Spatiotemporal characteristics of white mold and impacts on yield in soybean fields in South Dakota

Confiance Mfuka, Emmanuel Byamukama and Xiaoyang Zhang

ABSTRACT
White mold of soybeans is one of the most important fungal diseases that affect soybean production in South Dakota. However, there is a lack of information on the spatial characteristics of the disease and relationship with soybean yield. This relationship can be explored with the Normalized Difference Vegetation Index (NDVI) derived from Landsat 8 and a fusion of Landsat 8 and the Moderate Resolution Imaging Spectroradiometer (MODIS) images. This study investigated the patterns of yield in two soybean fields infected with white mold between 2016 and 2017, and estimated yield loss caused by white mold. Results show evidence of clustering in the spatial distribution of yield (Moran’s I = 0.38; p < 0.05 in 2016 and Moran’s I = 0.45; p < 0.05 in 2017) that can be explained by the spatial distribution of white mold in the observed fields. Yield loss caused by white mold was estimated at 36% in 2016 and 56% in 2017 for the worse disease pixels, with the most accurate period for estimating this loss on 21 August and 8 September for 2016 field and 2017 field, respectively. This study shows the potential of free remotely sensed satellite data in estimating yield loss caused by white mold.

ARTICLE HISTORY
Received 12 July 2019
Accepted 27 December 2019

KEYWORDS
Soybean; white mold; Landsat; MODIS; fusion; Kriging; NDVI

1. Introduction
Soybean (Glycine Max) is among the most important crops in North America and in the world, grown for beans that are processed to provide oil and meal. However, soybean production is affected by many diseases (Yang and Feng 2001), among which Sclerotinia Stem Rot (SSR), commonly known as white mold, is ranked among the top 10 diseases suppressing soybean yield (Hoffman et al. 1998; Wrather et al. 2010; Allen et al. 2017). White mold is caused by a fungal pathogen, Sclerotinia sclerotiorum, which overwinters as sclerotia in the soil and soybean residue. The disease generally develops at canopy closure, under wet and cool conditions (Boland and Hall 1988) and is more likely to occur in areas of the field with high yield potential.

Several aspects of white mold and its effect on soybean yield have been studied. Among these aspects are yield loss resulting from the effects of white mold inoculated at different growth stages (Danielson, Nelson, and Helms 2004). Other aspects include tillage and crop sequence (Kurle et al. 2001) as well as cultivars and herbicide selection that reduce white mold (Nelson, Renner, and Hammerschmidt 2002). Despite the importance of white mold and its influence on soybean yield, little is known on the spatial distribution of the disease at the field level and factors that might influence that distribution. Studies that employed remotely sensed images to investigate soybean stress have either focused on diseases and stresses other than white mold (Bajwa, Rupe, and Mason 2017; Behmann, Steinrucken, and Plumer 2014; Thompson and Wehmanen 1980), or employed expensive methods for collecting spatial data, such as hand-held spectrometers (Vigier, Pattey, and Strachan 2004), which are not freely accessible. Furthermore, recent studies that explored the spatial distribution of disease in soybean focused on other diseases and pests such as the analysis of spatial pattern of soybean cyst nematode (Avendaño et al. 2003), Japanese beetle (Sara, McCallen, and Switzer 2013), or Megacopta cribraria (Seiter, Rey-Jones, and Greene 2013). Hartman, Kull, and Huang (1998) conducted a field survey in Illinois to determine the spatial pattern of soybean yield. Their results suggested an aggregation of white mold in the surveyed fields, but the study involved traditional survey methods that remain costly in terms of time and labor. Rousseau, Rioux, and Dostaler (2006) examined the spatial distribution of white mold apothecia, but they focused on the soil physicochemical properties that are generally collected in plots and require intensive data collection. Our recent study quantified white mold with Landsat images but the study did not examine the temporal trend and the impact on yield (Mfuka, Zhang, and Byamukama 2019).

Current knowledge of spatial distribution of white mold in soybean fields is limited, and there is a need to investigate cost-effective alternative methods, including...
the use of free remotely sensed data to model the distribution of white mold and explain its impact on yield. This study has examined the pattern of yield in white mold infected soybeans fields, and employed a linear regression model to examine the relationship between soybean yield and the Normalized Difference Vegetation Index – NDVI (Rouse et al. 1973), which consists of a ratio between the difference and the sum of the reflectance for the Near-Infrared (NIR) and the red bands. Subsequently, the study estimated yield loss caused by white mold and used time-series NDVI surface maps generated by Ordinary Kriging (OK) to explore the spatio-temporal patterns of white mold and its influence on yield distribution.

2. Materials and methods

2.1. Study area

The study area consisted of two soybean fields that were monitored and confirmed to be affected with white mold in two different years. The first field was monitored in 2016 and was located in Moody county SD, while the second field, was monitored in 2017, and was located in Marshall county SD (Figure 1). No other soybean diseases were observed in these fields. The studied fields are located in the eastern and north-eastern regions of South Dakota, which are characterized by a heterogeneous land cover dominated by cropland and pasture/grassland, as well as wetland, forestland and open water (USDA 2017). Major crops in these regions include corn, soybean, alfalfa and spring wheat.

2.2. Data collection and preparation

2.2.1. Yield data collection

Field data consisted of soybean yield (expressed in Bushels per Acre Bu/ac: 1bu/ac = 67.251 kg/ha) collected by a combine at 1-s intervals, resulting to a total of 29,878 sample points in 2016 (Moody field) and 40,216 sample points in 2017 (Marshall field). The combine contained a Global Positioning System (GPS) that allowed to produce georeferenced data using a Universal Transverse Mercator (UTM) Zone 14 North projection with a datum (WGS 84) similar to the Landsat images. The harvest in the two fields occurred, respectively, on 10 October 2016 in Moody, and on 18 October 2017 in Marshall. These data collected at 1-s intervals resulted into a large point dataset comprised of several points in a single Landsat pixel; therefore, the yield data were resampled to match Landsat pixels because using raw data would produce an unnecessary redundancy that would not improve the spatial relationships. The selected statistical summary for resampling raw yield is the average, which represents the equivalent of average yield collected over each pixel. To obtain a regular and consistent dataset, raw yield data were spatially preprocessed by superimposing a regular 30 m × 30 m grid on the fields, matching original Landsat pixels and on which the average yield was computed for each cell. The grid was obtained using the fishnet tool from ArcMap (ESRI

Figure 1. Location of the study area. The scouted fields were located in Marshall and Moody counties, in South Dakota. The background image is a Landsat 8 false color bands 6-5-4 combination from 14 July 2017 (Marshall) and 20 July 2016 (Moody).
meaning they had the same projection with the Landsat images.

2.2.2. White mold imagery acquisition
Landsat Analysis-Ready Dataset (ARD) consisting of Surface Reflectance (SR) and Quality Assessment (QA) data were downloaded. These images are consistently processed using per pixel solar zenith angle corrections, gridded to a common cartographic (Albers Equal Area, D_WGS-84) projection (https://lta.cr.usgs.gov/USLSArdTile). We downloaded from earth explorer (https://earthexplorer.usgs.gov/), six cloud-free Landsat ARD images for the year 2016 covering the Moody field, and spanning the dates 17 May, 20 July, 5 and 21 August, and 14 and 30 September. Similarly, we downloaded 4 images for the year 2017 covering the Marshall field, and spanning the dates 11 May, 14 July, 31 August, and 08 September. In the analyses, however, only three 2016 images were used covering the Days of the Year (DOY) 217 (5 August), 233 (21 August) and 255 (12 September). Early images such as on May 17 were excluded because at this period, soybean has not developed yet and therefore white mold is not expected to be observed. In 2017 however, the ARD data were used as bracket images in the fusion process to obtain images for DOYs 217 and 233. Since there was no cloud-free image after 08 September 2017 to allow the data fusion, this 2017 DOY 251 (08 September) image was compared to the 2016 DOY 255 (14 September) image.

Daily Moderate Resolution Imaging Spectroradiometer – MODIS images (MDC43A4 Version 6 products) were downloaded for the year 2017, covering the DOYs 217 and 233. These products consist of Terra and Aqua Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance (NBAR) at 500 m spatial resolution in a sinusoidal projection containing 7 bands Albedo and 7 bands Nadir Reflectance. Detailed descriptions about MDC43A4 Version 6 products are provided by Schaaf and Wang (2015). Using the MODIS Reprojection Tool (MRT), selected bands (SWIR, NIR and RED) from the MDC43A4 images were extracted while conserving their native sinusoidal projection. The extracted bands were stacked in the Environment for Visualizing Image (ENVI) software version 5.0 (Exelis 2012), and the resulting composite image was projected (with the Map Projection Tool in ENVI) using the projection parameters from the Landsat images (Albers Equal Area, D_WGS-84); the pixels were resampled from 500 to 30 m using the nearest neighbor option in ENVI, to match the corresponding Landsat image and ease the fusion of the two datasets.

2.2.3. Data fusion
To obtain comparable datasets between the 2 years, we performed a data fusion for 2017 images; this process ensured that we obtain images for similar dates as those collected in the year 2016. High spatial resolution Landsat ARD (30 m) do not have the required temporal resolution (16 days) for crop monitoring; meanwhile, high temporal resolution daily MODIS do not have the necessary spatial resolution for local scale crop monitoring, thus the need to combine both datasets into consistent high spatial and temporal synthetic images for crop monitoring. Several blending techniques have been developed for time series analyses (Gao et al. 2006; Huang and Zhang 2014; Roy et al. 2008; Weng, Fu, and Gao 2014; Zurita-Milla, Clevers, and Schaepman 2008). In this study, we used the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) (Zhu et al. 2010) that requires two images pairs (fine and coarse) for the origin and the final dates, to estimate the reflectance at the prediction date between the two dates. ESTARFM was selected because of its advantages: it works better in heterogeneous regions such as the croplands in our study area, it improves the prediction with the use of several bands (Table 1 shows the bands that were combined in the algorithm) in selecting similar pixels, and it uses spectral similarities correlation coefficients between Landsat and MODIS in the weight calculation of similar pixels. Furthermore, this algorithm has outperformed many others (Emelyanova et al. 2013; Li et al. 2017; Wu et al. 2016). Details about the ESTARFM algorithm are provided by Zhu et al. (2010).

2.2.4. NDVI computation
The original images for 2016 and the resulting fused synthetic images for 2017 were used to compute the NDVI (Rouse et al. 1973) using the Red and the NIR bands (Equation 1). NDVI has been successfully used to monitor crops in several studies (Fan et al. 2014; Gao et al. 2017; Onojeghuo et al. 2018); it is a measure of healthy and green vegetation, combining the highest regions of chlorophyll absorption and reflectance. Its theoretical values range from −1 to +1, with the common vegetation values ranging from 0.2 to 0.8. A regular 30 × 30 m grid was superimposed on the original and the synthetic satellite images, with each cell representing an individual Landsat pixel. For each grid cell, NDVI values were extracted from the original (2016) and the synthetic (2017) Landsat images for all the dates.

| Band Designation | OLI (30 m) | MODIS (500 m) |
|------------------|-----------|---------------|
| Wavelengths (µm) | Band number | Wavelengths (µm) | Band number |
| RED | 0.650–0.680 | 4 | 0.630–0.670 | 1 |
| NIR | 0.845–0.885 | 5 | 0.841–0.876 | 2 |
| SWIR | 1.560–1.660 | 6 | 1.628–1.652 | 6 |

Table 1. Landsat Operational Land Imager (OLI) and the Moderate Resolution Imaging Spectroradiometer (MODIS) bands designation and wavelengths (µm) combined using the ESTARFM algorithm. (RED is the Red band, NIR is the Near Infrared band and SWIR is the Shortwave Infrared band).
\[ \text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \]  

(1)

where NIR is the Near Infrared Band and Red is the Red Band.

2.3. Analyses

2.3.1. Testing yield distribution for spatial autocorrelation

The spatial distribution of yield was tested for randomness, using the Moran’s I autocorrelation test (Moran 1950), which is a weighted correlation coefficient used to identify departures from spatial randomness. The purpose of this analysis is to test if yield distribution can be dictated by other factors, such as the white mold, or if there is no spatial pattern. Moran’s I is used to determine whether neighboring areas are more similar than would be expected under the null hypothesis; it considers yield locations and their respective attributes and integrates the attribute resemblance and location adjacency in the computed coefficient (Soltani and Askari 2017). Moran’s I coefficient is computed as shown in Equation (2) (Anselin 1995):

\[ I = \frac{n}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \sum_{j=1}^{n} w_{ij} (y_i - \bar{y})(y_j - \bar{y}) \]  

(2)

where \( w_{ij} \) is elements of a spatial binary contiguity matrix; \( y_i \) and \( y_j \) is variable values at specific locations \( i \) and \( j \) is average of the variable and \( n \) is total number of locations.

The coefficient values range from \(-1\) to \(1\), with a negative value meaning dispersion, positive value meaning clustering and 0 value meaning randomness (Soltani and Askari 2017). The results are interpreted within the context of the null hypothesis, which states that yield attributes are randomly distributed across the space (Blazquez et al. 2018).

The grid cells were first transformed to a neighborhood object based on the adjacency method, where all the eight pixels around the central pixel are considered neighbors. The spatial neighborhood object was then converted to a weighted matrix, in which elements (\( w \)) represent the connectivity relationship between location \( i \) and neighboring location \( j \) across the field, and on which the yield values distribution was tested for randomness. The significance of the resulting Moran’s I coefficient was tested using a Monte-Carlo simulation using the null hypothesis that there is no spatial autocorrelation in the spatial distribution of yield (or the resulting Moran’s I is due by chance). The same method was employed by Milne et al. (2019), where the resulting Moran’s I index value was compared against 99 independent permutations of the data generated by Monte-Carlo simulations of Moran’s I, and significance was established at \( p \leq 0.05 \). In our study, we increased the number of independent permutations to 999 to strengthen the Monte-Carlo simulation.

2.3.2. Ordinary Kriging (OK) NDVI and yield surface maps

The extracted NDVI and yield values for each grid cell were interpolated to create surface maps that could be compared and provide spatial trends. The interpolation was performed using the Ordinary Kriging (OK) method, which is an exact interpolator, meaning that the values of the input points do not change in the predicted model. OK is a geostatistical technique used to estimate values at unsampled areas, based on a limited number of observations; it has been used in many studies including the spatial distribution of disease in crops (Moral García 2006) or in modeling the spatial distribution of soil nutrients (Elbasiouny et al. 2014). Furthermore, OK has outperformed other geospatial methods in terms of accuracy (Bhunia, Shit, and Maiti 2018). To perform the Kriging interpolation on yield, data were first checked using the exploratory analysis tool of the Geostatistical Wizard module in ArcMap (ESRI 2016) to determine if they fit the basic conditions for a Kriging interpolation (normal distribution, stationarity and no trend). The best model was defined by the optimized semi-variogram, and a cross validation was performed to check the quality of the model.

2.3.3. Time-series NDVI and yield relationship

A linear regression model was computed in R-Studio (R-Studio 2016) to assess the relationship between yield and the NDVI. Correlation coefficients were computed for the time-series NDVI for all the available image dates of the growing season, to assess the change of the relationship between the two datasets. The purpose of this analysis is first, to compare how the relationship between NDVI and yield changed during the season, and second to examine the behavior of this relationship between the 2 years. This allowed to estimate the best period when yield can be predicted using NDVI, as explained by the strength of the relationship.

2.3.4. Impact of white mold on yield

The impact of white mold on yield was examined by assessing the yield loss caused by the disease. The yield in white mold pixels was compared to the maximum (expected) yield (yield from pixels with no white mold). The difference between the expected yield and the yield in the white mold pixels gave an estimation of the yield loss. The yield loss was compared and expressed as a percentage of the maximum expected yield as shown in Equation (3). The yield map used quantiles distribution in the symbology, with the lowest category representing the worst white mold cases, and the highest category representing the maximum yield in the nonwhite mold soybeans. For each range, the average yield was computed...
for comparison with the maximum yield. Furthermore, the relationship between white mold and yield was computed at different dates to assess the accuracy of each image in estimating yield loss at different crop stages.

\[ YL \% = \left( \frac{Y_{\text{max}} - Y_{\text{avg}}}{Y_{\text{max}}} \right) \times 100 \]  

(3)

where \( YL \% \) is the yield loss; \( Y_{\text{max}} \) is the maximum yield and \( Y_{\text{avg}} \) is the average yield for each range (class).

3. Results and discussion

3.1. Data fusion

A visual comparison between an original and a predicted image for DOY 163 is shown in Figure 2: the two images matched consistently, especially considering the heterogeneous character of the region. The computed NDVI from the predicted and the original images for the same date (DOY 163) were compared using a reflectance scatter plot. This method used by Huang and Zhang (2014) and Zhu et al. (2010) allows a comparison of the distribution of reflectance values along the 1:1 line to assess the accuracy of the blended images. Figure 3 shows the relationship between the original and the predicted NDVI values.

Zhu et al. (2010) and Huang and Zhang (2014) used the Average Absolute Difference (AAD) and the Average Difference (AD) to assess the accuracy of the blending techniques. These metrics, however, are suitable to compare reflectance difference between bracketing input images and an original existing image in-between, and the two bracketing images and the predicted image. This requirement fits best for comparison with the STARFM method, which requires two image pairs as inputs. The comparison between the original and the predicted image reflectance for ESTARFM by Huang and Zhang (2014) provided a correlation coefficient of 0.88–0.941 and 0.843–0.862, respectively, for the phenological and the land cover change datasets. These estimates are close to ours; however, our comparison was based on the NDVI instead of the individual bands. The computed correlation coefficient of 0.92 for the original and predicted NDVI shows how accurate the synthetic image is.

3.2. Yield spatial distribution

The results of the Moran’s I test show a significant positive spatial autocorrelation (Moran’s I = 0.38, \( p \)-value < 2.2e-16 for the Moody field and Moran’s I = 0.45, \( p \)-value < 2.2e-16 for the Marshall field, respectively) as evidence of clustering in the spatial distribution of yield. In other words, yield distribution was not random, and could be controlled by an underlying process, as expected. The Monte-Carlo simulation of Moran’s I provides statistically significant results, indicating that the spatial autocorrelation result is not due by chance (Table 2). The \( p \)-value indicates the probability that the obtained statistical results could be due by chance. In other terms, the smaller the \( p \)-value, the more significant is the test. In this case, the \( p \)-value is very small, meaning that the probability that obtained results could be random are very small.

A Moran’s I scatter plot of yield for the year 2016 is shown in Figure 4(a): an obvious cluster can be observed in the lower-left quadrant, denoting low-low values of yield that may explain stress. Clusters of high yield values can be seen, but those values are mostly close to the mean. A few outliers can be observed in the upper-left and lower-right quadrants, but they do not form any clusters. Similarly, in 2017, the Moran’s I yield...
scatter plot (Figure 4(b)) shows a significant clustering in the upper-right quadrant, denoting clusters of high values of yield, while the lower-left quadrant shows a moderate clustering of low values (low yield). However, some mixed values can be observed both in the upper-left and lower-right quadrants, but they are not as abundant as the clustered values.

These results are consistent with the literature, as Hartman, Kull, and Huang (1998) found spatial aggregation in soybean white mold in Illinois. However, their study used the Lloyd’s patchiness index, instead of the Moran’s I spatial autocorrelation used in our study. A cluster analysis of yield distribution conducted by Jaynes et al. (2003) and Jaynes, Colvin, and Kaspar (2005) provided similar findings to ours. Their study however analyzed multiyear yield data and grouped them into significant clusters that were examined using a Moran’s I analysis. Their greater Moran’s I statistics (0.61 \( p < 0.001 \) and 0.74 \( p < 0.001 \), respectively) can be explained by the characteristics of the inputs. Consistent yield data over several (five and six) years allowed to clearly distinguish clusters of different yields. In our study, we used two different fields from two different years, but still could detect the spatial pattern. Other data included in these studies (precipitations, electrical conductivity, elevation, soil color …) allowed a better investigation of factors susceptible to influence the detected pattern, while in our study field data limited our investigation to the presence/absence of white mold, especially since our objective was to reduce costs related to intensive field data collection. A similar study by Roel and Plant (2004)
analyzed yield clusters and investigated factors able to influence the observed patterns. They found that even though the patterns are consistent, it is hard to precisely specify factors underlying these patterns, especially if the most important factors are not quantified or included in the analyses. Therefore, even though white mold presence can explain the yield distribution, there are other factors that were not quantified in our study, that may have an influence on this pattern such as soil fertility, field elevation and others.

3.3. The yield ordinary kriging model

The optimized model (Model: 22*Nugget+61.733*Stable (280,0.57617)) was obtained through the Geostatistical Wizard module of ArcMap (ESRI 2016), which explores the input data and provides several options for exploring the desired interpolation in order to select the best model. The optimized models are shown in Figure 5. In 2016, the nugget is 22 which is less than a pixel size (30 m), meaning that there is no difference in the values of points that have the same location (distance = 0 m). The major range (280 m) represents the distance beyond which there is little or no autocorrelation among variables; it represents the x value at which the curve starts flattening out. In our case, the Landsat pixel size is 30 m, meaning that about 10 pixels aligned in the same directions have a decreasing similarity and this similarity may not exist from the eleventh pixel. The sill (65) is the maximum y value at which there is little or no autocorrelation. A cross validation method was used to assess the quality of the model: each observation is removed and is estimated from the remaining points. The fitted model achieved a root mean standardized error of 0.92, which is very high, as the aim is to get a value that is closest to 1. A similar model was developed for the 2017 yield and is shown in Figure 5.

The optimized modeled variogram is data-dependent, meaning that it can vary considerably from one crop to another (Vieira and Gonzalez 2003), and from 1 year to another. A study by Jaynes and Colvin (1997) modeled the spatiotemporal variability of yield. They used six years yield data and developed yield variogram that could not exhibit consistent trends in their model parameters (sill and range). The main reasons advanced are that yield spatial distribution is mostly controlled by soil properties; also, weather parameters can influence these properties from 1 year to another. The choice of the best variogram depends largely on the data being analyzed: Jaynes and Colvin (1997) opted for the spherical variogram while Kumhálová et al. (2011) chose an exponential variogram. In our study, the choice of the spherical variogram was dictated by the comparison with other models to select the one that best fitted our dataset.

3.4. Seasonal changes of the relationship between yield and NDVI

The correlation coefficients between yield and NDVI are different between the 2 years: in 2016, they are, respectively, 0.28, 0.65 and 0.33 while in 2017, they are 0.68, 0.71 and 0.77, respectively, for DOY 217, 233 and 255. Several factors such as the rainfall regime, the field management practices or field physicochemical properties may explain that difference. In 2016, the peak of the correlation coefficient is reached on 21 August ($R = 0.65$), according to the available images, and the relationship decreases later in the season, while for the same date, the correlation coefficient stayed steady for a while in 2017. The decreasing trend of this relationship (NDVI and yield) after the peak dates means that yield loss estimation is more accurate on these dates than later in the season where low NDVI may be due to natural senescing. However, in 2017, a subsequent image on 08 September showed a better relationship with yield ($R = 0.77$). A reasonable explanation of the difference in the behavior of the trend between the two years after 21 August could be attributed to the respective field management. One aspect that might explain this is the planting date: in fact, the 8 days’ difference in the harvest date might reveal a similar trend in the planting dates as well;

![Figure 5](image-url) Semivariogram of the yield Ordinary Kriging (OK) models. The optimized spherical models were obtained through the Geostatistical Wizard module of ArcMap, by which explores the input data and provides several options for exploring the desired interpolation in order to select the best model.
while the difference in the NDVI signal in the early season might be small, it can be amplified at the maturity stage.

Another possible explanation may be the shift in the growing season; in fact, the two fields are located one degree apart (North-South); even if the growing season might start a few days earlier in Moody county (South), the question in this case remains the comparison in the magnitude of the NDVI difference observed between the two fields; in other words, does the 8 days’ interval correspond to the one-degree latitude shift in the growing season? Previous studies have shown that the growing season has extended over the last century (Linderholm 2006; Piao et al. 2007) and this extension has varied effects (Tucker et al. 2001) that are exacerbated especially in high latitudes (Kozlov and Berlina 2002).

An alternative explanation could be the difference between the soybean maturity groups. In this case, the field in 2016 might have matured early and therefore reached senescence sooner. The soybean maturity can also be influenced by weather. Different soybean maturity groups can behave differently in different temperatures and soil conditions, and this may result to the difference in the seed quality (Dardanelli et al. 2006). Temperature decrease can have an influence on nitrogen uptake by decreasing the root growth and this can influence the maturation (George, Singleton, and Ben 1988). Inversely, temperature increase can also influence the soybeans yield quality by increasing seed oil quality and protein content (Piper and Boote 1999). Differences in the average temperature between the 2 years during the period June-July-August could be observed in our study area for air temperature and soil temperature. Indeed, the average air and soil temperatures were 20°C and 21.7°C in Marshall, and 21.1°C and 23.9°C in Moody, respectively (Mesonet 2018). Furthermore, the rainfall difference can explain the yield difference between the two fields: Marshall had a total of 182 mm while Moody had a total of 306 mm (Mesonet 2018), which may explain why yield is generally higher in Marshall than in Moody. In any case, the strength of the correlation between NDVI and yield is at its best between late August and early September, and these characteristics could be observed and accurately estimated with the use of optical Landsat images.

3.5. Estimating yield loss caused by white mold

The scouted white mold pixels and their corresponding yield are shown in Figure 6 in 2016 and Figure 7 in 2017. In general, both the NDVI and yield map exhibit some similarities, especially while identifying the worst white mold pixels (lowest NDVI). The strength of the relationships may be affected by the quality of the inputs: detailed information is lost while resampling yield data at pixel levels; in fact, the selected resolution is likely to miss small patches of diseases even though the disease characteristics can still be captured. Nevertheless, the observed relationships remain strong enough for the moderate 30 m spatial resolution of Landsat.

The estimated yield loss for each NDVI class (from the worst white mold pixels to the healthiest soybeans) computed using Equation (3), is summarized in Tables 3 and 4, and ranges from 36% to 56% of the maximum expected yield. Hoffman et al. (1998) examined yield loss across different soybeans maturity groups and noticed the group effect in the yield loss difference, as a function of the disease intensity. This can be explained by the susceptibility to different cultivars to the disease.

The relationship between yield loss and the corresponding average NDVI for each class is shown in Figure 8. The relationship is linear and negative, and has a correlation coefficient of 89% and 99%, respectively, in 2016 and 2017. This means that the higher the NDVI, the lower the yield loss. The strength of this relationship is affected by the time of the year the estimation is being made. Considering the aspect of the remotely sensed images inputs, this explains why the estimation on 21 August 2016 provides a lower correlation coefficient than the estimation made on 8 September 2017. The developed models provide good estimations of yield loss with the use of remotely sensed images; however, these models depend on the quality of the inputs image at the peak of the relationship between yield and NDVI. One limitation of such models is that they cannot be extrapolated to big areas because they represent specific field situations and can change from one field to another. An ideal yield loss model based on satellite images should include phenological differences between field locations, differences due to different crop maturity groups, and also differences due to local environmental factors such as soil temperatures and other physicochemical characteristics.

4. Conclusion

The results show that the yield distribution in the explored fields is not random and might be dictated by other parameters. Crop diseases such as white mold can exhibit spots of low yields that can explain the clustering in the distribution of yield. This study shows that the spatial distribution of yield can be predicted using the NDVI computed from freely available images; the strongest relationship between yield and NDVI can be observed in late August or early September. Yet, NDVI computed close to the harvest date are not always the best yield predictor, because in this period, plants have reached senescence and do not have high photosynthetic activity. Further investigation is needed to explore the inter-annual difference in the relationship between yield and NDVI, and also possible location (such as latitudinal) effects.

The yield loss caused by white mold is estimated between 36% and 56% for the worst white mold pixels,
and the most accurate estimations of this loss is between late August and early September. The pattern analysis can be improved by the use of high spatial resolution images such as Sentinel-2 or Planet images. However, we suggest that detailed field data such as disease severity throughout the growing season, as well as soil physical properties, even though expensive, be included in future analyses. This information could

Table 3. Yield loss estimates (%) for soybean infected with white mold for each Normalized Difference Vegetation Index (NDVI) class in the Moody field (2016). Yield loss was computed from the difference between maximum yield (high yield pixels) and each NDVI class.

| Class range (Kg/ha) | Yield average (Kg/ha) | NDVI (08/21) | Yield loss (%) |
|---------------------|-----------------------|--------------|---------------|
| 2421-3632           | 3026                  | 0.86-0.915   | 36            |
| 3699-3833           | 3766                  | 0.916-0.924  | 21            |
| 3901-3968           | 3934                  | 0.925-0.928  | 17            |
| 4035-4170           | 4102                  | 0.929-0.931  | 13            |
| 4237-4304           | 4270                  | 0.932-0.933  | 10            |
| 4371-5111           | 4741                  | 0.934-0.939  | 0             |

Table 4. Yield loss estimates (%) for soybean infected with white mold for each Normalized Difference Vegetation Index (NDVI) class in the Marshall field (2017). Yield loss was computed from the difference between maximum yield (high yield pixels) and each NDVI class.

| Class range (Kg/ha) | Yield average (Kg/ha) | NDVI (09/08) | Yield loss (%) |
|---------------------|-----------------------|--------------|---------------|
| 942-2623            | 1802                  | 0.62-0.76    | 56            |
| 2623-3026           | 2831                  | 0.76-0.77    | 30            |
| 3026-3295           | 3154                  | 0.78-0.79    | 22            |
| 3295-3497           | 3389                  | 0.8001-0.8009| 16            |
| 3497-3699           | 3584                  | 0.81-0.819   | 12            |
| 3699-4439           | 4055                  | 0.82-0.85    | 0             |
be useful in explaining some patterns that cannot be inferred directly and only with satellites images.

Acknowledgments

Financial support for the study came from USDA-NIFA Hatch grant #SD00H662-18 and South Dakota State University Agriculture Experiment Station. We wish to thank Mr. Jianmin Wang for his technical supports on data processing, and the anonymous reviewers whose comments and edits have greatly improved this article.

Funding

This work was supported by the National Institute of Food and Agriculture [SD00H662-18].

Notes on contributors

Confiance Mfuka is a recent PhD graduate from the Agronomy, Horticulture, and Plant Science Department at South Dakota State University. He holds a Master’s of Science in Geography from the same institution. His current research interests include the use of geospatial sciences tools and techniques for natural resources monitoring and management with an emphasis on vegetation analysis (natural forests and crop monitoring).

Emmanuel Byamukama is an Assistant Professor and Plant Pathology Extension Specialist in the Agronomy, Horticulture, and Plant Science Department at South Dakota State University. He has conducted research in plant disease epidemiology, yield loss assessment, pathogen characterization, and plant disease detection.

Xiaoyang Zhang is a Professor and senior scientist in the Geospatial Sciences Center of Excellence with a tenure home in the Geography department at South Dakota State University. His research interests include the developments of remote sensing algorithms and global products for investigating biomass burning emissions, land surface phenology, land cover and land use change, and climate-terrestrial ecosystem interaction.

ORCID

Confiance Mfuka http://orcid.org/0000-0002-9948-3721
Emmanuel Byamukama http://orcid.org/0000-0003-4914-5457
Xiaoyang Zhang http://orcid.org/0000-0001-8456-0547

References

Allen, T. W., C. A. Bradley, A. J. Sisson, E. Byamukama, M. I. Chilvers, C. M. Coker, A. A. Collins, et al. 2017. "Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014." Plant Health Progress 18: 19–27. doi:10.1094/PHP-RS-16-0066.

Anselin, L. 1995. "Local Indicators of Spatial Association - Lisa." Geographical Analysis 27: 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x.

Avendaño, F., O. Schabenberger, F. J. Pierce, and H. Melakeberhan. 2003. "Geostatistical Analysis of Field Spatial Distribution Patterns of Soybean Cyst Nematode." Agronomy Journal 95: 936–948. doi:10.2134/ajronj2003.9360.

Bajwa, S., J. Rupe, and J. Mason. 2017. "Soybean Disease Monitoring with Leaf Reflectance." Remote Sensing 9: 1–14. doi:10.3390/rs9020127.

Behmann, J., J. Steinrucken, and L. Plumer. 2014. "Detection of Early Plant Stress Responses in Hyperspectral Images." Isprs Journal of Photogrammetry and Remote Sensing 93: 98–111. doi:10.1016/j.isprsjprs.2014.03.016.

Bhunia, G. S., P. K. Shit, and R. Maiti. 2018. "Comparison of GIS-based Interpolation Methods for Spatial Distribution of Soil Organic Carbon (SOC)." Journal of the Saudi Society of Agricultural Sciences 17: 114–126. doi:10.1016/j.jssas.2016.02.001.

Blazquez, C. A., B. Picarte, J. F. Calderon, and F. Losada. 2018. "Spatial Autocorrelation Analysis of Cargo Trucks on Highway Crashes in Chile." Accident Analysis & Prevention 120: 195–210. doi:10.1016/j.aap.2018.08.022.

Boland, G. J., and R. Hall. 1988. "Epidemiology of Sclerotinia Stem Rot of Soybean in Ontario." Phytopathology 78: 1241–1245. doi:10.1094/Phyto-78-1241.

Danielson, G. A., B. D. Nelson, and T. C. Helms. 2004. "Effect of Sclerotinia Stem Rot on Yield of Soybean Inoculated at Different Growth Stages." Plant Disease 88: 297–300. doi:10.1094/PDIS.2004.88.3.297.

Figure 8. Yield loss as a result of the NDVI change within fields. The gray area represents the confidence interval of the modeled relationship.
Dardanelli, J. L., M. Balzarini, M. J. Martínez, M. Cuniberti, S. Resnik, S. F. Ramunda, R. Herrero, and H. Baigorri. 2006. “Soybean Maturity Groups, Environments, and Their Interaction Define Mega-environments for Seed Composition in Argentina.” *Crop Science* 46: 1939–1947. doi:10.2135/cropsci2005.12.0480.

Elbasioni, H. M., A. Obwalya, A. Abu Alkheir, and A. A. Gad. 2014. “Spatial Variation of Soil Carbon and Nitrogen Pools by Using Ordinary Kriging Method in an Area of North Nile Delta, Egypt.” *Cataea* 113: 70–78. doi:10.1016/j.cataea.2013.09.008.

Emelyanova, I. V., T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. van Dijk. 2013. “Assessing the Accuracy of Blending Landsat–MODIS Surface Reflectances in Two Landscapes with Contrasting Spatial and Temporal Dynamics: A Framework for Algorithm Selection.” *Remote Sensing of Environment* 133: 193–209. doi:10.1016/j.rse.2013.02.007.

ESRI. 2016. *Arcgis Desktop*. 10.5. Environmental Systems Research Institute.

Exelis. 2012. *Environment for Visualizing Images*. Boulder, Colorado: Exelis Visual Information Solutions.

Fan, C., B. Zheng, S. W. Myint, and R. Aggarwal. 2014. “Characterizing Changes in Cropping Patterns Using Sequential Landsat Imagery: An Adaptive Threshold Approach and Application to Phoenix, Arizona.” *International Journal of Remote Sensing* 35: 7263–7278. doi:10.1080/01431161.2014.967891.

Gao, F., J. Masek, M. Schwallier, and F. Hall. 2006. “On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance.” *IEEE Transactions on Geoscience and Remote Sensing* 44: 2207–2218. doi:10.1109/TGRS.2006.827801.

Gao, F., M. C. Anderson, X. Zhang, Z. Yang, J. G. Alfieri, W. P. Kustas, R. Mueller, D. M. Johnson, and J. H. Prueger. 2017. “Toward Mapping Crop Progress at Field Scales through Fusion of Landsat and MODIS Imagery.” *Remote Sensing of Environment* 188: 9–25. doi:10.1016/j.rse.2016.11.004.

George, T., P. W. Singleton, and B. Ben. 1988. “Yield, Soil Nitrogen Uptake, and Nitrogen Fixation by Soybean from Four Maturity Groups Grown at Three Elevations.” *Agronomy Journal* 80: 563–567.

Hartman, G. L., L. Kull, and Y. H. Huang. 1998. “Occurrence of Sclerotinia sclerotiorum in Soybean Fields in East-central Illinois and Enumeration of Inocula in Soybean Seed Lots.” *Plant Disease* 82: 560–564. doi:10.1094/PDIS.1998.82.5.560.

Hoffman, D. D., G. L. Hartman, D. S. Mueller, R. A. Leitz, C. D. Nickell, and W. L. Pedersen. 1998. “Yield and Seed Quality of Soybean Cultivars Infected with Sclerotinia sclerotiorum.” *Plant Disease* 82: 826–829. doi:10.1094/PDIS.1998.82.7.826.

Huang, B., and H. Zhang. 2014. “Spatio-temporal Reflectance Fusion via Unmixing: Accounting for Both Phenological and Land-cover Changes.” *International Journal of Remote Sensing* 35: 6213–6233. doi:10.1080/01431161.2014.951097.

Jaynes, D. B., T. C. Kaspar, T. S. Colvin, and D. E. James. 2003. “Cluster Analysis of Spatiotemporal Corn Yield Patterns in an Iowa Field.” *Agronomy Journal* 95: 1–13.

Jaynes, D. B., and T. S. Colvin. 1997. “Spatiotemporal Variability of Corn and Soybean Yield.” *Agronomy Journal* 89: 30–37. doi:10.2134/agronj1997.00021962008900010005x.

Jaynes, D. B., T. S. Colvin, and T. C. Kaspar. 2005. “Identifying Potential Soybean Management Zones from Multi-year Yield Data.” *Computers and Electronics in Agriculture* 46: 309–327. doi:10.1016/j.compag.2004.11.011.

Kozlov, M. V., and N. G. Berlina. 2002. “Decline in Length of the Summer Season on the Kola Peninsula, Russia.” *Climatic Change* 54: 387–398. doi:10.1023/A:1016175101383.

Kumhálová, J., F. Kumhala, M. Kroulik, and Š. Matějková. 2011. “The Impact of Topography on Soil Properties and Yield and the Effects of Weather Conditions.” *Precision Agriculture* 12: 813–830. doi:10.1007/s11119-011-9221-x.

Kurile, J. E., C. R. Gran, E. S. Oplinger, and A. Mengistu. 2001. “Tillage, Crop Sequence, and Cultivar Effects on Sclerotinia Stem Rot Incidence and Yield in Soybean.” *Agronomy Journal* 93: 973–982. doi:10.2134/agronj2001.935973x.

Li, L., Y. Zhao, Y. Fu, Y. Pan, L. Yu, and Q. Xin. 2017. “High Resolution Mapping of Cropping Cycles by Fusion of Landsat and MODIS Data.” *Remote Sensing* 9: 1–19. doi:10.3390/rs91211232.

Linderholm, H. W. 2006. “Growing Season Changes in the Last Century.” *Agricultural and Forest Meteorology* 137: 1–14. doi:10.1016/j.agrformet.2006.03.006.

Mesonet, S. D. 2018. South Dakota Mesonet Database [database]. ed. S. D. S. University.

Mfuka, C., X. Zhang, and E. Byamukama. 2019. “Mapping and Quantifying White Mold in Soybean across South Dakota Using Landsat Images.” *Journal of Geographic Information System* 11: 331–346. doi:10.4236/jgis.2019.1113020.

Milne, G. M., J. Graham, A. Allen, A. Lahuerta-Marin, C. McCormick, E. Presho, R. Skuce, and A. W. Byrne. 2019. “Spatiotemporal Analysis of Prolonged and Recurrent Bovine Tuberculosis Breakdowns in Northern Irish Cattle Herds Reveals a New Infection Hotspot.” *Spatial and Spatio-temporal Epidemiology* 28: 33–42. doi:10.1016/j.sste.2018.11.002.

Moral García, F. J. 2006. “Analysis of the Spatio–Temporal Distribution of Helicoverpa Armigera Hb. In a Tomato Field Using a Stochastic Approach.” * Biosystems Engineering* 93: 253–259. doi:10.1016/j.biosystemseng.2005.12.011.

Moran, P. A. 1950. “Notes on Continuous Stochastic Phenomena.” *Biometrika* 37: 17–23. doi:10.1093/biomet/37.1-2.17.

Nelson, K. A., K. A. Renner, and R. Hammerschmidt. 2002. “Cultivar and Herbicide Selection Affects Soybean Development and the Incidence of Sclerotinia Stem Rot.” *Agronomy Journal* 94: 1270–1281. doi:10.2134/agronj2002.1270.

Onojehgu, A. O., G. A. Blackburn, Q. Wang, P. M. Atkinson, D. Kindred, and Y. Miao. 2018. “Rice Crop Phenology Mapping at High Spatial and Temporal Resolution Using Downscaled MODIS Time-series.” *GI Science & Remote Sensing* 55: 659–677. doi:10.1080/15481603.2018.1423725.

Piao, S., P. Friedlingstein, P. Ciais, N. Viyov, and J. Demarty. 2007. “Growing Season Extension and Its Impact on Terrestrial Carbon Cycle in the Northern Hemisphere over the past 2 Decades.” *Global Biogeochemical Cycles* 21: 1–11. doi:10.1029/2006GB002888.

Piper, E. L., and K. J. Boote. 1999. “Temperature and Cultivar Effects on Soybean Seed Oil and Protein Concentrations.” *Journal of the American Oil Chemists Society* 76: 1233–1241. doi:10.1007/s11746-999-0099-y.

Roel, A., and R. E. Plant. 2004. “Factors Underlying Yield Variability in Two California Rice Fields.” *Agronomy Journal* 96: 1481–1494. doi:10.2134/agronj2004.1481.

Rouse, J. W., J. H. Haas, J. A. Schell, and D. W. Deering. 1973. “Monitoring Vegetation Systems in the Great Plains with ERTS.” Third ERTS Symposium, Washington, DC, December 10–14, 309–317.
Rousseau, G. X., S. Rioux, and D. Dostaler. 2006. “Multivariate Effects of Plant Canopy, Soil Physicochemistry and Microbiology on Sclerotinia Stem Rot of Soybean in Relation to Crop Rotation and Urban Compost Amendment.” Soil Biology & Biochemistry 38: 3325–3342. doi:10.1016/j.soilbio.2006.04.054.

Roy, D. P., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen, and E. Lindquist. 2008. “Multi-temporal MODIS–Landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data.” Remote Sensing of Environment 112: 3112–3130. doi:10.1016/j.rse.2008.03.009.

RStudio Team. 2016. RStudio: Integrated Development for R. Boston, MA: RStudio. http://www.rstudio.com/

Sara, S. A., E. B. McCallen, and P. V. Switzer. 2013. “The Spatial Distribution of the Japanese Beetle, Popillia Japonica, in Soybean Fields.” Journal of Insect Science 13: 1–9. doi:10.1673/031.013.3601.

Schaaf, C., and Z. Wang. 2015. “MCD43A4 MODIS/Terra +Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006 [Data Set].” NASA EOSDIS Land Processes DAAC. doi:10.5067/MODIS/MCD43A4.006.

Seiter, N. J., F. P. Reay-Jones, and J. K. Greene. 2013. “Within-Field Spatial Distribution of Megacopta Cribaria (Hemiptera: Plataspidae) in Soybean (Fabales: Fabaceae).” Environmental Entomology 42: 1363–1374. doi:10.1603/EN13199.

Soltani, A., and S. Askari. 2017. “Exploring Spatial Autocorrelation of Traffic Crashes Based on Severity.” Injury 48: 637–647. doi:10.1016/j.injury.2017.01.032.

Thompson, D. R., and O. A. Wehmanen. 1980. “Using Landsat Digital Data to Detect Moisture Stress in Corn-soybean Growing Regions.” Photogrammetric Engineering and Remote Sensing 46: 1087–1093.

Tucker, C. J., D. A. Slayback, J. E. Pinzon, S. O. Los, R. B. Myneni, and M. G. Taylor. 2001. “Higher Northern Latitude Normalized Difference Vegetation Index and Growing Season Trends from 1982 to 1999.” International Journal of Biometeorology 45: 184–190. doi:10.1007/s00484-001-0109-8.

USDA. 2017. National Agricultural Statistics Services. United States Department of Agriculture. https://www.nass.usda.gov/.

Vieira, S. R., and A. P. Gonzalez. 2003. “Analysis of the Spatial Variability of Crop Yield and Soil Properties in Small Agricultural Plots.” Bragantia, Campinas 62: 127–138. doi:10.1590/S0006-87052003000100016.

Vigier, B. J., E. Pattey, and I. B. Strachan. 2004. “Narrowband Vegetation Indexes and Detection of Disease Damage in Soybeans.” IEEE Geoscience and Remote Sensing Letters 1: 255–259.

Weng, Q., P. Fu, and F. Gao. 2014. “Generating Daily Land Surface Temperature at Landsat Resolution by Fusing Landsat and MODIS Data.” Remote Sensing of Environment 145: 55–67. doi:10.1016/j.rse.2014.02.003.

Wrather, A., G. Shannon, R. Balardin, L. Carregal, R. Escobar, G. K. Gupta, Z. Ma, W. Morel, D. Ploper, and A. Tenuta. 2010. “Effect of Diseases on Soybean Yield in the Top Eight Producing Countries in 2006.” Plant Health Progress 11: 29.

Wu, M., C. Wu, W. Huang, Z. Niu, C. Wang, W. Li, and P. Hao. 2016. “An Improved High Spatial and Temporal Data Fusion Approach for Combining Landsat and MODIS Data to Generate Daily Synthetic Landsat Imagery.” Information Fusion 31: 14–25. doi:10.1016/j.inffus.2015.12.005.

Yang, X. B., and F. Feng. 2001. “Ranges and Diversity of Soybean Fungal Diseases in North America.” Phytopathology 91: 769–775. doi:10.1094/PHYTO.2001.91.8.769.

Zhu, X., J. Chen, F. Gao, X. Chen, and J. G. Masek. 2010. “An Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model for Complex Heterogeneous Regions.” Remote Sensing of Environment 114: 2610–2623. doi:10.1016/j.rse.2010.05.032.

Zurita-Milla, R., J. Clevers, and M. E. Schaepman. 2008. “Unmixing-Based Landsat TM and MERIS FR Data Fusion.” IEEE Geoscience and Remote Sensing Letters 5: 453–457. doi:10.1109/LGRS.2008.919685.