

Advanced Course

STATISTICAL TOOLS FOR PLANT PHENOMIC DATA ANALYSIS

Zaragoza (Spain), 20-24 January 2020

1. Objective of the course

New high-throughput phenotyping techniques are changing plant sciences in general and plant breeding in particular. They produce huge volumes of data points through time requiring special statistical methods to extract meaningful information for plant breeding purposes.

The course introduces infrastructure needed for field and indoor platform phenomics. Then specific experimental designs and corresponding mixed models will be treated in detail together with spatial and longitudinal modelling. Statistical and machine learning techniques will be presented for pre-processing of phenomic data. Methodologies for the identification of the genetic basis of new phenotypic traits will be demonstrated. Finally, phenomic traits will be integrated in prediction models for yield. Examples and exercises will use real data from phenotyping platforms and field experiments.

At the end of the course participants will:

- Know the different platforms, sensors and carriers used for plant phenotyping.
- Be able to determine suitable experimental designs and perform analysis using mixed models.
- Be capable of correcting extracted plant features for spatial and temporal trends.
- Appreciate the potential of integrating secondary plant traits in genetic models for prediction of yield.
- Have acquired practical experience in applying statistical methods through the analysis of case studies and hands-on exercises.

2. Organization

The course is organized by the International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM), through the Mediterranean Agronomic Institute of Zaragoza

(IAMZ), with the collaboration of the European Plant Phenotyping Network (EPPN²⁰²⁰). The course will take place at the Mediterranean Agronomic Institute of Zaragoza and will be given by well qualified lecturers from universities, research centres and private companies in different countries.

The course will be held over a period of 1 week, from 20 to 24 January 2020, in morning and afternoon sessions.

3. Admission

This course is designed for 25 professionals with a university degree and with a background in plant breeding and biology who want a wider and integrated perspective in statistical data analysis and interpretation relevant to contemporary plant phenomics and genomics. Working knowledge on experimental design, analysis of variance and regression is required. Basic knowledge of R is expected. Familiarity with quantitative genetic theory and QTL mapping is also desirable.

Due to the limited number of available places, a selection will be made considering the interest of the training for the professional work carried out by the candidate.

English will be the working language of the course.

4. Registration

Candidates must apply online at the following address:
<http://www.admission.iamz.ciheam.org/en/>

Applications must include the curriculum vitae and copy of the supporting documents most related to the subject of the course.

The deadline for the submission of applications is 30 October 2019. The deadline may be extended for candidates not requiring a visa and not applying for a grant if there are free places available.

Applications from those candidates requiring authorization to attend the course may be accepted provisionally.



Registration fees for the course amount to 500 euro. This sum covers tuition fees only.

5. Scholarships

Candidates from CIHEAM member countries (Albania, Algeria, Egypt, France, Greece, Italy, Lebanon, Malta, Morocco, Portugal, Spain, Tunisia and Turkey) may apply for scholarships covering registration fees and full board accommodation.

Candidates from other countries who require financial support should apply directly to other national or international institutions.

6. Insurance

It is compulsory for participants to have medical insurance valid for Spain. Proof of insurance cover must be given at the beginning of the course. Those who so wish may participate in a collective insurance policy taken out by the IAMZ, upon payment of the stipulated sum.

7. Teaching organization

The course requires personal work and interaction among participants and with lecturers. The international characteristics of the course favour the exchange of experiences and points of view.

The course will be taught through a combination of lectures, case studies and supervised computer practicals.

8. Programme

1. Prediction, Prescription, Precision and Plant Phenotyping (1 hour lecture)
2. Introduction to phenomics (2 hours lectures)
 - 2.1. Indoor platforms
 - 2.2. Field phenomics including sensors and carriers
 - 2.3. Data processing pipeline for crop phenomics
 - 2.4. The European Plant Phenotyping Network and EMPHASIS

3. Choosing the design for field and platform experiments (3 hours lectures + 1 hour practicals)
 - 3.1. Fully replicated experiments
 - 3.2. Partially replicated experiments
 - 3.3. Design of multi-environment trials
 - 3.4. Practical work
4. Data collection and handling (1 hour lecture + 2 hours practicals)
 - 4.1. Introduction to feature extraction
 - 4.2. Case study and practical work on date-time formats and relational data tables
5. Mixed models analysis of extracted features (4 hours lectures + 4 hours practicals)
 - 5.1. Single experiment
 - 5.2. Multiple experiment
 - 5.3. Analysis of time series
 - 5.4. Case studies and practical work
 - 5.4.1. Correcting for spatial variation and temporal modelling
 - 5.4.2. Rerandomization in platforms
6. Statistical and machine learning techniques for feature extraction (4 hours lectures + 2 hours practicals)
 - 6.1. Dimension reduction and cluster analysis
 - 6.2. Penalized regression, classification and regression trees
 - 6.3. Deep learning
 - 6.4. Case study and practical work on feature extraction from hyperspectral canopy reflectance data using indices and multivariate analyses
7. Environmental data for modelling phenomic data (2 hour lectures + 2 hours practicals)
 - 7.1. Recording of environmental variates for plant phenotyping experiments
 - 7.2. Construction of environmental indexes
 - 7.3. Case study and practical work on modelling growth and developmental processes using environmental covariates
8. Integration of environmental, genomic and phenomic data (4 hours lectures + 4 hours practicals)
 - 8.1. QTL and association mapping
 - 8.2. Genomic prediction
 - 8.3. Use of crop growth modelling

GUEST LECTURERS

J. ANDEREGG, ETH Zürich (Switzerland)
J. BETRÁN, Bayer, Toulouse (France)
D. BUSTOS, Wageningen UR (The Netherlands)
J. HARTUNG, Univ. Hohenheim, Stuttgart (Germany)
A. HUND, ETH Zürich (Switzerland)
L. KRONENBERG, ETH Zürich (Switzerland)

E. MILLET, Wageningen UR (The Netherlands)
H. P. PIEPHO, University of Hohenheim, Stuttgart (Germany)
L. ROTH, ETH Zürich (Switzerland)
F. VAN EEUWIJK, Wageningen UR (The Netherlands)

