Field Crops Research 219 (2018) 106–112

Contents lists available at ScienceDirect

Field Crops Research

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

Predicting spatial patterns of within-field crop yield variability

Bernardo Maestrini, Bruno Basso*

Department of Earth and Environmental Sciences, Michigan State University, East Lansing, 48823, MI, USA

ARTICLE INFO

Keywords:
Yield map
NDVI
Aerial images
Plant surface temperature
Management zones
Precision agriculture

ABSTRACT

Over the last two decades, there has been significant advancements in the application of geospatial technologies in agriculture. Improved resolutions (spectral, spatial and temporal) of remotely sensed images, coupled with more precise on-the-ground systems (yield monitors, geophysical sensors) have allowed farmers to become more sensitive about the spatial and temporal variations of crop yields occurring in their fields. Previous research has extensively looked at spatial variability of crop yields at field scale, but studies designed to predict within-field spatial patterns of yield over a large number of fields as yet been reported. In this paper, we analyzed 571 fields with multiple years of yield maps at high spatial resolution to understand and predict within-field spatial patterns across eight states in the Midwest US and over corn, soybean, wheat and cotton fields. We examined the correlation between yield and 4 covariates, three derived from remote sensing imagery (red band spectral reflectance, NDVI and plant surface temperature) and the fourth from yield maps from previous years. The results showed that for spatial patterns that are stable over time the best predictor of the spatial variability is the historical yield map (previous years' yield maps), while for zones within the field that are more sensitive to weather and thus fluctuate from one year to the next the best predictor of the spatial patterns are the within-season images. The results of this research help quantify the role of historical yield maps and within-season remote sensing images to predict spatial patterns. The knowledge of spatial patterns within a field is critical not only to farmers for potential variable rate applications, but also to select homogenous zones within the field to run crop models with site-specific input to better understand and predict the impact of weather, soil and landscape characteristics on spatial and temporal patterns of crop yields to enhance resource use efficiency at field level.

1. Introduction

In order to apply variable rate input within a field (Schepers et al., 2004), it is essential to understand the drivers of the spatial distribution of yield at field scale. A number of studies have investigated the determinants of spatial variability of yield at the level of a single field (Basso et al., 2011; Koshla et al., 2010) however few studies have attempted to compare predictors of yield spatial patterns over a large number of fields.

Here we investigate factors that predict within-field yield spatial variability by dividing fields into stable and unstable portions, based on the yield temporal variability that each point of the field exhibits over three or more growing seasons (Basso et al., 2007; Blackmore, 2000). In the stable portions of a field, the main determinants of spatial distribution of yield are related to soil properties and landscape position. However, in areas where yield is unstable from year to year, spatial distribution of yield is the result of the interaction between the soil characteristics, position in the landscape and weather (i.e. the performance of an unstable area of the field will have stronger variation compared to the rest of the field depending on the year's weather).

In this study, we examined the correlation between yield and 4 covariates, three were derived from remote sensing imagery (red band spectral reflectance, NDVI and surface temperature) and the fourth entailed the use of yield maps from previous years. Each of these covariates is well-correlated to yield for various reasons. The red band reflects the amount of light that is not absorbed by the plant in the red portion of the electromagnetic spectrum and is therefore negatively correlated with the photosynthesis. In a similar fashion, NDVI (Tucker, 1979) represents the normalized difference between the near infrared (emitted by leaves) and red (absorbed by leaves) and is positively correlated to plant photosynthetic activity. Surface temperature is a proxy for plant transpiration and thus, soil water availability and plant photosynthetic rate.

We investigated the above-mentioned four covariates using a dataset that encompasses fields from eight states of the Midwest of the United States cultivated with maize (*Zea mays* L.), wheat (*Triticum*

* Corresponding author.
E-mail address: basso@msu.edu (B. Basso).

https://doi.org/10.1016/j.fcr.2018.01.028
Received 12 September 2017; Received in revised form 25 January 2018; Accepted 25 January 2018
0378-4290/

spp.L.), soybean (*Glycine max,* L.) and cotton (*Gossypium* spp. L).

We investigated the following research questions: 1) In what part of the growing season is the correlation between crop growth and plant spectral reflectance the highest in our 571 fields? We hypothesize that the best correlation for maize occurs in July, as reported by Johnson (2014) at the county level, because the processes governing the correlation (photosynthesis level) are the same at the two spatial scales (field and county); 2) Is the correlation between within-season images and yield stronger than the correlation between past yield maps and yield? We hypothesized that historical yield maps exhibit a stronger correlation because they are a proxy for the interaction between soil conditions and past weather along with crop phenology, whereas the individual within-season images reflect the effects of the weather on growth only at the time of the image (single crop stage).

To test our hypothesis under the most rigorous conditions, we compared the variable importance of the historical yield against the post-facto NDVI images (i.e. the NDVI image that showed the best correlation with the yield at harvest, although clearly in reality it is not possible to know beforehand which will be the within-season image that exhibit the best correlation). We further hypothesized that historical yield is the best predictor only in the stable zones whereas unstable zones have by definition poor correlation with the yield of previous years and therefore they can be better predicted using within season remotely sensed images.

2. Materials and methods

2.1. Yield data

We collected yield maps from 571 fields from 110 farmers, for a total number of 2009 fields-year maps. In 27% of the fields we had more than 4 years of yield maps. The fields were in 8 different states of the Midwest of the United States, as shown in Map 1. The distribution of the yields collected for each field and the number of yield maps collected for each state in shown in the Table SI 1.

For each harvest point dataset (i.e. the points recorded by the harvester monitoring system relative to one year), the median was used to define the lower (0.1 × median) and higher (3 × median) boundaries. All points below or above the boundaries were handled as outliers and deleted. Points with the same longitude and latitude were dissolved to avoid duplicates. The average minimum distance between points was 1.3 m with an average standard deviation between fields of 0.6 m and within field of 0.4 m. We applied to each harvest point dataset a spherical kriging model with a cell size of 2 by 2 m, and a fixed radius with a distance of 20 m and a minimum of 12 points to rasterize the point dataset.

For every field, we calculated the border of polygon representing the field, and removed the yield maps that covered less than 75% of the field. We calculated field boundaries first by merging all the georeferenced points into a unique dataset and then by creating a polygon around the points based on an aggregation distance varying depending on the number of years of harvest available. The aggregation distance was set to 15 m (3 or more years of yield data), 20 m (2 years of yield data) or 30 m (1 year of yield data). For each field, we resampled the yield maps to have all the same spatial extent to allow a pixel wise analysis using bilinear interpolation. Additionally, we removed the years for which more than one yield map for the same field was available because in those years there were two different crops cultivated in different sections of the field. Fig. 1 shows the geographical distribution of the fields.

2.2. Red band from aerial visual images

Visual imagery for 121 fields was collected in the red, green and blue bands (RGB) by Airscout, a commercial airborne image company operating in the Midwest US. Of the total number of fields, images were collected of 93 fields for one year, 25 fields for two years, and one field for 3 years. We only considered the red band, as this is a proxy of the light absorbed by plants. Images were taken between the 4th of April and the 10th of October (Fig. SI 2a) in 2014 (3 fields), 2015 (39 fields) and 2016 (102 fields). The flights hours were uniformly distributed between 9 a.m. and 6 p.m. (Fig. SI 2b). The resolution of the red band images was on average 0.30 m (sd 0.05), the resolution varied depending on the flying height of the airplane (Fig. SI 2c). In the few cases where multiple pictures of a field were taken at interval lower than one hour, raster images were averaged, under the assumption that either multiple pictures were taken by mistake or that each picture represents only a portion of the field. Raster images were resampled (using a bilinear interpolation method) and projected to match the resolution and projection of the yield maps.

2.3. Airborne plant temperature and visual images

Plant surface temperature and visual (RGB) images were taken simultaneously from 130 fields, in 9 fields the temperature image was available whereas the red band image was not available. The resolution of the temperature images was on average 2.2 m (sd 0.2, Fig. SI 3c). We resampled the temperature images to match the yield maps resolution, extent, and projection. This operation was necessary to perform a pixel-wise analysis of the correlation between the temperature image and yield image. The resampling method adopted was a bilinear interpolation method. As for the images of the red band reflectance, in the few cases where multiple pictures of the field were taken at a time distance lower than one hour the raster images were averaged, under the assumption that either multiple pictures were taken by mistake or that each picture represents only a portion of the field. We removed pixels indicating temperature values higher than 50 °C as they may indicate a measurement error.

2.4. Landsat 8 derived NDVI images

We downloaded all the images available for each field from April 1, 2014–November 1, 2016 using the python package *Landsat-util*. We screened all the images to mark as not available (NA) those pixels whose quality was affected by clouds, points that contained designated fills and dropped frames using the Landsat 8 Pre-Collection Quality Assessment. We then removed the images for which more than 25% of pixels in the field were marked as not available. We calculated the NDVI for each Landsat scene using the following formula:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

We report the distribution of the number of images available in the period July-August in the Fig. SI 4, the median of the distribution is 3. To measure the correlation between yield maps and NDVI images, we resampled the yield maps to match the resolution of the Landsat images using the bilinear interpolation method.

2.5. Historical yield

For the fields for which we had yield maps from at least four years, we calculated a *historical yield map* using the following algorithm: first, we normalized each yield map (i.e. centered and scaled to have mean = 0 and sd = 1); second, for each year we calculated the pixel-wise mean of the previous years' normalized yield maps. We calculated the historical yield map only for those years where at least three previous yield maps were available. For example, if there were yield maps for 2012, 2013, 2014 and 2015 for a field, the historical map was calculated only for 2015, whereas if there were yield maps only for 2013, 2014, and 2015, no historical map was calculated for that field. Conversely, if yield maps for 2012–2016 were available, we calculated the historical map for both 2015 and 2016. We used only maps from the