

ESTIMATING SUGAR BEET YIELD USING AVIRIS-DERIVED INDICES

Nahum Gat¹, Hector Erives¹, Glenn J. Fitzgerald², Steve R. Kaffka³, Stephan J. Maas²

1. INTRODUCTION

In this paper we discuss a preliminary analysis of the use of AVIRIS-derived vegetation indices to estimate sugar beet yield. Analysis was performed on scenes from a sugar beet field near Stratford, CA acquired by AVIRIS (20 meter pixel resolution) and images acquired with the Shafter Airborne Multispectral Remote Sensing System (SAMRSS) at 1 meter pixel size. A yield map was produced from yield monitor data to identify those areas of low and high yields. Hyperspectral remotely sensed imagery is proposed as a crop management aid in the estimation of yield early in the season to aid farmers in determining corrective actions that may need to be taken. The analysis is based on obtaining vegetation indices from a hyperspectral sensor that could be used to estimate yields. Various indices have the potential to indicate specific features of the crop that could be improved by management, such as nitrogen content and water stress. The two sensors mentioned were used in the analysis and their results combined and compared.

Preliminary analysis of the data from both sensors indicates that (i) remotely sensed data could be used to estimate yields in sugar beet fields, (ii) temporal variations in certain field-averaged and in spatial vegetation indices can be used to follow the growth of a crop and determine if such growth follows an anticipated schedule, (iii) new, and more effective indicators are needed that capture more of the factors that affect crop development, and that such parameters may be extracted from remote sensing data.

Agriculture will likely place greater demands on the deployment of hyperspectral remote sensing technology than any other field of endeavor. Due to the varied problems in agriculture (nitrogen, water stress, diseases, insects, weeds, soil types, growth cycle variations, climate and geography, etc.), the varying scales of spatial resolution required to resolve these (5 m may be sufficient for large field patterns but 1 m or less is required for disease and insect detection), the need for repeat overpasses (temporal resolution), and quick delivery of the information collected to the farming community, the pressures on successful utilization of hyperspectral remote sensing are great. Estimation of yield is one example of this. Crops are integrators of stresses present during the growth season. If stresses occur yield can suffer. While remotely sensed measurement of factors that can influence yield such as plant nitrogen (chlorophyll) and canopy water content have been shown (Gamon *et-al*, 1998; Green *et-al*, 1998; Kokaly and Clark, 1998), regular flights over fields are required especially early in crop development if corrective management is to be implemented. Thus, one image of a crop, even at a critical time of the season, may not be a good estimator of final yield if other stresses also occurred but were not recorded due to insufficient temporal coverage.

Vegetation indices (VIs) have been used for many years, since the advent of LandSat. These provide a simple measure of crop health but with wide band multispectral remote sensing few bands were available at important spectral regions for the calculation of many indices. Narrow band hyperspectral imaging allows the selection of specific spectral regions that may contain important physiological information about the crop. Indices calculated from these spectral regions have the potential to indicate types of stress the crop is undergoing. Although there are other data analysis methods, this is a simple method for initial assessment of crop studies.

2. MATERIALS AND METHODS

The following data sets were acquired and preprocessed:

¹Opto-Knowledge Systems, Inc. (OKSI), Torrance CA (nahum@oksi.com, hector@oksi.com)

²USDA-ARS, Western Integrated Cropping Systems Research Unit, Shafter, CA (gjfitzgerald@ucdavis.edu; sjmaas@ucdavis.edu)

³Dept. of Agronomy and Range Science, University of California, Davis (srkaffka@ucdavis.edu)

- AVIRIS flights over a sugar beet field located near Stratford, CA took place in 6/13/99, 8/28/99, 9/1/99, 9/10/99, and 9/24/99 (the 9/10 flight took place over a partially harvested field, while the 9/24 was post harvesting). Atmospheric compensation using ATREM (using bands that were selected based on the local vegetation spectra) and smoothing using the EFFORT algorithm were applied. False color images of the field are shown in Fig. 1.
- A high-resolution multispectral remote sensing image was acquired of this field on 6/1/99 after full canopy was established. The imagery was acquired using SAMRSS, a digital airborne imaging package developed by the USDA-ARS laboratory at Shafter, CA and OKSI. It consists of three digital cameras fit with narrow-band filters allowing transmission of light at 548 nm (green), 658 nm (red), and 850 nm (near-infrared [NIR]) wavelengths. The green and red bands Full Width at Half Maximum are 13 nm and the NIR band has a FWHM of 40 nm. Flights occurred within 1 hour of solar noon and at 7,500 feet (2,285 m) above ground surface resulting in a pixel resolution of 1 m (OKSI web reference).
- Additional SAMRSS flights and data were collected at various growth stages, including 3/29/99, 6/1/99, 6/30/99, 7/28/99, and 8/13/99. Image digital numbers were converted to percent reflectance using radiometer data collected from a 48% reflectance ground tarp visible within the images (Moran *et-al*, 1997) and bare soil. False color imagery of the field are shown in Fig. 2.
- Soil electrical conductivity maps were available for the field (Kaffka *et-al*).
- A Harvest Master yield monitor fit with a Differential GPS collected spatial data (longitude, latitude, altitude, and local yield) during harvesting in mid September, 1999. Harvesting, as noted took place over 10 days from Sept., 5 through Sept., 14, 1999. The vector data file was converted to a raster yield images at 6 m and 8 m pixel size. Anomalies and outlier data points (during harvesting the yield monitor suffered several DGPS drop outs) were masked out.
- The AVIRIS/SAMRSS images were co-registered to the yield map and masked to remove the same masked pixels in the yield map. Reflectance data were used to compute various vegetation indices.

The 8 m yield map is shown in Fig. 3. It is noted that the southeast half of the field appears brighter than the northwest half, indicating higher yields in the lower part. The SAMRSS and AVIRIS images also exhibit the pattern with a similar southwest to northeast division. However, even after masking the areas where the yield measurements are suspect, the yield data appear to exhibit high spatial variability, making the correlation with the spectral data less reliable.

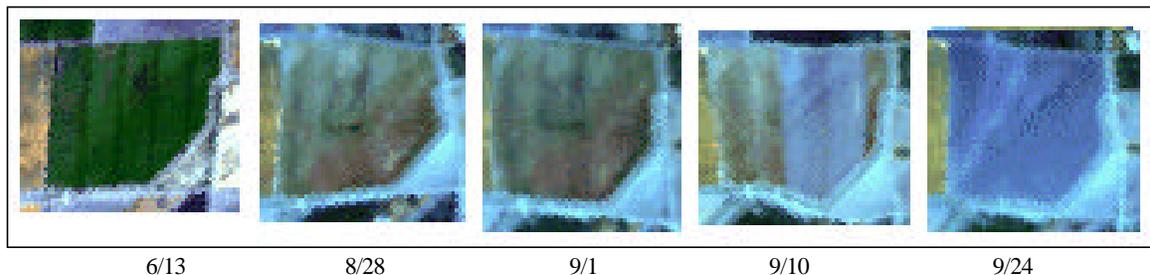


Figure 1. AVIRIS dated images of the sugar beets field.

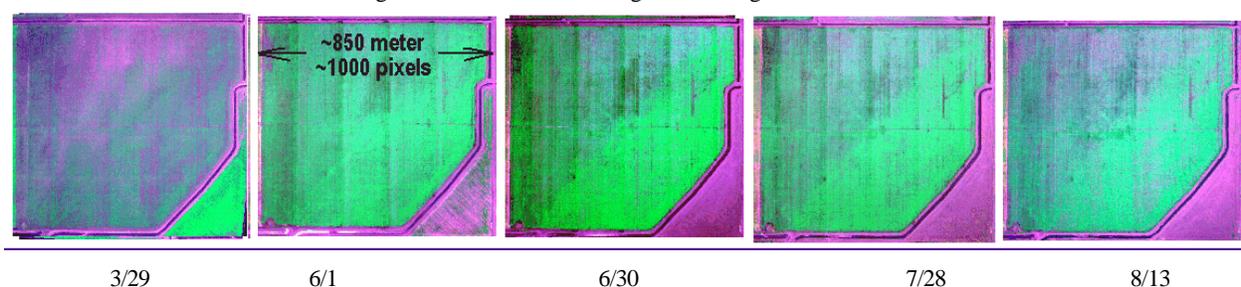


Figure 2. SAMRSS dated images of the sugar beets field. The 1st image shows primarily bare soil. The 2nd and 3rd show increased level of vegetation, while the 4th and 5th show an increase in senescing vegetation from the northwest corner moving later towards the southeast corner. (In this false color image using, B=548 nm, G=658 nm, and R=850 nm, vegetation appears green and soil is purple)



Figure 3. Yield data map at 8 meter resolution (brighter levels correspond to higher yields).

3. ANALYSIS AND RESULTS

The primary objective of the analysis is to establish a spatial relationship between remote sensing measurements and yields. Four aspects of the analysis are discussed:

1. Vegetation Indices
2. Statistics based yield correlations
3. Field averaged indices , and
4. An initial development of new crop health indicators and time integrated indices

From the outset we wish to point that the AVIRIS flight schedule was dictated by an overall AVIRIS season plan, and that the flights did not take place at an optimal time for the purpose of this work. Given the available AVIRIS data, however, we combine the information with the SAMRSS flight data to explore and demonstrate how these data can be used in precision farming applications.

3.1. Vegetation Indices

Typical spectra in the range of the SAMRSS/AVIRIS, and the interpretation of the indices are shown in Fig. 4. The figure exhibits the rationale for the computed indices. The basic indices, and brief comments, are listed in Table 1.

Table 1: Basic Vegetation Physiology, Health & Stress Indices

Index	Sensor	Bands	Wavelength, nm	Comments
NDVI	A	35, 46	0.6637, 0.7782	Common vegetation index indicator of plant health & biomass
	S	2, 3	0.660, 0.850	
NDWI	A	61, 65	0.9325, 0.9710	Water stress index
NDGI	A	19, 23	0.5468, 0.5863	Measure the height of the “green bump”, indicates chlorophyll level and photosynthetic activity.
Greenness	A	14-26	From 0.4981 to 0.6160	The peak wavelength in the “green bump”; senescing vegetation may exhibit a shift to the “yellow”; soil has no bump
HM	A	34-49	From 0.6637 to 0.8068	Wavelength at Half Max. of the red vegetation edge; a shift to longer wavelength may indicate higher photosynthetic activity
INFLEX	A	36-44	From 0.6827 to 0.7591	Wavelength at inflection point (derivative =0) in the red edge; similar to HM
SAVI	A	35, 46	0.6637, 0.7782	Similar to NDVI but with a factor that accounts for fraction of soil

Field-averaged NDVI computed for the AVIRIS and SAMRSS is plotted as a function of date in Fig. 5, along with the $\pm 1\sigma$ spread in the index. It is noted that the March SAMRSS flight took place while the field was still bare with soil (as can be seen from the low index values). Two SAMRSS and one AVIRIS flight took place during the peak canopy coverage, resulting in high NDVI indices.

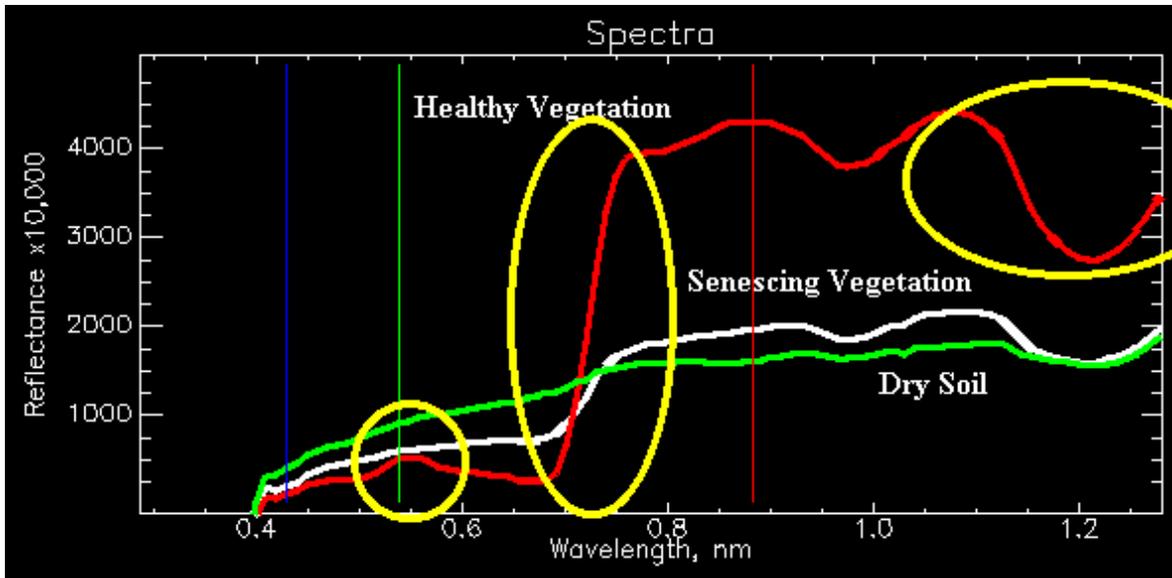


Figure 4. Typical reflectance spectra for healthy vegetation, dead or senescing vegetation, and bare dry soil. The vegetation indices are computed based on features shown in the encircled areas.

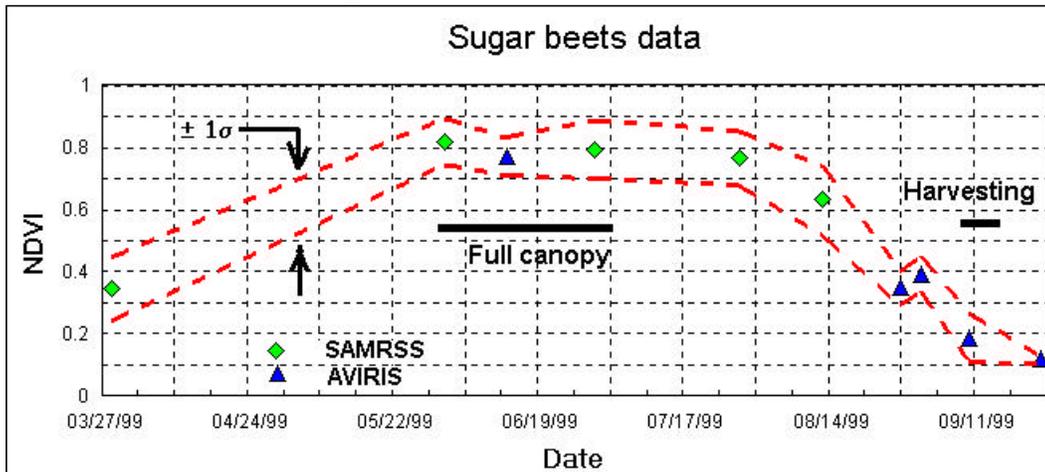


Figure 5. NDVI data from AVIRIS and SAMRSS. The NDVI expresses the height of the 700 nm rise in the reflectance (Fig. 4). Low values correspond to soil or to senescing vegetation.

3.2. Statistical correlations

A simple statistical correlation between yield and any of the indices, or the brightness at specific bands, produces a poor correlation, as depicted for NDVI in Fig. 6a. However, for farm management purposes, establishing general crop indicators may be sufficient. For that purpose the yield levels are classified into 4 management classes, low, medium-low, medium high, and high, as shown in Fig 6b (Fitzgerald *et-al*, 1999). Each level is represented by the mean yield value at that level and shown by triangles in the figure. In the next step, the NDVI data within each level are histogrammed, and the value corresponding to the peak of the histogram is selected to represent the “typical” NDVI value for the yield level. This approach results in the correlation shown by the regression line in Fig. 6c. It is, however, understood that this approach does not bridge the poor correlation exhibited by the full spatial data set in Fig. 6a. Finally, the predictive ability of this approach is depicted in Fig. 6d, where given a measured NDVI, the vertical axis shows the probability that the yield will fall within each of the 4 classes or management levels.

For instance, based on Fig. 6d, areas in the image that exhibit NDVI = 0.7, are 52% likely to be in the medium-low and 42% in the medium-high yield level, or 3.5% and 2.5% likely to be in the high or low yield groups, respectively.

The data distribution in Fig. 6d, suffers from a low number of data points for some classes (120, 830, 1243, and 306 points for the low, medium-low, medium-high, and high classes respectively), and does not exhibit a Gaussian trend. This may indicate that other predictive techniques might be more appropriate. Moreover, given the wide variability in soil conditions, and other parameters that affect the crop development and ultimately the yields, it becomes clear that no single index or brightness parameter can capture all the physiological conditions of the plants that affect ultimate yields. Different approaches are explored in the subsequent sections.

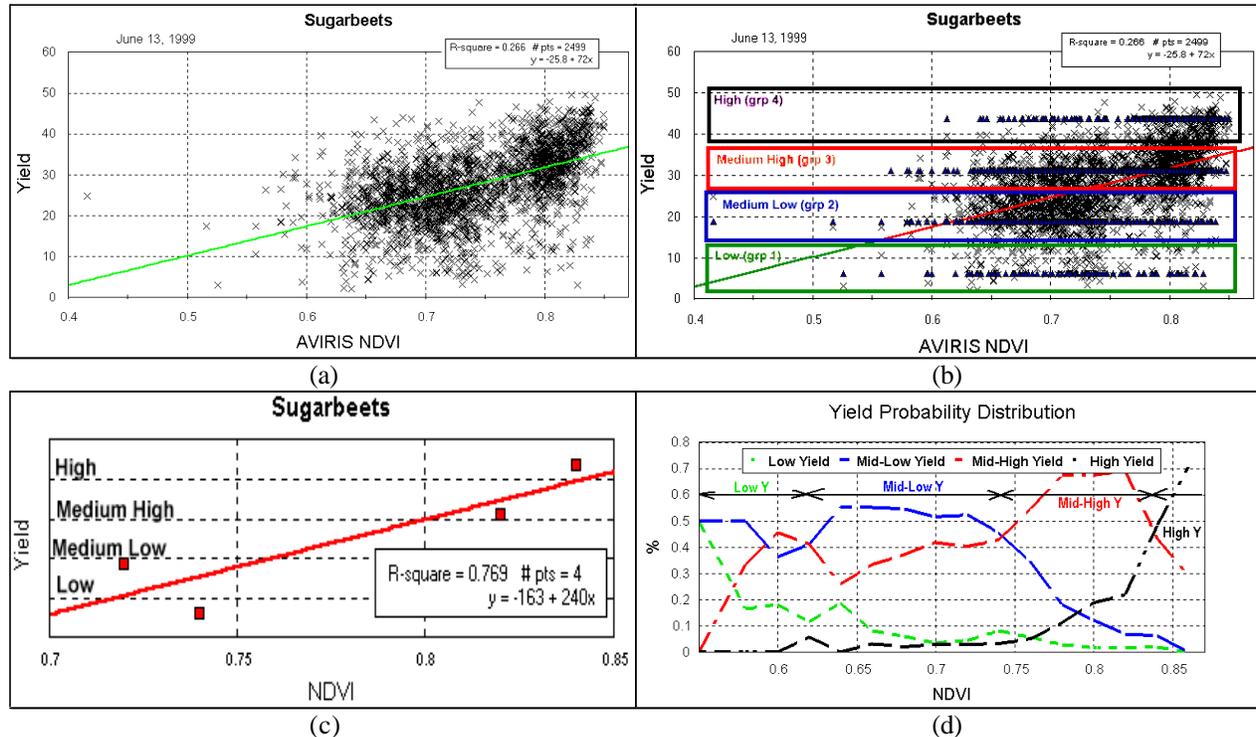


Figure 6. (a) Yield vs. NDVI showing relatively poor correlation, (b) yields are divided into 4 management levels, (c) within each yield management level, the peak of the NDVI index distribution is selected and a linear regression fitted to the data, and (d) shows the potential errors of the regression; at each NDVI level, the probability of achieving each of the yield levels is shown (total probability is 100% at each NDVI point).

3.3. Field-Averaged Indices

Field averaged VIs, although they cannot be used for spatial crop management, may be useful for an overall assessment of the crop status and crop growth stage, as shown in Fig. 5. In the present case the AVIRIS flights took place at less than optimal times, and the field averaged indices serve to demonstrate the crop status. For instance, the field averaged NDVI, Fig. 7a, shows a continuous decline between the June 13 (peak canopy) and the September flights. In early Sept., the NDVI may indicate partial vegetation, and no vegetation at the later dates. All other indices, including the Soil Adjusted Vegetation Index (SAVI) behave in a very similar manner.

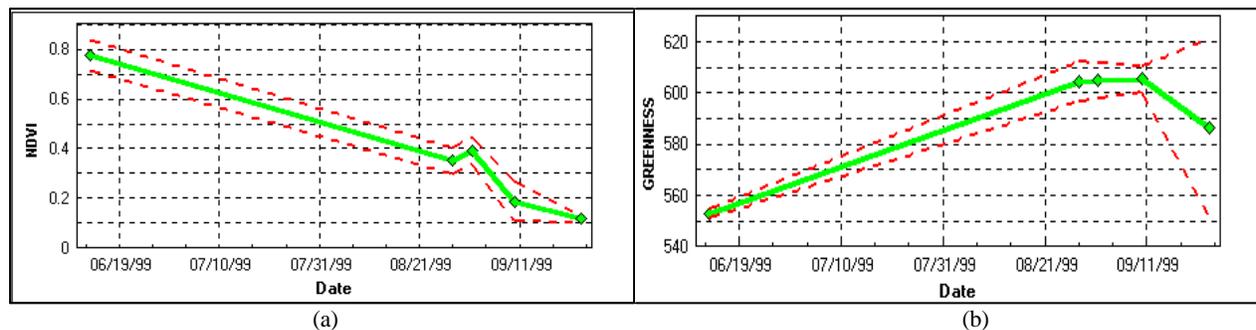


Figure 7. Field averaged indices from AVIRIS flights (a) NDVI, and (b) Greenness. Both indices show full canopy healthy vegetation in June and senescing vegetation in late August, and soil in September.

Some indices express related plant physiological parameters and are therefore closely correlated. For instance, The NDVI seems to correlate well with both the NDGI and the NDWI when plants are at their peak growth, as seen for the June 13 flight in Fig. 8. However this initial high correlation decreases during senescence, Fig. 8c.

Hence as seen from the AVIRIS flights, the field-averaged correlation between the various indices and their temporal change can serve as a gross indicator of the crop evolutionary state.

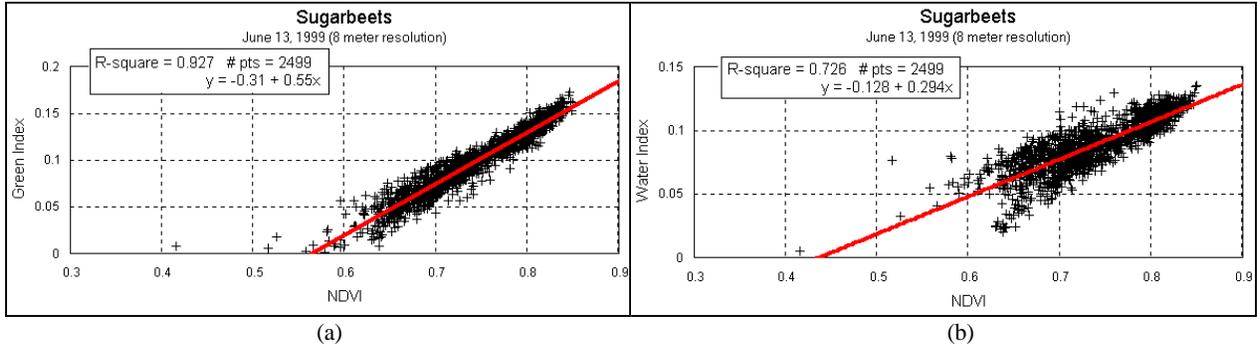


Figure 8. Correlation between the Green Index and NDVI (a), and Water Index and NDVI (b) for the June flight.

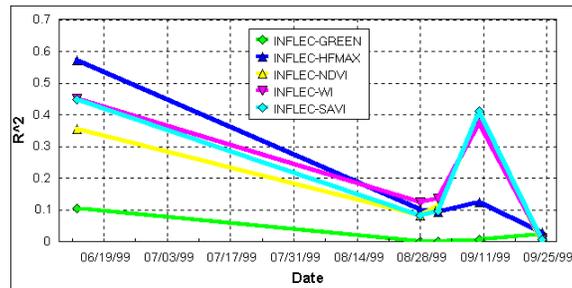


Figure 8c. Temporal change in correlation between several indices.

3.4. Preliminary development of Time-Integrated Indices

3.4.1. VIs and Eigen-Indices

In an ideal precision agriculture application, remote sensing should maximize the utility of data acquired from the air- or space-borne sensor, and minimize requirements for laborious supplemental ground measurements. And indeed this should be feasible since all ground parameters such as soil types, irrigation history, etc., affect the plant and are “reflected” one way or another in the plants’ signature. Hence, our goal is to try to extract the maximum number of independent quantities that characterize the factors that affect plant growth (e.g., soil parameters, nitrogen treatment, irrigation, pest insects infestation, diseases, weather related factors, etc.) via observable symptoms. Such parameters are growth stage dependent, and their cumulative (time-integrated) effects are important.

For instance vegetation indices that are based on the green, red and near infrared portion of the spectrum reveal various plant conditions. However, some indices express closely related information and are therefore correlated. Such correlated indices may be redundant and produce no new information, while other indices may be indicators of unrelated physiological conditions, and as such exhibit a low or no correlation. Such uncorrelated indices provide complementary data. As a plant’s ultimate yield is affected by its growth history and all the environmental parameters, we may need all the possible remote sensing data to develop a functional relationship between plant conditions and yields.

Further more, we are mainly seeking uncorrelated parameters and specifically those that capture the widest variance in the scene. This idea requires decorrelating the indices (Table 1, and perhaps other indices) for instance by means of a Karhunen-Loeve (K-L) transformation. This transformation produces a new set of compound indices, each made of a linear combination of the original indices. These new indices are “orthonormal” and therefore are decorrelated, and we term them eigen-indices (as they are related to the eigen vectors of the K-L transformation).

In a preliminary test of this approach, seven indices were computed for each pixel in the field (based on the 6/13/99 AVIRIS reflectance data) to form seven index images or maps. The index maps were then transformed using the K-L transformation. The correlation matrix and the eigen-values for the transformation are shown in Figs. 9 and 10, respectively.

Correlation Matrix							
	INFLEX	GREEN	HFMAX	NDVI	WI	GI	SAVI
INFLEX	1						
GREEN	-0.332586	1					
HFMAX	0.754731	-0.434009	1				
NDVI	0.593953	-0.392302	0.783831	1			
WI	0.641004	-0.379323	0.881697	0.878044	1		
GI	0.141265	-0.112624	0.191368	0.693186	0.461709	1	
SAVI	0.662248	-0.405109	0.851328	0.976618	0.898334	0.611623	1

Figure 9. Correlation matrix shows for instance that NDVI and SAVI are highly correlated, providing redundant data, while the Green Index is not well correlated with other indices, and perhaps providing independent information.

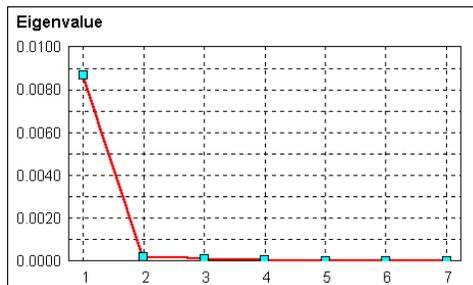


Figure 10. The K-L transformation eigenvalues indicate that most of the information provided by the seven indices is captured in the first principal component that, being a linear combination of all seven indices, captures most of the scene variance.

The question pursued in this case is whether the yield can be better predicted by the transformed eigen indices than any of the original indices. The NDVI index for the 6/13 flight, for instance, produces the correlation shown above in Fig. 6a, with a $R^2=0.266$. Fig. 11a is a scatter plot of the yield versus the first principal component (PC1) of the K-L transformed indices, showing some improvement in $R^2=0.315$. By adding more of the PCs into the yield versus indices regression analysis, the correlation improves further. When yield is correlated to all the PCs (using linear multivariate regression), a $R^2=0.408$ is achieved, as shown in Fig. 11b.

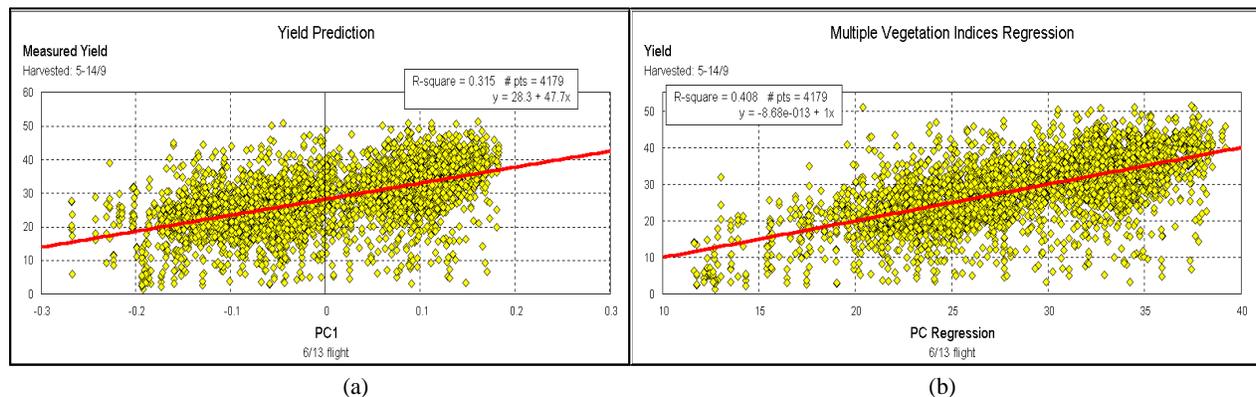


Figure 11. Yields correlation with PC1 (a), and with all seven PCs (note the slope of 1 and zero intercept of the regression curve) (b).

Since the PCs are linear combination of the vegetation indices, and the yield correlates best when all the PCs are included in the regression analysis, we can obtain the same results by simply correlating the yield with the original indices. Indeed, correlating the yield with the original indices produces the same identical result to those shown in Fig. 11b. This observation can be carried further. Since the vegetation indices are linear combinations of the original AVIRIS bands, the yield correlates with the original bands to produce the same results.

To summarize this point in other words, we first computed the VIs as $VI = F_1(R)$ where F_1 is a linear transformation matrix of the spectral reflectance (R). We then applied $PC = F_2(VI)$, where F_2 is a second linear transformation matrix. Finally we developed a regression analysis between yields and the PCs, $Y = f_3(PC)$, where f_3 is a linear transformation vector. Obviously we can state that $Y = F_4(R)$, where $F_4 = F_1 F_2 f_3$.

This points then to another approach. As stated earlier some of the VI are correlated and produce redundant information while there may be other indices that produce independent information. By computing seven indices we have obviously not extracted all the information contained in the AVIRIS data. For instance the lignin and protein in the plant produce signatures in the SWIR range that was not used here. The obvious implication here is “Why not try to correlate yield with all the AVIRIS bands?” Obviously there will be much redundant information and it is possible to find which bands make negligible contribution to the correlation and ignore those in future analysis.

Another question of significance is whether other parameters (measured from ground information) can improve the predictions. This question is tentatively answered by Fig. 12. No significant correlation between the electrical conductivity (EC, a parameter that is related to the soil salinity and hence an important factor in the plant’s growth) map and yield map ($R^2=0.024$) is shown in Fig. 12a. When the EC however, is added as an eighth map to the seven VI maps and the regression analysis between yields and these maps is repeated, a $R^2=0.411$ is obtained. The slight improved correlation from Fig. 11b, may be statistically insignificant. It may, however, indicate that the soil salinity variance, is reflected in the plant signatures and that the contribution of salinity is already accounted for by the reflectance signatures. This point must be further explored before conclusive evidence is produced.

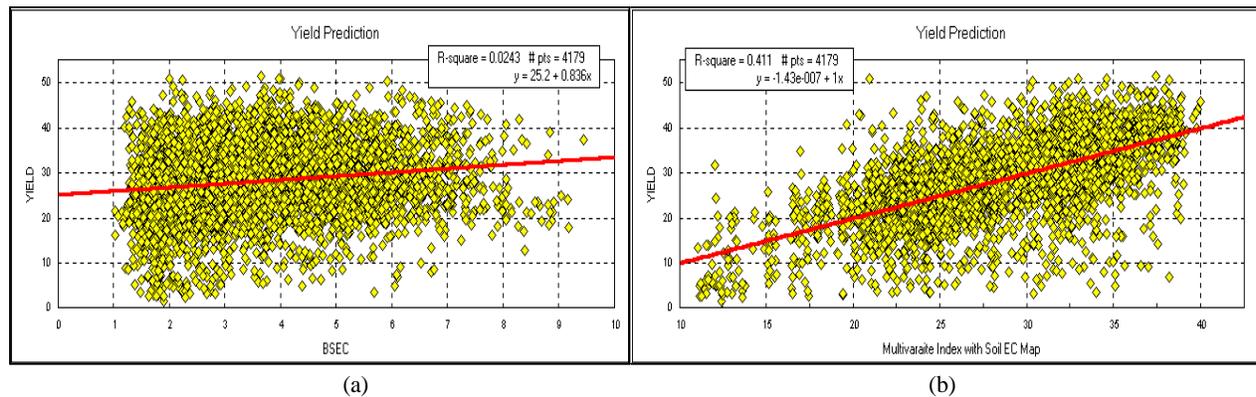


Figure 12. Yield correlation with Soil electrical conductivity (a), and with all eight indices including seven VIs and EC (b).

3.4.2. Time-Integrated Indices

While exploring correlations between yield at harvest time and plant conditions at any given time during growth season, two questions come to mind. First, “Can events that happen after the remote sensing data were collected drastically affect the yield?” For example higher than usual rain, or drought, or the onset of a disease, and many other factors, can completely change the outcome that present parameters can’t account for. The second more subtle question is: “Can stress residues from past events, that may have been corrected, that affect the yield?” How to capture all these considerations?

If the answer to the second question is negative, than monitoring present parameters should suffice for predicting yield. A parameter temporal profile may be established and if the crop is on schedule, than yields can be predicted. If the answer is affirmative though, the situation is a little more complex.

In the latter case the solution may be in establishing a time-integrated value of the indices, and then attempt to correlate the yield with such values. For instance, the AVIRIS computed VIs for June 13, Aug., 28, and Sept., 1 were used in the following analysis (the Sept., 10 and 24 flight data were discarded since the field was in the middle or completed of harvesting). Results, shown in Fig. 13, indicate as before that a single index may not be a sufficient predictor ($R^2=0.287$) while a combination of the time-integrated indices produce a better correlation ($R^2=0.406$). One caveat must be introduced, however. Since the elapsed time between the June and Aug. flights is 76 days, while that between the next two flights is only 4 days, the contribution of the latter period is negligible relative to the former.

This is why the correlation in Fig. 13b is practically identical to that in Fig. 11b. The yield predictions by the two methods (time-integrated and the VIs correlation for the June 13 flight) are compared in Fig. 14.

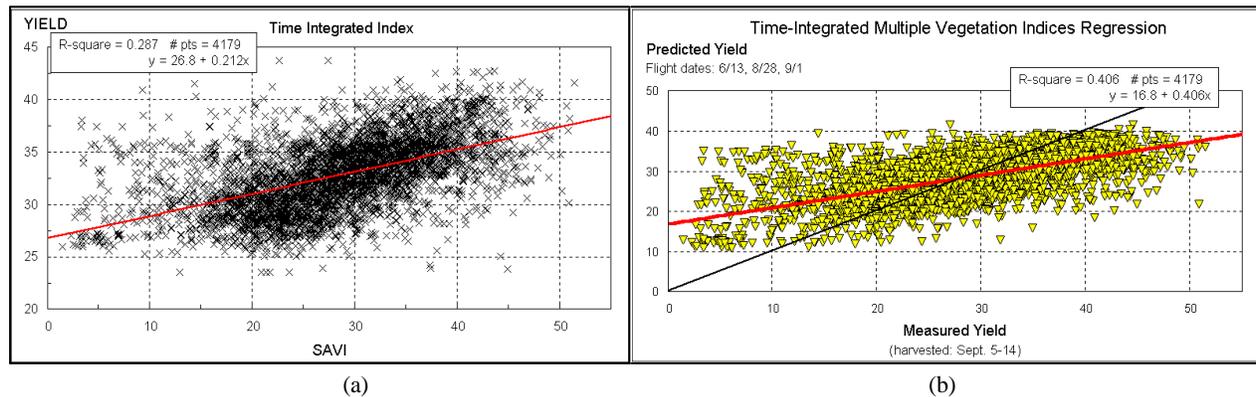


Figure 13. Yield correlation with time integrated SAVI (a), and with all seven time integrated indices (b) where the black line is a 45° expected result.

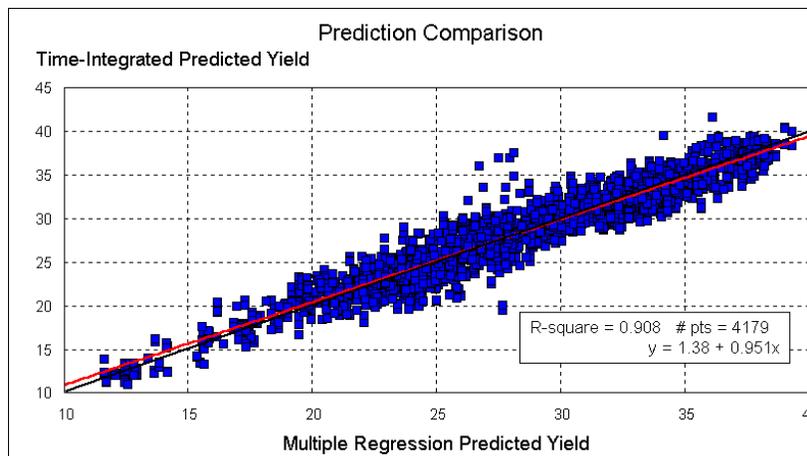


Figure 14. Comparison between the yield prediction by time-integrated indices over June 13-thru-Sept. 1, and the VIs for June 13. The small deviations from the straight line are due to the Aug. 28-thru-Sept. 1 contribution.

To improve this time-integrated analysis we would need the remote sensing data at frequent intervals during the more critical period of the crop growth. Moreover, we have already seen that the high correlation during June 13 between some VIs is lost at later dates. Thus instead of using the time-integrated VIs, perhaps time-integrated reflectance data might be used.

One way to use the time-integrated indices is in monitoring deviations from established evolution of the indexes. Such deviations may serve as an early warning.

4. SUMMARY DISCUSSION

This investigation suggests the development of crop health status indicators built upon time-integrated linear regression of hyperspectral reflectance crop data. The physical parameters that affect crop health and ultimately yield are manifested in physiological changes that are reflected across the spectrum. The yield may be predicted by a functional relationship (not necessarily linear) among many bands, in particular those that produce independent data. The traditional vegetation indices alone, may not capture all the required information and more bands may prove useful. If linear relationships are used, this investigation suggests working directly with the reflectance bands as any linear transformations to vegetation indices, or to an uncorrelated space, still expresses the same information.

The need for a hyperspectral (as opposed to a multispectral) sensor for commercial applications is supported by several considerations: (i) different crops may require different bands for allowing precision farming based on remote

sensing, (ii) even in a single crop, the optimal bands may change (1) during the growth season, and (2) when the crop grows in a different environment (e.g., geography, climate, geology). A commercial operation will not succeed when limited to a single crop because most growers deal with several crops, and the spatial distribution of crops provides an opportunity to cover everything within the flight path.

The ability to use remote sensing measurement to predict yield in this work has been limited by the fact that some areas of the field were affected by extensive weed growth. Without the ability to differentiate between weeds and actual crop, the indices that exhibit plant vigor are affected by the weeds, while the yield may be very low. This is a topic that needs to be addressed in further work.

Time-integrated correlations in this investigation suffer from lack of sufficient temporal coverage at the crucial growth period. Remote sensing flight schedule and the immediate availability of data are so crucial to investigations such as this, that future work should rely on the use of a dedicated sensor that can be flown on a weekly basis. The SAMRSS data, although limited in the number of bands, proved very useful in providing immediate temporal data.

A dedicated sensor is not only needed for commercial real time crop management, but also for development of the algorithmic bases and for building up the confidence level that is required by the agro-industry infrastructure of growers, farm advisors, university extension consultants, county, state, and federal monitoring agencies, crop insurers, and others.

ACKNOWLEDGMENT

The analyses conducted under this investigation were made possible under the NASA SSC Hyperspectral EOCAP project. AVIRIS flights were provided courtesy of Dr. Robert Green and the JPL AVIRIS crew; NASA SSC and HQ sanctions for additional AVIRIS collects are also acknowledged.

LITERATURE CITED

- Fitzgerald, G.J., Kaffka, S.R., Corwin, D.L., Lesch, S.M., and Maas, S.J., 1999. "Detection of soil salinity effects on sugar beets using multispectral remote sensing," Presented at the Agronomy Society meetings in Salt Lake City (available on line at: http://pwa.ars.usda.gov/uscrs/sugar_beets.htm)
- Gamon, J.A., Lee, L.F. Qiu, H.L. Davis, S. Roberts, D.A. Ustin, S.L. 1998. "A multi-scale sampling strategy for detecting physiologically significant signals in AVIRIS imagery." Summaries of the seventh JPL airborne earth science workshop, JPL publication 97-21, Vo1. 1, AVIRIS workshop, R.O. Green, Ed., Jet Propulsion Laboratory, Pasadena, CA, 111-120.
- Green, R.O., Pavri, B. and Roberts, D.. 1998. "Mapping agricultural crops with AVIRIS spectra in Washington state." Summaries of the seventh JPL airborne earth science workshop, JPL publication 97-21, Vo1. 1, AVIRIS workshop, R.O. Green, Ed., Jet Propulsion Laboratory, Pasadena, CA, 213-220B.
- Kaffka, S.R., Corwin, D., Lesch, S., and Fitzgerald, G.J., 2000 "Field-scale soil electrical conductivity characteristics and sugar beet emergence, growth and yield," Proceedings 28th California Plant and Soil Conference, Jan 19-20, 2000, CA chapter of American Society of Agronomy.
- Kokaly, R.F., and Clark, R.N.. 1998. "Spectroscopic determination of leaf biochemistry: Use of normalized band-depths and laboratory measurements and possible extension to remote sensing measurements." Summaries of the seventh JPL airborne earth science workshop, JPL publication 97-21, Vo1. 1, AVIRIS workshop, R.O. Green, Ed., Jet Propulsion Laboratory, Pasadena, CA, 235-244.
- Moran, M.S., Clarke, T.R., Qi, J., Barnes, E.M. and P.J. Pinter, Jr. 1997. "Practical techniques for conversion of airborne imagery to reflectances." 16th Biennial Workshop on Color Photography and Videography in Res. Assessment.

OKSI web site: <http://www.oksi.com>