

Characterizing yield through wheat's perception of time: Phenomics, enviromics and genomics-based predictions

Lukas Roth ^{a,b} , Juan M. Herrera ^c , Lilia Levy Häner ^c , Didier Pellet ^c , Dario Fossati ^c ,
Mike Boss ^a , Xiaoran Chen ^d , Paraskevi Nousi ^d  and Michele Volpi ^d 

^aETH Zürich, Institute of Agricultural Sciences, Zürich, Switzerland

^bUniversity of Queensland, Gatton Campus, Australia

^cAgroscope, Nyon, Switzerland

^dETH Zürich and EPFL, Swiss Data Science Center, Switzerland

Email: lukas.roth@usys.ethz.ch

Abstract:

Understanding crop performance across diverse and shifting environmental conditions is essential for addressing global food security challenges. Crop production systems are governed by three primary components: plant genetics (G), environmental conditions (E), and management practices (M). In breeding and variety evaluation contexts, management is typically optimized and held constant, focusing attention on the interaction between genetics and environment, commonly expressed as $P = G \times E$, where P represents the observed phenotype.

Modern molecular techniques enable comprehensive genetic characterization at the genome scale, while publicly available weather data from governmental open data initiatives provide detailed environmental information. However, a fundamental mismatch exists between data scales: while genetic markers and environmental time series usually comprise $>100k$ data points, multi-environment trial (MET) datasets to characterize phenotypes typically contain fewer than $10k$ observations. This creates a classic high-dimensional, low-sample-size scenario that complicates system modelling.

Field phenotyping technologies offer a solution by generating time-resolved phenotypic measurements P_t , representing the cumulative effect of genetic and environmental interactions over time (t), $P_t = \int_0^t G \times E_\tau d\tau$. These technologies can produce datasets exceeding $150k$ data points (Roth et al. 2025a), aligning phenotypic data volume with genetic and environmental datasets to enable robust empirical modelling. Yet, specific modelling approaches that can handle such multi-modal data are yet sparse.

Here, we introduce Plant Time Bender (PTB) (Roth et al. 2025b), a deep learning framework that integrates image time series, genetic markers, and environmental variables to model genotype-specific physiological responses to temperature and vapor pressure deficit. PTB outperforms conventional genomic prediction approaches when predicting yield across 48 European year-locations in previously unobserved environments. The model successfully captures nonlinear growth dynamics that vary with developmental stages and reveals distinct response signatures linked to yield performance and stability.

These response patterns support both historical analysis and future yield projections, establishing a framework for site-specific variety selection and breeding programs. By unifying phenomic, genomic, and environmental data streams, our approach offers significant potential for advancing climate adaptation research in agriculture, particularly in addressing the challenge of predicting crop performance under novel environmental conditions.

Keywords: Deep learning, Field phenotyping, Genotype-environment interaction, Yield prediction

REFERENCES

- Roth, Lukas et al. (2025a). “The FIP 1.0 Data Set: Highly resolved annotated image time series of 4,000 wheat plots grown in 6 years”. In: *GigaScience* 14, g1af051. DOI: 10.1093/gigascience/g1af051.
- Roth, Lukas et al. (2025b). “The plant-time-bender model: predicting yield through wheat’s perception of time”. In: *bioRxiv*. DOI: 10.1101/2025.05.12.653430.