

Research Article

The Impact of Simulated Spectral Noise on Random Forest and Oblique Random Forest Classification Performance

Na'eem Hoosen Agjee , Onesimo Mutanga, Kabir Peerbhay, and Riyad Ismail

School of Agriculture, Earth and Environmental Sciences, University of KwaZulu-Natal, Scottsville P/Bag X01, Pietermaritzburg 3209, South Africa

Correspondence should be addressed to Na'eem Hoosen Agjee; agjeen2@gmail.com

Received 30 August 2017; Accepted 9 January 2018; Published 13 March 2018

Academic Editor: Javier Garcia-Guinea

Copyright © 2018 Na'eem Hoosen Agjee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyperspectral datasets contain spectral noise, the presence of which adversely affects the classifier performance to generalize accurately. Despite machine learning algorithms being regarded as robust classifiers that generalize well under unfavourable noisy conditions, the extent of this is poorly understood. This study aimed to evaluate the influence of simulated spectral noise (10%, 20%, and 30%) on random forest (RF) and oblique random forest (oRF) classification performance using two node-splitting models (ridge regression (RR) and support vector machines (SVM)) to discriminate healthy and low infested water hyacinth plants. Results from this study showed that RF was slightly influenced by simulated noise with classification accuracies decreasing for week one and week two with the addition of 30% noise. In comparison to RF, oRF-RR and oRF-SVM yielded higher test accuracies (oRF-RR: 5.36%–7.15%; oRF-SVM: 3.58%–5.36%) and test kappa coefficients (oRF-RR: 10.72%–14.29%; oRF-SVM: 7.15%–10.72%). Notably, oRF-RR test accuracies and kappa coefficients remained consistent irrespective of simulated noise level for week one and week two while similar results were achieved for week three using oRF-SVM. Overall, this study has demonstrated that oRF-RR can be regarded a robust classification algorithm that is not influenced by noisy spectral conditions.

1. Introduction

Hyperspectral sensors capture detailed spectral information; however, they are often sensitive to noise from a range of different sources [1, 2]. The presence of spectral noise in hyperspectral data is of critical concern in classification approaches due to the performance of the classifier deteriorating with the presence of noise [3, 4]. Poor classification performance is unfavourable because it hinders the interpretability of models constructed and negatively influences the decisions gleaned from them [4]. Despite numerous researchers recognising the detrimental effects of noise on classifier performance, it is a concept that is poorly explored and understood [3–5]. Classifying hyperspectral data in a controlled environment, that is in the presence and absence of simulated noise, is essential to understanding and evaluating the impact of noise on classifier performance. In addition, classifying hyperspectral data under increasing levels of simulated noise is critical to evaluating if the deterioration

in classification performance is considerable or within acceptable limits. While significant attention has been placed on the quality of the training data, maximum accuracy can only be achieved through a combination of both high-quality training data and the implementation of a robust classification algorithm that is insensitive to noise [3, 4, 6]. Consequently, appropriate denoising and classification techniques should be explored to effectively deal with noisy data and enhance classification performance.

Over the past decade, several researchers have developed and implemented novel techniques to effectively minimize the effects of noise in hyperspectral data [7–11]. Denoising techniques implemented include minimum noise fraction transformation, noise-filtering methods, spectral smoothing algorithms, wavelet transformations, and a range of novel denoising algorithms [9–13]. Contrary to most researchers that advocate denoising as an important preprocessing step [1, 8, 11], denoising techniques are also limited in their approaches. For example, estimating the number of noisy

components is a challenge for the minimum noise fraction transformation. This estimation is different for bands with different signal to noise ratios as it is not possible to achieve an optimal denoising for all the bands at the same time [9]. While noise-filtering methods remove noisy attributes, they can also remove valuable spectral information that can be essential for specific applications [2]. Even though most denoising algorithms take into consideration some prior knowledge about the noise and remove one or two types of noise, real-world datasets incorporate a combination of noises [10]. Consequently, given these limitations, it is better suited to explore classification algorithms that do not require complex preprocessing denoising steps and are robust in dealing with spectral noise. This study attempts to classify hyperspectral data under increasing levels of simulated noise to evaluate the robustness of classification algorithms. One machine learning algorithm that is regarded as being a robust classifier that incorporates mechanisms to be less influenced by noise is random forest (RF).

RF is a popular tree-based ensemble classification algorithm that uses many classification trees to classify unknown samples [14–16]. For a detailed review on the applications and future directions of RF in remote sensing, see [17]. RF offers numerous advantages as a robust classification algorithm; however, one of the primary features of the algorithm is that it is less sensitive to noise and avoids overfitting [14]. Firstly, RF is able to efficiently process noisy data through randomization by constructing each decision tree in the ensemble using bootstrap samples which are created by resampling the original training dataset with replacement [14]. Secondly, RF selects a random subset of predictor variables at each node to grow each tree; therefore, it is most likely that individual decision trees avoid noise contributing input records and predictor variables [14]. Thirdly, RF learns many variable unbiased decision trees. Models that overfit model the noise in the training dataset and tend to have a high variance and low bias. RF reduces the variance by applying the majority vote rule, therefore producing a model that has a low bias and low variance [14]. Given the above, it is evident that RF could be well suited to deal with noisy data which can be associated with remotely sensed data. However, while RF can process noisy data efficiently, splitting the feature space using univariate hyperplanes that are orthogonal might be suboptimal for class separation and classifying spectral data [18]. In instances where collinearity exists (i.e., hyperspectral data), the marginal distributions of the input variables may lose their power to separate classes requiring complex decision boundaries and deeply nested trees [18]. Consequently, an algorithm such as the oblique random forest (oRF), that uses multivariate hyperplanes that are oblique, might be better suited for tasks when dealing with noisy data, thus offering better classification performances.

oRF is a novel classification algorithm that is an improvement over the standard RF algorithm. The algorithm employs the same underlying principles to effectively process noisy data as the standard RF algorithm; making it well suited to process remotely sensed data. The primary difference between RF and oRF is that oRF splits the feature space by using multivariate hyperplanes that are oblique [18, 19]. In

addition, oRF uses supervised linear models, for example ridge regression (RR) and support vector machines (SVMs), to perform multivariate node splitting at each node, thereby providing individual classifiers that are stronger in the standard forest and an overall improvement in classification accuracies [19]. Despite the numerous benefits associated with oRF, only a few studies have implemented oRF for classification tasks [15, 18–20]. For example, Bassa et al. [15] reported high classification accuracies when implementing oRF to classify the highly heterogeneous iSimangaliso wetland park, South Africa. Poona et al. [20] showed that oRF using three different node-splitting models outperformed traditional RF classifiers when classifying healthy and infected *Pinus radiata* seedlings using spectroscopic data. It is evident that the benefits of implementing oRF within a remote sensing context offer the potential to process noisy data efficiently as well as classify more effectively than the standard RF.

Water hyacinth (*Eichhornia crassipes*) is an exotic macrophyte that is typically controlled by two host-specific weevil species, the chevroned water hyacinth weevil (*Neochetina bruchi*) and the mottled water hyacinth weevil (*Neochetina eichhorniae*) [21–23]. To establish the efficacy of biocontrol agents, variable infestation levels can be monitored using hyperspectral remote sensing [24]. However, remotely detecting low infestation levels is a challenge because water hyacinth plants occur in an aquatic environment that is influenced by noise from a range of different sources, the effect of which may mask the detection of low infestation levels. To the authors' knowledge, no studies have attempted to detect low infestation levels on water hyacinth plants under noisy hyperspectral conditions. The robust nature of RF and oRF offers the potential to detect low infestation levels on water hyacinth under noisy spectral conditions.

Considering the above, the aim of this study was to determine the impact of hyperspectral spectral noise on RF and oRF classification performance. More specifically, the objectives of this study were (1) to discriminate healthy and low infested water hyacinth plants under simulated noise conditions, (2) to understand the effect of increasing levels (10%, 20%, and 30%) of simulated noise on classification accuracies, (3) to compare RF and oRF classification accuracies and robustness under simulated noise conditions, and (4) to compare oRF classification accuracies and robustness using RR and SVM node-splitting models.

2. Materials and Methods

2.1. Experimental Procedure. Healthy water hyacinth (*Eichhornia crassipes*) plants ($n=90$) were collected from the Amanzimtoti River located in the KwaZulu-Natal province of South Africa ($30^{\circ} 03' 29.44''$ S; $30^{\circ} 52' 38.53''$ E) and were transported to a laboratory at the University KwaZulu-Natal [24]. Water hyacinth plants that were collected were of the same size (i.e., phenostage five) [25] as well as free of any biocontrol agents or biocontrol damage. At the laboratory, individual circular plastic containers, 55 cm in diameter each, were filled with 20 L of water [24]. Thereafter, nitrogen

(potassium nitrate: 7.5 mg N L^{-1}) and phosphorus (dihydrogen orthophosphate: 1.37 mg P L^{-1}) were added to each container to simulate conditions found within highly eutrophic environments [26]. Commercial iron chelate (13% Fe) was also added to each container at a concentration of $11.2 \text{ mg Fe L}^{-1}$ [27, 28]. The nutrient medium was replaced on a weekly basis to maintain a constant nutrient concentration for plant growth and development [24]. Fifteen water hyacinth plants were placed in each container creating a dense mat within each container [24]. Water hyacinth plants were then acclimated to the surrounding environment for one week prior to weevil exposure [29].

After the acclimation period, each water hyacinth plant was cleaned of all debris by removing all dead leaves, dead petioles, and daughter plants to maintain the original stocking density [24]. Two adult male weevils per plant were considered in order to study the effect of low infestation levels on plant spectral characteristics [30]. The experiment was set up in a complete random design with one control and one treatment; each replicated three times [24]. After the weevils were introduced, each container was covered with a mesh ($3 \text{ m} \times 1.5 \text{ m}$; $2 \text{ mm} \times 2 \text{ mm}$ mesh cell size) to prevent weevils from escaping [24]. The number of male weevils in each container was maintained by replacing dead weevils on a weekly basis [31]. Water hyacinth plants were exposed to one week of weevil herbivory prior to plants being sampled for canopy reflectance spectra [24].

2.2. Canopy Reflectance Measurements. A FieldSpec® 3 spectroradiometer [32] was used to collect reflectance data of weevil herbivory. The ASD is a portable spectrometer that uses a fibre optic cable for reflectance measurements and a personal computer for data logging. The spectrometer has a spectral range of 350–2500 nm with a sampling interval of 1.4 nm in the 350–1000 nm range and 2 nm in the 1000–2500 nm range. Reflectance measurements were taken at an ambient air temperature of 21°C [24]. All reflectance measurements were taken within a black box to account for any background reflectance [24]. The fibre optic cable with a 10° field of view was pointed 0.5 m above each container with one 50 W halogen lamp across the container providing the only illumination [24]. The spectrometer was calibrated by measuring a “white reference” reading using a Spectralon panel before sample reflectance measurements were taken.

To measure canopy reflectance, the mesh from each container was removed and the container was placed on the target platform. Each container was rotated 45° eight times with reflectance measurements being captured at the centre of each container [24]. Four reflectance measurements were captured at each rotation of the container totalling 32 per container and 96 per treatment. After reflectance measurements were captured, the mesh on each container was replaced [24]. The first set of reflectance measurements were taken after one week of infestation, thereafter, on a weekly basis over a period of three weeks [24]. Prior to analysis, reflectance spectra captured at each rotation were averaged and atmospheric water absorption bands (1350–1450 nm; 1773–2020 nm; 2400–2500 nm) were removed [24].

2.3. Noise Simulation. A noise simulation algorithm was developed to add white noise to the reflectance spectra of 10%, 20%, and 30% of the training sample. The algorithm generates normally distributed random numbers that are added to the spectra achieving a signal to noise ratio of one. The noise simulation algorithm was developed in the R statistics package version 2.0.0 [33].

2.4. Statistical Analysis

2.4.1. The Random Forest Algorithm. RF is a classification algorithm that combines several univariate classification trees to build an ensemble that uses the entire forest as a complex composite classifier [14]. Random bootstrap samples are drawn from the original training data set with replacement, thereafter an unpruned classification tree is built for each bootstrap sample. To build each tree, RF randomly selects a subset of candidate predictors at each node and computes the Gini index for all possible orthogonal splits. The best orthogonal split, that is, the largest Gini index measure, is used to partition the data and generate child nodes. The final classification label of a new observation is decided by aggregating the predictions of the trees in the ensemble through majority voting [14]. In this study, the default hyperparameters were used for classification procedures as they produce acceptable results [34]. The randomForest software library [35] developed in the R statistics package version 2.5.1 [33] was used for all analysis.

2.4.2. The Oblique Random Forest Algorithm. oRF is a tree-based ensemble classifier similar to RF which grows many multivariate classification trees to a data set; combining the predictions of all the trees to classify the input data. oRF employs the same ensemble creating process (i.e., bootstrap aggregation and random selection of bands at each node) as RF [20]. However, oRF is differentiated from the standard RF algorithm by learning optimal oblique split directions using linear discriminative models at each node [18, 19]. Linear discriminative models implemented at each node include RR, SVM's, partial least squares regression and the logistic regression. In this study, two linear discriminative models were implemented, RR and SVM. These models are discussed in greater detail below.

RR is a multiple linear regression model that finds the minimum sum of squared prediction errors while limiting the sum of squares of the regression coefficients [36, 37]. RR employs a regularization parameter (λ) on the regression coefficients which is optimized using the OOB samples occurring at the same node. It is the regularization parameter that allows the classifier to adjust to an optimal split direction [18].

SVM is a binary supervised classification model that aims to find the optimal separating hyperplane in multidimensional space that linearly separates two classes by maximizing the margin and simultaneously minimizing misclassification errors [38, 39]. The trade-off between misclassification errors and margin maximization is controlled by the parameter C , which is optimized using the OOB samples occurring at the same node.

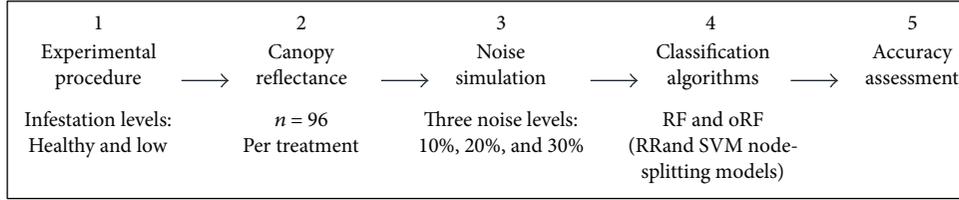


FIGURE 1: Flowchart outlining the process used to assess the impact of spectral noise on random forest and oblique random forest classification performance for three weeks.

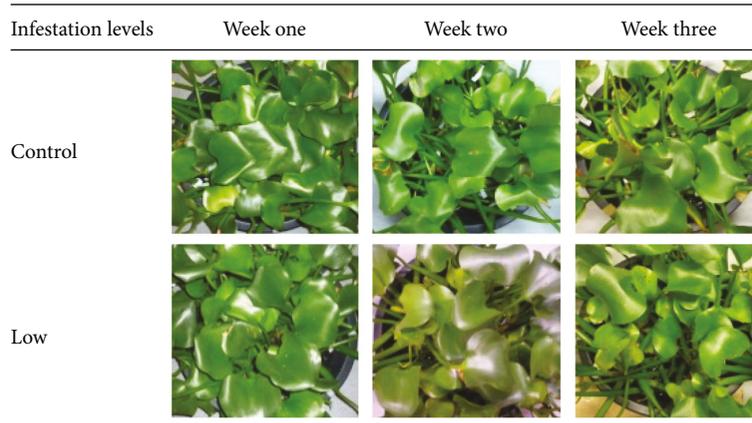


FIGURE 2: Photographs of healthy and low infested water hyacinth (*E. crassipes*) plants over three weeks of infestation (Adapted from Agjee et al. [24]).

Similar to the RF, oRF is also defined by two hyperparameters, n_{tree} and m_{try} . The default n_{tree} and m_{try} hyperparameters as determined by the oRF algorithm were used during classification. The obliqueRF library [40] developed in the R statistics package version 2.0.0 [33] was used for all analysis.

2.4.3. Accuracy Assessment. The classification performance of the RF and oRF classifiers was evaluated by calculating the overall accuracy and kappa coefficient. The overall accuracy is a measure of the entire classification, whereas the kappa coefficient is a measure of the difference between the training dataset and the test dataset arising due to chance [41, 42].

The workflow used to assess the impact of spectral noise on random forest and oblique random forest classification performance is presented in Figure 1.

3. Results

3.1. Description of *Neochetina* spp. Weevil Infestations. Healthy and low infested water hyacinth plants over three weeks of infestation is illustrated in Figure 2 graphically illustrates healthy and low infested water hyacinth plants over three weeks of infestation. Low infested water hyacinth plants were observed as vigorous and healthy after three weeks of infestation (Figure 2). A closer inspection of the low infested water hyacinth plants revealed that *Neochetina* spp. weevils were actively feeding from week one to week three. Feeding scar damage was observed on the leaves of the plants and progressively increased from week one to week three.

3.2. Reflectance Spectra of Healthy and Low Infestation Levels. Figure 3 documents the average spectral reflectance curves for healthy and low infested water hyacinth plants after three weeks of weevil infestation. Generally, water hyacinth plants with low infestation exhibited a slightly lower reflectance curve than the healthy water hyacinth plants.

3.3. Random Forest Classification Accuracies. Table 1 presents RF classification accuracies exclusive and inclusive of three simulated noise levels for three weeks of infestation. The highest RF test accuracies and test kappa coefficients were achieved for week one (test accuracy=91.07%; test kappa coefficient=82.14%) and week two (test accuracy=94.64%; test kappa coefficient=89.28%), excluding the addition of simulated noise. It is evident that RF is slightly influenced by noise due to classification accuracies slightly decreased with increasing levels of simulated noise. RF test accuracies decreased by 7.15% for week one and 10.72% for week two with the addition of 30% simulated noise. Similarly, RF test kappa coefficients decreased by 14.29% for week one and week two with the addition of 30% simulated noise. However, RF test accuracies and test kappa coefficients increased slightly and remained consistent with increasing levels of simulated noise for week three. In this regard, RF test accuracies and test kappa coefficients increased by 1.78% and 3.57%, respectively.

3.4. Oblique Random Forest-Ridge Regression Classification Accuracies. oRF-RR classification accuracies exclusive and inclusive of three simulated noise levels for three weeks of infestation are presented in Table 2. Overall, oRF-RR yielded

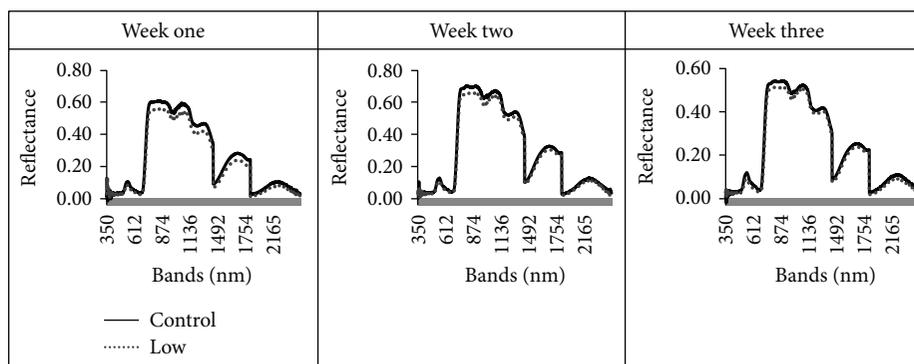


FIGURE 3: Reflectance spectra of healthy and low infested water hyacinth plants over three weeks of infestation (Adapted from Agjee et al. [24]).

TABLE 1: RF classification accuracies for three simulated noise levels for three weeks of infestation.

Noise (%)	Week one				Week two				Week three			
	0	10	20	30	0	10	20	30	0	10	20	30
Train accuracy	96.19	93.88	93.22	92.50	93.34	92.39	90.18	84.07	92.73	92.08	89.12	88.28
Test accuracy	91.07	87.50	85.71	83.92	94.64	91.07	91.07	83.92	87.50	89.28	89.28	89.28
Train kappa	92.30	87.68	86.35	84.92	86.94	84.76	80.14	68.34	85.47	84.18	78.30	76.66
Test kappa	82.14	75.00	71.42	67.85	89.28	82.14	82.14	67.85	75.00	78.57	78.57	78.57

TABLE 2: oRF-RR classification accuracies for three simulated noise levels for three weeks of infestation.

Noise (%)	Week one				Week two				Week three			
	0	10	20	30	0	10	20	30	0	10	20	30
Train accuracy	99.28	98.51	97.14	97.03	98.79	98.63	98.62	97.71	95.65	95.49	95.38	95.24
Test accuracy	96.42	97.04	96.42	96.42	100.00	100.00	100.00	100.00	91.07	91.07	91.07	92.85
Train kappa	98.57	96.42	94.28	93.97	97.59	97.27	97.26	95.47	91.23	90.92	90.69	90.54
Test kappa	92.85	92.85	92.85	92.85	100.00	100.00	100.00	100.00	82.14	82.14	82.14	85.71

higher classification accuracies than RF highlighting the utility of oRF-RR to classify low infestation levels in the absence of noise and under noisy conditions (Tables 1 and 2). For example, oRF-RR test accuracy and test kappa coefficient increased by 5.35% and 10.71%, respectively, for week one excluding the addition of simulated noise. From Table 2, it is clearly evident that oRF-RR test accuracies and test kappa coefficients remained consistent irrespective of the inclusion of increasing levels of simulated noise for weeks one and two. However, for week three, oRF-RR test accuracies and test kappa coefficients remained consistent up to 20% noise but improved slightly at 30% noise. Despite this result, the difference in test accuracy and test kappa coefficient was minimal (<3.57%) between noise levels. oRF-RR training accuracies and training kappa coefficients were higher than test accuracies and test kappa coefficients for all noise levels for weeks one and three. For week two, oRF-RR test accuracies (100%) and test kappa coefficients (100%) were consistently higher than the oRF-RR training accuracies and training kappa coefficients for all simulated noise levels.

3.5. *Oblique Random Forest-Support Vector Machine Classification Accuracies.* Table 3 presents the oRF-SVM

classification accuracies exclusive and inclusive of three simulated noise levels for three weeks of infestation. The implementation of oRF-SVM yielded higher classification accuracies than RF highlighting the utility of oRF-SVM to classify low infestation levels (Tables 1 and 3). Results indicate that oRF-SVM test accuracies and test kappa coefficients remained consistent at 100% across simulated noise levels for week three. In addition, oRF-SVM test accuracies and test kappa coefficients were higher than training accuracies and training kappa coefficients across simulated noise level. However, results show that oRF-SVM classification accuracies are influenced by noise for week one and week two. For week one, oRF-SVM test accuracies and test kappa coefficients decreased by 7.14% and 14.28%, respectively, with the addition of 30% of simulated noise. However, for week two oRF-SVM test accuracies and test kappa coefficients were variable with increasing levels of simulated noise.

4. Discussion

This study showed that oRF (oRF-RR and oRF-SVM) outperforms RF by consistently achieving better results under conditions of increasing levels of spectral noise. oRF employs

TABLE 3: oRF-SVM classification accuracies for three simulated noise levels for three weeks of infestation.

Noise (%)	Week one				Week two				Week three			
	0	10	20	30	0	10	20	30	0	10	20	30
Train accuracy	96.19	97.80	94.71	94.71	96.42	97.08	95.60	96.31	93.35	94.11	94.11	91.13
Test accuracy	94.64	91.07	89.28	87.50	94.64	96.42	94.64	94.64	100.00	100.00	100.00	100.00
Train kappa	92.30	95.57	89.23	89.33	92.85	94.14	91.26	92.69	86.66	88.23	88.23	82.28
Test kappa	89.28	82.14	78.57	75.00	89.28	92.85	89.28	89.28	100.00	100.00	100.00	100.00

the same underlying principles as RF to effectively deal with noise. However, oRF uses oblique decision boundaries to classify low infestation levels, therefore improving overall classification accuracies. oRF (oRF-RR and oRF-SVM) can, therefore, be regarded as a robust classification algorithm to classify low infestation levels on water hyacinth plants under noisy spectral conditions. This study has demonstrated that RF and oRF have the capability to discriminate between healthy and low infested water hyacinth plants using hyperspectral data.

4.1. Classifying Healthy and Low Infested Water Hyacinth Plants without Noise. RF and oRF classification accuracies achieved in this study confirm that healthy and low infested water hyacinth plants can be discriminated accurately in the absence of spectral noise. The results from this study showed that RF and oRF test accuracies were above 87.50% and 91.07%, respectively. Similar results were achieved by Adelabu et al. [43] who reported an overall RF classification accuracy of 82.42% when discriminating three insect defoliation levels (undefoliated, partly defoliated, and refoliated) in mopane woodland. In the present study, spectral discrimination could be attributed to the morphological and physiological damage caused by insect-induced stresses that altered canopy reflectance. Both oRF-RR and oRF-SVM yielded higher classification accuracies than RF without the addition of spectral noise for all three weeks. The utility of oRF-RR resulted in an increase in classification accuracies of 3.75%, 5.36%, and 3.50% for week one, week two, and week three, respectively. The results achieved in this study compare favourably with the results achieved by Menze et al. [18] and Poona et al. [20]. The improvement in classification accuracies can be attributed to the difference in which the feature space was split. RF splits the feature space using univariate hyperplanes that are orthogonal [14]. In contrast, oRF splits the feature space using multivariate hyperplanes that are oblique, therefore improving the generalization of individual trees and overall classification accuracy [18, 19]. Overall, oRF classification accuracies achieved in this study confirm that oRF performed better when classifying biocontrol damage on water hyacinth plants using hyperspectral data over the standard RF algorithm.

4.2. Classifying Healthy and Low Infested Water Hyacinth Plants with Noise. Many researchers advocate the utility of RF as a robust algorithm that is insensitive to noise [14, 44]. RF effectively deals with noisy data through (1) bootstrap aggregation, (2) random selection of bands at

each node, and (3) learning many variable unbiased decision trees [14]. However, results achieved in this study show that the RF results were influenced by noise due to the slight decrease in classification accuracies with increasing levels of simulated noise for all three weeks of the experiment. RF test accuracies decreased by 7.15% for week one and 10.72% for week two with the addition of 30% simulated noise. The decrease in classification accuracy with increasing simulated noise is not markedly high. Ross and Kelleher [45] conducted a comparative study of the effect of sensor noise on recognition models. Results showed that the RF classifier performance decreased by only 8.00% with the introduction of noise. RF can, therefore, be regarded as an effective classification algorithm to classify low infestation levels under simulated noise conditions. However, results from this study indicate that generally oRF-RR classification accuracies remained consistent with increasing levels of simulated noise confirming it is better suited to deal with noisy data. oRF is able to achieve this by employing the same underlying principles as RF to effectively deal with noise as well as splitting the feature space using multivariate hyperplanes that are oblique [19]. To the authors' knowledge, no studies have investigated the utility oRF for remote sensing applications under noisy conditions. The results achieved in this show that oRF is robust to effectively deal with noise and achieve higher classification accuracies.

4.3. Comparison between oRF-RR and oRF-SVM. oRF-RR and oRF-SVM classification accuracies confirm that the choice of the node-splitting model used does influence oRF performance. The results achieved in this study showed that oRF-RR produced higher classification accuracies than oRF-SVM for week one and week two, however, lower classification accuracies for week three. Similarly, Poona et al. [20] also found that oRF-SVM produced significantly higher classification accuracies than oRF-RR and oRF-PLS. In contrast, Menze et al. [18] demonstrated that oRF-RR yielded the best classification accuracy in comparison to oRF-lda and oRF-rnd. In this study, even though oRF-SVM produced higher classification accuracies than oRF-RR for week three, implementing oRF-SVM is time consuming, computationally demanding, and not practical. Bassa et al. [15] measured the computational speed of oRF-RR and oRF-SVM reporting that oRF-SVM was much slower to implement than oRF-RR. Consequently, it is better suited to implement oRF-RR at an operational level due to the difference between oRF-SVM

and oRF-RR is marginal and not markedly high as well as it being a more efficient algorithm to implement.

5. Conclusion

This study demonstrates the benefit of using oRF to classify healthy and low infestation levels. Generally, oRF outperforms RF for all three weeks in the absence of simulated noise. oRF classification accuracies do not deteriorate under increasing levels of simulated noise. oRF employs the same underlying principles as RF to effectively deal with noise. However, oRF uses oblique decision boundaries to classify low infestation levels, therefore improving overall classification accuracies. oRF-RR outperforms oRF-SVM for week one and week two, however, oRF-SVM outperformed oRF-RR for week three under conditions of simulated noise. While oRF-SVM slightly outperformed oRF-RR, implementing oRF-SVM is time consuming, computationally demanding, and not practical at an operational scale. Consequently, for operational use, water resource managers should implement oRF-RR. Future research should investigate the implementation of oRF-RR using noisy hyperspectral satellite imagery. Overall, this study showed that oRF-RR is a robust classification algorithm that is robust in the presence of noise.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank the National Research Foundation (NRF) for providing the necessary funds required to undertake this study.

References

- [1] D. G. Goodenough and T. Han, "Reducing noise in hyperspectral data — a nonlinear data series analysis approach," in *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pp. 1–4, Grenoble, France, August 2009.
- [2] A. Karami, R. Heylen, and P. Scheunders, "Band-specific shearlet-based hyperspectral image noise reduction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5054–5066, 2015.
- [3] X. Zhu and X. Wu, "Class noise vs. attribute noise: a quantitative study," *Artificial Intelligence Review*, vol. 22, no. 3, pp. 177–210, 2004.
- [4] J. A. Sáez, J. Luengo, and F. Herrera, "Evaluating the classifier behavior with noisy data considering performance and robustness: the equalized loss of accuracy measure," *Neurocomputing*, vol. 176, pp. 26–35, 2016.
- [5] D. F. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artificial Intelligence Review*, vol. 33, no. 4, pp. 275–306, 2010.
- [6] J. S. Sánchez, R. Barandela, A. I. Marques, R. Alejo, and J. Badenas, "Analysis of new techniques to obtain quality training sets," *Pattern Recognition Letters*, vol. 24, no. 7, pp. 1015–1022, 2003.
- [7] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779–785, 1994.
- [8] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, 2012.
- [9] D. Cerra, R. Muller, and P. Reinartz, "Noise reduction in hyperspectral images through spectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 109–113, 2014.
- [10] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.
- [11] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 296–308, 2015.
- [12] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
- [13] T. M. Khoshgoftaar and P. Rebour, "Improving software quality prediction by noise filtering techniques," *Journal of Computer Science and Technology*, vol. 22, no. 3, pp. 387–396, 2007.
- [14] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [15] Z. Bassa, U. Bob, Z. Szantoi, and R. Ismail, "Land cover and land use mapping of the iSimangaliso wetland park, South Africa: comparison of oblique and orthogonal random forest algorithms," *Journal of Applied Remote Sensing*, vol. 10, no. 1, article 015017, 2016.
- [16] W. Chen, X. Li, and Y. Wang, "Forested landslide detection using LiDAR data and the random forest algorithm: a case study of the Three Gorges, China," *Remote Sensing of Environment*, vol. 152, pp. 291–301, 2014.
- [17] M. Belgiu and L. Drăguț, "Random forest in remote sensing: a review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.
- [18] B. H. Menze, B. M. Kelm, D. N. Splitthoff, U. Koethe, and F. A. Hamprecht, "On oblique random forests," *Machine Learning and Knowledge Discovery in Databases*, vol. 6912, pp. 453–469, 2011.
- [19] T. N. Do, P. Lenca, S. Lallich, and N. K. Pham, "Classifying very-high-dimensional data with random forests of oblique decision trees," in *Advances in Knowledge Discovery and Management*, F. Guillet, G. Ritschard, D. A. Zighed, and H. Briand, Eds., vol. 292 of *Studies in Computational Intelligence*, pp. 39–55, Springer, Berlin, Heidelberg, 2010.
- [20] N. Poona, A. van Niekerk, and R. Ismail, "Investigating the utility of oblique tree-based ensembles for the classification of hyperspectral data," *Sensors*, vol. 16, no. 11, 2016.
- [21] T. D. Center, "Biological control of weeds: water hyacinth and water lettuce," in *Pest Management in the Subtropics: Biological Control — A Florida Perspective*, D. Rosen, F. D. Bennet, and J. L. Capinera, Eds., pp. 481–521, Intercept Publishing Company, Andover, UK, 1994.

- [22] M. H. Julien, "Biological control of water hyacinth with arthropods: a review to 2000," in *Proceedings of the Second Meeting of the Global Working Group for the Biological and Integrated Control of Water Hyacinth*, pp. 8–20, Beijing, China, October 2001.
- [23] S. Lowe, M. Browne, S. Boudjelas, and M. De Poorter, *100 of the World's Worst Invasive Alien Species: A Selection from the Global Invasive Species Database*, The Invasive Species Specialist Group (ISSG) a Specialist Group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), Auckland, New Zealand, 2004.
- [24] N. H. Agjee, R. Ismail, and O. Mutanga, "Identifying relevant hyperspectral bands using Boruta: a temporal analysis of water hyacinth biocontrol," *Journal of Applied Remote Sensing*, vol. 10, no. 4, article 042002, 2016.
- [25] T. D. Center, F. A. Dray Jr., G. P. Jubinsky, and M. J. Grodowitz, "Biological control of water hyacinth under conditions of maintenance management: can herbicides and insects be integrated?," *Environmental Management*, vol. 23, no. 2, pp. 241–256, 1999.
- [26] A. Bownes, M. P. Hill, and M. J. Byrne, "Assessing density–damage relationships between water hyacinth and its grasshopper herbivore," *Entomologia Experimentalis et Applicata*, vol. 137, no. 3, pp. 246–254, 2010.
- [27] J. A. Coetzee, M. J. Byrne, and M. P. Hill, "Impact of nutrients and herbivory by *Eccritotarsus catarinensis* on the biological control of water hyacinth, *Eichhornia crassipes*," *Aquatic Botany*, vol. 86, no. 2, pp. 179–186, 2007.
- [28] P. G. Soti and J. C. Volin, "Does water hyacinth (*Eichhornia crassipes*) compensate for simulated defoliation? Implications for effective biocontrol," *Biological Control*, vol. 54, no. 1, pp. 35–40, 2010.
- [29] A. Bownes, M. P. Hill, and M. J. Byrne, "Evaluating the impact of herbivory by a grasshopper, *Cornops aquaticum* (Orthoptera: Acrididae), on the competitive performance and biomass accumulation of water hyacinth, *Eichhornia crassipes* (Pontederiaceae)," *Biological Control*, vol. 53, no. 3, pp. 297–303, 2010.
- [30] R. A. Goyer and J. D. Stark, "The impact of *Neochetina eichhorniae* on water hyacinth in Southern Louisiana," *Journal of Aquatic Plant Management*, vol. 22, pp. 57–61, 1984.
- [31] M. O. Bashir, Z. E. El Abjar, and N. S. Irving, "Observations on the effect of the weevils *Neochetina eichhorniae* Warner and *Neochetina bruchi* Hustache on the growth of water hyacinth," *Hydrobiologia*, vol. 110, no. 1, pp. 95–98, 1984.
- [32] ASD, *Handheld Spectroradiometer: User's Guide Version 4.05*, Analytical Spectral Devices Incorporated, Boulder, CO, USA, 2005.
- [33] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012.
- [34] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 1, p. 3, 2006.
- [35] A. Liaw and M. Wiener, *Random Forest: Breiman and Cutler's Random Forests for Classification and Regression*, R Package Version 4.6-7, 2013.
- [36] E. A. Addink, S. M. de Jong, and E. J. Pebesma, "The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 8, pp. 905–912, 2007.
- [37] J. Hernandez, G. Lobos, I. Matus, A. del Pozo, P. Silva, and M. Galleguillos, "Using ridge regression models to estimate grain yield from field spectral data in bread wheat (*Triticum Aestivum* L.) grown under three water regimes," *Remote Sensing*, vol. 7, no. 2, pp. 2109–2126, 2015.
- [38] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, USA, 1995.
- [39] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.
- [40] B. Menze and N. Splitthoff, *Oblique Random Forests from Recursive Linear Model Splits*, R Package Version 0.3, 2012.
- [41] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environment*, vol. 37, no. 1, pp. 35–46, 1991.
- [42] G. M. Foody, "Thematic map comparison," *Photogrammetric Engineering & Remote Sensing*, vol. 70, no. 5, pp. 627–633, 2004.
- [43] S. Adelabu, O. Mutanga, E. Adam, and R. Sebege, "Spectral discrimination of insect defoliation levels in mopane woodland using hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 177–186, 2014.
- [44] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012.
- [45] R. Ross and J. Kelleher, "A comparative study of the effect of sensor noise on activity recognition models," *International Joint Conference on Ambient Intelligence*, vol. 413, pp. 151–162, 2013.



Hindawi

Submit your manuscripts at
www.hindawi.com

