration. For example, concentrated brines (Kaddah and Ghowail, 1964) or carbon dioxide reaching the surface (Male et al., 2010; Noomen and Skidmore, 2008) will most likely be toxic to natural vegetation and crops (Patil et al., 2010) and generate detectable damage. While conceptually straightforward, separating the typical patterns of spatiotemporal variability found in imagery-based crop monitoring (Kawabata et al., 2001; Moran et al., 1997; Peters et al., 2002; Singh, 1989)) from a localized event is a considerable challenge. Flooding, drought, disease, pests, and lodging, are common and difficult to separate from events arising from injection activities. In addition, crop growth is highly variable, with processes that operate across multiple temporal and spatial scales. At the inter-field scale, differences are introduced by planting time, crop type, and amendment practices. Within-field variability is driven by the interaction between soil properties, landform (Muñoz et al., 2014) (water, chemical, and sediment movement), engineering structures (tile drainage (Naz et al., 2009), grassed waterways, etc.) and the weather conditions for a given year, especially precipitation. Given this complexity, our detection scheme seeks to identify potential anomalous locations that require focused investigations (field work) to determine causation. In addition, the correlation of crop response to variations in weather during the growing season exposes the properties of the underlying soil and can serve as an important source of soil data in remote regions where soil maps and related data (digital elevation models (McBratney et al., 2000; McKenzie and Ryan, 1999)) are scarce or of coarse resolution.

We define a baseline crop condition from the normalized vegetation index (**NDVI**) calculated from 26 years of multispectral imagery from Landsat. The possibility of refining baseline estimates through statistical relationships with crop type and interannual variations in soil moisture is also investigated. Our primary hypotheses are a) that field-scale spatial patterns are statistically related to the type of crop and the cumulative soil moisture conditions over the growing season, and b) that these relationships can be used to refine estimates of expected (field relative) crop condition. We expect such relationships because precipitation, terrain, and soil texture interact to produce spatial variations in water accumulation and retention (Lin, 2003), and these can be used to map crop response patterns (Jaynes et al., 2003; Kaspar et al., 2004), which are accentuated in wet and dry years (Venter et al., 2004). For example, closed depressions in fields typically have elevated clay and organic matter contents and retain moisture and sediments relative to back slopes and ridges (Ritchie et al., 2007; Schumacher et al., 2005). In dry years this results in enhanced crop growth, but in wet years too much moisture is retained (leading to anoxic conditions, de-nitrification etc. (Zaidi et al., 2004; Zak et al., 1999)), resulting in impeded growth. Mapping these areas can help in the interpretation of crop condition anomalies and serve as a basis for the delineation of crop response units for precision management applications.

2. Methods

2.1 Study area

Key landscape characteristics surrounding the test site (a 20 km by 20 km area centered on -90.113° W, 39.792° N, northeast of Jacksonville, Illinois) influenced the conduct of our analyses. The climate is continental, with cold winters, hot summers, and an average rainfall of 94 cm (37 inches) per year (Gotsch, 1988). Peak precipitation occurs in summer and is associated with convective (thunderstorm) systems. The spatial variability in precipitation (Huff and Shipp, 1968) within a given year presents a

potential challenge for characterizing local soil moisture. The general geomorphic setting of the area is an Illinoisan- age till plain (the depth to bedrock ranges from less than 8 m (25 feet) in incised valleys to 60 m (200 feet) in the uplands (Piskin and Bergstrom, 1994)) overlain by thick (>152 cm (60 inches)) Wisconsinan-age loess deposits (Tama-Ipava-Sable soil association). Land use and land cover (LULC) patterns are strongly related to those of the underlying soils and geomorphology (Figure S1), with each land use category occupying a dominant soil association. The texture of the upland soils is generally silt loam and silty clay loam, with enhanced clay and organic matter in small closed depressions that are too small (generally less than 100 m in diameter) to be mapped explicitly in the soil survey (Gotsch, 1988). The dominant LULC of this setting is cropping in the form of corn-soybean rotation. The Illinoisan till plain has been exposed to erosion for ~130,000 years and so the landscape is incised with a well-developed drainage system that flows to the Illinois River (to the west-northwest of the study area). These incised bottom lands generally contain silty alluvial deposits (Rozetta-Hickory-Elco soil association) and the backslope and shoulder soils are typically highly eroded. The dominant LULC in these settings is wooded and pasture lands. Finally, there are several management practices that can affect growth patterns within fields in addition to corn-soybean rotation. Tillage is generally by chisel or moldboard and is conducted in the spring. Rill erosion is controlled by grassed waterways, which often flow into constructed drainage ditches. Brightness patterns in aerial photography reflecting soil moisture content indicate the use of drainage tile for some fields.

2.2 Landsat processing

Defining expected crop condition required long spatio-temporal records. Several years (at least a decade) with contrasting weather conditions were needed, which excluded many high spatial resolution commercial sensors. Sensors were evaluated on the tradeoffs between spatial resolution, temporal resolution, and length of record. MODIS provided images every 1 to 2 days since 1999 (for Terra) but the

250 m to 1000 m resolution was deemed insufficient. AWiFS (European Space Agency, 2014) had the positive qualities of moderate spatial resolution and a high recurrence rate (~5 days), but records only extended back to 2003. The best combination of long-term record and spatial resolution was obtained from the Landsat sensors, with records based on the Thematic Mapper instruments extending back to 1986. A challenge was developing a methodology compatible with the low spatial resolution (30 m) and recurrence interval (18 days), which was exacerbated by intermittent cloud cover.

This spatiotemporal analysis of crop condition was based on remote sensing data from the Landsat (NASA, 2014) family of instruments (Thematic Mapper, Enhanced Thematic Mapper+, and Operational Land Imager (for the 2013 test image), collectively abbreviated TM). A cloud-free image closest to the time of peak crop growth was selected for each year (see collection dates in Table S1) over the time span 1986 to 2011. Each image was radiometrically corrected using the dark-object subtraction technique (Chavez Jr, 1988) and projected to the UTM coordinate system. The normalized difference vegetation index (**NDVI**) for each grid cell (**u**) was calculated from

$$\mathbf{NDVI}(\mathbf{u}) = \frac{\mathbf{NIR}(\mathbf{u}) - \mathbf{RED}(\mathbf{u})}{\mathbf{NIR}(\mathbf{u}) + \mathbf{RED}(\mathbf{u})} \quad (1)$$

where **NIR** is the near infrared band (TM band number 4) and **RED** is the red band (TM band number 3). The metric (Figure 1) provides an estimate of the vigor of plant condition through scaling of the amount of infrared light reflected off the internal portion (mesophyll) of plant leaves.

2.3 Land use and land cover characterization

We categorized the LULC for each of the studied years so that differences between corn and soybeans could be tested. From 1997 to 2011, LULC was obtained directly from the Crop Data Layer (Johnson and Mueller, 2010; NASS (National Agricultural Statistics Service), 2013). For earlier dates, we developed classification signatures by correlating the 2010 Landsat data to land cover categories in the

corresponding Crop Data Layer. We used these signatures to identify land cover for the years 1986 to 1996 via a maximum likelihood classification algorithm. To eliminate the chance that an individual field would contain both corn and soybean (likely due to misclassification), we calculated the majority crop pixel within each field boundary for each year and assigned each field type accordingly.

To test for differences in patterns between soybeans and corn, we digitally delineated fields that were planted in a single crop (soybeans or corn) for each of the studied 26 years. The boundaries do not indicate a constant crop (most of the mapped fields were rotated several times over the span), but rather that the field was planted as a cohesive unit for each individual year. Also, pixels located on the edge of fields often exhibited a mixed response between crops and other land covers. Expected **NDVI** response was difficult to define for such locations so their inclusion was minimized (Figure 1). Fields with an area of 2.8 ha (7 acres) or more (approximately 30 Landsat pixels) were included in the analysis, as a minimum estimate of the required number of samples (n) to obtain a meaningful spatial representation.

2.4 Inter-annual variability in growing season soil moisture

Spatial patterns in peak crop condition vary from year to year due to the availability of soil water. To assess the impact of soil moisture, we used the Palmer Crop Moisture Index (CMI, (Palmer, 1968)) as calculated by the National Oceanic and Atmospheric Administration Climate Prediction Center (NOAA Climate Prediction Center, 2013). The CMI is a weekly index that captures short term soil moisture fluctuations and is derived from the longer term Palmer Drought Severity Index (PDSI, (Palmer, 1965)). The index uses precipitation, soil moisture, and temperature data to define abnormally wet and dry conditions. The Palmer computations are complex and their presentation here is beyond scope (a detailed discussion is available (Alley, 1984)).

The study area was located in the north central area of NOAA climate division 1106, near the borders of climate divisions 1104 and 1103 (Figure S2). Accordingly, the calculations were based on the average

CMI for the three divisions. The cumulative CMI (CCMI) was calculated from Palmer week 6 (the dates are in Palmer years, beginning the first Wednesday in March), the average early planting date for corn (Pioneer Hi-Bred, 2013), until the date of the Landsat image collected for each year (Table S1). The end result was a one dimensional grid with each year having a representative CCMI value to serve as a proxy for moisture conditions during the growing season.

2.5 Characterization of spatial variability within fields

Interannual comparisons of raw **NDVI** values are problematic due to differences in image collection dates and their interaction with crop growth stage. In addition, differences in mean **NDVI** occurred between fields due to variations in planting time and crop type. Figure 1 shows gross differences in **NDVI** soybeans and corn, but there were also contrasts between fields within the same crop, likely due to planting times. Definition of expected **NDVI** for each year required knowledge of both the planting time for each field and the integrated effect of growing conditions up to the image collection date.

As an alternative to using the **NDVI** values directly, we developed an approach to assess the relative spatial patterns in crop condition within each field. In addition to the temporal inconsistencies, the within-field distributions of **NDVI** were strongly skewed, introducing difficulties in hypothesis testing and anomaly definition (Figure S3). To address these issues, we transformed the data using the normal score transform (NST) (Deutsch and Journel, 1998). The technique is based on quantile matching, where a given data point is assigned a cumulative probability and converted to the equivalent normal score. We wrote custom software based on Python (*ArcPy*, *NumPy*, and *SciPy*) to apply the NST to the data on a field by field basis. The **NDVI** values were ranked within each field by increasing value

$$\mathbf{NDVI}^{(1)} \leq \mathbf{NDVI}^{(2)} \leq \dots \leq \mathbf{NDVI}^{(n)} \quad (2)$$

The original **NDVI** value of rank k was then transformed to its corresponding value (y) in the Gaussian normal distribution (G) according to-

$$y_{NDVI}^k = G^{-1} \left((kn^{-1} + (k-1)n^{-1})/2 \right) \quad (3)$$

where n is the number of cells within the field (the terms in parentheses are the average quantile of k and $k - 1$ as a simple approach to estimate the tail probabilities). The effect on the distribution for one field is illustrated in Figure S3. The method normalized the **NDVI** information within each field, and enhanced the display of within-field spatial variation (Figures 2 and 3). However, because the transform alters the statistical distribution significantly, comparisons between years of the absolute condition of plants in the field cannot be made with the transformed data.

2.6 Statistics for Mapping and Hypothesis Testing

We tested for predictive statistical relationships between crop type, CCMI, and y_{NDVI} . Grids were calculated for each crop and year combination. Analyses were based on a three dimensional masked *SciPy* array for each crop type (x and y horizontal cells, with the z dimension being time in years) and a one dimensional (z) CCMI array.

We conducted exploratory data analysis (EDA) to investigate the shape of the temporal distribution in y_{NDVI} for each cell. The critical issues were the large number of cells (over 150,000) which make typical graphics-based EDA impossible, and the small number of temporal samples (maximum $n = 26$, Figure S4) which required extra care in analysis and interpretation because there was considerable potential for biased fits, false hypothesis rejection, etcetera. However, the number of years in the Landsat record was sufficient to at least attempt statistical analyses, with the recognition that another decade of data would

increase rigor. The small sample size did not permit quantitative testing to confirm Gaussian or other distribution types. Instead, we calculated the range of moments in order to qualitatively evaluate how many cells might have problematic distributions. Our approach of using a single image for each year and CCMI (as opposed to multiple images and CMI) precluded the modeling of temporal autocorrelation (too few temporal samples). We tested the hypothesis that the mean y_{NDVI} response was significantly different between corn and soybeans using the two-sample t test with the degrees of freedom estimated using the Welch-Satterwaite equation (Welch, 1947).

The second set of statistical analyses sought to investigate the relationship between y_{NDVI} and CCMI. It is well-established (Denmead and Shaw, 1960, 1962; Robins and Domingo, 1953; Zaidi et al., 2004; Zaidi et al., 2003) that crop response is highly sensitive to moisture conditions. We expected the largest responses from the extremely wet and dry areas of each field. We initially explored these relationships through analysis of 60 relatively wet and 60 dry point locations paired within individual fields. The test sites were interpreted and digitized from a USGS panchromatic digital ortho quarter quad mosaic (DOQQ, Figure S5). The wettest areas were in relic depressions left over from the disintegration of the Pleistocene ice sheet. Such features commonly exhibited relatively sharp boundaries (Figure S5, inset). The boundaries of dry areas were less distinct, with most uplands showing bright returns in the DOQQs due to low wetness and soil organic carbon content (SOC) (Chen et al., 2000). However, light and dark areas in the photography required careful interpretation. Some topographic depressions and were bright due to precipitation of salts (CaCO_3 etc.), and patterns were masked in some fields by stubble left over from the previous growing season. Dark areas generally correlated to high moisture and SOC, but the relationship to crop growth made more complex through the occasional installation of drainage tiles, so that dark areas were not always reflective of current moisture conditions. Such problematic areas were excluded from the initial sample.

In addition to the wet/ dry test sites, we evaluated the relationships between y_{NDVI} and CCMI for all 166,990 grid locations (with the decision to nest by crop type to be based on the t -test results). We calculated the Pearson correlation and conducted a least-squares linear regression to test for relationships between crop growth and annual moisture conditions. Arbitrarily, grid locations with less than nine years of data were excluded from the analyses. The interpretation of linear regression results were based on slope, intercept, r^2 values, and the two-tailed p value to test the null hypothesis that the slope of the regression line is zero.

The final definition of normal and anomalous crop condition depended on the outcome of the previous described tests and models. If the correlative analyses failed, the simplest anomaly definition was to treat each cell in the new image (post 2011) as a sample from the temporal population distribution as defined by the sample mean and standard deviation. Anomalies were defined by a two-sided tolerance interval based on y_{NDVI}

$$T = \bar{x}_{y_{NDVI}} \pm k_2 s_{y_{NDVI}} \quad (4)$$

with the k_2 variable (Howe, 1969) estimated from

$$k_2 = \sqrt{\frac{\mathbf{v}_T \left(1 + \frac{1}{n}\right) \mathcal{Z}_{(1-p)/2}^2}{\chi_{1-\gamma, \mathbf{v}_T}^2}} \quad (5)$$

where $\mathcal{Z}$ is the critical value of the normal distribution for the cumulative probability $(1 - p)/2$, χ^2 is the critical value of the chi-square distribution based on the probability of exceedance γ and the degrees of freedom ($\mathbf{v}_T = n - 1$). The values from the normal and chi-square distributions were computed from their respective *SciPy* modules. For this work we specified the proportion of coverage of

the distribution at $p=0.95$ and a probability of exceedance of $\gamma=0.95$. A sample value lying outside T was considered an anomaly requiring further investigation.

3. Results

We characterized the sample distributions of y_{NDVI} in the time (z) axis to identify areas where statistical analysis could be problematic. Each moment was mapped and a histogram created (see Supporting Data). The typical range in y_{NDVI} within an individual field and year was -3.30 to +3.30, with small values indicating relatively poor condition compared other field locations. The mean y_{NDVI} values (Figure S6) were mostly near zero as expected. Some sites had consistently very small y_{NDVI} (up to -3.0) but none were consistently very large (greater than 3.0), resulting in a tail of low values in the histogram. The values of s (Figure S7) ranged from nearly zero to 2.25, with values commonly between 0.75 and 1.0. In general, the data had low skewness (Figure S8), with most values being within ± 0.5 , although there were grid cells with large values. The kurtosis values (Figure S9) were generally unremarkable, with the majority of the data within ± 1.0 . Overall, the analysis showed that the majority of sites within the fields were compatible with t -test and regression analysis, but the moments for border cells in and near grassed waterways, field boundaries, and around buildings indicated higher variability and increased likelihood of non-normal distributions. These noisy areas were unlikely to be useful for monitoring.

The y_{NDVI} response was not consistently different between corn and soybeans. The hypothesis of different means was rejected on the basis of the t -test. The proportion of p values (Table 1) at each significance level was approximately that expected from chance. However, this test was not definitive, given that the mean was only one statistical moment where the crop response could contrast, the number of temporal samples was small, and occasional cell locations with non-normal temporal

distributions. Therefore, we provide results where corn and soybeans were modeled separately (see Supporting Data), but focused on results where they were combined.

We calculated the average difference between extreme wet and dry locations for each year ($\Delta\bar{y}_{NDVI}$) and investigated its relationship with CCMI (Figure 4). The results at least partially confirmed that wet sites showed better condition than dry sites in dry years (negative CCMI) and vice versa. In general, there was good agreement between the trends in $\Delta\bar{y}_{NDVI}$ and CCMI, but with divergences in several years (for example 1987 and 1995). The reason for these departures was unclear, but they were likely related differences between the actual, on-site precipitation events and the regional-scale of the CCMI data. Occasional departures were not unexpected considering the dominance of convective systems for this region. Despite these anomalies, the overall correlation between the parameters was strong (Figure 4). However, in contrast to the *t*-test results, the correlation between $\Delta\bar{y}_{NDVI}$ and CCMI was 0.17 larger for soybeans than corn, suggesting a potential difference in response for at least the most wet and dry field locations. Of particular interest was the span from 1998 to 2003, where the soybean signal tracked CCMI fairly well, but corn showed a steadily decreasing trend. Here, there was a difference in crop response despite the same weather forcing, although overall soybean and corn $\Delta\bar{y}_{NDVI}$ were highly correlated ($r = 0.76$). In addition, the example wet (Figure 2) and dry (Figure 3) maps exhibited reversing of y_{NDVI} values as expected.

With favorable preliminary results, we calculated correlation coefficients and conducted regression analyses for corn and soybean fields both separately (see Supporting Data, Table S2 and S3) and combined. Pearson correlations ranged from -0.8 to 0.8 with positive correlation for dry areas and anti-correlation for wet areas (Figure 5), with the slopes from regression (Figure S10) following similar patterns. However, the *p* values for the hypotheses of non-zero slope demonstrated that there is a statistically significant linear relationship between y_{NDVI} and CCMI for only a portion of the study area

(Figure 6). The portion was far more than would be expected from chance (Table 2) with approximately 21% of the sites meeting the $p < 0.05$ significance level. However, statistical significance does not in itself indicate a useful prediction. Less than 8% of the data had an r^2 greater than 0.5 (Table 3), so few of the fit locations were useful for enhancing the spatial specificity of the anomaly definition.

Accordingly, the univariate definition of expected crop condition was used (equations 4 and 5) to identify potential anomalies. To illustrate the application to crop monitoring, we calculated y_{NDVI} for a 2013 Landsat image (2012 was a severe drought year, and the only cloud-free image available was from Landsat 7 with the scan line corrector issue). We mapped the grid cells in the 2013 image where y_{NDVI} values were outside the range of T (Figure 7), and inventoried the number of anomalous values for each field. The proportion of grid cells exceeding T was 0.0163, so overall there was little evidence for a larger amount of anomalies than expected from chance. However, grid cells exceeding T were not randomly distributed. Specifically, 62 of the 1483 fields had a proportion of cells exceeding $T > 0.05$. For monitoring applications, these fields would be the highest priority for causation investigations through inspection of local records, field work, and additional imagery.

4. Discussion

Our original motivation for the study was to define normal vs. anomalous patterns in crop **NDVI** response for use in monitoring geologic CO₂ injection sites. Landsat was the favored data source due to the long temporal record which gave the best opportunity for capturing interannual variability, but the direct use of **NDVI** was problematic due to the 18 day sampling interval (less when cloud cover is considered) and temporal confounding due to variations in planting time and image collection date. The NST transform is one potential solution to these issues, but we have not tested it against other potential methods. The transform has an additional benefit of minimizing the influence of strain improvements

(from breeding and genetics) when investigating the influence of weather variations on the spatial patterns in crop growth. A disadvantage of the approach was the loss of potential insights that could be obtained from analysis of multiple samples within each growing season. In addition, the current method required the mapping of field boundaries to delineate planting units and avoid of transition areas (border pixels) . While the similar response between corn and soybeans may lessen the need for field boundary mapping, the current dependency of the analysis on field boundaries requires automated field delineation for application to large areas (Yan and Roy, 2014).

We explored the potential for using crop type and CCMI as predictive variables to define expected crop conditions (y_{NDVI}) in future imagery. We found little evidence that crop type induced significant contrasts in y_{NDVI} . While raw **NDVI** was different between the crops (Figure 1), within-field spatial patterns were statistically similar. The *t*-test result was negative, and the linear regression results when stratified between crop type (Tables S2 and S3) were remarkably similar. However, some differences were noted in the analysis of extreme locations (Figure 4), so the hypothesis should be retested as more years of data become available. The statistical results between y_{NDVI} and CCMI (Tables 2 and 3) were more ambiguous. The spatial patterns in Pearson's R (Figure 5) were clearly related to those seen in the aerial photography with dry areas correlated to CCMI and wet areas anti-correlated as expected. In addition, there were areas where the linear regression is clearly statistically significant (Figure 6), although only 21% of the study area met the 0.05 significance criterion. This, coupled with the generally low r^2 of the linear regression, precluded a multivariate definition of expected crop condition. Currently, the best approach to anomaly detection is the tolerance interval based on y_{NDVI} (Equation 7, Figures S6 and S7). A useful multivariate baseline definition may be possible as more years are collected.

The statistical relationships between y_{NDVI} and CCMI are intriguing and have potentially significant applications for the mapping of soils and crop management optimization. Typically, spatial variations in

crop response at the field scale are mapped using precision agriculture techniques (McBratney et al., 2005; Venteris et al., 2004), employing farming equipment equipped with yield monitors and Global Positioning Systems, occasionally supplemented with high resolution soil data and remote imagery. Our Landsat-based method is not nearly as spatially precise (nor directly measures yield), but has the potential to investigate crop patterns over large geographic areas and long time periods, as the data area available nearly globally for over 20 years. Specifically, the strength of correlation with CCMI allows the influence of interannual variations in soil moisture conditions to be quantified. The example wet (Figure 2) and dry (Figure 3) years showed complex patterns. In general, sites that grew best in the wet years grew poorest in the dry, but the inversion patterns were not simple. Whereas the general relationships between crop condition and non-ideal soil moisture are well understood, mapping such as this provides insight into the complexity of water, soil, and crop interactions and provides the potential to separate soil moisture effects from other variability drivers such as pests that in many cases are less-correlated to soil and landform position.

In addition, the strength of correlation between y_{NDVI} and CCMI (Figure 5) is clearly related to soil properties and therefore has the potential for use in the mapping of soils and biogeochemical gradients at landscape scales (Ritchie et al., 2007; Venteris et al., 2004). The method could help bridge resolution gaps between SSURGO-scale and first order soil mapping in developed countries, and complement CLORPT-based mapping approaches (McBratney et al., 2000; McKenzie and Ryan, 1999) in developing countries and remote regions. The method also has a potential advantage for edaphological and biological applications over terrain-based mapping methods ((Moore et al., 1993; Ritchie et al., 2007; Venteris and Slater, 2005)) as crop response is directly incorporated.

5. Conclusion

We have demonstrated a new method of processing long Landsat records to define normal vegetation conditions at the field scale over large areas. Currently, the best estimate of expected crop condition is the tolerance interval based on y_{NDVI} , and we demonstrated the detection of fields with anomalous crop growth in a pre-injection image. The technique may be employed directly for the monitoring of geologic injection sites. The analysis technique also shows particular promise for the mapping the spatial variability of crop response to soil and weather variability in the developing world where data at resolutions relevant to crop management (ie field scale) are typically unavailable.

Acknowledgements

This material is based upon work supported by the Department of Energy under Award Number DE-FE0001882 and DE-FE0005054. Richard Skaggs and Vince Vermeul are thanked for providing early reviews of the manuscript.

References

- Alley, W.M., 1984. The Palmer drought severity index: limitations and assumptions. *J. Clim. Appl. Meteorol.* 23, 1100-1109.
- Chavez Jr, P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* 24, 459-479.
- Chen, F., Kissel, D.E., West, L.T., Adkins, W., 2000. Field-Scale Mapping of Surface Soil Organic Carbon Using Remotely Sensed Imagery. *Soil Sci. Soc. Am. J.* 64, 746-753.
- Denmead, O.T., Shaw, R.H., 1960. The Effects of Soil Moisture Stress at Different Stages of Growth on the Development and Yield of Corn. *Agron. J.* 52, 272-274.
- Denmead, O.T., Shaw, R.H., 1962. Availability of Soil Water to Plants as Affected by Soil Moisture Content and Meteorological Conditions. *Agron. J.* 54, 385-390.
- Deutsch, C., Journel, A., 1998. Geostatistical software library and user's guide (GSLIB). Oxford University Press Oxford, p. 369.
- European Space Agency, 2014. ESA, Earth Online, Browse Data Products, IRS-P6 AWiFS.
- Gotsch, K.A., 1988. Soil Survey of Morgan and Scott Counties, Illinois, in: Service, S.C. (Ed.). United States Department of Agriculture, p. 195.
- Howe, W.G., 1969. Two-Sided Tolerance Limits for Normal Populations—Some Improvements. *Journal of the American Statistical Association* 64, 610-620.

- Huff, F.A., Shipp, W.L., 1968. Mesoscale Spatial Variability in Midwestern Precipitation. *J. Appl. Meteorol.* 7, 886-891.
- Jaynes, D.B., Kaspar, T.C., Colvin, T.S., James, D.E., 2003. Cluster Analysis of Spatiotemporal Corn Yield Patterns in an Iowa Field. *Agron. J.* 95, 574-586.
- Johnson, D.M., Mueller, R., 2010. The 2009 cropland data layer. *Photogramm. Eng. Remote Sens.* November, 1201-1205.
- Kaddah, M.T., Ghowail, S.I., 1964. Salinity Effects on the Growth of Corn at Different Stages of Development. *Agron. J.* 56, 214-217.
- Kaspar, T.C., Pulido, D.J., Fenton, T.E., Colvin, T.S., Karlen, D.L., Jaynes, D.B., Meek, D.W., 2004. Relationship of Corn and Soybean Yield to Soil and Terrain Properties Joint contribution from USDA-ARS and Iowa State Univ. *Agron. J.* 96, 700-709.
- Kawabata, A., Ichii, K., Yamaguchi, Y., 2001. Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. *Int. J. Remote Sens.* 22, 1377-1382.
- Lin, H., 2003. *Hydropedology*. *Vadose Zone J.* 2, 1-11.
- Male, E., Pickles, W., Silver, E., Hoffmann, G., Lewicki, J., Apple, M., Repasky, K., Burton, E., 2010. Using hyperspectral plant signatures for CO₂ leak detection during the 2008 ZERT CO₂ sequestration field experiment in Bozeman, Montana. *Environ Earth Sci* 60, 251-261.
- McBratney, A., Whelan, B., Ancev, T., Bouma, J., 2005. Future Directions of Precision Agriculture. *Precision Agric* 6, 7-23.
- McBratney, A.B., Odeh, I.O., Bishop, T.F., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* 97, 293-327.
- McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89, 67-94.
- Moore, I.D., Gessler, P., Nielsen, G.A.e., Peterson, G., 1993. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* 57, 443-452.
- Moran, M.S., Inoue, Y., Barnes, E.M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens. Environ.* 61, 319-346.
- Muñoz, J.D., Steibel, J.P., Snapp, S., Kravchenko, A.N., 2014. Cover crop effect on corn growth and yield as influenced by topography. *Agriculture, Ecosystems & Environment* 189, 229-239.
- NASA, 2014. *Landsat Science*.
- NASS (National Agricultural Statistics Service), 2013. *CropScape - Cropland Data Layer*. USDA NASS.
- Naz, B.S., Ale, S., Bowling, L.C., 2009. Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes. *Agricultural Water Management* 96, 627-637.
- NOAA Climate Prediction Center, 2013. *Climate Prediction Center (CPC) Palmer Drought and Crop Moisture Indices*.
- Noomen, M.F., Skidmore, A.K., 2008. The effects of high soil CO₂ concentrations on leaf reflectance of maize plants. *Int. J. Remote Sens.* 30, 481-497.
- Palmer, W.C., 1965. *Meteorological drought*. US Department of Commerce, Weather Bureau Washington, DC, USA.
- Palmer, W.C., 1968. Keeping track of crop moisture conditions, nationwide: The new crop moisture index. *Weatherwise* 21, 156-161.
- Patil, R.H., Colls, J.J., Steven, M.D., 2010. Effects of CO₂ gas as leaks from geological storage sites on agro-ecosystems. *Energy* 35, 4587-4591.
- Peters, A.J., Walter-Shea, E.A., Ji, L., Vina, A., Hayes, M., Svoboda, M.D., 2002. Drought monitoring with NDVI-based standardized vegetation index. *Photogrammetric engineering and remote sensing* 68, 71-75.

- Pioneer Hi-Bred, 2013. 2013 Illinois Initial and Final Plant Dates - Corn.
- Piskin, K., Bergstrom, R.E., 1994. Glacial Drift in Illinois: Thickness and Character. Illinois State Geological Survey, Champaign, Illinois.
- Ritchie, J.C., McCarty, G.W., Venteris, E.R., Kaspar, T.C., 2007. Soil and soil organic carbon redistribution on the landscape. *Geomorphology* 89, 163-171.
- Robins, J.S., Domingo, C.E., 1953. Some Effects of Severe Soil Moisture Deficits at Specific Growth Stages in Corn. *Agron. J.* 45, 618-621.
- Schumacher, J.A., Kaspar, T.C., Ritchie, J.C., Schumacher, T.E., Karlen, D.L., Venteris, E.R., McCarty, G.W., Colvin, T.S., Jaynes, D.B., Lindstrom, M.J., Fenton, T.E., 2005. Identifying spatial patterns of erosion for use in precision conservation. *J. Soil Water Conserv.* 60, 355-362.
- Singh, A., 1989. Review Article Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 10, 989-1003.
- Venteris, E.R., McCarty, G.W., Ritchie, J.C., Gish, T., 2004. Influence of management history and landscape variables on soil organic carbon and soil redistribution. *Soil Sci.* 169, 787-795.
- Venteris, E.R., Slater, B.K., 2005. A Comparison Between Contour Elevation Data Sources for DEM Creation and Soil Carbon Prediction, Coshocton, Ohio. *Transactions in GIS* 9, 179-198.
- Welch, B.L., 1947. The generalization of student's 't' problem when several different population variances are involved. *Biometrika* 34, 28-35.
- Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sens. Environ.* 144, 42-64.
- Zaidi, P.H., Rafique, S., Rai, P.K., Singh, N.N., Srinivasan, G., 2004. Tolerance to excess moisture in maize (*Zea mays* L.): susceptible crop stages and identification of tolerant genotypes. *Field Crops Research* 90, 189-202.
- Zaidi, P.H., Rafique, S., Singh, N.N., 2003. Response of maize (*Zea mays* L.) genotypes to excess soil moisture stress: morpho-physiological effects and basis of tolerance. *European Journal of Agronomy* 19, 383-399.
- Zak, D.R., Holmes, W.E., MacDonald, N.W., Pregitzer, K.S., 1999. Soil Temperature, Matric Potential, and the Kinetics of Microbial Respiration and Nitrogen Mineralization. *Soil Sci. Soc. Am. J.* 63, 575-584.

Figure Captions

Figures

Figure 1 Portion of the study area showing *NDVI* for 1992, the field boundaries used in the study, and crop classifications for 1992.

Figure 2 Portion of the study area showing *NDVI* after application of the normal score transform for a relatively wet year (1992).

Figure 3 Portion of the study area showing **NDVI** after application of the normal score transform for a relatively dry year (1990). Fields in the central

Figure 4 Plot of the mean difference between y_{NDVI} for wet and dry sites and CCMI. The Pearson r value for corn and soybean fields is shown in the legend.

Figure 5 Isoline map showing Pearson r for correlation between y_{NDVI} and CCMI for selected fields. In general, high correlation is associated with bright (generally drier soil with less SOC) returns and high anti correlation is associated with dark (generally wetter soil with more SOC) returns.

Figure 6 Isoline map showing results of two-tailed t-test testing the hypothesis of non-zero slope.

Figure 7 Map of the entire study area showing cells where y_{NDVI} is outside the tolerance interval, based on a Landsat 8 image collected in 2013. Fields with a proportion greater than 0.10 of anomalous cells are highlighted in red. These fields are the highest priority for detailed field investigation.

TABLES

Table 1 Two tail p value results from the two sample t test for difference between y_{NDVI} between corn and soybeans.

p value	n cells	Proportion of cells meeting criterion
<0.05	5,403	0.041
<0.1	9,645	0.073
>0.1	122,866	0.927
Total Comparisons	132,511	1.000

Table 2 Compilation of two tail p value results for the test of non-zero slope (from linear regression).

Two Tailed Significance	Number Cells	Proportion of Cells Meeting Criterion
<0.005	4,799	0.029
<0.01	7,677	0.046
<0.05	35,212	0.211
<0.1	102,534	0.614
<1	166,990	1.000

Table 3 Compilation of the r^2 values from linear regression.

r Squared	Number Cells	Proportion of Cells Meeting Criterion
>0.9	23	0.000
>0.8	208	0.001
>0.7	1,052	0.006
>0.6	3,997	0.024
>0.5	13,018	0.078
>0.4	38,688	0.232
<0.4	166,990	1.000