Enviromics for wheat VCU testing in Switzerland: some ideas and some attempts

Herrera J. M., Levy Hänner L., Visse M., Pellet D.

May 16, 2022

Introduction

Farmers seek optimal combinations based on genotypes and agronomic management that can consistently deliver outputs close to the potential of their on-farm environments, while minimizing the risk of uneconomical outcomes. An improved understanding of the environmental context for attainable crop performance can enhance on-farm resource use efficiency. Genotype (G) performance often varies across environments (E), leading to variance differences and rank changes among genotypes (Crossa et al., 2004). Environmental information can be applied to identify appropriate groupings of the environments sampled in multi-environment trials (METs) and to quantify their relationships to the target production environment and to assist interpretations of plant responses to the environments, G×E interactions.

The fact that crop performance is strongly influenced by the environment, made researchers consider weather variables to better explain (Heslot et al., 2014) or predict (Jarquín et al., 2014) genotypic performance. Prediction of the genotypic value of a candidate genotype in a specific environment is especially desirable for unobserved environments, i.e. environments for which genotypic data do not exist. However, many important environmental parameters that are required for the interpretation of experimental results and the outcomes of prediction models are not captured routinely in METs.

The objectives of this study were to i) to determine critical points for the development of an approach to recommend wheat genotypes registered in a national catalog to farmers, ii) identify suitable predictors for the environmental dimension of G×E interactions and ways that they can be captured routinely in multi-environment trials (METs).

Materials and methods

The datasets used in this study to train and validate a model to predict genotypic performance came from three wheat (Triticum aestivum) varieties trial networks (referred to as W40, W42, and W43, hereafter) designed to evaluate genotypes in the context of different production strategies. The experimental sites were distributed across the wheat main production area of Switzerland. Field trials within each site and year were arranged as lattice designs with three replications (i.e. there were three plots with the same genotype at each site). The genotypes included released varieties, advanced lines, lines evaluated for registration in Switzerland and lines that were bred in Europe and registered or submitted for registration in Switzerland. These genotypes changed throughout the duration of the study, except for a core group of six (W40), five (W42), and four (W43) genotypes.

To estimate environmental limiting factors we used an approach that determines how suitable is a site for wheat cropping (Holzkämper et al. 2014). Six climatic factors (minimum temperature, average temperature, maximum temperature, average photothermal quotient, water availability, and period length) represent major crop limitations in each of four phenological phases. The four phenological phases considered to estimate the limiting factors were: i) planting to 3-leaf stage; ii) 3-leaf stage to double ridge; iii) double ridge to anthesis; and iv) anthesis to physiological maturity.

We use ridge regression (RR)-BLUP (Endelman, 2011) to account for the high-dimensional nature of environmental effects on the grain yield:

$$\boldsymbol{y} = u\mathbf{1} + \mathbf{W}\boldsymbol{e} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is a y is a vector of predicted/adjusted grain yields of a variety, u is the population mean, $\mathbf{1}$ is a vector of ones, $\boldsymbol{e}$ is a vector of environmental limiting factors that is distributed as $N(0, \sigma_e^2\mathbf{K})$ where $\mathbf{K}$ is the kernel similarity matrix, $\mathbf{W}$ is the matrix that relates $\boldsymbol{e}$ to the adjusted grain yield, and $\boldsymbol{\varepsilon}$ is a vector of residual errors. The model was fitted by maximizing the restricted log-likelihood and using the exponential option, among the possible kernels.

Cross-validation was done by considering the prediction of the performance of genotypes in environments that were not included for model-building which is also referred as leave-one-site-out validation (Saint Pierre et al., 2016). Prediction accuracy of the RR-BLUP models was assessed by computing Pearson's correlation (r) between models' prediction and observed values.

Results

Table 1 shows average Pearson correlation coefficients and standard errors between observed and predicted grain yields according to a leave-one-site-out cross-validation scheme. Averaged across all sites and genotypes in W40, the adjusted coefficient of determination (R^2) and RMSE were 0.18 and 14.80, respectively. Corresponding values in W42 were 0.12 and 13.02. In W43, R^2 and RMSE were 0.12 and 11.54. The results demonstrated a satisfactory level of accuracy (average r=0.45) for grain yield predictions within W40. Levels of accuracy were lower for the other varieties trial networks W42 (r=0.27) and W43 (r=0.34). When the prediction accuracies resulting from the cross-validation were considered in terms of genotypes within W40 (Table 2), correlation coefficients ranged from 0.32 (CH Camedo) to 0.62 (Zinal) in W40.

Although the level of accuracy depends on the used dataset, the average prediction accuracies (r=0.45 for W40, r=0.27 for W42, and r=0.34) we obtained with a leave-one-site-out cross-validation scheme are within the range of the results reported in recent studies using the same cross-validation scheme; r=0.53 (Lopez-Cruz et al., 2015), r=0.51(Jarqín et al., 2014), r=0.30 (Heslot et al., 2014), and r=0.21-0.52 (Saint Pierre et al., 2016). Although, higher prediction accuracies would be desirable, such approaches materialize the opportunity

Table 1: Average correlations (r) and standard error of the mean (SE) between observed and predicted values according to a leave-one-site-out validation. Prediction of grain yield were obtained using data from all other sites and excluding that for the site for which correlations are presented.

Site	W40		W42		W43	
	r	SE	r	SE	r	SE
Assens	0.52	0.14	ni		ni	
Changins	0.27	0.19	0.23	0.28	0.28	0.21
Ellighausen	0.60	0.12	ni		ni	
Grangeneuve	0.53	0.13	0.21	0.28	0.32	0.29
Liebegg	ni		0.12	0.34	0.10	0.25
Lindau	0.75	0.09	0.51	0.15	0.86	0.06
Moudon	0.74	0.07	0.33	0.19	0.80	0.07
Neuhausen	ni		0.28	0.28	0.06	0.27
Portalban	0.09	0.22	ni		ni	
Riedholz	ni		0.29	0.28	-0.03	0.27
Sulzkünten	0.66	0.10	ni		ni	
Vouvry	0.01	0.24	ni		ni	
Zollikofen	0.34	0.18	0.20	0.22	0.37	0.19

ni=genotype not included in the network.

Table 2: Average correlations (r) and standard error of the mean (SE) between observed and predicted values according to a leave-one-site-out validation presented according to genotypes.

Genotype	W40		W42		W43	
	r	SE	r	SE	r	SE
Cambrena	0.53	0.11	0.26	0.18	ni	
CH Camedo	0.32	0.15	0.31	0.17	0.58	0.25
Forel	0.50	0.01	0.13	0.20	0.10	0.20
Levis	0.41	0.13	0.14	0.20	0.13	0.20
Montalto	0.56	0.11	ni		ni	
Zinal	0.62	0.09	0.22	0.18	0.11	0.20

ni=genotype not included in the network.

to deliver recommendations with a quantitative basis. The way environmental variables are defined in genotype prediction models is not trivial in for environmental classification. More details about the latter will be given in the presentation.

Keywords

Environmental characterization, Envirotyping, Enviromics, G×E interactions

References

Crossa J., Yang R.-C., Cornelius P. L. (2004). Studying crossover genotype×environment interaction using linear-bilinear models and mixed models. *Journal of Agricultural, Biological, and Environmental Statistics* **9**, 362–380.

Endelman J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* **4**, 250–255.

Heslot N., Akdemir D., Sorrells M. E., Jannink, J. L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics* **127**, 463–480.

Holzkämper A., Fossati D., Hiltbrunner J., Fuhrer J. (2014). Spatial and temporal trends in agro-climatic limitations to production potentials for grain maize and winter wheat in Switzerland. *Regional Environmental Change* **15**, 109–122.

Jarquín D., Crossa J., Lacaze X., Du Cheyron P., Daucourt J., Lorgeou J., Piraux F., Guerreiro L., Pérez P., Calus M., Burgueño, J., de Los Campos G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics* **127**, 595–607.

Lopez-Cruz M., Crossa J., Bonnett D., Dreisigacker S., Poland J., Jannink J. L., Singh R.P., Autrique E., de Los Campos G. (2015). Increased prediction accuracy in wheat breeding trials using a marker x environment interaction genomic selection model. *G3* **5**, 569–82.

Saint Pierre C., Burgueño J., Crossa J., Fuentes Dávila G., Figueroa López P., Solís Moya E., Ireta Moreno J., Hernández Muela V. M., Zamora Villa V. M., Vikram P., Mathews K., Sansaloni C., Sehgal D., Jarquín D., Wenzl P., Singh S. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific Reports* **6**.