Mapping and Quantifying White Mold in Soybean across South Dakota Using Landsat Images

Confiance Mfuka[1*], Xiaoyang Zhang[2], Emmanuel Byamukama[1]

[1]Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, South Dakota, USA
[2]Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, South Dakota, USA
Email: *Confiance.Mfuka@jacks.sdstate.edu

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^2, 88 km^2, and 190 km^2 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

1. 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 [1], white mold was, more than a decade later, ranked in the top ten diseases that suppress soybean yields [2]. The apothecia 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˚C - 24˚C), wet and moist (enough rain: 70 - 120 hours of continuous wetness) conditions [3]. These conditions are favorable for optimal yield; therefore, incidence of white mold has been negatively correlated with yields [4] 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 [5].

The rationale behind the use of large scale imagery techniques is that they represent a fast, non-destructive method [6], and rely on biophysical characteristics that depend on the wavelength used for crop status monitoring. Malthus and Madeira [7] 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.* [8] 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.* [9] used multispectral radiometers and airborne multispectral scanner to identify the panicle blast rice. Qin and Zhang [10] collected ADAR (Airborne Data Acquisition and Registration) remote sensing images to map rice sheath blight. Further, Huang and Apan [11] used a portable spectroradiometer to collect hyperspectral data and detect Sclerotinia rot disease in celery. Naidu *et al.* [12] 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 [13] [14] and spectral indices are crucial in detecting and identifying plant diseases [14] [15] [16]. However, most of these studies required the use of portable spectro-radiometer or airborne remotely-sensed images, which represent costly resources and have reduced accessibility to common users and farmers.

While vegetation stress has received a lot of scientific attention, soybeans stress mapping has received little attention, and when it has, these studies focused either on other diseases than white mold [17], or in water stress [13] [18]. Vigier, Pattey and Strachan [14] 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 [19] [20], but these efforts have not addressed disease detection. In South Dakota which is one of the main soybean producing state, 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 [21] were used to extract spectral characteristics of soybean and white mold leading to mapping the spatial extent of the disease.

2. Materials and Methods

2.1. Study Area and Data Gathering

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 [22]. 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 1.

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 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 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
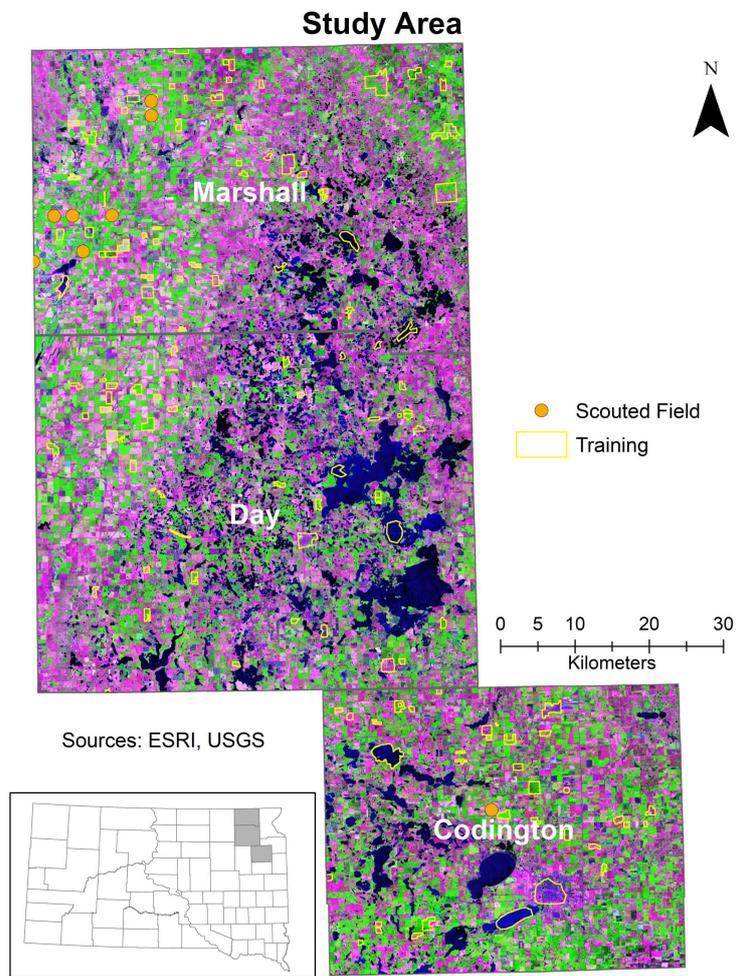
Study Area

Figure 1. Study area showing the three counties (Marshall, Day, and Codington) in Northeastern South-Dakota and the training polygons. The background image is a Landsat false color combination of bands 6-5-4 for July 14, 2017.

Table 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
	SWIR 1	B6_05	Band 1
	NIR	B5_05	Band 2
Landsat-8 May 11, 2017	RED	B4_05	Band 3
	GREEN	B3_05	Band 4
	BLUE	B2_05	Band 5
	SWIR 1	B6_07	Band 6
	NIR	B5_07	Band 7
Landsat-8 July 14, 2017	RED	B4_07	Band 8
	GREEN	B3_07	Band 9
	BLUE	B2_07	Band 10

for land cover mapping. This data also served in the comparison with our resulting land cover map.

2.2. Random Forest (RF) Classifiers for Mapping Soybean and White Mold

2.2.1. The Random Forest Algorithm for Image Classification

Methods that produce classifiers and aggregate their results have recently found many interests in the machine learning field [23]. The underlying principle is the same: based on a set of trainings used to extract spectral characteristics of different defined classes, these non-parametric classifiers (meaning that they require no statistical assumptions such as the normal distribution of the input dataset), build models that decide to which class to affect each observation. 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 [24], and bagging, in which successive trees are independent from earlier trees [25]. 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 [23].

The RF algorithm [21] 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 [23]. In the remote sensing field, especially in image or land cover classification, RF has shown to perform equally to Support Vector Machine (SVM) [26] [27] or to outperform Decision Tree (DT) [28]. Other studies have shown that RF outperformed SVM in term of robustness and stability [29] and in terms of accuracy [30]. The RF is preferred in our study because it can deal with classification problems of unbalanced, multiclass and small sample data [31]. In fact, when collecting training data, some classes may require more training than others 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.

2.2.2. 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. 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, and 3) 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 [32] [33]. 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 ω_i and ω_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.

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

where x is the feature vector of dimension k and $p\left(x|\omega_i\right)$ and $p\left(x|\omega_j\right)$ 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 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.

2.2.3. White Mold Mapping and 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 [34] 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)).

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{2}$$

$$TA = N \times A \tag{3}$$

where *TA* is the Total Area, *N* is the number of pixels, and *A* is the area of a pixel (30 m × 30 m).

3. Results and Discussion

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

The original input Landsat bands have been stacked in a color composite image combining both May and July bands. The corresponding output bands designations are listed in Table 1. Figure 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).

### 3.2. 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 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

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

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

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
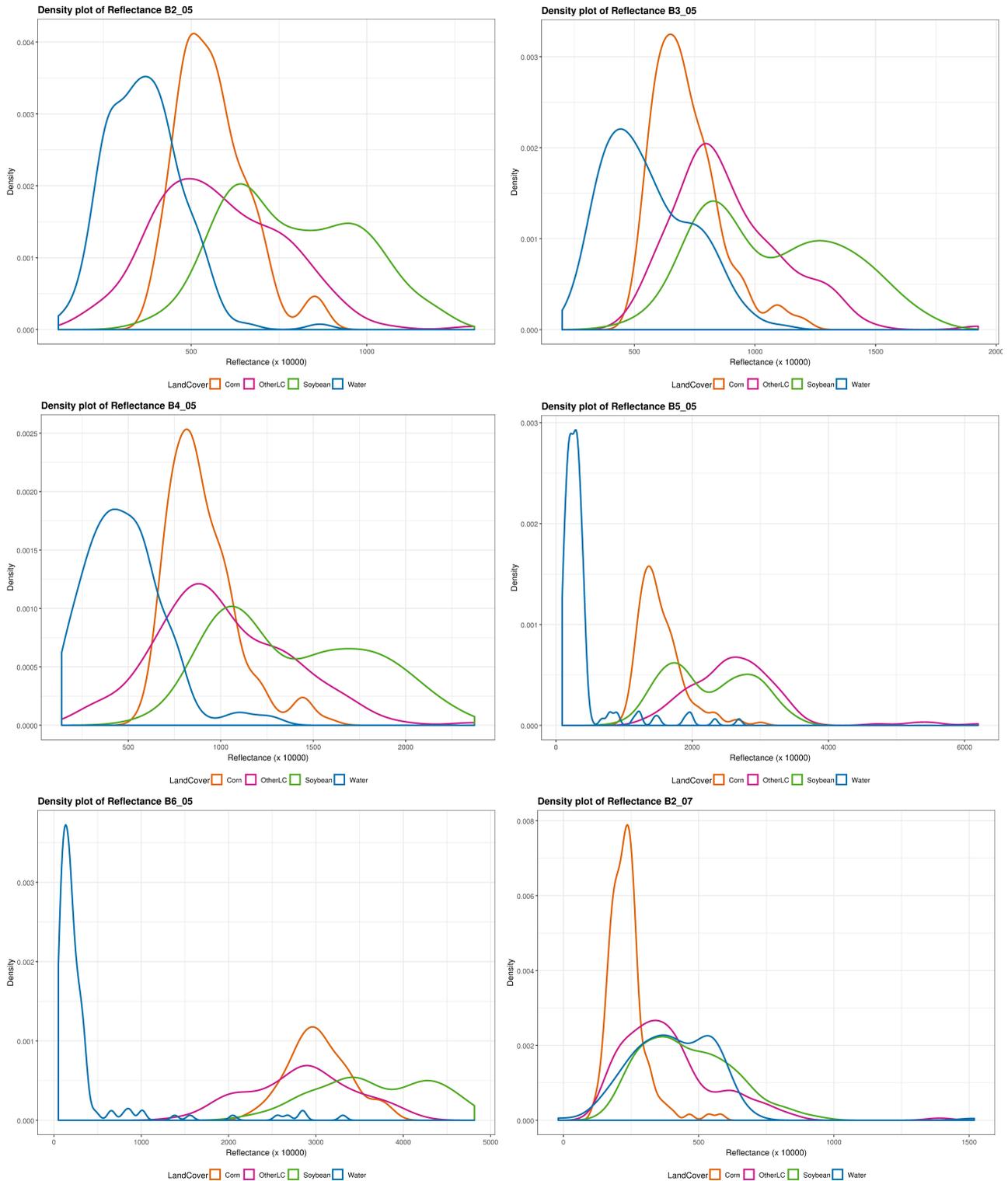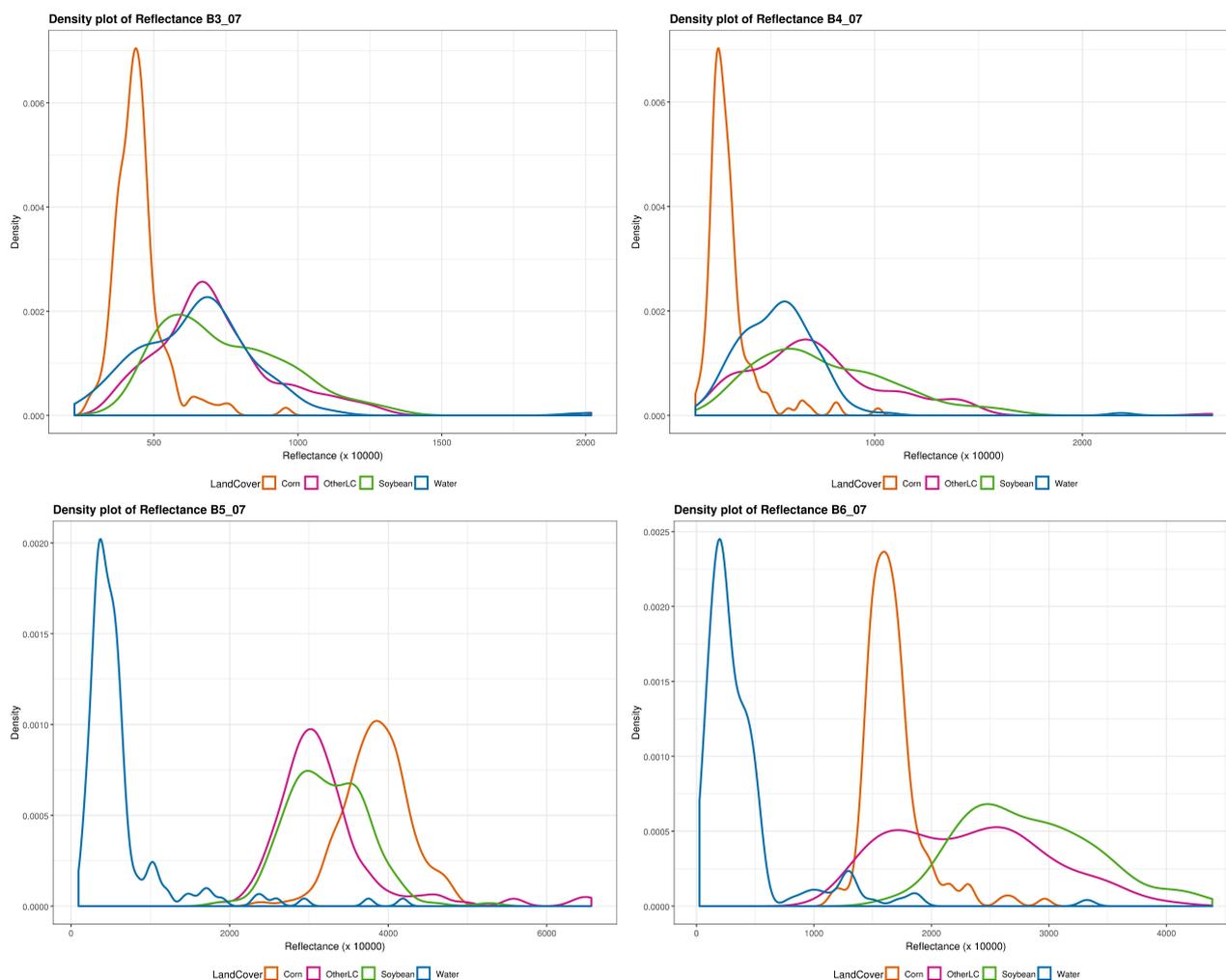
Figure 2. Classes spectral separability shown by the density plot of each band reflectance. The Near Infrared (NIR) band is very good in discriminating Water in both images (B5_05 and B5_07), while Other Land Cover can be distinguished using the July green band (B3_07); Corn is however well distinguished using the May NIR band; the shortwave infrared however is the remaining band susceptible to separate soybean, when the above mentioned classes are successfully extracted.

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.

3.3. Land Cover Map Accuracy Assessment

The accuracy of the resulting classification map was assessed using the confusion matrix (Table 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 3 reports the individual class accuracies as well as the errors. The commission and omission errors are reported as
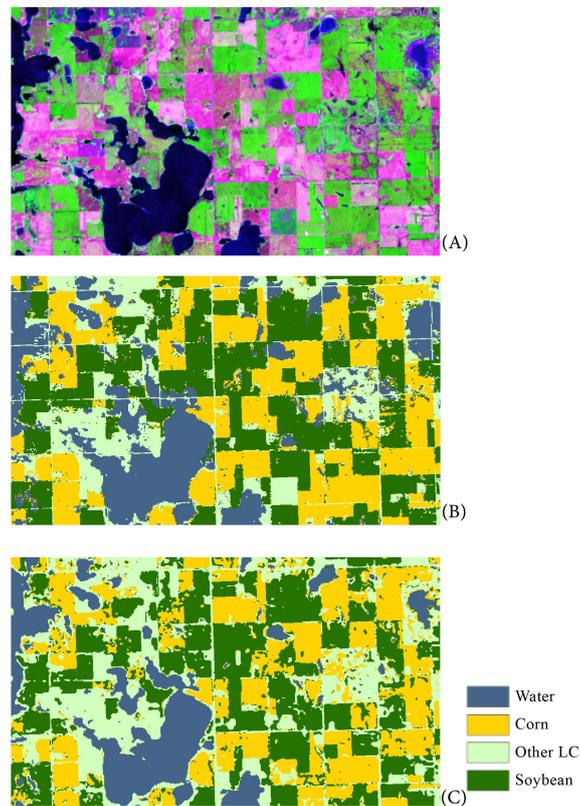
Figure 3. A comparison between a July 14 false color 6-5-4 Landsat 8 image composite (A), the Crop Data Layer (CDL) map (B) and the resulting classification (C) of the stacked May and July images. The classification image is very similar to the CDL; the classified map can clearly delineate soybean, especially in the roads between the fields.

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

Classification	Ground Truth				Commission Error	Producer's Accuracy
	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
Omission Error	0.01	0.05	0.05	0.07		
User's Accuracy	0.99	0.95	0.95	0.93		

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).
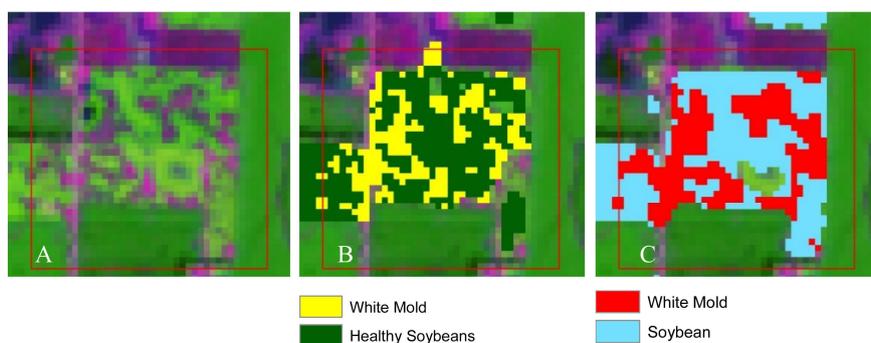
3.4. White Mold Mapping

Figure 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.

3.5. White Mold Map Accuracy Assessment

The accuracy of the resulting white mold map was assessed using the 30% set-apart samples that have not been used in the model building. The map achieved an overall accuracy of 99%. Table 4 reports accuracy and the commission/omission errors of the resulting white mold map. White mold is mapped with high accuracy (99%). These results can be explained by the quality of the independent variables that not only use individual bands, but also includes the NDVI in the modeling. However, this accuracy depends largely on the set-apart pixels used for the validation process, and considered as ground truth. Unfortunately, one limitation of the RF which is known as the black box, is that it cannot provide the contribution of each variable in the model. More importantly, we checked the known fields that were affected by white mold and all of them were correctly mapped. The resulting final white mold map is shown in Figure 5, as well as the classified Landsat images and the fields locations.

White Mold
Healthy Soybeans
White Mold
Soybean

Figure 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.

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

	Ground Truth			
Class	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		
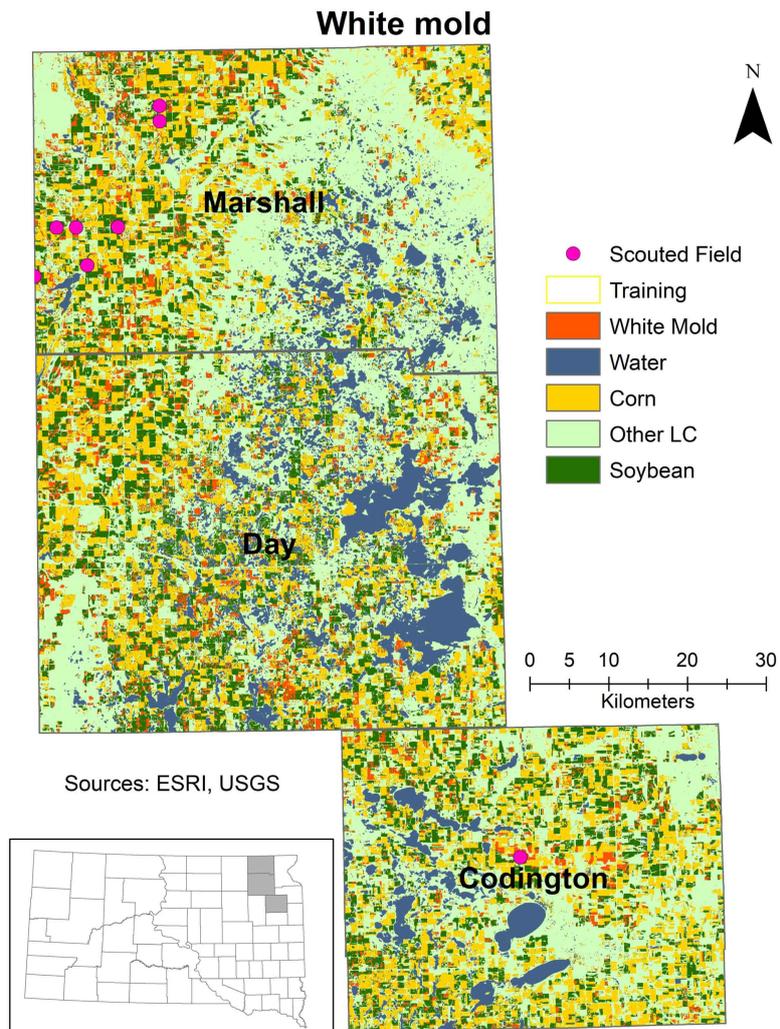
Figure 5. White mold in northeastern South Dakota: the map shows a classified image in background with the four important classes and the quantified white mold over the soybean mask.

3.6. Quantified Soybean and White Mold

Using the Landsat pixel size (30 m × 30 m), we estimated the total area of the classified soybean in the three counties based on the total number of pixels mapped. **Table 5** reports the total soybean areas estimation from both the classification

Table 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 (30 m × 30 m) and the total number of pixels.

County Name	County Area (km²)	Soybean (km²)	Soybean (USDA-km²)	White Mold (km²)
Marshall	2294.16	426.59	427.35	131.96
Codington	1856.90	403.70	439.89□	88.37
Day	2824.50	653.22	660.04□	189.93

and the USDA report [35], as well as the estimated white mold areas per county. The USDA estimated areas reported consist of the harvested statistics, but the values are very similar to those obtained by the classified Landsat images. The white mold area estimates are respectively 132 km², 88 km², and 190 km², and represent 31%, 22% and 29% of the total soybean area for Marshall, Codington and Day counties.

4. Conclusions

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 disease patches. 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 in better disease quantification.

Acknowledgements

#SD00H662-18 and South Dakota State University Agriculture Experiment Station. We wish to thank the anonymous reviewers for their valuable and useful suggestions and comments on the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1]   Marcinkowska, J., Tomala-Bednarek, J.W. and Schollenberger, M. (1982) Soybean Diseases in Poland. *Acta Agrobotanica*, **35**, 213-224. https://doi.org/10.5586/aa.1982.021

[2]   Wrather, A. and Koenning, S. (2009) Effects of Diseases on Soybean Yields in the United States 1996 to 2007. Plant Health Progress. https://doi.org/10.1094/PHP-2009-0401-01-RS

[3]   Boland, G.J. and Hall, R. (1988) Epidemiology of Sclerotinia Stem Rot of Soybean in Ontario. *Phytopathology*, **78**, 1241-1245. https://doi.org/10.1094/Phyto-78-1241

[4]   Hoffman, D.D., Hartman, G.L., Mueller, D.S., Leitz, R.A., Nickell, C.D. and Pedersen, W.L. (1998) Yield and Seed Quality of Soybean Cultivars Infected with *Sclerotinia sclerotiorum. Plant Disease*, **82**, 826-829. https://doi.org/10.1094/PDIS.1998.82.7.826

[5]   Lowe, A., Harrison, N. and French, A.P. (2017) Hyperspectral Image Analysis Techniques for the Detection and Classification of the Early Onset of Plant Disease and Stress. *Plant Methods*, **13**, 80. https://doi.org/10.1186/s13007-017-0233-z

[6]   Yeh, Y.-H.F., Chung, W.-C., Liao, J.-Y., Chung, C.-L., Kuo, Y.-F. and Lin, T.-T. (2013) A Comparison of Machine Learning Methods on Hyperspectral Plant Disease Assessments. *IFAC Proceedings Volumes*, **46**, 361-365. https://doi.org/10.3182/20130327-3-JP-3017.00081

[7]   Malthus, T.J. and Madeira, A.C. (1993) High-Resolution Spectroradiometry—Spectral Reflectance of Field Bean-Leaves Infected by Botrytis-Fabae. *Remote Sensing of Environment*, **45**, 107-116. https://doi.org/10.1016/0034-4257(93)90086-D

[8]   Polischuk, V.P., Shadchina, T.M., Kompanetz, T.I., Budzanivskaya, I.G., Boyko, A.L. and Sozinov, A.A. (1997) Changes in Reflectance Spectrum Characteristic of *Nicotiana debneyiplant* under the Influence of Viral Infection. *Archives of Phytopathology and Plant Protection*, **31**, 115-119. https://doi.org/10.1080/03235409709383221

[9]   Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K. and Torigoe, Y. (2001) Detection of Rice Panicle Blast with Multispectral Radiometer and the Potential of Using Airborne Multispectral Scanners. *Phytopathology*, **91**, 316-323. https://doi.org/10.1094/PHYTO.2001.91.3.316

[10] Qin, Z. and Zhang, M. (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. https://doi.org/10.1016/j.jag.2005.03.004

[11] Huang, J.F. and Apan, A. (2006) Detection of Sclerotinia Rot Disease on Celery Using Hyperspectral Data and Partial Least Squares Regression. *Journal of Spatial Science*, **51**, 129-142. https://doi.org/10.1080/14498596.2006.9635087

[12] Naidu, R.A., Perry, E.M., Pierce, F.J. and Mekuria, T. (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. https://doi.org/10.1016/j.compag.2008.11.007

[13] Behmann, J., Steinrücken, J. and Plümer, L. (2014) Detection of Early Plant Stress Responses in Hyperspectral Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **93**, 98-111. https://doi.org/10.1016/j.isprsjprs.2014.03.016

[14] Vigier, B.J., Pattey, E. and Strachan, I.B. (2004) Narrowband Vegetation Indexes and Detection of Disease Damage in Soybeans. *IEEE Geoscience and Remote Sensing Letters*, **1**, 255-259. https://doi.org/10.1109/LGRS.2004.833776

[15] Chappelle, E.W., Kim, M.S. and Mcmurtrey, J.E. (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. https://doi.org/10.1016/0034-4257(92)90089-3

[16] Mahlein, A.K., Rumpf, T., Welke, P., Dehne, H.W., Plümer, L., Steiner, U. and Oerke, E.C. (2013) Development of Spectral Indices for Detecting and Identifying Plant Diseases. *Remote Sensing of Environment*, **128**, 21-30. https://doi.org/10.1016/j.rse.2012.09.019

[17] Bajwa, S., Rupe, J. and Mason, J. (2017) Soybean Disease Monitoring with Leaf Reflectance. *Remote Sensing*, **9**, 127. https://doi.org/10.3390/rs9020127

[18] Thompson, D.R. and Wehmanen, O.A. (1980) Using Landsat Digital Data to Detect Moisture Stress in Corn-Soybean Growing Regions. *Photogrammetric Engineering and Remote Sensing*, **46**, 1087-1093.

[19] Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V. and Hansen, M.C. (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. https://doi.org/10.1016/j.rse.2017.01.008

[20] King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M. and Hansen, M.C. (2017) A Multi-Resolution Approach to National-Scale Cultivated Area Estimation of Soybean. *Remote Sensing of Environment*, **195**, 13-29. https://doi.org/10.1016/j.rse.2017.03.047

[21] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/A:1010933404324

[22] USDA (2010) Field Crops Usual Planting and Harvesting Dates. United States Department of Agriculture, National Agricultural Statistics Service.

[23] Liaw, A. and Wiener, M. (2002) Classification and Regression by RandomForest. *R News*, **2**, 18-22.

[24] Schapire, R.E., Freund, Y., Bartlett, P. and Lee, W.S. (1998) Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. *Annals of Statistics*, **26,** 1651-1686. https://doi.org/10.1214/aos/1024691352

[25] Breiman, L. (1996) Bagging Predictors. *Machine Learning*, **24**, 123-140. https://doi.org/10.1007/BF00058655

[26] Pal, M. (2007) Random Forest Classifier for Remote Sensing Classification. *International Journal of Remote Sensing*, **26**, 217-222. https://doi.org/10.1080/01431160412331269698

[27] Thanh Noi, P. and Kappas, M. (2017) Comparison of Random Forest, k-Nearest

Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* (*Basel*), **18**, 18.
https://doi.org/10.3390/s18010018

[28] Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. (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. https://doi.org/10.1016/j.isprsjprs.2011.11.002

[29] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M. (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.
https://doi.org/10.1016/j.oregeorev.2015.01.001

[30] Adam, E., Mutanga, O., Odindi, J. and Abdel-Rahman, E.M. (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.
https://doi.org/10.1080/01431161.2014.903435

[31] Liu, M., Wang, M., Wang, J. and Li, D. (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.
https://doi.org/10.1016/j.snb.2012.11.071

[32] Bruzzone, L., Roli, F. and Serpico, S.B. (1995) An Extension of the Jeffreys-Matusita Distance to Multielass Cases for Feature Selection. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 1318-1321. https://doi.org/10.1109/36.477187

[33] Ifarraguerri, A. and Prairie, M.W. (2004) Visual Method for Spectral Band Selection. *Ieee Geoscience and Remote Sensing Letters*, **1**, 101-106.
https://doi.org/10.1109/LGRS.2003.822879

[34] Rouse Jr., J.W., Haas, J.H., Schell J.A. and Deering, D.W. (1973) Monitoring Vegetation Systems in the Great Plains with ERTS. 3*rd ERTS Symposium*, Texas A&M University, College Station, TX, 1 January 1974, 309-317.

[35] USDA (2017) National Agricultural Statistics Services. United States Department of Agriculture. https://www.nass.usda.gov/