Identifying optimal measurement protocols for giving fertilization advice

Supervisors (TU Delft): Neil Budko, Corné Verburg
Supervisors (CropConsult): Jerke Eisma

Introduction

High-throughput phenotyping methods make it possible to collect an increasing amount of information about potato growth and environmental conditions. In practice, agronomists combine measurements, model outputs, weather information and expert knowledge to formulate fertilization advice. This decision process is often not explicitly documented, making it difficult to determine which measurements actually influence the advice and which provide little additional information. The goal of this project is to use historical data to make this advisory process more transparent, by identifying which measurements are most informative for reproducing fertilization advice and whether a smaller set of measurements could provide essentially the same information.

CropConsult, a company that provides advice to farmers, has collected field data containing soil measurements, crop biomass and nitrogen status, weather information, previous fertilizer applications, and the corresponding fertilization advice.

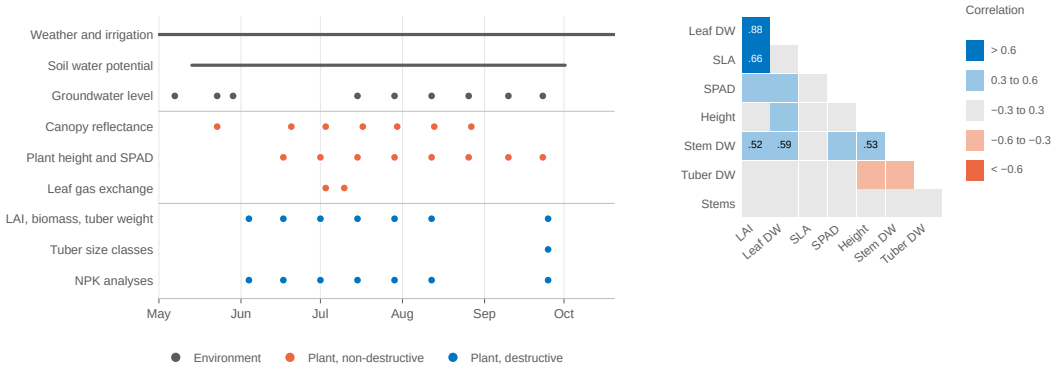


Figure 1. Sampling dates per measurement type at Vredepeel 2019 (left) and correlations between plant traits within sampling dates (right). Data: ten Den et al. (2024).

Let A denote a fertilization advice and let $\mathbf{x} = (x_1, \ldots, x_n)$ denote the information available at the time the advice was formulated. The advisory process can then be represented by the conditional distribution

$$p(A \mid \mathbf{x}). \tag{1}$$

The relevance of individual measurements can be studied by quantifying how much additional information a variable x_j provides about the advice once the remaining measurements are known. This can be done, for example, by comparing the conditional distributions

$$p(A \mid \mathbf{x}), \quad \text{and} \quad p(A \mid \mathbf{x}_{-j}), \tag{2}$$

which represent the advice given with and without the measurement x_j, respectively. The project will investigate how to quantify the difference between these two distributions and how to use this information to identify which measurements are most informative for reproducing the advice.

Reference for colloquium

J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, 24, 175–186, 2014.

Other references

I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, 3, 1157–1182, 2003.

T. ten Den, I. van de Wiel, F. van Evert, M. van Ittersum, A. de Wit, and P. Reidsma, "Agronomic dataset on potato growth and yield in the Netherlands," *Open Data Journal for Agricultural Research*, 10, 1–13, 2024.