

Question: What machine learning/deep learning have you used and what were the specific applications? Why did you select those techniques?

The main machine learning algorithm that I have incorporated in my work is partial least square regression (PLSR). I have used PLSR to retrieve canopy nitrogen content from hyperspectral data and compared different feature selection techniques. The main approach is to fit a statistical relationship between canopy nitrogen and spectral bands. PLSR has some properties that make a suitable choice for trait retrieval from hyperspectral sensors. Most often in remote sensing, the training data set has fewer observations than predictive features (e.g. hundreds of spectral bands in this project) and many of these features likely have a strong correlation with each other. These challenges require specific treatments such as dimension reduction. PLSR is a powerful dimension reduction technique that can handle these situations. Moreover, PLSR regression coefficients and feature selection techniques can provide an insight into the interpretability of the model such as light-canopy interactions. Finally, PLSR is fast to train and implement which makes it suitable for large-scale remote sensing mapping of vegetation traits. From a practical perspective, PLSR is widely implemented in different ecosystems to retrieve vegetation traits (e.g. <https://ecosml.org/>) and it is highly likely to be a standard algorithm for some vegetation traits retrieval in future NASA missions such as the Surface Biology, Geology designated observable.

Hamid Dashti