Integrating Machine Learning and mechanistic modeling towards adaptation of plants to climate change

Aalt D.J. van Dijk (aaltjan.vandijk@wur.nl), D. de Ridder (dick.deridder@wur.nl)

Background

The challenge we will address in this project is the development and seamless integration of mechanistic modeling and machine learning approaches to arrive at models for the response of plants to temperature change. As a demonstration, such models will be developed for response of Arabidopsis thaliana to elevated temperature. These models will combine quantitative, predictive power with the generation of detailed mechanistic insights. In this way, they enable us to extract actionable knowledge, delivering concrete opportunities for plant breeders to focus their efforts in breeding for better climate adaption. Models and machine learning approaches will be used to steer experimental efforts, critical to move from big to smart data approaches. Our goal is to develop our approach in a generalizable, modular and robust manner ensuring its suitability for future usage and extension.

Introduction

As a starting point for data-driven expansion of available mechanistic models (Fig. 1), an inventory of currently existing data and knowledge of Arabidopsis responses to temperature stresses will be generated, in particular related to expression levels of genes (Fig. 2). From these available experimental data, machine learning derived modules will be obtained (Fig. 3), which will be used as input for the mechanistic model.

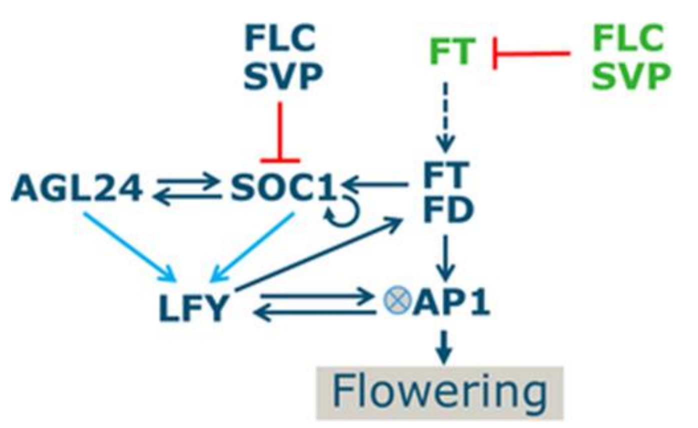

Figure 1. Available mechanistic model. This model uses a set of differential equations to describe how key genes regulate flowering time.

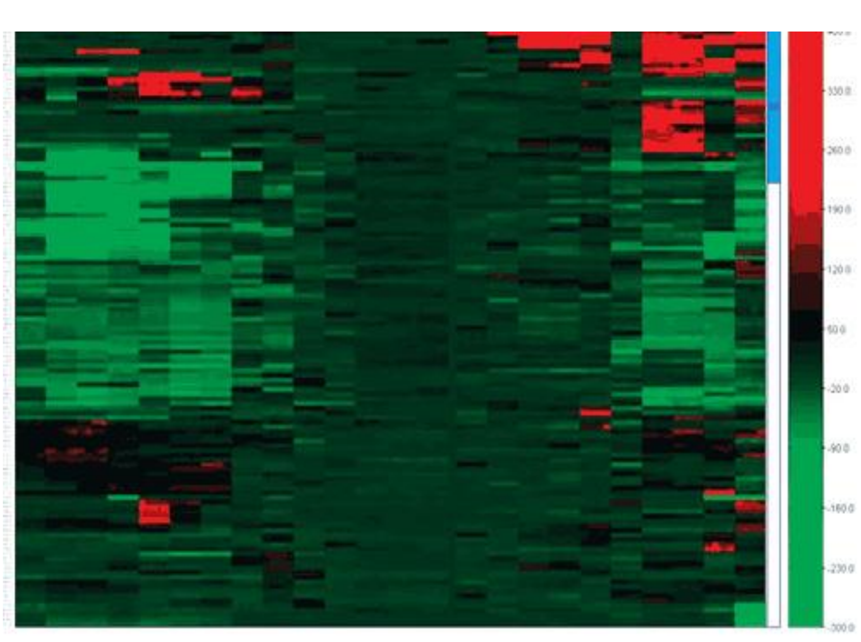

Figure 2. Example gene expression data. Rows represent different genes, columns different conditions (e.g. different temperatures).

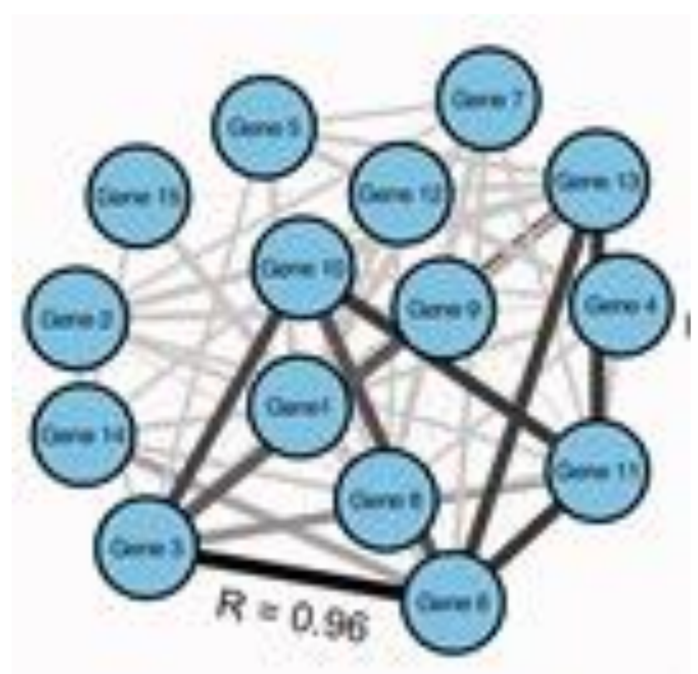

Figure 3. Module finding. Available large scale expression data will be used to define modules, key sets of genes involved in regulation of specific biological processes.

Methodology

Modules will be formulated as functions from experimental conditions, to a level of activity to serve as input in the mechanistic model, e.g. the level of a specific component. A subset of the inferred modules will then be selected for inclusion in the mechanistic model based on ML-based predicted association of module components with components of the mechanistic model (gene expression levels) or output of that model (flowering time).

Results

- Input dataset has been collected
- Differential equation model has been transferred from matlab to python in order to enable integration with pytorch
- Preliminary joint ML-mechanistic model has been implemented

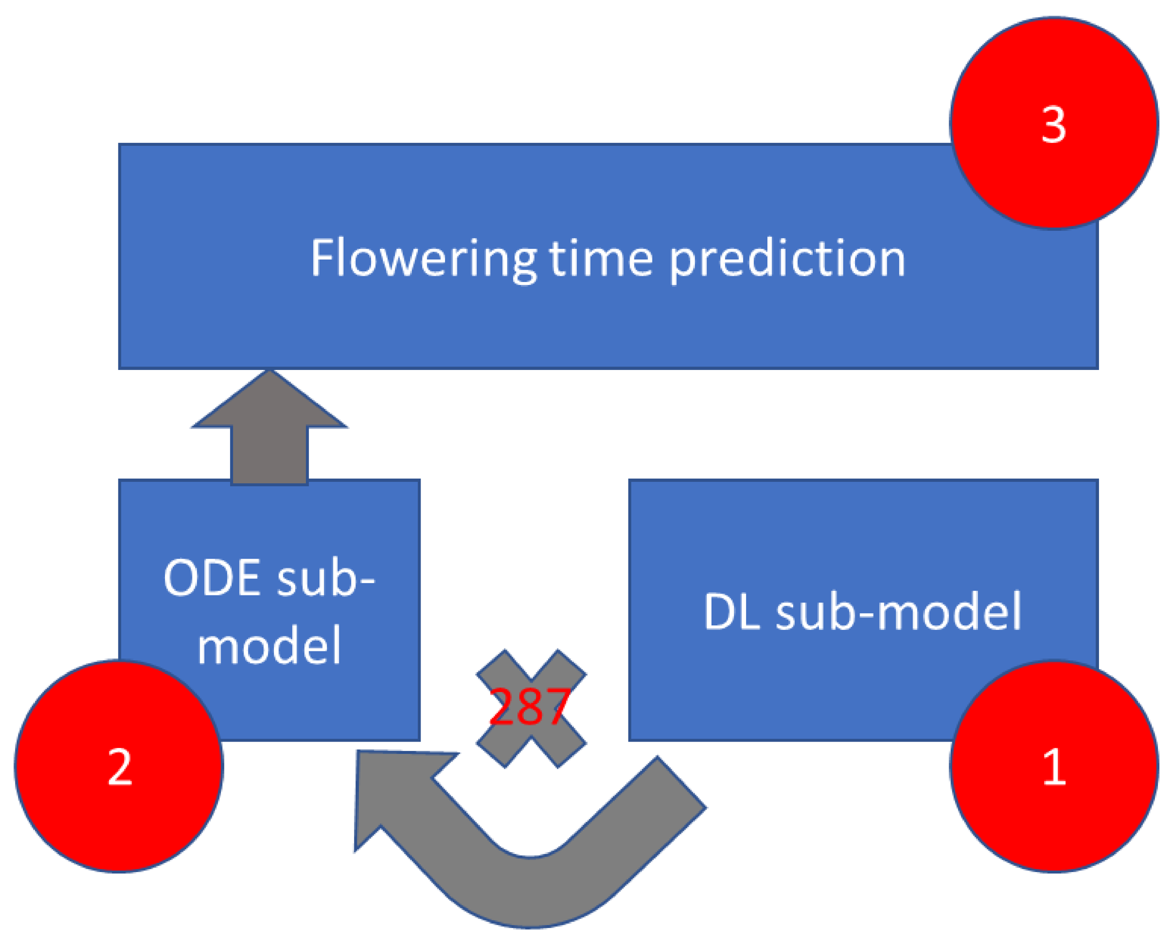

Figure 4. Preliminary version of model integrating differential equation and deep learning based model. Loss function containing three components is jointly optimized: **(1)** Deep learning regression model predicting the expression of a set of key genes based on expression level of other genes (cf. Fig. 2). **(2)** Prediction of expression levels in differential equation model. **(3)** Classification of 287 different mutants as early or late flowering, based on the effect that knocking out each of these genes has on the genes in the differential equation model.

Outlook

- Per 1 november, newly hired researcher will start working on the project
- Key next step will be to implement module finding
- Proposal has been sumitted at NWO (LTP programme) to continue working on the integration of mechanistic models with machine learning approaches towards prediction of plant resilicience

Conclusions

- Work in progress...
- But initial coupling between mechanistic model and machine learning model has been implemented
- Final model will enable predicting effect of temperature changes on flowering time

Wageningen University & Research
P.O. Box 123, 6700 AB Wageningen
Contact: name.surname@wur.nl
T + 31 (0)317 12 34 56, M +31 (0)6 12 34 56 78
www.wur.nl/xxxxx