A peer-reviewed version of this preprint was published in PeerJ on 28 February 2019.

NEON NIST data science evaluation challenge: methods and results of team Conor

Conor A McMahon Corresp. 1

1 Mechanical Engineering, University of Texas at Austin, Austin, TX, United States

Corresponding Author: Conor A McMahon
Email address: conor.mcmahon@utexas.edu

The NIST DSE Plant Identification challenge is a new periodic competition focused on improving and generalizing remote sensing processing methods for forest landscapes. To compete in the competition, I created a pipeline to perform three remote sensing tasks. First, a NDVI- and height-thresholded watershed segmentation was performed to identify individual tree crowns using LIDAR height measurements. Second, remote sensing data for segmented crowns was aligned with ground measurements by choosing the set of pairings which minimized error in position and in crown area as predicted by stem height. Third, species classification was performed by reducing the dataset's dimensionality through PCA and then constructing a set of maximum likelihood classifiers to estimate species likelihoods for each tree. Of the three algorithms, the classification routine exhibited the strongest relative performance, with the segmentation algorithm performing the least well.

NEON NIST data science evaluation challenge: methods and results of team Conor

Conor McMahon (conormcmahon22@gmail.com)

Department of Mechanical Engineering

University of Texas

Austin, Texas, United States

ABSTRACT

The NIST DSE Plant Identification challenge is a new periodic competition focused on improving and generalizing remote sensing processing methods for forest landscapes. To compete in the competition, I created a pipeline to perform three remote sensing tasks. First, an NDVI- and height-thresholded water-shed segmentation was performed to identify individual tree crowns using LIDAR height measurements. Second, remote sensing data for segmented crowns was aligned with ground measurements by choosing the set of pairings which minimized error in position and in crown area as predicted by stem height. Third, species classification was performed by reducing the dataset's dimensionality through PCA and then constructing a set of maximum likelihood classifiers to estimate species likelihoods for each tree. Of the three algorithms, the classification routine exhibited the strongest relative performance, with the segmentation algorithm performing the least well.

Keywords: Remote Sensing, Forestry, LIDAR, Hyperspectral Camera, Segmentation, Classification, Alignment, Ecology

1 BACKGROUND

Characterizing the structure and species makeup of forest systems is an essential step in many disciplines. These kinds of analysis are necessary for assessing the quality of a patch of habitat for conservation of particular target taxa(3)(8), for estimating system-level properties like primary productivity or capacity for carbon sequestration(11), and also for landowners interested in directly managing forests for wood or fruit production(10). Traditional methods of characterizing forests involved expensive, laborious, and time-consuming deployment of experts on foot to manually label individual trees with location, species, and structural data(1)(4). More recently, remote sensing technologies have emerged which show the potential to massively alter the scale and efficacy with which these characterizations can be performed. These technologies utilize cameras and LIDAR units mounted on aircraft (small planes or unmanned aerial vehicles) which can be flown over a forest collecting massive amounts of data on the height and spectral properties of the community(5)(10).

Extracting useful ecosystem parameters from this mass of generated data involves three primary steps: segmentation, alignment, and classification(4). In the segmentation step, individual tree crowns are automatically extracted from the scene so that they can be counted and analyzed separately. During alignment, individual trees from the segmented scene are automatically paired with stems labeled during traditional ground-based methods to improve the richness of the remote sensing dataset. This also allows assignment of species labels to some crowns, which provides training data for the classification step. During classification, species labels are estimated for remaining trees which were not already assigned labels by experts on foot.

In general the efficacy of different remote sensing processes depends strongly on the forest type being surveyed – in particular the degree of canopy openness and overall species diversity(6). When new methods are introduced in the literature there is often a lack of robust comparison to existing methods, and the comparisons which are included are difficult to apply broadly due to these inherent

35 differences in performance on different systems(4). As well, the formats in which remote sensing data are
36 saved and processed vary hugely across platforms and research disciplines, and have proven difficult to
37 standardize(4).
38   The NIST DSE Plant Identification Challenge(4) was introduced this year to try to combat these
39 problems and to increase standardization, methods benchmarking, and collaboration within the remote
40 sensing research community. During this competition, the three tasks outlined above (segmentation,
41 alignment, and classification) were performed on the same dataset by multiple competing teams. To keep
42 the results of each task from limiting the performance of subsequent tasks the input data for each task was
43 provided a priori by the competition organizers. In a real-world scenario the tasks could instead be arrayed
44 in a single coherent pipeline to perform more meaningful automatic forest characterization. More detailed
45 descriptions of the provided data will be provided in the task subsections below, and exact methods of
46 input data collection and preprocessing can be found in the parent paper on the overall competition pilot,
47 along with elaboration on the nature and goals of the competition.(4)

2 METHODS

2.1 Data Collection

50 All data used were provided by NEON, and included, according to the competition organizers:

51   1. Woody plant vegetation structure (NEON.DP1.10098);

52   2. Spectrometer orthorectified surface directional reflectance - flightline (NEON.DP1.30008);

53   3. Ecosystem structure (NEON.DP3.30015);

54   4. High-resolution orthorectified camera imagery (NEON.DP1.30010).

55 In plain speech, these datasets included hyperspectral and LIDAR-produced DSM images of the scenes in
56 question, as well as hand-labeled tree species, height, diameter, location, and stem-crown correspondences.
57 Again, more information regarding the data provided can be found in the parent paper (4).

2.2 Segmentation

2.2.1 Input Data

60 For this task, I used only the canopy height model provided as an input by the competition organizers and
61 the hyperspectral camera image of this same area. The dataset consisted of 30 training and 13 test plots,
62 each a pair of CHM and hyperspectral camera images of the forest in an 80x80m area, including a 20m
63 buffer on each edge of the 40x40m plot. Both rasters were gridded with 1x1m pixel sizes such that they
64 contained 6400 points each. The hyperspectral image contained 426 bands between 350 and 2500nm,
65 while the CHM image contained only a single band filled with height values. More details on the methods
66 of collection for the input data can be found in the parent paper (4).

2.2.2 Processing Overview

68 My approach started with applying an NDVI filter to the cloud in order to remove points that had NDVI
69 values too low to be plant matter. Then I applied a sliding linear window function to identify local maxima
70 (presumptive treetops). Finally, I performed a watershed segmentation to create tree polygons around
71 these top points.

2.2.3 Vegetation Filtering

The Normalized Difference Vegetation Index (NDVI) is an index used to determine the degree of plant
cover at a point in a spectral scene. It is given as:

$$NDVI = \frac{NIR - RED}{NIR + RED}, (5)$$

(1)

73 where NIR and RED are reflectances of the scene in the red and near infrared bands. For this filter I used
74 bands 50 and 70 of the hyperspectral image, which correspond to wavelengths of 628.1 and 728.3 nm
75 (respectively in the red and near-red IR ranges). High positive values of NDVI indicate substantial plant
76 cover. Low or negative NDVI indicates landcover by non-vegetative materials. I used a threshold of 0.5
77 for filtering because this removed most of the ground cover from the image while maintaining all of the
78 canopy material - I confirmed this using manual inspection of the RGB imagery pre- and post-filtering.

2.2.4 Maxima Search

A search for local maxima (treetops) was performed using the open source remote sensing package ForestTools (7). I used the TreetopFinder() function with a linear sliding window using the function

$$f = 0.25x + 1.2 \tag{2}$$

The maximum search was thresholded to ignore maxima below 5m, because the great majority of crowns were above this height and no ground points were. A linear window function was selected based on the recommendations of the package documentation, and the two parameters of the function were manually tuned, starting from their default values and ending with those presented above, until the segmentation results on the test data appeared to be appropriately segmenting the canopies to individual trees. This manual testing and tuning was performed on several input image plots to ensure that performance would be acceptable across different tree sizes and degrees of canopy openness.

2.2.5 Segmentation

Following identification of treetops, watershed segmentation was performed, again using the ForestTools package (7). The SegmentCrowns() function was used with the treetops found above, with the minimum canopy height set to 3m. This means that while tree crown maxima were not permitted to occur below 5m (in section 2.2.4), tree crown edges were allowed to extend down to as low as 3m. This height limit prevented inclusion of ground points while still retaining most of the canopy structure. The SegmentCrowns() function returns empty polygonal lines when the format parameter is set to "polygons," and these were saved to a shapefile using the function writeOGR().

2.3 Alignment

2.3.1 Input Data

Alignment was performed between the provided ground and individual tree crown (ITC) datasets. The ITCs present in the dataset were divided up into training and test data as described in the parent paper (4). The ground data consisted of stem IDs, locations in latitude and longitude, stem heights, and stem diameters. The remote sensing ITC data consisted of crown IDs, location in latitude and longitude, crown area, and plot IDs.

2.3.2 Algorithm

First, stems were divided into plots. This information was not provided in the input data, but because the plots were small relative to the distance between plots, it was possible to cluster stem data into plots by simply iterating through the list and assigning each stem to the plot containing its nearest crown neighbor in latitude and longitude. In order to correct for the possibility of systematic error in ground-based estimation of location, all the stems in each plot were transformed by the difference in average coordinates between stem and crown groups within the plot.

Next, for each crown an estimate of crown diameter was formed using the known crown area and the assumption that crowns are roughly circular:

$$D = 2\left(\frac{A}{\pi}\right)^{0.5} \tag{3}$$

A linear regression was performed to find the relationship between the crown and stem diameter. A similar regression was run to test the relationship between stem height and crown diameter, but this relationship was found to be weaker, and a multiple regression model including both stem diameter and height was not found to predict crown diameter meaningfully better than stem diameter alone. The RMS error between all crown diameter predictions based on this equation and all actual crown diameter values was determined across the training dataset. Analogously, the RMS error in latitude and longitude was found for all training crowns versus the corresponding values for their paired stems.

Within each plot, a list of all possible pairings of stems and crowns was created and iterated through, with every pairing possibility being given a cost which was determined as follows:

$$C = \left(\frac{D - \hat{D}}{D_{rms}}\right)^2 + \left(\frac{X - \hat{X}}{X_{rms}}\right)^2 + \left(\frac{Y - \hat{Y}}{Y_{rms}}\right)^2 \tag{4}$$

3/8

118 where C is the cost, D is crown diameter estimated from crown area, $\hat{D}$ is crown diameter predicated
119 from the stem using the regression equation, D_{rms} is the RMS error in diameter predictions, X is crown
120 longitude, $\hat{X}$ is stem longitude, X_{rms} is the longitude prediction RMS error, Y is crown latitude, $\hat{Y}$ is
121 stem latitude, and Y_{rms} is the latitude prediction RMS error. The RMS error values were used to create
122 dimensionless costs across the three variables, scaled by the overall noise in each signal. Within each plot,
123 the set of pairings which minimized the total cost across all pairs was taken as the best alignment.

2.4 Classification

2.4.1 Input Data

126 The input data for this task consisted of a dataframe of individual pixel observations within each scene.
127 Each pixel was labelled with its associated crown ID and also contained the height within the CHM at that
128 point and the response in all the hyperspectral bands. Species labels were also provided for each crown
129 ID within the training set. Again, see (4) for more information on the input dataset.

2.4.2 Algorithm

131 First, all the pixels in each tree were aggregated into a single observation of that tree, containing the
132 average reflectance value in each band, the average height, the minimum height, the maximum height,
133 and the square root of the total number of pixels in the tree (as an approximation of crown diameter, since
134 each pixel is 1x1m).
135 The resulting tree vectors within the training set were grouped by species class. Two separate principle
136 component analysis routines were run on the structural and spectral parameters. The dimensionality of
137 the set was reduced by retaining only the three most informative dimensions from the PCA result for
138 the structural matrix and the ten most informative dimensions from the spectral PCA result. The vectors
139 within each class group were assembled into a prediction matrix for that class.
A set of maximum likelihood classifiers was built on the training data for each species class. Maximum
likelihood methods work by assigning likelihoods L based on the following equation:

$$L = (2\pi)^{-\frac{N}{2}}|Y|^{-0.5}exp\big(-0.5(t-T)^{T}\big)Y^{-1}(t-T)(9) \tag{5}$$

140 where N is the total number of variables used in the classifier (here bands and structural parameters of
141 trees), Y is the covariance matrix of the entire set of training vectors for the target class, t is the vector of
142 parameters for the tree to be classified, and T is the average vector from the set of training vectors for the
143 target class.
144 Each individual tree vector in the test set was transformed by the PCA parameters found above and
145 used to determine a likelihood of assignment to each class. Trees were then assigned to the class with the
146 highest likelihood score.

3 RESULTS

3.1 Segmentation

149 The segmentation routine had the weakest results out of the three algorithms implemented here. The
150 scoring used by the competition was based on the Jaccard index, which measures the overlap between
151 two sets, and is given as follows:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| + |A \cup B|} \tag{6}$$

152 This index ranges between 0 and 1, with 0 representing a complete lack of overlap and 1 representing
153 complete overlap. The index was calculated on the output trees compared to models of tree crowns
154 produced by the competition coordinators hand-drawing crown segmentations on the data (4). My
155 algorithm yielded $J = 0.184$, compared to the baseline score found by the organizers of $J = 0.0863$ which
156 used a simpler segmentation routine.
157 An example segmentation of a plot scene is given in Figure 1 below. Note the robustness of the routine
158 to areas that do not contain tree crowns, with bare soil or vegetative ground cover visible.

4/8
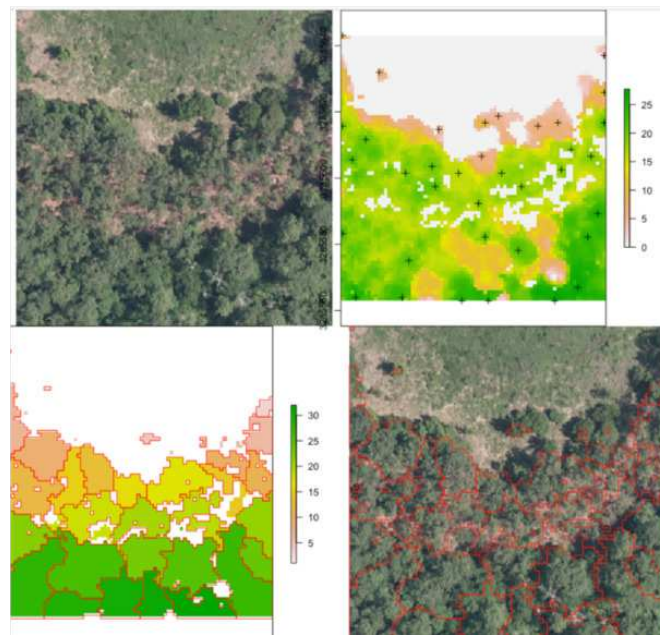
Figure 1. Segmentation results for one plot. a) Raw RGB image of the plot scene. b) CHM of the scene, filtered to exclude points with NDVI < 0.5 and height < 3m. Local maxima associated with presumptive treetops are shown in black marks. c) Output of the watershed segmentation routine showing polygonal crowns. d) Crown segmentations overlaid in red on the input RGB image.

3.2 Alignment

The regression equation found between crown diameter (as estimated by crown area) and stem diameter is given by $\hat{D_{crown}} = 0.148 D_{stem} + 0.549$. This equation was found to be significant (p < 0.0001) with $R^2 = 0.730$. A plot of crown diameters vs. stem diameters is given in Figure 2.
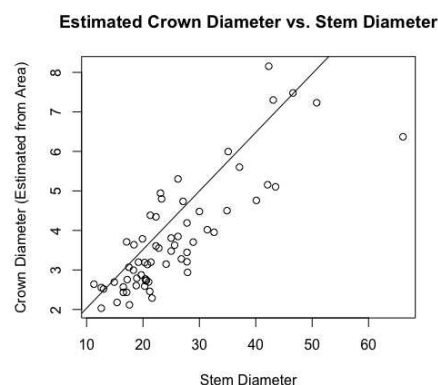
Figure 2. Linear regression for stem diameter (mm) versus crown diameter (m) as estimated from crown area. The point at far right in the plot was excluded from the regression as an outlier.

The alignment routine performed better overall than segmentation, and was able to correctly align 48% of the input crowns to the associated trees. However, an identical performance was yielded by the benchmark routine used by the competition organizers. That routine was very similar to the one I implemented, but used only latitude and longitude to align trees and ignored stem and crown diameters.

5/8

3.3 Classification

The classification performance of algorithms in the competition was measured by two metrics - rank-1 score (the percentage of all test trees correctly classified) and cross-entropy score (which rewards participants for expressing uncertainty about predictions). Higher rank-1 scores and lower cross-entropy scores are associated with "better" classification results.

The classification yielded a rank-1 score of 0.8226, indicating that it correctly classified 82% of all the trees in the set, and had a cross-entropy score of 1.2247. This can be compared to the 'baseline' method implemented by the competition organizers, which yielded a rank-1 score of 0.6667 and cross-entropy score of 1.1306.

My algorithm was the only method other than the baseline which had a 100% success rate at correctly identifying the most common tree in the dataset (the Longleaf Pine, *Pinus palustris*). It also yielded a very high success rate at identifying the second most common species (the Turkey Oak, *Quercus laevis*), at nearly 90%. However, it did not correctly identify any tree of any species other than these two most common species. Despite this, the two most common species represent a huge majority of the overall canopy in this system, covering 82.3% of all the trees in the training dataset.

4 DISCUSSION

4.1 Segmentation

Segmentation appears to be a major challenge within the remote sensing community, and no group participating in this competition was able to yield a segmentation Jaccard score of higher than 0.34. This problem is likely exacerbated by the heterogeneity of canopy structure between test plots in this test system. Segmentation algorithms are often sensitive to canopy structure differences such as degree of openness(6), and so in the future it might be beneficial to focus efforts on testing more algorithms which can automatically adjust their tuning based on the local openness and tree size within different areas of the canopy. My approach used only manual tuning, and work to automatically learn more effective parameters using training data represents a potential area for future improvement.

Because most of the classification results submitted were fairly powerful, it might be possible to perform pixel-wise classification to species first, and use this information to help inform the aggregation of those pixels into crowns (ie, trees of different species cannot be part of the same crown). An alternative possibility is an iterative process which first segments and classifies crowns at the pixel level and then splits crowns that appear to be combinations of two trees of differing species, or lumps adjacent crowns of the same species which might be different subcrowns within the same tree. It may also be worthwhile to try pulling in more data for the segmentation routine than just CHM results - especially the hyperspectral information.

4.2 Alignment

I was surprised that my alignment algorithm did not perform more strongly, especially because the strongest alignment algorithm implemented by another team used a method very similar to mine.(Dalponte and Gianelle) I think that there may have been a problem in my implementation of the crown area predictions, because my results were identical to those produced by the 'benchmark' algorithm which ignored crown area. Further investigation will be warranted in future work.

However, the broader applicability of these alignment algorithms is limited, because they rely on one-to-one datasets of perfectly labelled crowns and stems. They cannot perform on datasets in which not every stem is labelled, which is likely the case in a real forestry application. As well, the pairing algorithm I implemented scales very poorly with larger plot sizes - if it were to be implemented on a forest scale it would take an extremely long time to process all the potential pairings. Algorithms to automatically section the forest for alignment analysis into plots like those used here may be an area of potential interest for future work. Alternatively, it may be beneficial to rely instead on algorithms in which each target stem is compared to all of its nearest neighbor crowns, possibly with some cost to penalize multiple stems being mapped to the same crown.

4.3 Classification

My classification routine was extremely effective at identifying common species. In some contexts, this may be all that matters. For example, this would be entirely sufficient if the primary intent is to quantify the numbers in a community of a few very common "dominant" species. This would be the case for

efforts to take inventory of a wood production forest, or to calculate parameters related to gross system function like primary productivity, water filtration, or carbon sequestration. In these cases, the small number of misidentified rare species may not be important.

In other cases, this kind of result would not be acceptable. If the goal is to identify rare species in a community so that they can be managed for conservation, ability to recognize uncommon species may be important! However, while my algorithm was aimed primarily at identifying the most common species, it should be noted that no algorithm was able to yield strong performance on all the uncommon species. In fact, the best other submission for uncommon species identification was still only able to recognize four of the seven uncommon classes with greater than a 50% success rate - and this came at a loss in accuracy of about 10% at identifying the second most common species. It is clear that further work is warranted to develop algorithms that are capable of robustly recognizing rare species based on sparse training data.

5 CONCLUSION

My submission to the competition includes a tree segmentation, alignment, and classification pipeline which performs most strongly for common tree species. Consequently, it may be appropriate for applications such as maintenance of highly managed forestry plantations and efforts to estimate gross forest parameters in natural systems. Future work will focus primarily on improving the results of the segmentation algorithm, with emphasis also on improving the alignment of remotely collected and hand-labelled ground data. The latter will become especially important as the competition moves towards more realistic tree selection, potentially with overlapping plots and incomplete correspondence between the aerial and ground datasets.

As remote sensing methods continue to develop and the cost of deployment continues to decrease (with more and cheaper small aircraft), the technologies targeted by this competition may become increasingly important in a diverse array of disciplines, from agriculture to forestry to ecological research(5). The newly introduced NIST DSE Plant Identification Challenge may help to foster the development of systems for remote sensing analysis that are more streamlined and generalized across applications, which should aid their wider deployment across these fields. It is hoped that future competitions will continue to elaborate further on the methods developed here, and that this will aid in the expansion of remote sensing approaches into even more real world applications and fields.

6 ACKNOWLEDGEMENTS

7 SOFTWARE

The code developed for this analysis is available at: `https://github.com/conormcmahon/canopy_segmentation`

REFERENCES

[1] Barbosa, J. M. and Asner, G. P. (2017). Prioritizing landscapes for restoration based on spatial patterns of ecosystem controls and plant–plant interactions. *Journal of Applied Ecology*, 54(5):1459–1468.

[Dalponte and Gianelle] Dalponte, M. L. F. and Gianelle, D. In prep. neon nist data science evaluation challenge: methods and results of team fem.

[3] Fletcher, A. T. and Erskine, P. D. (2012). Mapping of a rare plant species (Boronia deanei) using hyper-resolution remote sensing and concurrent ground observation. *Ecological Management and Restoration*, 13(2):195–198.

[4] Marconi, S., Graves, S., Gong, D., Nia, M., Le Bras, M., Dorr, BJ. Fontana, P., Gearhart, J., Greenberg, C., Harris, D., Kumar, S., Nishant, A., Prarabdh, J., Rege, S., Bohlman, S., White, E., and Wang, D. (2018). A data science challenge for converting airborne remote sensing data into ecological information. *PeerJ Preprints*.

[5] Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4):358–371.

[6] Naidoo, L., Cho, M. A., Mathieu, R., and Asner, G. (2012). Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment.

[7] Plowright, A. (2018). Canopy analysis in r using foresttools.

[8] Rose, R. A., Byler, D., Eastman, J. R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H., Hansen, M., Headley, R., Hewson, J., Horning, N., Kaplin, B. A., Laporte, N., Leidner, A., Leimgruber, P., Morisette, J., Musinsky, J., Pintea, L., Prados, A., Radeloff, V. C., Rowen, M., Saatchi, S., Schill, S., Tabor, K., Turner, W., Vodacek, A., Vogelmann, J., Wegmann, M., Wilkie, D., and Wilson, C. (2015). Ten ways remote sensing can contribute to conservation. *Conservation Biology*, 29(2):350–359.

[9] Sisodia, P. S., Tiwari, V., and Kumar, A. (2014). Analysis of Supervised Maximum Likelihood Classification for remote sensing image. *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, pages 1–4.

[10] Tang, L. and Shao, G. (2015). Drone remote sensing for forestry research and practices. *Journal of Forestry Research*, 26(4):791–797.

[11] Vassallo, M. M., Dieguez, H. D., Garbulsky, M. F., Jobbágy, E. G., and Paruelo, J. M. (2013). Grassland afforestation impact on primary productivity: A remote sensing approach. *Applied Vegetation Science*, 16(3):390–403.