

FROM PIXEL TO YIELD: FORECASTING POTATO PRODUCTIVITY IN LEBANON AND IDAHO

H. Abou Ali¹, D. Delporte^{1,*}, L. M. Griffel²

¹ Idaho State University, Department of Geosciences, 83209 Pocatello, Idaho, USA - (abouhana, delporte)@isu.edu

² Biofuels and Renewable Energy Technology Dept, Idaho National Lab, 83415 Idaho Falls, Idaho, USA – mike.griffel@inl.gov

KEY WORDS: yield prediction, vegetation indices, PlanetScope, remote sensing, precision agriculture, crop monitoring

ABSTRACT:

Idaho and Lebanon rely on potatoes as an economically important crop. NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), SAVI (Soil Adjusted Vegetation Index), and MSAVI2 (Modified Soil Adjusted Vegetation Index 2) indices were calculated from PlanetScope satellite imagery for the 2017 growing season cloud free days. Variations in vegetation health were tracked over time and correlated to yield data provided by growers in Idaho. Based on ordinary least squares regression an Idaho yield forecast model was developed. Vegetation response during the growth stage at which potato tubers were filling out was significant in predicting yield for both the Norkotah and Russet potato variety. This corresponded to a week with high recorded temperatures that impacted the health status of the crops. The yield forecasting model was validated with a cross validation approach and then applied to potato fields in Lebanon. The Idaho model successfully displayed yield variation in crops for Lebanon. Spectral indices along with field topography allow the prediction of yield based on the crop type and variety.

1. INTRODUCTION

Agriculture is an important sector in the global economy and is a crucial component for fighting hunger and food insecurity. According to the United Nations Food and Agriculture Organization (2017), the expectation is that the world population will reach 10 billion by 2050 and there is a need to produce around 50% more food than in 2012. Idaho and Lebanon rely on potatoes as an economically important crop. Potatoes are responsible for employing 46% of the Idahoan agricultural processing workforce (Lewin et al. 2011) while in Lebanon potatoes account for 56% of vegetable production in the country, mainly in the Bekaa Valley and North Lebanon states (Hatoum 2005).

Precision agriculture uses information technology to better manage crop production by taking into consideration the variations within the field to increase profitability and sustainability (Zhang 2016). By utilizing remote sensing data from satellites systems, farmers can leverage cost-effective technologies to mitigate crop threats with targeted approaches for grower decision making such as variable rate fertilizer application, timely irrigation, early disease detection, and pest control. Introducing the concept of precision agriculture plays a major role in empowering local farmers and stakeholders by educating them about new technologies that have the potential to improve their crop productivity and enhance their economic sustainability.

New satellite systems have emerged in the past five years that have high spatial, temporal (PlanetScope), and spectral resolution (Sentinel-2) that have the potential to revolutionize the field of precision agriculture by aiding decision making. These advances in satellite imagery provide a valuable resource for crop monitoring and specifically yield modeling and prediction.

Crop yield forecasting is crucial for farmers as it helps with management decisions regarding harvesting, storage, pricing and marketing. The availability of satellite imagery with higher spatial and temporal resolution provides detailed information about crop yield over the duration of the growing season. A previous research study assessed the performance of different sensors with varying resolutions (5 m for RapidEye, 3 m for PlanetScope, 10 m for Sentinel-2 and 30 m for Landsat) for yield monitoring and concluded that higher resolution imagery gave more accurate results (Burke and Lobell 2017). Regression models are commonly used in yield prediction. Linear regression uses one variable at a time to explain and predict the yield using separate equations for each (Aditya Shastry 2017) assuming that there is a linear relation between the yield and the explanatory variable, model residuals are almost normally distributed such that there is no clustering in the data (Sellam and Poovammal 2016). In their paper, Rembold et al. (Rembold et al. 2013), used linear regression with NDVI (Normalized Difference Vegetation Index) from SPOT (Satellite Pour l'Observation de la Terre) imagery to explain yield values and obtained an R^2 of 0.930 and 0.799 for the Morocco and Egypt study areas, respectively. Al-Gaadi et al. (2016) also used linear regression but used Landsat-8 and Sentinel-2 imagery for computing both NDVI and SAVI (Soil Adjusted Vegetation Index). They generated a linear regression equation per index per sensor per field and concluded that higher resolution imagery yielded better regression models where the Landsat-8 imagery resulted in an R^2 range between 0.39 and 0.65 compared to 0.47 and 0.65 for the Sentinel-2 dataset. While previous studies on modelling yield prediction used Landsat (Song et al. 2016) and Sentinel-2 (Al-Gaadi et al. 2016), this paper offers a new approach utilizing high spatial and temporal resolution PlanetScope (3 m, daily) analytical scenes in comparison with Sentinel-2A (20 m, 10 days) to monitor potato crop health over the 2017 growing season and predict crop yield. The most common vegetation indices used for forecasting crop yield status are NDVI, Green Normalized Difference Vegetation Index (GNDVI), SAVI and Modified Soil Adjusted Vegetation Index 2 (MSAVI2). These indices were all tested and compared to potato yield data from ten fields (402.98 ha) in an ordinary

* Corresponding author

least squares (OLS) regression approach to choose the best performing variable for yield prediction. The critical week for each of the different potato varieties corresponded to the potato growth stage where tubers are in the process of filling out beneath the soil. Using the averaged SAVI and slope values for the critical weeks provided the best fit model for predicting yield in the Idaho fields. Upon considering the offset in the growing seasons between Idaho and Lebanon as well as the potato varieties, the adjusted Idaho model was applied to forecast potato yield in Lebanon.

2. METHODS

2.1 Study Areas

For the purpose of this research, we selected two study areas (Figure 1 and Figure 2) that shared topographic, geographic and environmental features including related crop threats. Both study areas have cold wet winters and hot dry summers and are within river plains and part of the North Temperate Zone. The Idaho study area is located at a latitude of 43.9° and altitude of 1502 m above mean sea level with a wet season from October through June with a high average temperature during the 2017 season of 24°C. The Lebanon study area is at 10° lower latitude than the Idaho study area, located at 33.9° latitude with an elevation of 872 m above sea level and a high average temperature of 34°C during the 2017 growing season. Moreover, potatoes are the dominant crop in both areas.

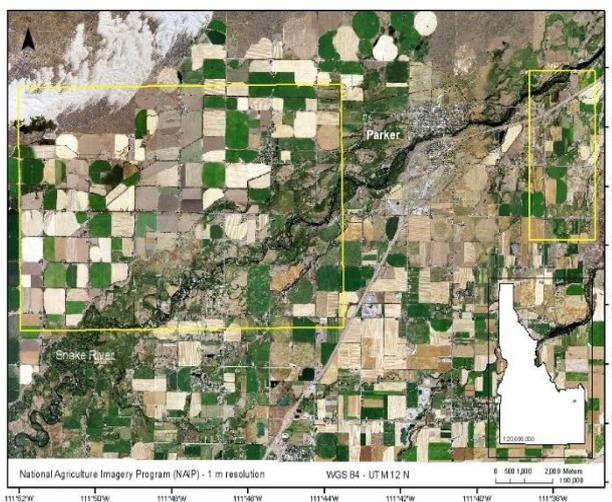


Figure 1. Idaho study area showing field locations near Parker, Idaho. (WGS 84 - UTM 12N)

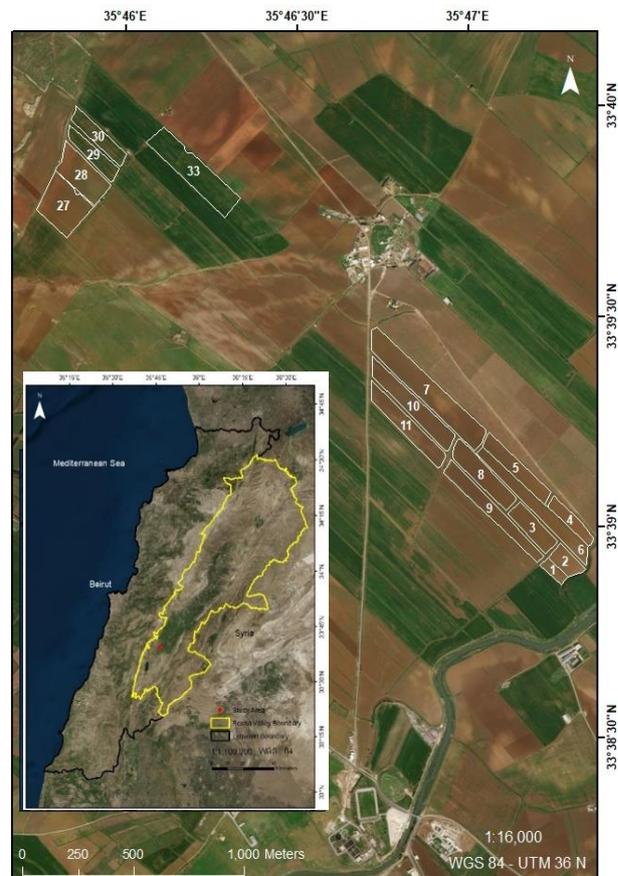


Figure 2. Lebanon study area field locations in the Bekaa Valley, Lebanon. (WGS 84 - UTM 36N)

2.2 Satellite Imagery

The satellite imagery used in this work includes Sentinel-2A and PlanetScope data. The USGS (United States Geological Services) Earth Explorer online interface allowed open access to the Sentinel-2A imagery where the data is freely available to the public via the ESA (European Space Agency). The PlanetScope imagery operated by Planet, a non-governmental privately-owned commercial company, was accessible through the Planet Application Program Interface using their Education and Research Program license. At the time of this study, in the summer of 2017, Sentinel-2B imagery was unavailable.

Sentinel-2A is a multispectral imager covering 13 spectral bands (443 nm – 2190 nm) at resolutions of 10-20 and 60 m with a five-day revisit period. The Sentinel-2A imagery was processed using ESA's SNAP software for atmospheric correction using the Sen2Cor Plugin. The PlanetScope imagery has four spectral bands (455 nm – 860 nm [Blue: 455-515, Green: 500-590, Red: 590-670, NIR: 780-860]) at a resolution of 3 m with a daily revisit. The downloaded PlanetScope imagery is the surface reflectance ready product provided by Planet so no further atmospheric correction was needed, however, individual bands had to be extracted.

After extracting red, green and NIR bands, and atmospherically correcting the Sentinel-2A bands (Figure 3), the next step was to calculate the various spectral indices that are described below.

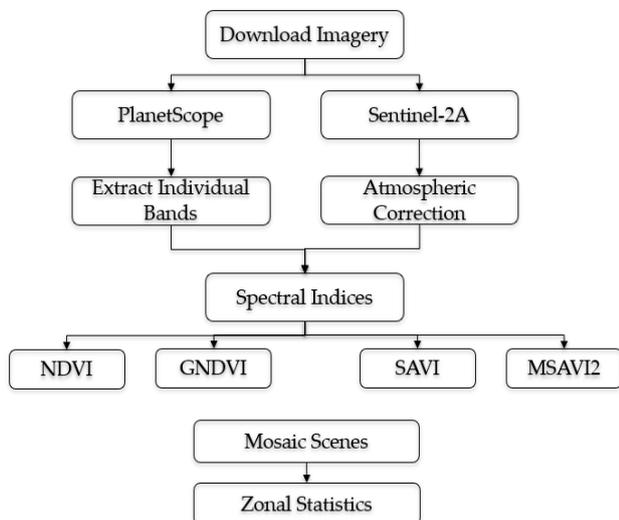


Figure 3. Satellite imagery processing workflow

NDVI (Rouse et al. 1973)

$$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

GNDVI (Gitelson, Kaufman, and Merzlyak 1996)

$$\frac{\rho_{NIR} - \rho_{Green}}{\rho_{NIR} + \rho_{Green}} \quad (2)$$

SAVI (Huete 1988)

$$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red} + L} * (1 + L) \quad (3)$$

MSAVI2 (Qi et al. 1994)

$$\frac{2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_{Red})}}{2} \quad (4)$$

In order to better organize the data, any dates that had multiple scenes for the same date were mosaicked to reduce the number of scenes for processing further. The final step for processing the satellite imagery was running zonal statistics that would be the input for the regression models to predict yield.

2.3 Yield Data

Idaho farmers supported the data analysis by providing information for each field that included planting dates and the potato variety. The farmers provided yield data from the 2017 growing season, which was collected using the GK Technology for Agriculture yield monitor. The sensor attaches to the harvester and records the speed, yield, and pounds per second along with the geographic coordinates every 0.8 meters by 10 meters.

Based on planting date and potato variety, the Idaho fields were sub grouped into three categories (G1, G2, G3) so that fields

within the same group have a similar growing timeline which influences the variables for yield forecasting models depending on date and variety. Different planting dates result in offset between models while each potato variety has its own growth cycle (different maturity periods).

2.4 Yield Forecasting

A Python script was written to extract all mean values for each index and bands red, blue and NIR along with the mean yield values for each grid cell making it suitable for running regression analysis to explain yield variability within the fields. The Python code calculated the average values within each grid cell for every date available. This resulted in multiple averaged cells for each index by growth week depending on cloud cover. Exploratory Regression in ArcGIS Pro helped reduce the number of potential explanatory variables for the yield model through several runs. Subsets of the explanatory variables included a selection of the Green, Red and Near Infrared bands while others used the different indices: NDVI, GNDVI, SAVI, and MSAVI2. In addition, a subset included a combination of bands with spectral indices. From each run, the most significant variables became the new subset for the next run. After three exploratory regression runs, the variables that had the highest correlations were used for input into the Ordinary Least Square (OLS) tool. The resulting models were validated using a cross validation leave-one-out approach (Figure 4). In addition, slope values were calculated and averaged within the fields based on the grid cells using the National Elevation Dataset (10m resolution).

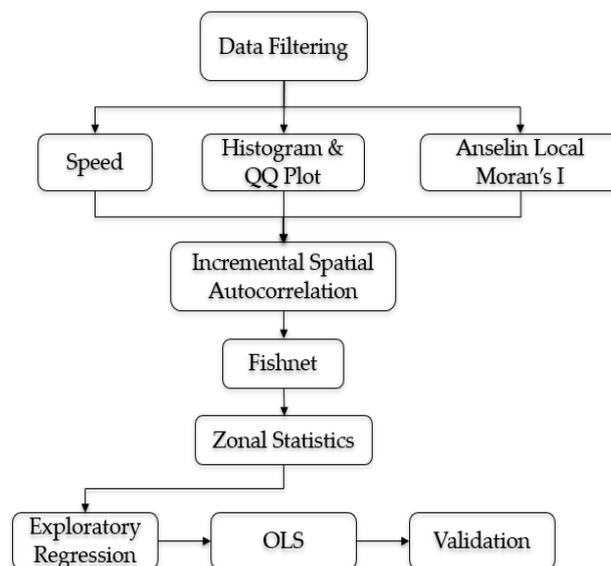


Figure 4. Yield data processing workflow

An incremental spatial autocorrelation approach determined the optimal grid size by examining the corresponding z-score of yield data. This approach utilizes Global Moran's I to determine yield clustering by measuring spatial autocorrelation based on yield points' location and value. An output of this method is the z-score which is the standard deviation returned by Global Moran's I where the peak in z-score corresponds to highest yield clustering density. The result of the incremental spatial autocorrelation method indicated that an 80x80 m grid size was the peak distance for yield value clustering.

3. RESULTS

3.1 Variation of Indices over Growing Season

All five indices used (NDVI, GNDVI, SAVI, and MSAVI2) showed similar variation over the fields throughout the season in both Sentinel-2 and PlanetScope data (Figure 5). However, the PlanetScope imagery gave a more detailed observation about what was happening in the fields along with determining critical dates and stages for the growing season due to its higher temporal resolution. The PlanetScope dataset corresponded to specific dates regarding plant emergence, row closure and full maturity in addition to dates with higher temperatures and increased evapotranspiration rates.

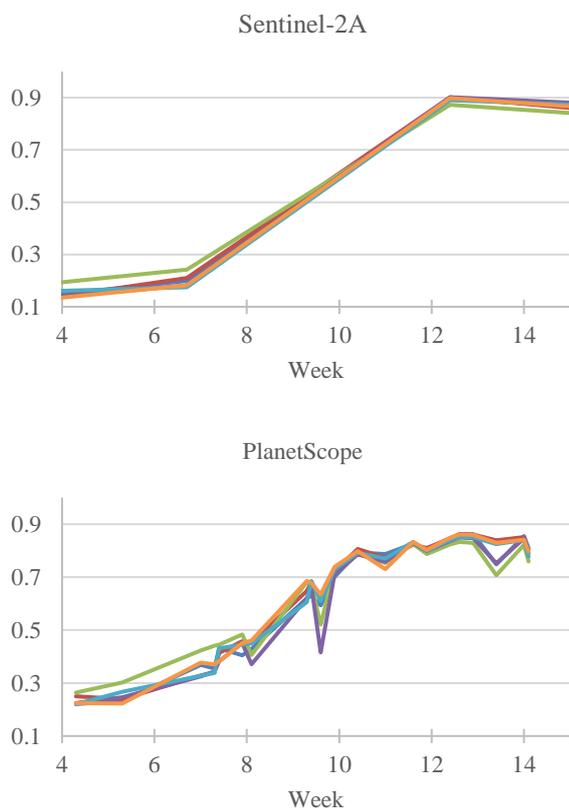


Figure 5. Variation of NDVI over the 2017 growing season Sentinel-2A vs. PlanetScope for Idaho fields

In both the Sentinel and Planet imagery for all the fields, NDVI and SAVI increase gradually during the season to reach a peak value of 0.9 and then decrease to around 0.8 (Figure 5). However, the limitation of Sentinel-2A availability due to cloud cover resulted in fewer data points between critical times over the growing season. PlanetScope imagery gave a more detailed interpretation of the variation between the weeks and the fields. For Lebanon, the cloud free Sentinel-2A dataset consisted of 10 dates, representing only 10 weeks during the growing season. While with PlanetScope there was imagery for 50 available dates for the same time period with an average of 4 or 5 images per week and covered the entire growing season. For Idaho, cloud free Sentinel-2A imagery consisted of 5 dates, representing only 5 growth weeks. For PlanetScope there was 22 available dates for the same time period with minimum of 2 images per week to provide average weekly vegetation response over most of the growing season.

3.2 Idaho Yield Models

The dependent variable, yield, was tested individually against each of the average weekly indices along with field slope. For all fields' groups, the different vegetation indices allowed the explanation and prediction of yield values. With access to only Sentinel-2A (Sentinel-2B was unavailable for the 2017 growing season), there were only five cloud free Sentinel-2A images, representing 5 growth weeks (weeks 1, 3, 7, 11, and 13) which was not sufficient to test a regression model. On the other hand, there were 24 available PlanetScope dates, covering 14 growth weeks, which was enough for building the yield forecasting model. The model calculated R^2 values by growth week over the entire growing season for each index plus slope. The best-fitting model corresponding to growth week was identified for each potato variety.

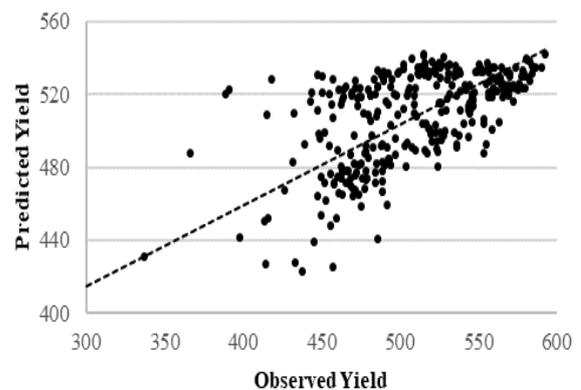


Figure 6. Russet potato fields yield model results

The Russet potato variety yield prediction model (Figure 6) with SAVI response during week 12 as the most significant for yield prediction. The model obtained has an R^2 value of 0.44 and follows the equation of: $y = 0.4435x + 281.6$

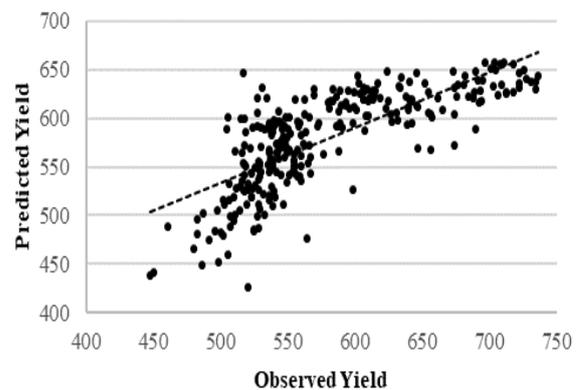


Figure 7. Norkotah potato fields yield model results

The Norkotah potato variety yield prediction model (Figure 7) with SAVI response during week 10 as the most significant for yield prediction. The model obtained has an R^2 value of 0.57 and follows the equation of: $y = 0.570x + 248.8$

3.3 Growth Degree Days and SAVI

Potatoes are an irrigated crop and water is a crucial element in the productivity of the fields. Although irrigation rate was

unavailable for this study, we retrieved precipitation and temperature records (Figure 8) to relate to indices' and yield values. A better understanding of these weather variables provided vital information related to critical weeks in the growth stages has the potential to help farmers with irrigation decisions.

In addition to the relationship between SAVI and yield, there was also a relationship between SAVI and GDD (Growth Degree Days). The correlation coefficient between SAVI and averaged weekly GDD (Equation 1) values for the individual fields was between 0.45 and 0.67 (Figure 9).

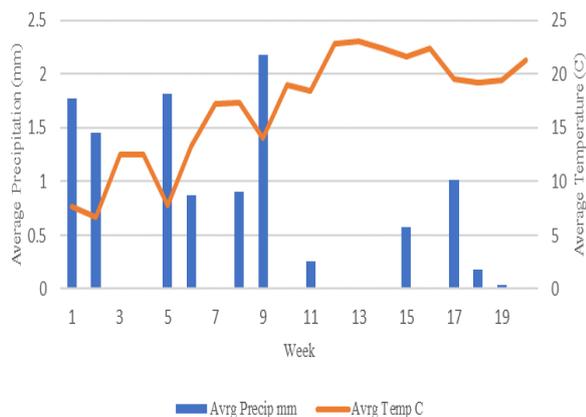


Figure 8. Precipitation and Temperature during the 2017 growing season provided by U.S. Bureau of Reclamation Agrimet weather station.

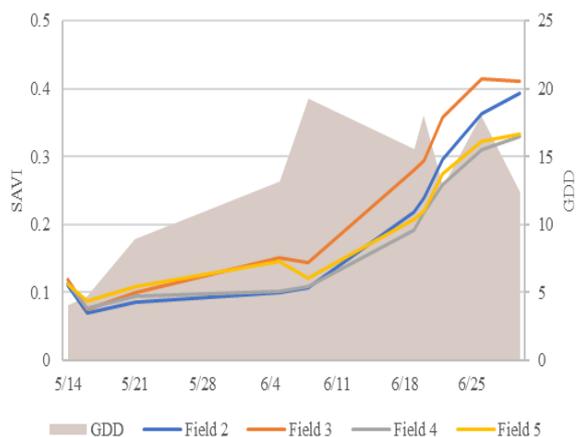


Figure 9. The variation of SAVI and GDD for the 2017 range of critical growth weeks (6 to 13)

Growth Degree Days:

$$(\text{MaxAirTemp} + \text{MinAirTemp}) / 2 - \text{Base Temp} \quad (6)$$

[Base Temp = 50F]

3.4 Yield Forecasting in Lebanon

The regression model derived from the Idaho data predicted yield for fields in Lebanon. Due to the difference in potato variety, the approach was to calculate yield using the different models. Though the Idaho model used averaged values of the variables within a fishnet cell size of 80m x 80m, the fishnet used to predict yield in the Lebanon data was 32m x 32m. The reason for that choice is the fact that the Lebanese farmer gave an approximation

of the yield to have had a range of 3 ton to 5 ton per 1,000 m². Hence, a 32m x 32m cell grid is the closest approximation to what the farmer provided.

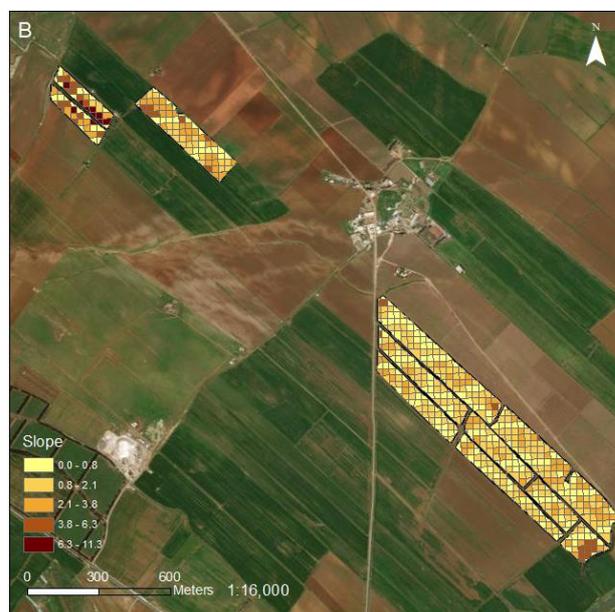


Figure 10. Predicted yield values in kg/1000m² (A) and the corresponding slope values in degrees (B)

The micro topography of the fields affected the predicted yield values due to the variation in slope leading to uneven irrigation. The results obtained from the yield model were verified qualitatively by showing the farmer the yield variation within the fields and confirming with his records. The farmer also mentioned that there are three sprinklers every 100 meters and as the yield values are clustered in a 100-meter neighborhood this could potentially show the positions of the installed sprinklers.

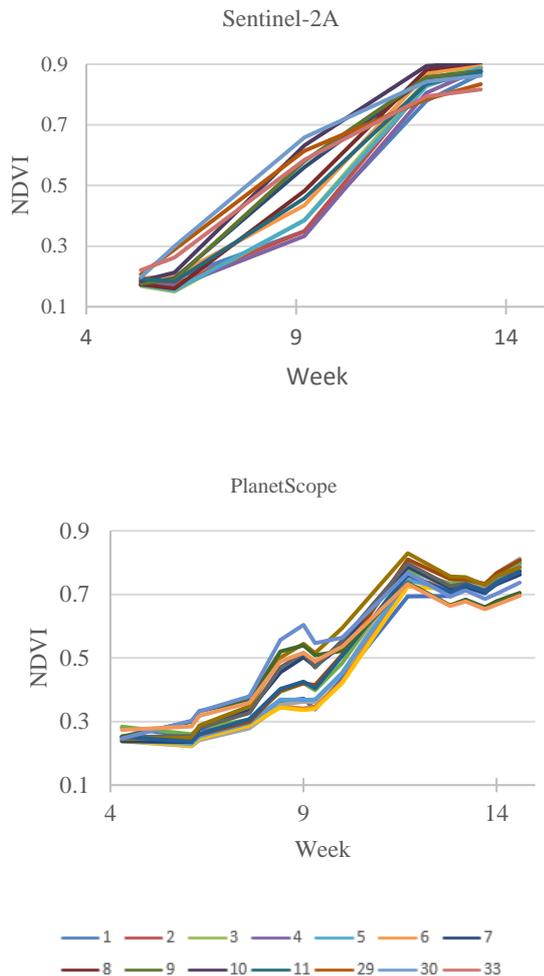


Figure 11. Variation of NDVI over the 2017 growing season Sentinel-2A vs. PlanetScope for Lebanon fields

4. DISCUSSION/CONCLUSION

Over the past five years, new offerings of satellite imagery are providing high spatial and temporal resolution data that advances opportunities for monitoring crops and predicting yield. The high resolution and daily return frequency of PlanetScope allows farmers to apply precision agriculture practices that take in to consideration the variation within the field as opposed to the assumption that variables such as microtopography and soil types within the field are uniform. The higher spatial resolution of PlanetScope is particularly valuable with smaller fields such as those in Lebanon where fields are usually small in size as most of the fields are inherited from father to sons and thus get subdivided when passed from generation to another. With each spectral band delivering specific information about the plants based on the reflectance signature, combining slope data with vegetation index values – particularly SAVI or NDVI daily, growers can rapidly respond to crop health threats. Vegetation indices provide information about the health status of the crop and can help with decision making regarding adding nutrients and water to plants when needed based on the index value. Further, our study demonstrated that SAVI values averaged by week from PlanetScope imagery revealed the growth stage and growth week that was the most critical for impacting potato yield. Though the critical week varied between potato varieties (ranging

between week 10 and 12), it corresponded to when the tubers were filling, while the later weeks that corresponded to the crops reaching full maturity had no role in forecasting yield. On the contrary, as the crops reached full row closure and continued to mature before harvesting showed a decrease in the correlation between the indices and yield values.

Our study also highlighted the importance of frequent satellite revisit periods. With only 5 cloud free image dates across the entire 2017 growing season in Idaho with no dates during the critical growth weeks as indicated by PlanetScope and Sentinel 2B data not yet available, Sentinel 2A alone was insufficient to identify critical growth stages or plant response to predict yield. With PlanetScope’s daily revisit there was a minimum of 2 cloud free images available per week for Idaho. The Lebanon study area had an even larger number of available PlanetScope images due to more cloud free days (50 over the growing season compared to Idaho’s 24) that provided a more frequent observation period of crop variation over the growing season. The graphs from the PlanetScope imagery clearly corresponded to the different stages of growth for potato plants within the study area. As the plants started to emerge around week 4, the indices increased gradually to show a slight peak. The maximum peak within the graphs represented the full maturity of the crops, indicating that they were ready for harvest. A significant drop in the curves also appeared at week 9. This drop indicated high stress levels in the crops related to high recorded temperatures.

Norkotah potato variety crops reached full maturity two weeks earlier than the Russet crops. This was evident by the difference in the critical growth week based on the correlation values and the regression models for the two varieties where the critical week for Norkotah groups was week 10 while the critical week for the Russet group was week 12. All the regression models indicated that SAVI, along with average slope, best explained and predicted yield. The indices’ values of the weeks beyond week 12 showed a decreased relationship to yield and thus not considered a good indicator for yield prediction. SAVI proved to be a better indicator of yield over NDVI due to the limitation of NDVI and its relationship with the leaf area index. The variable L in the calculation of SAVI takes into consideration the soil within the fields unlike NDVI and this is crucial in fields with early season soil exposure between plants.

Access to high resolution and accurate yield data collected from the GTK sensor installed on potato harvesters in Idaho (0.8 m x 10 m) coupled with high resolution PlanetScope imagery (3 m) facilitated the development of a yield forecasting model that was suitable to apply in another region (Lebanon) with smaller fields. With agricultural lands being passed down between generations through inheritance, mainly in developing countries such as Lebanon, yield prediction models for smaller fields are a global need and an essential tool towards food security and sustainability. This model can be modified for other crop types at any given geographic location. The key information for adapting this model and applying to other crops is identifying the critical week as it differs between crops and varieties of the same crop as demonstrated for the Norkotah and Russet potato varieties within this research.

ACKNOWLEDGEMENTS

This research received funding from the Fulbright Program by the US Department of State, Planet Education and Research Program by Planet Team, Idaho State Department of Agriculture and U.S. Geological Survey under Grant/Cooperative Agreement No. G18AP00077.

We would like to thank the Geosciences Department at Idaho State University for their assistance and support. Special thanks to the farmers who contributed to this work by providing yield data and information about their fields and growing practices.

Zhang, Qin. 2016. *Precision Agriculture Technology for Crop Farming*. Prosser, Washington: Taylor & Francis Group.

REFERENCES

Aditya Shastry, H.A. Sanjay and E. Bhanusree. 2017. "Prediction of Crop Yield Using Regression Techniques." *International Journal of Soft Computing* 12(2): 96–102.

Al-Gaadi, Khalid A. et al. 2016. "Prediction of Potato Crop Yield Using Precision Agriculture Techniques." *PLoS ONE* 11(9): 1–16. <http://dx.doi.org/10.1371/journal.pone.0162219>.

Burke, Marshall, and David B. Lobell. 2017. "Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems." *Proceedings of the National Academy of Sciences* 114(9): 2189–94. <http://www.pnas.org/lookup/doi/10.1073/pnas.1616919114>.

Gitelson, Anatoly A., Yoram J. Kaufman, and Mark N. Merzlyak. 1996. "Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS." *Remote Sensing*.

Hatoum, Talal. 2005. "مديرية الدراسات الزراعية في لبنان 2005 والتسويق:" http://www.agriculture.gov.lb/Arabic/DataAndAgriStatistics/AgriInLebanon/Documents/Enq_2005.pdf (January 11, 2017).

Huete, A. R. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25(3): 295–309.

Lewin, Paul et al. 2011. "The Role of Agricultural Processing in Idaho's Economy: Status and Potential."

Qi, J et al. 1994. "A Modified Soil Adjusted Vegetation Index." 126: 119–26.

Rembold, Felix, Clement Atzberger, Igor Savin, and Oscar Rojas. 2013. "Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection." *Remote Sensing* 5(4): 1704–33.

Rouse, J. W., R. H. Hass, J.A. Schell, and D.W. Deering. 1973. "Monitoring Vegetation Systems in the Great Plains with ERTS." *Third Earth Resources Technology Satellite (ERTS) symposium* 1: 309–17.

S.K., McFeeters. 1996. "The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features." *International Journal of Remote Sensing* 17(7): 1425–32.

Sellam, V, and E Poovammal. 2016. "Prediction of Crop Yield Using Regression Analysis." *Indian Journal of Science and Technology* 9(October): 1–5.

Song, Renzhong et al. 2016. "Evaluation of Landsat 8 Time Series Image Stacks for Predicting Yield and Yield Components of Winter Wheat." *IEEE*: 6300–6303.

United Nations Food and Agriculture Organization. 2017. *The Future of Food and Agriculture – Trends and Challenges*. Rome.