CLASSIFICATION OF MULTISOURCE SATELLITE IMAGERY USING RANDOM FOREST IN NORTHERN ALBERTA

Bogoljub Stankovic[1], Ridha Touzi[2], Jinkai Zhang[1]

[1] Wildfire Management Brunch, Alberta Agriculture and Forestry
[2] Canada Centre for Mapping and Earth Observation, Natural Resources Canada

Abstract
This work was conducted to improve our understanding of the complex boreal forest environment intermittent with wetlands and ravaged by wildfires, while using multisource imagery and machine learning algorithm to improve land cover classification. The data set contains a mixture of optical, liDAR and multisource SAR data including the fully polarimetric Advanced Land Observing Satellite (ALOS2)data provided kindly by Japan Aerospace Exploration Agency (JAXA). We examined the capability C-band and fully polarimetric L-band data for the land cover classifications. Adding liDAR and SAR data of different wavelengths did improve classification accuracy. Among various scenarios, it was found that the one combining optical, Sentinel-1, liDAR derivatives and Touzi α component yielded the best predictive score.

Keywords: wetland, polarimetry, incidence angle, backscatter, ASLC, ASLC code

1. BACKGROUND

Classifying remotely sensed data into a thematic map is very challenging because of many factors, including complexity of the landscape, selected remotely sensed data, selection of training samples, image processing and classification approaches. All of which may affect the success of a classification. Generally, wetlands are a challenging target for classification due to their inherent dynamism and natural range of variability [3]. The depth of water and the duration of its presence vary considerably from wetland to wetland and from year to year or throughout a season [2], which necessitates remotely sensed data with a higher temporal resolution than typically required to map other less dynamic land cover types. Accurate wetland classification becomes challenging as the distinction between wetland and upland boundaries blurs as water recedes and above ground vegetation develops. Wetlands reach their maximum inundation during the early spring after snow is gone, during which time boundaries are clearly visible and wetlands are most easily classified. Some wetland features particularly emergent and wet meadow marshes increase in coverage and density, in parallel with a decrease in open standing water as the season progresses. This early leaf-off green-up period likely contributes to the higher classification accuracy achieved during the spring seasons due to better visibility of wetlands. However, during summer wetlands experience a decrease in water availability and depth, and an increase in land surface temperature and vegetation growth. This reduction in open standing water, and increase in wetland vegetation extent and density contributes to the lower accuracy of classification work based on late summer imagery.

Therefore, for better wetland mapping, multispectral optical images [7] are supplemented with synthetic aperture radar (SAR) images. Optical sensor data are not able to capture forest stand structure information because they cannot penetrate forest canopy. In contrast, SAR can penetrate the forest canopy depending on frequency and polarization mode and may capture more structural information than optical data. Another advantage is that SAR measurements are independent of weather conditions. The amount of backscatter energy is primarily affected by the wavelength, polarization, and incidence angle of the microwave energy in combination with several key characteristics of the Earth's surface including dielectric property, size/roughness, and structure [8]

With more available multi source imagery, more complexity comes into classification process. Numerous features can be extracted from multisource imagery, which are being inputted into the process. To process potentially huge number of bands, new techniques are needed. Non-parametric approaches are suggested for the classification of multi-source data in complex environments[17]. This may not be the best approach due to computational intensity and a lot of redundant information. To avoid this complexities, data dimension reduction techniques comes to play its role.

In wildfire management it is critically important to have accurate fuel type maps. The overall objective is to characterize the land cover in more accurate manner while combining lidar, SAR and optical remote sensing data products using a Random Forest (RF) classifier (Breiman, 2001). Specifically, we seek to answer which Sentinel-1 or Alos2 features provide the greatest utility for discriminating among various landcover classes.

2.STUDY AREA

The area is located North of the town of Peace River in the province of Alberta,Canada.It is at -119.2 longitude and 56.7 degrees of latitude. It is Boreal Mixedwood and wildfire prone landscape dominated by Lodgepole Pine (Pinus contorta), White Spruce (*Picea glauca*), Aspen (*Populus tremuloides)* , Black Spruce (Picea Mariana) , bogs and fens.

Fig. 1. Study area

3. DATA

3.1 Field and Remote Sensing Data

Training data for image classification were gathered in summer 2018 during a ground survey. Several sampling areas were established and on each 3 plots were located. The image stacks we work with include Landsat 8 bands excluding 1st "Coastal band", liDAR derivatives , Tassel cap components, Sentinel-1 backscatter coefficients and ALOS2 components from different polearimetric decompositions.

The Sentinel-1 satellites (A and B) carry C-band SAR instruments observing the Earth with a repetitive cycle of 12 days. The main IW acquisition mode of the Sentinel-1 mission supports operations in dual polarization VV and VH. The instrument transmits microwave energy in linear vertical (V) polarization and detecting returns in both linear vertical (V) and horizontal (H) polarization. We obtained the Sentinel-1 IW Level-1 GRD products from the Copernicus Open Access Hub [9]. Nominal incidence angle for this particular product is 28.3°. There is a slight difference in incidence angle between Alos2 data and Sentinel -1 data.The Government of Alberta (GOA) has received a multi temporal set of Alos-2 Palsar data in 2019 from JAXA (*Japan* Aerospace Exploration Agency). SLC format PALSAR. Full-polarimetry image that has been used for the study area was in Level 1.1 geo-coded format. Nominal incidence angle is 31 degrees. Polarimetric SAR systems measure a complete information of radar target in a form of scattering matrix (S). The matrix has four components describing the transformation of the polarization of a wave pulse incident upon reflective medium to the polarization of the backscattered wave.

Optical images used in this study were geometrically and radiometrically corrected, and only those optical images with minimum cloud coverage were selected for this work. Landsat 8, data were calibrated to TOA reflectance. Lidar and its derivatives were obtained in the house and they are part of an earlier acquisition campaign. From our lidar dEM, we derived three DEM derivatives - slope, compound topographic positional index (CTI) and topographic Position index (TPI).

4. METHODOLOGY

Part of our job is to build a model to find out which Alberta Satellite Land cover (ASLC) value might be appropriate for each cluster of pixels in our image stack bands. Before, we come to that point, data has to be processed and these processing steps are predominantly described for SAR images. Sentinel-1 data were radiometrically corrected using Alaska Satellite Facility DAAC [10] a program written in python utilizing Sentinel Toolbox (SNAP). The rest of data processing took place in ESA SNAP toolbox environment and PWS software [11] which were made freely available by European Space Agency (ESA) and Natural Resources Canada respectively.

4.1 SAR Data Preprocessing

Generation of Sentinel-1 radar backscatter coefficients (σVV and σVH) included the following processing steps:
(1) Applying precise orbital information
(2) Performing Calibration
(3) Applying Speckle-Filter to reduce systematic speckle noise using Lee Sigma Filter
(4) Terrain flattening and finally
(5) Terrain Correction using Alberta Digital Elevation Model.

The backscattering coefficient (σ_0) from ALOS–PALSAR level 1.1 products can be calculated using the following equation given by JAXA. Calibration is a process that converts SLC data to backscattering coefficients (power value). Calibration coefficients for SAR data are often defined in the decibel (dB) scale due to the high dynamic range of the imaging system. For the L-band ALOS 2 PALSAR data at hand, the conversion from raw values to calibrated radar cross section values in dB scale is performed by applying a standard formula

$$\sigma^0 = 10\log_{10}\left(I^2 + Q^2\right) + CF - 32$$

σ^0: Backscattering coefficient [unit: dB]

I(real) and Q (an imaginary part) are in-phase, and their sum of squares creates an Intensity band. CF_1 and A are Calibration factors in dB units.

Calibration factor	Value [dB]
CF_1	-83.0
A	32.0

Table 1. Values of CF1 and A (March 28,2017)

After the data at hand are radiometrically corrected, we subject them to speckle suppression using the Lee with window size 7x7. Finally, speckled images were geometrically corrected and we had terrain corrected sigma naught backscattering coefficients. **Correct scale** in which computing operations are performed **is the** power scale**.** While **dB**-scaled images are often visually pleasing, they are often not a good basis for mathematical operations on data. For instance, when we compute the mean of observations, it makes a difference whether we do that in power or dB scale. Since dB scale is a logarithmic scale, we cannot simply average data in that scale. To convert from dB to power we apply:

$$\gamma^o_{pwr} = 10^{\frac{\gamma^o_{dB}}{10}}$$

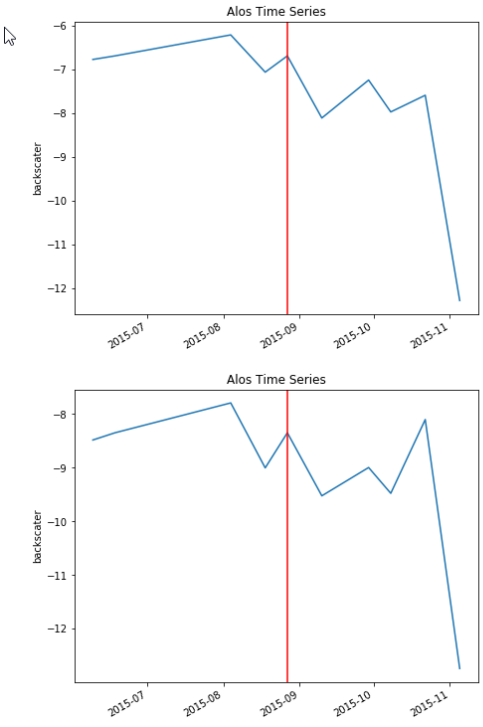


Fig 2. Fen's (upper graph) and bog's (lower graph) backscatter coefficient temporal changes expressed in dB scale.

When working with SAR imagery, it is necessary to determine whether a particular pixel is a pure target or distributed scatterer. Nevertheless, the analysis of distributed scatterers must be performed by means of the so-called incoherent decompositions. A qualitative way to differentiate pure from distributed scatterers is to consider their physical nature. A rough division would be to consider the man-made targets as pure targets, whereas natural targets can be considered as distributed. PolSAR data allow employment polarimetric decomposition techniques to identify different backscattering mechanisms of the ground targets. Unlike coherent decompositions, which are only useful for man-made structures, incoherent decompositions determine the relative contributions from different scattering mechanisms.

4.2 Polarimetric Decomposition

When ground objects and radar signals interact, backscattering of objects is affected by their characteristics and the type of polarization used. Therefore, ground objects appear different depending on the chosen polarization, allowing the various objects to be distinguished and classified. After calibration of ALOS2 image, we created a polarimetric matrix, which was followed by spatial filtering to reduce speckle noise(Refined Lee 7by 7). The reflectivity of the area being observed at a given radar wavelength can be represented by "scattering matrix"**.** As Touzi [12]has expressed, the goal of incoherent decomposition is to break down the average scattering mechanisms to associate a physical mechanism to each component. From ALOS-2 images we extracted various polarimetric decompositions including Freeman-Durden [13], H-A-α [14], and, Touzi [12] decompositions. Freeman-Durden decomposition divides the total energy into surface, volume, and double-bounce scatterings. The surface scattering component contains only ground soil information (thus, no vegetation information) . Therefore, the scattering information from vegetation and soil is completely separated.

H/A/α Decomposition

The Cloude–Pottier polarimetric decomposition expresses the 3 x 3 complex coherence matrix as a sum of eigenvalues and eigenvectors. Three components can be obtained from the T3 matrix: entropy (H), angle (α) and anisotropy (A). The entropy, degree of randomness, ranges between 0 and 1; The polarimetric anisotropy A is a complementary parameter of the polarimetric entropy H. Anisotropy measures the relative importance of the second and the third eigenvalues of coherency matrix. The average **α** component ranges between 0°and 90° and depends on the dominant scattering mechanism.

In Touzi decomposition, the phase of the symmetric scattering type is used to discriminate between wetland vegetation and forest species. So, the dominant scattering type phase with the dominant scattering eigenvalue will enhance the separation of classes, leading to a better wetland classification[12].
In this work, the coherent and incoherent decompositions were all included in the classification such that the feature selection algorithm selects the best features automatically.

4.3 Supervised classification

Our methodology includes supervised classification approach which incorporates RandomForest [15] ensemble decision tree algorithm. Random forest method makes use of several randomized decision trees when making its prediction. It can be used both for classification and regression. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. Each tree within the forest only gets to train on some subset of the full training dataset.

RF works in four steps:
1. Select random samples from a given dataset.
2. Construct a decision tree for each sample and get a prediction result from each decision tree.
3. Perform a vote for each predicted result.
4. Select the prediction result with the most votes as the final prediction.

Unlike other algorithms, RF possesses ability to work on many different spatial bands stacked in to an image. It gives also a measure of "variable importance" which relates how useful individual input spectral bands in the classification are. The main disadvantages is that Random forests is slow in generating predictions because of multiple decision trees. Whenever it makes a prediction, all the trees in the forest have to make a prediction for the same given input and then perform voting on it. The more data we stack into image it will take more processing time and computing power. This whole process is time-consuming. The model is difficult to interpret compared to a decision tree, where you can easily make a decision by following the path in the tree. Random forest requires training, or using a dataset where we know the true answer to fit a predictive model. Based on the accuracy of the results, we can apply the prediction model to the entire image and save the output. So the way we divide the predictor space (all image bands) into smaller regions is recursive in nature. The main layout of classification process is stated below:

a) Image collection, process and import images into a stack

b) Collection of areas of interest (AOI) for training and validation and processing them. The ratio of training to validation is 2/3 to 1/3.
c) Initialize and train the model using the training data.
d) Apply model Classifier to validation Data set. Training part of the dataset is used to build/train a model (decision tree) , and then use the same model to predict the classes for the validation data set. This validation data set is being used to assess the accuracy of our model where we compare true values that we collected against predicted values.
e) Evaluation of the Classifier. Here we seek to select the best set of tuning parameters for the model and estimate the performance of our model on validation data set.

The accuracy metrics calculated on validation dataset should give us a reasonable estimate of the performance of our model on unseen/new sample data. Confusion matrix and the accuracy score enable us to compare this model with other models. The number of correct and incorrect predictions is presented as count values are broken down by each ASLC code. We can calculate accuracy for different number of decision trees (estimators). When we put this info on a graph, we can estimate how many estimators (decision trees) we need to come up with a stable estimate of desired accuracy. This is important because the increased number of trees increases the processing time.

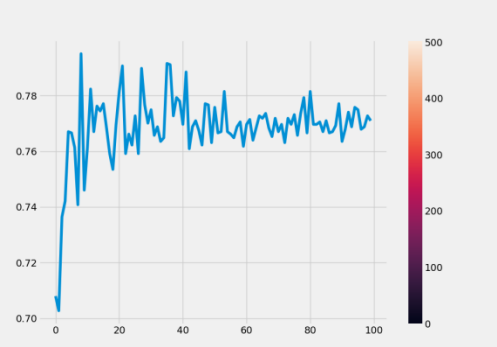
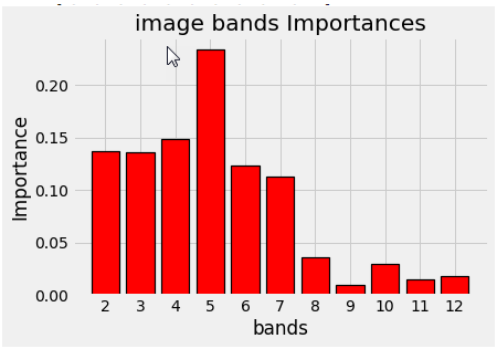


Fig 3 Number of random trees and its effect on correct prediction. In this case we do not need more than 50 trees to come up with a 77% of accuracy.

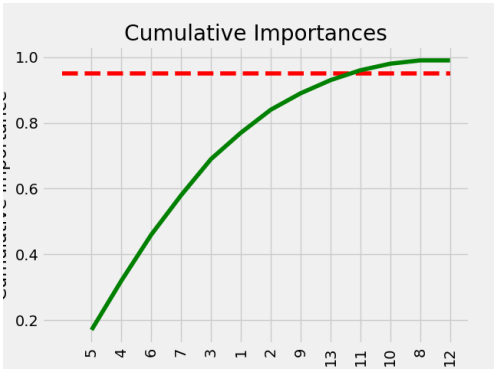
f) Dimension Reduction

In fact, not all the derived features are informative for land cover classification, or some of them may contain redundant information. At this stage we convert a set of data having vast dimensions into data with lesser dimensions ensuring that it conveys similar information concisely. Dimension Reduction process enables us to compress data, reduce the storage, fastens the time required for performing same computations and removes redundant features. The model produces a score telling us

how important each feature was in classifying. In order to maximize the classification accuracy, it is necessary to identify the best combination of bands including spectral indices, optical, SAR and Lidar data.



4 a)



4b)

Fig 4 a). Individual Importance for each band in our stacked data set.
Fig 4b) Cumulative importance with cut off value at 95%. X axis is individual band in a stack and Y axis is band importance.

Those bands at the end adding reminding 5% of importance are being taken out to improve computing time.

4.4. Experimental design

To examine the utility of the dataset we designed multiple experiments in which RF classifiers used different sets of predictors comprising different combinations of features derived from Sentinel-1, ALOS2, Landsat 8 and liDAR features. Experimental design included following various scenarios with stack combinations:

1st scenario: Landsat 8 without 1st band.

2nd scenario: Landsat 8 + lidar derivatives (TPI, TRI, slope)

3rd scenario: Landsat 8 + lidar derivatives (TPI, TRI, slope) + Sentinel 1 (σ_{HV}, σ_{VV})

4th scenario: Landsat 8 + lidar derivatives (TPI, TRI, slope) + Sentinel 1 (σ_{HV}, σ_{VV}) + Tassel cap (brightness,greenness)

5th scenario: Landsat 8 + lidar derivatives (TPI, TRI, slope) + ALOS2 (σ_{HH}, σ_{HV}, σ_{VH} and σ_{VV} backscatter coefficients) + Freeman Durden polarimetric components

6th scenario: Landsat 8 + lidar derivatives (TPI, TRI, slope) + Sentinel 1 (σ_{HV}, σ_{VV}) + ALOS2 Touzi α decomposition components.

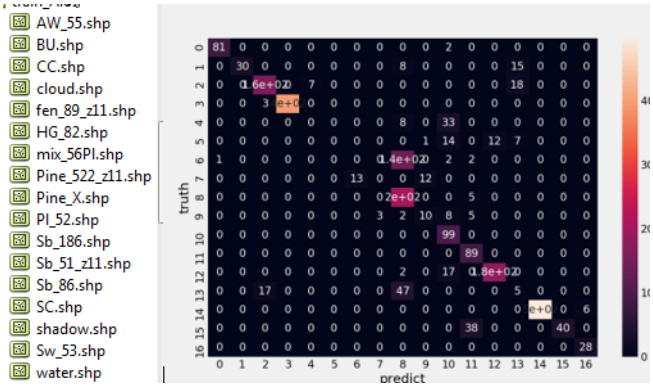
7th scenario: Landsat 8 + lidar derivatives (TPI, TRI, slope) + Sentinel 1 (σ_{HV}, σ_{VV}) + ALOS2 Touzi α decomposition components + ALOS2 H,A decomposition (entropy ,anisotropy and angle)

5. RESULTS

To examine the utility of Sentinel-1 and ALOS 2 features in classifications, we grouped and compared the accuracy measures of RF classification results from all the experiments.

scenario:	Correct Prediction (%)	Optical	Lidar	SAR	ALOS2	Decomposition
1st scenario:	73.33	L8				
2nd scenario:	78.41	L8	TPI,TRI,slp			
3rd scenario:	79.62	L8	TPI,TRI,slp	σ_{HV}, σ_{VV}		
4th scenario:	71.14	L8, Brigh/Green	TPI,TRI,slp	σ_{HV}, σ_{VV}		
5th scenario:	75.6	L8	TPI,TRI,slp		σ_{HH}, σ_{HV}, σ_{VH} and σ_{VV}	FD
6th scenario:	80	L8	TPI,TRI,slp	σ_{HV}, σ_{VV}		Touzi (α)
7th scenario:	76.32	L8	TPI,TRI,slp	σ_{HV}, σ_{VV}		Touzi (α) ,H, A

Table 2. Different scenarios and their predicted accuracies.



Graph 5. Classification report

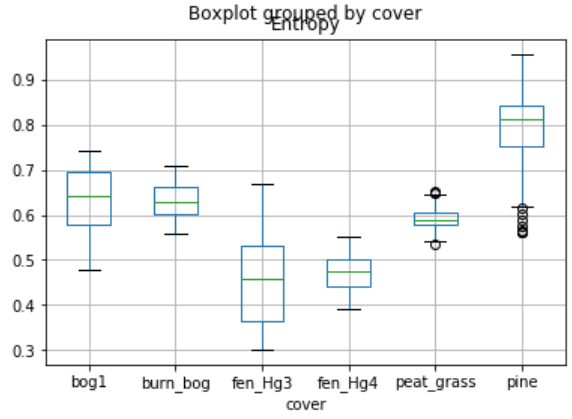
6. DISCUSSION

The predicted accuracies for different scenarios, tell us that bringing more features into the model does not create a better model. Lidar derivatives increased the baseline (Landsat only) scenario accuracy, and therefore they seem to be an important component to the model. Lidar crown

height based model (CHM) was excluded on the purpose due to having no data for a portion of the study area. From all Sentinel-1 possible features we used only σVH and σVV. Adding them to the model, slightly increased predicted accuracy. Across all feature combinations, the inclusion Sentinel-1 σVH & σVV were consistently associated with the higher accuracies indicating the utility of these polarizations in discriminating among our land cover classes. On the other hand, brightness and greenness caused accuracy to decrease, and they were removed from other scenarios. Adding into the mix different polarimetric decompositions did not create the output we were hoping. Only adding Touzi α to Sentinel -1 polarizations produced the best score. The accuracies did not change significantly among the different scenarios. Seven different scenarios raised accuracy from 73 to 80%. Similar finding came from Wulder et all study [1]. The more work is required to find an optimal composition of different features producing the best score. One set of features may be optimal for one type and size of landscape, and different one for another type of landscape. Each wavelength has its own advantages and disadvantages. The selection of an appropriate SAR wavelength depends on the land cover classes since the interaction of SAR wavelengths varies widely with different vegetation types depending on their size. For example,L-band can pass through the vegetation canopy and detect water beneath the flooded, while C-band will better depict non-treed classes[6]. When it comes to temporal dimension, it is necessary to select May or October image if there is a wetland on landscape.

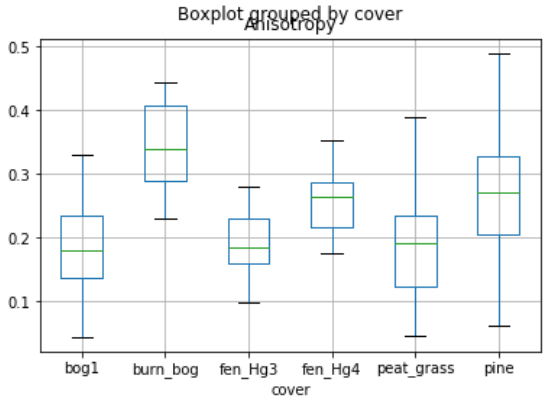
The major issue in the classified output was immature Lodgepole Pine ending in to conifer dominated mixedwood. Scattering over forested areas is dominated by volume scattering.
Clearly, it is ambiguous how combination of C and L-band features into an image stack affects the output. This difference between incidence angles may also cause some artifacts. In a scene with shallow angle more volume scattering will appear, while in a scene with a steeper angle we may see more double bounce in the same immature pine stand. The scattering mechanism of a radar target could be affected by the penetration depth of the radar signal. Dense flooded vegetation could present volume scattering mechanism in C-band SAR(return from canopy), but double bounce in L-band due to scattering process from trunk and water interaction.
As Freeman and Durden found the immature regrowth and upland forest results are virtually identical at C-band [16]. At L-band, the level of canopy scatter is only slightly lower (by 1 dB) for the immature regrowth, which cannot be regarded as significant. Only P-band could differentiate between these two classes since the level of HV backscatter for these two classes, can be interpreted as a difference in the level of biomass.

Let's look at several of our samples through Cloude–Pottier polarimetric decomposition. The entropy indicates depolarization of the scattering ensemble. If entropy is low (H < 0.3), then scatterers are weakly depolarized, when entropy is high, then scattering is highly depolarized.

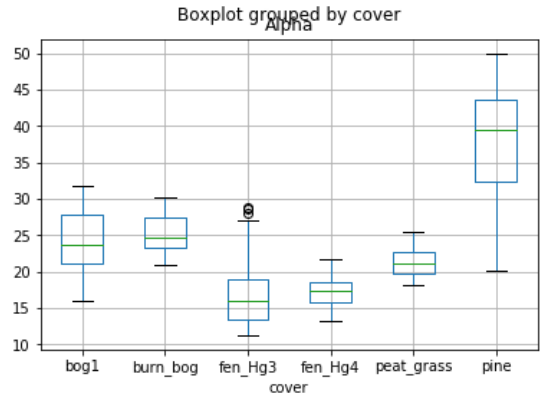


Graph 6. Entropy for some training samples.

The graph above depicts that our immature pine samples are associated with values near 1 indicating high depolarization and random combination of different scatterers(two non-negligible secondary mechanisms). Both our fens have values near 0 indicating a single-surface scattering mechanism. From an application point of view, the anisotropy A can be applied as a means of discrimination when entropy H >0.7. Only immature Pine has Entropy > 0.7 and therefore, it has two main scattering mechanisms with the same probability of occurrence and a third mechanism of minor importance



Graph 7 Anisotropy for some training samples.



Graph 8 Alpha some training samples.

The average **α** component indicates that dipole scattering is characteristic for immature Pine samples. There is no double bounce in these Pine stands which indicates Pine being immature. Also, a shallow incidence generates more volume scattering at expense of double bounce. The fen samples on a graph indicate surface as a main scatterer.

7. CONCLUSION

Resulting classification accuracies of 80% for project area in NW Alberta confirmed that combination of optical, liDAR and SAR data provides adequate input to use in a classification process based on machine learning algorithms such as Random Forest. As the study [1] concluded, the accuracies do not change significantly among the different scenarios. We conclude that using a C-Band shallow incidence angle scene is more accurate for the study of non treed cover types, while L band is more suitable for forested landscape. Detection of immature forest stands may require longer wavelengths or higher resolution of optical data.

FUTURE WORK

There is a plenty of scenarios we could have used here and explore it in more detail. We also need explore in more detail Touzi decomposition componenets and their influence on these scenarios.

8. REFERENCES

[1] ZhanLiHaoChenJoanne C.WhiteMichael A.WulderTxominHermosilla. Discriminating treed and non-treed wetlands in boreal ecosystems using time series Sentinel-1 data. International Journal of Applied Earth Observation and Geoinformation, Volume 85, March 2020, 102007

[2] By William J. Mitsch, James G. Gosselink, Li Zhang, Christopher J. Anderson. Wetland Ecosystems

[3] Alisa L. Gallant ,The Challenges of Remote Monitoring of Wetlands. Remote Sens. 2015, 7, 10938-10950; doi:10.3390/rs70810938

[4] Gordana Kaplan, Zehra Yigit Avdan and Ugur Avdan. Mapping and Monitoring Wetland Dynamics Using Thermal, Optical, and SAR Remote Sensing Data

- [5] Elhadi M.I Adam,Onisimo Mutanga,Denis Rugege. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. Wetlands Ecology and Management 18(3):281-296 · June 2010

[6] Bahram Salehi, Masoud Mahdianpari, Meisam Amani, Fariba M. Manesh, Jean Granger, Sahel Mahdavi and Brian Brisco. A Collection of Novel Algorithms for Wetland Classification with SAR and Optical Data

[7] Elhadi Adam • Onisimo Mutanga • Denis Rugege. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. Wetlands Ecol Manage (2010) 18:281–296

[8] M.W. Lang, G.W. McCarty Remote sensing data for regional wetland mapping in the United States: trends and future prospects. R.E. Russo (Ed.), Wetlands: Ecology, Conservation and Restoration., Nova Science Publishers, Inc., Hauppauge, New York (2008), pp. 73-112

[9] Copernicus Open Hub. https://scihub.copernicus.eu/dhus/#/home

[10] ASF MapReady User Manual Version 3.1 Alaska Satellite Facility. https://www.asf.alaska.edu/wp-content/uploads/2019/03/mapready_manual_3.1.22.pdf

[11] PWS Function Description. https://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/tools-applications/polarimetric-workstation/9753?wbdisable=true

[12] R. Touzi, A. Deschamps, G. Rother Phase of target scattering for wetland characterization using polarimetric C-band SAR IEEE Trans. Geosci. Remote Sens., 47 (9) (2009), pp. 3241-3261

[13] A. Freeman. A three-component scattering model for polarimetric SAR data. IEEE Transactions on Geoscience and Remote Sensing (Volume: 36 , Issue: 3 , May 1998)

[14] Shane Robert Cloude,An entropy based classication scheme for land applications of polarimetric SAR. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 35, NO. 1, JANUARY 1997

[15] LEO BREIMAN. Random Forests. Machine Learning, 45, 5–32, 2001 c 2001 Kluwer Academic Publishers

[16] Anthony Freeman, and Stephen L. Durden. A Three-Component Scattering Model for Polarimetric SAR Data. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 36, NO. 3, MAY 1998 Canadian wetlands: Environmental gradients and classification . Vegetatio 118: 131-137, 1995.

[17] Sara Attarchi and Richard Gloaguen. Classifying Complex Mountainous Forests with L-Band SAR and Landsat Data Integration: A Comparison among Different Machine Learning Methods in the Hyrcanian Forest. *Remote Sens.* **2014**, *6*, 3624-3647; doi:10.3390/rs6053624