



Selecting hyperspectral bands for leaf mass per area prediction by means of neural networks

Catherine Baskiotis, Josselin Aval, Marwa El Bouz, Ayman Alfalou

► To cite this version:

Catherine Baskiotis, Josselin Aval, Marwa El Bouz, Ayman Alfalou. Selecting hyperspectral bands for leaf mass per area prediction by means of neural networks. IGARSS (IEEE International Geoscience and Remote Sensing Symposium) 2022, Jul 2022, Kuala Lumpur, Malaysia. hal-03638244

HAL Id: hal-03638244

<https://hal.science/hal-03638244>

Submitted on 1 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

SELECTING HYPERSPECTRAL BANDS FOR LEAF MASS PER AREA PREDICTION BY MEANS OF NEURAL NETWORKS

Catherine Baskiotis, Josselin Aval, Marwa El Bouz, Ayman Al Falou

L@bISEN, Light Scattering Learning Team, Yncrea Ouest, 29200 Brest, France

ABSTRACT

Accurate prediction of Leaf Mass per Area (LMA) from leaf-level hyperspectral reflectance measurements is fundamental for plant health watching.

Limiting the number of required bands to a very low number while maintaining a good accuracy of LMA prediction is of very high interest for cost-effective implementations. For such purpose, the best current model consists in a combination of vegetation indices and have been obtained by performing computationally intensive brute force approach. Further accuracy improvements are, at present, limited by current computational capacity limit.

We propose to use a neural network-based methodology for selecting a very low number of hyperspectral bands for further improving the LMA prediction accuracy. Testing on the publicly available Angers dataset, we diminished the Root Mean Square Error of prediction by 26% and 19% as compared to the best literature vegetation index and combination of vegetation indices, respectively.

Index Terms— Leaf Mass per Area, hyperspectral data, band selection, neural-networks, vegetation index

1. INTRODUCTION

Leaf Mass per Area (LMA) is a functional trait whose precise knowledge is fundamental for numerous plant health monitoring activities such as, for example, growth plant analysis, drought detection or nutrient stress detection [1].

Despite the possibility of using destructive bio-chemical methods for accurate LMA determination, the only route for cost-effective and intensive LMA watching consists in non-destructive methods, such as hyperspectral remote sensing.

Numerous current hyperspectral remote sensing methods are employing more than 100 entry bands for LMA prediction, such as the PROSPECT physical model [2] or Machine Learning models [2]. Reducing the number of required bands is of very high interest for cost reduction. For such purpose, vegetation indices are the simplest approach but suffers from a lack of precision [3,4].

Current most precise vegetation indices have been developed by a brute force approach, consisting in extensive one by one try/catch explorations of numerous band combinations [3,4]. At present, the further improvements of vegetation indices are, therefore, limited by current

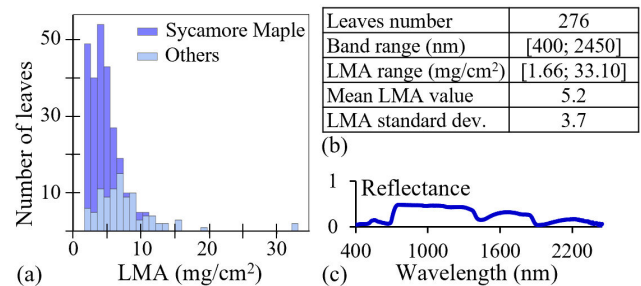


Figure 1: (a) Statistical repartition of LMA of the Angers dataset. (b) Table of characteristics of the Angers dataset. (c) Indicative reflectance spectrum of one leaf.

computational capacity limit.

Other approaches for LMA prediction from only a few bands, founded on numerical methods, such as Partial Least Square Regression (PLSR), have been tested but have shown to be of poorest accuracy as compared to current vegetation indices [4].

We propose a new neural network-based methodology for selecting a very few bands in the purpose of LMA prediction and demonstrate that using the selected bands as entries of a neural network allows to reach better precision than current literature vegetations indices.

2. MATERIAL

We consider the Angers publicly available database acquired in Angers (France) in 2003 [5,6]. This database contains $N=276$ leaves of 49 different dicotyledon species. One specie: Sycamore Maple, with 181 samples, is representing 65% of the total leaves number.

The LMA was measured by extracting small disks of fresh leaves, immediately weighting them, then drying them in an oven and immediately weighting them at the withdraw out of the oven [2]. LMA covers a large range of values from 1.66 mg/cm^2 to 33.10 mg/cm^2 (cf. figure 1.(a-b)).

The available leaf reflectance measurements span over the wavelength range [400nm, 2450nm], cf. figure 1.(b-c).

3. EVALUATION PROTOCOL AND METRICS

We divided the dataset into two non-overlapping subsets: the calibration and the test subset with, respectively 207

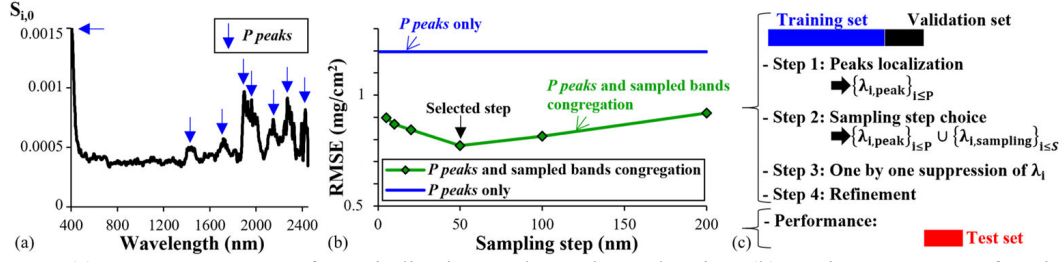


Figure 2: (a) LMA $S_{i,0}$ spectrum for an indicative random subsets drawing. (b) Tuning RMSE as a function of different sampling steps. (c) Summary of our band selection algorithm.

samples and 69 samples (75% and 25% of the dataset). The calibration subset was used to construct the models, while the test subset was used to evaluate the models.

We assessed the accuracy of the fit and the prediction on respectively the calibration and test subset through the Root Mean Square Error (RMSE) [3].

We performed 20 evaluations randomly drawings the calibration and test subsets. We determined, for each of the constructed model, the mean RMSE obtained over all the drawings on the calibration and test subsets.

4. NEURAL NETWORK BAND SELECTION

4.1 Choices for the neural network

The hyperspectral bands were set to be the multiple inputs of a neural network, with one unique output: the LMA value of the sample under consideration.

We used feed-forward neural networks with a number of hidden layers comprised in between 2 and 4. We used, respectively, the “tanh” and the “sigmoid” function as transfer function for, respectively, the hidden layers and the output layer. The trainings were performed using the back-propagation procedure [7].

4.2 Overfitting management

To avoid the over-fitting phenomenon, we further divided the calibration subset into training and validation subset with 138 samples and 69 samples, respectively [7].

The parameters of the neural network (namely the weight and bias of each neuron) were assessed by minimizing the RMSE on the training subset. The hyper-parameters (number of hidden layers, number of neurons on each layer, number of epochs, batch and learning rate) were chosen by minimizing the mean value between the RMSE on the training subset and the RMSE on the validation subset (*tuning RMSE*). In particular, we choose the number of epochs with the early-stopping procedure [7].

4.3 Band selection scheme

4.3.1 Input weights definition

We firstly trained the network by using all the available hyperspectral bands. We assessed the relative weight $S_{i,0}$ of

each band i on the output of the neural network following the methodology proposed in [8].

The figure 2.(a) presents an indicative $S_{i,0}$ spectrum obtained, for one of the 20 subsets drawings. We observe several bands at which $S_{i,0}$ exhibits a peak of relatively high value in a background where $S_{i,0}$ is almost constant and very small. We called *P peaks* the set constituted by these bands.

4.3.2 Band selection algorithm

We followed the protocol described in [8], which consists in the following backward elimination loop:

1. Train the network and stop the training by early stopping.
2. Compute $S_{i,0}$ for all the input bands.
3. Remove the band with the lowest coefficient $S_{i,0}$.
4. Iterate to step 1 until reaching one band only.

4.3.3 $S_{i,0}$ spectrum peaks and sampled bands congregation

Following the above-detailed algorithm, we should suppress one by one the bands starting by the whole reflectance spectrum (2050 bands). To save computing time, we decreased the number of starting bands by first selecting the *P peaks* (step 1 of our algorithm, cf. figure 2.(c)).

We, then, as second step, performed a sampling of the remaining bands with various sampling steps. After congregating the sampled bands and the *P peaks*, we evaluated the tuning RMSE for the different sampling steps and selected the sampling step which exhibits the minimum tuning RMSE (cf. figure 2.(b)). It is worth noting that, due to the “curse of dimensionality” [7], a too small sampling step leads to a relatively high value for the tuning RMSE.

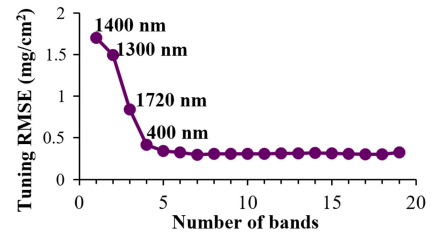


Figure 3: Details of the step 3 for an indicative random subsets drawing.

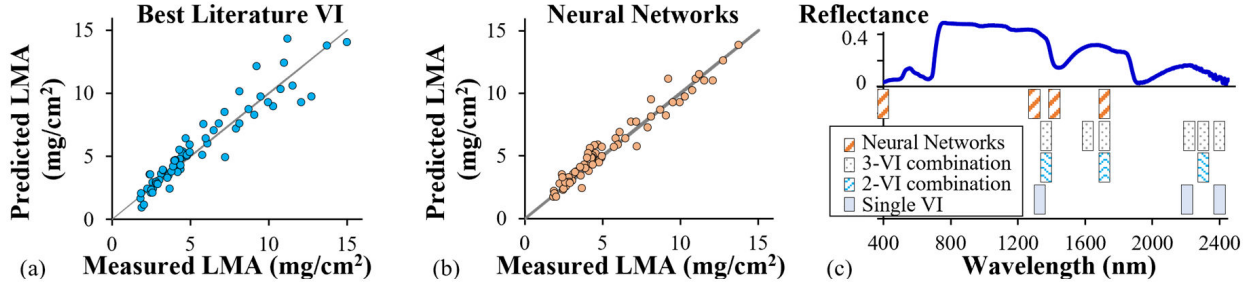


Figure 4: (a-b) Predicted to measured LMA values comparison (for the samples of the test subset) of an indicative random subsets drawing for the best literature vegetation index (a) and our neural network-based scheme (b). (c) Bands used by the different models.

We, thirdly, started, with the congregation of the P peaks and the S sampled bands, the band selection scheme described at the section 4.3.2 (step 3, cf. figure 2.(c)). When removing one by one the bands at the step 3, the tuning RMSE is keeping approximately the same value until reaching a given number of bands below which it increases dramatically (cf. figure 3).

We defined a stopping test for the step 3: does the suppression of one band implies an increase of the tuning RMSE greater than 5% as regard to its minimum value? In case of a positive answer the step 3 is stopped, and the remaining bands are selected.

4.3.4 Ultimate sets of bands

After performing the step 3, in the remaining set of bands, redundant bands may still be present [9]. To remove potential redundant bands, we applied a fourth step which consists in, first, testing the impact on the tuning RMSE of the removing of each of the remaining bands and then remove the band which produces the lower increase in RMSE. This step is ended with the same stopping criterion as those of the step 3.

At the end of the step 4, we obtain the *ultimate set* of bands which is specific to each of the 20 subsets drawings and contains from 3 to 9 bands.

For each ultimate set, the optimum neural network architectures were exhibiting 2 hidden layers with 6 to 10 neurons in the first hidden layer and 7 to 9 neurons in the second hidden layer.

We evaluated the mean RMSE over all the drawings on the training, validation and test subset (*mean training, validation and test RMSE*) using the ultimate sets and the optimum neural network architectures.

4.4 Unifying the set of selected bands

We constructed an *unified set* of bands by taking the four most frequent bands: 1720 nm, 1300 nm, 1400 nm and 400nm with an occurrence number of respectively 18, 16, 14 and 11 among the 20 subsets drawings.

5. VEGETATION INDICES REFERENCE

A vegetation index (VI) which predicts the LMA is a combination of the reflectance of the leaf at few different

bands, chosen in such a way that the LMA can be approximated by a simple function (*fitting function*) of this combination [3,4,10].

We tested 24 VI dedicated to LMA extracted from references [3,4,10], which consist either in a single index (single VI) either in a combination of two or three indices (2-VI or 3-VI combination).

For the fitting function, we used a polynomial function of second degree in one dimension when fitting a single index and in two (or three) dimensions when fitting a combination of two (or three) indices. We do not observe any improvement when increasing the degree of the polynomial function.

For each of the 20 subsets drawings, we determined the polynomial function coefficients of each VI by fitting the calibration subset.

We selected the VI which leads to the lower mean test RMSE over the 20 drawings. Among, respectively, the single vegetation indices, the 2-VI combinations and the 3-VI combinations, we found out, respectively, MND ($[R_{2285} - R_{1335}] / [R_{2285} + R_{1335} - 2R_{2400}]$), ND-R2300 (combination of $ND = [R_{1368} - R_{1722}] / [R_{1368} + R_{1722}]$ and R_{2300}) and MSR-ND-R2300 (combination of $MSR = [R_{2265} - R_{2400}] / [R_{1620} - R_{2400}]$, $ND = [R_{1368} - R_{1722}] / [R_{1368} + R_{1722}]$ and R_{2300}).

6. RESULTS

The mean calibration and test RMSE obtained, over the 20 drawings, with the best vegetation indices and our neural network-based scheme are depicted in table 1. The figure 4.(a-b) presents the predicted to measured LMA values comparison for the samples of the test subset of an indicative random subsets drawing.

It is worth noting that the best literature VI is the 2-VI

Model	Calibration	Test	Nb. of bands
Single VI	1.15	1.25	3
2 VI-combination	0.97	1.14	3
3 VI-combination	0.94	1.16	6
Neur. Net. (ultimate sets)	0.77	0.92	3-9
Neur. Net. (unified set)	0.79	0.94	4

Table 1: RMSE in mg/cm² obtained with the best literature vegetation indices and neural networks.

combination ND-R2300 and requires only 3 bands.

The neural networks tested with the ultimate sets of bands allow to diminish the RMSE on prediction (mean test RMSE) by 26%, 19% and 21% as compared to single VI, to 2-VI combination and to 3-VI combination, respectively.

7. DISCUSSION

7.1 Neural network superiority analysis

It is worth noting that our neural networks achieved better accuracy in LMA prediction than literature vegetation indices when PLSR failed [4]. We explain this property by the fact that PLSR is restricted to the investigation of linear relationships between band reflectances and LMA, while neural networks include non-linear relationships in their exploration.

Moreover, we explain the superiority of our neural network-based scheme on literature vegetation indices by the fact that current literature vegetation indices have been built by investigating a very little number of relationship types between band reflectance and LMA [3,4], while the neural networks explore a dramatically wider space of possible functions linking band reflectances and LMA [7].

7.2 Subset of predictive bands

The bands used by each model are detailed in figure 4.(c). All the bands of best literature LMA-VI (whatever the number of indices they are combining) are belonging to Short Wave InfraRed (SWIR) domain. Moreover, one band close to the 1450 nm water peak absorption is systematically present in the best literature LMA-VI (1335nm, 1368nm and 1368nm for the best single VI, 2-VI combination and 3-VI combination, respectively).

Neural network selected band set and best literature VI (ND-R2300) band set share one band in common: 1720 nm. This band is close to an absorption peak of the lignin, the starch and the nitrogen [11]. As compared to the band set used in the best literature VI (ND-2300), the neural network selected band set presents two specific characteristics: firstly, the presence of two bands (1300 nm and 1400 nm) and not only one band near the 1450 nm water absorption peak; secondly the presence of a band in the visible domain (400 nm) at which the absorption of pigments is high and the absorption of water and dry matter is low.

8. CONCLUSION

We propose to use a neural-network based methodology for selecting an over-restrained number of leaf-level hyperspectral reflectance bands for the prediction of LMA. This methodology takes accounts of the nonlinear relationship between the band reflectances and the LMA.

By performing tests in cross-prediction configuration on the Angers dataset, we found out that the neural-network based methodology allows to improve the RMSE of prediction by 26%, 19% and 21% as compared to current

literature single VI, 2-VI combination and 3-VI combination vegetation indices. Very simple neural networks architectures with only two hidden layers are enough for achieving such result.

The eventually selected bands by our neural network scheme are: 1720nm, 1300nm, 1400nm and 400nm. Current works targeting to build our own measurements database are ongoing to further investigate the involvement of those bands in LMA modelling.

9. ACKNOWLEDGMENTS

This work was funded by the Brittany Region (France) and the L@bISEN, an Yncrea Ouest laboratory (France).

10. REFERENCES

- [1] Poorter H, Niinemets Ü, Poorter L, Wright IJ, Villar R., "Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis", *New Phytologist*, **182**, pp. 565-588, 2009.
- [2] J.-B. Féret, G. le Maire, S. Jay, D. Berveiller, R. Bendoula, G. Hmimina, A. Cheraïet, J.C. Oliveira, F.J. Ponzoni, T. Solanki, F. de Boissieu, J. Chave, Y. Nouvellon, A. Porcar-Castell, C. Proisy, K. Soudani, J.-P. Gastellu-Etchegorry, M.-J. Lefèvre-Fonollosa, "Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning", *Remote Sensing of Environment*, **231**, pp. 110959, 2019.
- [3] G. le Maire, C. François, K. Soudani, D. Berveiller, J.Y. Pontailler, N. Bréda, H. Genet, H. David, E. Dufrêne, "Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass", *Remote Sensing of Environment*, **112**, pp. 3846-3864, 2008.
- [4] Y. Chen, J. Sun, L. Wang, S. Shi, W. Gong, S. Wang, T. Tagesson, "Optimized Estimation of Leaf Mass per Area with a 3D Matrix of Vegetation Indices", *Remote Sensing*, **13**, pp. 3761, 2021.
- [5] Database available at the website: <http://opticleaf.ipgp.fr/index.php?page=database>
- [6] Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., "PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments", *Remote Sens. Environ.*, **112**, pp. 3030-3043, 2008.
- [7] I. Goodfellow, Y. Bengio and A. Courville, "Deep Learning", MIT Press, MIT, 2016.
- [8] M. Yacoub, Y. Bennani, "Features selection and architecture optimization in connectionist systems", *International Journal of Neural Systems*, **10**, pp. 379-95, 2000.
- [9] B.-H. Adil, Y. Ghanou, El. Abderrahim, "Hybrid Method HVS-MRMR for Variable Selection in Multilayer Artificial Neural Network Classifier", *International Journal of Electrical and Computer Engineering*, **7**, pp. 2773-2781, 2017.
- [10] T. Erudel, S. Fabre, T. Houet, F. Mazier, and X. Briottet, "Criteria Comparison for Classifying Peatland Vegetation Types Using In Situ Hyperspectral Measurements", *Remote Sensing*, **9**, pp. 748, 2017.
- [11] S. Jacquemoud and S. Ustin, *Leaf Optical Properties*, Cambridge University Press, Cambridge, 2019.