Remote Sensing of Environment 141 (2014) 116–128

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States

David M. Johnson *

United States Department of Agriculture, National Agricultural Statistics Service, 3251 Old Lee Highway, suite 305, Fairfax, VA 22030, United States

ARTICLE INFO

Article history:
Received 15 July 2013
Received in revised form 24 October 2013
Accepted 26 October 2013
Available online xxxx

Keywords:
Crop yield forecasting
Corn
Soybeans
Precipitation
NDVI
Land surface temperature
MODIS
Regression tree modeling

ABSTRACT

Four timely and broadly available remotely sensed datasets were assessed for inclusion into county-level corn and soybean yield forecasting efforts focused on the Corn Belt region of the central United States (US). Those datasets were the (1) Normalized Difference Vegetation Index (NDVI) as derived from the Terra satellite's Moderate Resolution Imaging Spectroradiometer (MODIS), (2) daytime and (3) nighttime land surface temperature (LST) as derived from Aqua satellite's MODIS, and (4) precipitation from the National Weather Service (NWS) Nexrad-based gridded data product. The originating MODIS data utilized were the globally produced 8-day, clear sky composited science products (MOD09Q1 and MYD11A2), while the US-wide NWS data were manipulated to mesh with the MODIS imagery both spatially and temporally by regridding and summing the otherwise daily measurements. The crop growing seasons of 2006–2011 were analyzed with each year bounded by 32 8-day periods from mid-February through late October. Land cover classifications known as the Cropland Data Layer as produced annually by the National Agricultural Statistics Service (NASS) were used to isolate the input dataset pixels as to corn and soybeans for each of the corresponding years. The relevant pixels were then averaged by crop and time period to produce a county-level estimate of NDVI, the LSTs, and precipitation. They in turn were related to official annual NASS county level yield statistics. For the Corn Belt region as a whole, both corn and soybean yields were found to be positively correlated with NDVI in the middle of the summer and negatively correlated to daytime LST at that same time. Nighttime LST and precipitation showed no correlations to yield, regardless of the time prior or during the growing season. There was also slight suggestion of low NDVI and high daytime LST in the spring being positively related to final yields, again for both crops. Taking only NDVI and daytime LST as inputs from the 2006–2011 dataset, regression tree-based models were built and county-level, within-sample coefficients of determination (R^2) of 0.93 were found for both crops. Limiting the models by systematically removing late season data showed the model performance to remain strong even at mid-season and still viable even earlier. Finally, the derived models were used to predict out-of-sample for the 2012 season, which ended up having an anomalous drought. Yet, the county-level results compared reasonably well against official statistics with $R^2 = 0.77$ for corn and 0.71 for soybeans. The root-mean-square errors were 1.26 and 0.42 metric tons per hectare, respectively.

1. Introduction

1.1. Crop yield statistics

Accurate and timely estimation of local and regional crop yield statistics is important for a variety of reasons. On the macroeconomics level they allow societies to understand the food and fiber supply which in turn helps the demand side plan for and better utilize the finite crop resources. In the most developed countries this is manifested through futures contract markets which are most efficient and fair for price discovery when transparent and current statistics are available. Local, direct to consumer markets work similarly in that statistics help both parties understand the value of the crop. From a management standpoint, yield information gives a farmer a baseline of what is typically expected to be produced and thus can be used to best establish risk, insurance premiums or the value of input costs. Established yield information also highlights the impact to crops from natural events such as severe weather or changing climatic conditions. Likewise, regional yield statistics help quantify how strategies such as planting methodologies, irrigation, fertilizer and pesticide use are playing out in aggregation and can identify regions that are chronically underperforming, or have a "yield gap."

The United States Department of Agriculture (USDA) spends considerable effort in determining United States (US) crop yields in service to the agricultural community. The statistical arm of the USDA, the National Agricultural Statistics Service (NASS), conducts two large panel surveys (USDA, 2012) that are annually ongoing throughout the growing season (USDA, 2010) to establish state- and national-level yield estimates. The first is known as the Agricultural Yield Survey

* Tel.: +1 703 877 8000x169.
E-mail address: dave.johnson@nass.usda.gov.

0034-4257/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2013.10.027

which is based on a maintained “list frame” of farmers and the results are directly reliant on the information they provide. Each year thousands of those farmers are randomly selected, contacted monthly by phone during the growing season and asked to report expected yields for their crops grown. Information from all those sampled is then combined and summarized to derive a set of regional yield “indications.” Run in parallel is the Objective Yield Survey which derives an independent set of indications through biophysical crop measurements. For it, hundreds of small plots are randomly sampled from fields throughout the major growing areas and visited by an enumerator a few times during the crop season. Attributes collected include plant counts per unit area, grain size, grain weight, etc. The information from all of the plot-level data is ultimately aggregated into a model to derive this second set of yield indications. The Objective Yield Survey is more limited in scope over the Agricultural Yield Survey in that it only focuses on the dominant commodity crops like corn, soybeans, wheat, potatoes and cotton. Ultimately, the results from both surveys, along with any relevant ancillary information, are analyzed by the NASS Agricultural Statistics Board (ASB) to establish the monthly published yield forecasts.

After the season is complete late in the fall, an additional widely cast survey is undertaken which documents agricultural production statistics down to the county-level. For it questionnaires are sent to a much larger sample of producers asking for responses on many agricultural facets of their operation including estimates of their crop yields. Finally, these county-level statistics are assessed and published to reconcile with the previously established ASB national- and state-level yields.

Any further independent, error assessable and cost effective measures of crop yield indications that can be provided to the ASB are welcome. Real-time measuring of crop yields from remote sensing technologies has been promoted as a feasible methodology but has been fairly limited in implementation (Allen, Hanuschak, & Craig, 2002; Baruth, Royer, Klisch, & Genovese, 2008; Reynolds et al., 2000; Rojas, 2007). Reasons for lack of uptake are likely many but probably lead by a perception that results are not being seen as accurate, timely, or objective enough. Furthermore, remote sensing estimation of crop yields has potentially been hindered due to the unknown availability, cost and capacity of future imagery data combined with the highly specialized nature of the work for which it may be hard to find skilled and experienced labor.

In terms of US crop statistics themselves, corn and soybeans are the two largest commodities grown by land area and the planted acreage has steadily expanded in reach by about 25% over the last couple of decades (USDA/NASS Quick Stats). Yield trends for these crops have been increasing at a similar rate but see more relative variability year to year. Corn and soybeans from the US are high value and significant commodities on global export markets and of late they have been volatile in pricing. This suggests, at least in part, that the true amount produced has not been fully understood at all times.

1.2. Remote sensing of crop yields

Monitoring crops via satellite remote sensing is not a new idea or one in which there is a lack of research. Funk and Budde (2009) showed a summary of the work in a variety of sensor, location, and crop type contexts. Gallego, Carfagna, and Baruth (2010) also presented a history targeted specifically to crop production estimation. Even with all this aggregated work, assimilating the results to summarize to a best practice is confounding because the research has targeted different ecoregions of the globe, does not use the exact same type of input datasets or has varying methodologies. Furthermore, the specific crop type of focus has varied across the studies making the outcomes further difficult to compare. However, in general there has been more emphasis on corn, soybeans and wheat. Reasons why these three crops in particular have been the most studied are unknown but likely because they are found wide spread and in large quantities geographically. An alternative reason could be that there has been found better remote sensing yield estimation success with them versus other potential crops (and failures tend not to get published).

Regardless of crop being investigated, a common and central theme of this type of remote sensing research involves the reduction of the sensor's multispectral channels into a single metric known as the Normalized Difference Vegetation Index (NDVI), analyzing its response throughout the crop growing season and then relating it again to in situ collected crop information. NDVI is calculated from the red and near-infrared (NIR) spectral channels as

$$\mathrm{NDVI} = (\mathrm{NIR}-\mathrm{red})/(\mathrm{NIR}+\mathrm{red}).$$

NDVI exploits the large difference seen between the red and NIR bands for heavily vegetated land cover types and has been shown to be strongly correlated with plant productivity in both in situ (Hatfield, 1983; Shanahan et al., 2001; Viña et al., 2004) and remote sensing applications (Basnyat, McConkey, Lafond, Moulin, & Pelcat, 2004; Tucker, 1979). NDVI is often preferred over the independent use of the red and NIR channel in that it simplifies data analysis into a single metric while at the same time it is a normalization which helps to reduce data errors due to poor viewing geometry or hazy atmospherics. The normalization also allows for easier comparison across different sensors.

Many sensors with the ability to provide NDVI have been utilized throughout the past few decades. Long-term, ubiquitous, and freely available US satellite assets are reviewed here. The program with the longest lineage is that of the Advanced Very High Resolution Radiometer (AVHRR) sensor. Variants of it have been aboard over a dozen operational polar orbiter meteorological satellites that were first launched in the late 1970s. The last was placed into orbit in 2009. AVHRR is decently suited from monitoring vegetation dynamics given its daily revisit rate and reasonable spatial resolve of a little over 1 km which is fine enough for monitoring relatively homogenous crop areas. Assessment of AVHRR NDVI phenology and the general relation local crop yields has been performed (Ferencz et al., 2004; Maselli & Rembold, 2001) in addition to analysis targeted toward the specific commodities of corn, soybeans or wheat (Benedetti & Rossini, 1993; Hays & Decker, 1996; Mkhabela, Mkhabela, & Mashinini, 2005; Salazar, Kogan, & Roytman, 2007; Wall, Larocque, & Léger, 2008). Modeling results have been shown reasonable for all but there are limitations on the yield estimation precision. Errors may be a function of the native coarse pixel size of AVHRR, which is larger than most fields even in heavily mechanized agricultural regions, or due to the sensor itself which may not be able to provide adequate spectral information with low noise.

Technology has progressed and a newer and more sophisticated sensor called the Moderate Resolution Imaging Spectroradiometer (MODIS) improves on AVHRR (Fensholt & Sandholt, 2005; Huete et al., 2002) particularly in terms of spectral response, spatial resolution, and having more emphasis placed on land related observations (Justice et al., 2002). MODIS is aboard two earth science research oriented satellites, Terra and Aqua, which were launched in 1999 and 2002, respectively. MODIS carries a total of 36 spectral bands with most having a nadir ground resolution of about 1 km, which is similar to AVHRR. However, two of the key bands for land observations, the red and NIR, have a much finer resolution of about 250 m. MODIS derived crop and productivity yield work has again often relied on NDVI phenology with a focus on wheat (Becker-Reshef, Vermote, Lindeman, & Justice, 2010; Mkhabela, Bullock, Raj, Wang, & Yang, 2011; Reeves, Zhao, & Running, 2005) and corn or soybeans (Bolton & Friedl, 2013; Doraiswamy et al., 2004; Doraiswamy et al., 2005; Funk & Budde, 2009; Guindin-Garcia, Gitelson, Arkebauer, Shanahan, & Weiss, 2012; Sakamoto, Gitelson, & Arkebauer, 2013). Yield research which use MODIS data have proven better results than those that use AVHRR data. There are no identical MODIS follow-on mission planned but a similar meteorology focused polar orbiting sensor called the Visible Infrared Imaging Radiometer Suite (VIIRS) program has begun. A prototype VIIRS instrument was launched in 2011 aboard a satellite platform