

South Dakota State University

# Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange

---

Electronic Theses and Dissertations

---

2019

## Characterizing Spatiotemporal Patterns of White Mold in Soybean across South Dakota Using Remote Sensing

Confiance L. Mfuka

Follow this and additional works at: <https://openprairie.sdstate.edu/etd>



Part of the [Agronomy and Crop Sciences Commons](#), [Geographic Information Sciences Commons](#), [Physical and Environmental Geography Commons](#), [Plant Pathology Commons](#), [Remote Sensing Commons](#), and the [Spatial Science Commons](#)

---

### Recommended Citation

Mfuka, Confiance L., "Characterizing Spatiotemporal Patterns of White Mold in Soybean across South Dakota Using Remote Sensing" (2019). *Electronic Theses and Dissertations*. 3396.  
<https://openprairie.sdstate.edu/etd/3396>

This Dissertation - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact [michael.biondo@sdstate.edu](mailto:michael.biondo@sdstate.edu).

CHARACTERIZING SPATIOTEMPORAL PATTERNS OF WHITE MOLD IN  
SOYBEAN ACROSS SOUTH DAKOTA USING REMOTE SENSING.

BY

CONFIANCE L. MFUKA

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Plant Science

South Dakota State University

2019



*A ma progéniture, pour que ce travail constitue un repère dans leur avenir scientifique.*

## ACKNOWLEDGEMENTS

This work would not have been possible without the financial supports from the Department of Agronomy, Horticulture and Plant Science, as well as the Geography Department of South Dakota State University. I am deeply indebted to Dr. Emmanuel Byamukama, whose patience, trust, understanding, and encouragements have been more than vital for the completion of this research and for my personal growth as a responsible human. I would also like to express my deepest appreciation to Dr. Xiaoyang Zhang, whose scientific and unwavering insights have been essential to this achievement.

I would like to extend my deepest gratitude to each of my committee members, namely Dr. Bruce Millett and Dr. Jayarama Gunaje, for their valuable and enriching inputs, as well as the whole SDSU scientific community for their diverse contributions along my scientific journey. I am also grateful to Dr. Trisha Jackson, who encouraged me to start and pursue this program, for her profound belief in my abilities, and Dr. David Roy who motivated me throughout the difficult steps that I came across.

Finally, I would like to thank the Congolese community of Brookings SD, my Pastors, David and Jeanne Kaufman for spiritual guidance and support, and the whole Church of Holylife Tabernacle. Most importantly, I wish to thank my loving and supportive wife, Rose, and my wonderful children, Latifah-Rose, Loghan-Boniface, Lincoln-Michael, and Lola-Christine who provided an endless inspiration.

## CONTENT

ABBREVIATIONS .....	ix
LIST OF FIGURES .....	xiii
LIST OF TABLES .....	xvii
ABSTRACT.....	xix
CHAPTER 1: LITTERATURE REVIEW.....	1
1.1. Soybean ( <i>Glycine max</i> ) generalities.....	1
1.2. Soybean diseases and impacts on yield.....	3
1.3. Crop disease detection and mapping with Remote Sensing.....	7
1.3.1. Disease detection with remote sensing .....	7
1.3.2. Landsat data fusion for crops time-series analyses.....	10
1.4. Research goal, objectives and hypotheses.....	15
1.5. Significance of the study .....	16
1.6. Organization of the dissertation .....	17
CHAPTER 2: MAPPING AND QUANTIFYING WHITE MOLD IN SOYBEAN ACROSS SOUTH DAKOTA USING LANDSAT IMAGES.....	19
2.1. Abstract .....	19
2.2. Introduction .....	20
2.3. Materials and methods .....	23

2.3.1.	Study area.....	23
2.3.2.	Data gathering.....	25
2.3.3.	Random Forest classifiers for mapping soybean and white mold .....	26
2.4.	Results and discussion.....	30
2.4.1.	Land cover spectral separability .....	30
2.4.2.	Land cover classification results .....	33
2.4.3.	Land cover map accuracy assessment.....	34
2.4.4.	White mold mapping.....	36
2.4.5.	White mold map accuracy assessment.....	37
2.4.6.	Quantified soybean and white mold .....	38
2.5.	Conclusion.....	39
 CHAPTER 3: SPATIOTEMPORAL CHARACTERISTICS OF WHITE MOLD AND IMPACTS ON YIELD IN SOYBEAN FIELDS IN SOUTH DAKOTA .....		 42
3.1.	Abstract .....	43
3.2.	Introduction .....	43
3.3.	Materials and Methods .....	45
3.3.1.	Study area.....	45
3.3.2.	Data collection and preparation .....	47
3.3.3.	Analyses .....	50

3.4.	Results and discussion.....	54
3.4.1.	Data fusion .....	54
3.4.2.	Yield spatial distribution.....	56
3.4.3.	The yield ordinary kriging model .....	59
3.4.4.	Seasonal changes of the relationship between yield and NDVI .....	61
3.4.5.	Estimating yield loss caused by white mold .....	63
3.5.	Conclusion.....	67
CHAPTER 4: MODELING TEMPORAL PATTERNS OF WHITE MOLD IN SOUTH DAKOTA USING FUSION OF LANDSAT AND MODIS DATA.....		69
4.1.	Abstract .....	70
4.2.	Introduction .....	71
4.3.	Materials and Methods .....	72
4.3.1.	Study area.....	72
4.3.2.	Data collection and preparation .....	74
4.3.3.	Data fusion .....	74
4.3.4.	NDVI computation.....	75
4.4.	Analyses .....	75
4.5.	Results and Discussion.....	77
4.5.1.	Time-series NDVI for healthy soybeans and white mold.....	77

4.5.2.	Temporal change in the NDVI difference between healthy soybeans and white mold .....	80
4.5.3.	Temporal modeling of white mold .....	82
4.6.	Conclusion.....	85
CHAPTER 5: SUMMARY OF THE RESEARCH.....		87
5.1.	Hypothesis #1:.....	87
5.1.1.	Summary of the methods .....	87
5.1.2.	Results and Conclusions .....	87
5.1.3.	Implications and future work .....	88
5.2.	Hypothesis #2:.....	89
5.2.1.	Summary of the methods .....	89
5.2.2.	Results and Conclusions .....	90
5.2.3.	Implications and future work .....	90
5.3.	Hypothesis #3:.....	91
5.3.1.	Summary of the methods .....	91
5.3.2.	Results and Conclusions .....	92
5.3.3.	Implications and Future Work .....	92
REFERENCES .....		94

## ABBREVIATIONS

ADAR	Airborne Data Acquisition and Registration
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
ARD	Analysis-Ready Dataset
AUC	Area Under the Curve
AVIRIS	Airborne Visible Infrared Imaging Spectrometer
BRDF	Bi-directional Reflectance Distribution Function
BT	Brightness Temperature
CDL	Crop Data Layer
DAS	Days After Sowing
DOY	Day of the Year
DTC	Decision Tree Classifiers
ENVI	Environment for Visualizing Images
ESTARFM	Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model
ETM+	Enhanced Thematic Mapper +
FAO	Food and Agriculture Organization
FSDAF	Flexible Spatiotemporal Data Fusion
GEIFM	Global Empirical Image Fusion Model
GLD	Grapevine Leafroll Disease
GLM	General Linear Model
HSD	Honest Significant Differences
ISTDFA	Improved Spatial Temporal Data Fusion Approach

JM	Jeffries-Matsusita Index
LDC	Late Diseases Complex
LDF	Linear Discriminant Function
LIM	Linear Interpolation Model
LST	Land Surface Temperature
MLC	Maximum Likelihood Classifiers
MODIS	Moderate Resolution Imaging Spectroradiometer
MP	Moldboard Plowing
MRT	MODIS Reprojection Tool
MT	Mulch Tillage
NASA	National Aeronautic and Space Administration
	Nadir Bidirectional Reflectance Distribution Function -Adjusted
NBAR	Reflectance
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared Band
NT	No-Till
OLI	Operational Land Imager
OS	Other Stress
PP	Total Precipitations
PRR	Phytophthora Root Rot
QA	Quality Assessment
ROC	Receiver Operating Characteristic
SADFAT	Spatiotemporal Adaptive Data Fusion Algorithm for Temperature

SCN	Soybean Cyst-Nematode
SDS	Sudden Death Syndrome
SMA	Spectral Mixing Analysis
SMV	Soybean Mosaic Virus
	Sparse Representation-based Spatiotemporal Reflectance Fusion
SPSTFM	Model
SR	Surface Reflectance
SSC	Soybean Stem Canker
SSR	Sclerotinia Stem Rot
	Spatial Temporal Adaptive Algorithm for mapping Reflectance
STAARCH	Change
STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
STDFA	Spatial Temporal Data Fusion Approach
STRUM	Spatial and Temporal Reflectance Unmixing Model
SVM	Support Vector Machine
SWIR	Shortwave Infrared Band
TM	Thematic Mapper
TOA	Top-Of-Atmosphere
UAV	Unmanned Aerial Vehicle
UBDF	Unmixing-Based Data Fusion
USA	United States of America
USDA	United States Department of Agriculture
U-STFM	Unmixing-based Spatio-Temporal Reflectance Fusion Model

WM	White Mold
YLV	Yellow Leaf Virus

## LIST OF FIGURES

Figure 1. 1. World's top five Soybeans producers. Source:FAOSTATS <a href="http://www.fao.org/faostat/en/#rankings/countries_by_commodity">(<a href="http://www.fao.org/faostat/en/#rankings/countries_by_commodity">http://www.fao.org/faostat/en/#rankings/countries_by_commodity</a>)</a> .....	2
Figure 2. 1. Study area showing the three counties (Marshall, Day, and Codington) in Northeastern South-Dakota. The background image is a Landsat false color combination of bands 6-5-4.....	24
Figure 2. 2. Class spectral separability: each plot shows the ability and the contribution of each band in separating land cover classes. ....	32
Figure 2. 3. A comparison between a July 14 false color 6-5-4 Landsat 8 image (A), the Crop Data Layer (CDL) map (B) and the resulting classification (C) of the stacked May and July images. showing similarities between the resulting classification and the CDL. ....	34
Figure 2. 4. August Landsat composite (A), August Landsat NDVI with white mold range (B), and mapped soybean and white mold (C): White mold is accurately mapped from the soybean mask, using the appropriate NDVI signal.....	37
Figure 2. 5. White mold in northeastern South Dakota: the map shows a classified image in background with the four important classes and the quantified white mold over the soybean mask. The white circles indicate ground trothed white mold fields.....	41
Figure 3. 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 false	

- color bands 6-5-4 combination from July 14, 2017 (Marshall) and July 20, 2016 (Moody)..... 46
- Figure 3. 2. An original Landsat image (A) and an ESTARFM predicted image (B) for the same date (DOY 163). The two images look very similar, a few differences in the brightness can be noticed due to the ENVI software enhancement for visualization. The two images are a false color composite using a combination of bands 6-5-4. .... 55
- Figure 3. 3. Comparison of the actual and predicted band reflectance in a heterogeneous region. The upper-left and the lower-right quadrats show the Normalized Difference Vegetation Index (NDVI) respectively from the original Landsat and the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) image. The lower-left quadrat shows the NDVI distribution along the 1:1 line and the upper-right quadrat shows the correlation coefficient between the two images NDVI. .... 56
- Figure 3. 4. Moran's I scatter plot showing evidence of clustering in both years: clusters are mostly located in the lower-left quadrat, which explains low yield (and probably crop stress) and in the upper-right quadrat which explains a high yield. Evidence of clustering is more pronounced in 2017 than in 2016. .... 58
- Figure 3. 5. Semivariogram of the 2016 yield Ordinary Kriging (OK) model. The blue crosses depict the computed predicted values, while the model is indicated by the blue line. .... 60
- Figure 3. 6. Normalized Difference Vegetation Index (NDVI) classes (left) compared with the corresponding yield (right) for each class for the soybean field in

Moody County, 2016. White mold pixels correspond to the three lowest NDVI classes. .... 64

Figure 3. 7. Normalized Difference Vegetation Index of a soybean field that had white mold (top) compared with the corresponding yield (bottom) in Marshall County in 2017. .... 65

Figure 3. 8. Yield loss as a result of the Normalized Difference Vegetation Index (NDVI) change. .... 67

Figure 4. 1. Location of the study area showing different scouted fields with their respective weather station names. The background image is a July 14, 2017 false color composite of Landsat 8 using bands 6-5-4. .... 73

Figure 4. 2. Average Normalized Difference Vegetation Index (NDVI) values computed for each weather station. Values distribution is similar with no outliers. Brit=Britton, Grot=Groton, Web=Webster, SS= South Shore. .... 77

Figure 4. 3. Scatter plot of the frequency distribution (y axis) of the Normalized Difference Vegetation Index (NDVI, x axis) for healthy soybeans (blue) and white mold (pink). .... 78

Figure 4. 4. Seasonal Normalized Difference Vegetation Index (NDVI) of healthy soybean as compared to white mold for each site. White mold starts declining earlier than healthy soybean and has lower NDVI after Day of the Year (DOYS) 207. .... 80

Figure 4. 5. Box plots of Normalized Difference Vegetation Index (NDVI) difference by site. .... 80

Figure 4. 6. Area Under the Curve computed by the two models, showing a better (bigger area) estimation for the model that includes the Normalized Difference Vegetation Index (NDVI) and Day of the Year (DOY)..... 85

## LIST OF TABLES

Table 2. 1. Original Landsat 8 bands including the Shortwave Infrared (SWIR), the Near Infrared (NIR), the red (RED), the green (GREEN) and the blue (BLUE) bands, and their corresponding names used in the Random Forest (RF) classification, and in the stacked image. ....	28
Table 2. 2. Jeffries-Matusita (JM) spectral separability index showing the goodness of the trainings .....	31
Table 2. 3. Confusion matrix of the land cover map accuracy assessment .....	35
Table 2. 4. White mold accuracy assessment: Confusion matrix table comparing the mapped classes with ground truth. ....	38
Table 2. 5. Comparison between soybean area estimates from the United States Department of Agriculture (USDA) and the classified map in this study, as well as white mold extent estimated for each county, based on the calculations from the Landsat pixel size (30mx30m) and the total number of pixels. ....	38
Table 3. 1. Landsat Operational Land Imager (OLI) and the Moderate Resolution Imaging Spectroradiometer (MODIS) bands designation and wavelengths ( $\mu\text{m}$ ) combined using ESTARFM. ....	49
Table 3. 2. Monte-Carlo simulation of Moran I.....	57
Table 3. 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.....	65
Table 3. 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.....	66
Table 4. 1. Tukey Honest Significant Differences test for detecting similarities in the NDVI difference temporal change between sites, showing the 95% lower and upper limit values. ....	81
Table 4. 2. White mold modeled as a function of the Normalized Difference Vegetation Index (NDVI) only. ....	83
Table 4. 3. White mold modeled as a function of the Normalized Difference Vegetation Index (NDVI) and Day of the Year (DOY). ....	83
Table 4. 4. Comparison between the Normalized Difference Vegetation Index (NDVI) model and the NDVI + (Day of the Year) DOY models.....	84

## ABSTRACT

CHARACTERIZING SPATIOTEMPORAL PATTERNS OF WHITE MOLD IN  
SOYBEAN ACROSS SOUTH DAKOTA USING REMOTE SENSING.

CONFIANCE L. MFUKA

2019

Soybean is among the most important crops, cultivated primarily for beans, which are used for food, feed, and biofuel. According to FAO, the United States was the biggest soybeans producer in 2016. The main soybean producing regions in the United States are the Corn Belt and the lower Mississippi Valley. Despite its importance, soybean production is reduced by several diseases, among which *Sclerotinia* stem rot, also known as white mold, a fungal disease that is caused by the fungus *Sclerotinia sclerotiorum* is among the top 10 soybean diseases. The disease may attack several plants and considerably reduce yield. According to previous reports, environmental conditions corresponding to high yield potential are most conducive for white mold development. These conditions include cool temperature (12-24 °C), continued wet and moist conditions (70-120 h) generally resulting from rain, but the disease development requires the presence of a susceptible soybean variety. To better understand white mold development in the field, there is a need to investigate its spatiotemoral characteristics and provide accurate estimates of the damages that white mold may cause. Current and accurate data about white mold are scarce, especially at county or larger scale. Studies that explored the characteristics of white mold were generally field oriented and local in scale, and when the spectral characteristics were investigated, the authors used

spectroradiometers that are not accessible to farmers and to the general public and are mostly used for experimental modeling. This study employed free remote sensing Landsat 8 images to quantify white mold in South Dakota. Images acquired in May and July were used to map the land cover and extract the soybean mask, while an image acquired in August was used to map and quantify white mold using the random forest algorithm. The land cover map was produced with an overall accuracy of 95% while white mold was mapped with an overall accuracy of 99%. White mold area estimates were respectively 132 km<sup>2</sup>, 88 km<sup>2</sup>, and 190 km<sup>2</sup>, representing 31%, 22% and 29% of the total soybean area for Marshall, Codington and Day counties. This study also explored the spatial characteristics of white mold in soybean fields and its impact on yield. The yield distribution exhibited a significant positive spatial autocorrelation (Moran's I = 0.38, p-value < 0.001 for Moody field, Moran's I = 0.45, p-value < 0.001, for Marshall field) as an evidence of clustering. Significant clusters could be observed in white mold areas (low-low clusters) or in healthy soybeans (high-high clusters). The yield loss caused by the worst white mold was estimated at 36% and 56% respectively for the Moody and the Marshall fields, with the most accurate loss estimation occurring between late August and early September. Finally, this study modeled the temporal evolution of white mold using a logistic regression analysis in which the white mold was modeled as a function of the NDVI. The model was successful, but further improved by the inclusion of the Day of the Year (DOY). The respective areas under the curves (AUC) were 0.95 for NDVI and 0.99 for NDVI+DOY models. A comparison of the NDVI temporal change between different sites showed that white mold temporal development was affected by the site location, which could be influenced by many local parameters such as the soil

properties, the local elevation, management practices, or weather parameters. This study showed the importance of freely available remotely sensed satellite images in the estimation of crop disease areas and in the characterization of the spatial and temporal patterns of crop disease; this could help in timely disease damage assessment.

## CHAPTER 1: LITTERATURE REVIEW

### 1.1. SOYBEAN (*GLYCINE MAX*) GENERALITIES

Soybean (*Glycine max*) is an important crop, primarily grown for beans, which can be processed into human food, animal feed, oil and other bioproducts. In the United States, the main soybean producing area is in the Corn Belt and the lower Mississippi Valley (USDA 2010). The usual planting dates are between May 8 and June 21, with the most active dates between May 15 and June 11, while the harvesting dates are reported between September 22 and November 3, with the most active dates between September 28 and October 24. (USDA 2010).

According to FAO, the world's top five soybeans producers in 2016 are respectively USA, Brazil, Argentina, India, and China (Figure 1), with the USA producing about 36% of the total world's production (117M tons). Brazil ranks second with 96M tons, Argentina third with 58M, India fourth with 14M and China Mainland fifth with about 11M tons. Masuda and Goldsmith (2009) estimated country-level production and projection through the year 2030, using the yield and the harvested area (Box-Jenkins model employing exponential smoothing with a damped trend) as components of production in their model. Their results suggest annual growth rates of 2.5% from 2010 to 2020 and 1.8% from 2020 to 2030, reaching respectively 311.1 million metric tons in 2020 and 371.3 million metric tons in 2030. These rates seem decreasing though crop production needs to double by 2050 to meet world's population food demands. Projections suggest therefore, that the trends are insufficient (Ray et al. 2013).

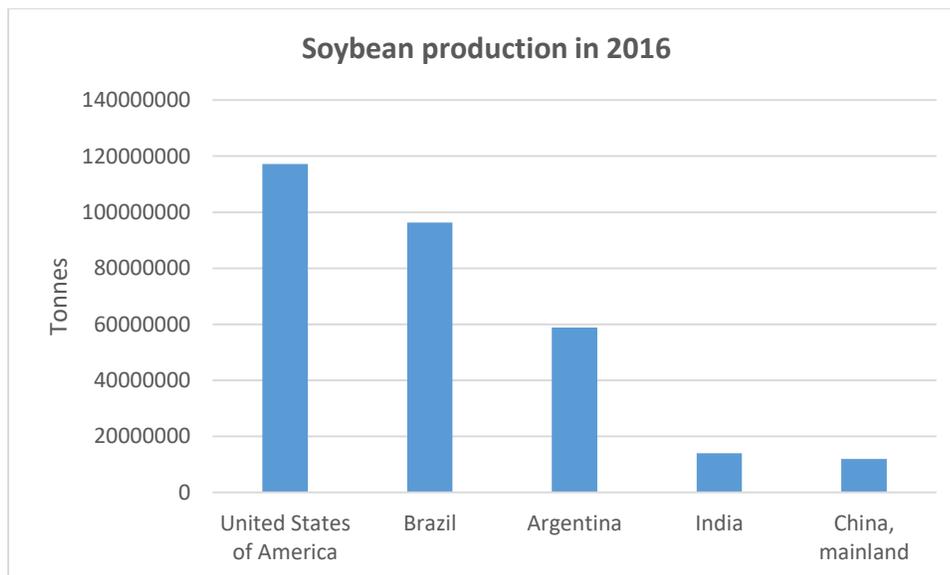


Figure 1. 1. World's top five Soybeans producers. Source:FAOSTATS (

[http://www.fao.org/faostat/en/#rankings/countries\\_by\\_commodity](http://www.fao.org/faostat/en/#rankings/countries_by_commodity))

Lobell and Asner argued that the previously predicted trends have been overestimated because they analyzed the impact of climate change in the crop yields (especially corn and soybean), and observed a 17% yield decrease for each degree increase in growing season temperature (Lobell and Asner 2003). Further, climate impacts on yields have been studied and estimations showed a worldwide balance in which the USA exhibited a negative impact on yield (Lobell, Schlenker and Costa-Roberts 2011).

Beside the climate change influence, many other environmental or agricultural factors can affect soybean yields. Soil properties such as slope and texture (fine sandy soil), along with base saturation, pH, clay content, and elevation considerably explain soybean yield variability (Jiang and Thelen 2004). Kravchenko and Bullock (2000) found

similar results and argued that these topographic influences on yield variabilities can differ from one site to another. Water table management contributes to 32-37% increase in soybeans yields, as compared to a free drainage, and the increase is exacerbated in drier years (Mejia, Madramootoo and Broughton 2000). In addition, while crop rotation is more favorable to increased yields than a monoculture (Peterson and Varvel 1989), a higher rotation sequence including three or more crops is even more suitable for increased soybeans yields (Crookston et al. 1991).

## 1.2. SOYBEAN DISEASES AND IMPACTS ON YIELD

Several diseases cause significant soybean yield losses annually (Allen et al. 2017). Among these, *Sclerotinia* stem rot also known as white mold has been consistently ranked among the top ten major soybean diseases across soybean growing states (Wrather et al. 2010, Allen et al. 2017). Many authors have studied soybean diseases, but few have focused on white mold of soybean. White mold is a fungal disease, caused by *Sclerotinia sclerotiorum* (Boland and Hall 1988). Most of the studies on white mold focused on factors influencing diseases development and effect on soybean yield. These factors include fungicide applications and difference in the timing of fungicide applications (Bradley 2008, Carmona et al. 2011), planting dates, inoculation stage (Danielson, Nelson and Helms 2004), and agricultural practices such as tillage or crops sequence (Kurle et al. 2001). Attention around white mold developed in the 1980s, when the disease epidemiology was documented in Ontario by Boland and Hall (1988). According to their observations, the development of crop canopy between mid and late July was followed by apparition of apothecia. The study highlighted important events that favors the occurrence of white mold: crop canopy closure, flowering, apothecia appearance and

ascospores germination, and environmental conditions essentially corresponding to wetness lasting 40-112 after rain and air temperature between 12-24 C. These conditions are determined as suitable for high yield potential, but also are favorable to development of white mold.

Fungicides and some herbicides have been used to manage white mold. The frequency and timing of the application, as well as inoculation stage influence the level of white mold development and the resulting yield. The effect of two treatments (2,6-dichloroisonicotinic acid- INA and benzothiadiazole- BTH) on soybean yields and on disease severity caused by *Sclerotinia sclerotiorum* was examined by Dann et al. (1998). The authors found that the severity of white mold was reduced by 20-70% by three or four applications of INA as compared with the control, and by 20-60% by two or four applications of BTH. In both cases, the greatest reduction of disease severity was observed in susceptible cultivars and yield was negatively correlated with disease severity. Danielson et al. (2004) examined effects of white mold on soybeans yields inoculated at two different growth stages (R3 and R5); the authors measured seed weight, seed and pod numbers, seed protein, and oil content; they found that the disease resulted in significant seed weight loss, and the yield loss was estimated to an average of 136.6 kg/ha per 10% disease increase. In some experiments, the disease resulted to a reduction in the number of seeds and pods per plant and seed oil content.

The importance of cultivar selection has been highlighted in several studies, as some cultivars might be more susceptible to certain diseases than other. Buzzell, Welacky and Anderson (1993) conducted a study in which they found that soybeans cultivars selection had a significant importance in the occurrence of white mold some cultivars produced

more sclerotia per diseased plant than others. They also noticed that white mold occurred less in cultivars that matured earlier than in late maturing crops, which justified partially the cultivar selection. Hoffman et al. (1998) examined yield and seed quality response of different soybean cultivars to white mold. They found that the disease incidence was a function of the choice of the soybean cultivar; also, the white mold incidence increase corresponded with yield reduction with significant difference among cultivars. They also noted that the disease incidence was positively correlated with the number of sclerotia for all the cultivars, and negatively correlated to seed weight and seed oil content. These results were confirmed by Yang, Lundeen and Uphoff (1999).

Management practices are also an important factor that influences the development of white mold. The effects of tillage, and crop sequence on white mold incidence and yield in soybean have been investigated by Kurle et al. (2001). The authors found a simple linear relationship between white mold incidence and yield with a correlation coefficient of 35%, a significant effect of crop rotation (especially when soybean was planted after corn and oat) and tillage in the reduction of white mold, and the cultivar effect due to their susceptibility to white mold. Similarly, Mueller, Hartman and Pedersen (2002) examined the effects of crop rotation and tillage on white mold by comparing moldboard plowing (MP), mulch tillage (MT) and no-till (NT) and investigating their effects on sclerotia and apothecia counts and distribution, and on yield. They found that white mold incidence was not significantly affected by crop rotation, while tillage affected both the number and distribution of sclerotia in the soil profile. MP was more effective than MT and NT, and lowered the number of sclerotia by burying them deeper than 10cm under soil; MP also delayed apothecia germination when

compared to NT and MT. Milas et al. (2003) modeled the prevalence of white mold using management practices (tillage, herbicide, manure and fertilizer application, and seed treatment with fungicide) and weather data (summer air temperature and precipitation) and established relationship between the two datasets and soybean yield, with a multiple linear regression coefficient of 0.27. The authors suggested that the occurrence of white mold was associated with environments of high potential yields. These environmental conditions were associated with variables of significance such as average air temperature during July and August, precipitation during July, tillage, seed treatment, liquid manure, fertilizer, and herbicide applications.

The above reviewed studies established a knowledge benchmark on several aspects of white mold; however, they focused on field observations and experiments that do not provide a large spatial picture of the disease. The field methods developed however reproducible, are laborious if replicated over large areas. Furthermore, local and regional differences can not easily be accounted for while analyzing sparse fields, which makes it hard to draw conclusions for a large spatial scale. With remote sensing, satellite images can be used to provide consistent observations on relatively larger areas in order to extract and interpret field differences. However, the use of remote sensing for interpreting a given phenomenon requires a strong knowledge of the modeled phenomenon, as well as an informed understanding in the choice of the data able to explain the phenomenon being modeled.

### 1.3. CROP DISEASE DETECTION AND MAPPING WITH REMOTE SENSING

#### 1.3.1. Disease detection with remote sensing

Remote sensing has played a significant role in detecting and modeling vegetation stress, especially in crops. A milestone in the use of images for crop monitoring was the establishment of the relationship between the measured reflectance and a determined pigment (i.e. chlorophyll a, b, carotenoid) quantity. One of the first attempts to characterize plant disease with remote sensing analyzed the change in the spectral reflectance of field bean leaves (*Vicia faba*) infected by the *Botrytis fabae* (Malthus and Madeira 1993). The researchers observed an increase of the lesion area corresponding to the increase in the reflectance in the blue (470-500nm) and the red (590-700nm) regions of the visible, and a decrease of the reflectance in the green region (550nm); however, the highest correlations of reflectance with infection area were found in the near infrared region (>720nm). These changes also corresponded to the decrease of chlorophyll a, xanthophyll and carotenoids, and an increase of chlorophyll b. Other similar studies showed the decrease in chlorophyll content related to the change in reflection spectra, (Polischuk et al. 1997, Kobayashi et al. 2001).

Most of the subsequent studies employed the same principles and similar methods: they measured field disease severity that they related to reflectance data. However, the reflectance data were mostly collected from field measurements using hand-held spectrometers (Grisham, Johnson and Zimba 2010), spectro-radiometers (Muhammed 2005), spectrographs (Moshou et al. 2005) over a wavelength range in regular intervals, or using digital cameras (Fitzgerald, Maas and Detar 2004). Results are complementary to each other as they pertain to different crops and diseases, and the different studies

highlight the importance of given wavelengths over others for different disease detection. Bravo et al. (2003) studied early yellow rust (*Puccinia striiformis*) detection in wheat fields, suggesting that the two best wavebands used to detect disease were at 750 and 630nm, while in a study investigating the sugar cane Yellow Leaf Virus (YLV), the discrimination of YLV was important in several spectral regions including the ultraviolet (220–320nm), the violet and blue (400–500nm), the green and yellow (500–590nm), orange and red (590–650nm), and the near-infrared (740–850nm)(Grisham et al. 2010). Similarly, Naidu et al. (2009) investigated the change in leaf spectral reflectance between virus-infected and uninfected grapevine (*Vitis vinifera L.*) to detect grapevine leaf roll disease (GLD); differences between the diseased and healthy grapevines were observed at wavelengths 550nm (green), 900nm (near infrared), and 1600nm (mid-infrared)

While the field reflectance measurements represent a viable data for modeling, it also has many limitations: (1) the reflectance can only be collected on a limited area; (2) the measurements need to be made in similar environmental conditions, which is very critical for data consistency; (3) collected data cannot be extrapolated over larger areas, limiting the data only for the explored fields; (4) measurement tools are expensive and generally not accessible to the public. These constraints reduce the usability of such reflectance field measurements. These limitations can be solved by the use of remotely sensed images, which can provide observations spanning big areas, can be calibrated based on sensor characteristics, can incorporate different fields in a single observation, and in some cases, can be free of charge. A pioneering use of Landsat TM (Thematic Mapper) images for disease detection was performed to study the infestation of take-all disease (caused by the fungus *Gaeumannomyces graminis var tritici*) in wheat (Chen et

al. 2007), in combination with field data collected using a spectrophotometer (with 3.5nm spectral resolution, ranging from 0.3 to 1.08 $\mu$ m). The authors computed Normalized Difference Vegetation Index (NDVI) for healthy and unhealthy wheat samples and noticed the vegetation index of the latter was lower due the reduction in the chlorophyll; they also found that Landsat band 5 played an important role in differentiating wheat health. Similarly, multispectral remote sensing images were used to detect rice sheath blight in Arkansas (Qin and Zhang 2005). The study employed broadband high spatial resolution Airborne Data Acquisition and Registration (ADAR) images and field observations consisting of disease severity.

The technological development and the advances in sensors recently increased the use of high spatial resolution images for crop monitoring. While these images provide better details than free moderate-resolution images (i.e. Landsat), they also are expensive and not generally accessible to the general public. Franke and Menz (2007) employed high resolution QuickBird (2.4m spatial resolution) and hyper-spectral HyMap (4m spatial resolution at wavelengths ranging from 450nm to 2480nm) data to detect powdery mildew (*Blumeria graminis*) and leaf rust (*Puccinia recondita*) in winter wheat; their method employed the decision tree that used the spectral mixture analysis (SMA) and NDVI to classify different levels of disease severity. The validation included disease severity recorded in the field at different stages and spectroradiometer measurements. The results showed classification results matching the disease severity with respect to the dates of the image acquisition. The aggregated binary map (healthy versus unhealthy) were validated and produced an overall accuracy of 56.8%, 65.9%, and 88.6% respectively for the dates 04/22, 05/28, and 06/20. Nutter et al. (2002) therefore

suggested the combination of multisource remote sensing images such as ground reflectance measurements, aerial and satellites images for disease detection and monitoring.

### 1.3.2. Landsat data fusion for crops time-series analyses

Landsat has remained consistent over the four decades of its different missions, in terms of spatial and temporal resolutions. The spectral improvements of recent Landsat enhanced Thematic Mapper Plus (ETM+) and the Operational Land Imager (OLI) have provided Landsat with more capabilities in vegetation monitoring. However, the temporal resolution (16 days revisiting period) has remained a major limitation, especially when clouds further extend the delays in observations. Several techniques have been developed to overcome these long gaps in Landsat observations, relying on the fusion of high temporal resolution sensors such as the Moderate Resolution Imaging Spectroradiometer-MODIS (but with low spatial resolution) to produce consistent observations over a single growing season, or several years. Several authors have used data fusion techniques to overcome the low temporal resolution of Landsat. Many factors however, need to be considered while choosing the desired fusion algorithms. These factors include the nature of the study area (including the land cover change), the assumptions behind the chosen algorithm, the computation requirements, the quality of the inputs, and the objectives of the study.

The nature of the area to be studied needs to be considered in the choice of the fusion algorithm to be used. The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al. 2006) for instance performs better in homogeneous regions such as forest areas, while it's improved version, the Enhanced STARFM

algorithm (ESTARFM) (Zhu et al. 2010), performs better in heterogeneous regions. However, STARFM is an open source algorithm, and represents an important benchmark that could be explored and further improved. Examples include Spatiotemporal Adaptive Data Fusion Algorithm for Temperature mapping (SADFAT) algorithm (Weng, Fu and Gao 2014), that predicted Land Surface Temperature (LST), and the Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH), which was improved for detecting the land cover change, especially in the forested landscapes (Hilker et al. 2009) because forest cover generally looks homogeneous. A comparison of both STARFM and ESTARFM to two other simple techniques (Linear Interpolation Model - LIM and Global Empirical Image Fusion Model-GEIFM), suggested that ESTARFM is more suitable when and where the land cover and the spectral bands exhibit high spatial variances, and STARFM when there is high temporal variance (Emelyanova et al. 2013).

As land cover change represents an important factor in the choice of the fusion algorithm, several improvements have been made to account for this aspect. Methods such as the linear unmixing (Zurita-Milla, Clevers and Schaepman 2008, Amorós-López et al. 2013) and the unmixing-based data fusion (UBDF) (Zurita-Milla et al. 2009) approaches are not able to capture land cover change, and may not be suitable in studies and areas with high land cover dynamics. Both STARFM and ESTARFM can capture changes at a finer resolution if they are permanent (such as burns) and if they can be detected by MODIS. Yet, other algorithms addressed the land cover temporal change limitations in the fusion techniques by developing methods such as the Spatial Temporal Data Fusion Approach (STDFA) (Wu et al. 2012), which was further improved by the Improved Spatial Temporal Data Fusion Approach (ISTDFA)(Wu et al. 2016). These two

algorithms were able to capture the land cover change if these changes could be observed in the input MODIS images. This condition is similar to the requirements for change detection with STARFM and ESTARFM. Other advanced methods considering the land cover change include the Unmixing-based Spatio-Temporal Reflectance Fusion Model (U-STFM) (Huang and Zhang 2014), which takes into account the phenological and land cover changes, and the sparse representation-based spatiotemporal reflectance fusion model (SPSTFM) (Huang and Song 2012).

The assumption behind the chosen algorithm is very important in the consideration of the fusion technique. A review of different fusion methods (Zhang et al. 2015) allowed their classification as contexture methods, temporal and spatial variance related fusion methods, and non-linear methods such as sparse representation-based spatiotemporal reflectance fusion model (SPSTFM). Contexture methods depend on existing maps, including land cover maps resulting from images classifications (ISODATA or K-means algorithms). This category includes unmixing methods (Zurita-Milla et al. 2008, Amorós-López et al. 2013) algorithms such as UBDF (Zurita-Milla et al. 2009) or STRUM (Gevaert and García-Haro 2015). Temporal and spatial variance methods include algorithms such as STARFM, ESTARFM, STDFA, ISTDFA and U-STFM. Some of these methods assume that the temporal variation of land cover is constant (STDFA and ISTDFA), while others assume that the land cover change is linear (STARFM and ESTARFM). Advanced methods, instead include the notion of change-ratio (U-STFM and SPSTFM), which consider both phenological and land cover changes in the fusion.

Computation (and automation) requirements can sometimes be a limiting factor, especially when fusion methods use image-pairs as inputs. STARFM has an advantage of

low computation because it can work with a single pair of inputs (Landsat and MODIS) while ESTARFM has a reduced automation (by manually setting the size of the search window and the number of classes), and an intense computation (because it requires at least two pairs of fine-coarse images for the same data). U-STFM requires significantly less time than ESTARFM or STARFM; however, it requires extra-steps in the pre-processing including image segmentation and an unmixing step, which might be a disadvantage. The Spatial and Temporal Reflectance Unmixing Model (STRUM) (Gevaert and García-Haro 2015) has a reduced computational time and is less dependent on the number of fine resolution images as inputs which is an improvement if compared to STARFM.

The quality of the inputs depends mostly on the types of the study, but data being fused need to be calibrated in order to be comparable. While some fusion techniques used individual bands, other used basic generated products such as LST or advanced generated products such as NDVI. In any case, the pre-processing steps should always include the geometric correction in order to ensure that analyzed pixels align correctly and to reduce the biases in the method, atmospheric and radiometric corrections in order to reduce bands reflectance variability. The temporal characteristics of the input images constitutes another important factor in several aspects: selecting images at key phenological phases (Greenup, maturity, or senescence) as inputs is more likely to capture the changes and accurately reconstitute the growing season; also, when the analyses are extended over several years, data collected by the same dates allow better comparisons and consistent data predictions.

While choosing the appropriate fusion algorithm, the objectives of the study are a non-negligible factor. While most of the methods have been developed to provide consistent time-series images for land cover monitoring, some methods might have very specific objectives. An example includes an algorithm that was primarily developed to overcome the Landsat ETM+ scan line corrector failure but can also be used to predict cloud and shadow pixels (Roy et al. 2008). This method developed an approach that uses the MODIS Bi-directional Reflectance Distribution Function (BRDF)/ Albedo and Landsat ETM+ to predict reflectance at an ETM+ scale for the same date and the dates before and after the observation. Other methods have been developed as improvements of existing algorithms in order to address specific limitations such as the land cover change (STDFA, U-STFM), or the computation requirements (STRUM). More recently, a Flexible Spatiotemporal Data Fusion (FSDAF), was developed based on spectral unmixing analysis and a thin plate spline interpolator (Zhu et al. 2016); the method aimed to provide a more accurate prediction of fine-resolution images in heterogeneous areas by capturing all land cover type changes (gradual and fast) while requiring minimal input data (only one fine resolution input image).

Considering the aspects mentioned above and the fact that ESTARFM usually outperforms STARFM (Li et al. 2017), we selected ESTARFM as the fusion algorithm to fill the gap in the images to be analyzed over our study area. In fact, the scarcity of consistent Landsat observations prompted us to consider a fusion of the available Landsat 8 images with the daily MODIS MDC43A4 Version 6 products. Resulting images are expected to inherit the 30m spatial resolution of Landsat at a daily frequency in order to

allow an accurate analysis of white mold by capturing its spatial and temporal characteristics.

#### 1.4. RESEARCH GOAL, OBJECTIVES AND HYPOTHESES

The overall goal of this research is to improve our understanding of the white mold occurrence through the exploration of satellite images. This goal is pursued through the following three objectives:

- 1) Quantify the occurrence of white mold in several counties of South Dakota through the use of three remotely sensed Landsat 8 images;
- 2) Investigate the spatiotemporal characteristics of white mold by exploring its spatial distribution in soybean fields and its impact on yield;
- 3) Investigate the temporal characteristics of white mold by modeling the white mold as a function of the vegetation index through time.

To achieve these three goals in this research, the following three hypotheses have been developed and addressed:

**Hypothesis #1:** Based on field knowledge and available Landsat images, the occurrence of white mold can be accurately mapped and quantified.

**Hypothesis #2:** Spatiotemporal characteristics of yield in white mold infected soybean fields can be modeled using NDVI computed from a fusion of Landsat and MODIS images, and the relationship can provide estimates of yield loss caused by white mold.

**Hypothesis #3:** White mold temporal characteristics can be modeled as a function of the Normalized Difference Vegetation Index obtained by a fusion of Landsat and MODIS.

### 1.5. SIGNIFICANCE OF THE STUDY

Despite the scientific attention given to white mold of soybean, there is still a need to investigate the contribution of satellite imagery in better understanding the spatial and temporal characteristics of the disease. Thanks to the technological advances, remote sensing can be used to collect large scale data that can be modeled to provide good insight on the crops conditions and a better understanding of the disease development. While remote sensing is advantageous in terms of labor and time involved, it also presents some limitations in terms of the quality of the input data, the timing of data acquisition, and sometimes the level of details needed to better interpret the investigated phenomena. Field scouting has proven to be an accurate disease identification method, but the labor does not allow timely and accurate identification of the disease over large areas.

This study provides improvement in the understanding of the spatiotemporal characteristics of white mold and also sets a benchmark in using remote sensing images to understand other threats that might affect soybean production. First, the mapping and quantification of white mold using free remote sensing images provide accurate estimates of the impacts of the diseases when compared to the damage as reported by USDA. However, these results can still be improved by the use of higher spatial and temporal resolution images such as SENTINEL-2, in the optics of capturing subtle patches that might have not been observed in the medium Landsat images. Second, the assessment of spatiotemporal characteristics of white mold development over time using NDVI allows to understand the critical period for predicting yield and estimating yield loss, as well as

the spatial development of white mold in a field. This information is beneficial to farmers as it may allow them to take appropriate measures in the mid-season to prevent yield loss such as spraying fungicides based on previous white mold extent in their locations. The disease extent estimation method can be improved by incorporating timely and accurate images such as those provided by commercial drones, in case satellite images gaps are hard to fill. Finally, modeling the temporal evolution of white mold as a function of the Normalized Difference Vegetation Index (NDVI) allows to assess the importance of non-destructive methods in analyzing the disease occurrence. This study highlights the importance of the time component expressed as the Day of the Year in modeling the temporal development of the white mold. Furthermore, the comparison of the NDVI temporal change between different sites show how the disease can be influenced by local parameters that are important to capture differences from one location to another. The models developed in this study, are limited in space and time, and might not necessarily apply to regions with different environmental characteristics; however, they constitute an important source of information for the development of future models in the regional and global scales, provided ground truth data are accessible.

## 1.6. ORGANIZATION OF THE DISSERTATION

This dissertation contains five chapters. Chapter 1 (Literature review) provides generalities on the soybeans, white mold, information on the use of Remote Sensing for crop monitoring and disease detection, information on the data fusion algorithms, and frames the “aim, objectives, and hypotheses” that this research pursued and addressed. Chapter 2 through Chapter 4 separately addresses three hypotheses enumerated above. Finally, Chapter 5 concludes the whole research.

Chapter 2 addresses **hypothesis #1**: The use of free available Landsat images to map and quantify the occurrence of white mold in three northeastern counties of South Dakota; it uses the random forest algorithm to map the Land Cover and white mold occurrence.

Chapter 3 addresses **hypothesis #2**: Spatiotemporal characteristics of yield in white mold infected soybean fields can be modeled using NDVI computed from a fusion of Landsat and MODIS images. It also examines the spatial distribution of yield in a soybean field infected with white mold, and investigates the change in the relationship between yield and NDVI.

Chapter 4 addresses **hypothesis #3**. White mold temporal characteristics can be modeled as a function of the Normalized Difference Vegetation Index obtained from fusion of Landsat and MODIS. It models the temporal evolution of soybean as a function of weather parameters, in order to investigate factors susceptible to explain white mold occurrence.

Chapter 5 presents major conclusions of the study. It summarizes the key findings in the tests of the three hypotheses and relates them back to the aim and specific objectives. It also discusses future research directions and recommendations for improving the key findings.

## CHAPTER 2: MAPPING AND QUANTIFYING WHITE MOLD IN SOYBEAN ACROSS SOUTH DAKOTA USING LANDSAT IMAGES

Mfuka, C., Zhang, X. and Byamukama, E. (2019) Mapping and Quantifying White Mold in Soybean across South Dakota Using Landsat Images. *Journal of Geographic Information System*, 11, 331-346. <https://doi.org/10.4236/jgis.2019.113020>

This chapter addresses **Hypothesis #1**: Based on field knowledge and available Landsat images, the occurrence of white mold can be accurately mapped and quantified.

### 2.1. ABSTRACT

White Mold of soybeans (*Glycine Max*), also known as *Sclerotinia stem rot* (*Sclerotinia sclerotiorum*), is among the most important fungal diseases that affect soybean yield and represents a recurring annual threat to soybean production in South Dakota. Accurate quantification of white mold in soybean would help understand white mold impact on production; however, this remains a challenge due to a lack of appropriate data at a county and state scales. This study used Landsat images in combination with field-based observations to detect and quantify white mold in the northeastern part of South Dakota. The Random Forest (RF) algorithm was used to classify the soybean and the occurrence of white mold from Landsat images. Results show an estimate of 132 km<sup>2</sup>, 88 km<sup>2</sup>, and 190 km<sup>2</sup> of white mold extent, representing 31%, 22% and 29% of the total soybean area for Marshall, Codington and Day counties, respectively, in 2017. Compared with ground observations, it was found that soybean and

white mold in soybean fields were respectively classified with an overall accuracy of 95% and 99%. These results highlight the utility of freely available remotely sensed satellite images such as Landsat 8 images in estimating diseased crop extents, and suggest that further exploration of consistent high spatial resolution images such as Sentinel, and Rapid-Eye during the growing season will provide more details in the quantification of the diseased soybean.

Keywords: Random Forest, Landsat, White Mold, Soybean

## 2.2. INTRODUCTION

White mold of soybeans (*Glycine Max*), also known as “*Sclerotinia Stem Rot*” (SSR), is among the most important fungal diseases affecting soybean yields and represents a recurring annual threat to soybean production in South Dakota. Initially reported in Poland in 1982 as a disease of local importance (Marcinkowska, Tomala-Bednarek and Schollenberger 1982), white mold was, more than a decade later, ranked in the top ten diseases that suppress soybean yields (Wrather and Koenning 2009). The apothecia (fungal structures that produce inoculum of the pathogen) of white mold generally appear after the crop canopy develops, around mid to late July and the environmental conditions corresponding to the development of white mold are cool (air temperature around 12-24 °C), wet and moist (enough rain: 70-120 hours of continuous wetness) conditions (Boland and Hall 1988). These conditions are favorable for optimal yield. Therefore, incidence of white mold has been negatively correlated with yields

(Hoffman et al. 1998) because the disease is more likely to develop where there is high yield potential. Thus, mapping and quantifying the disease is crucial to understand its impact on yields, and two options can be used: field scouting represents an accurate assessment, but remains time-consuming and does not provide a global view of the variations in the field, while remote sensing represents the best solution because it provides a synoptic view and allows observations to span large areas in a short period (Lowe, Harrison and French 2017).

The rationale behind the use of large scale imagery techniques is that they represent a fast, non-destructive method (Yeh et al. 2013), and rely on biophysical characteristics that depend on the wavelength used for crop status monitoring. Malthus and Madeira (1993) highlighted the interest of using image to detect crop diseases by examining the spectral leaf reflectance properties of field bean infected by the fungus *Botrytis fabae*. Later, Polischuk et al. (1997) studied the correlation between chlorophyll content and spectral reflectance in virus affected plants. In the 2000s, several authors explored diverse options for disease detection: Kobayashi et al. (2001) used multispectral radiometers and airborne multispectral scanner to identify the panicle blast rice. Qin and Zhang (2005) collected ADAR (Airborne Data Acquisition and Registration) remote sensing images to map rice sheath blight. Further, Huang and Apan (2006) used a portable spectroradiometer to collect hyperspectral data and detect *Sclerotinia* rot disease in celery. Naidu et al. (2009) later identified grapevines viral infections by using the leaf spectral reflectance collected with a portable spectrometer. The use of hyperspectral images is necessary to characterize plant stress (Behmann, Steinrucken and Plumer 2014, Vigier, Pattey and Strachan 2004) and spectral indices are crucial in detecting and

identifying plant diseases (Chappelle, Kim and McMurtrey 1992, Mahlein et al. 2013, Vigier et al. 2004) . However, most of these studies required the use of portable spectroradiometer or airborne remotely-sensed images, which represent costly resources and have a reduced accessibility to common users and farmers.

While vegetation stress has received a lot of scientific attention, soybeans stress mapping has received little consideration, and when it has, these studies focused either on other diseases than white mold (Bajwa, Rupe and Mason 2017), or in water stress (Behmann et al. 2014, Thompson and Wehmanen 1980). Vigier et al. (2004) used hyperspectral reflectance to compute several vegetation indices to detect white mold, but the study focused on inoculated disease, rather than in-situ observation, and reflectance was collected using a field spectrometer. Recent studies have focused on mapping soybean at national scale (Song et al. 2017, King et al. 2017) , but these efforts have not addressed disease detection. In South Dakota which is one of the main soybean producing states in the US, no studies have been conducted for the quantification of soybean diseases, especially white mold using remote-sensing approaches.

There is still a knowledge gap in the effectiveness of free of charge moderate-resolution remotely sensed images such as Landsat in accurately mapping crop diseases, especially the occurrence and evolution of white mold in the Midwest. The current study employs free Landsat 8 images to map and quantify white mold in selected counties in South Dakota. Random forest (RF) classifiers (Breiman 2001) were used to extract spectral characteristics of soybean and white mold leading to mapping the spatial extent of the disease.

## 2.3. MATERIALS AND METHODS

### 2.3.1. Study area

The study was located in northeastern South Dakota and includes three counties: Marshall, Day and Codington. Soybeans are planted in South Dakota between May 8 and June 21, with the most active period between May 15- June 11 (USDA 2010). The harvest occurs between September 22 and November 3, with the most active period between September 28 and October 24. Field data consisted of scouting and reporting on the presence/absence of white mold during the months of July and August in the year 2017. In the study area, a total of 11 fields were scouted, where white mold was reported and confirmed as shown in Figure 2.1.

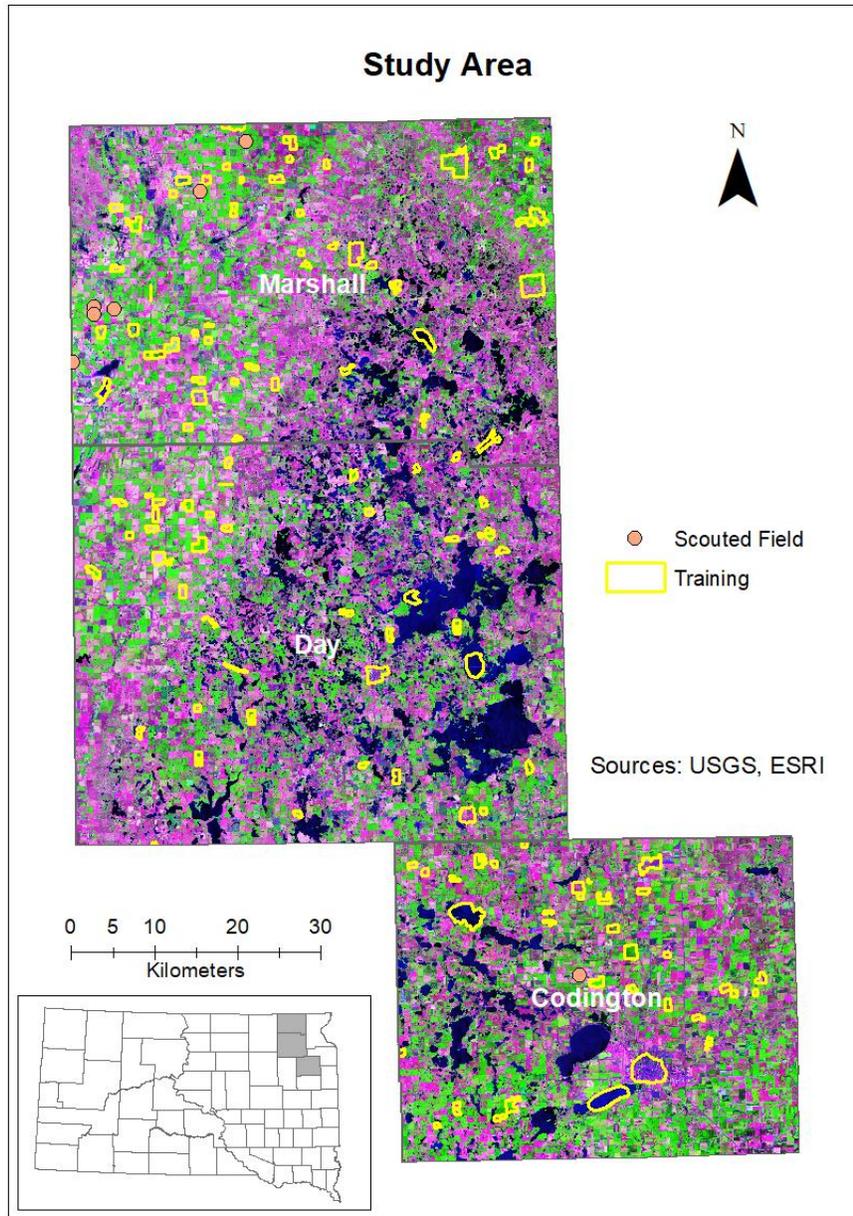


Figure 2. 1. Study area showing the three counties (Marshall, Day, and Codington) in Northeastern South-Dakota. The background image is a Landsat false color combination of bands 6-5-4.

### 2.3.2. Data gathering

We downloaded the 30-meter spatial resolution Landsat Analysis Ready Data (ARD) from Earth Explorer (<https://earthexplorer.usgs.gov/>) for the growing season of the year 2017, and covering the three counties in the northeastern South Dakota (Marshall, Day, and Codington counties) as shown in Figure 2.1. These Cloud-free images were respectively from May 11, July 14, and August 31 and were derived from Landsat Collection 1 Level-1 precision and terrain-corrected scenes consisting of Top-of-Atmosphere (TOA) Reflectance, Surface Reflectance (SR), Brightness temperature (BT) and Quality Assessment (QA). In our study, the products of interest consisted of SR and the selected bands are summarized in Table 2.1. Yet, Landsat images were particularly hard to obtain during the growing season, due to persistent clouds that often extend the 16-day revisiting period of Landsat. This situation allowed to collect only two Landsat images (May and July) for soybean classification and one image (August) for white mold mapping.

The Crop Data Layer (CDL) is a land cover dataset developed by the National Agricultural Statistics Services (NASS) of the United States Department of Agriculture (USDA). This dataset can be used to extract soybean masks or other land cover of interest; however, the timing in the publication of CDL might not always match the needs to map the land cover within the growing season. The CDL is generally produced early in the year, for the land cover map of the previous year. We used CDL as a reference data in our study, guiding the trainings for land cover mapping. This data also served in the comparison with our land resulting cover map.

### 2.3.3. Random Forest classifiers for mapping soybean and white mold

#### a) The Random Forest (RF) algorithm

Methods that produce classifiers and aggregate their results have recently found many interests in the machine learning field (Liaw and Wiener 2002). The underlying principle is the same: based on a set of trainings used to extract spectral characteristics of different defined classes, these classifiers build models that decide to which class to affect each observation. The methods are called non-parametric classifiers, meaning that they require no statistical assumptions such as the normal distribution of the input dataset. Among them are methods such as boosting, that use successive trees to assign extra weight to samples that have been incorrectly predicted by earlier predictors (Schapire et al. 1998), and bagging, in which successive trees are independent from earlier trees (Breiman 1996). In the end of the prediction process, a weighted vote is taken in the boosting while a simple majority vote is taken in the bagging (Liaw and Wiener 2002).

The RF algorithm (Breiman 2001) is one of the learning methods that adds an additional layer of randomness to the bagging: each node is split using the best among a subset of predictors randomly chosen at that node, which is different from standard trees (i.e Decision Tree-DT), where each node is split using the best split among all variables (Liaw and Wiener 2002). In the remote sensing field, especially in image or land cover classification, RF has shown to perform equally to Support Vector Machine (SVM) (Pal 2007) (Thanh Noi and Kappas 2017) or to outperform Decision Tree (DT) (Rodriguez-Galiano et al. 2012). Other studies have shown that RF outperformed SVM in terms of robustness and stability (Rodriguez-Galiano et al. 2015) and in terms of accuracy (Adam et al. 2014). The RF is preferred in our study because it can deal with classification

problems of unbalanced, multiclass and small sample data (Liu et al. 2013). In fact, when collecting training data, some classes may require more training than other in order to capture the maximum variability in their spectral differences. This type of data collection can be dealt with by RF which does not require further processing.

b) Soybean mapping and validation.

To classify land cover, we collected a set of trainings (about 183,810 pixels) used to extract spectral characteristics of different classes in ArcMap. We particularly trained four classes namely: Water, Corn, Soybean and Other Land Cover (OtherLC). To guide the trainings, three types of information could be displayed to better interpret the land cover in digitizing the training polygons: (1) Landsat-8 composites, (2) Crop Data Layer (CDL) serving as a cross-reference, (3) and high resolution Google Earth images. The quality of the training samples was evaluated using the Jeffries-Matusita's (JM) spectral separability index, which provides a good mean of estimating the difference between the classes (Bruzzone, Roli and and Serpico 1995, Ifarraguerri and Prairie 2004). This index is a measure of statistical separability for two-class cases based on distance, and can be extended in the separability of multiple classes. The JM distance between classes  $\omega_i$  and  $\omega_j$  is formulated as shown in Equation 1. In general, a JM of greater than 1.9 represents a good difference, while JM of less than 1 implies a combination of the classes (no difference); a JM between 1 and 1.8 generally suggests improvement of training classes. The JM index was computed in ENVI.

$$J_{ij} = \int \left\{ \sqrt{p(x|\omega_i)} - \sqrt{p(x|\omega_j)} \right\}^2 dx \quad (1)$$

Where  $x$  is the feature vector of dimension  $k$  and  $p(x|\omega_i)$  and  $p(x|\omega_j)$  are class conditional probability distributions of  $x$ .

The training polygons were imported in R, and seventy percent of the pixels (128,667) were used to build the model while thirty percent (55,143) were used for validation. The two early images (May and July) bands were stacked using ENVI 5.0, and the resulting stacked image was classified using the RF algorithm in R. The ten Landsat bands (Table 2.1) were used as independent variables, while the land cover (four classes) to predict represented the response variable. The soybean mask was extracted from the resulting land cover classification map. The set-apart thirty percent of the samples were used to assess the accuracy of the land cover map. A confusion matrix was built to assess the accuracy of each class as well as the overall accuracy, and to estimate the classification errors.

Table 2. 1. Original Landsat 8 bands including the Shortwave Infrared (SWIR), the Near Infrared (NIR), the red (RED), the green (GREEN) and the blue (BLUE) bands, and their corresponding names used in the Random Forest (RF) classification, and in the stacked image.

Image	Original Band	RF Name	Stacked Band
Landsat-8 May 11, 2017	SWIR 1	B6_05	Band 1
	NIR	B5_05	Band 2
	RED	B4_05	Band 3
	GREEN	B3_05	Band 4
	BLUE	B2_05	Band 5

Landsat-8 July 14, 2017	SWIR 1	B6_07	Band 6
	NIR	B5_07	Band 7
	RED	B4_07	Band 8
	GREEN	B3_07	Band 9
	BLUE	B2_07	Band 10

c) White mold mapping, validation, and areas estimates

The August 31 Landsat image was used to evaluate soybeans health and to characterize white mold. Field locations of well-known white mold occurrence were used to extract the spectral characteristics of white mold using the computed Normalized Difference Vegetation Index -NDVI (Rouse et al. 1973) from the same image. NDVI is a measure of the vegetation health and greenness, computed as the ratio between the difference and the sum of the Near Infrared (NIR) band and the Red band, which respectively represent the regions of high chlorophyll absorption and reflectance (Equation 2). Locations presenting similar NDVI than the known fields were targeted to train the data for modeling; a total of 3981 pixels were collected in the trainings. Classes consisted of white mold (unhealthy) and other soybean (healthy), representing the response variables, while the explanatory variables consisted of the 5 individual Landsat bands and the NDVI. To maximize the accuracy of white mold detection and reduce the false positive, all pixels with low NDVI that do not correspond to white mold were excluded from the soybean mask. In fact, soybean disturbances occurring in July are not white mold because at this stage, there is not yet canopy closure. While healthy soybean

in mid-July has and expected NDVI around 0.5, all pixels with NDVI lower than 0.45 within the soybean mask were excluded.

The RF algorithm was run on the soybean mask extracted from the LC classification; as with the land cover, seventy percent (2787 pixels) of the total sample pixels were used to build the model while thirty percent (1194 pixels) were used for accuracy assessment. To assess the accuracy of the results, the set-apart thirty percent of the samples were used to produce the confusion matrix, estimate the individual classes errors and the overall map accuracy. The resulting mapped white mold pixels were used to estimate areas by using the pixel counts and pixel size as it pertains to Landsat (Equation 3).

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (2)$$

$$TA = N \times A \quad (3).$$

Where TA is the Total Area, N is the number of pixels, and A is the area of one pixel (30mx30m)

## 2.4. RESULTS AND DISCUSSION

### 2.4.1. Land cover spectral separability

The performance of the trainings was assessed using the computed Jeffries-Matusita index, which assesses the classes' spectral separability. Overall, all the classes exhibited good spectral separability ( $JM > 1.9$ ) while the pair soybean/corn exhibits the lowest index (1.86) and water showing the highest separability ( $JM = 2$ ). Table 2.2. provides

different values of JM index between classes as trained for the Landsat bands in the northern part of the study area.

Table 2. 2. Jeffries-Matusita (JM) spectral separability index showing the goodness of the trainings

Classes	Water	Corn	Soybean
OtherLC	1.9998	1.9419	1.9681
Water		2	2
Corn			1.8335

The original input Landsat bands were stacked in a color composite image combining both May and July bands. The corresponding output bands designations are listed in Table 2.1. Figure 2.2 provides a visual display of each band's ability to discriminate individual classes. Both NIR and SWIR bands in May and July separated water successfully; corn tended to stand out particularly in July using the visible bands (Blue, Green, and Red), while soybean (areas where soybean will grow) was distinguished in the visible bands in May. In fact, soybean is not visible in the fields at this period, but their areas can be distinguished with corn. The "OtherLC" class looks particularly difficult to extract because of the high variability of the land covers included (grass, pasture, other crops).

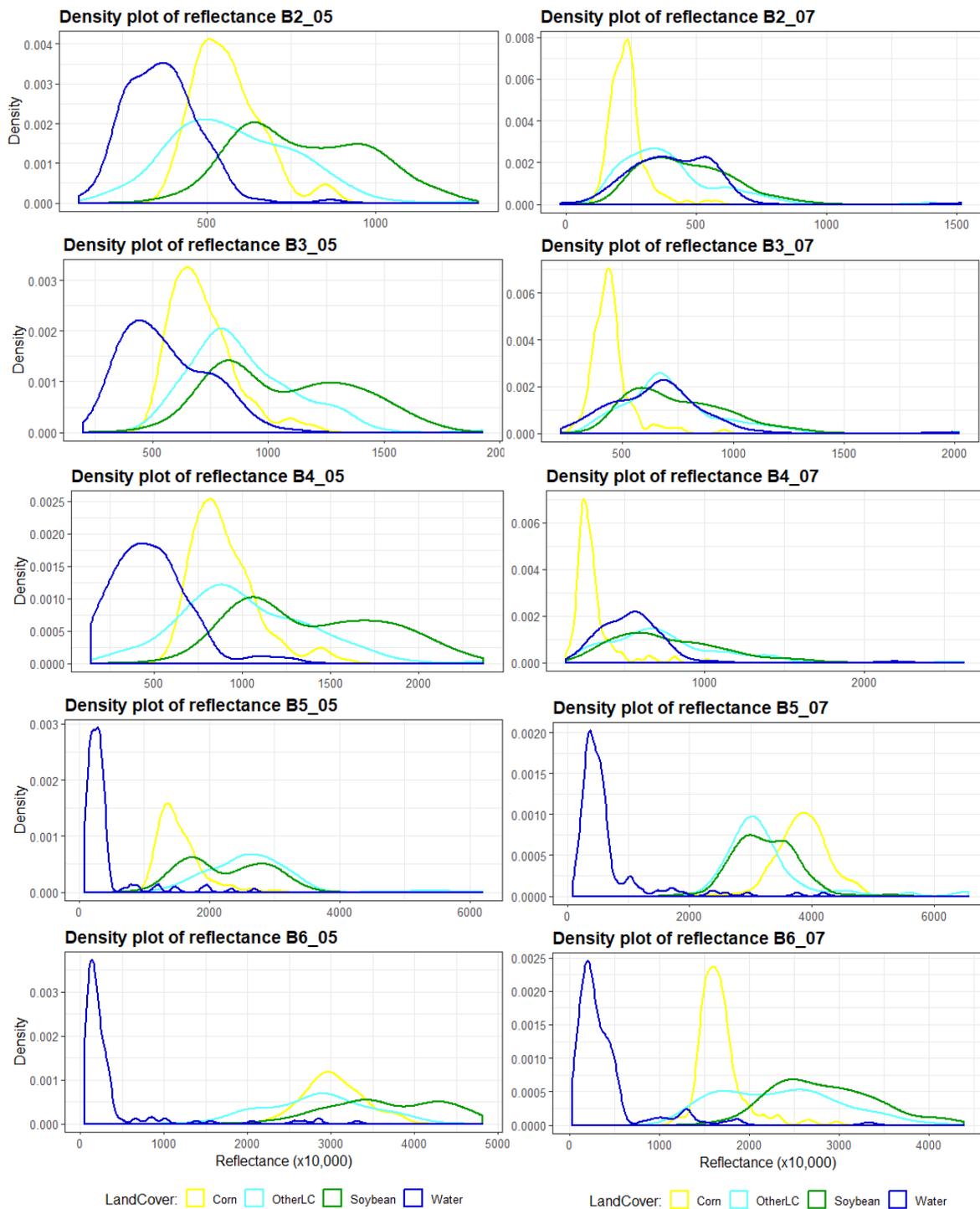


Figure 2. 2. Class spectral separability: each plot shows the ability and the contribution of each band in separating land cover classes.

#### 2.4.2. Land cover classification results

The stacked May and July images were classified using the RF algorithm and the land cover map was generated using the R software. The four classes (Water, Corn, OtherLC and Soybean) were labeled and colored to match the Crop Data Layer (CDL) dataset. Figure 2.3 shows a comparison between the July False color (6-5-4) Landsat composite, the CDL and the classified images. Water (Upper-right) is in some cases classified as other land cover, especially when it corresponds to swamps as mapped by CDL. Overall, the classified image is close to the CDL but reflects more what is observed in the composite Landsat image, especially the field roads in-between soybean fields that are excluded from the classified map, thus excluding the false positive when mapping the disease. The rationale behind computing a land cover map instead of using existing datasets such as the CDL is the timing: The release date of the CDL for a given year occurs early the following year, while the estimate the disease extent may be needed earlier than that. However, extracting the mask of interest from CDL is a good alternative provided it is released on time.

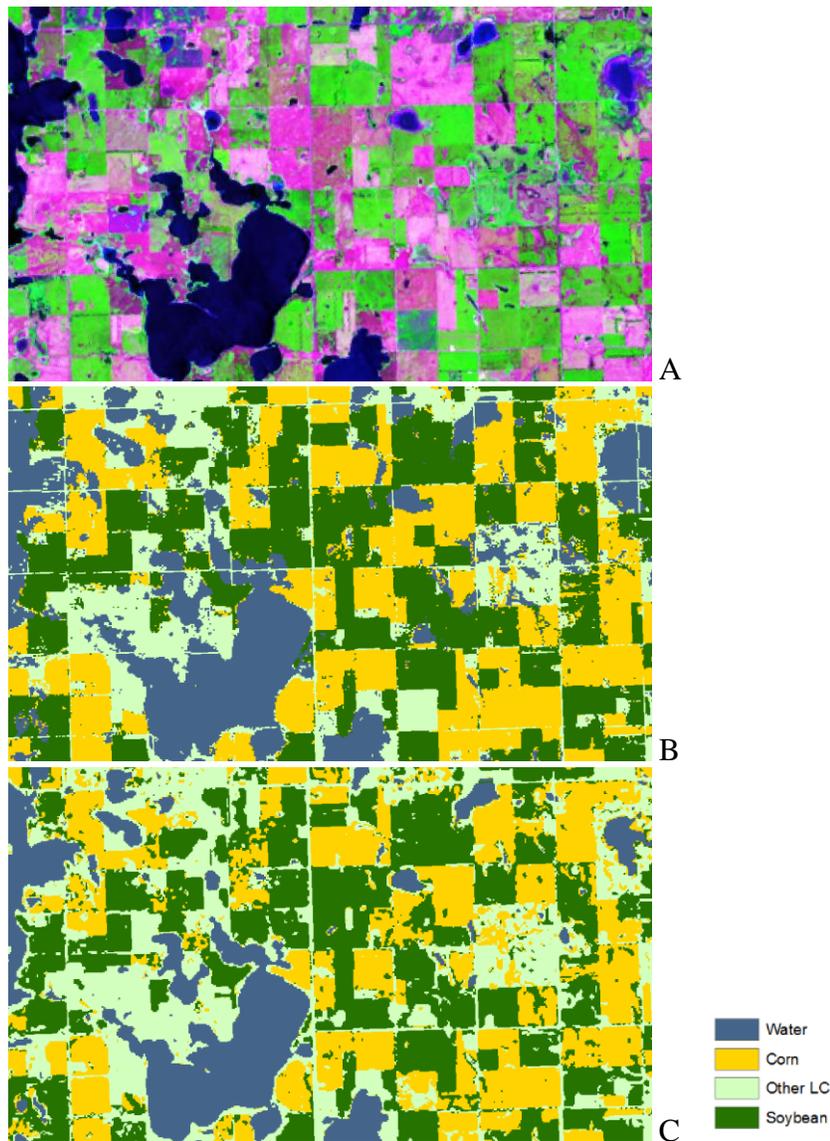


Figure 2. 3. A comparison between a July 14 false color 6-5-4 Landsat 8 image (A), the Crop Data Layer (CDL) map (B) and the resulting classification (C) of the stacked May and July images. showing similarities between the resulting classification and the CDL.

#### 2.4.3. Land cover map accuracy assessment

The accuracy of the resulting classification map was assessed using the confusion matrix (Table 2.3), with the 30 % set-apart pixels that were not used in the RF

classification process. The classification results achieved an overall accuracy of 95%. The “Water” class performed the best (98% accuracy) while “Corn” performed the least (91% accuracy); OtherLC was classified with 97% accuracy while soybean achieved an accuracy of 94%. Table 2.3 reports the individual class accuracies as well as the errors. The commission and omission errors are reported in Table 2.3 as well. Soybean is accurately classified with a 94% producer’s (meaning that approximately 94% of the soybean ground truth pixels also appear as soybean pixels in the classified image) and 93% user’s accuracy (meaning that 93% of the soybean pixels in the image actually represent soybean in the ground).

Table 2. 3. Confusion matrix of the land cover map accuracy assessment

Classification	Ground Truth				Com. Err.	Prod. Acc.
	Water	Corn	OtherLC	Soybean		
Water	459	8	0	0	0.02	0.98
OtherLC	5	1042	10	27	0.03	0.97
Corn	1	23	356	13	0.09	0.91
Soybean	0	24	8	524	0.06	0.94
Om. Err.	0.01	0.05	0.05	0.07		
User’s Acc.	0.99	0.95	0.95	0.93		

The accuracy of the mapped land cover classes is reported, with the Commission Error (Com. Err.) and the corresponding Producer’s Accuracy (Prod. Acc.): Water has the

highest accuracy while corn has the lowest, with soybean being 94% accurate. The omission error (Om. Err.) and the corresponding User's accuracy (User's Acc) are reported as well, with water being the most accurate (99%) while soybean has a 93% accuracy.

#### 2.4.4. White mold mapping

Figure 2.4 shows the computed NDVI (B) on the August Landsat image (A), and the resulting mapped soybean and white mold (C). In late August, the soybean crops are mature and therefore the vegetation index is high. The detected white mold NDVI ranges between 28%-78% while the healthy soybean exhibits a high NDVI of more than 79%.

Some unhealthy areas can also be detected with very low NDVI values, corresponding to early soybean damages that are not white mold. However, these cases represent sparse and isolated pixels and were not included in the training. Despite the efforts to accurately detect white mold, some other disturbances can also present similar spectral index, especially since the white mold mapping is only using one image. Including several images in the white mold mapping would allow exclusion of disturbances that have the same index with white mold while representing something else. Information on the timing of white mold is crucial in excluding such disturbances in the presence of several images. Yet, unplanned disturbances such as drought or hail damages would not exhibit similar spatial patterns as white mold in the field, and can therefore be distinguished from the mapped disease.



Table 2. 4. White mold accuracy assessment: Confusion matrix table comparing the mapped classes with ground truth.

Class	Ground Truth			
	Healthy	Unhealthy	Commission Error	Producer's Accuracy.
Healthy	1834	22	0.01	0.99
Unhealthy	1	643	0.01	0.99
Omission Error	0.01	0.03		
User's Accuracy.	0.99	0.97		

#### 2.4.6. Quantified soybean and white mold

Using the Landsat pixel size (30mx30m), we estimated the total area of the classified soybean in the three counties based on the total number of pixels mapped. Table 2.5 reports the total soybean areas estimation from both the classification and the USDA report (USDA 2017), as well as the estimated white mold areas per county. The USDA estimated areas consist of the harvested statistics, but the values are very similar to those obtained by the classified Landsat images. The diseased soybean area estimates are respectively 132 km<sup>2</sup>, 88 km<sup>2</sup>, and 190 km<sup>2</sup>, and represent 31%, 22% and 29% of the total soybean area for Marshall, Codington and Day counties.

Table 2. 5. Comparison between soybean area estimates from the United States Department of Agriculture (USDA) and the classified map in this study, as well as white mold extent estimated for each county, based on the calculations from the Landsat pixel size (30mx30m) and the total number of pixels.

County Name	County Area (km <sup>2</sup> )	Soybean (km <sup>2</sup> )	Soybean (USDA- km <sup>2</sup> )	White Mold (km <sup>2</sup> )
Marshall	2294.16	426.592	427.348	131.9589
Codington	1856.90	403.700	439.893	88.3692
Day	2824.50	653.220	660.042	189.9252

## 2.5.CONCLUSION

This study demonstrated that free of charge remotely sensed images could be used to detect and quantify white mold. The RF algorithm used was efficient in mapping the land cover and detecting white mold as reflected in the accuracy assessment. To improve the accuracy in the disease detection, this study combined both Landsat individual bands and NDVI. Including NDVI in the model provides more information, especially since the index puts together the strengths of the NIR band and the Red band.

A good knowledge of the investigated fields is necessary to complement images processing and ensure a proper validation. Constraints such as the images availability, or the timing of the disease should be addressed carefully in mapping the disease. To improve the classification results, more images can be obtained by the fusion of medium spatial resolution Landsat (30 m, 16 days) with high temporal resolution Moderate Imaging Spectroradiometer –MODIS (500 m, 1 day) for instance. Disease extents may be underestimated because of the Landsat pixel size that may not capture small patches of the disease. The use of satellite images with short revisiting period and a higher spatial resolution such as Sentinel-2 (10 m, 5 days revisiting period) or daily Rapid-eye may

provide a better way of quantifying the disease, but the extent or the coverage might require many scenes according to the size of the study area.

The disease rating might also represent an important factor in mapping the occurrence of white mold, as according to the latitude and the difference in the planting dates for instance, some phenological differences might be observed in the signal of white mold. The disease severity can help accounts for these differences while mapping the crop stress, which may result to a better disease quantification.

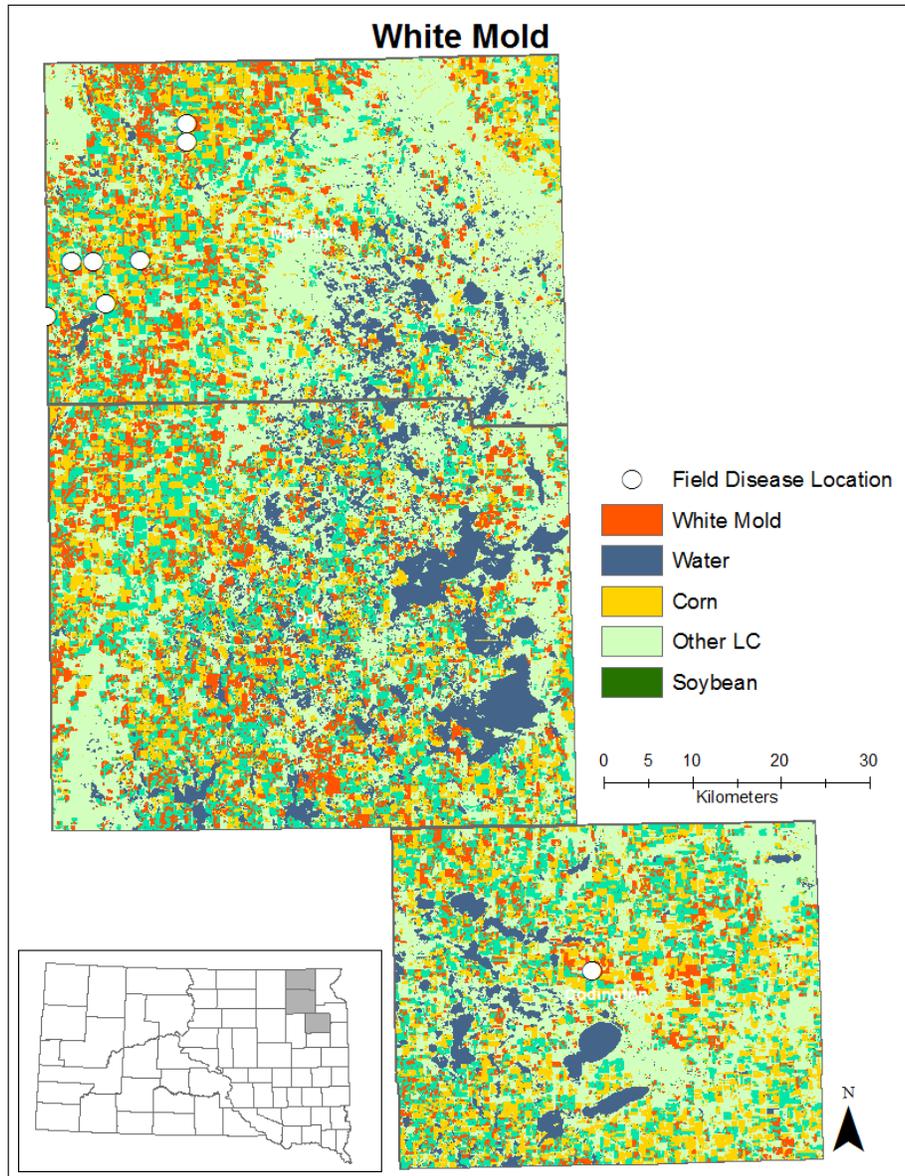


Figure 2. 5. White mold in northeastern South Dakota: the map shows a classified image in background with the four important classes and the quantified white mold over the soybean mask. The white circles indicate ground trothed white mold fields.

### CHAPTER 3: SPATIOTEMPORAL CHARACTERISTICS OF WHITE MOLD AND IMPACTS ON YIELD IN SOYBEAN FIELDS IN SOUTH DAKOTA

Mfuka, C., Byamukama, E. and Zhang, X. (2019) Spatiotemporal characteristics of white mold and impacts on yield in soybean fields in South Dakota. *Geo-Spatial Information Science*. (Paper accepted and under revision).

This chapter addresses **Hypothesis #2**: Spatiotemporal characteristics of soybean yield in white mold infected soybean fields can be modeled using NDVI computed from a fusion of Landsat and MODIS images, and the relationship can provide estimates of yield loss caused by white mold.

### 3.1. 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 in August 21st and September 8th for 2016 field and 2017 field, respectively. This study shows the potential of free remotely-sensed data in predicting yield and estimating yield loss caused by white mold.

Keywords: Soybean, White Mold, Landsat, MODIS, Fusion, Kriging, NDVI, Time-series

### 3.2. 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 ten 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 et al. 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 soybean white mold and its influence on 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 et al. 2017, Behmann et al. 2014, Thompson and Wehmanen 1980), or employed expensive methods for collecting spatial data, such as hand-held spectrometers (Vigier et al. 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, Reay-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 physico-chemical properties that are generally collected in plots and require intensive data collection.

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 examined the pattern of yield in white mold infected soybeans fields, and employed a simple 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 and the red bands. Subsequently, the study estimated yield loss caused by white mold and used time-series NDVI surface maps generated by the ordinary kriging (OK) to explore the spatio-temporal patterns of white mold and their influence on yield distribution.

### 3.3. MATERIALS AND METHODS

#### 3.3.1. Study area

The study area consisted of two soybean fields that were monitored and confirmed 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 3.1). The studied fields are located in the eastern and northeastern 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 include corn, soybean, alfalfa, and spring wheat.

### Study Area: Soybean Fields in Marshall and Moody Counties (SD)

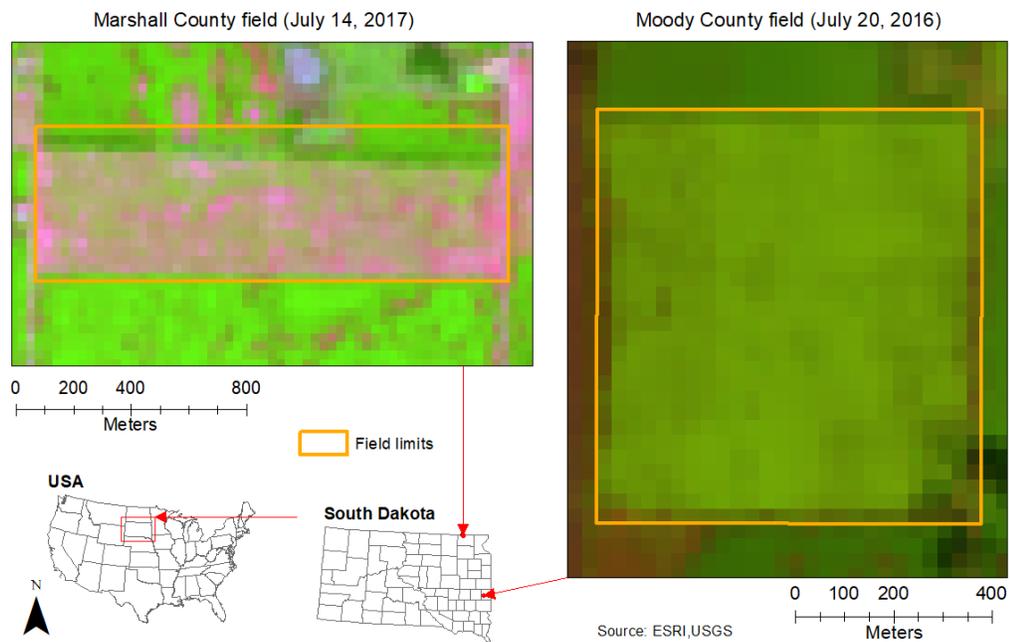


Figure 3. 1. Location of the study area. The scouted fields were located in Marshall and Moody counties, in South Dakota. The background image is a Landsat false color bands 6-5-4 combination from July 14, 2017 (Marshall) and July 20, 2016 (Moody).

### 3.3.2. Data collection and preparation

#### a) Data collection

Soybean yield data were collected by a combine at one second intervals, resulting to a total of 29878 sample points in 2016 (Moody field) and 40216 sample points in 2017 (Marshall field). The harvest in the two fields occurred respectively on October 10, 2016 in Moody, and on October 18, 2017 in Marshall. These data collected at one second intervals resulted into a large point dataset; therefore, the yield data were aggregated to match Landsat pixels because using raw data would produce an unnecessary redundancy that would not improve the spatial relationships. The selected statistical summary of the yield aggregation 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 x 30 m grid on the fields, matching original Landsat pixels and on which the average yield was computed for each cell.

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 May 17, July 20, August 5 and 21, and September 14 and 30. Similarly, we downloaded 4 images for the year 2017 covering the Marshall field, and spanning the dates May 11, July 14, August 31, and September 08.

Daily Moderate Resolution Imaging Spectroradiometer -MODIS images (MDC43A4 Version 6 products) were downloaded from the period May 11 to September 8 of the year 2017. 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 500m to 30m using the nearest neighbor option in ENVI, to match the corresponding Landsat image and ease the fusion of the two datasets.

#### b) Data fusion

To obtain comparable datasets between the two years, we performed a data fusion for 2017 images; this process ensured that we obtain images for similar dates than those collected in the year 2016. High spatial resolution Landsat ARD (30m) do not have the required temporal resolution (16 days) for crop monitoring; meanwhile, high temporal resolution daily MODIS do not have the necessary spatial accuracy 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 et al. 2014, Zurita-Milla et al. 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 within the two dates. ESTARFM was selected because of its advantages: it works better in heterogeneous regions such as the crop lands in our study area, it improves the prediction with the use of several bands (Table 3.1 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).

Table 3. 1. Landsat Operational Land Imager (OLI) and the Moderate Resolution Imaging Spectroradiometer (MODIS) bands designation and wavelengths ( $\mu\text{m}$ ) combined using ESTARFM.

<b>Band Designation</b>	<b>OLI (30 m)</b>		<b>MODIS (500 m)</b>	
	Wavelengths	Band number	Wavelengths	Band number
RED	0.630 – 0.680	4	0.620 – 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

RED is the Red band, NIR is the Near Infrared band and SWIR is the Shortwave Infrared band.

### c) NDVI computation

The original images for 2016 and the resulting fused synthetic images for 2017 were used to compute the NDVI 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.

The computed NDVI images had a 30 m spatial resolution, matching the originally downloaded ARD Landsat images. A regular 30 x 30 m grid was superimposed on the 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.

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad \text{Eq. 1}$$

Where: NIR is the Near Infrared Band and Red is the Red Band

### 3.3.3. Analyses

#### a) Testing yield distribution for spatial autocorrelation:

The spatial distribution of yield was tested for randomness, using the Moran's I autocorrelation test, 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):

$$I = \frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad \text{Eq.2}$$

Where:

$w_{ij}$ : elements of a spatial binary contiguity matrix;

$y_i / y_j$ : variable values at specific locations  $i$  and  $j$

$\bar{y}$  : average of the variable

$n$ : 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 8 pixels around the central pixel are considered neighbors. The spatial neighborhood object was then converted to a weighted matrix, 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. (2018), 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.

b) 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 (Bhunias, 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.

c) Time-series NDVI and yield relationship

A linear regression model was computed in R-Studio (RStudio 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. Non-white mold pixels (low NDVI resulting from diseases other than white mold, based on the time of onset and ground trothing scouting) were not included in the dataset. 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 two years. This allowed to estimate the best period when yield can be predicted using NDVI, as explained by the strength of the relationship.

d) Impacts 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 non-white 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.

$$YL (\%) = \frac{(Y_{max} - Y_{avg})}{Y_{max}} * 100 \quad \text{Eq. 3}$$

Where YL is the yield loss; Ymax is the maximum yield and Yavg is the average yield for each range (class).

### 3.4. RESULTS AND DISCUSSION

#### 3.4.1. Data fusion

A visual comparison between an original and a predicted image for DOY 163 is shown in Figure 3.2: the two images matched consistently, especially considering the heterogeneity character of the region. The computed NDVI from the predicted image 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 3.3 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 images 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.

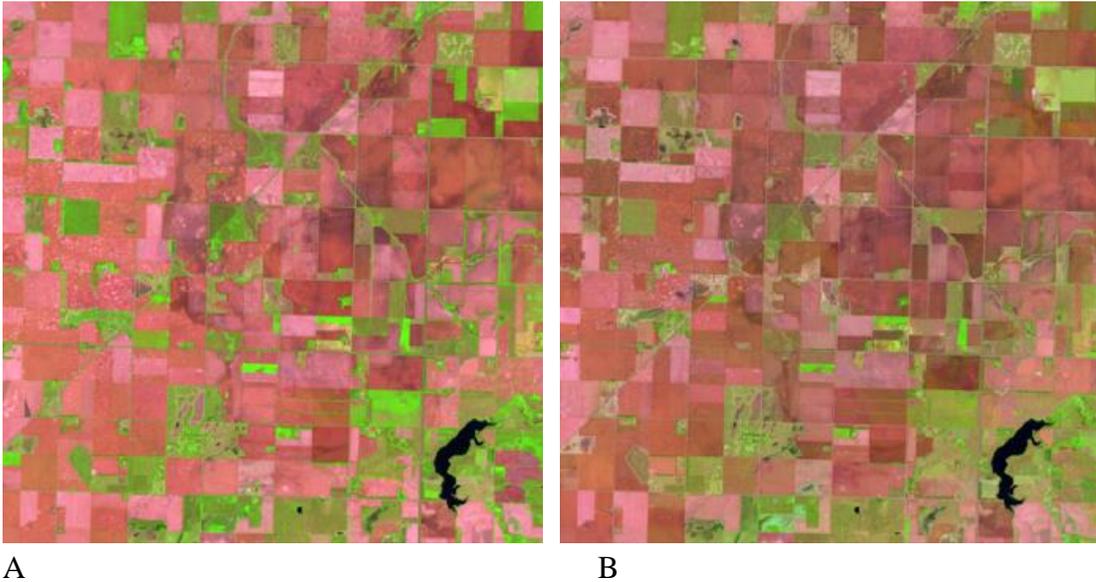


Figure 3. 2. An original Landsat image (A) and an ESTARFM predicted image (B) for the same date (DOY 163). The two images look very similar, a few differences in the brightness can be noticed due to the ENVI software enhancement for visualization. The two images are a false color composite using a combination of bands 6-5-4.

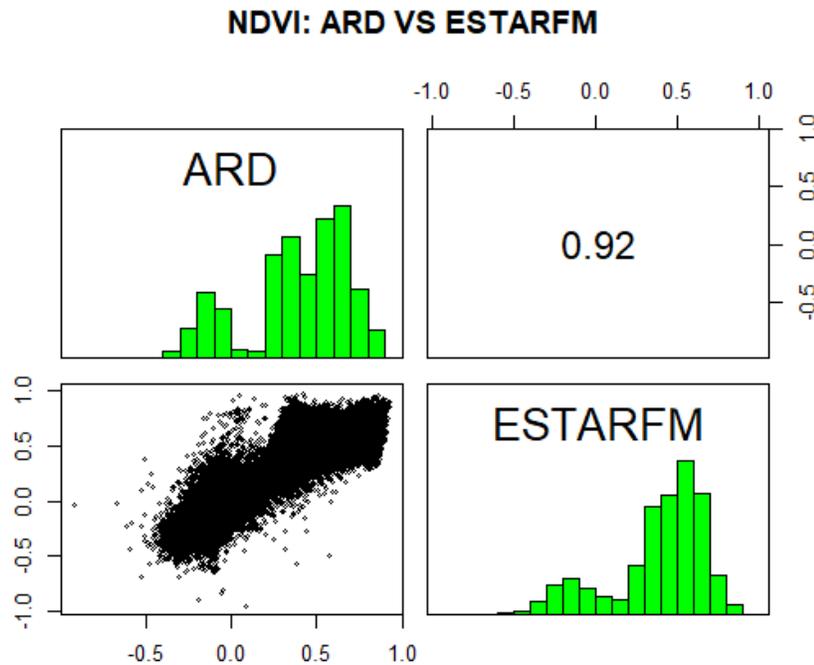


Figure 3. 3. Comparison of the actual and predicted band reflectance in a heterogeneous region. The upper-left and the lower-right quadrats show the Normalized Difference Vegetation Index (NDVI) respectively from the original Landsat and the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) image. The lower-left quadrat shows the NDVI distribution along the 1:1 line and the upper-right quadrat shows the correlation coefficient between the two images NDVI.

#### 3.4.2. 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) as an 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. The Monte-Carlo simulation of Moran's I provides statistically

significant results, indicating that the spatial autocorrelation result is not due by chance (Table 3.2).

Table 3. 2. Monte-Carlo simulation of Moran I

	2016	2017
number of simulations + 1	1000	1000
statistic	0.38	0.45
observed rank	1000	1000
p-value	0.01	0.01
alternative hypothesis	greater	greater

A Moran's I scatter plot of yield for the year 2016 is shown in Figure 3.4 (left): an obvious cluster can be observed in the lower-left quadrat, denoting low-low values of yield that may explain a 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 quadrats, but they do not form any clusters. Similarly, in 2017, the Moran's I yield scatter plot (Figure 3.4 – right) shows a significant clustering in the upper-right quadrat, denoting clusters of high values of yield, while the lower-left quadrat shows a more relaxed clustering of low values (low yield). However, some mixed values can be observed both in the upper-left and lower-right quadrats, but they are not as abundant as the clustered values.

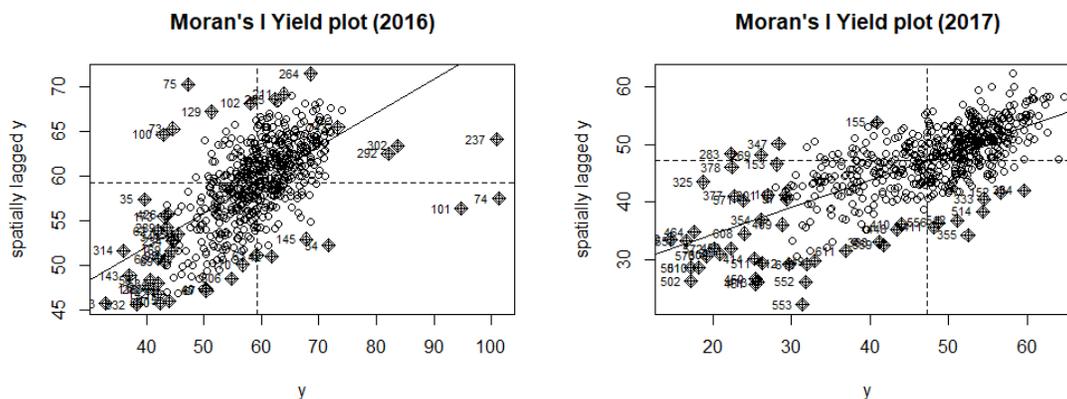


Figure 3. 4. Moran's I scatter plot showing evidence of clustering in both years: clusters are mostly located in the lower-left quadrat, which explains low yield (and probably crop stress) and in the upper-right quadrat which explains a high yield. Evidence of clustering is more pronounced in 2017 than in 2016.

These results are consistent with the literature, as Hartman et al. (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 a big influence on this pattern such as soil fertility, field elevation and others.

#### 3.4.3. The yield ordinary kriging model

The optimized model (Model:  $0 \cdot \text{Nugget} + 61.733 \cdot \text{Stable} (330.36, 0.57617)$ ) parameters are shown in the Figure 3.5. The nugget is zero, meaning that there is no difference in the values of points that have the same location (distance = 0 m). The major range (330 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 eleven pixels aligned in the same directions have a decreasing similarity and this similarity may not exist from the eleventh pixel. The sill (61.7) 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. The same process was completed for 2017 as well.

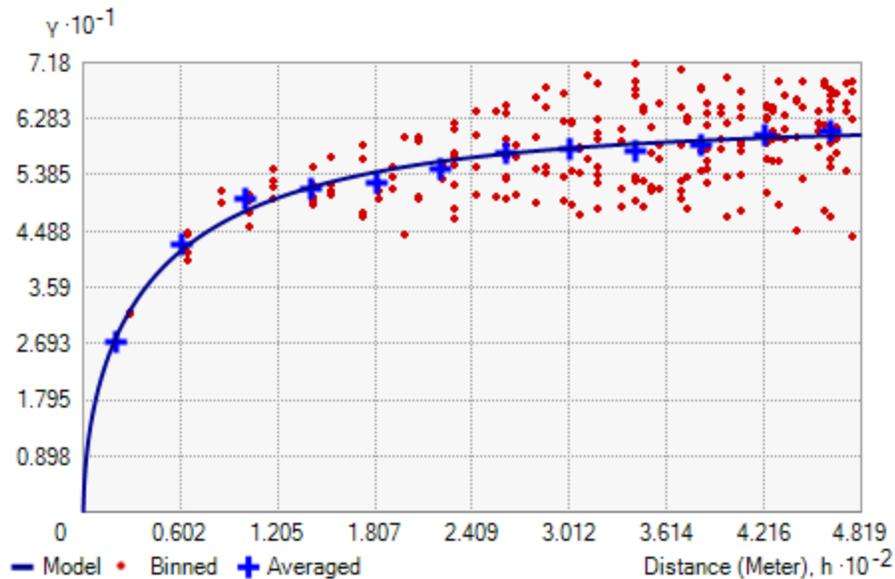


Figure 3. 5. Semivariogram of the 2016 yield Ordinary Kriging (OK) model. The blue crosses depict the computed predicted values, while the model is indicated by the blue line.

The optimized modeled variogram is data-dependent, meaning that it can vary considerably from one crop to another (Vieira and Gonzalez 2003), and from one year to another. A study by Jaynes and Colvin (1997) modeled spatiotemporal variability of yield. They used 6 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 one year to another. For example, nitrate availability may limit yield in a year with adequate rainfall while in a dry year, soil water capacity can be the most important factor. Further, the influence of rainfall on yield spatiotemporal variability was highlighted by Bakhsh et al. (2000) and Kumhálová et al. (2011) who suggested that yield was reduced in wet years as compared to dry years, according to their data. 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 variable variogram was dictated by the comparison with other models as they best fitted our dataset.

#### 3.4.4. Seasonal changes of the relationship between yield and NDVI

The correlation coefficients between yield and NDVI are different between the two 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. Yet, the source of the difference in the strength of the coefficients between the two years is unknown. Several factors such as the rainfall regime, the field management practices or field physico-chemical properties may explain that difference, but further investigation is needed and more data should be explored to better understand this difference. In 2016, the peak of the correlation coefficient is reached around August 21<sup>st</sup> ( $R = 0.65$ ), 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. However, in 2017, a subsequent image on September 08 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 August 21 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, while the difference in the NDVI signal in the early season might be small, the same difference can be seen at the maturity stage.

Another possible explanation may be the shift in the growing season; however, this hypothesis is not strong enough as the two fields are located just 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 in this difference; in other words, if the 8 days' interval corresponds to the one-degree latitude shift in the growing season. This aspect requires further investigations with more accurate and consistent inputs including high spatial and temporal resolutions images and field data for interpretation.

An alternative possibility 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). The difference in the average temperature between the two years during the period June-July-August was 2 °F for air temperature and 4 °F for soil temperature. Indeed, the average air and soil temperatures were 68 °F and 71 °F in Marshall, and 70 °F and 75 °F in Moody, respectively (Mesonet 2018). Furthermore, the rainfall difference can explain the yield difference between the two fields: Marshall had a total of 7.16 inches while Moody had a total of 12.05 inches (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.

#### 3.4.5. Estimating yield loss caused by white mold

In 2016, the best relationship between yield and NDVI was found on August 21<sup>st</sup>, while in 2017, the best relationship was found on September 8<sup>th</sup>. The scouted white mold pixels and their corresponding yield are shown in Figure 3.6 in 2016 and Figure 3.7 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 by the aggregation of yield data at pixel levels, while the selected resolution is likely to miss small patches of diseases. Nevertheless, the observed relationships remain strong enough for the moderate 30 m spatial resolution of Landsat.

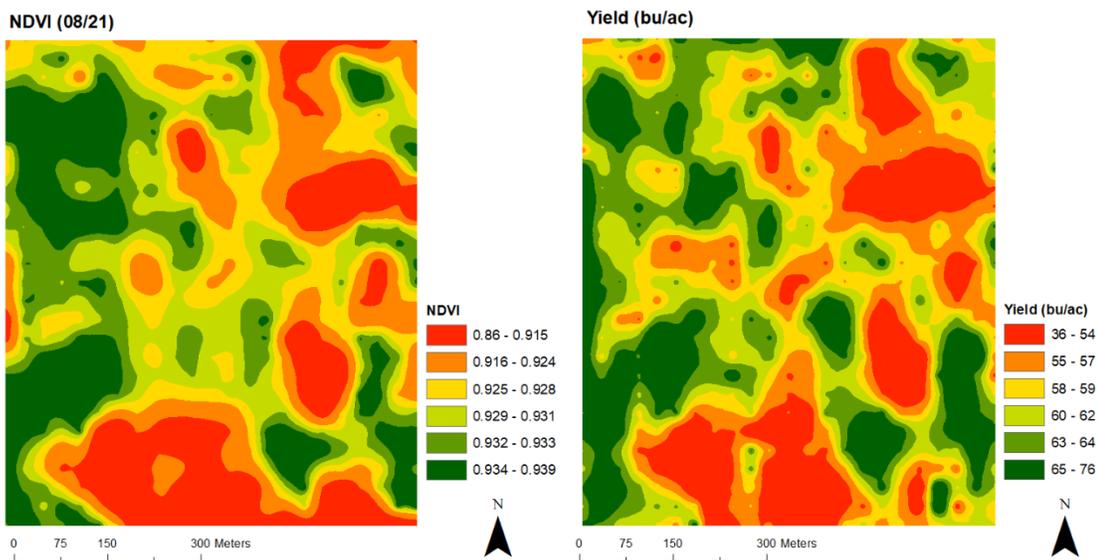


Figure 3. 6. Normalized Difference Vegetation Index (NDVI) classes (left) compared with the corresponding yield (right) for each class for the soybean field in Moody County, 2016. White mold pixels correspond to the three lowest NDVI classes.

The overall relationship between NDVI and yield was 65% on August 21<sup>st</sup>, 2016 and 77 % on September 8<sup>th</sup> 2017. The difference in the strength of this relationship between the two years, especially in common dates (August 21) may be due to many factors such as the rainfall regime, the local field managements or other factors that need to be investigated. However, the general trend in the impact of white mold on yield is similar for the two fields: white mold impact is weak in early August (disease on-set stage), increases and reaches the peak towards late August/early September, and starts decreasing mid-September (maturity to senescence stages). 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. 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 3.3 and 3.4. Hoffman et al. (1998) examined yield loss across different soybeans maturity groups and noticed that 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.

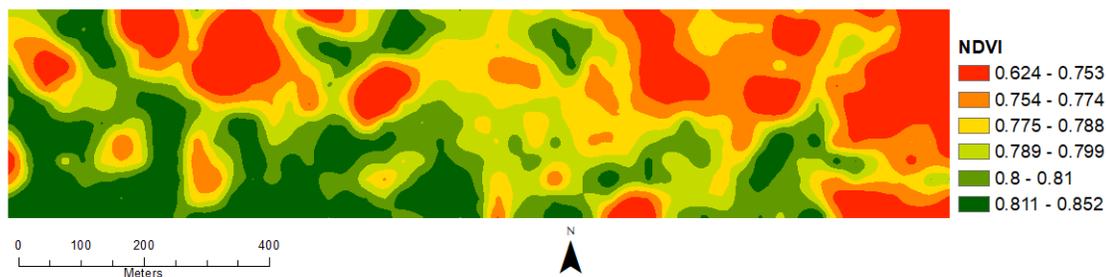
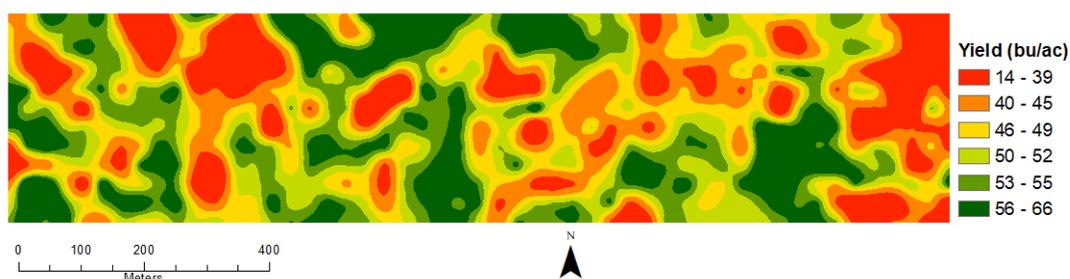
**NDVI (09/08)****Yield (bu/ac)**

Figure 3. 7. Normalized Difference Vegetation Index of a soybean field that had white mold (top) compared with the corresponding yield (bottom) in Marshall County in 2017.

Table 3. 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.

<b>Class range (Bu/ac)</b>	<b>Yield average (Bu/ac)</b>	<b>NDVI (08/21)</b>	<b>Yield loss (%)</b>
36-54	45	0.86-0.915	36
55-57	56	0.916-0.924	21
58-59	58.5	0.925-0.928	17
60-62	61	0.929-0.931	13

63-64	63.5	0.932-0.933	10
65-76	70.5	0.934-0.939	0

Table 3. 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.

<b>Class range (Bu/ac)</b>	<b>Yield average (Bu/ac)</b>	<b>NDVI (09/08)</b>	<b>Yield loss (%)</b>
14 – 39	26.8	0.62-0.76	56
39 – 45	42.1	0.76-0.77	30
45 – 49	46.9	0.78-0.79	22
49 – 52	50.4	0.8001-0.8009	16
52 – 55	53.3	0.81-0.819	12
55 – 66	60.3	0.82-0.85	0

The relationship between yield loss and the corresponding average NDVI for each class is shown in Figure 3.8. 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. This explains why the estimation on August 21<sup>st</sup> in 2016 provides a lower correlation coefficient than the estimation made on September 8<sup>th</sup> 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 satellites images should include varieties that may be due to 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 physico-chemical characteristics.

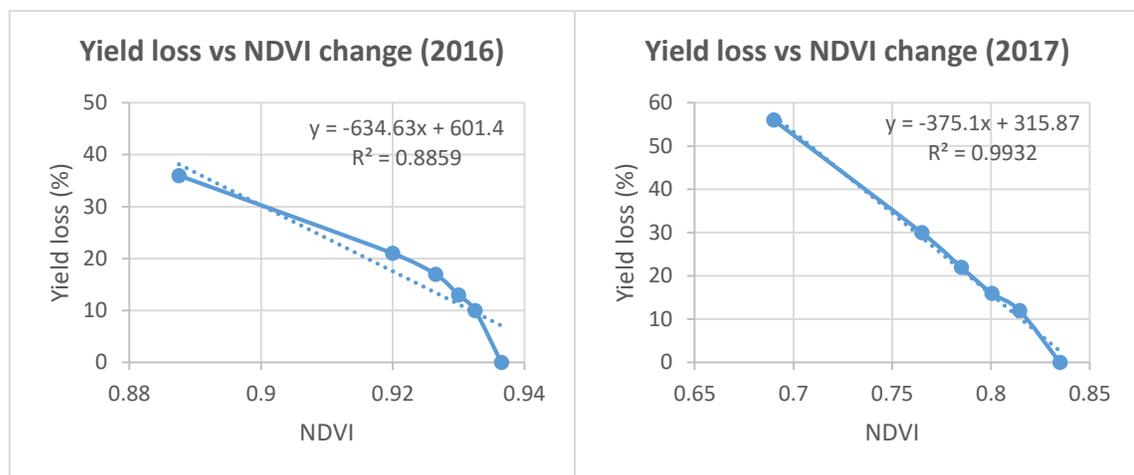


Figure 3. 8. Yield loss as a result of the Normalized Difference Vegetation Index (NDVI) change.

### 3.5. 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 on 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% to 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 be useful in explaining some patterns that cannot be inferred directly and only with satellites images.

CHAPTER 4: MODELING TEMPORAL PATTERNS OF WHITE MOLD IN SOUTH  
DAKOTA USING FUSION OF LANDSAT AND MODIS DATA

Mfuka, C. and Byamukama, E. Modeling Temporal Patterns of white mold in South Dakota using Fusion of Landsat and MODIS Data. Manuscript in preparation.

This chapter addresses **Hypothesis #3**: White mold temporal characteristics can be modeled as a function of the Normalized Difference Vegetation Index obtained by a fusion of Landsat and MODIS.

#### 4.1. ABSTRACT

Soybean production in South Dakota is affected by many diseases, among which an important fungal disease, *Sclerotinia Stem Rot* (SSR), also known as white mold. Currently, information on the temporal characteristics of the disease is scanty. Several healthy and white mold infected fields were scouted in Marshall, Day and Codington counties, from which spectral characteristics were extracted. This study used the Normalized Difference Vegetation Index (NDVI) derived from a fusion of free remotely-sensed images (Landsat 8 and MODIS) to model the temporal progress of white mold using a logistic regression analysis. White mold could be predicted using the NDVI; however, the model was improved by the inclusion of the temporal component expressed as the Day of the Year (DOY). The change in the temporal development, expressed as the difference in the NDVI between the healthy soybeans and white mold, was statistically different ( $p < 0.0001$ ) between sites, especially in five of the six paired sites comparison. This study shows the potential of free remotely-sensed data in monitoring and predicting crop health.

**Keywords:** White mold, soybean, temporal progress, Landsat, MODIS, Fusion, NDVI, logistic Regression

## 4.2. INTRODUCTION

Sclerotinia stem rot (SSR), also known as white mold, is among the top diseases that affect soybean (*Glycine Max*) production in North America. White mold is ranked among the top 10 soybean diseases between 2010 and 2014 (Allen et al. 2017). The disease develops in the lower portion of the soybean stem and leads to the entire plant wilting. Yield loss due to white mold has been as high as 50% depending on the time of infection and the susceptibility of the cultivar planted. Although a number of studies on white mold have been done, most of the studies focused either on the epidemiology of the disease (Boland and Hall 1988), or in studying specific field aspects suspected of influencing yield (Hoffman et al. 1998). Danielson et al. (2004) studied the effects of white mold on yield of soybean inoculated at different growth stages. The effects of herbicides or time of fungicides application on white mold have been studied by Nelson et al. (2002) and Milas et al. (2003). Other factors, including row spacing (Buzzell et al. 1993), tillage, crop sequence, and cultivar have also been studied (Kurle et al. 2001, Mueller et al. 2002).

Recent studies that modelled the temporal occurrence of white mold either were limited to the observation of a few factors or employed experimental field settings and intensive data collection. Fall et al. (2018) examined the temporal occurrence of apothecia, but the study only focused on two factors: canopy closure and air temperature. Willbur et al. (2018) modelled the risk of occurrence of apothecia, but their methodology required intensive field data collection. There is still a need to investigate the disease development on a large scale and during the extended growing season by using cost-

effective and non-destructive methods such as available remotely sensed images, to better understand the white mold development.

This study employed the Normalized Difference Vegetation Index -NDVI (Rouse et al. 1973) computed from the fusion of two freely available remotely sensed satellite data (Landsat and MODIS) to investigate the temporal development of white mold in selected counties in South Dakota.

### 4.3. MATERIALS AND METHODS

#### 4.3.1. Study area

The study area was in northeastern South Dakota and included Marshall, Day and Codington counties. The region is characterized by a heterogeneous land cover dominated by cropland and pasture/grassland, as well as wetland, forestland and open water (USDA 2017). Major crops include corn, soybean, alfalfa, and spring wheat. A total of 67 fields were analyzed in this study, including 11 fields that were scouted for the presence of white mold, and 56 fields that were mapped in the Chapter 2 of this dissertation, all in the same year 2017. These 67 fields consisted of 37 healthy fields and 35 white mold infected fields, which were each associated to the closest of the four weather stations (Britton, Groton, Webster, and South Shore) in terms of geographical location. Figure 4.1 shows the location of the study area and the fields that were analyzed in this research.

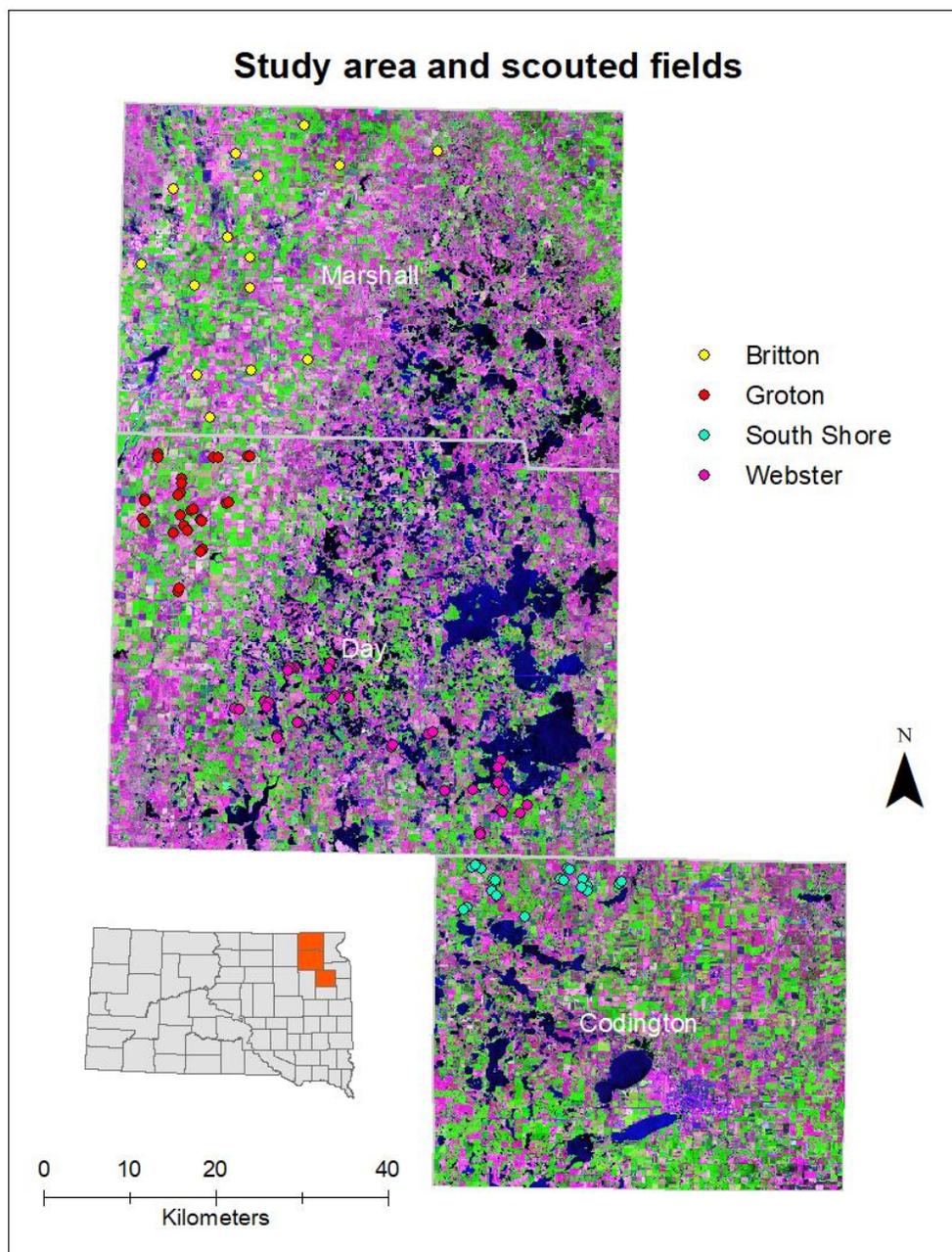


Figure 4. 1. Location of the study area showing different scouted fields with their respective weather station names. The background image is a July 14, 2017 false color composite of Landsat 8 using bands 6-5-4.

#### 4.3.2. Data collection and preparation

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 3 cloud-free (on our study area) Landsat 8 ARD in the 2017 growing season from earth explorer (<https://earthexplorer.usgs.gov/>), covering the dates May 11, July 14, and August 31.

Daily MODIS images (MDC43A4 Version 6 products) were downloaded from the period May 11 to September 8 of the year 2017, and included 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 and stacked in ENVI, 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 500m to 30m using the nearest neighbor algorithm in ENVI, to match the corresponding Landsat image and ease the fusion of the two datasets.

#### 4.3.3. Data fusion

The data fusion process was used to combine the strengths of both Landsat and MODIS images into a daily high spatial resolution (30 m) synthetic dataset. The

algorithm employed was ESTARFM (Zhu et al. 2010); the details about the resulting images and their quality assessment were provided in the Chapter 3 of this dissertation. The same dataset was used in this chapter to analyze the temporal occurrence of white mold in selected fields.

#### 4.3.4. NDVI computation

The resulting fused synthetic images were used to compute the NDVI using the Red and the NIR bands (Equation 1). NDVI 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. The computed NDVI images had a 30 m spatial resolution, matching the originally downloaded ARD Landsat images. For each healthy and white mold infected field, a NDVI of the healthy/infected pixel was extracted throughout the growing season. For each weather station, an average NDVI was computed for both the healthy and the white mold pixel categories.

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

Where: NIR is the Near Infrared Band and Red is the Red Band

#### 4.4. ANALYSES

The spectral characteristics of healthy soybeans and white mold pixels were extracted using the NDVI for each individual field. For each weather station, a mean NDVI was computed for both healthy soybeans and white mold pixels. The resulting NDVI was plotted against time (Day of the year –DOY) to examine the temporal

difference between healthy soybeans and white mold. The difference between healthy soybeans and white mold NDVI was further computed. The change in the NDVI difference was assessed and examined between sites. An analysis of variance (ANOVA) was conducted to examine statistical differences between sites. While the ANOVA tells us only if there is difference or not, a Tukey Honest Significant Differences test was computed to identify the sources of these differences. It compares groups pairs (weather station) between them to detect which ones are statistically different from others.

A logistic regression model was computed to model the soybean status (healthy and white mold) as a function of the NDVI (Press and Wilson 1978). The logistic regression can be considered as a special case of linear regression, where the outcome variable is categorical, generally binary and in which the probability of occurrence of an event is predicted by fitting data to a logit function. In our study, the outcome variables included healthy soybeans and white mold. An empirical formula for the logistic function is shown in Equation 2 (Battilani et al. 2008). NDVI from the four sites were analyzed with the GLM (General Linear Model) function in RStudio (RStudio 2016). The initially developed model (NDVI model) was compared with another model that included NDVI and DOY as independent variables in terms of the difference between the Null deviance and the Residual Deviance, and the Akaike Information Criterion (AIC). Moreover, the Receiver Operating Characteristic (ROC) was used to assess the Area Under the Curve. An ANOVA of the two models was computed to assess the significance of the respective variables.

$$Prob(\text{white mold}) = \frac{1}{1+e^{-z}} \quad (2)$$

With  $e$ : the base of the natural logarithms and  $Z$  is the linear combination defined as

$$Z = B_0 + B_1X_1 + \dots + B_pX_p \quad (3)$$

With  $B_0, B_1, B_p$  the coefficient estimates of the independent variables  $X_1, X_p$  (NDVI, DOY).

#### 4.5.RESULTS AND DISCUSSION

##### 4.5.1. Time-series NDVI for healthy soybeans and white mold

For each site, an average NDVI was computed for both healthy soybeans and white mold affected areas. Figure 4.2 shows the distribution of the average NDVI for each weather station. NDVI values present similar distribution with the mean around 0.8 (or less); there is no observable outlier candidates and the values are comparable.

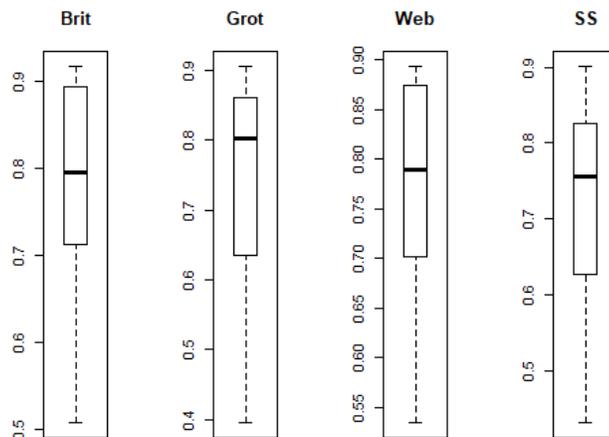


Figure 4. 2. Average Normalized Difference Vegetation Index (NDVI) values computed for each weather station. Values distribution is similar with no outliers. Brit=Britton, Grot=Groton, Web=Webster, SS= South Shore.

A scatter plot of the NDVI frequency distribution for each site (Figure 4.3) shows that there are many overlaps (especially in South Shore), which makes it hard to model white mold based on the NDVI alone using simple modeling techniques such as linear regression. An important part of the healthy observations (blue) are located on NDVI values above 0.8 while the white mold observations (pink) are mostly located around lower values. This data distribution makes it hard to model soybean health.

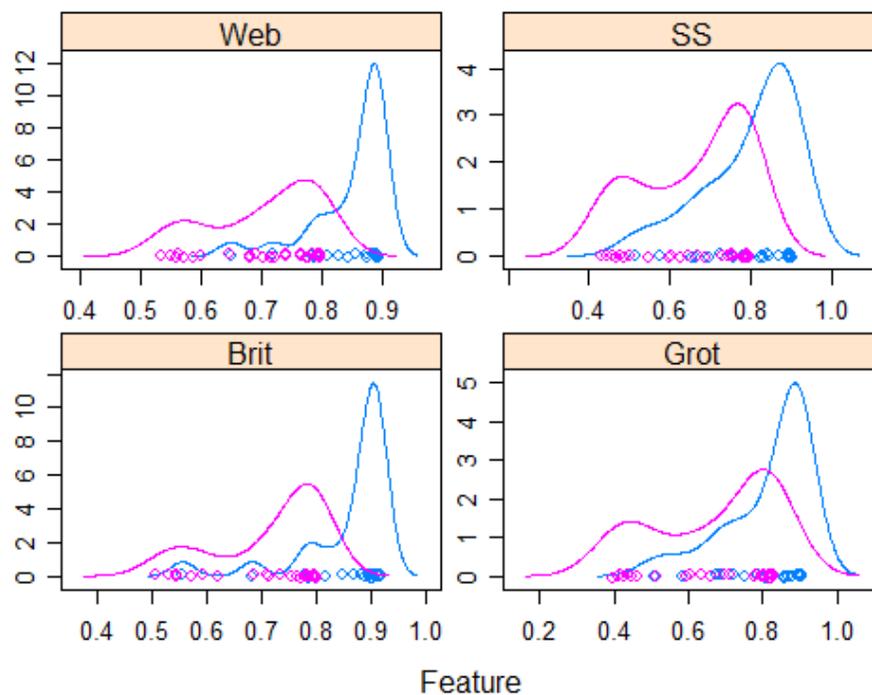


Figure 4. 3. Scatter plot of the frequency distribution (y axis) of the Normalized Difference Vegetation Index (NDVI, x axis) for healthy soybeans (blue) and white mold (pink).

The temporal trend in the NDVI values between the healthy soybeans and white mold pixels are shown in Figure 4.4. In all the sites, there is a clear difference between healthy soybean and white mold, where the first has a greater NDVI value than the latter.

In all sites, the healthy soybeans exhibit a similar trend to white mold, but the difference resides in the magnitude of the observed NDVI. However, maximum NDVI can occur at different times as a function of the difference in the soybeans maturity groups (Dardanelli et al. 2006), or the management practices including the planting dates (Hershman et al. 1990). In healthy soybeans, maximum NDVI is the highest in Britton and occurs earlier than in other sites, around DOY 221, while this maximum NDVI occurs around DOY 228 in Groton and Webster, and around DOY 233 in South Shore. The observed maximum NDVI follows the same trend in white mold fields.

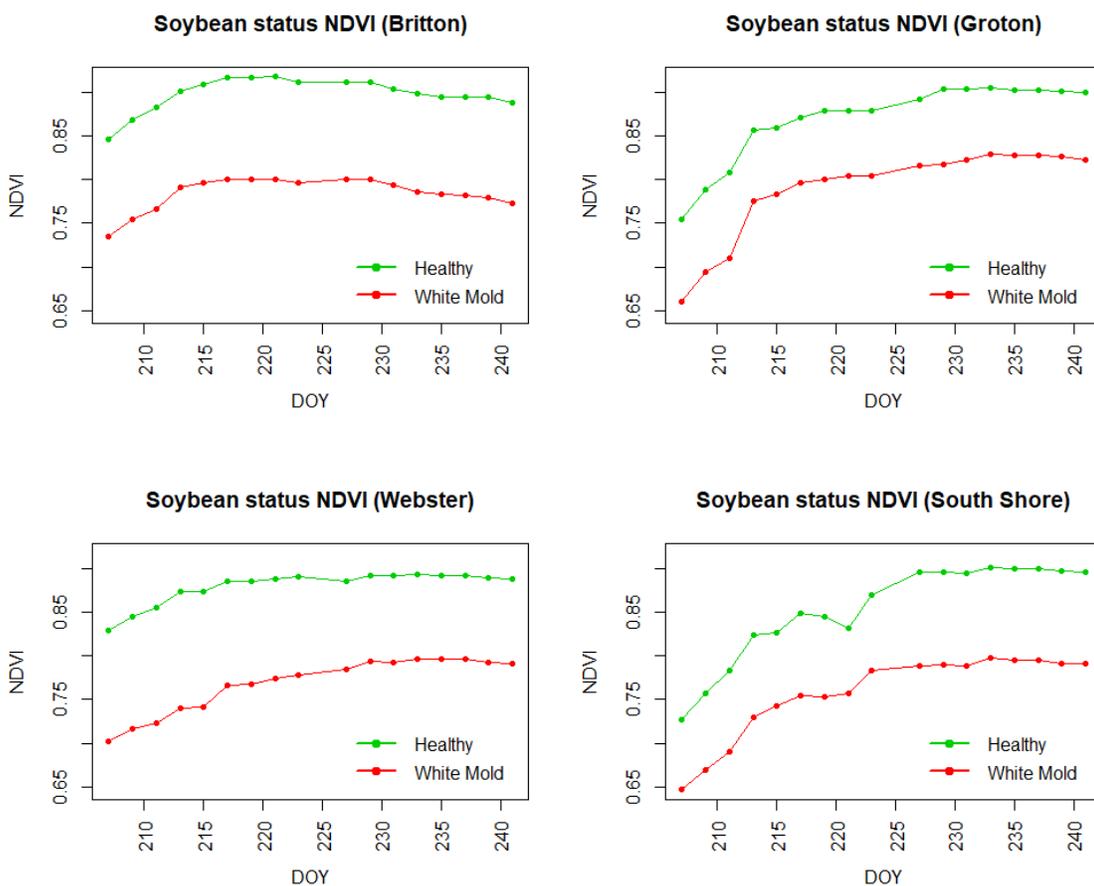


Figure 4. 4. Seasonal Normalized Difference Vegetation Index (NDVI) of healthy soybean as compared to white mold for each site. White mold starts declining earlier than healthy soybean and has lower NDVI after Day of the Year (DOYS) 207.

#### 4.5.2. Temporal change in the NDVI difference between healthy soybeans and white mold

For each site, we computed the NDVI difference between healthy soybeans and white mold. The NDVI difference allows to assess the magnitude of the NDVI variation from one site to another. The boxplot shown in Figure 4.5 shows values distribution, highlighting some differences between sites. Some potential outliers can be observed in Groton and similarities can be observed in the means between Britton and Webster, while other sites exhibit dissimilar NDVI differences.

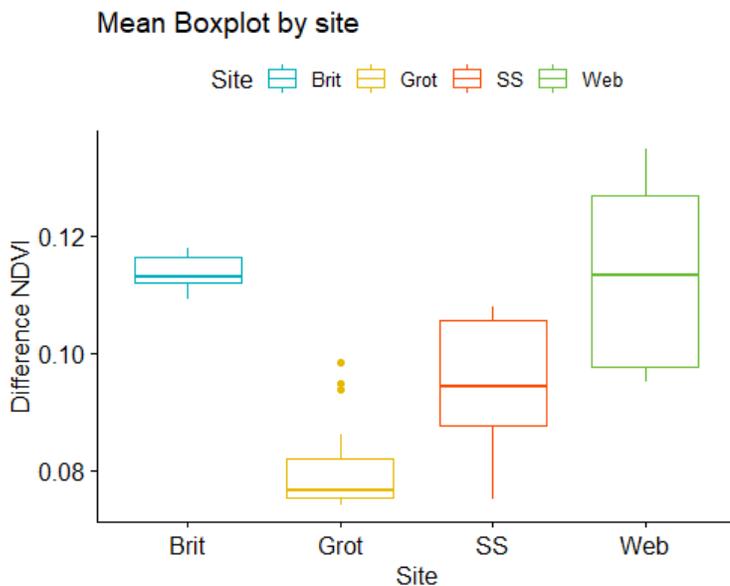


Figure 4. 5. Box plots of Normalized Difference Vegetation Index (NDVI) difference by site.

An analysis of variance (ANOVA) was conducted to investigate the variances in the difference between healthy soybeans and white mold NDVI as observed in different sites. The computed p-value is  $8.77e-15$ , which is significant at level 0.0001, meaning that at least two sites are statistically different. The site location is a significant factor in the temporal NDVI difference, however, we need to investigate the source of the differences by testing sites pairs. The Tukey Honest Significant Differences (Tukey HSD) test was used to compare site-pairs and assess the differences. Results (Table 4.1) show that all the pairs but Webster-Briton are statistically different, meaning that the temporal change in the white mold signal in comparison to healthy soybean can be different from one location to another. While these observations are made at county scales, similar trends are expected to be observed at larger scales, especially if environmental factors such as temperature or rainfall exhibit important differences.

Table 4. 1. Tukey Honest Significant Differences test for detecting similarities in the NDVI difference temporal change between sites, showing the 95% lower and upper limit values.

	<b>NDVI Difference</b>	<b>Lower limit</b>	<b>Upper limit</b>	<b>p-adj</b>
Grot-Brit	-0.033276573	-0.042437089	-0.024116057	0.0000000
SS-Brit	-0.017863014	-0.027023530	-0.008702498	0.0000162
Web-Brit	-0.001679337	-0.010839853	0.007481179	0.9624744
SS-Grot	0.015413559	0.006253043	0.024574075	0.0002100
Web-Grot	0.031597236	0.022436720	0.040757752	0.0000000
Web-SS	0.016183677	0.007023161	0.025344193	0.0000955

#### 4.5.3. Temporal modeling of white mold

A logistic regression was computed to model the white mold as a function of NDVI progression. However, as shown in Figure 4.3, it is hard to model white mold based on NDVI alone, since the NDVI values overlap between healthy soybeans and white mold. A subsequent model was developed, that included the DOY as an additional input. The two models are summarized in tables 4.2 for NDVI only, and table 4.3 for NDVI+DOY respectively. Both models show that the independent variables are highly significant at  $p < 0.0001$  for NDVI only in the first model, and  $p < 0.001$  and  $p < 0.01$  respectively for NDVI and DOY for the second model. The NDVI estimate is negative in all models, meaning that the disease development is explained by the NDVI decrease, which can be explained by the reduction of the photosynthetic activity in infected crops. However, the inclusion of the DOY in the logistic regression improves the disease modeling: the positive estimate of DOY shows that the disease develops with time, especially within the studied period. These results are in agreement with those of Byamukama, Robertson and Nutter (2010) who found that the DOY explained about 94-98% of the variation in the Bean pod mottle virus (BPMV) logit model. The big difference between the two studies resides in the quality and the quantity of the data collected. In fact, Byamukama et al. (2010) collected intensive field data including disease incidence over different years, while our study focused mainly on NDVI extracted from the synthetic images. Similarly, Byamukama, Robertson and Nutter (2011) estimated the disease rate development based on a logistic regression.

Table 4. 2. White mold modeled as a function of the Normalized Difference Vegetation Index (NDVI) only.

	<b>Estimate</b>	<b>Std. Error</b>	<b>Z Value</b>	<b>Pr(&gt; z )</b>
Intercept	55.53	11.02	5.038	4.70e-07***
NDVI	-67.38	13.49	-4.994	5.91e-07***

Table 4. 3. White mold modeled as a function of the Normalized Difference Vegetation Index (NDVI) and Day of the Year (DOY).

	<b>Estimate</b>	<b>Std. Error</b>	<b>Z Value</b>	<b>Pr(&gt; z )</b>
Intercept	12.9461	12.5175	1.034	0.30102
NDVI	-131.6652	44.4090	-2.965	0.00303**
DOY	0.4273	0.1814	2.355	0.0185*

A comparison between the two models is summarized in table 4.4. The null deviance is a prediction made by a model that uses only the intercept while the residual deviance is a prediction made by a model which adds independent variables. As a rule, the lower these two values, the better the model. The null deviance (composed only of the intercept) is the same for the two models, while the residual deviance is lower for the NDVI+DOY model while compared to the NDVI model. The AIC however is considered as metric similar to  $R^2$ , with the difference that the lower the better the model.

Subsequently, the inclusion of the DOY reduces the AIC from 47.56 to 24.91. The ROC

summarizes the model's performance by evaluating the trade-offs between true positive and false positive rates. The area under the ROC curve explains how well the ROC performs; therefore, the higher the AUC, the better the model. The computed AUC is higher for the NDVI+DOY (0.99) model than for NDVI (0.95) only model as shown in Figure 4.6. A recent study showed similar results where the vegetation index exhibited strong relationship ( $R^2=0.90$  for NDVI) with the AUC (Loladze et al. 2019). This highlights the importance of using time-series images for vegetation monitoring, whether for low (Zhang et al. 2003) or moderate spatial resolution images (Bhandari, Phinn and Gill 2012, Dong et al. 2016). The use of multiple images allows to characterize a particular disturbance and distinguish the observed signal from other disturbances that may present similar characteristics in a single common date.

Table 4. 4. Comparison between the Normalized Difference Vegetation Index (NDVI) model and the NDVI + (Day of the Year) DOY models.

	<b>NDVI</b>	<b>NDVI + DOY</b>
Null deviance	141.402	141.40
Residual deviance	43.555	18.91
AIC	47.555	24.91
ROC AUC	0.957	0.993

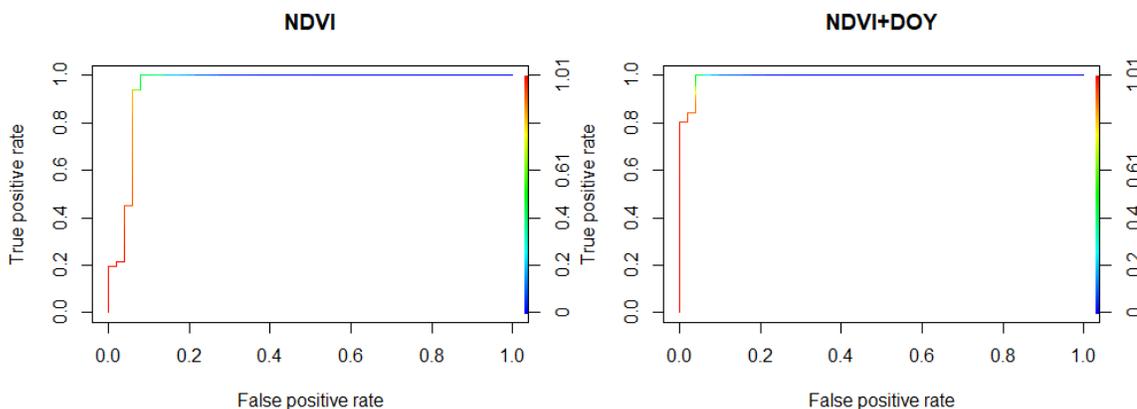


Figure 4. 6. Area Under the Curve computed by the two models, showing a better (bigger area) estimation for the model that includes the Normalized Difference Vegetation Index (NDVI) and Day of the Year (DOY).

#### 4.6. CONCLUSION

The study shows that white mold can be modelled using the NDVI. However, a better model would include timing consideration such as the Day of the Year (DOY). This highlights the importance of using remotely-sensed satellite images to help keep a close look on the development of crop diseases. The difference in the temporal development of white mold can be influenced by the sites location. This means that several local factors can influence the temporal development of the disease, including local field management, crop varieties, soil properties, or weather. Other factors could include change in the local elevation. While the models developed with remotely sensed satellite data are accurate enough, they still could be improved with complementary observations (i.e field surveys or scouting) that might represent an important factor in validating or controlling the remotely-sensed data and the evolution of crop health.

The temporal development of white mold can be further monitored by the use of fine spatial resolution images such as SENTINEL-2 or Planet images, in order to capture subtle spatio-temporal development of the disease, which could help in estimating the disease development rate. Punctual observations such as those acquired by the use of unmanned aerial vehicles (UAV), can be a time-efficient input for valid assessment of crop health.

## CHAPTER 5: SUMMARY OF THE RESEARCH

### 5.1. HYPOTHESIS #1:

*Based on field knowledge and available Landsat images, the occurrence of white mold can be accurately mapped and quantified.*

#### 5.1.1. Summary of the methods

Landsat images were collected for the year 2017 covering three counties in north-eastern South Dakota. Three images in total have been collected, covering the dates May 11, July 14 and August 31. The first two images were used to map the land cover and extract the soybean mask, while the last image was used to map white mold. The classification was conducted using the random forest (RF) algorithm for both the land cover and white mold. In the first step, trainings were collected for the four mapped classes (Water, Corn, Soybean and Other Land Cover) and the RF algorithm used the Landsat bands from the two dates as independent variables, and the four classes as response variable. In the second step, training have been collected on known white mold fields and the RF algorithm used the five Landsat bands and the NDVI as independent variables and the forest status (healthy vs unhealthy) as response variables. Spectral characteristics of white mold were used to reduce false positive, by excluding from the soybean mask, early stress that might occur before the timing of white mold. White mold areas have been extracted from the mapped white mold pixels.

#### 5.1.2. Results and Conclusions

Hypothesis #1 was verified: The Land cover map was classified with an overall accuracy of 95%, with the soybean class mapped with Producer's and User's Accuracies

of respectively 94% and 93%. The mapped soybean areas compared well with the soybeans areas reported by the USDA, meaning that the classified map was accurate enough to map the land cover. Additionally, soybean health map was mapped with an overall accuracy of 99%, with the white mold producing Producer's and User's accuracy of respectively 99% and 97%. Furthermore, the known fields have been checked visually on the resulting maps for accuracy and they all have been correctly mapped with white mold.

### 5.1.3. Implications and future work

The results obtained in this study show that white mold can be accurately mapped with freely available remotely sensed satellite images. Our study used the Random Forest algorithm for the classification, but other algorithms such as Artificial Neural Network (ANN) can also be used as some studies showed that it outperformed RF (Raczko and Zagajewski 2017). The Crop Data Layer dataset can also represent an alternative; however, the release date of this dataset may represent a limitation if disease estimation needs to be made sooner or within the growing season. When more accurate images are available such as the SENTINEL-2 images, such a study might be able to capture smaller patches of white mold that could not be captured by the medium/low resolution Landsat (30m). In fact, in the field, some occurrence of white mold is smaller than the smallest detectable unit by the satellite, which results to an underestimation of the diseased areas. A more extensive field scouting could also improve the accuracy of the mapped white mold by incorporating regional differences. Disease rating might also represent an important factor mapping the occurrence of white mold, as according to the latitude and the difference in the planting dates for instance, some differences might be observed in

the signal of white mold. Including all those differences would allow to better capture the disease.

## 5.2. HYPOTHESIS #2:

*Spatiotemporal characteristics of yield in white mold infected soybean fields can be modeled using NDVI computed from a fusion of Landsat and MODIS images, and the relationship can provide estimates of yield loss caused by white mold.*

### 5.2.1. Summary of the methods

Yield data collected in a two different soybean fields in two different years (2016 and 2017), was harmonized in a regular 30mx30m grid, with individual cells corresponding to Landsat pixels, and on which the average yield was computed for each cell. The regular grid was converted into a neighborhood object from which yield spatial distribution was analyzed using the Moran's I statistics. A validation of the computed Moran's I was assessed with a Monte-Carlo simulation. The yield and NDVI values extracted from each cell for the two years were used to generate surface maps using the kriging algorithm. For the 2017 soybean field, NDVI were computed from images resulting from the fusion of Landsat and MODIS to match the 2016 images. These time-series NDVI were analyzed to estimate the best yield predictor using a simple linear regression analysis. The relationship between white mold pixels and yield was examined to estimate the yield loss caused by white mold, and the best period for accurately estimating this loss.

### 5.2.2. Results and Conclusions

Hypothesis #2 has been verified: The correlation coefficients extracted from the simple linear regression between the time-series NDVI and yield suggested that yield can be predicted from the NDVI from DOY 218 (August 21) and DOY 251 (September 8). As expected, the correlation coefficient increased throughout the growing season, reaching the peak in the interval described above, and declining as the crops reached senescence. The Moran's I test for the spatial distribution of yield suggested a significant positive autocorrelation (Moran's I = 0.38, p-value < 2.2e-16 in 2016 and Moran's I = 0.45, p-value < 2.2e-16 in 2017), meaning that there is evidence of clustering (the distribution of yield is not random). The Monte-Carlo simulation of this distribution suggested that evidence of clustering was not due by chance. Yield loss caused by the worst white mold pixels was estimated between 36% and 56% for the two fields, with the most accurate estimation occurring between late August and early September.

### 5.2.3. Implications and future work

The results obtained in this study may be applied in other fields: producers can have an idea of their expected yield by using NDVI computed around the dates of importance. However, in the presence of consistent images, there is no need to employ data fusion in order to obtain NDVI. Also, the use of high spatial resolution (10 m) and high temporal resolution (5days) such as SENTINEL-2 can open an interesting venue for predicting yield by using frequent and accurate vegetation indices computed on the required areas. Another alternative would be the inclusion of high resolution (<1m, daily) commercial images such as Planetscope, as these images can be now obtained with authorized access. Farmers can also use personal drones to capture images as needed, if

they can produce important products such as the vegetation index that may help them get an idea of the expected yield.

It would be important to explore elevation influence in the occurrence of white mold in soybeans field. This may require acquisition of yield data, local elevation or LiDAR data, and high spatio-temporal images in different years, as this influence can vary as a function of the raining regime. As demonstrated in this study, yield spatial distribution is not random, and may be influenced by other factors; soil physico-chemical properties, when available, can also be used in the investigation of factors influencing the yield spatial distribution. However, these data require intensive and systematic collection.

### 5.3. HYPOTHESIS #3:

*White mold temporal characteristics can be modeled as a function of the Normalized Difference Vegetation Index obtained by a fusion of Landsat and MODIS.*

#### 5.3.1. Summary of the methods

Several soybean fields were scouted and confirmed with white mold in Marshall, Day and Codrington counties in 2017. Daily NDVI were computed from the synthetic fused images (from Landsat 8 and MODIS). For healthy fields and fields infected with white mold, NDVI was extracted from the central pixel of each field and the average NDVI signal was calculated. Time-series NDVI allowed to assess the temporal change in the difference between healthy soybeans and white mold. A logistic model was used to model white mold as a function of NDVI only, and NDVI+DOY.

### 5.3.2. Results and Conclusions

Hypothesis #3 has been confirmed: white mold was modelled as a function of NDVI resulting from the fusion of Landsat and MODIS. The model was improved by the inclusion of the DOY as an independent variable as shown by several model parameters. While the null deviation is the same for both models (141.1) the residual deviance is lower for the NDVI+DOY model (18.91) than for the NDVI only model (43.56). The NDVI+DOY model reduces the AIC from 47.55 to 24.91. Finally, the computed AUC is improved from 0.95 to 0.99 by the NDVI+DOY model. The ANOVA of NDVI changes between healthy soybeans and white mold showed that at least one location was different from the others. A paired comparison between locations showed the temporal development of white mold was different in five of the six pairs comparison.

### 5.3.3. Implications and Future Work

These results show the importance of using remotely-sensed images in monitoring crop health, particularly white mold in this study. NDVI was able to model white mold using a logistics regression model, but the model was further improved by including the DOY as an independent variable, highlighting the importance of the timing of the disease for its characterization. White mold is a complex disease, and several factors can influence its temporal development. These factors include local conditions that can range from soil physico-chemical properties, local elevation, to local weather. A more accurate white mold analysis needs therefore to consider these local and regional differences. An ideal and robust model should be able to capture the maximum of variabilities by including different sites characteristics. This remains however a complex task that can be

achieved by the development of regional collaborations that could include many universities in order to capture most of the disease variabilities.

## REFERENCES

- Adam, E., O. Mutanga, J. Odindi & E. M. Abdel-Rahman (2014) Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35, 3440-3458.
- Allen, T. W., C. A. Bradley, A. J. Sisson, E. Byamukama, M. I. Chilvers, C. M. Coker, A. A. Collins, J. P. Damicone, A. E. Dorrance, N. S. Dufault, P. D. Esker, T. R. Faske, L. J. Giesler, A. P. Grybauskas, D. E. Hershman, C. A. Hollier, T. Isakeit, D. J. Jardine, H. M. Kelly, R. C. Kemerait, N. M. Kleczewski, S. R. Koenning, J. E. Kurle, D. K. Malvick, S. G. Markell, H. L. Mehl, D. S. Mueller, J. D. Mueller, R. P. Mulrooney, B. D. Nelson, M. A. Newman, L. Osborne, C. Overstreet, G. B. Padgett, P. M. Phipps, P. P. Price, E. J. Sikora, D. L. Smith, T. N. Spurlock, C. A. Tande, A. U. Tenuta, K. A. Wise & J. A. Wrather (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.
- Amorós-López, J., L. Gómez-Chova, L. Alonso, L. Guanter, R. Zurita-Milla, J. Moreno & G. Camps-Valls (2013) Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 23, 132-141.
- Anselin, L. (1995) Local Indicators of Spatial Association - Lisa. *Geographical Analysis*, 27, 93-115.

- Avendaño, F., O. Schabenberger, F. J. Pierce & H. Melakeberhan (2003) Geostatistical Analysis of Field Spatial Distribution Patterns of Soybean Cyst Nematode. *Agronomy Journal*, 95.
- Bajwa, S., J. Rupe & J. Mason (2017) Soybean Disease Monitoring with Leaf Reflectance. *Remote Sensing*, 9, 1-14.
- Bakhsh, A., D. B. Jaynes, T. S. Colvin & R. S. Kanwar (2000) Spatio-Temporal Analysis of Yield Variability for a Corn-Soybean Field in Iowa. *Transactions of the ASAE*, 43, 31-38.
- Battilani, P., A. Pietri, C. Barbano, A. Scandolaro, T. Bertuzzi & A. Marocco (2008) Logistic regression modeling of cropping systems to predict fumonisin contamination in maize. *J Agric Food Chem*, 56, 10433-8.
- Behmann, J., J. Steinrucken & L. Plumer (2014) Detection of early plant stress responses in hyperspectral images. *Isprs Journal of Photogrammetry and Remote Sensing*, 93, 98-111.
- Bhandari, S., S. Phinn & T. Gill (2012) Preparing Landsat Image Time Series (LITS) for Monitoring Changes in Vegetation Phenology in Queensland, Australia. *Remote Sensing*, 4, 1856-1886.
- Bhunja, G. S., P. K. Shit & 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.
- Blazquez, C. A., B. Picarte, J. F. Calderon & F. Losada (2018) Spatial autocorrelation analysis of cargo trucks on highway crashes in Chile. *Accid Anal Prev*, 120, 195-210.

- Boland, G. J. & R. Hall (1988) Epidemiology of Sclerotinia Stem Rot of Soybean in Ontario. *Phytopathology*, 78, 1241-1245.
- Bradley, C. A. (2008) Effect of fungicide seed treatments on stand establishment, seedling disease, and yield of soybean in North Dakota. *Plant Disease*, 92, 120-125.
- Bravo, C., D. Moshou, J. West, A. McCartney & H. Ramon (2003) Early Disease Detection in Wheat Fields using Spectral Reflectance. *Biosystems Engineering*, 84, 137-145.
- Breiman, L. (1996) Bagging predictors. *Machine Learning*, 24, 123-140.
- Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32.
- Bruzzone, L., F. Roli & S. B. and Serpico (1995) An Extension of the Jeffreys-Matusita Distance to Multiclass Cases for Feature Selection. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 1318-1321.
- Buzzell, R. I., T. W. Welacky & T. R. Anderson (1993) Soybean cultivar reaction and row width effect on Sclerotinia stem rot. *Canadian Journal of Plant Science*, 73.
- Byamukama, E., A. E. Robertson & F. W. Nutter (2010) Quantification of Temporal and Spatial Dynamics of Bean pod mottle virus at Different Spatial Scales. *Plant Health Progress*, 11.
- Byamukama, E., A. E. Robertson & F. W. Nutter, Jr. (2011) Quantifying the Within-Field Temporal and Spatial Dynamics of Bean pod mottle virus in Soybean. *Plant Dis*, 95, 126-136.

- Carmona, M., F. Sautua, S. Perelman, E. M. Reis & M. Gally (2011) Relationship between Late Soybean Diseases Complex and Rain in Determining Grain Yield Responses to Fungicide Applications. *Journal of Phytopathology*, 159, 687-693.
- Chappelle, E. W., M. S. Kim & J. E. McMurtrey (1992) Ratio Analysis of Reflectance Spectra (Rars) - an Algorithm for the Remote Estimation of the Concentrations of Chlorophyll-a, Chlorophyll-B, and Carotenoids in Soybean Leaves. *Remote Sensing of Environment*, 39, 239-247.
- Chen, X., J. Ma, H. Qiao, D. Cheng, Y. Xu & Y. Zhao (2007) Detecting infestation of take-all disease in wheat using Landsat Thematic Mapper imagery. *International Journal of Remote Sensing*, 28, 5183-5189.
- Crookston, R. K., J. E. Kurle, P. J. Copeland, J. H. Ford & W. E. Lueschen (1991) Rotational Cropping Sequence Affects Yield of Corn and Soybean. *Agronomy Journal*, 108-113.
- Danielson, G. A., B. D. Nelson & T. C. Helms (2004) Effect of Sclerotinia stem rot on yield of soybean inoculated at different growth stages. *Plant Disease*, 88, 297-300.
- Dann, E., B. Diers, J. Byrum & R. Hammerschmidt (1998) Effect of treating soybean with 2,6-dichloroisonicotinic acid (INA) and benzothiadiazole (BTH) on seed yields and the level of disease caused by *Sclerotinia sclerotiorum* in field and greenhouse studies. *European Journal of Plant Pathology*, 104, 271-278.
- Dardanelli, J. L., M. Balzarini, M. J. Martínez, M. Cuniberti, S. Resnik, S. F. Ramunda, R. Herrero & H. Baigorri (2006) Soybean Maturity Groups, Environments, and

Their Interaction Define Mega-environments for Seed Composition in Argentina.  
*Crop Science*, 46.

Dong, J., X. Xiao, M. A. Menarguez, G. Zhang, Y. Qin, D. Thau, C. Biradar & B. Moore, 3rd (2016) Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens Environ*, 185, 142-154.

Elbasiouny, H., M. Abowaly, A. Abu\_Alkheir & 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. *Catena*, 113, 70-78.

Emelyanova, I. V., T. R. McVicar, T. G. Van Niel, L. T. Li & A. I. J. M. 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.

ArcGIS Desktop. 10.5, Environmental Systems Research Institute.

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

Fall, M. L., J. F. Willbur, D. L. Smith, A. M. Byrne & M. I. Chilvers (2018) Spatiotemporal Distribution Pattern of *Sclerotinia sclerotiorum* Apothecia is Modulated by Canopy Closure and Soil Temperature in an Irrigated Soybean Field. *Plant Dis*, 102, 1794-1802.

Fan, C., B. Zheng, S. W. Myint & 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.
- Fitzgerald, G. J., S. J. Maas & W. R. Detar (2004) Spider Mite Detection and Canopy Component Mapping in Cotton Using Hyperspectral Imagery and Spectral Mixture Analysis. *Precision Agriculture*, 5.
- Franke, J. & G. Menz (2007) Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture*, 8, 161-172.
- Gao, F., M. C. Anderson, X. Zhang, Z. Yang, J. G. Alfieri, W. P. Kustas, R. Mueller, D. M. Johnson & 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.
- Gao, F., J. Masek, M. Schwaller & 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.
- George, T., P. W. Singleton & B. Ben (1988) Yield, Soil Nitrogen Uptake, and Nitrogen Fixation by Soybean from Four Maturity Groups Grown at Three Elevations. 80.
- Gevaert, C. M. & F. J. García-Haro (2015) A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sensing of Environment*, 156, 34-44.
- Grisham, M. P., R. M. Johnson & P. V. Zimba (2010) Detecting Sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes. *J Virol Methods*, 167, 140-5.

- Hartman, G. L., L. Kull & 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.
- Hershman, D. E., J. W. Hendrix, R. E. Stuckey & P. R. Bachi (1990) Influence of Planting Date and Cultivar on Soybean Sudden Death Syndrome in Kentucky. *Plant Disease*.
- Hilker, T., M. A. Wulder, N. C. Coops, J. Linke, G. McDermid, J. G. Masek, F. Gao & J. C. White (2009) A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113, 1613-1627.
- Hoffman, D. D., G. L. Hartman, D. S. Mueller, R. A. Leitz, C. D. Nickell & W. L. Pedersen (1998) Yield and seed quality of soybean cultivars infected with *Sclerotinia sclerotiorum*. *Plant Disease*, 82, 826-829.
- Huang, B. & H. Song (2012) Spatiotemporal Reflectance Fusion via Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*, 50.
- Huang, B. & 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.
- Huang, J. F. & A. Apan (2006) Detection of *Sclerotinia* rot disease on celery using hyperspectral data and partial least squares regression. *Journal of Spatial Science*, 51, 129-142.
- Ifarraguerri, A. & M. W. Prairie (2004) Visual Method for Spectral Band Selection. *Ieee Geoscience and Remote Sensing Letters*, 1, 101-106.

- Jaynes, D. B. & T. S. Colvin (1997) Spatiotemporal Variability of Corn and Soybean Yield. *Agronomy Journal*, 89.
- Jaynes, D. B., T. S. Colvin & T. C. Kaspar (2005) Identifying potential soybean management zones from multi-year yield data. *Computers and Electronics in Agriculture*, 46, 309-327.
- Jaynes, D. B., T. C. Kaspar, T. S. Colvin & D. E. James (2003) Cluster Analysis of Spatiotemporal Corn Yield Patterns in an Iowa Field. *Agronomy Journal*, 95.
- Jiang, P. & K. D. Thelen (2004) Effect of Soil and Topographic Properties on Crop Yield in a North-Central Corn–Soybean Cropping System. *Agronomy Journal*, 252-258.
- King, L., B. Adusei, S. V. Stehman, P. V. Potapov, X. P. Song, A. Krylov, C. Di Bella, T. R. Loveland, D. M. Johnson & M. C. Hansen (2017) A multi-resolution approach to national-scale cultivated area estimation of soybean. *Remote Sensing of Environment*, 195, 13-29.
- Kobayashi, T., E. Kanda, K. Kitada, K. Ishiguro & Y. Torigoe (2001) Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. *Phytopathology*, 91, 316-323.
- Kravchenko, A. N. & D. G. Bullock (2000) Correlation of Corn and Soybean Grain Yield with Topography and Soil Properties *Agronomy Journal* 75-83.
- Kumhálová, J., F. Kumhála, M. Kroulík & Š. Matějková (2011) The impact of topography on soil properties and yield and the effects of weather conditions. *Precision Agriculture*, 12, 813-830.

- Kurle, J. E., C. R. Gran, E. S. Oplinger & A. Mengistu (2001) Tillage, crop sequence, and cultivar effects on Sclerotinia stem rot incidence and yield in soybean. *Agronomy Journal*, 93, 973-982.
- Li, L., Y. Zhao, Y. Fu, Y. Pan, L. Yu & Q. Xin (2017) High Resolution Mapping of Cropping Cycles by Fusion of Landsat and MODIS Data. *Remote Sensing*, 9.
- Liaw, A. & M. Wiener (2002) Classification\_and\_Regression\_by\_RandomForest. *R News*, 2, 18-22.
- Liu, M., M. Wang, J. Wang & D. Li (2013) Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical*, 177, 970-980.
- Lobell, D. B. & G. P. Asner (2003) Climate and Management Contributions to Recent Trends in U.S. Agricultural Yields. *Science*, 1032.
- Lobell, D. B., W. Schlenker & J. Costa-Roberts (2011) Climate Trends and Global Crop Production Since 1980. *Science*, 616-620.
- Loladze, A., F. A. Rodrigues, Jr., F. Toledo, F. San Vicente, B. Gerard & M. P. Boddupalli (2019) Application of Remote Sensing for Phenotyping Tar Spot Complex Resistance in Maize. *Front Plant Sci*, 10, 552.
- Lowe, A., N. Harrison & A. P. French (2017) Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods*, 13, 1-12.

- Mahlein, A. K., T. Rumpf, P. Welke, H. W. Dehne, L. Plumer, U. Steiner & E. C. Oerke (2013) Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment*, 128, 21-30.
- Malthus, T. J. & A. C. Madeira (1993) High-Resolution Spectroradiometry - Spectral Reflectance of Field Bean-Leaves Infected by Botrytis-Fabae. *Remote Sensing of Environment*, 45, 107-116.
- Marcinkowska, J., J. W. Tomala-Bednarek & M. Schollenberger (1982) Soybean diseases in Poland. *Acta Agrobotanica*, 35, 213-224.
- Masuda, T. & P. D. Goldsmith (2009) World Soybean Production: Area Harvested, Yield, and Long-Term Projections. *International Food and Agribusiness Management Review*, 143-162.
- Mejia, M. N., C. A. Madramootoo & R. S. Broughton (2000) Influence of water table management on corn and soybean yields. *Agricultural Water Management*, 73-89.
- Mesonet, S. D. 2018. South Dakota Mesonet Database [database]. ed. S. D. S. University.
- Milas, A. L., A. L. Carriquiry, J. Zhao & X. B. Yang (2003) Impact of Management Practices on Prevalence of Soybean Sclerotinia Stem Rot in the North-Central United States and on Farmers' Decisions Under Uncertainty. *Plant Disease*, 87.
- Milne, G. M., J. Graham, A. Allen, A. Lahuerta-Marin, C. McCormick, E. Presho, R. Skuce & A. W. Byrne (2018) Spatiotemporal analysis of prolonged and recurrent bovine tuberculosis breakdowns in Northern Irish cattle herds reveals a new infection hotspot. *Spatial and Spatio-temporal Epidemiology*.

- 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.
- Moshou, D., C. Bravo, R. Oberti, J. West, L. Bodria, A. McCartney & H. Ramon (2005) Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps. *Real-Time Imaging*, 11, 75-83.
- Mueller, D. S., G. L. Hartman & W. L. Pedersen (2002) Effect of crop rotation and tillage system on sclerotinia stem rot on soybean. *Canadian Journal of Plant Pathology- Revue Canadienne De Phytopathologie*, 24, 450-456.
- Muhammed, H. H. (2005) Hyperspectral Crop Reflectance Data for characterising and estimating Fungal Disease Severity in Wheat. *Biosystems Engineering*, 91, 9-20.
- Naidu, R. A., E. M. Perry, F. J. Pierce & T. Mekuria (2009) The potential of spectral reflectance technique for the detection of Grapevine leafroll-associated virus-3 in two red-berried wine grape cultivars. *Computers and Electronics in Agriculture*, 66, 38-45.
- Nelson, K. A., K. A. Renner & R. Hammerschmidt (2002) Cultivar and herbicide selection affects soybean development and the incidence of *Sclerotinia* stem rot. *Agronomy Journal*, 94, 1270-1281.
- Nutter, F. W., G. L. Tylka, J. Guan, A. J. D. Moreira, C. C. Marett, T. R. Rosburg, J. P. Basart & C. S. Chong (2002) Use of remote sensing to detect soybean cyst nematode-induced plant stress. *Journal of Nematology*, 34, 222-231.
- Onojeghuo, A. O., G. A. Blackburn, Q. Wang, P. M. Atkinson, D. Kindred & Y. Miao (2018) Rice crop phenology mapping at high spatial and temporal resolution

- using downscaled MODIS time-series. *GIScience & Remote Sensing*, 55, 659-677.
- Pal, M. (2007) Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217-222.
- Peterson, T. A. & G. E. Varvel (1989) Crop Yield as Affected by Rotation and Nitrogen Rate. I. Soybean. *Agronomy Journal*, 727-731.
- Piper, E. L. & 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.
- Polischuk, V. P., T. M. Shadchina, T. I. Kompanetz, I. G. Budzanivskaya, A. L. Boyko & A. A. Sozinov (1997) Changes in reflectance spectrum characteristic of nicotiana debneyi plant under the influence of viral infection\*. *Archives Of Phytopathology And Plant Protection*, 31, 115-119.
- Press, S. J. & S. Wilson (1978) Choosing Between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association*, 73.
- Qin, Z. & M. Zhang (2005) Detection of rice sheath blight for in-season disease management using multispectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 7, 115-128.
- Raczko, E. & B. Zagajewski (2017) Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing*, 50, 144-154.
- Ray, D. K., N. D. Mueller, P. C. West & J. A. Foley (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. *Plos One*, 1-8.

- Rodriguez-Galiano, V., M. Sanchez-Castillo, M. Chica-Olmo & M. Chica-Rivas (2015) Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804-818.
- Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo & J. P. Rigol-Sanchez (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
- Roel, A. & R. E. Plant (2004) Factors Underlying Yield Variability in Two California Rice Fields. *Agronomy Journal*, 96.
- Rouse, J. W., J. H. Haas, J. A. Schell & D. W. Deering. 1973. Monitoring vegetation systems in the great plains with ERTS. In *Third ERTS Symposium*, 309-317.
- Rousseau, G. X., S. Rioux & D. Dostaler (2006) Multivariate effects of plant canopy, soil physico-chemistry and microbiology on Sclerotinia stem rot of soybean in relation to crop rotation and urban compost amendment. *Soil Biology & Biochemistry*, 38, 3325-3342.
- Roy, D. P., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen & 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.
- RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL <http://www.rstudio.com/>.

- Sara, S. A., E. B. McCallen & P. V. Switzer (2013) The spatial distribution of the Japanese beetle, *Popillia japonica*, in soybean fields. *J Insect Sci*, 13, 36.
- Schaaf, C. & Z. Wang. 2015. MCD43A4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006 [Data set]. In *NASA EOSDIS Land Processes DAAC*, ed. N. E. L. P. DAAC.
- Schapire, R. E., Y. Freund, P. Bartlett & W. S. Lee (1998) Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, 26, 1651-1686.
- Seiter, N. J., F. P. Reay-Jones & J. K. Greene (2013) Within-field spatial distribution of *Megacopta cribraria* (Hemiptera: Plataspidae) in soybean (Fabales: Fabaceae). *Environ Entomol*, 42, 1363-74.
- Soltani, A. & S. Askari (2017) Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, 48, 637-647.
- Song, X. P., P. V. Potapov, A. Krylov, L. King, C. M. Di Bella, A. Hudson, A. Khan, B. Adusei, S. V. Stehman & M. C. Hansen (2017) National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sensing of Environment*, 190, 383-395.
- Thanh Noi, P. & M. Kappas (2017) Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel)*, 18, 1-20.
- Thompson, D. R. & 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.

- USDA. 2010. Field Crops Usual Planting and Harvesting Dates. ed. U. S. D. o. Agriculture. National Agricultural Statistics Service.
- . 2017. National Agricultural Statistics Services. <https://www.nass.usda.gov/>: United States Department of Agriculture.
- Vieira, S. R. & A. P. Gonzalez (2003) Analysis of the spatial variability of crop yield and soil properties in small agricultural plots. *Bragantia, Campinas*, 62.
- Vigier, B. J., E. Pattey & 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 & F. Gao (2014) Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sensing of Environment*, 145, 55-67.
- Willbur, J. F., M. L. Fall, C. Bloomingdale, A. M. Byrne, S. A. Chapman, S. A. Isard, R. D. Magarey, M. M. McCaghey, B. D. Mueller, J. M. Russo, J. Schlegel, M. I. Chilvers, D. S. Mueller, M. Kabbage & D. L. Smith (2018) Weather-Based Models for Assessing the Risk of *Sclerotinia sclerotiorum* Apothecial Presence in Soybean (*Glycine max*) Fields. *Plant Dis*, 102, 73-84.
- Wrather, A. & S. Koenning (2009) Effects of Diseases on Soybean Yields in the United States 1996 to 2007. *Plant Health Progress*.
- Wrather, A., G. Shannon, R. Balardin, L. Carregal, R. Escobar, G. K. Gupta, Z. Ma, W. Morel, D. Ploper & A. Tenuta (2010) Effect of Diseases on Soybean Yield in the Top Eight Producing Countries in 2006. *Plant Health Progress*.

- Wu, M., Z. Niu, C. Wang, C. Wu & L. Wang (2012) Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing*, 6.
- Wu, M., C. Wu, W. Huang, Z. Niu, C. Wang, W. Li & 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.
- Yang, X. B. & F. Feng (2001) Ranges and diversity of soybean fungal diseases in north america. *Phytopathology*, 91, 769-75.
- Yang, X. B., P. Lundeen & M. D. Uphoff (1999) Soybean varietal response and yield loss caused by *Sclerotinia sclerotiorum*. *Plant Disease*, 83, 456-461.
- Yeh, Y.-H. F., W.-C. Chung, J.-Y. Liao, C.-L. Chung, Y.-F. Kuo & T.-T. Lin (2013) A Comparison of Machine Learning Methods on Hyperspectral Plant Disease Assessments. *IFAC Proceedings Volumes*, 46, 361-365.
- Zhang, H. K., B. Huang, M. Zhang, K. Cao & L. Yu (2015) A generalization of spatial and temporal fusion methods for remotely sensed surface parameters. *International Journal of Remote Sensing*, 36, 4411-4445.
- Zhang, X. Y., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. F. Hodges, F. Gao, B. C. Reed & A. Huete (2003) Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84, 471-475.
- Zhu, X., J. Chen, F. Gao, X. Chen & J. G. Masek (2010) An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114, 2610-2623.

- Zhu, X., E. H. Helmer, F. Gao, D. Liu, J. Chen & M. A. Lefsky (2016) A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment*, 172, 165-177.
- Zurita-Milla, R., J. Clevers & M. E. Schaepman (2008) Unmixing-Based Landsat TM and MERIS FR Data Fusion. *IEEE Geoscience and Remote Sensing Letters*, 5, 453-457.
- Zurita-Milla, R., G. Kaiser, J. G. P. W. Clevers, W. Schneider & M. E. Schaepman (2009) Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sensing of Environment*, 113, 1874-1885.