Machine learning for selecting crop varieties as climate adaptation measure

George A.K. van Voorn, Bernardo Maestrini, Ponraj Arumugam, Martin Boer, Shota Gugushvili

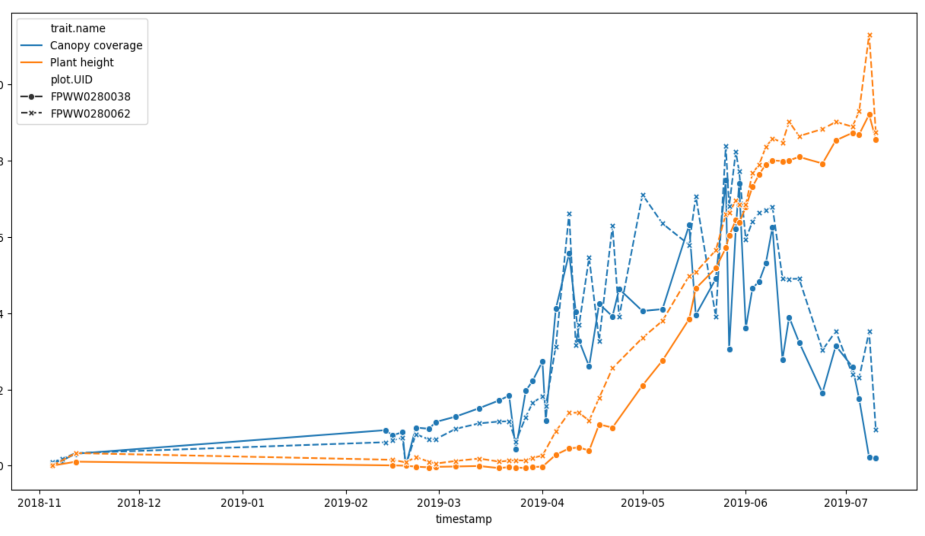
D3-C2 2023 project

Objective

This project aims to develop hybrid methodologies based on machine learning, statistics, and (dynamic) process-based modelling as a proof-of-principle for a tool for the selection of optimal crop genotypes for changing climates.

Background

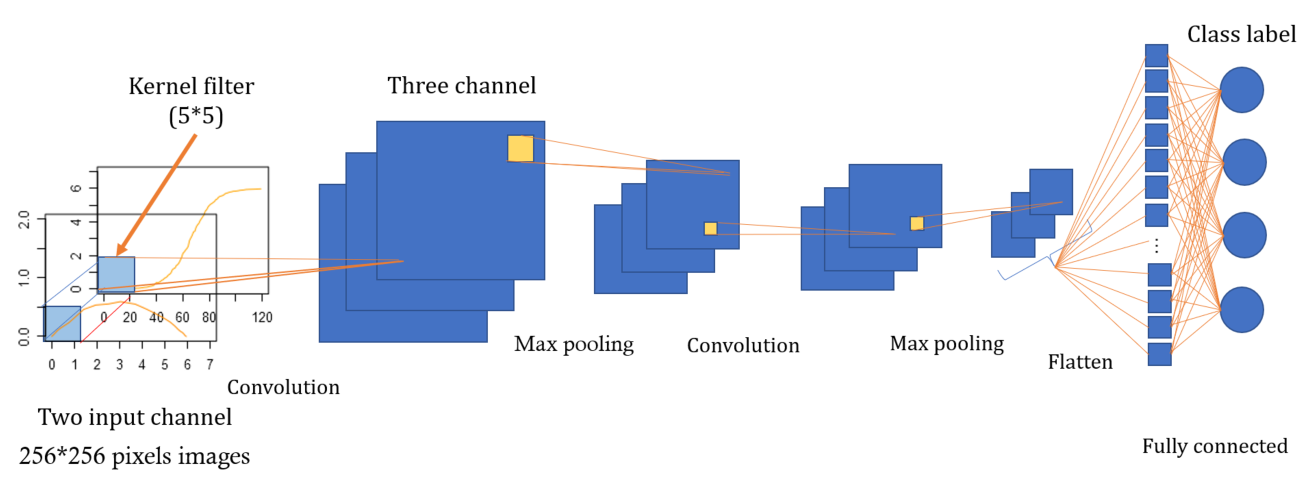
- Climate change affects conditions for crop growth.
- Crop traits like yield result from time-dynamic GxE (genotype-by-environment interactions).
- Selecting new crop genotypes suitable for new conditions **requires forecasting**.
- Current crop modelling lacks reliable descriptions of essential GxI under heat conditions.
- Increasing availability of time series data of crop phenotypic traits of multiple genotypes in multiple environments (High Throughput Phenotyping, remote and proximal monitoring).
- We aim to combine statistical, Machine Learning (ML), and process based modelling approaches to obtain GxI decribing heat responses for crop models from these time series data.



Above, left. An experimental monitoring setup.
Above, right. Typical time series data from experiments by partners (courtesy of Lukas Roth, ETH Zürich)

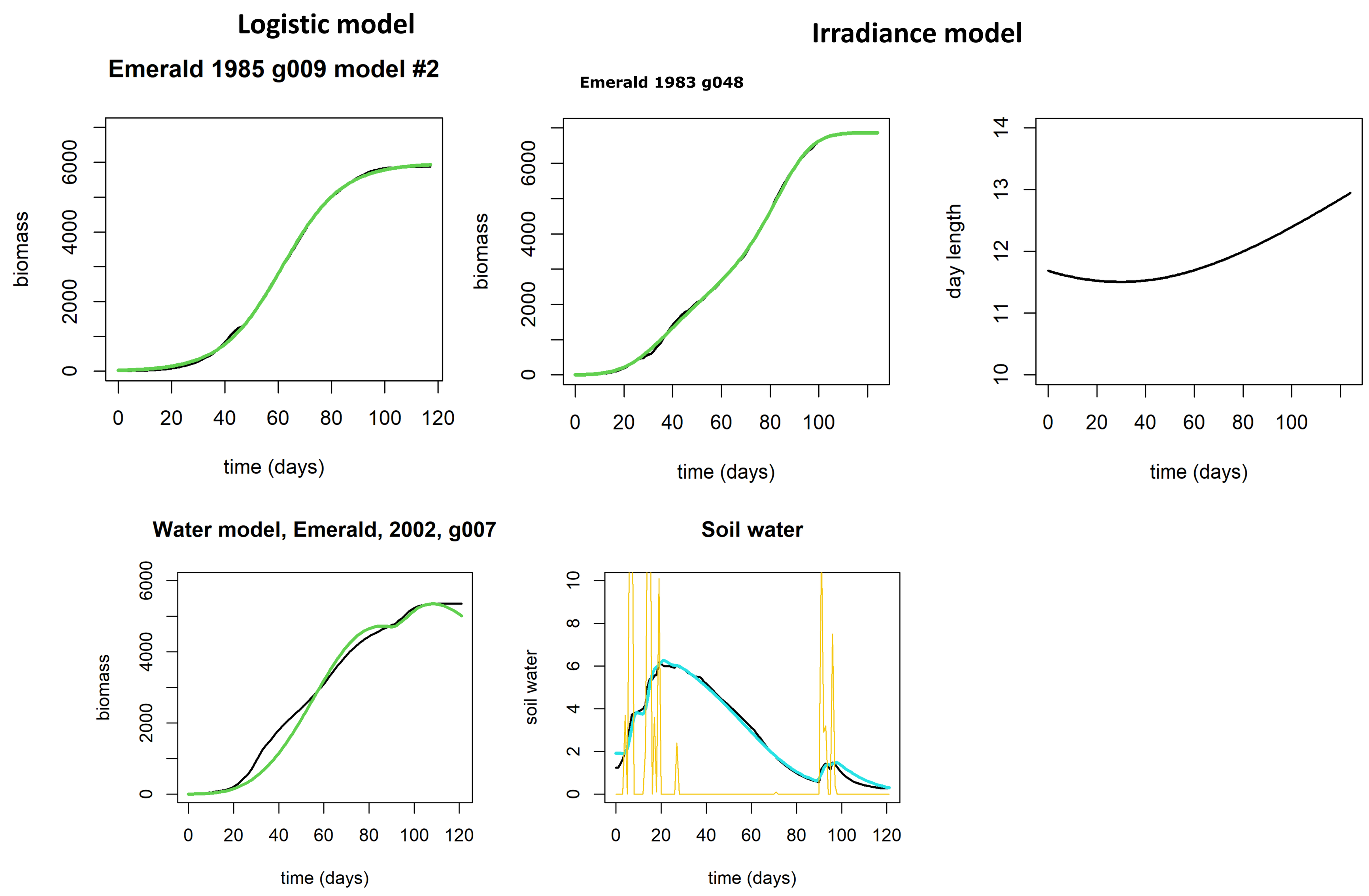
Main activities

- Low-complexity Differential Equation (DE) based crop growth models were fitted to time series data (in silico as well as experimental) of different genotypes in diferent environments.
- Different ML methodologies were tested for classification of time series data to identify critical growth processes with the inclusion of different types of noise.
- The model Tipstar for potatoes was coupled with Prosail (a canopy reflectance model) for crop disturbance classification with ML.

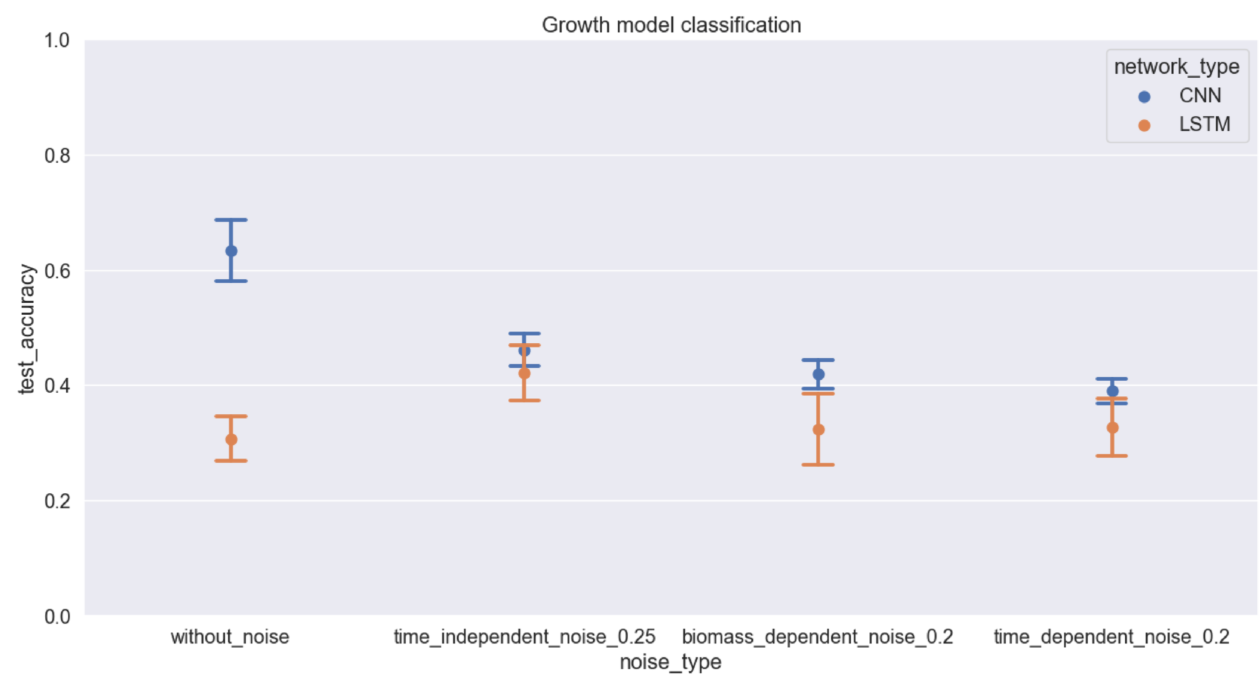


Above. Convolute Neural Network (CNN), one of the used methods for classifying time series.

Key results



Above. The best-fitting model differs per genotype and per environment.
This suggests there are different limitations for the various genotypes in these environments, including around climate-affected attributes such as water and temperature response.
These need to be included in future work. (Van Voorn et al., Frontiers in Plant Science 14, 2023)



Above. Classification results for LSTM and CNN models dealing with different sources of noise.
Models were trained with and without additional features from DE-based models.
The CNN with additional features outperformed the LSTM in the case of noiseless data.
Noise considerably affects correct classification by ML (courtesy of Yingjie Shao).

Next steps

Research will continue in

- KB DDHT2 programme, likely resulting in an app on Farmmaps;
- Follow-up D3-C2 project 'Hybrid Machine Learning process-based modelling approaches for climate adaptation strategies' focusing on applications of ML to assess climate adaptation measures in the agri-food value chain;
- Two PhD projects and several MSc topics on using ML for classification of time series and reconstruction of dynamic systems (such as GxI in crops) from such data.

Wageningen University & Research
George.vanVoorn@wur.nl
Biometris