A Pilot Study of the Effects of Soil Moisture and Training Data Diversity on the Accuracy of Machine-Learning-Based Microwave Sensing for Cranberry Yield Estimation

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

Accurate and efficient tools for estimating cranberry yield are vital to farmers, cranberry genetics researchers, and industry handlers who gauge crop supply. Previously, we conducted a feasibility study of the predictive accuracy of a near-field active microwave sensing approach to cranberry yield estimation (A. Haufler, J. Booske, S. C. Hagness, and B. Tilberg, IEEE International Symposium on Antennas and Propagation, Boston, MA, July 2018, pp. 1147-1148). We refer to cranberry yield as the total mass of the cranberries within the cranberry canopy volume for one square foot surface area of canopy. The study's results provided evidence for the feasibility of cranberry yield estimation using machine learning in conjunction with the measured backscatter.

The performance of machine learning algorithms depends on the diversity of data in the training data set. The characteristics of the microwave backscatter depend not only on the cranberry yield but also on soil moisture content, which can vary significantly depending on rainfall and irrigation practices. Thus, variable soil moisture content may be a source of potentially undesirable diversity in the backscatter signatures for a fixed yield. On the other hand, diversity in the set of distinct cranberry yields used to train the machine-learning algorithm is essential for ensuring that the algorithm performs well in estimating yield from new backscatter input.

In this work, we explore the effects of soil moisture as well as training-data yield diversity on the predictive accuracy of our approach to cranberry yield estimation. First, we developed full-wave computational electromagnetic simulation testbeds to model the microwave response for varying canopy scenarios in terms of the soil moisture and the cranberry yields. We also acquired backscatter measurements from 100 one-square-foot sites within cranberry beds in central Wisconsin. In our experiments, we measured the microwave backscatter with our near-field sensor and volumetric water content with a commercial soil moisture meter. The simulation testbed offers the advantage of a controlled environment while the experimental testbeds provide the most realistic data sets for our study. We then analyzed how changes in simulated and measured soil moisture and training-data yield diversity impact the predictive accuracy of our approach. Our findings suggest that both parameters play an important role and may be leveraged to improve performance.