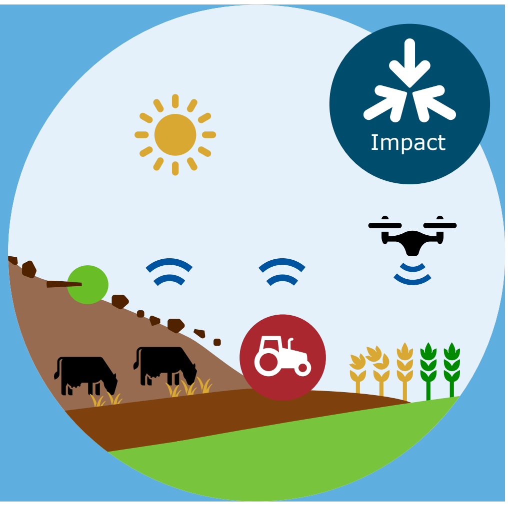
Machine learning for selecting crop varieties as climate adaptation measure

Selecting new crop genotypes to maintain food production

Emerging DS/AI methods

WAGENINGEN
UNIVERSITY & RESEARCH



Data Driven Discoveries in a changing climate (D3C2)

Objective: An actionable adaptation strategy to maintain food production is selecting new crop genotypes optimised for new climates. Essential in this strategy is a quantitative approach to forecast how genotypes will perform in these new environments. The aim of this project is to develop hybrid methods involving machine learning approaches and process-based dynamic modelling to improve our forecasting ability of how genotypes will perform in new environments.

Activities

The project provided important funding to invest time in trying out and developing methodologies. And, second, in collaborations with companies and other partners for obtaining data. Important partners were Corteva Agriscience, Gro Intelligence, Universidad Austral de Chile, University of Florida, ETH Zurich and the NPEC facility at WUR.

These data include high throughput phenotyping time series data of different genotypes of crops grown in different environments (locations, years). These and needed to train machine learning methods and identify relevant GxE (gene-by-environment) interactions in crops. In these collaborations (mostly via regular online Zoom meetings) we also developed candidate time-dynamic crop models and a Bayesian approach for the simultaneous fitting of crop models to multiple genotypes grown in different environments.

In addition, a crop growth model (Tipstar inside the Digital Future Farm) was coupled with a canopy reflectance model (PROSAIL) to be used for hybrid machine learning in crop disturbance classification. The crop model was used to generate in silico data for training a CNN (Convoluted Neural Network) which was subsequently fine-tuned with observational data. We were able to couple the two models using the high-performance cluster Anunna to simulate large datasets for machine learning training.

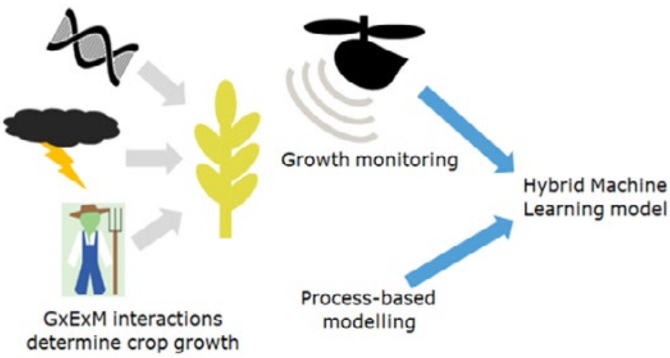
Achievement

The cross development between machine learning and domain knowledge is challenging. Machine learning methods generally are not developed for WUR domain applications such as crop growth. These methods are rarely exactly what is needed from the domain point-of-view. In particular, it is a challenge to get actual high-resolution time series data of crop traits and appropriate machine learning methods for analysing these data. This was a key topic in a workshop on 'From impact to solutions, data, data science and machine learning for climate adaptation' held at WUR on November 27–29, 2023, which we participated in. We also have been participating in meetings organised by the SciML (Scientific Machine Learning) group hosted by Joost Iwema and Dennis Walvoort at WUR. MSc students have been working on machine learning techniques for the identification and classification of crop growth models from time series data. We have not yet found methods that are always successful in obtaining valid crop growth models. Two PhD candidates have started in November

2023 on the development of hybrid machine learning methodologies. One of the two projects is specifically aimed at crop modelling methods.

Outlook

A follow-up project 'Identifying climate adaptation strategies for agri-food value chain actors using hybrid Machine Learning and process-based modelling approaches' was accepted for funding in the D3-C2 2024 call. It involves a collaboration between PSG, ESG, SSG, WECR, WENR, WFBR, WFSR, and WPR. The collaboration was organised in four work packages on climate adaptation strategies for different actors in the agri-food value chain to help alleviate negative impacts of climate change by applying machine learning techniques. Our research will also continue in the KB DDHT2 programme. If proven successful, it will eventually become an app on Farmmaps. Moreover, we are working on proposals involving several parties on hybrid GxE crop modelling.



Deliverables

- Van Voorn, Boer, et al. (2023). A conceptual framework for the dynamic modelling of time-resolved phenotypes for sets of genotype-environment-management combinations: a model library. Frontiers in Plant Science, 14, 1172359, doi.org/10.3389/fpls.2023.1172359
- https://git.wur.nl/maest001/ddht2_anomaly_detection/-/tree/main/synthetic?ref_type=heads. The folder synthetic contains the recipe to couple canopy reflectance model and crop growth model (inside DFF) using the hpc and singularity for containerization
- Kallenberg, Maestrini, et al., Integrating processed-based models and machine learning for crop yield prediction. Accepted after peer-review at the first workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023, https://edepot.wur.nl/634811

Lessons learned

Crop yield forecasting usually involves either theory-driven, process-based crop growth modelling or machine learning techniques based on data. Both approaches are data intensive and suffer from issues like equifinality, over-fitting, and lack of parameter and process identifiability. This project is intended to merge machine learning and process-based models in a 'best of both worlds' approach. In this approach, the hybrid forecasting models should outperform models solely based on machine learning or process-based models.

A previous paper (Maestrini et al., 2022, https://doi.org/10.1016/j.eja.2022.126569) listed three categories of hybrid approaches:

1. using crop models for engineering features for ML approaches;
2. estimating missing inputs for process-based models by using machine learning;
3. producing metamodels of process-based models by using machine learning methods.

An important lesson from this project has been that there are many machine learning methods but not tailored approaches for our domain applications. Time-dependent approaches differ from more classical approaches like image and speech analysis. They also have additional data requirements, namely time series of adequate coverage (here: traits, genotypes, environments). These series contain sufficient information about (auto) correlations. The collection of adequate data is equally challenging, as there are few data sets available with an adequate coverage.

There is limited data that explicitly links genotype to traits like biomass and yield. If they do, they rarely cover multiple genotypes. We have some promising results, but the development of suitable methodologies as well as data collection must continue. It is likely that a practical approach to data collection and usage continues to be needed, in which small data sets from different sources are combined for machine learning model training. One result we are proud of is that we nicely containerised the coupled models Tipstar and Prosail with a database on the HPC.

The collaboration in the team was fruitful and led to a follow-up, larger proposal that considerably expands the project team and ambition. The collaboration with parties from outside WUR takes time to develop. It helps to have regular meetings in smaller groups which a tangible goal, such as collaboratively working on actual candidate crop models.

Some other, more specific lessons learned include:
1 containerise your application to facilitate portability,
2 keeping updated with the latest version of models, costs time
3 generating synthetic datasets can be a useful way to use ML for overcoming a lack of experimental data and in the design of future field experiments.

Simulating canopies of crops at early stages (emergence to full ground coverage) was difficult because one of the assumptions of canopy reflectance models are not met (homogenous canopy cover). This issue needs to be further evaluated. The ambition of this project proved to be too much given the available funding and time-span of one year. Another issue that we intend to address in the follow-up project is the description of an appropriate heat response function in crop models.

Contact

George A.K. van Voorn
Functieomschrijving
george.vanvoorn@wur.nl

Wageningen University & Research
P.O. Box 47
6700 AB Wageningen
The netherlands
T +31 317 48 07 00
www.wur.eu