Session 2: Machine Learning

The keynote speaker was **Jose Crossa**, Distinguished Scientist working at the Biometrics and Statistics Unit at the CIMMYT, with a talk about "Linking genomics, phenomics and enviromics for modern plant breeding".

learnMET, an R package to apply machine learning methods on multi-environment trials data – Cathy Westhues

The key note was followed up by the talk of **Cathy Westhues** about "learnMET, an R package to apply machine learning methods to multi-environment trials data". Analyzing large datasets from various sources (e.g. multi-environmental trials (MET) phenotypic data, genomic data, historical weather and soil data) with data-driven machine learning algorithms can represent a pertinent strategy to predict performance of plant genotypes grown under various environmental conditions. Here, we introduce the R-package learnMET, which provides a flexible framework to incorporate environmental characterization and to perform machine learning-based predictions of quantitative traits in MET. The package enables the processing of raw daily weather data, which can be directly given by the user. Alternatively, the package allows the retrieval of daily weather records from the NASA POWER database given geographic coordinates and planting and harvest date. The core functionality of the package is the possibility for the user to evaluate several types of machine learning approaches (gradient boosted trees, stacked ensemble models…) with respect to their predictive ability using different cross-validation schemes. The primary objective of this package is to enhance breeding decisions aiming at developing stable crop varieties in untested environmental conditions. Three toy datasets are incorporated in the package to illustrate the various functionalities on real datasets.

*[Text based on the abstract provided by the speaker]*

GitHub repository : https://github.com/cjubin/learnMET/
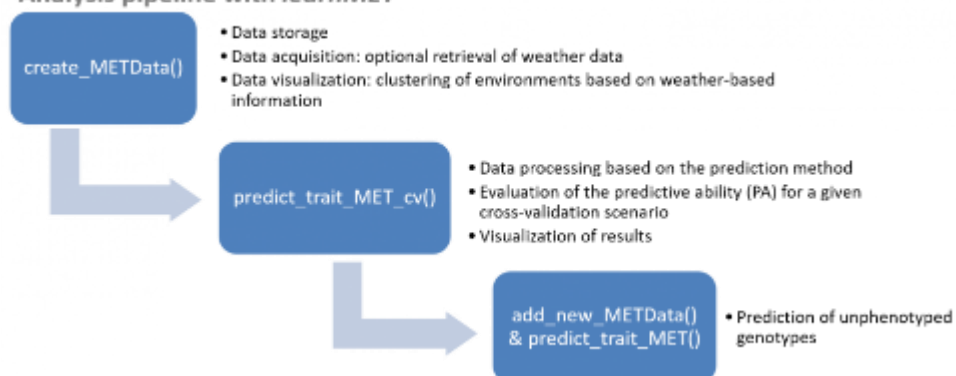
Available documentation: https://cjubin.github.io/learnMET/

In silico identification of the complex interplay between regulatory SNPs, transcription factors, and their related genes in Brassica napus L. using multi-omics data – Selina Klees

A further talk given by **Selina Klees** focused on regulatory SNPs (rSNPs), which defines a class of SNPs that are located within promoter regions and that have a high potential to affect the phenotype due to their impact on the binding of transcription factors (TFs). With the release of their recently published database of rSNPs for agricultural species (agReg-SNPdb), they opened the way for new studies analyzing, prioritizing, and interpreting SNPs and their effects on TF binding and, hence, gene expression.

In the presented study, they use a multi-omics approach by combining genomics,
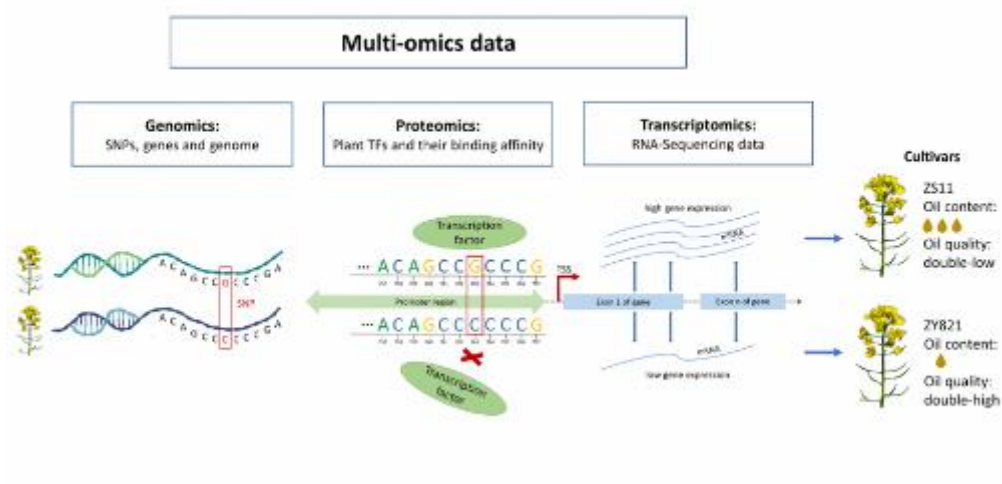
transcriptomics, and proteomics data of two different *Brassica napus* L. cultivars, namely Zhongshuang11 (ZS11) and Zhongyou821 (ZY821), with high and low oil content, respectively, to monitor the regulatory interplay between rSNPs, TFs and their corresponding differentially expressed genes in the tissues flower, leaf, stem, and root. They identified a total of 41,117 rSNPs in *B. napus*, and, by applying a random forest machine learning algorithm measuring their association with the phenotype, they selected 1141 oil content-associated rSNPs. Exemplary, they showed the complex interplay of the LACS gene regulation via DOF1 binding.

In this work, they provide the first genome-wide collection of rSNPs for *B. napus* and their impact on the regulation of gene expression in vegetative and floral tissues, which will be highly valuable for future studies on rSNPs and gene regulation.

[Text based on the abstract provided by the speaker]

Machine learning in a plant breeding company – three examples: phenoBob, BlueBob and BeetControl – Christian Hügel

In the presentation "Machine learning in a plant breeding company – three examples: phenoBob, BlueBob and BeetControl" **Christian Hügel** addressed the different applications of machine learning in a plant breeding company. He presented three examples where Strube is using machine learning:

PhenoBob is an autonomously navigating phenotyping robot used for the measurement of field emergence and early plant development in sugar beet microplots. The robot drives through the plots and records the position of every plant with RTK precision. These measurements are repeated multiple times from emergence until row closure, so plant count and early growth development is recorded over time. Every plant is classified with artificial intelligence into sugar beet or weed plant, to be able to evaluate the development of crop plants over time, and ignore weed plants.

BlueBob is an autonomously navigating field robot for mechanical weed management in sugar beet cultivation. The robot is based on the phenoBob and detects and classifies sugar beet and weed plants with artificial intelligence. In the crop rows weeds are removed with electrically driven active weeding tools which are moved around the crop plants, thus only removing the weeds. Additionally weeds between the rows are removed with conventional static weeding tools. With this approach a full field mechanical weed management is possible. This is Strubes contribution to making sugar beet cultivation more sustainable and environmentally friendly as an alternative to chemical weed management.

BeetControl is a smartphone app that is available in iOS and Android and is developed in cooperation with Strube D&S and Südzucker. The application will be able to recognize images of Cercospora via an artificial intelligence and then make a prediction via another artificial intelligence how the development of the disease will be within the next 14 days.

Within the smartphone APP, from 100 individual leaves a picture has to be taken in the field frontally. These leaves are then evaluated individually by the artificial intelligence and determined how many leaves are infested with Cercospora. Furthermore, in addition to the frequency of infestation, it is estimated how high the infestation intensity is. The advantage of this is monitoring warnings can be given earlier, because the infestation intensity also includes the different degree of infestation of the individual leaves.

Based on a threshold model applied in the respective country, the infestation frequency is determined time-specifically. The farmer receives an indication whether the threshold for treatment has been exceeded or not. The integration of the infestation level is intended to enable the farmer to be warned earlier than the current threshold values if the infestation intensity is significantly different.

[Text based on the abstract provided by the speaker]

Machine learning methods for interpretation of hyperspectral plant signatures – Dr. Stefan Paulus

The final talk of the session was given by **Dr. Paulus** about "Machine learning methods for interpretation of hyperspectral plant signatures". Since several years hyperspectral cameras are used to measure plants and to describe specific plant signatures. These cameras extend the recorded area compared to RGB cameras that image the area of visible light by the NIR area. Here, between 700-1000 nm, plants show an increased light reflectance. Thus changes in the plant state are supposed to be tracked here first.

Hyperspectral images often are reduced to two or three wavebands that form a vegetation index like NDVI or OSAVI. Nevertheless, for identification of a disease infection or the presence of stress the recorded hyperspectral images and thus the complete spectrum have to be analysed. Machine learning methods like support vector machines or decision trees use the direct spectral reflectance input to build a model. This model is able to classify infection classes or to estimate a disease rating scale. By analysing the importance of the single input wavebands linking back the technical spectral wavebands to the biological process is possible. In an experiment with the soil-based disease *Rhizoctonia solani* sugar beet has been inoculated and imaged using a hyperspectral camera in a time series. The spectrum has been analysed to differentiate between the inoculated group of plants and the non-inoculated control group.

It could be shown that the combination of hyperspectral cameras and machine learning is a useful tool for detection of a disease infection in sugar beet. The algorithm showed a high accuracy regardless that symptoms have not been visible even for human specialists.

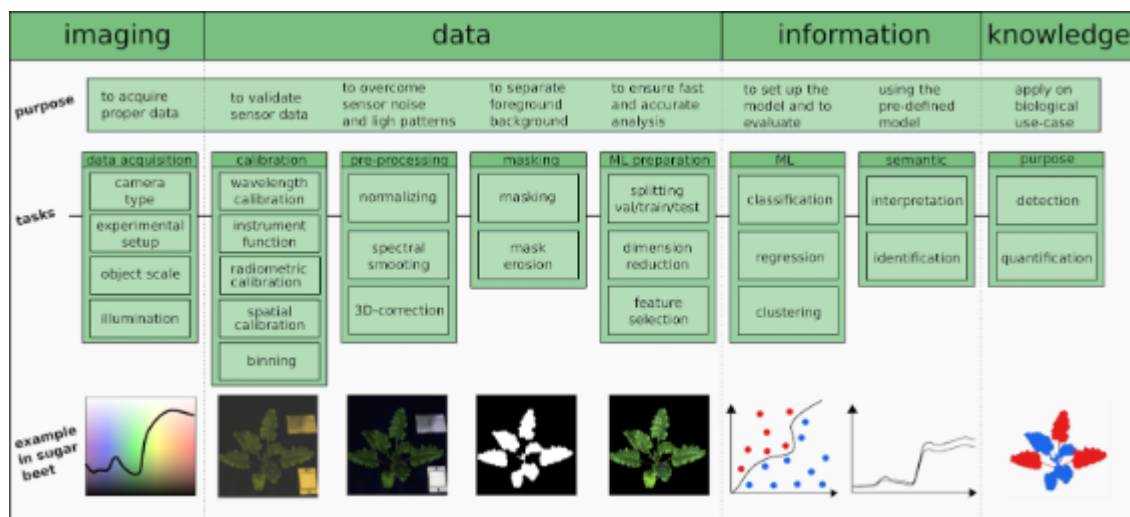*[Text based on the abstract provided by the speaker]*

Figure from Paulus *et al.* (2020). A generalization of a hyperspectral workflow.