Grapevine nitrogen retrieval by hyperspectral sensing at the leaf and canopy level

Authors: Yuto Kamiya[1,2], Alireza Pourreza[1*],Sirapoom Peanusaha[1], Matthew W. Fidelibus[3,4]

[1]*Department of Biological and Agricultural Engineering, University of California Davis, 3042 Bainer Hall, Davis, CA, 95616, United States*
[2]*Kubota Tractor Corporation, 1000 Kubota Drive, TX, 76051, United States*
[3]*Kearney Agricultural Research and Extension Center, 9240 S. Riverbend Avenue, Parlier, CA 93648, USA*
[4]*Department of Viticulture and Enology, University of California, Davis, 595 Hilgard Ln, Davis, CA 95616, USA*

*Corresponding author:*apourreza@ucdavis.edu*

Abstract:

Context and purpose of the study – Grapevine nitrogen (N) monitoring is essential for efficient N management plans that optimize fruit yield and quality while reducing fertilizer costs and the risk of environmental contamination. Unlike traditional vegetative-tissue sampling methods, remote sensing technologies, including hyperspectral imaging, have the potential to allow monitoring of the N status of entire vineyards at a per-vine resolution. However, differential N partitioning, variable spectral properties, and complex canopy structures hinder the development of a robust N retrieval algorithm. The present study aimed to establish a solid understanding of vine spectroscopic response at leaf and canopy levels by evaluating the different nitrogen retrieval approaches, including the radiative transfer model.

Material and methods – At the leaf level, N content and its relative position within a shoot were measured along with the proximal hyperspectral reflectance (350nm-2500nm) from 'Flame Seedless' vines grown in pots as well as 'Solbrio' vines in a vineyard. At the canopy level, leaf nitrogen concentration, and hyperspectral images (400nm-1000nm) of 'Valley Pearl' vines were collected using a hyperspectral camera mounted on an uncrewed aerial vehicle. At leaf and canopy levels, we evaluated the N retrieval performance of several spectral analytics approaches, including empirical data-driven models, a physical-based model (radiative transfer model), and hybrid models.

Results – At the leaf level, the performance of data-driven approaches using the entire 350-2500 nm spectrum (chemometrics and machine learning) outperformed (R^2=0.76-0.78) the use of vegetation index, physical-based modeling, and hybrid approaches. However, collecting and analyzing hyperspectral data within visible, near-infrared, and shortwave infrared is unrealistic for large-scale monitoring. Protein, one of the variables retrieved by a physical-based approach, showed high potential to be used as a predictor of N content because protein, unlike chlorophyll, remained consistently correlated with N content regardless of leaf age. At the canopy level, the performance of data-driven and hybrid approaches was competitive (R^2=0.61-0.69) except for the combination of physical-based parameters and random forest regression (R^2=0.50). However, the performance of N content retrieval models varies widely across datasets, and it is not yet clear what factors determine the performance of models. Further data processing and calibration to extract more reliable spectral features from hyperspectral images are required to scale N retrieval from the leaf level to the canopy level by leveraging the knowledge acquired at the leaf level analysis.

Keywords: Grapevine, Nitrogen Retrieval, Hyperspectral, Radiative Transfer Model, Unmanned Aerial Vehicle, Proximal Hyperspectral

1. Introduction

Efficient management of vineyard Nitrogen (N) status is essential for optimizing yield and quality while simultaneously reducing the environmental effect of excessive use of N fertilizer, such as Nitrate contamination in groundwater (Chlingaryan et al., 2018; Fu et al., 2021; Harter et al., 2012; Jiang et al., 2019). Remote sensing technology offers promising crop N monitoring results (Berger et al., 2020; Zheng et al., 2018). However, several factors, such as phenological stages, environmental conditions (Jafarbiglu & Pourreza, 2023), and complex canopy structure, pose challenges in developing a robust N retrieval algorithm that can be applied across different phenological stages and cultivars.

Berger et al. (2020) conducted a comprehensive review of N retrieval algorithms used in N monitoring. Most of the studies relied on the vegetation index (VI). Machine learning and chemometrics were among the top-performing approaches but lacked interpretability and generalizability. The hybrid model offers interpretability from the radiative transfer model (RTM) while providing computational efficiency and accuracy offered by machine learning regression (Berger et al., 2020). At the leaf level, this study aimed to establish a solid understanding of leaf spectroscopic response to N content by conducting per-leaf nutrient analysis using a proximal sensor to eliminate environmental effects from flight and canopy structure. At the canopy level, this study aimed to observe the spectral analysis performances of major N retrieval algorithm approaches and their trends among data collected from several phenological stages.

2. Material and methods

Plant materials – Experiments were conducted at the University of California Agricultural and Natural Resources Kearney Research and Extension Center in Parlier and a commercial table grape vineyard in Shafter, California. Leaf level data were collected from potted ‘Flame Seedless’ vines and ‘Solbrio’ grafted on Freedom rootstock grown in a vineyard. At the canopy level, the data were collected from ‘Valley Pearl’ table grapes in a commercial vineyard near Shafter, California. Potted vines (Flame Seedless) were fertigated with nutrient solutions containing different levels of N (approximately 1 to 7.5 mM). Field-grown vines (Solbrio and Valley Pearl) were supplied with different quantities of ammonium sulfate and calcium nitrate (between 0–72 g N/vine, annually), to induce different N sufficiency levels. The vines were irrigated as needed using estimated daily evapotranspiration as a guide for water replacement.

Plant measurement – At the leaf level, N content and its relative position within a shoot were measured along with the proximal hyperspectral reflectance (HR-1024i, Spectra Vista Corp, NY, USA, 350nm-2500nm). For ‘Flame Seedless’, data was collected in June 2021 and July 2021. For ‘Solbrio’, the data was collected during Veraison stage in June 2022. Individual leaf fresh weight, dry weight, leaf area, and total N (TN) were collected. At the canopy level, imaging and tissue sample data were collected in 2020 and 2021 using an aerial hyperspectral camera (Pika L, Resonon Inc. Bozeman, MT, USA, 400 – 1000 nm). The UAV flew at a height of 15-30m around solar noon and captured the vineyard in the nadir direction. The radiance images were converted to reflectance images using irradiance captured by irradiance sensor with a cosine corrector mounted on the UAV. The region corresponding to each tree was extracted from the hyperspectral image to obtain the average reflectance spectrum of the canopy. Within a few days of imaging, approximately 40 leaf blades per vine were collected at each of four data collection campaigns (Prebloom, Bloom, Fruit set, Veraison). Leaf blades were collected from opposite the clusters at Prebloom, Bloom, and Fruit set, and from recently expanded leaves at Veraison. Leaf blades were separated from petioles before being washed, dried, ground, and sent to a commercial lab to determine TN values expressed in the percentage of leaf dry weight.

Data processing and analysis – We evaluated several spectral analytics approaches for retrieving N, including a data-driven, physical-based, and hybrid model. The data-driven model is vegetation index (VI)-based N prediction and machine learning-based N prediction using reflectance spectrum as inputs. The physical-based model relied on the RTMs to estimate biochemical parameters of the leaf and canopy from the reflectance spectrum, which were then used to predict N. The hybrid model combined a machine learning algorithm with biochemical parameters estimated by RTMs to predict N. We selected partial linear regression (PLSR) (Wold et al., 2001), random forest regression (RFR) (Segal, 2004), and Gaussian process regression (GPR) (Rasmussen , 2004) algorithms for machine learning algorithms. We used the inversion of PROSPECT-PRO (Féret et al., 2021) and PROSAIL to estimate the biochemical parameters of the leaf and canopy, respectively.

3. Results and Discussion

3.1 Protein estimated by PROSPECT can predict leaf nitrogen levels consistently across different leaf ages.

Chlorophyll content estimated by PROSPECT is shown to be highly correlated with widely used vegetation indexes, such as the Normalized Difference Red Edge index (NDRE) (Rodriguez et al., 2006). However, chlorophyll shows an inconsistent relationship with N as the leaf age increases due to the inverted bell curve trend between chlorophyll and leaf age (Poni et al., 1994). Younger leaves with high N content contain low chlorophyll. Conversely, protein shows a consistent linear relationship with N content regardless of leaf age. In the hybrid model using PROSPECT predicted leaf biochemical parameters and RFR, the feature importance shows that protein content is critical in the decision tree to estimate the leaf N content.

3.2 Hybrid method shows potential of reducing required wavebands by 50% while maintaining acceptable accuracy and increasing model interpretability.

PLSR and GPR showed high accuracy of 0.76 and 0.78, respectively, using 5-fold cross-validation and the entire waveband from 350 to 2500 nm. However, collecting visible, near-infrared, and short-wave infrared is impractical for large-scale monitoring. A hybrid model that combined RFR and PROSPECT predicted leaf biochemical parameters could achieve an accuracy of about 0.54 using test and training data from all three data collection campaigns. PROSPECT required about 50% fewer wavebands compared to PLSR and GPR. Comparatively, when the wavebands used in PLSR and GPR were reduced from 350-2500 nm to 350-900 nm, the R^2 value decreased dramatically to 0.41 and 0.38, respectively.

3.3 Performance of N prediction in canopy level

N prediction models were only effective for the dataset collected at the bloom stage, with poor performance for the other three datasets ($R^2 < 0.4$). At the bloom dataset, the data-driven approaches, specifically PLSR and GPR using reflectance spectra as inputs, as well as PLSR using all VIs as inputs, and the hybrid model (PLSR with all VIs and biochemical parameters as inputs) achieved the best N prediction performance (R^2=0.68-0.69). However, other approaches also showed comparable N prediction performance. Therefore, the N retrieval approach had minimal impact on the prediction performance. The high correlation between VI and N was observed only in the bloom dataset, explaining the moderate prediction performance and unsuccessful predictions in other datasets. Improving N prediction requires identifying the conditions leading to high VI-N correlation and conduct remote sensing under those conditions. Additionally, developing techniques to predict N without relying on the high correlation between VI and N is necessary. One possible next step is to leverage the 3D RTM to develop an analysis technique to extract important optical traits with varying light conditions, sun-object-sensor geometry, and canopy structure.

4. Conclusions

At the leaf level, the hybrid models combining RTM and machine learning show potential in increasing model interpretability and reducing the number of bands required while maintaining acceptable prediction accuracy ($R^2 \geq 0.50$). Mass-based protein content shows high potential to be used as a surrogate variable to predict mass-based N content. Nonetheless, based on the correlation pattern of the biochemical parameters and feature importance calculated RFR, chlorophyll still plays a significant role in N prediction. At the canopy level, N prediction was only successful when there was a high correlation between VI and N, and the choice of retrieval approach had minimal impact on prediction performance. Improving N prediction would require a method that robustly extracts critical optical traits considering the variations in light conditions, sun-object-sensor geometry, and canopy structure.

5. Acknowledgements

This work was in part supported by USDA-NIFA Specialty Crop Research Initiative Award No. 2020-51181-32159, and the California Table Grape Commission. Additionally, the authors would like to thank Mario Salinas, and

Hamid Jafarbiglu from UC Davis Digital Agriculture Lab, and Larry Williams and Jaclyn Stogbauer from the Department of Viticulture and Enology, UC Davis, for their help, advice, and directing related papers and materials.

6. Literature cited

Berger, K., Verrelst, J., Féret, J. B., Hank, T., Wocher, M., Mauser, W., & Camps-Valls, G. (2020). Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. International Journal of Applied Earth Observation and Geoinformation, 92(April), 102174. https://doi.org/10.1016/j.jag.2020.102174

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151(November 2017), 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Féret, J. B., Berger, K., de Boissieu, F., & Malenovský, Z. (2021). PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sensing of Environment, 252(October 2020). https://doi.org/10.1016/j.rse.2020.112173

Fu, Y., Yang, G., Pu, R., Li, Z., Li, H., Xu, X., Song, X., Yang, X., & Zhao, C. (2021). An overview of crop nitrogen status assessment using hyperspectral remote sensing: Current status and perspectives. European Journal of Agronomy, 124(December 2020), 126241. https://doi.org/10.1016/j.eja.2021.126241

Harter, T., Lund, J. R., Darby, J., Fogg, G. E., Howitt, R., Jessoe, K. K., Pettygrove, G. S., Quinn, J. F., & Viers, J. H. (2012). Addressing Nitrate in California's Drinking Water with a focus on Tulare Lake Basin and Salinas Valley Groundwater. Report for the State Water Resources Control Board Report to the Legislature, 1–78. http://groundwaternitrate.ucdavis.edu/files/138956.pdf

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. Remote Sensing of Environment, 34(2), 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z

Jafarbiglu, H., & Pourreza, A. (2023). Impact of sun-view geometry on canopy spectral reflectance variability. ISPRS Journal of Photogrammetry and Remote Sensing, 196(January), 270–286. https://doi.org/10.1016/j.isprsjprs.2022.12.002

Jiang, Z., Zhong, Y., Yang, J., Wu, Y., Li, H., & Zheng, L. (2019). Effect of nitrogen fertilizer rates on carbon footprint and ecosystem service of carbon sequestration in rice production. Science of the Total Environment, 670, 210–217. https://doi.org/10.1016/j.scitotenv.2019.03.188

Poni, S., Intrieri, C., & Silvestroni, O. (1994). Interactions of Leaf Age, Fruiting, and Exogenous Cytokinins in Sangiovese Grapevines Under Non-Irrigated Conditions. II. Chlorophyll and Nitrogen Content. American Journal of Enology and Viticulture, 45(3), 278–284. http://www.ajevonline.org/cgi/content/abstract/45/3/278

Rasmussen, C. E. (2004). Gaussian Processes in Machine Learning. In O. Bousquet, U. von Luxburg, & G. Rätsch (Eds.), Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures (pp. 63–71). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-28650-9_4

Rodriguez Glenn J.; Belford, R.; Christensen, L. K., D. F. (2006). Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts. Australian Journal of Agricultural Research, 57(7), 781–789. https://doi.org/10.1071/ar05361

Segal, M. R. (2004). Machine Learning Benchmarks and Random Forest Regression Publication Date Machine Learning Benchmarks and Random Forest Regression. Center for Bioinformatics and Molecular Biostatistics, 15. https://escholarship.org/uc/item/35x3v9t4

Wold Michael; Eriksson, Lennart, S. S. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130. https://doi.org/10.1016/s0169-7439(01)00155-1

Zheng Tao; Li, Dong; Zhou, Xiang; Yao, Xia; Tian, Yongchao; Cao, Weixing; Zhu, Yan, H. C. (2018). Evaluation of RGB, Color-Infrared and Multispectral Images Acquired from Unmanned Aerial Systems for the Estimation of Nitrogen Accumulation in Rice. Remote Sensing, 10(6), 824-NA. https://doi.org/10.3390/rs10060824

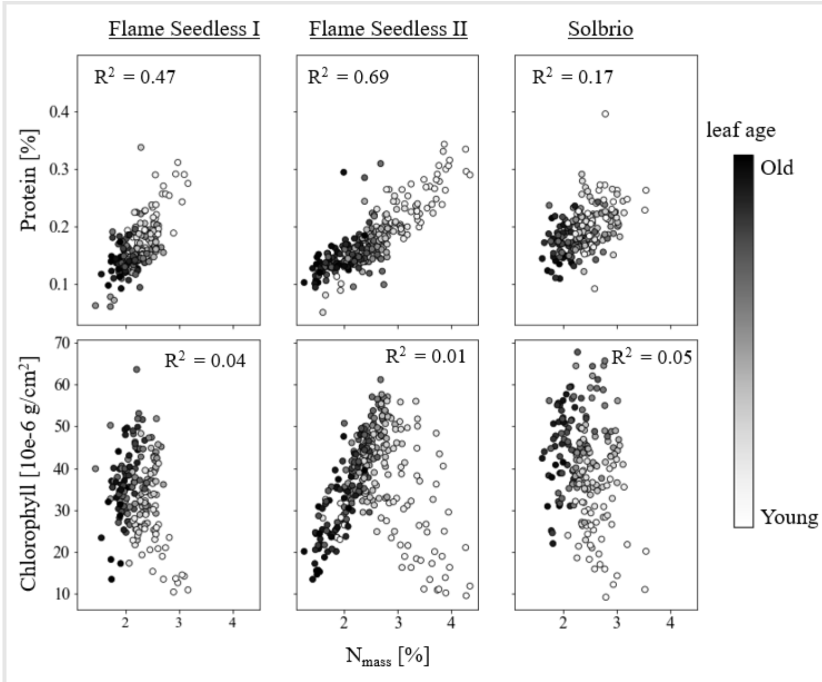

Figure 1: Relationship between N, protein and chlorophyll content across different range of leaf age. R^2 values were determined using either protein or chlorophyll as a single predictor in a simple linear regression. Leaf age was estimated based on the corresponding leaf position on the shoot.

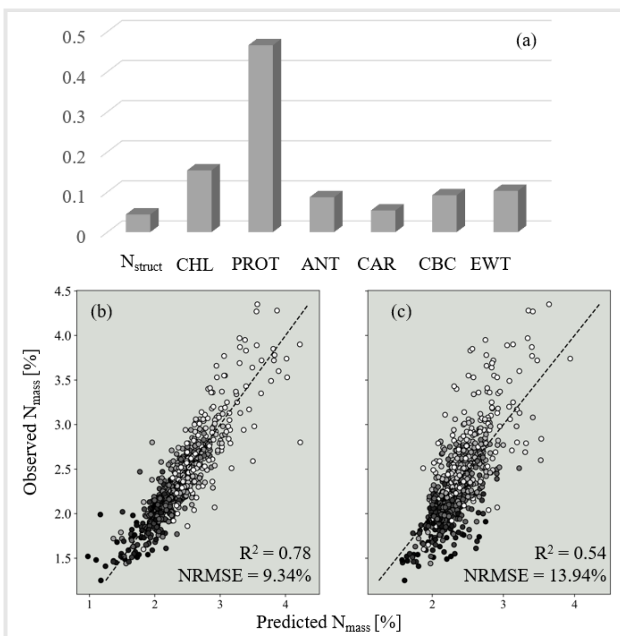

Figure 2: Feature importance of RTM parameters in hybrid mode using RTM parameters in RFR in (a). Predicted versus observed N_{mass} using GPR in (b) and the hybrid model in (c). The color of the data points in (b) and (c) correspond to leaf age using the same color scale as Figure 1. The coefficient of determination (R^2) and normalized root-mean-squared error (NRMSE) of each model were calculated using combined data from all three campaigns with Flame Seedless and Solbrio.