
Etude de l'impact des covariables sur la qualité de prédiction des modèles d'interactions génotype x environnement

Vanessa Choisi^{*†1} and Célia Pontet^{*‡1}

¹Terres Inovia – TERRES INOVIA – Centre de recherche INRA de Toulouse, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France

Résumé

L'évolution des conditions climatiques et la diversification des pratiques culturales génèrent des conditions de culture de plus en plus hétérogènes, avec une plus grande expression des stress environnementaux. Dans ce contexte, il devient de plus en plus important de comprendre la variabilité du comportement des variétés face à ces stress afin de pouvoir mieux adapter le choix variétal au milieu de culture. Pour ce faire, il est nécessaire d'être capable de comprendre et de prédire les interactions entre génotype et milieu (IGE) dans les réseaux d'essais variétaux.

Des travaux sur l'analyse des IGE ont été réalisés. Différentes méthodes statistiques telles que la régression PLS ou les forêts aléatoires ont permis jusqu'à maintenant d'expliquer une part significative des IGE. Cependant, une évaluation des performances de prédiction de ces modèles montre au travers d'indicateurs divers qu'elles ne sont pas satisfaisantes. Or, pour que ces modèles soient utiles au préconisateur, il est nécessaire de pouvoir prédire le comportement des variétés face aux stress dans des environnements non expérimentés. On se propose donc dans cette étude d'évaluer la qualité prédictive des modèles statistiques d'IGE en s'intéressant particulièrement aux covariables utilisées pour mesurer les niveaux des stress dans les environnements.

Les données exploitées dans cette étude sont issues du réseau d'essais variété tournesol 2015. Il est composé d'une dizaine de lieux, sur lesquels sont cultivées une dizaine de variétés, réparties de manière aléatoire en 3 répétitions. Pour chaque variété sur un lieu et une répétition, on mesure le rendement, qui correspond au poids de graines de tournesol par surface récoltée.

Dans un premier temps, un modèle mixte est réalisé. Les termes d'interactions sont ensuite extraits de ce modèle pour être modélisés. Afin d'expliquer ces IGE, chaque environnement est caractérisé à l'aide de covariables quantifiant les stress qui se sont appliqués sur chacun d'eux (déficit ou excès d'eau, de températures, de rayonnement, d'azote).

On compare dans cette étude trois types de calcul de covariables :

1) Calculées sur les stades sensibles du cycle du tournesol : levée, végétation, floraison,

*Intervenant

†Auteur correspondant: v.choisi@terresinovia.fr

‡Auteur correspondant: c.pontet@terresinovia.fr

remplissage

- Les dates clé du cycle sont estimées à partir des sommes de températures nécessaires pour atteindre chaque stade. Les covariables calculées rendent compte surtout des stress climatiques
 - Le modèle de culture SUNFLO est utilisé afin de simuler l'évolution de la culture au cours de son cycle. Les dates clés sont estimées par le modèle de culture, qui donne également accès à des stress importants comme le stress azoté
- 2) Calculées de manière automatique, sur des périodes indépendantes des stades, par décades ou par mois.

Le terme " environnement " du modèle est alors décomposé en somme des covariables. La qualité prédictive de ce modèle sera évaluée grâce à l'analyse des erreurs de prédictions (procédure de validation croisée), résumée à l'aide de différents critères tels que le coefficient de Spearman ou la RMSEP (Root Mean Square Error of Prediction). Nous présenterons alors les résultats de cette évaluation en fonction des trois listes de covariables étudiées. Plusieurs méthodes statistiques seront également utilisées : régression PLS, régression Lasso, régression Ridge et forêts aléatoires.