Data Averaging Enhancements of the Predictive Accuracy of Machine-Learning-Based Microwave Sensing for Estimating Cranberry Fruit Yields

A. Haufler, J. Booske, and S. C. Hagness
Department of Electrical and Computer Engineering,
University of Wisconsin-Madison, Madison, Wisconsin, USA

We are developing a machine-learning-based microwave sensing approach to estimate the total mass of cranberries within a cranberry canopy volume for one square foot surface area of canopy (the cranberry yield). In our approach, multiple microwave backscatter measurements are acquired for each unique site in the cranberry bed by rotating or translating the sensor above the same one square foot surface area of canopy. Previously, we treated each measurement as a unique scattering signature and determined the probability distribution of error in yield prediction for these single scattering-signature inputs to the machine-learning algorithm (Haufler et al., IEEE International Symposium on Antennas and Propagation, Boston, MA, July 2018, pp. 1147-1148). In this work, we take advantage of the availability of multiple measurements for the same canopy location and explore accuracy enhancements of our approach for cranberry yield estimation using different data averaging schemes.

There are multiple points within the data processing framework that allow for data averaging to enhance the yield prediction at a specific location. First, multiple frequency-dependent scattering signatures (normalized reflected power measurements at many discrete frequencies) obtained from a specific location can be averaged at each frequency, prior to analysis by the machine-learning algorithm. Second, our algorithm reduces the high dimensional scattering signatures at many discrete frequencies to a low dimensional space. These dimension-reduced scattering signatures can be averaged after the reduction step. Third, the individual scattering signatures can be individually processed by the algorithm, to produce individual yield estimates that are subsequently averaged into a final yield estimate. At the third point, the independent estimates can be obtained from a single machine-learning algorithm or from multiple, independently trained algorithms.

We investigated the impact of these three data-averaging schemes on the predictive accuracy of our approach. We used microwave scattering signatures obtained from full-wave simulated cranberry canopies and commercial cranberry beds in central Wisconsin with the near-field sensor from the prior investigation. Our findings indicate that significant improvement in accuracy is achieved with appropriate averaging of data.